Proceedings of the Fifth Mediterranean Morphology Meeting

On-line Proceedings of the
Fifth Mediterranean Morphology Meeting (MMM5)
Fréjus, 15-18 September 2005

Edited by G. Booij, L. Ducceschi, B. Fradin, E. Guevara, A. Ralli and S. Scalise

URL: http://mmm.lingue.unibo.it/

On-line Proceedings of the Mediterranean Morphology Meetings (MMM)
ISSN: 1826-7491

Contents

The Lexical Integrity Hypothesis in a New Theoretical Universe

Rochelle Lieber
University of New Hampshire

Sergio Scalise
University of Bologna

0. Introduction

Twenty five years ago, at the outset of generative morphology, the Lexical Integrity Hypothesis was a widely accepted part of the landscape for morphologists. The LIH, or the Lexicalist Hypothesis came in a number of different forms:

(1) Lapointe (1980:8) *Generalized Lexicalist Hypothesis*
No syntactic rule can refer to elements of morphological structure.

(2) Selkirk (1982:70) *Word Structure Autonomy Condition*
No deletion of movement transformations may involve categories of both W-structure and S-structure.

(3) Di Sciullo & Williams (1987:49) *The Atomicity Thesis*
Words are "atomic" at the level of phrasal syntax and phrasal semantics. The words have "features," or properties, but these features have no structure, and the relation of theses features to the internal composition of the word cannot be relevant in syntax – this is the thesis of the atomicity of words, or the lexical integrity hypothesis, or the strong lexicalist hypothesis (as in Lapointe 1980), or a version of the lexicalist hypothesis of Chomsky (1970), Williams (1978; 1978a), and numerous others.

Although there are slight differences in these formulations, they have a common effect of preventing syntactic rules from looking into and operating on the internal structure of words.

It should also be pointed out that at this early stage of discussion, a Weak Lexicalist Hypothesis and a Strong Lexicalist Hypothesis were often distinguished, the former merely stating that transformations could not look into word structure (i.e., derivation and compounding), the latter adding inflection to the domain of the LIH (Spencer 1991:73, 178–9).

Another possible statement of the earliest versions of LIH in terms of ordering can be found in Borer (1998:152–3):

(4) "The way in which LIH is enforced in many of these models is by assuming that the WF component, as a block of rules, is ordered with respect to the syntax. The WF component and the syntax thus interact only in one fixed point. Such ordering entails that the output of one system is the input to the other. This

G. Booij, *et al.* (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)*
Fréjus 15-18 September 2005, University of Bologna, 2007.
URL http://mmm.lingue.unibo.it/

> notion of the autonomy of the syntax and the WF component, and the restricted interaction between them, thus mimics the notion of autonomy developed for the interaction between the syntax and the phonology, where it is the output of the former which interacts with the latter."

We note that the earliest versions (Selkirk, both of Lapointe's versions, Borer's interpretation of the LIH) are of course stated in terms of the theoretical universe of the late seventies and early eighties. For all intents and purposes, syntax was a matter of ordered components, phrase structure rules and transformations, so to say that morphology and syntax did not interact was essentially to say that transformations could not look down into word structure and manipulate it. As Borer points out, if components are ordered and Morphology precedes Syntax, Lexical Integrity follows from the structure of the theory.

By the time we get to the mid-eighties, things are already a bit more complicated, with Government Binding / Principles and Parameters models expanding what we thought of as syntactic rules, principles, etc. Ordering is no longer a theoretical device, and by this time even the notion of what we mean by "component" is in question. In Bresnan and Mchombo's statement in the mid-nineties, all mention of transformations has of course been dropped. Nevertheless, the notion that words are unanalyzable units is still firmly entrenched.

(5) "A fundamental generalization that morphologists have traditionally maintained is the *lexical integrity principle*, which states that words are built out of different structural elements and by different principles of composition than syntactic phrases. Specifically, the morphological constituents of words are lexical and sublexical categories – stems and affixes – while the syntactic constituents of phrases have words as the minimal, unanalyzable units; and syntactic ordering principles do not apply to morphemic structures." (Bresnan & Mchombo 1995:181)

All of these statements of LIH still have in common that they assume a firewall between morphology and syntax, in whatever form syntax takes.

We should also mention that the No Phrase Constraint (NPC), proposed by Botha (1983), is related to the LIH, in that it prohibits root compounds from containing syntactic phrases. However, as we will see below, there is ample evidence that this sort of data exists, and is derived productively, at least in Germanic languages.

In fact the two principles LIH and NPC were intended to prevent intramodular interactions: the LIH preventing syntax from "looking into" morphology and the NPC preventing morphology from "looking into" syntax. Both directions of possible interaction were forbidden. This separation historically was well motivated. Its main effect was to delimit specific and consistent fields of research. Now that morphology is a well-established domain, however, this aprioristic separation cannot be maintained, at least not in the form it was proposed, because the data discovered and studied in the meantime have proved that it was too strong.

Even very recent statements of the LIH are subtly different, and as we will show, still appear to be not quite right. Given the formulation of LIH of Anderson: "the syntax neither manipulates nor has access to the internal structure of words" (Anderson 1992:84), Booij (2005) proposes to split the LIH in two parts:

(6) a. Syntax cannot manipulate internal structure of words.

b. Syntax cannot enter into the internal structure of words.

Spencer (2005: 81) also proposes a two-part statement of the LIH:

(7) *Revised Lexical Integrity*:
syntactic rules cannot alter the lexical meaning of words (including argument structure); syntactic rules have no access to the internal structure of X^0 categories.

The data we will review here suggest that while (6a) is true, (6b) is not, and that although the first clause of (7) may be correct (syntactic rules cannot alter the semantics of words), the second one cannot be (syntactic rules have no access to the internal structure of X^0 categories).

We have a number of reasons to think that this is a good time to reassess the LIH, and that even the most recent statements of the LIH are not quite right. The first is that over the years a number of sorts of data have been uncovered that call the LIH into question. We need to highlight what these data are, and just what kind of threat they pose to the LIH. The second is that the theoretical universe in which we find ourselves is vastly more complex these days. Not only are there a number of competing models of syntax to assess – all roughly within the generative rubric – but there are also a number of different models of morphology which embrace or deny the LIH to different extents. In this theoretical universe we have any number of ways in which we can understand the notions of syntactic rule and morphological rule.

We will first take up the nature of the problematic data, and then we will try to untangle this new theoretical universe. Both are of course tasks too big for this talk, but we will at least try to sort out some of the issues that arise within different possible frameworks. In the end, we will try to assess what the LIH amounts to a quarter century after its first statement.

Given the data that we will show, we think it's likely that we need to say that morphology and syntax can interact in both directions; we wish to take a more 'mature', less ideological look at data that have been amassed in this period of intense morphological research. We will show in what follows that we know that morphology and syntax interact, and that this interaction is not a one way affair: morphology sees syntax and syntax sees morphology. Nevertheless this two way interaction is highly constrained. Our goal, then is not so much to come up with a theoretical framework that will predict this interaction – although we will make a tentative suggestion – but rather, as closely as we can to assess what the scope and limits of a new LIH should be. We start by assessing some relevant data, and then proceed to assess theoretical possibilities and where they leave us.

1. Data

One difficulty we have at the outset is that the amount of data that could enter into this debate is potentially vast. In this talk we cannot hope to highlight all of it. We have therefore chosen some areas that we think are particularly important for our assessment, fully realizing that there might be other types of data that will again force us to change our minds. Specifically, we have chosen a number of cases that we consider to be very strong challenges to the LIH, as well as a few others that we feel to be quite revealing. We leave aside others which we feel are less convincing, among them cases like noun incorporation in various languages (e.g. Baker 1988 and much other literature) or resultative verbs in Chinese (Li 2005) which – although they may be consistent with syntactic analyses – yield nicely to morphological analyses as well (cf. Borer 1998:166). Similarly, we leave aside cases like English synthetic compounds or conversion verbs, which have been analyzed syntactically, but clearly need not be.

1.1. Phrasal Compounds in English

One kind of data that has offered a strong challenge to the LIH is the phrasal compound. Phrasal compounds (cf. Lieber 1992) are compounds headed by a noun with a phrasal non-head. They can occur quite freely in Germanic languages (and in fact most of the known data come from English, German, Dutch and Afrikaans). Consider the data in (8):

(8)

a *pipe and slipper* husband	a *floor of a birdcage* taste
over the fence gossip	*in a row* nests
a *slept all day* look	a *who's the boss* wink
God is dead theology	

Lieber argued in her 1992 book that phrasal compounds offer a strong argument in favor of abandoning the LIH entirely, but even she now thinks that that position is too strong. Nevertheless, we think that phrasal compounds still tell us a great deal about what is right about the LIH and where it is wrong.

One way in which theorists have sought to maintain the LIH in the face of this data is by explaining it away – in some sense denying that these are compounds, or that they are productively formed. As an illustration, let us give Bresnan and Mchombo's take on phrasal compounds.

While they have no trouble in acknowledging that these are indeed compounds, they argue that the phrasal constituent either is lexicalized or can be lexicalized on the spot:

(9) "In sum, we suggest that true phrasal recursivity is lacking in word structure. Where syntactic phrases appear to undergo morphological derivation, it is by virtue of their being lexicalized. Although lexicalization can be innovative, the non-syntactic status of lexicalized phrases embedded in word structure can be detected in properties such as lexical gaps, and can be confirmed by the other lexical integrity tests." (Bresnan & Mchombo 1995:194)

Bresnan & Mchombo (henceforth B&M) argue that phrases within compounds often have the flavor of quotations, and can include foreign phrases. Some phrases do seem quotative or contain a foreign phrase (or both!, e.g., *Ich bin ein Berliner speech*), but certainly not all of them, and as for the presence of foreign phrases or words, these can occur in syntactic collocations as well, as the literature on code switching shows us. B&M also point out that pronouns used within the phrasal part of the compound do not refer, as pronouns usually do – but again, this is not surprising, since it is a well-known characteristic of compounds that the non-head constituent of a compound has no independent reference (so in a compound like *catfood*, the stem *cat* cannot refer to any particular cat). Finally, they point to what they call "lexical gaps" citing forms that they've made up and asserting that they are much less plausible as phrasal compounds than other examples. We are skeptical of this argument as well, as it often turns out that words that seem odd to us out of context can be used given sufficient context.

The most problematic part of B&M's claim is that where phrases are not obviously fixed or quotative, they can be innovatively lexicalized. This is a truly odd claim: it amounts to saying that any phrase that is used in a compound is instantaneously listed (before the creation of the compound?), and therefore can qualify to be incorporated into the compound. It's hard to see what this claim amounts to. And in any case it requires a completely new view of what we mean by "lexicalization".

In contrast to B&M, Hohenhaus (1998) has argued that certain sorts of compounds – phrasal compounds among them – are largely *non-lexicalizable*. Non-lexicalizable formations, according to Hohenhaus, are ones that are context-dependent; often they are nonce formations. Further, they frequently occur only in a restricted realm – say, conversation rather than written forms. They occur on the fly. We are rarely aware of them. And as they tend to be conversational, we are possibly less aware of them than words we see in print.

Our conclusion is that the existence of phrasal compounds still offers a challenge to the LIH, or at least to some forms of it.

1.2. Conjunction in English Derivation and Compounding

Germanic languages have other expressions which might be analyzed as phrases within words, specifically conjunctions within synthetic compounds:

(10) a. a truck driver

b. a [[car and truck] driver]

As with the phrasal root compounds discussed above, they still seem to be compounds, and as such, they pose a problem for LIH.[1] Interestingly, the type of phrase that can occur as the non-head is far more limited than those that occur in phrasal root compounds. While conjunctions are permitted, it is not the case that just any NP can be the non-head. For example, it is not possible to have a non-head which is modified by and adjective, as in (11):

[1] Interestingly, Ackerman & LeSourd (1997) suggest that constructions of this sort in Hungarian do not contain phrasal constituents, but rather that conjunction can happen at any level, including the X^0 level. If so, we would remove this case from our list of challenges to the LIH.

(11) a [[red truck] driver]

We will return to what this means below.

Spencer (2005:82) (cf. also Strauss 1982), points out a similar problem for derivation in English, citing examples such as those in (12)

(12) a. pre- and even to some extent post-war (economics)

b. pro- as opposed to anti-war

c. hypo-but not hyper-glycemic

Spencer notes that the prefixes that allow this sort of conjunction are relatively limited (one can't conjoin *un-* and *re-*), for example, and considers using the traditional designation of "prefixoids" for them. For our purposes, it doesn't matter whether we call these prefixes, prefixoids, or something else (semi-words). The inevitable point seems to be that they constitute a clear violation of the LIH, as do cases of so-called *Gruppeninflection* or "suspended affixation" (Spencer 2005:83) which seem to constitute a similar phenomenon, albeit concerned with inflection rather than word formation.

1.3. Italian "trasporto latte"-type *Constructions*

Italian also has a sort of construction that poses a strong challenge to the LIH. Consider the following data:

(13)

arruolamento volontari	'volunteers enlistment'
produzione scarpe	'shoes production'
trasporto merci	'goods transportation'
asporto rifiuti	'litter removal'
elaborazione dati	'data processing'

We can subject these forms to some of the classic tests (cf. ten Hacken 1994) that distinguish lexical forms from syntactic ones. These tests are:

(14) i. conjunction
ii. wh-movement of the head and the non-head constituent
iii. non-head topicalization
iv. pronominal reference

(15) Test i.: conjunction:
a. **il trasporto passeggeri e il* ___ *merci sono fallimentari*
the transportation passengers and ___ goods are not convenient
in questa stagione
in this season
'passenger and freight transportation are not convenient in this season'

b. **la produzione carta e la* ___ *schede sono in piena attività*
the production paper and ___ cards are in full service
'paper and card production are in full service

c. **l' elaborazione dati e la* ___ *programmi richiedono tempi lunghi*
the elaboration data and ___ programs demand long times
'data and program elaboration demands long times'

(16) Test ii.: wh-movement of the head (a-b) and the non-head constituent (c-d)

a. **che cosa* ___ *carta è stata sospesa*?
what paper has been stopped?

b. **che cosa* ___ *passeggeri è efficiente*?
what passengers is efficient?'

c. **cosa produzione* ___ *è stata sospesa*?
what production has been stopped?'

d. **cosa trasporto* ___ *è efficiente*?
what transportation is efficient?'

(17) Test iii.: non-head topicalization

a. **carta, è stata sospesa la produzione* ___.
paper has been stopped the production

b. **passeggeri, è efficiente il trasporto* ___.
passengers, is efficient the transportation'

(18) Test iv.: pronominal reference

a. **il trasporto passeggeri$_i$ è efficiente, e noi li$_i$ conosciamo*
the transportation passengers is efficient, and we them know
'passenger transportation is good, and we know them'

b. **la produzione carta$_i$ ha avuto un notevole sviluppo, ma noi non*
la$_i$ compriamo
the production paper has had a good growth, but we do not
it buy
'paper production had a good growth but we do not buy it'

As illustrated, three of the four tests give ungrammaticality and we are therefore tempted to conclude that the constructions under examination are compounded words. But if we apply the test of insertion/modification, which is traditionally the main test of cohesiveness, we find the following situation:

(19) a. *produzione accurata scarpe*
production accurate shoes
'accurate shoe production'

b. *produzione* *scarpe* *estive*
production shoes summer
'summer shoe production'

c. *produzione* *accurata* *scarpe* *estive*
production accurate shoes summer
'accurate production summer shoes'

d. **produzione* *limitata* *nel tempo* *scarpe*
production limited in time shoes'
'a limited in time shoe production'

Items of this sort are thus transparent to insertion/modification (though with some restrictions, since insertion/modification seems to be limited to only one adjective). An adjective can modify the first noun (19a), the second one (19b) or both (19c). What is impossible, however, is the insertion of anything other than an Adjective. Even a parenthetical is not allowed (19d).

Summarizing, the test results given for the three different kinds of expressions are illustrated in the following table:

(20)

	PHRASES	COMP-LIKE PHRASES	COMPOUNDS
CONJUNCTION	+	–	–
WH-MOVEMENT	+	–	–
TOPICALISATION	+	–	–
INSERTION	+	+	–
PRON. REFERENCE	+	–	–

According to four of the five tests these constructions are more similar to compounds than to phrases. Nevertheless, our sense is that these are not phrasal compounds akin to the Germanic cases – the types of phrases that occur internal to the compound are restricted to ones in which either the head or the non-head noun is modified by an adjective. If anything, they are more akin to the English phrasal synthetic compounds, in which the non-head is highly restricted in its nature – in the English case, limited to conjoined Ns. Although we don't have the space here for a full analysis, we believe that these Italian compound-like structures would best be dealt with as a sort of construction in the sense of Booij (2005), involving a fixed template for the phrasal element, which is then down-graded to a word, following the schema in (21):

(21) $[[[N]\ [N(A)]]_{NP}]_N$ or $[[[N(A)]\ [N]]_{NP}]_N$

$[[[$produzione$]\ [$scarpe estive$]]_{NP}]_N$ $[[[$produzione accurata$]\ [$scarpe$]]_{NP}]_N$

production shoes summer production accurate shoes

We will return below to what this analysis means in terms of a theoretical model. To summarize to this point: it seems so far that there is good evidence from compounds and compound-like expressions in several languages which indicates interaction between morphology and syntax, specifically that morphology must be able to "see" into syntax or that syntax may "feed" morphology.

1.4. Phrasal Derivation in Various Languages

The jury is still out, however, on the possibility of derivation on phrasal bases. Certainly there is little support for it in Italian, where the phrases to which affixes can be attached really do seem to be lexicalized, cf. (22):

(22) a. menefreghismo 'I don't care-ism'
*cenefreghismo '*we don't care-ism'

b. cento metrista 'runner of the hundred meters'
*centoventimetrista '*runner of a hundred twenty meters'

As for English, it appears that phrasal derivation is at very best limited. There are certainly a few examples in which the suffix *-ish* seems to attach to a phrase. Spencer (2005:83), for example, says the following:

(23) "Another type of deviation from strict lexical integrity is found when an affix apparently attaches to a whole phrase, as in *a why-does-it-have-to-be-me-ish expression*. I am not aware of any serious study of such formations, and their status is unclear to me. A cursory internet search reveals large numbers of such coinings, though it also reveals that for some speakers *-ish* has become a free morpheme with roughly the meaning 'approximately'"

It's not hard to find fixed expressions with the affix *-ish* such as the one in (24a), but (24b) sounds less plausible, and our search of the internet turned up fairly few *-ish* forms on phrases (24c):

(24) a. old maidish

b. ?young girlish

c. self-sufficient-ish, $50M-ish, New Years Day-ish, Space Hulkish Report Don's Long-Awaited Photo Tutorial-ish Thing

If anything, it seems possible to attach *-ish* to names – for example *Shelly Lieber-ish* doesn't sound all that bad, or even *Lieber and Scalise-ish* if it were known, for example, that we were in the habit of working together. But it's hard to document productivity on real phrases in this case. At best we would have to say, with Spencer, that not enough work has been done.

As for other languages, Ackema & Neeleman (2004:11) mention a case in Quechua where a particular affix attaches to phrases to nominalize them. This case is

perhaps a promising example of derivation on a phrasal base, but as yet we know too little about it to assess how much of a threat it poses to the LIH.

Finally, it is less difficult to find examples where inflections seem to attach to the outsides of phrases, for example, the English genitive. In any case, to the extent that phrasal derivation (or inflection) is attested, we will assume that it has a structure like that in (25):

(25) $[[YP]\ X]_X$ *or* $[X\ [YP]]_X$ *or* $[[YP]\ X]_Y$ *or* $[X\ [YP]]_Y$

In the first two cases, we assume that X is the head, and in the last two that X is the non-head.

1.5. Kageyama's Word Plus (W+) Data

Another set of data that poses a challenge to LIH can be found in the work of Kageyama (2001). The so called Word-Plus cases are complex phrase-like expressions involving prefixation and compounding, or a mixture of both which show properties of both morphological and syntactic constructs:

(26) *Morphological properties of W+*:
- syntactic atomicity – lexical integrity (neither modification nor deletion of constituents)
- absence of case marking

Syntactic properties of W+:
- syntactic internal semantic relations, sentence level anaphora, phrase-like accent contour
- internal pauses indicating constituent boundaries

(27) a. *zén* | *gaimu- dáizin* — *hán* | *taiséi*
ex- foreign minister — anti- establishment
'ex-foreign minister' — 'anti-establishment'

b. *Yomiuri-Kyozingun* | *toosyu* — *booeki- gaisya* | *syatyoo*
Yomiuri Giants pitcher — trading-company president
'a pitcher of the Yomiuri Giants' — 'a president of a trading company'

c. *Yomiuri-Kyozingun-no toosyu* — *booeki-gaisya-no syatyoo*
Yomiuri Giants-GEN pitcher — trading-comp.-GEN president

In (27a) we find examples of W+ with prefixes (note that here the prefix bears a main stress of its own, and it is followed by a pause, indicated by " | "); in (27b) we find examples of compound W+, which, in contrast with the corresponding phrases in (27c) lack case marking and are subject to syntactic atomicity.

Kageyama proposes to treat these contradictory expressions as a new category of words, Word Plus (W+), which are larger than words (X^0) but nonetheless constitute a morphological object distinct from phrases (X'). The category W+, rather than X^0, is the maximal projection of morphological structure. We suspect, however, that the former

two cases might plausibly be analyzed as prefixation (27a) and compounding (27b) on a phrasal base, along the lines shown in (25).

1.6. Scope in Spanish Prefixation

Another sort of data challenging LIH has been noted for Spanish prefixation (Rainer & Varela 1992; Felíu & Fábregas 2003; Kornfeld & Saab 2003):

(28) a. *el* [***ex***- [*futbolista del Barça*]$_{NP}$]
the ex- footballer of Barça
'the former Barça footballer'

a'. *comisión* [***pro***- [*legalización de las drogas*]$_{NP}$]
committee pro- legalization of the drugs
'pro- drug-legalization committee'

b. ***inter****-comunicación departamental*
(argumental reading) 'reciprocal communication between departments'
(locative reading) 'communication between people in a department'

b'. *comunicación* ***inter****-departamental*
(argumental reading) 'reciprocal communication between departments'

Work on these affixes has noted two types of phrasal scope: as shown in (28a-a') although phonologically prefixation takes place on the N head of an NP, semantically the prefix affects the whole NP. It also appears that when the prefix *inter-* is attached to a predicate's argument (cf. (28b') vs. (28b)), it has obligatory semantic scope over the whole NP. Similar data can be found in English:

(29) a. postinfluenzal, premeditiation
[[post-influenza]al]

b. post dog show coffee
[[post [dog show]] coffee]

c. post digestive disorder complications
[[post[digestive disorder]]complications]

d. pre and post FDR era
[[pre and post] [FDR era]]

e. pre and post synaptic dopamine function
[[[pre and post]synapt]ic]

(30) my ex-car

In (29a), *post-* seems to have scope over the stem *influenza*; *postinfluenzal* means 'pertaining to the period after influenza'. In (29b), *post-* has scope over the compound *dog show*. In (29c) it appears to have scope over an entire phrase *digestive disorder*. Further, note that in (29d) the conjunction of *pre and post* has scope over a compound and in (29e) it has scope over the stem *synapse*. We have already mentioned the challenge to the LIH posed by conjoined prefixes. The fact that such prefixes (or whatever we call them) can take scope over phrases further suggests strongly that there is interaction – in this case between morphology and semantics.

The behavior of the prefix *ex-* also exhibits scopal properties in English, and strengthens the conclusion we draw from *pre-* and *post-*: *ex-* can take scope over a possessive pronoun rather than the stem to which it attaches. So, for example, in (30), in the phrase *my ex-car* the car in question is still a car, but it is no longer mine.

All this data suggests that we cannot build an absolute firewall between morphology and syntax. Neither, however, do we want to allow free interaction between morphology and syntax or to say that morphology can be reduced to syntax. We therefore need to consider what our options are.

1.7. Sublexical Co-reference

Another area in which there has been dispute about lexical integrity concerns coreference into complex words, the so-called Anaphoric Islands type of data. Lieber (1992) argues that this data also falsifies the LIH, at least if one speaks a dialect in which coreference is allowed in sentences like (31):

(31) Clinton_{i}-ites no longer believe in $\text{him}_{i,j}$.

Of course, there has been long dispute over both the judgments and the analyses of this case, so we will not go further into it here. We note simply that further investigation is needed on various factors which seem to influence judgments, including differences between derivation and compounding, the type of syntactic construction involved, the typology of the language in question, the productivity of forms, and so on. This is just to point out the scope of the data that might need to be accounted for.

2. Theory

One of the things that makes a reassessment of the LIH so challenging at this point is that – as we have noted above- we have available to us not only multiple theories of syntax to consider, but also multiple theories of word formation. It is impossible to reassess the LIH without considering a multitude of possibilities.

As an example, certain syntactic theories virtually force an abandonment of the LIH. One of these, as we see it, is Minimalism. Minimalism (where what we mean is the strict Minimalism of Chomsky 1995) requires words to come fully inflected from the lexicon. As Marantz (1995:379) expresses one view of minimalism:

(32) "Under the MP, all inflected words are formed in the lexicon. The question arises, then, whether all explanations of the distribution of morphemes within words should be left to whatever principles govern affixation in the lexicon. In addition, Chomsky explicitly limits the computational system to the path from lexical resources to LF. If any operations on tree structures occur between Spell-Out and PF, they would seem to fall outside the mechanisms, if not the principles, characteristic of the computations in the syntax proper."

Chomsky himself has varied on the location of morphology, at various points (1995:133) claiming that inflection is syntactic, but more recently (1998) maintaining that inflection is in the lexicon. In his (1995) book, Chomsky says in an offhand way that derivation is probably in the lexicon. Nevertheless, this claim is at odds with what Chomsky claims elsewhere. Specifically, if computation is restricted to the syntax proper, and we assume that some sort of computation is involved in the generation of derived words or compounds, it appears that derivation and compounding must be syntactic. Minimalism is incompatible with the LIH, then.[2]

Distributed Morphology (DM) is a Minimalist-sympathetic framework that rejects the LIH outright (cf. Harley & Noyer 1999), but rejects as well the notion that forms are pulled fully inflected from "the lexicon." Within DM, inflection and derivation are separated from what Halle & Marantz (1993) call Vocabulary Insertion. The former involves the manipulation of features within a syntactic tree, with Merge and Move as possible operations. The latter is a matter of Spell-out, which happens late in the derivation (Marantz 1995:379):

(33) "In proposing and detailing the theory of 'Distributed Morphology,' Halle and Marantz (1992) suggest that functional heads like Tense and AGR should in fact serve as the locus of lexical ('Vocabulary') insertion and that all word formation should occur in the syntax, as a result of the syntactic combination of heads. In addition, we argue that the operations taking place in a derivation between Spell-Out and PF are of the same sort and obey the same principles as the operations in the rest of the syntax."

Our chief objection to DM is that it is untested with respect to the kind of data we have outlined in this talk. As Spencer (2005) points out, much of the literature on DM is unpublished, or available only from websites, and we would add that very little of what

[2] Note too that this turns on its head a weaker version of LIH advocated by Anderson in which derivation and compounding are lexical, but inflection is syntactic!

is published is explicitly concerned with derivation and compounding, as opposed to inflection. Of course, proponents of DM are free to explicitly reject the LIH, but in doing so they must explain why the data for syntax/morphology interaction appears to be so limited. In frameworks such as DM we should expect a much freer interaction between word formation and syntax than we actually find.

Other syntacticians with clear Minimalist sympathies continue to embrace the LIH, or at least to accept it in a modified form, however. We have in mind here the recent work of Yafei Li (2005). Li allows some minimal interaction between morphology and syntax, proposing a principle which he calls *The Morphology-Syntax Mapping Hypothesis (MSMH)*:

(34) Li (2005:4) *The Morphology-Syntax Mapping Hypothesis* (MSMH)
If morphological components X and Y are in a word W and there is a relation R between X and Y, then R is reflected in syntax if and only if:
a. R is thematic, and
b. the representation of R in syntax obeys all syntactic principles.

In effect, words are invisible to syntax unless there is a thematic relation expressed sublexically. Things like noun incorporation and the formation of applicative verbs are therefore legitimate fodder for the syntax. Li goes into great detail on the ways in which such constructions obey syntactic principles, but he is conspicuously silent on just what the morphological component looks like, what form morphological rules take, and indeed whether there is actual computation involved in non-syntactic morphology. Significantly, his proposal for the limited interaction between morphology and syntax does not help us with some of the data sets that we outlined above.

For example, with respect to phrasal compounds, since there is no thematic relation between the phrase that forms the first element of the compound and the second stem, we must assume that Li's theory would still prevent syntax-morphology interaction. This leaves open what sort of derivation we could propose for phrasal compounds within his theory.

Also roughly within the realm of Minimalist Syntax falls the work of Ackema & Neeleman (2004)(henceforth A&N). However, their work is much more explicit about the contents of the morphological component (or subcomponent) and the nature of the interaction between morphology and syntax.

A&N (elaborating a proposal put forth by Jackendoff 1997:39) propose that the grammar is constituted by three modules (syntax, semantics and phonology) but that each contains "a submodule that generates phrasal representations and a submodule that generates word-level representations" (2004:3). The main idea is that morphology is a "set of submodels within these bigger modules" (2004:6).

Morphology and syntax can thus share common principles – for example, a vocabulary of features and a process of merger -- but they can at the same time be based on different principles. Nevertheless, A&N argue that there can be a number of different types of intramodular interactions between morphology and syntax: first, words and sentences consist of a certain amount of shared vocabulary (certain features, the notion of Merge, etc.); second, word syntax and phrasal syntax are in competition (2004:9); and finally, the process of insertion works both ways between morphology and syntax: words can of course be inserted into syntactic structures, but it is also possible for phrases to be inserted into words (2004:10).

A&N maintain that the LIH is basically correct (2004:109). Their arguments are largely theory internal, in the sense that the structure of complex words – according to them – is invisible to syntax because syntax builds up the "host structure" and morphological complex words are inserted into this structure. But principles operative in the host structure are insensitive to the structure of words. For example, they claim that the head of a complex predicate cannot itself be complex (35a) (what they refer to as a "complexity constraint"), but this principle is not valid for morphological complex words (35b) (2004:33–4):

(35) a.

dat	Jan	en	Piet	[samen werken]	
that	Jan	and	Piet	together work	
dat	Jan	en	Piet	het voorstel	[uit werken]
that	Jan	and	Piet	the proposal	elaborate
*dat	Jan	en	Piet	het voorstel	[uit [samen werken]]
that	Jan	and	Piet	he proposal	out together work

b.

dat	Jan	de	foto's	[ver groot]
that	Jan	the	pictures	enlarge
dat	Jan	de	foto's	[uit [ver groot]]
that	Jan	the	pictures	up enlarge

A&N therefore make a good case for differentiating between syntactic and morphological complex heads. Nevertheless, it is not clear to us that their treatment either of insertion or of competition is sufficiently restrictive. For example, with respect to insertion, their analysis of phrasal compounds seems to us to have a number of problems. A&N's rule for root compounding is that in (36)

(36) Structure for compounds (A&N 2004:80)
$<_M \alpha \beta >$

This basically says to merge two objects and count them as a morphological object. Presumably this means that these can be any two things – lexical or phrasal. A&N also seem to allow the notion of feature matching between a lexical node in a compound and a phrase that is inserted into that compound. But given the range of phrases that can occur in phrasal compounds, it's hard to see how this would work. For example, phrasal compounds allow PPs and CPs in the non-head position. In the former case, we would have to allow a sort of compounding that is all but ruled out in English. In the latter case, it's unclear what would be matched at all.

As for securing adequate restrictions on the insertion of phrases as the base of affixation, A&N rely on their notion of distributed selection. If an AFFIX (a morpho-syntactic affix) corresponds to a phonological word, then it should be able to attach to any phrase, and not just to the head of the phrase. If an AFFIX does not correspond to a phonological word, it will only be able to attach to the head of the phrase. There are two problems with this. First, A&N give us no criteria for determining what AFFIXES count as phonological words – certainly an affix like *-hood* should in English. Yet *-hood*

certainly does not allow a phrasal base. Further we have the impression that for AFFIXES that are not phonological words, A&N's theory still overgenerates wildly. It should be possible on their story to generate words like those in (37).

(37) [[untrained military]ize] [[highly humid] ify]

It is also not clear to us that A&N's notion of competition is the right way to breach the firewall between morphology and syntax. According to A&N, morphology and syntax are in competition, and all other things being equal, syntax takes precedence. This means that, for example, the verbal compound *truckdrive* is blocked by the existence of a syntactic phrase *to drive trucks*.

Our problem with A&N's notion of competition is that it drives them to a rather odd analysis of derivation and compounding in English. A&N argue that derivation – i.e., word formation by means of affixes – is lexical because the semantic effect of affixation is unpredictable and because derivation does not always preserve the argument structure of the base. Root compounds are derived in the lexicon because the semantic effect of root compounding is also indeterminate. From this we can conclude that one criterion for a word formation process being lexical is semantic unpredictability.

Oddly, however, A&N then argue that synthetic compounds are also derived lexically – they must be, in their system, because they involve the addition of an affix to a [NV] compound, even though the semantic effect is utterly predictable. Further, we can point to various affixation processes which are semantically utterly regular and which preserve argument structure (*-er* affixation in English for example). Given the existence of such affixes, it appears that the alignment of semantic irregularity with morphological derivation and semantic regularity with syntactic derivation is an arbitrary move on A&N's part, without which the notion of competition would not work. We are forced to conclude that although A&N make an excellent attempt to rework the LIH in such a fashion that it makes sense of some of the examples that have plagued morphologists over the last twenty years, they still have not arrived at the right restrictions.

Of course, not all syntacticians embrace the Minimalist paradigm – far from it. And it's clear that our discussion of the morphology/syntax interaction now needs to be attentive to these theories as well. One theory that comes to mind is that of Construction Grammar.

Goldberg (1995) argues that the trend in G-B/P&P and later Minimalist syntax to consider constructions an epiphenomenon of the interaction of more general rules and principles attributes too much semantic weight to individual lexical items. In particular, Goldberg argues that the traditional notion of construction is still an important one, and that individual constructions are associated with a meaning that is independent of particular lexical items. For example, the intrinsic meaning of the "caused motion" construction allows us to extend the use of a non-motion verb like *sneeze*, as in (34) (Goldberg 1995:3):

(38) Pat sneezed the napkin off the table.

That is, the verb *sneeze* itself has no motional component in its meaning. This part of the meaning of the sentence must be attributed to the construction itself.

Constructions need not only be syntactic objects, however. Goldberg in fact recognizes no firewall between syntax and morphology (1995:7):

(39) "In Construction Grammar, no strict division is assumed between the lexicon and syntax. Lexical constructions and syntactic constructions differ in internal complexity, and also in the extent to which phonological form is specified, but both lexical and syntactic constructions are essentially the same type of declaratively represented data structure: both pair form with meaning. It is not the case, however, that in rejecting a strict division, Construction Grammar denies the existence of any distinctly morphological or syntactic constraints (or constructions). Rather it is claimed that there are basic commonalities between the two types of constructions, and moreover, that there are cases, such as verb-particle combinations, that blur the boundary."

It is not entirely clear what constitute particular morphological as opposed to syntactic constraints, as Goldberg is not concerned with morphology in her book, but it seems clear that Construction Grammar pushes us towards at least a partial rejection of the LIH.

Goldberg's ideas have recently been extended to morphology by Booij. In particular in a recent article, Booij (2005) argues that there are constructions in Dutch which involve both an idiosyncratic syntactic element and an idiosyncratic morphological element. One such construction is the Quantifier Adjective-s construction, as illustrated in (40):

(40) $[Q\ldots[A\text{-}s]_N\ ]_{NP}$

iets leuks	'something nice'
niets moois	'nothing beautiful'
wat zoets	'something sweet'

Booij points out that there are a number of odd things about this construction. First, there are a limited number of quantifiers that can occur in it. Second, the suffix *-s* does not normally attach to adjectives. Booij argues that in this case the *-s* is a category-changing inflection. The fact that the construction seems to have both idiosyncratic syntactic aspects and idiosyncratic morphological aspects therefore calls into question the LIH, at least insofar as it rules out syntax having access to word internal structure. (Booij still agrees that the LIH should exclude syntax overtly manipulating word internal structure – for example movement).

A major issue that remains unresolved in Construction Morphology, however, is the issue of computation or generativity. Goldberg is fairly explicit that constructions are created and learned on the basis of analogy (1995: 70–2, 123), rather than generated. We would assume the same to be true of morphological constructions. In contrast to this, generative syntax is based on the notion of computation, rather than analogy, and computation involves some sort of notion of 'rule'.

Similarly, generative morphology embraces the notion of rules, at least for highly productive morphological processes. There may be and have been fierce

disagreements over the formal nature of these rules, but the basic idea is that both syntax and morphology involve rule-based computation, some version of what has come to be called in recent parlance "Merge". This is such a fundamental difference between the two types of frameworks that it's difficult to see the impact this would have on the LIH.

At this point, we are willing to propose that there are indeed things in both syntax and morphology that we would call "constructions", for example, our case of the Italian compounds and Booij's construction for morphological examples, and the "Verb X's way PP" construction in (41):

(41) *V X's way PP:*
Fred sneezed his way out the door

We are not, however, willing to concede that because there are some constructions in morphology and syntax that *all* of morphology and syntax should be constructional. It has yet to be shown that independent meaning can be attributed to every syntactic or morphological pattern. Rather, we believe that a constructional approach should be reserved for cases like the ones we've described which are open in certain respects, but quite limited in others.

Meanwhile, for those of us who have become skeptical of syntactic analyses of word formation and who also reject the premise of Construction Grammar that morphology is analogical and declarative rather than generative, what theories do we have at our disposal? Where does all of this leave us with respect to the LIH? The data show us that the interaction between word formation and syntax goes both ways, but that nevertheless it is quite restricted. This is the biggest problem that faces us, and to resolve it we must take stock of what our data really tell us.

We use the table in (42) to summarize a range of cases that have been brought to bear on the LIH, even beyond the ones we have mentioned here, categorizing them according to our assessment of the strength of the challenge they pose. We summarize the cases we think are strong in (43), leaving out cases such as Booij's where what is at stake is arguably inflectional. Our hope in doing this is that by putting together many different examples that have been discussed independently, a clearer pattern will emerge from which we will be able to assess what a new version of the LIH should look like:

(42)

Strong challenges (i.e. concerns productive class, not lexicalized, no good uni-modular analysis)	**Possible/weak challenges** (i.e. class is not productive, examples tend to be lexicalized, or judgments tend to be murky)	**Probably not a challenge** (i.e., class is clearly not productive, or a reasonable uni-modular analysis is available)
NON-INFLECTIONAL		
Phrasal compounds in English, German, Dutch, Afrikaans	Phrasal derivation in English	Resultative V-compounds in Chinese (can be analyzed morphologically)
Insertion/modification into *trasporto latte* constructions in Italian	Compound prefixes in English	Most synthetic compounds in English (can be analyzed morphologically)
Kageyama's W+	Sublexical co-reference in English	Clitics (can be analyzed syntactically)
Quechua nominalizations (A&N)	Phrasal stress on English root compounds (Giegerich)	Noun incorporation (morphological analyses are reasonably strong)
Scope in prefixes		Phrasal derivation in Italian
Conjunction in compounds and prefixed words in English		Causatives (some analyzed as syntax, others as word formation, on language specific basis)
		Verb-particle constructions (some analyzed as syntax, others as word formation, even in the same language, cf. A&N on Dutch)
INFLECTIONAL		
Agreement features in East Netherlandic dialects (A&N: 11)		English genitive (can be analyzed as a clitic)
iets leuks construction in Dutch		
Gruppeninflection or "suspended affixation" (like English conjoined compounds above, but inflectional)		

If, now, we take only those cases that we think are the strongest challenges to LIH, we can categorize them according to the type of inter-component interaction that they imply:

(43) a. Morphology has access to Syntax
- syntactic phrases within words (phrasal compounds)
- insertion/modification into trasporto latte constructions (Italian data)
- conjunction in compound and prefixed words in English
- Quechua nominalizations on phrases (nominalizing suffix selects for phrasal base)(Lefebvre & Muysken 1988)

b. Syntax has access to Morphology
- Quechua nominalizations : position of verb is dependent on whether VP is nominalized or not.
- Agreement in East Netherlandic dialects: word order is dependent upon choice of inflection (or vice versa ?).

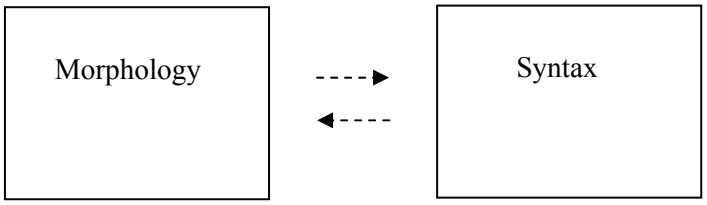
c. Morphology/Semantics interactions
- scopal properties that go beyond the boundaries of a word
- anaphoric properties of sublexical elements (Kageyama 2001)

d. Morphology/Phonology interactions
- phrase level phonology operating within word (Kageyama 2001)

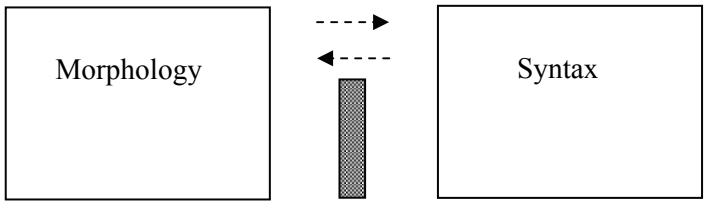
One interesting thing to note about these examples is that only examples in (43a) and (43b) have to do with the morphology/syntax interface with which the LIH was originally concerned. The (43c) examples might conceivably be characterized as concerning the morphology/ semantics interface and the (43d) example the morphology/phonology interface. If so, the original LIH was silent on these, and we might consider whether any new formulation of the LIH should pertain to them at all. Perhaps the jury is still out on this sort of interaction in the absence of systematic study.

The examples in (43) point to the fact that there is interaction between morphology and syntax, but that it is not free, as illustrated in option (44a). Rather, it is it circumscribed (44b). If we can identify the nature of this circumscription, we can come closer to having a new formulation of the LIH.

(44) a.



b.



Descriptively, a generalization might go something like this:

(45) *Towards a new LIH:*
Syntax and morphology are normally blind to each other. However, limited intermodular access may be allowed by virtue of allowing configurations like:

a. $[[XP]\ Y]_Y$ / $[Y\ [XP]]_Y$
b. $[[XP]\ [Y]]_Y$ / $[[Y]\ [XP]]_Y$
c. $[[XP]\ Y]_X$ / $[Y\ [XP]]_X$
d. $[[XP]\ [Y]]_X$ / $[[Y]\ [XP]]_X$

where Y may be null. If any of these configurations is selected for, morphology will obviously have limited access to syntax.

We assume, then that the principles needed to construct phrases and sentences are distinct from the principles needed to construct complex words: in current parlance let us say that Syntactic Merge is different than Morphological Merge. Syntactic Merge produces phrases and sentences, and Morphological Merge produces words. However, there is a point of contact between them, in that languages can allow word formation of certain sorts to Merge syntactic phrases. It is possible, as well, that sentences and phrases can be 'downgraded' to words as part of a process of grammaticalization. The interaction that we seem to need might be stated in the following principle:

(46) *The Limited Access Principle*
Morphological Merge can select on a language specific basis to merge with a phrasal/sentential unit. There is no Syntactic Merge below the word level.

We might go further and try to formalize what we mean by Morphological Merge (leaving Syntactic Merge to the syntacticians!). A tentative statement might be something like (47):

(47) *Morphological Merge*
Let there be items α, β, such that α is a base and β a base or affix. MM takes α, β (order irrelevant)[3] and yields structures of the form < α, β>γ
a. where γ is an X^0, categorically equivalent to α or β, and
b. α or β can be null.

Morphological Merge, together with the Limited Access Principle, yields the sorts of structures that we have highlighted in (45). English phrasal compounds, and probably the compounds with conjoined non-heads in English have the structure (45a). We assume that the Italian compound-like structures have (45b) or (45d), although we must also assume that Y is null in this case, as the structure involves an NP that has been degraded to an N.[4] This sort of down-grading of phrases to words is permitted by clause (b) of Morphological Merger.

[3] Note that making the order irrelevant allows us to cover cases of prefixation as well as cases of suffixation.
[4] As we have seen above, the first constituent or the second constituent can be modified by an adjective, and even both of them (*trasporto veloce latte* / *trasporto latte fresco* / *trasporto veloce latte fresco*) but

Quechua nominalizations seem to have something like the structure in (45a). Kageyama's examples could also potentially be analyzed as cases of prefixation to a phrase or compounding with a phrase (with the absence of Genitive marking being explained by the phrase-internal status of the phrase). And Booij's *iets leuks* analysis might yield to an analysis in which the plural ending *–s* is attached to semi-fixed phrases like *iets leuk*. Many of the examples of scope in prefixes – both the Spanish and the English examples – appear to be analyzable as cases like this as well, where the prefix is syntactically outside a phrase, and therefore has semantic scope over that phrase. As yet, we do not know if all the cases in (45) are actually attested. For example, we do not know if phrasal compounds exist in languages in which the morphology is left-headed as well as in languages that are right-headed. We will have to leave this point open for the time being. Also open is whether it is correct that all structures of this sort are endocentric. We assume so, as the data so far suggests that this is correct.

We believe that some of the examples mentioned in (42) show that syntax may also be allowed a limited view below word level in that syntax may be sensitive to the nature of the non-phrasal constituent in a word with a structure like that in (44) – as in Quechua – but as yet we have too few examples, and too little knowledge of how they work to say with confidence how interaction in this direction must be constrained. In any case, this "view" would not involve Syntactic Merge below the word level.

3. Concluding Remarks

As yet, there are also too few examples to understand fully the interaction of morphological and phrasal semantics, and we confess we have not scoured the literature on the morphology/phonology interface as closely as we could for examples of interaction, so restatements of LIH in terms of other modules will have to wait. Nevertheless we feel that an examination of both the relevant data and the theoretical possibilities leads us in the right direction. We know now that any adequate statement of the LIH must be sensitive to interactions between morphology and other components as well. The data tell us that we do not need to sanction a complete collapse of morphology into syntax or other components, for that matter – this possibility predicts far more interaction than we find. Nor can we explain away the data and maintain that morphology is an island unto itself. There is a point of contact – a small one – between morphology and syntax (and probably between morphology and phrasal semantics and phonology), and our theory must eventually allow for that point of contact.

We feel that our restatement of the LIH as the Limited Access Principle, together with the statement of Morphological Merge that we suggest allows us to loosen the original strictures of the LIH without vitiating it entirely. In other words, we neither deny any contact between morphology and syntax, nor allow free access. Ultimately one would want our statement of the LIH and the Limited Access Principle to follow from something in the architecture of our theory, but at present we are not yet prepared to offer such a theory. We therefore leave this as a goal for future research.

---

the acceptability of these expressions is not the same, the second one being the most acceptable (intended as a fixed compound-like expression).

References

Ackema, P. and A. Neeleman (2004), *Beyond Morphology*, Oxford, Oxford University Press.

Ackerman, F. and P. LeSourd (1997), *Toward a Lexical Representation of Phrasal Predicates,* in A. Alsina *et al.* (eds.) *Complex Predicates,* CSLI Publications.

Anderson, S.R. (1992), *A-morphous Morphology,* Cambridge, Cambridge University Press.

Baker, M. (1988), *Incorporation: A Theory of Grammatical Function Changing*, Chicago, The University of Chicago Press.

Bisetto, A. and S. Scalise (1999), *Compounding: Morphology and/or Syntax?*, in L. Mereu (ed.), *Boundaries of Morphology and Syntax*, Benjamins, 31–48.

Booij, G. (2005), *Construction-dependent Morphology*, "Lingue e Linguaggio" IV.2, 163–78.

Borer, H. (1998), *Morphology and Syntax*, in A. Spencer and M. Zwicky (eds.), *The Handbook of Morphology*, Oxford, Blackwell, 151–90.

Botha, R. (1983), *Morphological Mechanisms*, Pergamon Press, Oxford.

Bresnan, J.W. and S.A. Mchombo (1995), *The Lexical Integrity Principle; Evidence from Bantu.* "Natural Language and Linguistic Theory", 13, 181–254.

Chomsky, N. (1970), *Remarks on Nominalization*, in R. Jacobs and P. Rosenbaum (eds.) *Readings in English Transformational Grammar*, Waltham (MA), Blaisdell.

Chomsky, N. (1995), *The Minimalist Program*, Cambridge (MA), MIT Press.

Di Sciullo, A.M. and E. Williams (1987), *On the Definition of Word*, Cambridge (MA), MIT Press.

Felíu, E. and A. Fábregas (2003), *Phrasal Scope and Argument Constraints of Spanish inter-Prefixation*, in G. Booij *et al*. (eds.), *Topics in Morphology. Selected Papers from the Third Mediterranean Morphology Meeting*, Barcelona, Institut Universitari de linguistica aplicada, Universitat Pompeu Fabra, 165–76.

Goldberg, A. (1995), *A Construction Grammar Approach to Argument Structure*, Chicago, The University of Chicago Press.

Halle, M. and A. Marantz (1993), *Distributed Morphology and the Pieces of Inflection*, in K. Hale and S.J. Keyser (eds.), *The View from Building 20. Essays in Linguistics in Honor of Sylvain Bromberger*, Cambridge (MA), MIT Press, 111–76.

Harley, H. and R. Noyer (1999), *Distributed Morphology*, "Glot International", 4:4, 3–9.

Hohenhaus, P. (1998), *Non-lexicalizability*, "Lexicology", 4:2, 237–80.

Jackendoff, R. (1997), *The Architecture of the Language Faculty,* Cambridge (MA), MIT Press.

Kageyama, T. (2001), *Word Plus: The Intersection of Words and Phrases*, in J. Van de Weijer and N. Nishihara (eds.), *Issues in Japanese Phonology and Morphology*, *Studies in Generative Grammar 51*, Berlin, Mouton de Gruyter. 245–76.

Kornfeld, L. and A.L. Saab (2003), *Morphology and Syntax: Prepositional Prefixes in Spanish*, G. Booij *et al*. (eds.), *Topics in Morphology. Selected Papers from the Third Mediterranean Morphology Meeting*, Barcelona, Institut Universitari de linguistica aplicada, Universitat Pompeu Fabra, 227–40.

Lapointe, S. (1980), *A Theory of Grammatical Agreement*, PhD. diss., Amherst, University of Massachusetts.

Lefebvre, C. and P. Muysken (1988), *Mixed Categories*, Dordrecht, Kluwer.

Li, Y. (2005), X^0 *A Theory of the Morphology-Syntax Interface*, Cambridge (MA), MIT Press.

Lieber, R. (1992), *Deconstructing Morphology*, Chicago, University of Chicago Press.

Marantz, A (1995), *The Minimalist Program*, in G.Webelhuth (ed.) *Government and Binding Theory and the Minimalist Program,* London, Blackwell, 351–82.

Rainer, F. and S. Varela (1992), *Compounding in Spanish*, in Scalise (ed), special issue of "Rivista di Linguistica", 117–40.

Scalise, S. and E. Guevara (2005), *The Lexicalist Approach to Word-Formation and the Notion of Lexicon*, in P. Štekauer and R. Lieber (eds.), *Handbook of Word-Formation*, Amsterdam, Springer.

Spencer, A. (1991), *Morphological Theory*, Oxford, Blackwell.
Spencer, A. (2005), *Word-Formation and Syntax*, in P. Štekauer and R. Lieber (eds.), *Handbook of Word-Formation*, Amsterdam, Springer.
Selkirk, E. (1982), *The Syntax of Words*, Cambridge (MA), MIT Press.
Strauss, S. (1982), *On 'Relatedness Paradoxes' and Related Paradoxes*, "Linguistic Inquiry" 13, 694–700.
Ten Hacken, P. (1994), *Defining Morphology*, Hildesheim–Zurich–New York, Olms.
Williams, E. (1978), *Across-the-board Rule Application*, "Linguistic Inquiry" 9–1, 31–43.

The Lexical Integrity Hypothesis and the Notion of Irregularity: The Case of Spanish Participles*

Antonio Fábregas
Instituto Universitario Ortega y Gasset
webfabregas@gmail.com

Elena Felíu Arquiola
Universidad de Jaén
efeliu@ujaen.es

Soledad Varela
Universidad Autónoma de Madrid
soledad.varela@uam.es

0. A New Approach for the Lexical Integrity Hypothesis

A problem that has always been present in the history of linguistics is to determine what relationship exists between morphology and syntax. The Lexical Integrity Hypothesis (LIH, Siegel 1974, Bauer 1978, Williams 1981, DiSciullo & Williams 1987) proposes that syntax is blind to the internal constituents of a word – i.e., its morphemes –. This principle limits the morphological information that can be read by syntax to the outer layer of the word. In a widely quoted article, Williams (1981) observes that the information perceived by syntax is contained in the head of the word as a whole, which always is in the outer layer. He notes that the past tense of the verb *under-stand* is the irregular *under-stood*, just like the past tense of the verb *stand* is *stood*. From here it follows that syntax must be sensitive to the information contained in the base *stand*.

(1) [under [stand]]

The LIH defines the level whose information is accessed by syntax in a relational way: the last step of the process of derivation is the only one that syntax can see. There are no inherent properties of the layer that make it a 'special' domain, distinct from the rest of the components of the word.

In this paper, we make an alternative proposal in which what can be considered 'the outer layer of a word' is defined by its intrinsic characteristics and a particular configurational status.

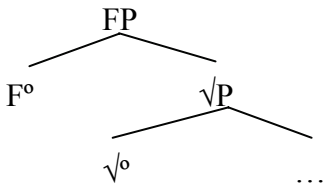
* The research that underlies this article is supported by projects DGI BFF2003-06053 ("Léxico-sintaxis del español. Clases de predicados verbales") and HUM2004-04235-C02-01/FILO ("Límites intercategoriales: las categorías híbridas. Teoría, descripción y aplicaciones"). The authors gratefully acknowledge Mark Aronoff, Olivier Bonami, Gilles Boyé, Carlos Piera, Angeliki Ralli and the audience of the *5th Mediterranean Morphology Meeting* for comments and insightful remarks.

0.1. Word Structure

Our proposal builds over a specific theory about the internal structure of words. We assume that words are constructed through combination of smaller units, morphemes, in a binary branching structure. Words have an internal hierarchical structure which (at least) shares with syntax some properties, such as the fact that heads project in phrases, where relations of c-command and domination hold (Cf. DiSciullo 1993, 1997; Lieber 1992; Borer 2004).

About the classes of morphemes combined, we assume that a morphological word is constructed through combination of roots and functional projections (Embick 2000, 2004; Marantz 1997, 2001; Borer 1999, 2004; Fábregas 2005). In consequence, we assume that the structure of a word is composed of at least two layers of different nature.[1] In the inner layer, we have the root (√) – take for instance English √DOG –, which is the part of the word which contains conceptual semantics. It is also the part of the word responsible of the fact that *dog*, *die*, *intelligent* and *repeat* are different lexical items, with different properties, but lacks crucial pieces of information, such as grammatical category. This root is selected by another head, which heads the second layer of the word: the functional head (F), which assigns a grammatical category to the root and therefore is responsible for the fact that the whole structure is a noun, an adjective or a verb –among other characteristics–, so it distinguishes between the noun *house* and the verb *to house*. In languages such as Spanish, English or French, the functional head may project as a suffix. This second layer of the word is shared by those structures which belong to the same category or subcategory.[2] The set formed by the root layer and the functional layer is what has been traditionally called the stem (2).[3]

(2)



We follow Fábregas (2005: 269 and ff.) in his proposal that the structure in (2) has a special status in the structure of a word. The structure in (2), that Fábregas calls 'Morphological Local Domain' (MLD),[4] is a domain whose information from a

[1] It may be the case that some words are constructed without roots and only with functional heads. This may be the case of the Spanish verb *ser*, according to Fábregas (2005: 271-273).

[2] We use the term subcategory to refer to the different subclasses of words inside the same category, such as transitive vs. intransitive verbs, count vs. mass nouns, qualitative vs. relational adjectives, and so on.

[3] Following Chomsky (2004: 110-111), when two heads are merged, the one that projects is the one whose semantic properties 'select' the other. The root, even though it lacks a category, contains semantic information which, among other things, is enough to select an internal argument. Therefore, when the root is merged with its internal argument, it projects its label, √, absent of categorial features but not of semantic ones, to the complete structure. In the absence of an internal argument, or any other complement of √P, both √ and F are heads when they are merged together, but only F may project as a full phrase, as it semantically selects the other.

[4] It is important to note that MLDs are not syntactic Phases (Chomsky 2001). They are morphological entities, whose existence is not related to syntactic conditions, but to the fact that every independent word needs to have a grammatical category. Let us mention some specific differences between the two. First, Phases may have an edge position which can be used to extract constituents from inside them (Chomsky op.cit.). MLDs do not have this position; this explains that roots cannot be extracted from inside words.

morphological point of view is complete, because it has a grammatical category and conceptual semantics. A Morphological Local Domain is a structural space where phonological, semantic and specifically morphological information is defined. Inside a MLD, all the information is equally accessible. From outside the MLD, the information contained can only be accessed if the higher head of the MLD transmits it, so that the outer heads can read it. Therefore, we propose the principle in (3).

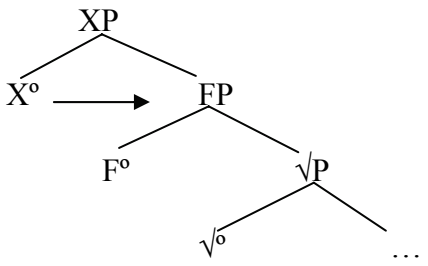
(3) *MLD Information Policy*

a. All the information inside a Morphological Local Domain α is accessible from inside the Morphological Local Domain.

b. From outside a MLD, only the information contained in the higher head is accessible.

If the information is not contained in the projection of the higher head –in the case of (2), in FP–, the information inside the MLD is not available for the rest of the structure. In (4), as FP defines a MLD, X can be sensitive to the information contained in FP, but not in the root phrase √P or the root head √º.

(4)



In our proposal we use the concept of MLD in order to explain whether, inside a word, the morphological information contained in its base is accessible to an affix or not. In particular, we propose the following idea:

(5) The sharing of morphological information between an affix and its base is an epiphenomenon which covers different kinds of relations between heads.

Our alternative for the LIH has, then, some characteristics that differentiate it from other proposals. First, it does not necessarily imply that morphology and syntax are different in nature; it only acknowledges that some information is too far from some heads for them to access it. Being too far means having a head between the higher layer of the MLD and itself. Secondly, what constitutes the domain for information accessing is not defined relationally, as in the case of Williams' proposal, but through the ontological internal properties of the structure.

We will provide evidence for this proposal studying the loss of irregularity in derived verbs. Irregularity is an idiosyncratic property of some roots that has to be

---

Secondly, according to Chomsky (op. cit.), only a special type of little v, i.e., the one with a causative meaning and able to check accusative case, can define a Phase. On the other hand, every functional head able to categorise a root –i.e., little v, little a, and little n– defines a MLD. In fact, the type of little v intervening in verbal participles is not able to define a syntactic Phase, because it is not causative, but passive-inchoative. For further differences between Phases and MLDs Cf. Fábregas (2005: 294 and ff.).

accessible to certain functional heads (such as Tense or Aspect), so it is a pertinent phenomenon to determine the availability of word internal information. We will concentrate on four groups of prefixed verbs in Spanish.

1. Classes of Prefixed Verbs

In this section we will present the data that we will use to substantiate our proposal for the LIH, presenting four different classes of prefixed verbs in Spanish. Let us consider first the Spanish verb *decir*, 'to say'. This verb is irregular, so that in some components of its paradigm special forms are found. We concentrate our interest on the participle in (6).

(6) *decir – dicho*
say – said

As can be seen, the form *dicho* implies special forms of the root, which is spelled out as *d(i)-* instead of the usual *dec-*, and the aspectual head, transformed in *-cho*. However, the behaviour of the participle is not the same in the derivates that are obtained through combination of the verb with different types of prefixes.

1.1. Verbs with Demotivated Meaning that Do Not Keep the Regularity

In the case of the verbs *bendecir*, 'to bless', and *maldecir*, 'to curse', which are constructed from the verb *decir* through the addition of the forms *ben-*, related with the adverb *bien*, 'well', and *mal-*, related with *mal*, 'badly', the irregular form of the verbal participle is simply impossible.

(7) a. *he maldecido,* *he bendecido*
have.1st.sg cursed, have.1st.sg. blessed

b. **he maldicho,* **he bendicho*

1.2. Verbs that Keep the Irregularity with Compositional Meaning

In contrast, in the case of the verb *contradecir*, 'to contradict', which is constructed from the verb *decir* through the addition of the form *contra-*, 'against', related with the preposition *contra*, the irregular form is possible and is the one preferred by speakers.

(8) a. *he contradicho*
have.1st.sg. contradicted

1.3. Verbs without a Straightforward Regular or Irregular Participle

To make things more complex, the case of the verbs *desdecir*, 'to step back', and *predecir*, 'to predict', formed with the prefixes *pre-* and *des-*, is somewhat puzzling. Speakers of contemporary Spanish, as for instance the authors of this paper, feel that none of the forms, the regular or the irregular, are completely perfect.

(9) a. *?he desdecido*, *?he predecido*
have.1st.sg. stepped back, have.1st.sg. predicted

b. *?he desdicho, ?he predicho*

Speakers tend to avoid the use of the participle or, if they have to employ it, they prefer the irregular form, for fear of being considered ignorant of the rules of Spanish grammar. However, none of the forms is perfect.

1.4. Verbs with a Demotivated Meaning that Keep the Irregularity

There is a fourth possibility, which is not documented with a form of the verb *decir*, but is still real and is a needed piece to complete the puzzle of Spanish participles. Some verbal bases with an irregular form of the verbal participle still have the irregular form when combined with certain prefixes, such as *in-*. This is the case of the base *–scrib-*, related to the verb *escribir*, 'to write'.

(10) a. *he inscrito* (cf. *he escrito*)
have.1st.sg. inscribed

b. **he inscribido*

Let us note that prefixed verbs such as this one cannot be classified with (8), *contradecir*, even though both share the property of keeping the idiosyncratic form of the participle. While both the prefix and the base of *contradecir* are independent words on their own, neither *in-* nor *-scrib-*, the two constituents of this verb, are full words in contemporary Spanish.

1.5. A Note on the Possibility of Becoming Regular

The four classes of verbs may exhibit in colloquial Spanish occurrences with the regular participle, including *contradecido* and *inscribido*. One famous case in contemporary Spanish is the verb *proveer*, which shows the regular *proveído* in addition to the irregular *provisto*, and there are even some cases in which the participles in 1.3.4. are regularised[5]. These data are interesting to the extent that they may show that regular

[5] "Hay veces que tú bajas y lo que te quedas es en la playa cogiendo sol porque no hay olas. O sea, te te devuelves, yo me he devolvido, yo me he devuelto como bueno, como cien veces". CSHC-87 Oral interview, Venezuela, 1987; "Un solo rebaño ha poseído el 27% de los genes de todo el Herd Boock de la

inflection is in a sense more basic that irregular inflection (Cf. Pinker 2001), but they are not crucial for our argumentation, even though we recognise that they exist. Relevant to our argumentation is the fact that some of the four classes of verbs may have a strong verbal participle, while others cannot, with the consequences that this fact has on the understanding of the LIH.

1.6. A Note on Accidents: Why the Verb "decir"?

Basically, we consider that the fact that this pattern can be seen with the verb *decir* and not with another verb in contemporary Spanish is a matter of historical accident which may have a motivation, not an explanation. The fact that an entity is regular or irregular is a matter of historical accident, so there is not –in our mind- any synchronic reason that explains why something is irregular or regular, apart from the restrictions on MLD's which we explore in this paper.

There may be a motivation, however, for the verb *decir* to have a special status. The verb *decir* is the most basic *verbum dicendi* in Spanish. Semantically, this gives it a special position in Spanish. Universally, verba dicendi have a particular status. For example, morphemes derived from the most basic verbum dicendi in Hungarian are used to express epistemic modality. There is, moreover, some type of semantic operation which can turn verbs of emission into verba dicendi, such as those which express animal sounds (*maullar*, 'to mew', *ladrar*, 'to bark', *barritar*…). This suggests that the nature of a verb as a verb of saying has a special status. From here it can follow a special historical consideration which makes it distinct from other elements.

Nonetheless, we would like to note that the phenomena studied here are not exactly unique to the verb *decir*, even though this is the most puzzling case. Let us consider, for example, the verb *v**e**nir*, 'to come'. The verb venir has a derivative *v**i**niente*, 'coming', where the vowel /e/ from the root has become an /i/. However, when this verb is the base of a prefixed verb *pro-venir*, 'to come from', and *con-venir*, 'to be convenient', this irregularity is lost, in such a way that the forms are not **prov**i**niente* or **conv**i**niente*. In the verb *convenir*, which belongs to the class of *bendecir*, because its meaning is demotivated, the regular form *conv**e**niente*, 'convenient', is the only one possible. In the case of *provenir*, which belongs to the class of *predecir*, neither the irregular form mentioned nor the regular *prov**e**niente* are considered entirely grammatical. Therefore, the phenomenon we are describing is not restricted to only one verb.

2. Two Preliminary Problems

In this section we provide an analysis of the different classes of prefixed verbs introduced in the previous sections. Crucial to our analysis is the internal structure of a participle and the positions occupied by the different prefixes in Spanish, so we begin with these two problems.

---

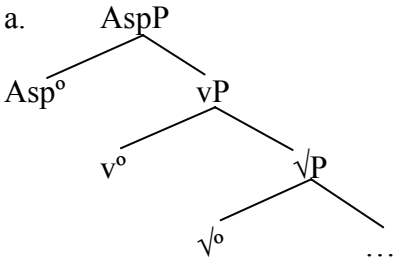
raza Polied Hereford Australiana registrado, en el Angus Neozelandés un solo rebaño ha proveído el 22% de los genes". Raunelli Sander, José W. J., *Genética de la calidad de la carne bovina*, 1994, Perú.

2.1. The Structure of the Participle

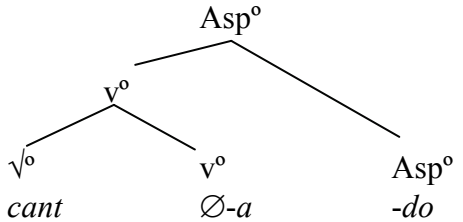
As we said, we assume a distinction between roots and functional projections. Taking the participle *cantado*, 'sung', as an illustration, the root is CANT-, shown also in the nouns *cant-o*, 'song', and *cant-or*, 'singer', and in the adjective *cant-oso*, 'notorious'. Several reasons lead us to propose that there is a little *v* projection dominating this root in the participle (Cf. also Embick 2000, 2004). Semantically, the existence of an event presupposed by the participle makes it necessary that there is an event-denoting category in its structure, and this category is little *v* (Chomsky 1995, 2001, Marantz 1997, Kratzer 1996, Van Hout & Roeper 1998).[6] Formally, the participle contains a theme vowel, which is a morphological property tightly associated with verbal heads, as Oltra (1999) and Oltra & Arregi (2005) argue convincingly. To conclude, paradigmatically, participles are forms which in general are contained inside verbal paradigms.

The combination of the two heads mentioned, little *v* and the root, produces a verbal stem, but there is some additional information needed to obtain a participle. As has already been noted, participles are associated to stative aspect, and are therefore aspectual in nature. Following Embick's previous work, and to a large extent traditional grammars, we propose that there is an aspectual component in participles. Driven by the morpheme order shown and the Mirror Principle (Baker 1985), we propose that the aspectual head selects little *v* (11).[7] (11a) represents the syntactic configuration, and (11b), the morphological configuration after Morphological Merger (Marantz 1984).

(11) a.



b.



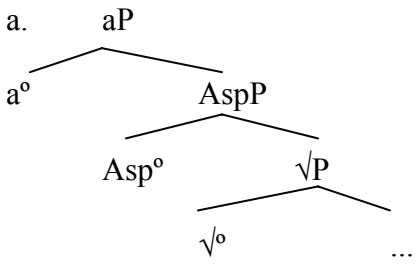
This is the structure that we assume for a verbal participle. As for the adjectival participle, the lack of eventive information and the impossibility of licensing an agent or a manner adverb (Cf. Bosque 1999, Varela 2003, 2004, 2005), implies –in our mind– that the little *v* projection is missing from inside the structure. In contrast, as it denotes a state which can become resultative, AspP is present. As it is an adjective, we propose that a little *a* projection, which categorises the word as an adjective (Marantz 2001, Fábregas 2005) is dominating AspP (12).

---

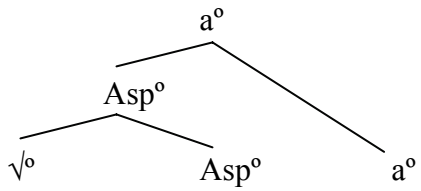
[6] This characterisation of the meaning of little *v* is valid, in principle, for eventive verbs, but it seems inadequate for state verbs, which do not denote events. The analysis of state verbs is a matter on its own and we do not intend to explore it in this article.

[7] As one of the referees observes, this structure does not give account of the ordering of some of the aspect morphemes in Slavic languages, which are prefixes: Cf. Russian *pisa-*, 'to write' vs. ***na****pisa-*, 'to have written'. Even though we must recognise that we don't have an answer to this question, we would like to note that there is, in principle, nothing in our analysis that prevents that Slavic aspect morphemes are different from the Spanish ones –for example, they may be phrasal (Cf. Svenonius, to appear)- and they undergo a syntactic movement that changes their position with respect to the rest of the word.

(12) a.



b.



2.2. Classes of Prefixes

At first blush, the main difference between the four classes of verbs presented in (1) – *inscribir, bendecir, desdecir* and *contradecir*– is to be found in the different classes of prefixes that combine with the base. Therefore, crucial to the analysis is the classification of prefixes in natural languages.

One very appealing classification of prefixes in natural languages that has been successfully applied to the study of Greek (Ralli 2002) and Spanish (Varela & Haouet 2001) is found in DiSciullo (1997), where there is a division between external and internal prefixes. Internal prefixes are those which can change the argument structure of the base, while external prefixes express adverbial notions which operate over an already constructed argument structure. One example of the later would be French iterative *re-* in *réorganiser*, 'to reorganise', while an example of the former is to be found in French *a-* in *apporter*, 'to fetch'.

In this paper we will propose a more complex classification of prefixes. We will analyse all cases of prefixes as category-less constituents which are adjoined to different positions inside word structure – which implies to leave aside, for the sake of the exposition, the possibility that some prefixes are constituents that take a complement[8]. Starting from here, we will consider two factors which combine to produce four different classes of prefixes. From one side, we make a difference between those prefixes which are adjoined to the root layer and those which are adjoined to the functional layer. Following Di Sciullo's (1997) Adjunct Identification Criterion, we will assume that an adjunct must identify –this is, operate on– a characteristic found in the projection to which it is adjoined. If the root layer contains conceptual semantics, we expect that those prefixes which alter the conceptual semantics of the word are adjoined to the root; in other words, we expect that the combination of this kind of prefixes with the base gives as a result a word with demotivated meaning. Following this criterion, we must consider that prefixes such as those that take part in the verbs in (13) (Cf. Aronoff 1976) are adjuncts to the root layer. [9]

(13) *in-ferir, re-ferir, pre-ferir, di-ferir*…
to infer, to refer, to prefer, to differ…

[8] DiSciullo (1997) also proposes that some prefixes are heads. A good candidate for this status are the prefixes of parasynthetic formations, for example *en-* in *encarcelar 'to put in jail'*. Cf. also Varela & Haouet (2001).

[9] A second possibility is to analyse these prefixes as heads which are under the root layer, as Marantz (2003) does. We will not pursue here this track, which may imply problems for the isomorphism between the argument structure and the category definition of constituents. At this point of the argumentation, it is only relevant for us to make the point that the prefixes must be associated to the root layer.

On the other hand, prefixes which are adjoined to the functional layer are expected not to be able to alter the conceptual semantics of the base, but to operate on the formal properties which are defined by the functional head itself, such as case assignment or the number of arguments selected (cf. for example, the distinction between two types of little v heads proposed in Chomsky 1995).

At this point it is important to make explicit our assumptions about what counts as conceptual meaning. In configurational theories, meaning is divided into two classes: conceptual and structural. Structural meaning consists of the aspects of meaning that derive from the syntactic configuration, and, therefore, depends on the formal properties of the structure, while conceptual meaning is encyclopaedic, unpredictable and related to knowledge of the world. Let us consider, as an illustration, theta roles. The number of theta roles of a predicate depends on its structural configuration (Hale & Keyser 1993, Mateu 2002), so the fact that, for instance, a causative verb has two arguments counts as structural semantics. In contrast, the semantic selection of the specific entities that can be a felicitous argument of a verb depends on the speaker's knowledge of the world (Harley & Noyer 2000): for example, we know that *John*, *a boy* or *the writer* are good external arguments for a verb such as *to think*, while *the daisy*, *freedom* or *the construction* make much less sense in that context, even though we may imagine another situation, such as a fairy tale, where their adequacy would improve. The semantic selection of arguments is a fact of conceptual semantics.

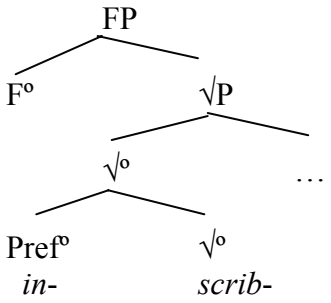
There is a second criterion which will also be employed in this paper, and it is whether the prefix is an adjunct to the head or to the phrase. As we know, both possibilities exist in the grammar of natural languages, but, in accordance with the Linear Correspondence Axiom (Kayne 1994), it is only possible that heads are adjoined to heads and phrases to phrases. From here it follows that we have two types of prefixes, one adjoined to the phrase and the other adjoined to the head, and that those adjoined to heads must be heads and those adjoined to phrases must be phrases. This fact provides us, by implication, with a criterion to determine which adjuncts are heads and which phrases. Let us assume that affixes are heads[10] (Cf. Lieber 1980, Zwicky 1985, DiSciullo & Williams 1987), and that independent words are constructed by the combination of affixes in a meaningful structure. In the Distributed Morphology framework, where the internal structure of a word is a syntactic object, from the previous two assumptions it follows that independent words are structures where syntactic heads are combined, this is, syntactic phrases. The internal logic of this theory, which is, of course, arguable, leads us to consider that those prefixes whose form is that of a complete word[11] are phrases and, therefore, adjuncts to phrases. In addition to this, if the prefix intervenes in the definition of those properties which have to be defined by the head – for example, in the case of the verbal projection, whether the verb selects an internal argument or not –, the prefix is, plausibly, adjoined to the head. By combination

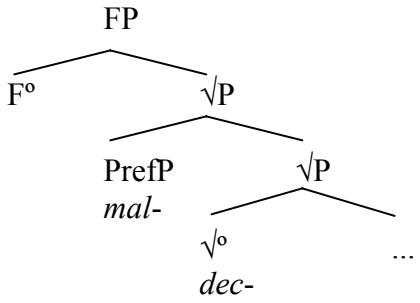
[10] At least, in those theories which consider morphemes to be units and not the result of processes (Siegel 1974, Lieber 1980, Scalise 1984 vs. Aronoff 1976, 1994, Anderson 1992, Spencer 1999).

[11] Note that the reasoning does not force all phrase-adjunct prefixes to be complete words. It is still possible that some prefixes are adjoined to a phrase without themselves being a complete word, due –for example– to prosodic licensing conditions. There are other criteria which can be used, as for example the type of information over which it must have scope. For example, if argument structure is determined, at least partially, by a head, a prefix such as *inter*-, which operates on an already defined argument structure imposing a semantic condition on it (in this case, reciprocity) is a good candidate for being adjoined to the phrase, not to the head.

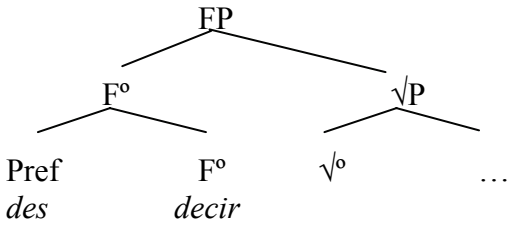
of these two criteria, we obtain four classes of prefixes – which, not accidentally, coincides with the number of different classes of prefixed verbs.

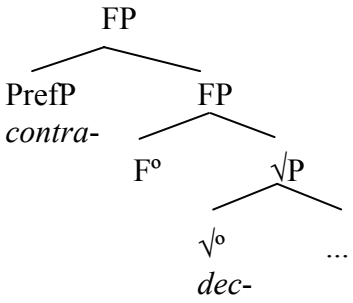
(14) a. *Prefixes adjoined to the root head*. b. *Prefixes adjoined to the root phrase*





c. *Prefixes adj. to the functional head* d. *Prefixes adj. to the functional phrase*





In the following section we will propose that verbs such as *inscribir* are instances of the structure in (14a), while verbs such as *maldecir* are represented as in (14b). On the other hand, verbs such as *desdecir* are analysed as in (14c), and verbs such as *contradecir*, as in (14d).

3. Irregularity and MLD's

In the following section, we will analyse each of the four classes of prefixed verbs according to the structures proposed and we will discuss the impact that the configuration has on the possibility of accessing information contained in the root.

3.1. "Bendecir": A Case of a Prefix Adjoined to √P

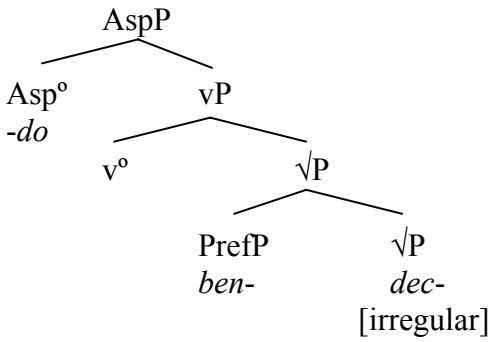
We propose that the elements *ben-* and *mal-* of the verbs *bendecir* and *maldecir* are adjoined to the root layer. We have several pieces of evidence to back this claim.
In the first place, the prefix alters the conceptual meaning of the base. The conceptual meaning of the verbs *maldecir* and *bendecir* is not compositional. It is a new meaning which does not derive from the meaning of the separate constituents. *Maldecir* is not to speak badly of someone, but a specific action which means to curse someone. *Bendecir* is not to speak well, but to bless.

In the second place, the prefix alters the selectional requisites which depend on the root. We follow Chomsky (1995, 2000, 2001, 2004) and Kratzer (1996) with respect to the proposal that the external argument is selected by the functional heads that dominate the root, but the object is associated to the root (Marantz 2003). If the prefix is added to the root layer, we expect that its adjunction may change the semantic selectional requisites of the direct object, but it will not change the number of arguments. *Decir* selects propositional entities as DO, for example CP's or NP's with propositional meaning, such as *que vino*, 'that he came', or *la pregunta*, 'the question'; however, *maldecir* selects individual entities as DO, such as *el niño*, 'the child', or *Pedro*, but never propositional entities.

As can be deduced from the exposition of the internal structure of the word, roots are dependent entities, because they have to be associated to functional heads in order to have some of their properties defined, such as category. From here it follows that we expect entities in the root layer to show signs of dependency. In the case of the elements considered, we can see that they are phonologically dependent, as they do not have stress of their own. The constituent *ben-*, in fact, is associated to the adverb *bien*, but, unlike it, it does not have stress, and is therefore undiphthongised. This behaviour is expected from any element contained in the root layer.

Finally, these prefixes are associated with otherwise independent words, so we will consider them adjuncts to a phrase, and therefore themselves phrases.

(15)



In this structure, vP heads a MLD whose information is the only accessible to the head Aspº, on which it depends whether the participle is regular or irregular. The configuration makes it impossible for the head vº to reflect the meaning that the root is irregular, so the participle must be necessarily regular.

We propose that prefixes are ontologically the same element as roots, for they both lack category information, select their combination semantically and do not have desinences by themselves. The difference between a root and a prefix is structural: roots are bases and prefixes are adjuncts to different positions. From here it follows that ontologically a prefix and a root are undistinguishable by a functional head. The functional head will be sensitive to the information contained in the element which is, in structural terms, nearer to it.

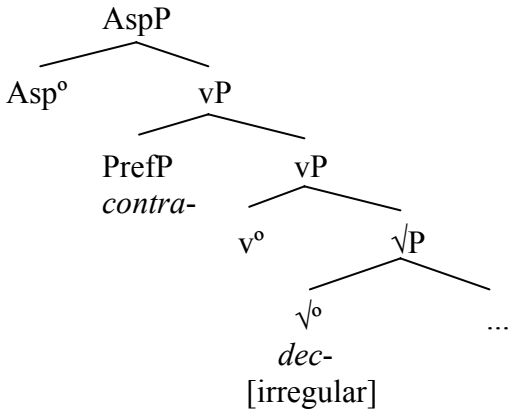
As the prefix is not c-commanded by the root (because it is only dominated by a segment of the category, Cf. Kayne 1994), it c-commands the root. If the crucial relation in syntax is c-command, this means that the prefix is nearer to the functional head than the root. The problem is that the information that the root is [irregular] is present in the root, not in the prefix. As prefixes and roots are ontologically the same type of elements, the functional head chooses the nearest of them, which is the prefix. The

prefix does not have the feature [irregular], so when Aspº, from outside the MLD, checks the features of little v, it does not see a feature [irregular], with the result that it inserts the morpheme by default, i.e., *-do*.

3.2. "Contradecir": A Case of Prefix Adjoined to FP

Now we will consider cases such as *contradecir*. We propose for them the structure in (16), where the prefix is adjoined to the functional phrase.

(16)



In these verbs, the prefixes are related with full words of Spanish, specifically prepositions which may appear as full morphological forms: *sobre mí,* 'over me', *contra mí*, 'against me'. This is not to claim that the prefixes are in fact prepositions, as we do not claim that *ben-* or *mal-* are adverbs in the previous examples. However, the fact that they are formally identical to prepositions shows, in our framework, that they have a phrasal status and that they are not simply dependent heads.

The meaning of the prefixed word is compositionally derived from the meaning of the root and the meaning of the prefix. The semantics of the verb *contradecir* is, to a wide extent, the one that we expect from the phrase *decir X contra Y*, 'to say X against Y'. Similar observations can be made from verbs such as *sobrevolar,* 'to overfly', or *contraindicar*, 'to contraindicate', showing that the behaviour of this kind of prefixes is quite regular.

The prefix changes one property of the verb: case checking. A verb such as *volar* is unable to check accusative case (17a), but in combination with the prefix *sobre-* it acquires the capacity to check this kind of case (17b).[12]

(17) a. *El pájaro voló (*el campo) – *el pájaro lo voló.*
the bird flew (*the countryside) - *the bird itACC flew

[12] It may be in order that case checking and argument selection are considered independent processes in most generative frameworks. Stowell (1981) notes that every category may select a subject in a special structure called Minimal Clause, but only some of them are able to check their subject's case. This is clear in so called Exceptional Case Marking contexts, where a higher verb assigns accusative case to the subject of a lower predicate: *la considero {inteligente / en buena forma}*, 'I consider her (acc.) {intelligent / in good shape}.

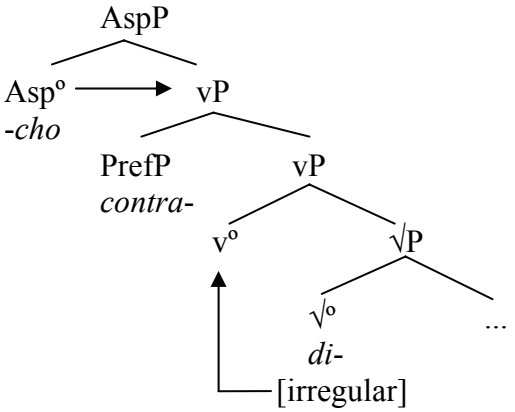
b. *El pájaro sobrevoló el campo – el pájaro lo sobrevoló.*
the bird over-flew the countryside – the bird itACC over-flew
'The bird flew over the countryside'

Case checking is a characteristic of functional projections, which contain the formal features necessary to trigger this syntactic operation (Chomsky 2004: 113-115); roots lack these features, so they are not expected to assign case by themselves. Therefore, if a prefix changes this property of a predicate, we expect that it is adjoined to its functional layer, not to the root.

Finally, these prefixes express notions which can be considered adverbial, in DiSciullo's sense. *Contra-* can be assimilated to a meaning of opposition – *contradecir* is to say something in the opposite direction of what had been previously said–, and *sobre-* usually has a locative meaning – *sobreimprimir* is to print something in a certain position, and so on –. This is the semantics which is associated to external prefixes in DiSciullo's (1997) theory.

If we concentrate now on the structure in (18), we will see that in this configuration the prefix does not prevent the head little *v* to access the information that the root is [irregular]. The irregularity of the root is transmitted unambiguously to v°, and from here it projects to the whole phrase, in such a way that the head Asp° has access to it.

(18)



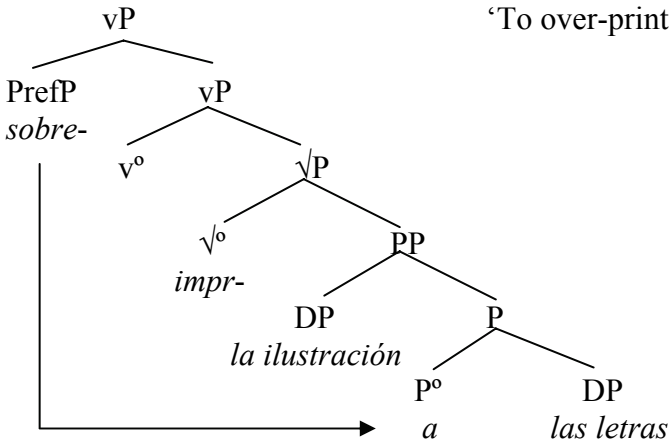
An alternative to our analysis could be to claim that the preposition has been incorporated à la Baker. We have several reasons to reject an analysis where *sobre-* is a preposition that assigns accusative case to the direct object and has been incorporated (Baker 1988) to the verb. First, Spanish prepositions assign oblique case to pronouns (*sobre mí*), but, in these verbs, the direct object exhibits accusative case, which is the case assigned by verbs: *sobreimprimir-**lo***, 'over-print **it** (acc.)'.

Secondly, an incorporated element leaves a trace or copy in its base position, which precludes the insertion of another element. However, it is possible to find a preposition in the place where the trace of the incorporated preposition is expected to be, as in ***sobre**-imprimir una letra **a** otra*, 'to **over**-print one letter **to** the other'.[13]

It is still true, however, that there is a meaning relation between the phrases *decir algo contra algo*, 'to say something against something' and *contradecir algo*, 'to

[13] For the many pieces of semantic evidence to make a distinction between prefixes and prepositions, we refer to Dal (2003).

contradict something', which is, in our mind, the most powerful reason to pursue an incorporation analysis. To give account of this property of the structure, instead, we propose that the prefix is in this case a preposition which does not select an external argument. It has conceptual semantics, in such a way that, when there is a PP in the structure, as this preposition is semantically weak and is c-commanded by the prefix, the prefix imposes the semantic interpretation on the dative. Let us note that the preposition which is materialised in the oblique argument is the weakest possible, *a*. In this way, *a* means *sobre* in *sobreimprimir*.

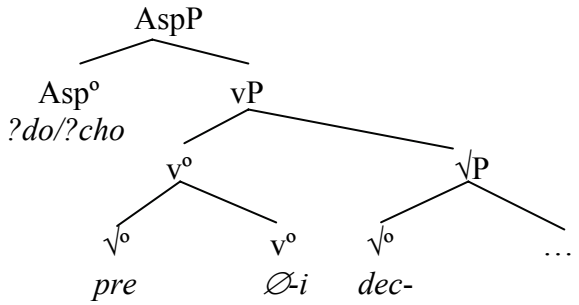
(19) 

'To over-print the picture to the letters'.

This control is semantic, and in fact it is present even in those cases in which the only argument of the verb is a DO, as in *contradecir* or *sobrevolar*. This shows that the prefix is semantically active, imposing an adverbial semantics on the event, but an incorporation analysis must be rejected.

3.3. "Predecir": A Prefix Adjoined to Fº

In this case, we propose that the prefix is adjoined to the functional head:

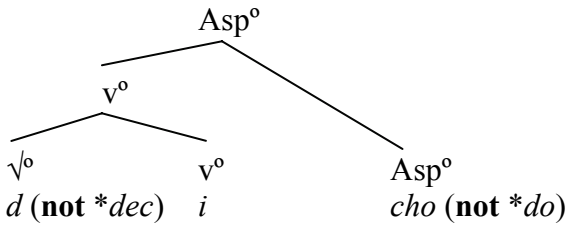
(20) 

Let us note that these prefixes do not change the conceptual meaning of the root, because the words exhibit compositional meaning –*pre-decir* is to say something in advance and to *des-decir* is to go back on what was said–. Instead, they change the case checking properties of the verb: for example, the verb *decir* is transitive and assigns accusative case to its internal argument, while the verb *desdecir* is intransitive and needs a preposition to express the internal argument: *desdecirse* ***de*** *lo dicho*, 'to step back from what was said'. Also, these prefixes have a strong relationship with the verbal

event, contained in the head little *v*: *pre-* implies anteriority; *des-* may imply reversative action. Following the Adjunct Identification Criterion and taking into account the characteristics that these prefixes change, we expect that they are adjoined to little v.

Let us consider now the behaviour of the participle. In principle, the participle of a verb like this should be irregular, because the prefix is merged in a position where it does not intervene between the irregular root and the functional head. That is the reason why the regular form ?*desdecido* is not completely right. The reason why the irregular form is not right either is the following: the irregular participle is built merging three heads, √º, vº and Aspº, that show special allomorphs (21).

(21)



This morphological merging is possible in the verb *contradecir*, because the prefix *contra-* is not merged between two heads. However, in the case under discussion, *pre-* is adjoined to the head, and, therefore, intervenes between the heads vº and Aspº, making impossible the operation of Morphological Merger, which is necessary to build the irregular form. We propose that the reason why a prefix in that position interrupts morphological merger is that prefixes have properties of roots, because, like roots, they lack a grammatical category and contain conceptual semantics. Let us assume the following principle (22).

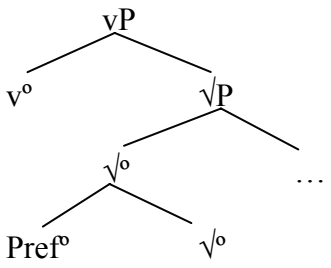
(22) A structure headed by a functional head cannot be morphologically merged with a root.

If this is correct, we expect that *pre-* interrupts merger between Aspº and vº, because vº is a functional head; in contrast, in a verb such as *bendecir*, *ben-* does not interrupt merger between the root and vº, because the prefix is adjoined to a root. Therefore, in these verbs, the irregular participle cannot be formed, either.

3.4. "Inscribir": A Prefix Adjoined to √

The fourth possibility is that in which the prefix is an adjunct to the root head. We propose that this is the case with *inscribir*. This verb is not compositionally derived from the meaning of *escribir*, 'write', and *in-* –and in fact here there are phonological differences which cannot be derived unless idiosyncratically stipulated, as in *deponer,* 'to depose'–. In contrast with prefixes such as *ben-*, *in-* is not an independent word in Spanish, so we propose that it is adjoined to the root head.

(23)



In this position, the prefix does not c-command √P, so little *v* has access to the information that the root is [irregular]. Therefore, little *v* can have the information that the root is irregular and Asp° can read this information from outside the MLD. The result is that the irregular form of the verbal participle is accessible.

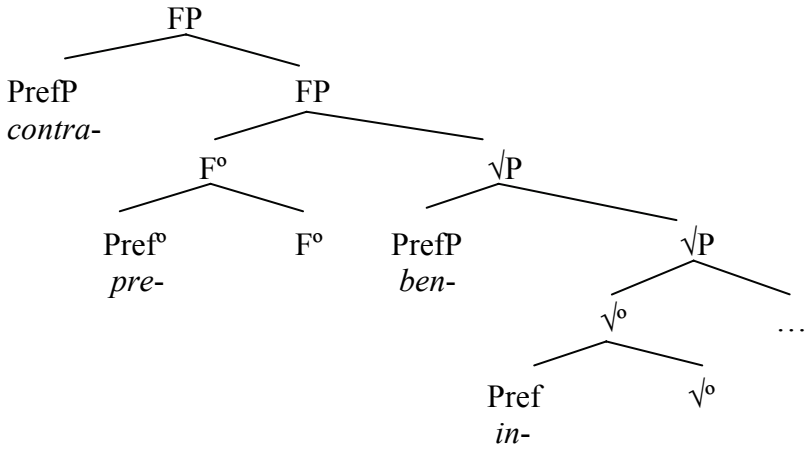
(24) *inscrito* (Engl. *inscribed*)

3.5. The Order of Morphemes

Additional evidence that the system of prefixes that we are proposing is correct comes from the position of the different prefixes. In accordance with the LCA (Kayne 1994), we expect the following ordering (25a), as a reflection of the structure (25b).

(25) a. Prefix adj. to FP – prefix adj. to F° – prefix adj. to √P – prefix adj. to √°

b.



This ordering is in fact verified.

(26) a. *contra-pre-decir* 'to predict something against what someone predicted before'
b. *contra-mal-decir* 'to curse someone back (in resp. for a previous cursing)'
c. *contra-in-scribir* 'to inscribe someth. in response for a previous inscription'
d. *pre-ben-decir* 'to bless someone in advance'
des-ben-decir 'to reverse a previous action of blessing'
e. *des-in-scribir* 'to reverse an event of inscription'
f. *mal-in-scribir* 'to inscribe someone in a bad way'

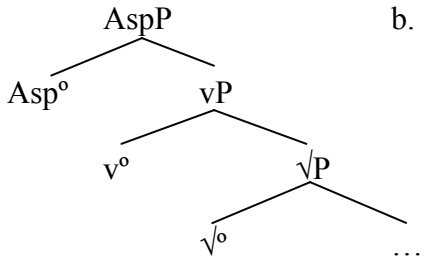
4. Adjectival Participles and Verbal Participles

Many verbs which do not display an irregular verbal participle, have nevertheless an adjectival irregular participle.

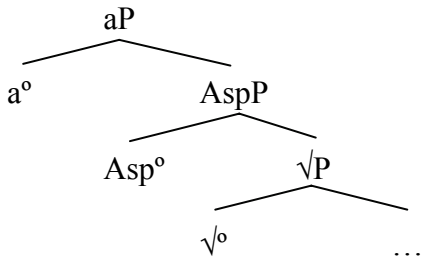
(27) *bendito, maldito, corrupto*
lit. blessed, cursed, corrupted

This is a fact which also needs to be explained. The difference between a verbal and an adjectival participle is the lack in the second structure of a little *v* projection. (28a) represents the verbal participle; (28b), the adjectival one.

(28) a.

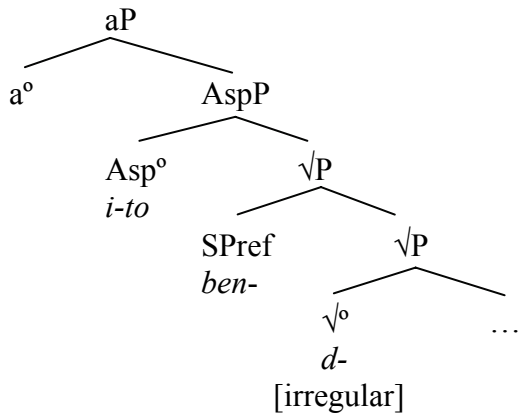


b.



Let us note that MLD's theory predicts that a suffix unable to assign a grammatical category may appear inside and outside the MLD, depending on the order in which the different heads are merged in the structure. This is the case of appreciative morphemes, as argued in Fábregas (2005: 289-292). If Aspº is a functional head unable to categorise a root we expect that it may appear inside or outside the MLD. Indeed, there are reasons to propose that Asp does not assign a grammatical category, because the notion that it expresses is trans-categorial and Aspº combines with verbs, adjectives, nouns and prepositions. Apart from verbs, adjectives also have aspectual information, as evidenced by the opposition between stage-level adjectives (such as *descalzo,* 'barefoot') and individual-level adjectives (such as *mortal*, 'mortal') (Luján 1980). Some nouns contain also aspectual information (Cf. Musan 1995), for there are nouns which can be combined with aspectual adverbs such as *ya*, 'already', or *dos veces*, 'twice', and, finally, prepositions may be telic (central coincidence) or atelic (terminal coincidence) (Hale & Keyser 1993, 1998, 2002). From here it follows that Aspº is a head whose information is functional, but unable to determine the grammatical category of the word with the consequence that in (25b) Asp does not define a MLD; the MLD is defined, instead, by the head little *aº*. This means that Asp is inside the MLD in an adjectival participle. The immediate consequence of this is that, by virtue of the principle that claims that inside an MLD all the information is accessible, Asp will always be able to read the information that the root is [irregular].

(29)



There are independent reasons to propose that Aspº is external in a verbal participle, and internal in an adjective. Let us note that a verbal participle is part of one of the forms of the verb in a regular paradigm: every verb, independently of its Aktionsart, has a verbal participle. In contrast, aspect is not part of an adjective's paradigmatic information. Each adjective is associated with a particular aspectual value: some are stage-level and some are individual-level. Even if some adjectives may be stage- or individual-level, this is usually related with a difference in meaning, as is shown, for example, by the constrast between *aburrido* (individual-level), 'boring', and *aburrido* (stage-level), 'bored' (Cf. Varela 2003).

5. Other Forms of the Paradigm

In this section we will take a view to the rest of the paradigm of these verbs. We will focus precisely in two contrasts. The first one is the fact that the aorist ('pretérito indefinido' in the traditional grammars of Spanish) behaves differently from the verbal participle, even though they are semantically related. The aorist is irregular, while the verbal participle isn't (30).

(30) a. *bendecido* vs. *bendije, bendijiste, bendijo, bendijimos*…
blessed (regular) vs. bless.past.irregular…

b. *?predicho* vs. *predije, predijiste, predijo, predijimos*…
predicted (irregular) vs. predict.past.irregular

The second is the fact that the future behaves exactly like the verbal participle. The verb *decir* has an irregular future (31a). The verbs from the first group make a regular future and the irregular is impossible (31b), while the verbs from the second group have a regular future (31c), the verbs from the third group don't admit any of them, regular or irregular (31d), and the verbs from the fourth group produce an irregular future (31e).

(31) a. *decir – diré*
say – say.future.irregular

b. *bendeciré – *bendiré*
bless.future.regular – bless.future.irregular

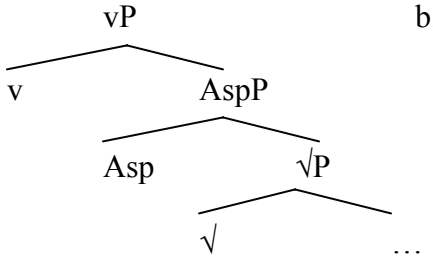
c. *contradiré – *contradeciré*
contradict.future.irregular – contradict.future.regular

d. ?*predeciré – ?prediré*
predict.future.regular – predict.future.irregular

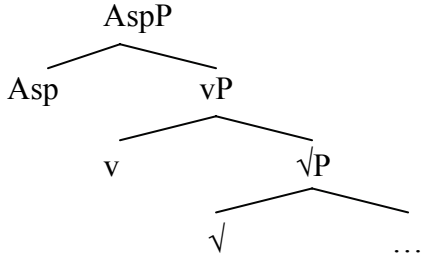
e. *contendré – *conteneré*
contain.future.irregular – contain.future.regular

These forms should be explained according to our theory of MLD's. Let us consider first the difference between the two perfects. In this case, we will propose that the difference is to be found in the two positions which can be occupied by the node Asp, which is [perfective] in both cases. In the case of the aorist, we propose that there is an internal Aspº head, which is inside the MLD and therefore can have access to the information that the root is [irregular]. In contrast, the participle has an external Aspº, this is, an Aspº which is outside the MLD.

(32) a.



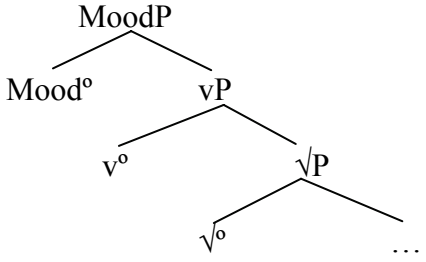
b.



This situation is not surprising; every verb has a verbal participle, but some verbs do not have an aorist. This depends partially from the properties of the verb, in such a way that a verb such as *saber,* 'to know', when it appears in the aorist, has another meaning. Another property of aorists is that they can be expressed with suppletive roots, as *ser – fui*, *ir – fui*, etc.

As for the future, we assume Oltra's (1999) proposal that the future is modal in nature and therefore is the result of the structure in (33), where the future morpheme has to be expressed in Moodº. Let us note that future tenses are opaque contexts where, among other things, it is possible to have an unspecific interpretation of indefinite arguments, as in *encontraré un lápiz*, 'I wil find a pencil', which does not even imply that there exists a pencil. The configuration in (33) is the same as in the case of the participle, so we expect precisely the same pattern of forms.

(33)



Imperative forms are also 'modal' forms where indefinite arguments may have a non-specific interpretation: in *encuentra un lápiz*, 'find a pencil', there is no implication that there even exists a pencil. It is not implausible to think that the structure of the imperative may be parallel to the structure of the future. Let us note that the relationship of these forms with irregularity is similar to the one that we find with verbal participles:

(34) a. *decir – di*
to say – say!

b. *bendecir – bendice, not *bendí*
to bless – bless!

c. *contradice – contradí, not *contradice*
to contradict – contradict!

d. *predecir – ?predice / ?predí*
to predict – predict!

However, our proposal is not a general proposal about irregularity inside the verbal paradigm, because there are other irregularity phenomena that are, at least at this point, different from the cases we are considering, such as irregularity in imperfect past tenses, or in different forms of the present tense.

6. Consequences for Irregularity

Our proposal presents irregularity as a phenomenon that interacts with the internal structure of the word. The same base behaves regularly or irregularly depending on the internal structure, arguably syntactic, so, in our mind, is not clear how to account for these data in a paradigmatic approach. Consequently, in our proposal irregularity is rather a characteristic of individual items that is inherited by the word as a whole if its internal structure allows it. In those cases in which irregularity cannot be projected to the whole, it is lost, because it is not accessed by the elements that have to spell out the items.

References

Anderson, S. (1992), *A-morphous Morphology*, Cambridge, Cambridge University Press.
Aronoff, M. (1976), *Word Formation in Generative Grammar*, Cambridge (Mass.), MIT Press.
Aronoff, M. (1994), *Morphology by Itself*, Cambridge (Mass.), MIT Press.
Baker, M.C. (1985), *The Mirror Principle and morphosyntactic explanation*, «Linguistic Inquiry», 16 (3), 373-416.
Baker, M.C. (1988), *Incorporation. A Theory of Grammatical Function Changing*, Chicago, University of Chicago Press.
Bauer, L. (1978), *The Grammar of Nominal Compounding. With Particular Reference to Danish, English and French*, Odense, Odense University Press.

Bonami, O. & G. Boyé (2003), *Supplétion et classes flexionnelles dans la conjugaison du français*, «Langages», 152, 102-126.

Borer, H. (1999), *Endo-Skeletal vs. Exo-Skeletal Explanations in Linguistics*, ms., USC.

Borer, H. (2004), *Structuring Sense. An Exo-Skeletal Trilogy. Vol. 1: In Name Only*, Oxford, Oxford University Press.

Bosque, I. (1999), *El sintagma adjectival. Modificadores y complementos del adjetivo. Adjetivo y participio*, in Bosque, I. & Demonte, V. (dirs.), *Gramática Descriptiva de la Lengua Española*, Madrid, Espasa, 217-311.

Cartstairs-McCarthy, A. (2003), *Paradigm Irregularity as the Basis for Morphology's Existence*, electronic ms. available at:
URL: http://lings.ln.man.ac.uk/Info/staff/KEB/Conferences/Irregularities. html

Chomsky, N. (1995), *The Minimalist Program*, Cambridge (Mass.), MIT Press.

Chomsky, N. (2000), *Minimalist Inquiries: the framework*, in Martin, R. *et alii* (eds.), *Step by Step: Essays in Honour of Howard Lasnik*, Cambridge (Mass.), MIT Press, 89-155.

Chomsky, N. (2001), *Derivation by Phase*, in Kenstowicz, M. (ed.), *Ken Hale: A Life in Language*, Cambridge (Mass.), MIT Press, 1-52.

Chomsky, N. (2004), *Beyond explanatory adequacy*, in Belletti, A. (ed.), *Structures and Beyond. The Cartography of Syntactic Structures*, Vol.3, Oxford, Oxford University Press, 104-131.

Dal, G. (2003), *Arguments pour un préfixe contre-*, «Recherches Linguistiques», 26, 172-201.

DiSciullo, A.-M. (1993), *The Complement Domain of a Head at Morphological Form*, «Probus», 5, 95-125.

DiSciullo, A.-M. (1997), *Prefixed Verbs and Adjunct Identification*, in Di Sciullo, A.M. (ed.), *Projections and Interface Conditions*. New York: Oxford University Press, 52-73.

DiSciullo, A.-M. & E. Williams (1987), *On the Definition of Word*, Cambridge (Mass.), MIT Press.

Embick, D. (2000) *Features, syntax and categories in the Latin Perfect*, «Linguistic Inquiry», 31 (2), 185-230.

Embick, D. (2004) *On the structure of resultative participles in English*, «Linguistic Inquiry», 35 (2), 355-392.

Fábregas, A. (2005), *The Definition of the Grammatical Category in a Syntactically Oriented Morphology. Nouns and Adjectives*, unpublished doctoral dissertation, Madrid, IUIOG / UAM.

Fábregas, A., E. Felíu Arquiola & S. Varela (2006), *The Lexical Integrity Hypothesis and the notion of irregularity: the case of Spanish participles*, in G. Booij *et alii*, *Electronic Proceedings of the 5th Mediterranean Morphology Meeting.*

Hale, K. & S. J. Keyser (1993), *On argument structure and the lexical expression of syntactic relations*, in Hale, K. & S. J. Keyser (eds.), *The View from Building 20. Essays in Honor of Sylvain Bromberger*, Cambridge (Mass.), MIT Press, 111-176.

Hale, K. & S. J. Keyser (1998), *The basic elements of argument structure*, MITWPL, 32, 73-118.

Hale, K. & S. J. Keyser (2002), *Prolegomena to a Theory of Argument Structure*, Cambridge (Mass.), MIT Press.

Harley, H. & R. Noyer (2000), F*ormal vs. encyclopaedic properties of Vocabulary. Evidence from nominalisations*", in Peeters, B. (ed.), *The Lexicon/Encyclopedia Interface*, Amsterdam, Elsevier Press, 349-374.

Kayne, R. (1994), *The Anti-Symmetry of Syntax*, Cambridge (Mass.), MIT Press.

Kratzer, A. (1996), *Severing the external argument from its verb*, in Rooryck, J. & L. Zaring (eds.), *Phrase Structure and the Lexicon*, Dordrecht, Kluwer.

Lieber, R. (1980), *On the Organization of the Lexicon*, doctoral dissertation, Cambridge (Mass.), MIT.

Lieber, R. (1992), *Deconstructing Morphology*, Chicago, University of Chicago Press.

Luján, M. (1980), *Sintaxis y semántica del adjetivo*, Madrid, Cátedra.

Maiden, M. (2005), *Morphological autonomy and diachrony*, in Booij, G. & J. van Marle (eds.), *Yearbook of Morphology 2004*, Dordrecht, Kluwer, 137-177.

Marantz, A. (1984), *On the Nature of Grammatical Relations*, MIT Press, Cambridge MA.

Marantz, A. (1997), *No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon*, in Dimitriadis, A. *et alii* (eds.), *Penn Working Papers in Linguistics,* 4.2., 201-225.

Marantz, A. (2001), *Words*, ms., Cambridge (Mass.), MIT.

Marantz, A. (2003), *Subjects and Objects*, ms., MIT.

Mateu, J. (2002), *Argument Structure. Relational Construal at the Syntax-Semantics Interface*, doctoral dissertation, Barcelona, Universidad Autónoma de Barcelona.

Musan, R. (1995), *On the Temporal Interpretation of Noun Phrases*, doctoral dissertation, MIT.

Oltra, I. (1999), *On the Notion of Theme Vowel: A New Approach to Catalan verbal Morphology*, master's thesis, MIT.

Oltra, I. & K. Arregi (2005), *Stress-by-Structure in Spanish*, "Linguistic Inquiry", 36:1, 43-84.

Ralli, A. (2002), *Prefixation vs. compounding* , in DiSciullo, A.M. (ed.), *Asymmetry in Grammar*, vol. 2., Amsterdam & Philadelphia, John Benjamins, 37-63.

Rothstein, S. (1999), *Fine Grained Structure in the Eventuality Domain: the Semantics of Predicative Adjective Phrases and Be*, "Natural Language Semantics", 7, 347-420.

Scalise, S. (1984), *Generative Morphology*, Dordrecht, Foris.

Siegel, L. (1974), *Topics in English Morphology*, doctoral dissertation, Cambridge (Mass.), MIT.

Spencer, A. (1999), *Transpositions and argument structure*, in Booij, G. & J. van Marle (eds.), *Yearbook of Morphology 1998*, Dordrecht, Kluwer, 73-102.

Stowell, T. (1981), *Elements of Phrase Structure*, doctoral dissertation, Cambridge (Mass.), MIT.

Stump, G. (2001), *Inflectional Morphology. A Theory of Paradigm Structure*, Cambridge, Cambridge University Press.

Svenonius, P. (to appear), *Russian Prefixes are Phrasal,* "Proceedings of FSDL 5".

Van Hout, A. & T. Roeper (1998), *Events and Aspectual Structure in Derivational Morphology*, in *MITWPL* 32, 175–200.

Varela, S. (2003), *Lexical Morphology revisited: Form / meaning correspondences in Psych adjectival participles*, in Booij, G. *et alii* (eds.), *Topics in Morphology. Selected Papers from the 3rd MMM*, Barcelona, Universitat Pompeu Fabra, 51-75.

Varela, S. (2004), *La regularización léxica como proceso funcional y formal: cambio diacrónico, variación dialectal y préstamo*, "Hispanismo*"*, I, 17-36.

Varela, S. (2005), "Lexicalización y cambio categorial", en L. Santos Río *et alii*, (eds.), *Palabras, norma, discurso. En memoria de Fernando Lázaro Carreter*, Salamanca, Ediciones Universidad de Salamanca, pp. 1133-1145.

Varela, S. & L. Haouet (2001), *For a morphological analysis in the privacy of the lexicon: prefixed verbs*, "Cuadernos de Lingüística del Instituto Universitario Ortega y Gasset"*,* 8, 53-69.

Williams, E. (1981), *On the notions 'lexically related' and 'head of a word*, "Linguistic Inquiry", 12: 2, 245-274

Williams, E. (1994), *Remarks on lexical knowledge*, "Lingua"*,* 92, 7-34.

Zwicky, A. (1985), *Clitics and Particles*, "Language", 61, 283-305.

Morphology and Syntax Inside the Word: Pronominal Participles of Headless Relative Clauses in Turkish

Aslı Göksel
Boğaziçi University, Istanbul & SOAS, London
gokselas@boun.edu.tr

Abstract

This paper investigates the structure of pronominal participles in Turkish and their implications for the *Lexical Integrity Hypothesis* (Anderson 1992, Bresnan and Mchombo 1995, Mohanan 1995). I will argue that the expression of grammatical functions in participles is constrained by the formal properties of word structure. This claim is based on the observation that in Turkish the participles of headless relative clauses (pronominal participles) have a fixed size. Grammatical relations are then expressed within the space made available by conditions on word structure. Some interesting aspects of word structure emerge from this. One of these is that although morphological conditions determine the formal properties of words, the interpretation of word internal elements shows partial sensitivity to combinatorial ordering restrictions. The data thus support the weaker version of the *Lexical Integrity Hypothesis* as discussed by Booij (2005) where syntactic mechanisms have access to word internal elements but cannot manipulate them. Secondly, pronominal participles embody a form-function mismatch whereby the relevant suffixes that occur on pronominal participles belong to the nominal inflectional paradigm, yet they are associated with syntactic functions typical of the verbal paradigm. Finally, these participles allow the marking of non-subject arguments on a form, a phenomenon which is otherwise unattested in Turkish.[1]

1. Introduction

Regarding the competing views on the nature of word structure, the investigation into whether a separate word formation component is warranted has been fed by two sources of research. One of these sources has to do with syntactic categories and whether head movement can give all and only the words attested in languages (Baker 1985, 1988). The other source from which the notion of a separate word formation component has been fed is the discussion surrounding the place of the lexicon in grammar as a component of word formation rules, research which dates back to Jackendoff (1975) and Aronoff (1976). A sub-branch of the views attributing word structure to a component other than syntax posits a separation of morphology from the lexicon, the former being a system of word-formation and the latter a list of items (Di Sciullo and

[1] I would like to thank the audiences at the Linguistics Seminar, Boğaziçi University, April 2004 and the Fifth Mediterranean Morphology Meeting, Frejus, September 2005 for their feedback, where earlier versions of this paper were presented. For their comments, I am also indebted to Cem Bozşahin and two anonymous reviewers for *Lingue e Linguaggio* where a shortened version of this paper has appeared (Special Issue on Lexical Integrity, *Lingue e Linguaggio* IV.2 (2006)). Thanks also go to Hasan Mesut Meral for technical and editorial help. Needless to say, all errors are mine.

Williams 1987, Ackema and Neeleman 2004). Within this constellation, the notion of the impenetrability of word structure by syntactic mechanisms has a central role, as this turns out to be one of the decisive factors in the choice between these models. The structure of pronominal participles in Turkish, the participles of headless relative clauses, is one such testing ground for the place of morphology and syntax in word formation.

The structure of a pronominal participle given in (1) is three ways ambiguous in Turkish, as illustrated in (2):[2]

(1) VERB-….-RELATIVISER-PLURAL-POSSESSIVE

(2) a. sev-di-k-**ler-imiz**
like-T-REL-LAR-1PL.POSS
'those who we like/liked

a'. [SUB$_j$ **___i (OBJ)** sev-di-k-**ler$_i$**-imiz$_j$]

b. sev-en-**ler-imiz**
like-REL-LAR-1PL.POSS
'those who like/liked us'

b'. [**___i (SUB)** OBJ$_j$ sev-en-**ler$_i$**-imiz$_j$]

c. (köpek) sev-en-**ler-imiz**
dog like-REL-LAR-1PL.POSS
'those among us who like/liked dogs'

c'. [**___i (SUB)** OBJ sev-en-**ler$_i$-imiz$_i$**]

Some speakers have a fourth interpretation for the sequence in (1), which is contingent on a number of factors to be elaborated in section 4.3:

d. köpek ısır-an-**lar-ımız**
dog like-REL-LAR-1PL.POSS
'those among us who dogs bite/bit'[3]

d'. [SUB **___i (OBJ)** ısır-an-**lar$_i$-ımız$_i$**]

The words in (2a-d) are nominalised non-finite verb forms that function as participles in headless relative clauses. In (2a) the plural morpheme *-lAr* is coindexed with the gap in the object position, and the possessive marker agrees with the subject which may be overtly expressed, as illustrated in (2a'). In (2b) these roles are reversed. While the plural morpheme is coindexed with the gap in the subject position, the possessive marker refers to the object. The situation in (2c) and (2d) is somewhat different. In both of these, the plural morpheme indicates a subset of the denotation of the possessive marker. But in the two cases the possessive marker refers to different

[2] The abbreviations used in this paper are: ACC: accusative, CNJ: conjunctive suffix, COMP: complementiser, CT: contrastive topic, DAT: dative, FRC: relative clause with overt lexical head, FUT: future, GEN: genitive, LAR: the set whose members are *-lAr* (3rd person plural pronominal suffix) and *-Ø* (3rd person singular pronominal suffix), NSR: non-subject relativiser and/or constructions containing it, OBJ: object, PASS: passive, PL: plural, P.COP: past copula, POSS: possessive (nominal agreement), PRC: pronominal/headless relative clause, REL: relativiser, SG: singular, SR: subject relativiser and/or constructions containing it, SUB: subject, T: tense

[3] The speakers for whom this interpretation is unavailable find the parallel interpretation in the corresponding full relative clause construction also unacceptable. Although this interpretation is highly marked, I shall include it in the investigation here, as it shows an interesting regularity in the interpretation of the relevant suffixes.

grammatical functions. While the denotation of the possessive marker is the subject in (2c), it is the object in (2d). These observations are summarised below:

(3)	VERB	-REL	*-lAr*	-POSS
			a. non-subject[4]	subject
			b. subject	non-subject
			c. part	whole (subject)
			d. part	whole (non-subject)

The four-way ambiguity in (2) is partly resolved by the relativising suffixes. Relativisation in Turkish employs two different suffixes which belong to a group of nominalisers: *-K-* in (2a), and *-(y)An* in (2b-d), hence the choice of possessive markers for indicating person. The internal structure of non-finite nominalised verb forms is partly determined by the lexical specifications of these relativisers, discussed in section 6. The second source is syntactic and has to do with the type of grammatical function each relativiser targets. If a non-subject relativiser is used, the following *-lAr* is interpreted as referring to the non-subject gap in the relative clause. However, neither of these constraints explains the interpretation in (2b-d). These forms show that the position of *-lAr* and the possessive marker remains fixed, but their syntactic associations are different. This leads to another well-formedness condition, one that is imposed by constraints on the formal properties of the word and the inability of syntactic operations relating to argument structure to change it.

The data also show that the participles in (2a) and (2b-d) behave differently with respect to co-ordination. As we shall argue, this has implications for the interpretation of the various versions of the *Lexical Integrity Hypothesis* as discussed by Anderson (1992), Bresnan & Mchombo (1995), Mohanan (1995) and Booij (2005). The discussion below will specifically try to bring to light the role of morphology in word-formation and the weight it has with respect to syntax in the organisation of the word. It will be claimed here that the model which best accommodates the present data is a tripartite model with a separation of the lexicon, morphology and syntax, where morphology and syntax are distinct components as suggested by Di Sciullo and Williams (1987) and Ackema & Neeleman (2004).

These claims will be based on the following points:

i. PRC participles have a fixed ordering of affixes, irrespective of their syntactic function

ii. PRC participles have a fixed maximal size, irrespective of whether the expression of more functions is required syntactically

iii. PRC participles use affixes from the nominal paradigm irrespective of the fact that these fulfil grammatical functions

However, in addition to these factors which highlight the sensitivity of word structure to morphological constraints and which cannot be explained by syntactic operations, the lexical specifications of the relativisers show at the same time the

[4] Here the term 'non-subject' will be used instead of 'object', as the observations above also apply to adjuncts.

presence of word internal compositionality typical of syntactic phrases. The topic of this paper is the interaction of these properties.

The paper is organised as follows. In section 2 I lay out the general properties of full relative clauses (FRCs, by which is meant relative clauses with lexical heads), the properties of the relativizing suffixes and headless relative clauses which I shall refer to as pronominal relative clauses (PRCs).[5] Section 3 discusses why the PRC construction is not a reduced version of an FRC construction with a deleted head, as one might be led to think from the suffixes they share. It will be argued here that *-lAr* in PRCs is not a plural marker but a 3rd person plural pronominal suffix. In section 4, further discussion on person marking in PRCs is presented. This is followed in section 5 by the role of morphology in shaping the word, and in section 6 by the permeation of syntactic factors into the word. In section 7 we look at suspended affixation and discuss to what extent it forms a diagnostic with respect to syntactic intervention. The paper concludes with a discussion of the implications of these mixed findings for the *Lexical Integrity Hypothesis*.

2. Pronominal relative clauses and their relation to full relative clauses

2.1. Relative clauses

A proper analysis of the internal structure of PRCs rests on whether they are head-deleted versions of FRCs. FRCs are right-headed constructions which are of two types, marked by the two distinct relativising morphemes given below:[6]

(4) Subject relativiser (SR): *-(y)An*
Non-subject relativiser (NSR): *-K-*[7]

SR occurs under two conditions:

i. Where the relative clause contains no subject (which means that the subject is relativised):[8]

(5) a. ___$_{i\ (SUB)}$ çiçek-ler-i sev-en Semra$_i$
flower-PL-ACC like-SR Semra
'[the] Semra, who likes flowers'[9]

[5] The choice of the term 'pronominal relative clause' rather than the commonly used term 'headless relative clause' is for reasons of convenience. As I refer to the participles of these clauses as 'pronominal participles' I have chosen to use the abbreviation 'PRC' for its more transparent link with its participle.

[6] Capital letters in the lexical representation of affixes indicate variability due to phonological processes (except for the abbreviation LAR which denotes a set, see footnote 2 above). The bracketed consonant 'y' is a buffer consonant occurring where two vowels would otherwise be adjacent. Hence *-lAr*: -ler/lar, *-K-* k/ğ, *-(y)An*: -en/-an/-yen/-yan.

[7] The internal structure of the NSR suffix is a matter of debate. Kural (1993), Göksel (1997), Göksel (2001), Tekin (2001), Kelepir (2007) analyse *-K-* as a separate morpheme which attaches to the tense/aspect/modality-related suffixes *-DI* and *-(y)AcAK*, while in the remainder of the literature it is implicitly assumed to be an unanalysable part of the suffixes *-DIK* and *-(y)AcAK*. The choice between these two claims has no bearing on the issues discussed here.

[8] Except where the predicate is passive. The details of this analysis are not relevant to the discussion here and its implications will not be addressed.

b. ___$_{i\ (SUB)}$ çok konuş-an insan-lar$_i$
a.lot talk-SR person-PL
'people who talk a lot'

ii. Where the relative clause does contain a subject but one that is categorial/generic (which means that a non-subject constituent is relativised).[10] The non-subject constituent which is relativised is usually the direct object as in (6a), or the specifier inside a non-subject constituent, as in (6b):[11]

(6) a. ___$_{i\ (OBJ)}$ köpek ısır-an kız$_i$
dog bite-SR girl
'the girl who a dog/dogs bit'[12]

b. [___$_{i\ (SPEC)}$ el-in-i] köpek ısır-an kız$_i$
hand-3SG.POSS -ACC dog bite-SR girl
'the girl whose hand dogs/a dog bit'

Otherwise NSR are used. This covers cases where the direct object, indirect object or adjunct is relativised and the relative clause does not contain a categorial/generic subject. The participles of non-subject relative clauses obligatorily contain nominal agreement markers for subject agreement and they can optionally have a subject in the genitive case:

(7) (Semra-nın) $_j$___$_{i\ (OBJ)}$ sev-di-ğ-i$_j$ çiçek-ler$_i$
Semra-GEN like-T-NSR-3SG.POSS flower-PL
'the flowers that Semra likes'[13]

There is at least one case, however, where the two strategies overlap and irrespective of which relativiser is used, the interpretation of the relative clause is the same. This happens when a constituent within a sentential subject is relativised. In this case, either strategy can be used without any effect on the interpretation (adapted from Csató (1985) cited in Barker et al (1990)):

(8) a. [[biz-im ___$_i$ güven-eceğ-imiz] şüpheli ol]-**an** adam$_i$[14]
we-GEN trust-COMP-1PL.POSS doubtful be-SR man
'the man that it is doubtful we will trust'

---

[9] The forms with *-(y)An* and *-K-* are underspecified for tense and aspect, hence in this article the translations of these predicates will variably be perfective or imperfective.

[10] While it has been recognised in the literature that *-(y)An* can occur in clauses that contain a subject, it is still commonly referred to as 'SR'. I shall therefore use this term for practical reasons.

[11] See Göksel and Kerslake (2005) for a list of such constituents.

[12] As mentioned above, this interpretation is not accepted by some native speakers.

[13] Where the participle contains a tense suffix as in (7), the relative clause is ambiguous with respect to tense. Where the participle is made up of two words, one with a lexical verb and the other with a buffer stem, the full array of tense specifications can be expressed. See Göksel (2001) for details.

[14] Note that this example, example (42) in Barker et al., has a genitive suffix on the predicate of the embedded clause, which, for the native speakers I have consulted and for myself, is ungrammatical. I have therefore taken the liberty of using my dialect in (8a), which does not affect the analysis here.

b. [[biz-im ___$_i$ güven-eceğ-imiz-in] şüpheli ol-du]-**ğ**-u adam$_i$
we-GEN trust-COMP-1PL.POSS-GEN doubtful be-T-NSR-3SG.POSS man
'the man that it is doubtful we will trust'

The head of the relative clause *adam* 'man' is the oblique object of *güven* 'trust', the predicate of the clause which is the sentential subject of *şüpheli ol* 'is doubtful'. This pair is particularly pertinent to the claim made here regarding the difference between FRCs and PRCs, and we shall return to this example shortly.[15]

2.2. Structure of PRCs

It is tempting to think of PRCs as FRCs with deleted lexical heads. After all they both seem to share the same suffixes, i.e. the plural marker and a possessive marker. Indeed, the morpheme *-lAr* in PRCs seems to be 'left over' from the deleted head of an FRC:

(9) a. (ben-im) $_j$___$_i$ (OBJ) *gör-dü-ğ-üm* $_j$ [konuk-lar]$_i$ (FRC-NSR)
I-GEN see-T-NSR-1SG.POSS guest-PL
'the guests who I saw'

b. (ben-im) ___$_i$ (OBJ) *gör-dü-k-ler*$_i$*-im* (PRC-NSR)
I-GEN see-T-NSR-LAR-1SG.POSS
'those who I saw'

(9a) is an example of a non-subject relative clause with a plural head. The same parallelism in (9) is found between subject FRCs with plural heads and subject PRCs:

(10) a. ___$_i$ (SUB) ben-i *gör-en* [konuk-lar]$_i$ (FRC-SR)
I-ACC see-SR guest-PL
'the guests who saw me.'

b. ben-i ___$_i$ (SUB) *gör-en-ler*$_i$ (PRC-SR)
I-ACC see-SR-LAR
'those who saw me.'

FRCs that have singular heads and PRCs with a singular interpretation also seem formally identical, irrespective of whether they use the NSR strategy as in (11) or the SR strategy as in (12):[16]

[15] Various analyses have been proposed for the syntax of Turkish FRCs, among which are Hankamer & Knecht (1976), Kornfilt (1984a), Kornfilt (1997), Csató (1985), Barker et al (1990), Özsoy (1994a) Özsoy (1994b), Haig (1997), Erkman-Akerson & Ozil (1998), Çağrı (2005), Ulutaş (2005). Here I shall not give an evaluation of these analyses. Testing these against the data provided here for PRCs might prove to favour one of them over the other, but such an undertaking is outside the scope of this paper.

[16] -Ø refers to cases where the lack of a plural suffix indicates singularity, evidence for which is given in section 6. This is not the only interpretation of forms without *-lAr*, among which are transnumeral, categorial and indefinite interpretations, these being relevant also to forms containing *-lAr*. Since what interests us here are the formal properties of pronominal participles, the various interpretations of either *-lAr* or the lack of it are not relevant to the issues discussed here.

(11) a. (ben-im) ___$_{i\ (OBJ)}$ *gör-dü-ğ-üm* [konuk-Ø]$_i$ (FRC-NSR)
I-GEN see-T-NSR-1SG.POSS guest-SG
'the guest who I saw'

b. (ben-im) ___$_{i\ (OBJ)}$ *gör-dü-ğ-Ø$_i$-üm* (PRC-NSR)
I-GEN see-T-NSR-Ø-1SG.POSS
'the one who I saw'

(12) a. ___$_{i\ (SUB)}$ ben-i *gör-en* [konuk-Ø]$_i$ (FRC-SR)
I-ACC see-SR guest-SG
'the guest who saw me'

b. ___$_{i\ (SUB)}$ ben-i *gör-en-Ø$_i$* (PRC-SR)
I-ACC see-SR-Ø
'the one who saw me'[17]

Indeed, there is little difference between FRCs and PRCs in terms of the grammatical function of the relativised head in the former case, and what the pronoun is coindexed with in the case of a PRC. A head noun in an FRC can stand in a direct object, oblique object or adjunct relationship with the verb in the relative clause. It can also be the complement of a postposition in the relative clause. Hence the form in (13a) can have an interpretation where *-lAr* is coindexed with the direct object gap as in (13bii), or where it has an oblique object/adjunct relationship with the predicate (i.e. where it is coindexed with the oblique object or adjunct gap), as in (13cii):

(13) a. sor-du-k-lar-ımız
ask-T-NSR-LAR-1PL.POSS

b. *Direct object*
(i) sor-du-k-lar-ımız
ask-T-NSR-LAR-1PL.POSS
'those that you ask'

(ii) ___$_{i\ (OBJ)}$ sor-du-ğ-umuz [soru-lar]$_i$
ask-T-NSR-1PL.POSS question-PL
'the questions that we ask'

---

[17] The head noun in FRCs (ia)-(iia) and the participles in PRCs (ib)-(iib) can also contain other markers (e.g. case suffixes and clitics).

FRC	*PRC*
NSR	
(i) a. (Ben-im) gör-dü-ğ-üm konuk-**lar-dan-dı**-Ø.	b. (Ben-im) gör-dü-k-**ler**-im-**den-di**-Ø.
I-GEN see-T-NSR-1SG.POSS guest-PL-ABL-P.COP-3SG	I-GEN see-T-NSR-**lar**-1SG.POSS-ABL-P.COP-3SG
'S/he was on of the guests whom I had seen.'	'S/he was one of the ones whom I had seen.'
SR	
(ii) a. ben-i gör-en konuk-**lar-ın-sa**	b. ben-i gör-en-**ler-in-se**
I-ACC see-SR guest-PL-GEN-CT	I-ACC see-SR-*LAR*-GEN-CT
'as for the guests who see/saw/have seen me'	'as for those who see/saw/have seen me'

c. *Oblique object/adjunct*
(i) (soru) sor-du-k-lar-ımız
(question) ask-T-NSR-LAR-1PL.POSS
'those whom we ask (questions)'

(ii) ___$_{i\ (OBJ/ADJ)}$ (soru) sor-du-ğ-umuz [polis-ler]$_i$
ask-T-NSR-1PL.POSS police.officer-PL
'the police officers whom we ask (questions)'

Similarly, *-lAr* can be coindexed with the oblique object gap where the predicate is a psychological verb, as in (14):

(14) *Oblique object*
a. ürk-tü-k-ler-imiz
fear-T-NSR-LAR-1PL.POSS
'those whom we fear'

b. ürk-tü-ğ-ümüz görevli-ler
fear-T-NSR-1PL.POSS official-PL
'the officials that we fear'

Only where *-lAr* refers to a place or temporal adverbial is a PRC marginal, indicating that there is a preference for LAR-POSS to refer to arguments in PRCs:

(15) a. (i) *? gör-ül-dü-k-ler-iniz
see- PASS-T-NSR-LAR-2PL.POSS
'those (places) at which you were seen'

(ii) gör-ül-dü-ğ-ünüz pastane-ler
see-PASS-T-NSR-2PL.POSS pastry.shop-PL
'the pastry shops at which you were seen'

b. (i) *? çarşı-ya git-ti-k-ler-iniz
market-DAT go-T-NSR-LAR-2PL.POSS
'those (times) at which you went to the market'

(ii) çarşıya git-ti-ğ-iniz gün-ler
market-DAT go-T-NSR-2PL.POSS day- PL
'the days on which you went to the market'

The assumption that PRCs are head-deleted versions of FRCs, however falls short of explaining certain properties of these constructions which ultimately set them apart from FRCs. The most important difference between the two is the function of the suffix *-lAr* (and *-Ø* which indicates singularity). For this reason, in the remainder of this

paper, the set *-lAr/Ø* which occurs in PRCs will be referred to as LAR to distinguish it from the set of the number suffixes *-lAr/Ø* in FRCs.[18]

3. Why PRCs Are not FRCs with Deleted Lexical Heads

There are a number of reasons why it does not seem to be the case that PRCs are reduced versions of FRCs where the lexical head has been deleted. An obvious piece of evidence comes from ordering facts. If the plural suffix on the head noun were to attach to the participle, the ordering would have been POSS-*lAr*, rather than the order that is attested, which is LAR-POSS, as indicated in (16a). If, on the other hand, the plural marker and the possessive were directly attached to the participle, the ensuing construction would have been POSS-*lAr*-POSS, which is also ungrammatical. These are illustrated in (16a) and (16b) respectively:

(16) a. * ara-dı-ğ-ımız ~~insan~~-lar
call-T-NSR-1PL.POSS ~~people~~-PL

b. * ara-dı-ğ-ımız ~~insan~~-lar-ımız
call-T-NSR-1PL.POSS ~~people~~-PL-1PL.POSS

We return to the constructions in (16) in section 3.4. But first we shall look at additional evidence in favour of separating PRCs from FRCs. These are listed below and elaborated in sections 3.1-3.4:

i. the nature of *-lAr/Ø* in FRCs as opposed to the properties of LAR in PRCs

ii. the asymmetry between referents of *lAr/Ø* in FRCs and LAR in PRCs

iii. the unavailability of a non-restrictive reading in PRCs

iv. the unavailability of the presence of a genitive-marked NP in PRCs

3.1. -lAr/Ø in FRCs vs. LAR in PRCs

One of the reasons why PRCs cannot be FRCs where the lexical head has been deleted has to do with the function of the seemingly identical suffixes *-lAr* and *-Ø* in these two types of clause. While in FRCs *-lAr/Ø* only marks number, in PRCs, this cannot be the sole function of LAR. Being a placeholder for the lexical head while at the same time indicating number, the members of LAR are more likely to be pronominal suffixes with number specification.[19] Hence I suggest that *-lAr* in PRCs is a 3rd person plural pronoun

[18] Hence *LAR* refers to the set of 3rd person pronominal suffixes, *-lAr* is the lexical form of *-ler* and *-lar* irrespective of their function, and *-ler* and *-lar* refer to the citation form of these.

[19] This is the reverse of the function of *-lAr* as described by Kornfilt (1984a: 52) as a member of (one set of) the verbal agreement paradigm. There Kornfilt identifies *-lAr* as the marker for number and not person. Here we identify it as the marker of number and 3rd person. Although Kornfilt's remark is intended for the verbal paradigm whereas LAR here is taken as a member of the nominal paradigm as will

and -Ø is a 3rd person singular pronoun, rather than plural and singular number suffixes respectively. One indication of the pronominal nature of LAR has to do with its mutual exclusivity with a lexical head. In FRCs the participle cannot contain *-lAr* even where the lexical head is plural:

(17) a. *gör-dü-k-ler-im konuk(-lar)
see-T-NSR-LAR-1SG.POSS guest(-PL)
Int. interpretation: 'the guests whom I saw'

b. *gör-en-ler konuk(-lar)
see-SR-LAR guest(-PL)
Int. interpretation: 'the guests who saw me'

The observation that the members of LAR are pronominal suffixes specified for number is supported by other data. *-lAr* can attach to simple adjectives with the result being ambiguous between *number* specification as in (18ai) and *number+pronominal* as in (18aii). Further, the addition of a possessive suffix disambiguates these constructions as shown in (18b) which has only a number+pronominal interpretation. *-lAr* in forms such as (18b) does not quantify over events or states:

(18) a. kırmızı-lar
red-*lAr*
(i) 'types/shades/varieties of (the colour) red
(ii) 'the red ones'

b. kırmızı-lar-ı
red-*lAr*-3SG.POSS
'the red ones [of…]' (as in *bunların kırmızıları daha tatlı oluyor*
'the red ones [of these] are sweeter)

Note that the sequence in (18b) is exactly what is found in PRCs and, as expected, these also have a single interpretation:

(19) ara-dı-k-lar-ımız
call-T-NSR-LAR-1PL.POSS
(i) *'the one we called many times'
(ii) 'the ones we called'

3.2. The Reference of lAr/Ø in FRCs and LAR in PRCs

It was noted above in section 2.1 that the two strategies of relativisation, namely SR and NSR, converge when the head of the relative clause corresponds to a gap inside a sentential subject (see 8). The result of this convergence was that whichever strategy was used, the interpretation was the same. This does not carry over to PRCs and only the SR strategy in this case is grammatical:

---

be discussed in section 5.3, below, whether these two claims can exist side by side is a matter for further research.

(20) a. [[biz-im ___$_{i(OBJ)}$ güven-eceğ-imiz] şüpheli ol]-**an**-lar$_i$
we-GEN trust-COMP-1PL.POSS doubtful be-SR-LAR
'the ones that it is doubtful we will trust'

b. *[[biz-im ___ $_{(OBJ)}$ güven-eceğ-imiz-in] şüpheli ol-du]-**k**-lar-ı
we-GEN trust-COMP-1PL.POSS-GEN doubtful be-T-NSR-LAR-3SG.POSS
Int. interpretation: 'the ones that it is doubtful we will trust'

If PRCs were FRCs without lexical heads there would be no reason for such an asymmetry. If it were the case that the suffix *-lAr* on the lexical head straightforwardly got realised on the participle, (20b) should have been grammatical on a par with (20a).

3.3. The Unavailability of a Non-restrictive Reading in PRCs

Although non-restrictive relative clauses are not common in Turkish, FRCs can be ambiguous between a restrictive and a non-restrictive reading. PRCs, on the other hand, only allow a restrictive reading:

(21) a. ___$_{i\ (OBJ)}$ çok sev-di-ğ-im [komedi-ler]$_i$ (FRC)
much like-T-NSR-1SG.POSS comedy-PL
(i) 'the comedies that I like very much'
(ii) 'comedies, which I like very much'

b. ___$_{i\ (OBJ)}$ sev-di-k-ler$_i$-im (PRC)
like-T-NSR-LAR-1SG.POSS
(i) 'the ones that I like'
(ii) *'these that I like'

(22) a. ___$_{i\ (OBJ)}$ hiç kaçır-ma-dı-ğ-ım [komedi-ler]$_i$ (FRC)
never miss-NEG-T-NSR-1SG.POSS comedy-PL
(i) 'the comedies that I never miss'
(ii) 'comedies, which I never miss'

b. ___$_{i\ (OBJ)}$ hiç kaçır-ma-dı-k-lar$_i$-ım (PRC)
never miss-NEG-T-NSR-LAR-1SG.POSS
(i) 'the ones that I never miss'
(ii) *'these that I never miss'

Again, under the view that PRCs are head-deleted versions of FRCs, this is unexpected.

3.4. The Unavailability of Overt Genitive-marked NPs in PRCs

FRCs and PRCs also differ in terms of the genitive NP functioning as the specifier in constructions that have possessive markers. Possessive markers are associated with genitive NPs and are considered to be their licensors. This is exemplified by an FRC

below where the genitive NP agrees with the possessive marker on the head of the relative clause. This construction contains two possessive markers, one on the lexical head which licenses the genitive NP, the other one on the participle, referring to the subject of the relative clause:[20]

(23) Tolstoy-**un** sık sık oku-du-ğ-**um** roman-lar-**ı**
Tolstoy-GEN often read-T-NSR-**1PL.POSS** novel-PL-**3SG.POSS**
(i) those of Tolstoy's novels that I often read
(ii) Tolstoy novels, which I often read

However, although genitive NPs occur in FRCs, they cannot be overtly expressed in PRCs:

(24) a. *Tolstoy'**un** sık sık oku-du-k-lar-**ım-ı**
Tolstoy-GEN often read-T-NSR-LAR-**1PL.POSS-3SG.POSS**
Int. interpretation: 'those of Tolstoy which I often read'

b. *Tolstoy'**un** sık sık oku-du-k-lar-**ım**
Tolstoy-GEN often read-T-NSR-LAR-1PL.POSS
Int. interpretation: 'those of Tolstoy which I often read'

As can be seen, (24a) which contains the same suffixes as (23) is ungrammatical. The ungrammaticality of this form can be explained on the grounds that the participle is not a well-formed word, a point we shall return to in section 5.2. (24b) has a morphologically well-formed participle, but the construction is still not well-formed as it does not have a possessive marker that licenses the genitive NP. So in PRCs, there is no possibility of expressing a genitive NP which is in agreement with the possessor of the denotation of the head noun. In other words, there is no possibility of construing the possessive on the participle as being copied from the lexical head. Hence, neither of the markers on the participle of the PRC corresponds to the markers on the head of the FRC.

These facts show that the participle in the PRC is not a combination of an FRC participle with the suffixes *-lAr* and POSS copied from the head noun.

[20] It is interesting to note that the overt expression of two genitive NPs, one licensed by the possessive marker on the lexical head, the other by the possessive marker on the participle sounds rather contrived:

(i) ?Tolstoy-**un**$_i$ ben-**im**$_j$ sık sık oku-duğ-**um**$_j$ roman-lar-**ı**$_i$
Tolstoy-GEN I-GEN often read-NSR-**1PL.POSS** novel-PL-**3SG.POSS**
(a) those of his novels that I often read
(b) his novels, which I often read

This may be related to the topic status of pronominal genitive NPs, see Enç (1986) and Öztürk (1999), since corresponding constructions where the genitive NPs are omitted are grammatical, see (32).

4. Person Marking in PRCs (SR Strategy)

A further difference between FRCs and PRCs manifests itself in the SR strategy of relativisation. It is a well-known property of the participles of FRCs that they cannot contain person markers when the SR strategy is used. The marker of the SR strategy, which is *-(y)An*, excludes the occurrence of a person marker in FRCs. It should be remembered that person marking is realised through possessive markers as a result of the nominal nature of participles.

(25) biz-i gör-en(*-imiz) kişi-ler (FRC)
we-ACC see-SR(*-1PL.POSS) person-PL
'the people who see us'

The unavailability of possessive marking in FRC participles does not carry over to PRCs. Participles of PRCs where the SR strategy has taken place *can* have possessive marking:

(26) gör-en-ler-imiz (PRC)
see-SR-LAR-1PL.POSS

As mentioned in section 1 above, these constructions are ambiguous where LAR and POSS can have the following interpretations:

(26')	LAR	POSS
a.	subject	non-subject
b.	part	whole (subject)
c.	part	whole (non-subject)

We shall now take each one of these in turn.

4.1. LAR: Subject, POSS: Non-subject

The possibility of interpreting LAR as being coindexed with the gap in subject position and POSS as the non-subject is concomitant with the predicate of the relative clause receiving an individual-level interpretation. In (27) below, the verb *gör* 'see' denotes a permanent state of 'seeing/visiting' (e.g. as a habit), rather than the stage-level interpretation of 'seeing' as a temporary action:[21]

(27) ___$_i$ (SUB) ___$_j$ (OBJ) gör-en-ler$_i$-imiz$_j$
see-SR-LAR-1PL.POSS
'the ones who see us'

[21] See Diesing (1992) for the differences between these predicate types.

The fact that these predicates do not allow temporal specification further supports this point.[22] The non-subject in these constructions is usually a direct object, but oblique objects are also marginally acceptable.

4.2. LAR: Part, POSS: Whole (Subject)

In the second instance where the participle of a PRC contains a possessive suffix, it gives part of the information relating to the gap in the subject position of the relative clause:

(28) ___$_{i\ (SUB)}$ Amerika-yı gör-en-[ler-imiz]$_i$
America-acc see-SR-LAR-1PL.POSS
'*those among us* who have seen America'

The full interpretation of the gap hinges on the presence of *-lAr* which selects a subset of this group of 1st, 2nd or 3rd persons. As such, it is still a plural pronoun referring to a group. Hence in (28) *-lAr* indicates a non-singleton subset of a set whose person specification is marked by the possessive marker. Thus, these two suffixes jointly define the identity of the gap that they are coindexed with. Note that (27) and (28) are distinguished from each other only by the absence of a direct object in the former.[23]

4.3. LAR: Part, POSS: Whole (Non-subject)

The third interpretation associated with (26) above is similar to the one in (28) in terms of the function of LAR. Again LAR denotes a subset, and the possessive marker gives the value of the group in terms of its person specification but this time LAR-POSS is coindexed with a gap that is the non-subject, rather than the subject. This is given in the first interpretation below. This interpretation is accepted only by some speakers and even then is possible only where a number of conditions are met. It occurs only with a handful of verbs denoting an aggressive action perpetrated by non-human agents, such as *ısır* 'bite', *sok* 'sting' and *tırmala* 'scratch' and is contingent on the presence of a bare NP subject which is categorial/generic:[24]

(29) ___$_{i\ (OBJ)}$ köpek ısır-an-[lar-ımız]$_i$
dog bite-SR-LAR-3PL.POSS
'*those among us* who dogs bit'

[22] These constructions with the given reading are somewhat similar to deverbal nouns containing one of the deverbalizing suffixes *-(A/I)r* (as in *oku-r* 'reader') or *-(y)IcI* (as in *oku-yucu* 'reader'), which might seem suggestive of a lexical phenomenon. Attributing the phenomenon to the lexicon, however, does not change matters in terms of explaining the internal make-up of these participles.

[23] Examples such as (28) can contain a possessive-marked adjunct coreferential with LAR-POSS, see Göksel & Kerslake (2005: 283)

[24] See Öztürk (2005) for a recent analysis of such subjects as pseudo-incorporated NPs.

Note that (29) has a second and universally acceptable interpretation parallel to the one discussed in relation to (28) above, where LAR-POSS is coindexed with the subject gap.

The revised summary of the suffixes in pronominal participles in line with what has been discussed above is as follows:

(30)

VERB	-REL	-LAR	-POSS
	-K-	3rd p. non-subject pronoun	subject agreement/pronoun
	-(y)An	3rd p. subject pronoun	object agreement/pronoun
	-(y)An	pronoun denoting part	pronoun denoting whole (subject)
	-(y)An	pronoun denoting part	pronoun denoting whole (non-subject)

5. Morphological Aspects of the Structure of Pronominal Participles

What do the data discussed above tell us about the factors that underlie the structure of the participles of PRCs? First, from an empirical point of view, this is the only instance in Turkish where a verb form contains an affix that corresponds to a constituent other than the subject. Non-subject marking on a verbal form, finite or non-finite, is otherwise unattested in Turkish. As for the issue of whether pronominal markers exist in Turkish at all, the reader is referred to arguments put forth in Enç (1986) and Öztürk (1999) regarding the pronominal nature of the agreement morphemes in Turkish. A question to be raised at this point would be why LAR, the set of pronominal affixes, only contains 3rd person forms and not 1st and 2nd person forms, given that person paradigms in Turkish contain, with the exception of the imperative paradigm, forms for all persons. This will be discussed below in 5.3. But first we shall look at two factors that are crucial for an understanding of the effect of morphology as displayed by participles: the fixed order of affixes and the fixed amount of space allocated to affixation in a word.

5.1. Fixed Order

One point which is obvious from the data, yet has not been pointed out explicitly so far, is the fixed order of LAR and POSS. Any other order is ungrammatical:

(31) *gör-dü-ğ-üm-ler
see-T-NSR-1SG.POSS-LAR
Int. interpretation: 'the ones that I have seen'

The restrictions on the ordering of the plural morpheme *-lAr* with respect to the possessive markers are well-known (Lees 1962, Kornfilt 1984a, Kornfilt 1984b, Göksel 1988, Schroeder 1999). Not only does it have to occur before the possessive markers, it also cannot occur twice in the same word even if it is required for semantic reasons. It is also well-known that *-lAr* has various semantic functions whether it be in the verbal paradigm or the in nominal paradigm (Yükseker 1995, Ketrez undated, Göksel and Kerslake 2005). The observation here adds to the diversity of the functions of this morpheme. The fact that it is, in this case, a member of the pronominal set LAR does

not affect its positioning, even when it can correspond to various grammatical functions such as the subject, non-subject or a part of either of these. Similarly POSS, from its fixed position, can correspond to the subject, non-subject or the set from which the denotation of either of these is chosen. In addition, LAR and POSS can be coindexed with gaps separately as in (2a-b) or jointly, as in (2c-d). These are serious counterexamples to approaches where the morphological ordering of affixes has a one-to-one correspondence to syntactic operations as in Baker (1985), among many others.

5.2. Fixed Size

Another significant property of the participles of PRCs is that they can only have one POSS marker, although double person marking in FRCs is possible. By virtue of having an overt lexical head, FRCs have two locations on two separate words for expressing person marking. One of these positions, the one on the lexical head, denotes possession and the other one which is located on the participle itself denotes agreement with the subject of the relative clause:[25]

(32) a. gör-dü-ğ-**ün** bütün kitap-lar-**ım**
see-T-NSR-**2SG.POSS** all book-PL-**1SG.POSS**
'all my books that you have seen'

The iteration of two possessive markers on the participle of a PRC is ungrammatical:[26]

(33) *gör-dü-k-ler-**im-in**
see-T-NSR-LAR-**1SG.POSS-2SG.POSS**
Int. interpretation: (i) 'all the ones of mine that you have seen'
(ii) 'all the ones of yours that I have seen'

Thus, it is not possible to express both of the functions associated with possessive marking on the same form concurrently.

As a result of the conditions requiring affixes to appear in a fixed order and within a fixed space (see also Göksel 1998, Göksel 2001), morphemes may be associated with various functions yet still appear in fixed positions. This in itself is an indication that syntactic requirements may not override formal conditions on word structure, a point which will be elaborated in section 8.

5.3. Nominal Paradigm Functioning as Verbal Paradigm

A final indication that morphological constraints override syntactic constraints has to do with the class membership of LAR. The suffix *-lAr* has a ubiquitous character and appears in more than one paradigm, given below:[27]

[25] As mentioned above, FRCs have possessive marking on the participle only in the NSR strategy.
[26] See also Kornfilt (1984b) and Inkelas & Orgun (1998).
[27] For the occurrence of *-lAr* in other paradigms see Göksel & Kerslake (2005). For the functions of *-lAr*, see Lewis (1968), Kirchner (2001), Ketrez (undated), among others. For the usage of possessive morphemes see Schroeder (1999).

(34)

	(i) *nominal agreement paradigm (possessive markers)*	*(ii) verbal agreement paradigm (participle group)*	*(iii) number paradigm*
			sg. **-Ø**
			pl. **-lAr**
1sg	-(I)m	-(y)Im	
2sg	-(I)n	-sIn	
3sg	**-(s)I(n)**	**-Ø**	
1pl	-(I)mIz	-(y)Iz	
2pl	-(I)nIz	-sInIz	
3pl	**-lArI**	**-lAr**	

Let us remind ourselves that participles are nominal constructions. It would therefore be expected for LAR to belong to the nominal agreement paradigm. However, it does not. The forms for third person in (i) (i.e. *-(s)I(n)* and *-lArI*) do not match the members of LAR (namely *-lAr* and **-Ø**). What about (ii)? Here there is a match between the members, however, two points would then be inexplicable. One of these is why, given that LAR has a pronominal function, we see only the forms *-lAR* and -Ø in PRCs and none of the other persons. Although rare in Turkish, heads of relative clauses can be 1st or 2nd person pronouns. In principle then, there would be no reason why a 1st or a 2nd person pronominal affix should not appear in a PRC. But, as mentioned in section 3 above, this does not happen:

(35) *gör-dü-ğ-ünüz-ümüz
see-T-NSR-2PL.POSS.1PL.POSS
Int. interpretation: (i) 'you (all) who we have seen'
(ii) 'we, who you (all) have seen'

It would also be unexpected to have verbal agreement suffixes attaching to a nominal stem, ruling out (ii) as the possible paradigm that LAR belongs to.

The only paradigm that is left as a possible option is thus (iii), the number paradigm. What is important to note is that this paradigm is selected for purely morphological reasons. The possible order of any type of inflection following a nominal stem is number-possessive:

(36) kitap-lar-ım
book-PL-1SG.POSS
'my books'

This is exactly what occurs on any type of nominal, including participles that we have been discussing.

The properties of PRCs discussed above show that they are subject to principles of morphological well-formedness overriding syntactic factors relating to the expression of grammatical functions. These points indicate that word structure is subject to independent principles of morphology. So far, syntax has had no say in the internal make-up of participles. Except that there are two instances where syntax does seem to play a role. One of these is the incremental nature in the ordering of the affixes with

respect to the relativisers *-K-* and *-(y)An*, and the other one is the behaviour of participles under suspended affixation, to which we now turn.

6. Syntactic Effects in the Word Domain

Up to now we have looked at ways in which morphological constraints shape the structure of participles in PRCs in terms of defining the positions for the expression of grammatical functions. The area of the word covered by LAR-POSS seems to be opaque to syntactic operations. If we move one step to the left and include the relativiser into the picture, syntactic effects do seem to play a role.

There are two aspects of participial forms which point to the possibility of syntactic intervention. One of these has to do with the combinatorial properties of the relativisers and constitute examples of how syntactic mechanisms affect word structure. The other one, the behaviour of pronominal participles under suspended affixation does not give clear results with respect to how or whether syntax has access to word structure. We look at each one of these in turn.

6.1. Combinatorial Properties of the Relativisers

There is a striking dissimilarity between the combinatorial properties of the relativisers and how they induce an interpretation on the following suffix. (30), repeated below, illustrates this point:

(37)

VERB	-REL	-LAR	-POSS
(i)	-K-	3rd p. non-subject pronoun	subject agreement/pronoun
(ii)	-(y)An	3rd p. subject pronoun	object agreement/pronoun
(iii)	-(y)An	pronoun denoting part	pronoun denoting whole (subject)
(iv)	-(y)An	pronoun denoting part	pronoun denoting whole (non-subject)

As shown in the table above, the suffix following the NSR relativiser *-K-* is interpreted as a non-subject. *-K-* which, by virtue of being a relativiser introduces an operator and a gap, requires that the value of this gap be assigned a partial interpretation (partial in the sense that it is a pronominal item) immediately.

However, there are two counterexamples to this generalisation and not all affixes that follow *-K-* are interpreted as objects. Firstly, FRC participles typically contain *-K-*POSS sequences where POSS is obligatorily interpreted as the subject, as in (7) repeated below:

(7) (Semra-nın)$_j$ ___$_{i\ (OBJ)}$ sev-di-ğ-i$_j$ çiçek-ler$_i$
Semra-GEN like-T-NSR-3SG.POSS flower-PL
'the flowers that Semra likes'

Secondly, the interpretation of LAR-POSS is affected by the complexity of the clause. In multiple embeddings, LAR can be interpreted as the subject (38b) or as the direct object (39b) of a sentential complement:

(38) a. ben-im$_j$ [____$_{i(SUB)}$ kitab-ı oku-du-ğ-un$_i$-u] san-dı-ğ-ım$_j$ [adam-lar]$_i$[28]
I-GEN book-ACC read-T-NSR-3SG.POSS think-T-NSR-1SG.POSS man-PL
'the men who I thought read the book'

b. ben-im$_j$ [____**$_{i(SUB)}$** kitab-ı oku-du-ğ-un$_i$-u] san-dı-k-**lar**$_i$-ım$_j$
I-GEN book-ACC read-T-NSR-3SG.POSS think-T-NSR-*LAR*-1SG.POSS
'the ones who I thought read the book'

(39) a. ben-im$_j$ [adam-ın____$_{i(OBJ)}$ oku-du-ğ-un$_i$-u] san-dı-ğ-ım$_j$ [kitap-lar]$_i$
I-GEN man-GEN read-T-NSR-3SG.POSS think-T-NSR-1SG.POSS book-PL
'the books that I thought the man read'

b. ben-im$_j$ [adam-ın ____**$_{i(OBJ)}$** oku-du-ğ-un$_i$-u] san-dı-k-**lar**$_i$-ım$_j$
I-GEN man-GEN read-T-NSR-3SG.POSS think-T-NSR-*LAR*-1SG.POSS
'the ones that I thought the man read'

What LAR following an NSR participle cannot refer to is a constituent inside a sentential subject, witnessed by the ungrammaticality of (20b) above.[29] The NSR relativiser then does not automatically constrain the grammatical function of the affix that follows it. Rather, it is sensitive to the presence of *a pair of* affixes. This means that the grammatical function of a suffix that follows NSR is not predictable unless these affixes form some kind of a template.

The combinatorial properties of *-(y)An* are less obvious. The affix to the right of *-(y)An* is not interpreted as a non-subject under any circumstances. It is either interpreted as the subject, as in (37ii), or it is assigned an interpretation which has nothing to do with grammatical functions, as in (37iii-iv). Hence, *-(y)An* also seems to be sensitive to the presence of a pair of affixes, although it not in the same way as *-K-*. The difference is that the affix adjacent to *-(y)An* can never be identified as the non-subject. This can further be supported by a form such as (38), which can only be analysed as having a *-Ø* affix which occupies a position but is not phonologically overt. An analysis which does not posit a null affix would fall short of accounting for its plural counterpart in the interpretation where the understood subject is plural:

(40) a. sev-en-Ø-im
love-SR-Ø(=subject)-1SG.POSS(= object)
'the one who loves me'

b. sev-en-ler-im
love-SR-ler(=subject)-1SG.POSS(= object)
'those who love me'

These observations can be summarised as follows:

[28] I would like to thank Cem Bozşahin for bringing this example to my attention.
[29] The asymmetry between extraction from a subject and object is discussed in Hankamer and Knecht (1976). The data here is in conformity with the analysis presented there.

(i) The combinatorial properties of the relativisers constrain the interpretation of the suffixes that follow them in terms of what grammatical function they may fulfil, hence word-internal items must have access to syntactic information

(ii) However such syntactic information is sensitive to whether the affixes that follow a relativiser appear as part of in a pair or not

These points indicate that the accessibility of PRC participles to syntactic information is contingent on the presence of a morphological template.

7. Co-ordination under Suspended Affixation

The next piece of data, the co-ordination of word internal elements, give mixed results with respect to whether co-ordination, in so far as it is considered a syntactic phenomenon, is a reliable test for understanding the opacity of word structure to syntax.[30] In suspended affixation (Lewis 1968), identical suffixes on all but the last of consecutive co-ordinated constituents can be deleted in Turkish:[31]

(41) dere ve ırmak-lar-da
stream and river-PL-LOC
'in streams and (in) river[s]'

The relevant insight that suspended affixation can provide with respect to the data at hand is that constituents ending in *-(y)An* behave differently from those ending in *-K-*. This is elaborated below.

7.1. The Constituent Ending in -K- Cannot Be 'Severed'

When a co-ordinator such as *ve* 'and' is used, SR participles can be co-ordinated by suspending LAR-POSS as in (40a) but not NSR participles:[32]

(42) a. [sev-en ve anla-yan]-lar-ımız
love-SR and understand-SR-LAR-1PL.POSS
(i) 'those who love us and those who understand us'
(ii) 'those who love and understand us'

[30] Bresnan and Mchombo (1995) take co-ordination as a test for the lexical integrity of words containing derivational morphemes. I It is not clear whether co-ordination can be applied as a reliable test for words containing inflectional morphemes, especially in view of the fact that inflection may create separate word domains (see Kabak and Vogel 2001). However, the asymmetry in pronominal participles with respect to co-ordination is an interesting phenomenon which may bring more light to the notion of word and issues relating to lexical integrity.

[31] Suspended affixation is subject to certain conditions (for various views see Lewis (1968), Orgun (1995), Kornfilt (1996), Inkelas & Orgun (1998), Kabak (2006)). As discussed in these works, the point of cut-off on the conjuncts is not arbitrary and is subject to certain conditions. We shall touch upon these below but will not provide a full rendition of the analyses in these works for reasons of space.

[32] Similar observations are made in Kornfilt (1984a:149) for sentential complements ending in *–K-* (*-DIK* in her terms) and conjoined with the co-ordinating clitic *-(y)lA*.

b. *[sev-dik ve anla-dı-k]-lar-ımız
love-NSR and understand-T-NSR-*LAR*-1PL.POSS
Int. interpretation: 'those who we love and understand'

Constituents carrying NSR and SR can also be co-ordinated as long as the 'severed' constituent is an SR participle:

(43) [ben-i sev-en ve ben-im gör-dük]-ler-im
I-ACC love-SR and I-GEN see-NSR-LAR-1SG.POSS
'those who love me and who I have seen'

If the severed constituent is an NSR participle, the construction is ungrammatical:

(44) *[ben-im gör-dük ve ben-i sev-en]-ler-im

This observation supports the analysis presented in Kabak (in press), namely that the main condition that applies to suspended affixation is that the severed form must be 'word' (see also Ido 2003 for relevant discussion).

7.2. Suspended Affixation is only Sensitive to the Formal Properties of Affixes

The possibility of co-ordinating NSR and SR participles shows that suspended affixation is not sensitive to the identity of the grammatical functions of the suspended affixes, and is an operation which is only sensitive to the formal properties of affixes. The availability of co-ordination in SR and NSR participles above shows that the operation takes place irrespective of the functions of the affixes. The co-ordination of SR participles where LAR and POSS do not have unique functions further supports this claim.

(45) [orman-da gez-en ve aslan ısır-an]-lar[33]
forest-loc stroll-sr and lion bite-sr-lar
'those who were strolling through the forest and who lions bit'

In the example above, *-lar* is interpreted as the object in the second conjunct but the recovered interpretation for the first and severed conjunct is as the subject of the predicate.

[33] I would like to thank İlhan Çağrı for bringing this example to my attention. This example is grammatical on the intended interpretation only for some speakers, see FN 3.

7.3. Suspended Affixation can 'Skip' an Affix

Another difference between SR and NSR constructions has to do with the inability of an NSR participle to be severed at the point where it ends in the suffix *–K-*, as mentioned above. This requirement forces the presence of other material on the participle for grammaticality. Interestingly, this can lead to forms where affixes are not suspended as groups. In NSR constructions individually suspended affixes can occur. One such example is *-lAr* below, which is sandwiched between two unsuspended affixes:

(46) [dil-in-i bil-di-ğ-im ve anla-dı-k]-lar-ım
language-3SG.POSS-ACC know-T-NSR-1SG.POSS and understand-T-NSR-LAR-1PL.POSS
(i) 'those whose language I know and understand'
(ii) 'the one whose language I know and those I understand'

Notice that *-lar* is missing from the first conjunct, yet the possibly more prominent interpretation of the construction is the one where it is interpreted as occurring there. The facts with SR participles are not the same. The SR construction parallel to (46) only has the interpretation where *-lAr* is *not* interpreted as part of the first conjunct.

(47) [aslan ısır-an-ımız ve arı sok-an]-lar-ımız
lion bite-SR-1SG.POSS and bee-SR-LAR-1PL.POSS
'the one lions/a lion bit and those who bees/a bee stung'

The accounts of suspended affixation do not account for the interpretation of suspended suffixes sandwiched between unsuspended ones. Firstly, the analysis whereby groups of affixes can either be suspended together or not at all does not apply to LAR and POSS, although it applies to other suffixes (contra. Orgun 1995, Inkelas & Orgun 1998).[34] It is also surprising that NSR participles cannot be co-ordinated using a free co-ordinator, although *-DIK* can occur word finally in lexicalised forms (as in e.g. *[tanıdık ve akadaş]larımız* 'our acquaintances and friends'): However, the data suggest that the condition put forth by Kabak (in press) regarding the presence of agreement suffixes as a condition for guaranteeing the well-formedness of the severed word seems to be supported by (46).

The transparency of the combinatorial properties of the relativisers as discussed in section 6.1 indicates that syntax is accessible to word structure. The ability of parts of participles to be conjoined under suspended affixation as shown in section 7, on the other hand indicate that although co-ordination is generally thought to be a syntactic process, here too the formal properties of the words come into play. The split behaviour of conjoined participles further indicates that the factors intervening in word structure do not present a uniform picture.

[34] As pointed out in Kabak (2006), these suffixes do not readily fall under an analysis of suspended affixation.

8. Conclusion: The Interaction of Morphology and Syntax

It has been argued above that relative clauses with pronominal participles are structurally different from full-fledged relative clauses, leading to the observation that a unitary syntactic structure for both is unwarranted. Further, it was shown that pronominal participles have the following morphological structure:

i. The grammatical functions of the suffixes following NSR and SR cannot be predicted by isolating them. The suffix following NSR can be coindexed with a subject (9a), an object (9b), an embedded subject (38b) and an embedded object (39b). The suffix following SR can be coindexed with the gap of a subject (27), a part of a subject (28) and that of an object (29).

ii. The grammatical functions of the suffixes following NSR and SR are contingent on whether such suffixes occur in pairs or not.

iii. PRC participles have a fixed ordering of affixes, irrespective of their syntactic function.

iv. PRC participles have a fixed maximal size, irrespective of whether the expression of more functions is required syntactically.

v. PRC participles use affixes from the nominal paradigm irrespective of the fact that these fulfil syntactic functions.

The following questions can be raised concerning the nature of the interaction of morphology and syntax in pronominal participles and how this relates to the *Lexical Integrity Hypothesis*:

1. *What is the nature of the syntactic aspects of word structure?*

2. *What is the nature of the morphological constraints that determine the formal properties of words (i.e. fixed order of affixes, fixed size, and the form-function mismatch)?*

3. *How do the morphological constraints interact with syntactic constraints?*

The observations in (i) and (ii) relate to the first question above, the nature of syntactic intervention in word structure. The internal structure of PRCs show that the ordering of functional categories in a pronominal participle is only partially predictable and is sensitive to the difference between the occurrence of a single affix and a pair of affixes as discussed in section 6. Such a difference does not correspond to a property that can be conceptualised in terms of syntactic mechanisms. It therefore challenges hypotheses that posit an ordering constraint on affixes which is based on the ordering of syntactic operations. The prime example for such a hypothesis is the *Mirror Principle* (Baker 1985), but any analysis which incorporates head movement is likely to run into similar problems.

Turning to the first three points which have been elaborated in section 5, these taken together show that the formal properties of words are determined by mechanisms internal to morphology. What we mean by 'mechanisms internal to morphology' are the formal constraints on word structure imposed by neither of the phonological and syntactic components, nor by the lexicon. The fixed nature of the form of such words as pronominal participles indicates that there is a template which forms the basis for the expression of syntactic, lexical and phonological material, reminiscent of word structure that hosts position class affixes (cf. Stump 1992, Inkelas 1995).

Once these factors are taken into account, the interaction of morphology and syntax becomes clearer. Morphology as a separate system of rules (as suggested in Di Sciullo and Williams 1987, Ackema and Neeleman 2004, among others) is the source of providing the form of a word in terms of the space that is available. This space is then used for the expression of grammatical functions. If this is the case, word structure does seem to be transparent to processes external to morphology, but only after the morphological component provides the template in which these operations can take place. How the analysis of word structure in terms of a template fits in with morphophonological conceptions of the 'word' (cf. Kabak and Vogel 2001) is a research area yet to be explored.

The data thus show that the versions of the *Lexical Integrity Principle* (e.g. Bresnan and Mchombo 1995) where word structure is opaque to syntactic mechanisms is too strong. The partial transparency of word structure to the combinatorial specifications of its internal elements supports the weakening of this principle along the lines suggested in Booij (2005).

References

Ackema, Peter and Ad Neeleman (2004). *Beyond Morphology*. Oxford:OUP.

Anderson, Steven R. (1992). *A-morphous Morphology*. Cambridge, CUP.

Aronoff, Mark (1976*). Word Formation in Generative Grammar*. Cambridge (Mass.): The MIT Press.

Baker, Mark (1985). "The Mirror Principle and morphosyntactic explanation." *Linguistic Inquiry* 16, 373-415.

Baker, M. (1988). *Incorporation, A Theory of Grammatical Function Changing*. Chicago: University of Chicago Press.

Barker, Chris, Jorge Hankamer and John Moore (1990). "Wa and Ga in Turkish." In K. Dziwirek, P. Farrell and E. Mejías-Bikandi (eds.), *Gramatical Relations; A Cross-Theoretical Perspective*. Stanford: CSLI Publications, 21-43.

Booij, Geert (2005). "Construction-dependent morphology." *Lingue e Linguaggio* 2, 163-178

Bresnan, Joan and Sam A. Mchombo (1995). "The lexical integrity principle: evidence from Bantu." *Natural Language and Linguistic Theory* 13, 181-254.

Çağrı, İlhan (2005). "Turkish Relative Clauses." Ms., University of Maryland.

Csató, Éva Ágnes (1985). "A syntactic analysis of participle constructions in Modern Turkish." In *Beşinci Milletlerarası Türkoloji Kongresi*, İstanbul.

Di Sciullo, Anna-Maria and Edwin Williams (1987) *On the definition of Word*. Cambridge (Mass.): MIT Press.

Enç, Mürvet (1986). "Topic switching and pronominal subjects in Turkish." In D. I. Slobin and K. Zimmer (eds.), *Studies in Turkish Linguistics*. Amsterdam: John Benjamins, 195-208.

Erkman-Akerson, Fatma and Şeyda Ozil (1998). *Türkçe'de Niteleme: Sıfat İşlevli Yan Tümceler*. İstanbul: Simurg Yayınları.

Göksel, Aslı (1988). "Bracketing paradoxes in Turkish nominal compounds." In S. Koç (ed.), *Studies on Turkish Linguistics*. Ankara: METU, 287-298..

Göksel, Aslı (1997). "Morphological asymmetries between Turkish and Yakut." In K. İmer and N. E. Uzun (eds.), *Proceedings of the VIIIth International Conerence on Turkish Linguistics*. Ankara: Ankara Üniversitesi Basımevi.

Göksel, Aslı (1998). "Word Size." In G. Booij, A. Ralli and S. Scalise (eds*.), Proceedings of the First Mediterranean Conference on Morphology*. Patras: University of Patras,190-200.

Göksel, Aslı (2001). "The auxiliary verb ol at the morphology-syntax interface." In E. E. Erguvanlı-Taylan (ed.), *The Verb in Turkish*. Amsterdam: John Benjamins, 151-181.

Göksel, Aslı and Celia Kerslake (2005). *Turkish; A Comprehensive Grammar*. London: Routledge.

Haig, Geoffrey (1997). *Relative Constructions in Turkish*. Wiesbaden: Harrassowitz.

Hankamer, Jorge and Laura Knecht (1976). "The role of the subject/non-subject distinction in determining the choice of relative clause participle in Turkish." *Harvard Studies in Syntax and Semantics*, 2, Boston (Mass.), 197-219.

Ido, Shinji (2003). *Agglutinative Information: A study of Turkish Incomplete Sentences*. Wiesbaden: Harrassowitz.

Inkelas, Sharon (1995). "Nimboran position class morphology." *Natural Language and Linguistic Theory*, 11, 559-624.

Inkelas, Sharon and Orhan Orgun (1998). "Level (non)ordering in recursive morphology: Evidence from Turkish." In S. G. Lapointe, D. K. Brentari and P. M. Farrell (eds.) *Morphology and its Relation to Phonology and Syntax*. Stanford: CSLI Publications, 360-410.

Jackendoff, Ray (1975). "Morphological and semantic regularities in the lexicon." *Language* 51, 639-671.

Kabak, Barış (in press). "Turkish suspended affixation". (To appear in *Linguistics*.)

Kabak, Barış and Irene Vogel (2001). "The phonological word and stress assignment in Turkish." *Phonology* 18, 315-360.

Kennelly, Sarah (1996). "Turkish Subordination: [-CP, -Tense, +Case]." In A. Konrot (ed.), *Studies in Turkish Linguistics*. Eskişehir: Anadolu Üniversitesi, 55-75.

Ketrez, Nihan (undated). "Multiple readings of the plural morpheme in Turkish." Ms. USC.

Kelepir, Meltem (2007). "Copular forms in Turkish, Turkmen and Noghay." (Kelepir, M and B. Öztürk (eds.), *Proceedings of the 2nd Workshop on Altaic Formal Linguistics*, 11-13 October 2004, Boğaziçi University. Cambridge (Mass.): MIT Working Papers in Linguistics.

Kirchner, Mark (2001) "Plural agreement in Turkish." *Turkic Languages* 5. 216-225.

Kornfilt, Jaklin (1984a). *Case marking, agreement and empty categories in Turkish*. Doctoral dissertation, Harvard University.

Kornfilt, Jaklin (1984b). "The stuttering prohibition and morpheme deletion in Turkish." In A. Aksu-Koç and E. Erguvanlı-Taylan (eds.), *Proceedings of the Turkish Linguistics Conference*. Istanbul: Boğaziçi University Publications, 59-84.

Kornfilt, Jaklin (1996). "On copular clitic forms in Turkish." In A. Alexiadou, N. Fuhrfop, P. Law and S. Löhken (eds.), *ZAS Papers in Linguistics* 6, Berlin, Zentrum für Allgemeine Sprachwissenschaft, 96-114.

Kornfilt, Jaklin (1997). "On some constraints governing free relative clauses in Turkish." In K. İmer and N. E. Uzun (eds.), *Proceedings of the VIIIth International Conerence on Turkish Linguistics*. Ankara: Ankara Üniversitesi Basımevi, 17-24.

Kural, Murat (1993). "V-to(-I-to)-C in Turkish." *UCLA Occasional Papers in Linguistics* 11.

Lees, Robert B. (1962). "A compact analysis for the Turkish personal morphemes." In N. Poppe (ed.), *American Studies in Altaic Linguistics*. Bloomington (Ind.): Indiana University Publications, 141-176

Lewis, Geoffrey (1968). *Turkish Grammar*. Oxford: OUP.

Mohanan, Tara (1995). "Wordhood and Lexicality: Noun Incorporation in Hindi." *Natural Language and Linguistic Theory* 13, 73-134.

Orgun, C. Orhan. (1995). "Flat versus branching morphological structures: The case of suspended affixation." In *Proceedings of the Berkeley Linguistic Society*, 21. Berkeley (Ca.): Berkeley Linguistic Society, 252-261.

Özsoy, A. Sumru (1994a). "Türkçe'de ortaç yapısı." *Dilbilim Araştırmaları*, 21-30.

Özsoy, A. Sumru (1994b) "Relative clauses in Turkish." In L. Johanson, in cooperation with É. Á. Csató, V. Locke, A. Menz and D. Winterling (eds.), *The Mainz Meeting, Proceedings of the Seventh International Conference on Turkish Linguistics*. Wiesbaden: Harrassowitz.

Öztürk, Balkız (1999). "Turkish as a non-pro-drop language." In Emine Eser Erguvanlı-Taylan (ed.), *The Verb in Turkish*. Amsterdam: John Benjamins.

Öztürk, Balkız (2005). *Case, Referentiality and Phrase Structure*. Amsterdam: John Benjamins.

Schroeder, Christoph (1999). *The Turkish Nominal Phrase in Spoken Discourse*. Wiesbaden: Harrassowitz.

Stump, Gregory (1992). "Position classes and morphological theory." In G. Booij and J. van Marle (eds.), *Yearbook of Morphology*, 129-180.

Tekin, Şinasi (2001). *İştikakçının Köşesi*. Istanbul: Simurg Yayınları.

Ulutaş, Süleyman (2005). *Feature percolation: Evidence from Turkish Relative Clauses*. Paper presented at the Linguistics Seminar.

Yükseker, Hitay (1995). "How many '-ler/lar's are there in Turkish." In P. Koskinen (ed.), *Proceedings of the 1995 Annual Conference of the Canadian Linguistic Association*. Toronto: University of Toronto, 638-649.

External Modifiers in Georgian

Alice C. Harris*
SUNY Stony Brook
alice.harris@stonybrook.edu

Abstract

This paper addresses the issue of stranded modifiers and null heads through two otherwise unrelated constructions in Georgian. In each construction, a word in the oblique form modifies part of the complex word following it. It is shown that null modifiers in Georgian have a form different from that of the modifiers in the constructions at issue, and the latter cannot have null heads. However, Baker's (1988) approach is not easily compatible with the derivational morphology of these examples. I propose an analysis in terms of Beard (1991), which addresses other bracketing paradoxes by permitting "the semantic features of an attribute [to] subjoin with one and only one semantic feature of its head" (1991: 208). In this way I suggest a unified analysis of the two construction types, drawing on a mechanism that must be included in the grammar for non-derived words as well.

1. Introduction

This paper addresses the issue of external modifiers through two otherwise unrelated constructions in Georgian. The constructions are illustrated below.

(1) sam tit-moč'r-il-i (k'aci)[1]
three.OBL finger-cut.off-PTCPL-NOM man.NOM
'(a man) with three fingers cut off'

(2) or-ze-met' marcvl-ian-i (sit'q'va)
two-on-more.OBL syllable-PROP-NOM word.NOM
'(a word) of more than two syllables'

The problem is that in each construction, a word in the oblique form appears on semantic grounds to modify part of a word following it. Thus, *sam* 'three' in (1)

* The research reported here was supported in part by the National Science Foundation under grant BCS-0091691. I am grateful to Marina K'enč'ošvili, Ramaz KurdaʒE, and Šukia ApridoniʒE for their intuitions and help with this paper. Earlier versions of the paper were presented at the University of Canterbury in 1999 and at Cornell University in 2000, and this version was presented at the 5th Mediterranean Morphology Meeting in Fréjus. I am grateful to all three audiences for helpful comments; I would especially like to thank Andrew Spencer and Albert Ortman.

[1] Abbreviations used in glossing include the following: DAT dative, ERG ergative, GEN genitive, INCHO inchoative, INDIC indicative, INST instrumental, MAS masdar, NOM nominative, OBL oblique, PL plural, PRIV privative, PROP proprietive, PRPTCPL present participle, PRX proximate, PTCPL past participle, PTV partitive, SG singular, SM series marker, SCM screeve marker, TRLV translative. Each element of a circumfix is glossed, and they are linked with subscripted numbers.

apparently modifies *tit-* 'finger', and *or-ze-met'* 'more than two' appears to modify *marcvl-* 'syllable'. There has been a debate in the literature about whether similar constructions involve stranded modifiers or null heads. I argue in this paper that things are just as they appear to be; these are stranded modifiers, not modifiers with null heads.

Baker (1988, especially 92-105, 1996) described stranding of a similar kind in noun incorporation and used this as an argument for syntactic movement and incorporation in the syntax. This approach violates the Lexical Integrity Principle, and analyses more in keeping with this Principle are suggested in Mithun (1984) and Rosen (1989). However, Baker's syntactic approach is not easily compatible with the derivational morphology illustrated in (2), and there are other problems with the approaches of Mithun and Rosen. In this paper I argue that, while words are formed in the morphology, under some circumstances parts of words are accessible to parts of the syntax, contrary to the Lexical Integrity Principle. Specifically, I argue that a demonstrative, adjective, or participle may modify the left-hand element of a compound or derived word in Georgian and that these constructions with "stranded modifiers" are actually simple bracketing paradoxes. In developing this I follow Beard (1991), which addresses other bracketing paradoxes by permitting "the semantic features of an attribute [to] subjoin with one and only one semantic feature of its head" (1991: 208). In this way I propose a unified analysis of both construction types.

I begin with an introduction to the issue of stranded modifiers vs. null heads in the linguistic literature and then go on to describe each of the constructions illustrated above. In the sections devoted to these individual constructions, (§§3-4), the first subsection provides a general introduction, the second presents evidence for wordhood as indicated by spacing in (1-2), and a later subsection includes specific arguments for an external modifier. §5 provides a discussion of Beard (1991) and a unified analysis based on this, and §6 offers a conclusion.

2. Stranded Modifiers and Null Heads in Linguistics

The issue of stranded modifiers has been addressed primarily in connection with noun incorporation. Mithun (1984) presents a large typological study of noun incorporation (NI) types, arguing that noun incorporation takes place in the morphology or lexicon, not in the syntax. Mithun argues that verbs with incorporated nouns (INs) are compounds, and particularly notes that in general left-hand constituents of such compounds do not refer, do not introduce discourse referents, and are not marked for definiteness or number (1984: 849). Thus, although we find the sentences in (3-4a), we do not find the corresponding (b) sentences.

(3) a. Bob went berry-picking.

b. *Bob went {the, ripe} berry-picking.

(4) a. I am baby-sitting.

b. *I am {those, some, three} baby-sitting.

(The examples in (3) and (4) are from Mithun 1984 or inspired by that paper.)

Her discussion of stranded modifiers of the incorporated noun is also of immediate interest here. She notes that (5) in Mohawk appears to involve a stranded modifier, but adds that "languages which exhibit structures like the above [(5)] also exhibit structures like [(6)]" (1984: 870).

(5) kanekwarúnyu wa'-k-akya'tawi'tsher-ú:ni.
it.dotted.DIST PAST-I-dress-make
'I dress-made a polka-dotted one.' ('I made a polka-dotted dress.')

(6) kanekwarúnyu wa'katkáhtho. (Mithun 1984: 870)
it.dotted.DIST PAST.I.see
'I saw a polka-dotted (one).'

She concludes that constructions such as (5) involve null heads.

Sadock (1980, 1986) suggests that INs in Greenlandic Eskimo can, contrary to Mithun's claim, be referential and may introduce discourse referents. In Greenlandic, these elements may also be quantified.

(7) ...Paliitsit 276-inik ammassattortoq (adapted from Sadock 1986: 28)
P. 276-INST.PL sardine.eat.NOM.PTV.3SG
'...that Paliitsit ate 276 sardines...'

Sadock argues that in Greenlandic quantifiers and other modifiers can be stranded, and he shows that Greenlandic lacks constructions with null heads, contrary to Mithun's claim quoted above. Greenlandic stranded modifiers also include both possessors (illustrated in (8)) and adjectives (Sadock 1986: 26-27).

(8) kunngip panippassuaqarpoq.
king-ERG daughter.many.have.INDIC.3SG
'There are many king's daughters (i.e. princesses).'

In (8), the ergative case expresses possession of the incorporated noun, 'daughter' (or of 'daughter.many').

Mithun (1986) argues that the constuctions at issue in Greenlandic are not noun incorporation, at least not in the usual sense of compounding. Rather, the construction that Sadock refers to as incorporation in Greenlandic is, according to Mithun, derivation. The bases for her analysis are (i) that the construction at issue is denominal verb derivation, and (ii) that independent verbs do not incorporate nouns (and the derivational affixes cannot exist independently as verbs). For Mithun, the fact that the construction at issue in Greenlandic does not meet the definitional criteria of noun incorporation means that it is irrelevant that Greenlandic does not conform to her generalizations. But the present paper focuses on external modifiers, not on noun incorporation; so the Greenlandic construction is entirely relevant here.

Like Mithun (1984, 1986), Baker (1988) considers incorporated noun constructions to be compounds (1988: 84). Unlike Mithun, he considers the external modifiers of Mohawk to be stranded, and he accounts for this in the syntax by a process

of movement with incorporation of the noun head, leaving behind (stranding) the modifiers.

Rosen (1989) suggests that two kinds of noun incorporation need to be distinguished, and that in a language in which noun incorporation is a kind of compounding, stranded modifiers will not occur. Rosen associates stranded modifiers with occurrence of a direct object outside the verb (called "doubling" or "possessor raising").

Baker et al. (2005) continue to look upon modifier stranding as a characteristic limited to NI, and they continue to argue for a syntactic account for NI, and thus for stranding.

Looking beyond NI, it has been observed that it is impossible to quantify the left-hand member of synthetic compounds in English: **the-man hater, *every-cat chaser, *some-dog lover* (Sproat and Ward 1987: 326). They suggest (FN 3) that this has something to do with the size, in X' terms, of the left-hand member: **The-Bronx hater, Bronx hater*.

Bresnan and Mchombo (1995), while acknowledging the considerable literature on syntactic phrases in derivation and compounding, state that "word-internal constituents generally differ from word-external phrases in disallowing the arbitrarily deep embedding of syntactic phrasal modifiers" (1995: 192). As examples, they cite (9).

(9) a. [$_{A}$ happy]-ness

b. *[$_{AP}$ quite happi]-ness

c. *[$_{AP}$ more happy [than sad]]-ness

For Bresnan and Mchombo, this is one of five tests for lexical integrity, and in this way the ban on external modifiers has become an integral part of the Lexical Integrity Principle.

Other violations of aspects of the Lexical Integrity Principle have been attributed to a variety of factors. Booij (1985) and Nespor (1985) suggest that conjunction reduction may apply to part of a word if it is a phonological word. Sproat and Ward (1987) suggest that outbound exceptions to anaphoric islandhood such as (10) are due to the "discourse salience" of part of a word.

(10) After painting the house I had *enough left over* for the dog-kennel.
(Simpson 1991: 56)

Simpson (1991: 61-62) agrees with the latter and adds that the acceptability of *pre- and post-World War II* is related to the relative transparency of the argument structure. She suggests that acceptability of gapping may also be related to stress. Finally, Spencer (1988, 1991: 414-417) has suggested that a prerequisite for the acceptability of *baroque flautist*, where *baroque* modifies a proper part of *flautist*, is the lexicalization of the phrase *baroque flute*.

I argue here that the Georgian construction in (1) is not noun incorporation, and the construction in (2) is not even similar to noun incorporation. Therefore the Greenlandic data are as relevant as the Mohawk data, whether one agrees with Mithun or with Sadock. In the sections below, I argue that in neither of the Georgian

constructions illustrated here does the modifier have a null head, nor does either involve word-internal modifiers. The analysis that the external modifiers are stranded by head movement encounters the problem that case markers and affixes that derive nouns and adjectives are not ordinarily viewed as projecting phrase structure. I argue (following Beard 1991) that instead we have a word-external modifier of only part of the word, and that this analysis is needed in any case for English expressions such as *old friend* that do not involve morphological complexity.

3. Participial Compounds

3.1. General Description

Modern Georgian no longer has productive noun incorporation with finite verb forms, as Old Georgian had.[2] However, compounding of a noun with a participial head is highly productive. These formations must be viewed as compounds of a noun and a participle for two reasons. First, these are adjectival, not part of a periphrastic verb form, such as English *is baby-sitting* or *has baby-sat*.[3] Second, since noun incorporation with finite verbs does not exist in Georgian, these cannot be the participles of such constructions.

The left-hand member of the compound is usually the theme of the verb expressed as a participle. Thus, in (11), *xel-* 'hand', is the theme of *gašlili* 'spread out'.

(11) xel-gašl-il-i
hand-spread.out-PTCPL-NOM
'(with) hand(s) spread out' i.e. 'generous'

The word in (11) also illustrates the fact that the meaning of the compound is not entirely compositional, and the compound may take on a particular meaning of its own.[4]

Additional examples with external modifiers are provided below.

(12) garšemo šav q'aitan-movleb-ul-i (Šaniʒe 1973:160-161)
outside black.OBL silk.cord-encircled-PTCPL-NOM
'edged on the outside with black silk cord'

(13) or tit-gašl-il-i (Šaniʒe 1973:160-161)
two.OBL finger-spread-PTCPL-NOM
'two fingers spread'

---

[2] A description of the syntax of productive noun incorporation in Old Georgian can be found in Harris (1985: 331-337).

[3] See Booij (1993: 39) for similar reasoning regarding Dutch.

[4] Similarly, the first element in the compound is not always the theme. Šaniʒe (1973: 160) makes the point that (i) refers to a person or thing carried away by water, not the person or thing that carried away water.

(i) c'q'al-c'aγeb-ul-i
water-carry.away-PTCPL-NOM
'[someone or something] carried away by water'

(14) [deda-s] cxare creml-mtovi-are-sa (Šanidze 1973:160-161)
mother-DAT hot tear-stream-PRPTCPL-DAT
'[the mother with] hot tears streaming'

(15) i šemomt'k'iceb-ul puʒe-eb-ši (A. Lomtaʒe 1987: 9, 6 up)
i retain-PTCPL stem-PL-in
'in *i*-retaining stems' 'in stems that retain *i*'

(16) p'ap'-is cxvr-is t'q'av-is kud-čamopxat'-ul-i
grandfather-GEN sheep-GEN skin-GEN hat-down.pull-PTCPL-NOM
'wearing grandfather's sheep skin hat pulled down low' (Šanidze 1973:160-161)

The last example, (16), shows that the external modifier may itself have a modifier and that the external modifier may be compound. Here, both 'grandfather's and 'sheep skin' modify *kud-* 'hat'. The modifier *t'q'avis* 'skin-GEN' is itself modified by *cxvris* 'sheep-GEN'.

A detail of Georgian grammar that is illustrated here and below is that when a quantifier is in construction with a noun, the noun does not bear an overt plural marker. Thus we find *or-i marcval-i* 'two syllables' [two-NOM syllable-NOM], not **or-i marvl-eb-i* [...syllable-PL-NOM]. For the same reason, the plural of *tit-* 'finger' in (13) would be ungrammatical, even if the words were independent: *or-i tit-(*eb)-i* 'two fingers'.

It is often noted that in noun incorporation the incorporated noun is non-referential, and more generally, left-hand members of compounds are non-referential (Mithun 1984, Spencer 1991: 312). Note in the examples above that the left-hand member may be non-referential (e.g. (11)); but when this element is modified, it appears to be referential.

3.2. Wordhood

In this section I argue that the participle and its theme are parts of a single word. For example, in (1), the participle *moč'rili* 'cut off' and its theme *tit-* 'finger' are parts of a single word. Arguments in this section do not prejudice the question addressed later, whether *sam* 'three' is also part of the same word, except that obviously the wordhood of the participle plus theme is a prerequisite to the more extensive wordhood.

In (11) above, *xel-gašl-il-i* '(with) hand(s) spread out', the participle, *gašl-il-i* 'spread out', can stand alone as a word. The left-hand element, *xel-* 'hand', on the other hand, cannot. In Georgian every independent noun must occur in a case form, and for consonant-final stems, all cases require a non-null case suffix. (For the declension of vowel-final stems see §3.4.) Thus, (17b), (18b), and (19b) are impossible.

(17) a. es marǰ vena xel-i aris
this.NOM right hand-NOM is
'This is the right hand.'

b. *es marǰ vena xel aris

(18) a. marǰ vena xel-is črdil-i
right hand-GEN shadow-NOM
'the right hand's shadow' 'shadow of the right hand'

b. *marǰ vena xel črdil-i

(19) a. xel-i da tav-i
hand-NOM and head-NOM

b. *xel da tav-i

The stem form and noun is used only in compounding and derivation:

(20) *Compounding*

xel-axla	hand-now	'again, renewed'
xel-burt-i	hand-ball-NOM	'handball'
xel-tepš-i	hand-plate-NOM	'wooden plate of medium size'
xel-saxoc-i	hand-towel-NOM	'handtowel'
xel-mk'lav-i	hand-arm-NOM	'hand and arm'

(21) *Derivation*

xel-eur-i	hand-PROP-NOM	'sheaf of ears (e.g. of corn)'
xel-v-a	hand-SM-MAS	'working, taking in hand, finding'
u-xel-o	$PRIV_1$-hand-$PRIV_2$	'[person, statue, ...] without hands'

Because *xel-* cannot be an independent word, *xel-gašl-il-i* '(with) hand(s) spread out' must be a single word. The same reasoning applies to the other examples quoted above, with the exception of (15), which is indeterminate in this respect and will not concern us further.

A second argument that the string in (11) is a word is that it has a single stress, *xel-gašl-íl-i* or *xel-gášl-il-i,* whereas the closest corresponding phrases, (22), have two.

(22) a. xél-i gašl-íl-i-a
hand-NOM spread-PTCPL-NOM-is
'The hand is open, outspread.'

b. gášl-il-i xél-i
spread-PTCPL-NOM hand-NOM
'open, outspread hand'

Third, the string in (11) cannot be split by another element.

(23) *xel-ve-gašl-il-i
hand-indeed-spread-PTCPL-NOM
'(with) that very hand outspread'

In this section I have given three arguments that the participle and its theme constitute a word, not a phrase, and I conclude that this is correct.

3.3. Two Analyses of (1)

In this subsection I lay out two possible analyses of (1) and of examples like it; in later subsections I argue that both are wrong. In §5 I propose a different analysis. For convenience, (1) is repeated here as (24).

(24) sam tit-moč'r-il-i (k'aci)
three.OBL finger-cut.off-PTCPL-NOM man.NOM
'(a man) with three fingers cut off'

According to the null head analysis, the structure of (1/24) is as in (25), where *sam* 'three' has a null head, indicated by N. Constituency is indicated with brackets, word boundaries with spaces.

(25) *Null Head Analysis*
[sam N] [tit-moč'r-il-i]
three N finger-cut.off-PTCPL- NOM

This is parallel to Mithun's analysis of NI in Mohawk, which the Georgian examples superficially resemble.

Section 3.2 above showed that the participle and its theme, here *moč'rili* 'cut off' and *tit'* 'finger' respectively, are a single compound word; it was silent on the inclusion of *sam* 'three' in that word. The second analysis, which I refer to as the internal modifier analysis, posits that *sam* 'three' is part of the word with the participle and its theme.

(26) *Internal Modifier Analysis*
sam-tit-moč'r-il-i
three-finger-cut.off-PTCPL- NOM

In the following subsections I argue against both analyses.

3.4. Arguments Against the Null Head Analysis

Georgian does have null heads; for example, the sentences in (27) are fully grammatical.

(27) a. sam-i vnaxe.
three-NOM I.see
'I saw three.'

b. sam-s vxedav.
three-DAT I.see
'I see three.'

c. sam-i makvs moč'rili.
three-NOM I.have.it cut.off
'I have three [e.g. fingers] cut off.' 'I have cut off three [e.g. fingers].'

Each sentence in (27) assumes a null head, whose referent has been established in discourse; (27a) might mean, for example, 'I saw three horses' or 'I saw three bridges', according to the referent in discourse. I argue below that (1) cannot be interpreted as an example of the null head construction illustrated in (27).

The first argument is based on the form of the modifier. The paradigm in (28a) shows the forms of an adjective, numeral, participle, or other modifier of a nominal, when the modifier precedes the head, in the ordinary order. Here *sam-* is 'three', *moxuc-* is 'old', and the head is *k'ac-* 'man'.[5]

(28)

	a. 'three old men'			b. 'three'
Nominative	sam-i	moxuc-i	k'ac-i	sam-i
Ergative	sam-ma	moxuc-ma	k'ac-ma	sam-ma
Dative	sam	moxuc	k'ac-s	sam-s
Genitive	sam-i	moxuc-i	k'ac-is	sam-is
Instrumental	sam-i	moxuc-i	k'ac-it	sam-it
Translative	sam	moxuc	k'ac-ad	sam-ad

The forms in (28b) show the forms for a numeral, adjective, or participle with a null head. While null heads are very common in Georgian, a consonant-final stem, such as *sam* 'three', cannot occur in this construction without an overt case suffix, as illustrated in (29).

(29)

a. sam-i / *sam vnaxe.
three-NOM three I.see
'I saw three.'

b. sam-s / *sam vxedav.
three-DAT three I.see
'I see three.'

c. sam-i / *sam makvs moč'rili.
three-NOM three I.have.it cut.off
'I have three [e.g. fingers] cut off.'

d. lamaz-eb-i / *lamaz-eb vnaxe.
pretty-PL-NOM pretty-PL I.see
'I saw the pretty ones.'

e. lamaz-eb-s / *lamaz-eb vxedav.
pretty-PL-DAT pretty-PL I.see
'I see the pretty ones.'

The grammatical versions of (29a-e) differ from the ungrammatical versions only in that the latter have no case suffix. ((29b,e) differ from (29a,d) in tense and require an object in a different case.) The examples in (29) show that in Georgian, a modifier with a null head cannot be in bare one of a few cases, e.g. dative (see (28)). and the modifiers in (1), (12), and (13), which are in stem form, cannot have null heads.

---

[5] The paradigm in (28a) represents the standard; an alternative pattern is as in (ii).

(ii)

Nominative	sam-i	moxuc-i	k'ac-i	'three old men'
Ergative	sam	moxuc	k'ac-ma	
Dative	sam	moxuc	k'ac-s	
Genitive	sam	moxuc	k'ac-is	
Instrumental	sam	moxuc	k'ac-it	
Translative	sam	moxuc	k'ac-ad	

That is, in all cases except nominative, pre-nominal modifiers may occur in stem form. This does not affect the form of modifiers with null heads, in (28b), which are those at issue here. However, the forms in (ii) are relevant to one of the arguments given in §4.4.

Vowel-final stems in Georgian are invariant when they precede a head, as illustrated by *cxare* 'hot' in (30a) ; when they have null heads, they are declined regularly, as in (30b). Because the stem form of a vowel-stem is indeterminate, examples with these are avoided here.

(30)	a.					b.	
		Nominative	cxare	creml-i	'hot tears'		cxare
		Ergative	cxare	creml-ma			cxare-m
		Dative	cxare	creml-s			cxare-s
		Genitive	cxare	creml-is			cxar-is
		Instrumental	cxare	creml-it			cxar-it
		Translative	cxare	creml-ad			cxare-d

For this reason, the examples in (14) and (15) are consistent with either analysis, and they are included here only to provide a variety of examples.

A second, related argument, is that, although null heads are common in Georgian, it is not possible to express a meaning parallel to that of Mohawk (5) using one. Recall that Mithun translates the NI example in (5) as "'I dress-made a polka-dotted one.' ('I made a polka-dotted dress.')" The semantic parallel for Georgian (1) would be 'with fingers-cut-off three of them', but 'three of them' cannot be expressed in a form parallel to Mohawk (5). Rather, if we take the nominative case as our example, the Georgian form most closely parallel to Mohawk (5) has the meaning in (31).

(31) sam-i tit-moč'r-il-i
three.NOM finger-cut.off-PTCPL-NOM
'three (men, people, statues, ...) with fingers cut off'

If *sam-* 'three' is in a case other than the nominative, it is difficult to assign any meaning to a null head here.

The formal and semantic differences between (1) and (31), show that (31) has a null head, while (1) has some other structure.

3.5. Arguments Against the Internal Modifier Analysis

In formal terms, the status of *sam* 'three' is indeterminate. That is, as shown in the preceding section, the stem form of a modifier is formally consistent either with its being part of a larger word, as in (20-21), or with its having a head that is itself in stem form.

In an example such as (16), while *kud-čamopxat'-ul-i* 'with hat pulled down low' has a single main stress (*kud-čamopxát'-ul-i* or *kud-čamopxat'-úl-i*), the other elements have their own main stresses. This indicates that these do not all form a single word.

As discussed above, Bresnan and Mchombo have argued that "word-internal constituents generally differ from word-external phrases in disallowing the arbitrarily deep embedding of syntactic phrasal modifiers" (1995: 192), and this statement would probably be accepted by most linguists. We find both a multiple modifier and a recursive modifier in example (16). Both *p'ap'is* 'grandfather's' and *cxvris t'q'avis*

'sheep skin' modify *kud-* hat, and *cxvr-is* 'sheep-GEN' modifies *t'q'avis* 'skin', just as the phrase modifies *kud-* 'hat'. Most linguists do not accept that words could have this complexity of internal structure of the types found in syntax, and such an analysis is undesirable.

3.6. Conclusion

We have seen that the stem form of a modifier is not consistent with its having a null head; null heads in Georgian cannot have bare stem form, which the modifiers in examples (1) and (12-13) have. The depth of embedding and complexity of examples such as (16) suggest that the internal modifier analysis is inappropriate. We must thus conclude that the participial construction described here has some other structure.

4. The Man in the Panther's Skin

4.1. General Description

The great Georgian epic poem that is translated into English as *The Man in the Panther's Skin* is in Georgian *vepxistq'aosani* or *vepxis tq'aosani*. Note in (32) that *vepxis* 'panther's' appears to be an external modifier of *t'q'a* 'skin, pelt'. In examples in this section, the root is in bold.

(32) vepx-is **tqa'**-osan-i
panther-GEN skin-PROP-NOM
'[the one] having a panther's skin'

The derivational suffix *-osan-* attaches to nouns and forms proprietive adjectives meaning 'having', 'characterized by', or, as here, 'wearing'. (Note that this phrase itself has a null head.) The suffix *-osan-* is not highly productive today, but a similar suffix with the same range of meanings, *-ian-*, **is** very productive, and it is illustrated in (2).

4.2. Wordhood

In this section I argue, regarding (2), repeated here for convenience as (33), that the root, *marcv(a)l-* 'syllable', and the derivational suffix, *-ian-* PROPRIETIVE, are parts of a single word.

(33) or-ze-met' **marcvl**-ian-i (sit'q'va)
two-on-more.OBL syllable-PROP-NOM word.NOM
'(a word) of more than two syllables'

Arguments in this section do not prejudice the question addressed later, whether *or-ze met'* 'more than two', is also part of the same word, except that the wordhood of these two parts is a prerequisite to the more extensive wordhood.

Evidence that *marcvl-ian-(i)* 'syllabic, having a syllable' in (2) and (24) is a word, not a phrase, is similar to that discussed in connection with the participial construction, in §3.2. The string *-ian(-i)* 'PROP' cannot stand as an independent word, nor can the noun stem *marcv(a)l-* 'syllable'.

(34) a. **marcvl**-is dasac'q'is-i
syllable-GEN beginning-NOM
'syllable onset'

b. ***marcv(a)l**

Independent nouns with consonant-final stems cannot stand in bare stem form. Since neither element can exist as an independent word, I conclude that ***marcvl**-ian(-i)* is a word, not a phrase.

4.3. Three Analyses of (2)

In this subsection I lay out three possible analyses of (2) and of examples like it; in later subsections I argue that none of these is correct. In §5 I propose a different analysis. For the sake of simplicity, I use the example in (35), instead of (2/33).

(35) or-**marcvl**-ian-i puʒe-eb-i (Jeiranišvili 1971: 23, 7 up, 4 up)
two-syllable-PROP-NOM stem-PL-NOM
'bisyllabic stems'

A null head analysis of (35) would be as in (36), where *or* 'two' has a null head, indicated by N.

(36) *Null Head Analysis*
[or N] [marcvl-ian-i [puʒe-eb-i]]
two.OBL N syllable-PROP-NOM stem-PL-NOM

While we have shown that *marcvliani* 'syllabic' is a word, not a phrase, we have not yet considered the possibility that *or* 'two' may be part of the same word. The internal modifier analysis in (37) posits exactly this.

(37) *Internal Modifier Analysis*
[or-marcvl-ian-i [puʒe-eb-i]]
two-syllable-PROP-NOM stem-PL-NOM

For (2) or (35) we need to consider an additional analysis; in (38) the head has nested modifiers.

(38) *Nested Modifier Analysis*
[or [marcvl-ian-i [puʒe-eb-i]]]
two syllable-PROP-NOM stem-PL-NOM

In the following subsections I argue against each of these analysis.

4.4. Evidence Against these Analyses

The arguments against the null head analysis in the participial construction also apply here. As shown in section 3.4, a modifier with a null head must bear a case marker (which can have a zero-form only with vowel-final stems). The modifiers *or-ze met'-* 'more than two' in (2), and *or-* 'two' in (35), do not bear cases; therefore they cannot have null heads, and we must rule out the analysis in (36).

The form of the modifier is likewise incompatible with the nested modifier analysis. As shown in (28a), a modifier in bare stem form is never grammatical with a head in the nominative case. In example (35), the only available heads are the nominals in nominative case and the stem form *marcv(a)l-* 'syllable'. The modifier in the stem form, *or* 'two' in (35), cannot modify *puʒeebi* 'stems' or the derived word as a whole, *marcvliani* 'syllabic', since each of these is in the nominative and is thus incompatible with the stem form of the modifier. There is a grammatical example similar to (35), but with nested modifiers; this requires the nominative form of the numeral, as in (39), where 'two' now modifies the phrase 'syllabic stems', as in (38).

(39) or-i **marcvl**-ian-i puʒe-eb-i
two-NOM **syllable**-PROP-NOM stem-PL-NOM
'two syllabic stems'

On the basis of the form of the numeral, as well as the meaning, we must rule out the nested modifier analysis of (35).

Regarding the internal modifier analysis, where modifiers of the root are part of the same word, orthography provides ambiguous evidence. Like many other examples, (40) shows the modifier of the root written together with the root and derivational suffix *-ian* as one word; in the original, it is written without hyphens.

(40) tanxmovan-**puʒ**-ian-eb-tan (A. Lomtaʒe 1987: 39, l. 6)
consonant-**stem**-PROP-PL-with
'with [ones] having consonant stems'

However, in other examples the modifier of the root is written as a separate word, as in (41) below.

(41) nac'armoeb-i (čveul. nates[a]obit brunva-ši dasm-ul-i) puʒ-isa da
derived-NOM ordinarily genitive.OBL case-in place-PTCPL-NOM stem-GEN and

"c'minda" **puʒ**-ian-i arsebiti saxel-isa-gan šedgen-il-i
pure **stem**-PROP-NOM substantive-NOM nominal-GEN-from compose-PTCPL-NOM
'composed of a derived stem (ordinarily placed in the genitive case) and a substantive nominal [i.e. noun] having a "pure" stem' (Jeiranišvili 1971: 30, 15)

In (41), the modifier of the root is *"c'minda"*. In yet other examples, part of the modifier is written with the derived word and part separately.

(42) ert-ze met'-**marcvl**-ian-i ʒir-eul-i morpem-eb-i
one-on more-**syllable**-PROP-NOM root-PROP-NOM morpheme-PL-NOM
'polysyllabic root morphemes' 'root morphemes having more than one syllable'
(Gamq'reliʒe 1983:16)

Here the modifier of *marcv(a)l-* 'syllable' is itself a complex phrase, *ert-ze met'-* 'more than one'.[6] We can see from these examples that the orthography is not consistent and thus provides no help in determining whether the internal modifier analysis, (37), is correct.

The form of the modifier likewise provides no information relevant to (37). On formal grounds, the number of words in the string *or-marcv-ian-i* 'bisyllabic' in (35) or even in *or-ze met' marcvl-ian-i* 'polysyllabic' in (2) is indeterminate. A quantifier, such as *or* 'two' or *met'* 'more', is in bare stem form preceding heads in certain cases (see the paradigm in (28a)). In addition, the stem form occurs in compounds and derivation, as illustrated in (43-44).

(43) *Compounds*

or-as-i	two-hundred-NOM	'two hundred'
or-zoma	two-size	'double-size'
or-sul-a	two-soul-PROP	'pregnant'
or-gul-a	two-heart-PROP	'liar, traitor'

(44) *Derivation*

or-d-eb-a	two-INCHO-SM-3SG	'it doubles, reduplicates'
or-eul-i	two-PROP-NOM	'double, shadow'

Thus, the form of the numeral in (35) cannot distinguish between the internal modifier analysis, (37), and the external modifier analysis proposed in §5.

We do find relevant evidence in (45), where the modifier is more complex than any we have seen.[7]

(45) ert-ze met' (γia an daxurul) **marcvl**-ian arsebit saxel-eb-ši
one-on more open or closed **syllable**-PROP substantive nominal-PL-in
'in substantive nominals [i.e. nouns] having more than one (open or closed) syllable' (Jeiranišvili 1971: 44, 23)

---

[6] This phrase illustrates one productive way of stating comparison in Georgian, where the standard of comparison is expressed as the object of the enclitic postposition *-ze* 'on', and where the comparitive adjective is expressed without special morphology, as in (iii).

(iii) šen-ze did-i
you-on big-NOM
'bigger than you'

(iv) or-ze met'-i minda
two-on more-NOM I.want.it
'I want more than two.'

[7] (45) is from the literature, but it is infrequent enough that I checked it with three consultants. Two found it fully acceptable, but the third wanted hyphens inserted in unspecified locations.

A string such as *ert-ze met' (ɣia an daxurul) marcvl-ian* 'having more than one (open or closed) syllable' in (45) is best viewed as a phrase, not a single word because, even if words may be based on phrases, it is generally assumed that a word may not include a syntactic construction as complex as a comparative phrase, such as *ert-ze met'* 'more than one', or a parenthetical remark, such as *(ɣia an daxurul)* '(open or closed)'. On this basis we must rule out also the internal modifier analysis of this construction.

4.4. Conclusion

In this section we have examined words derived with the proprietive suffix *-ian-*, which forms a word with a base and a case suffix. For examples in which the root of the derived lexeme is modified, we have considered a null head analysis, an internal modifier analysis, and a nested modifier analysis. The form of (2) and other related examples is not consistent with the first- and last-named analyses, but is indeterminate with respect to the internal modifier analysis. The fact that modifiers may include complex phrases, such as comparative constructions, and parenthetical comments suggests that the internal modifier analysis is also incorrect. I conclude that the modifiers here are external modifiers of the root.

5. Beard's Solution to Bracketing Paradoxes and Its Application to Georgian

Beard (1991) shows that a variety of bracketing paradoxes in English cannot be dealt with effectively by means of the various devices previously proposed for this, including bracket erasure. He argues that a construction such as *transformational grammarian* is actually semantically compositional and proposes a formal semantics which permits a modifier to take wide or narrow scope. For example, the modifier *transformational* can, in principle, take wide scope over *grammarian*, or narrow scope over *grammar* alone, the latter corresponding to the ordinary reading of this phrase. Beard shows that structural analyses are inherently incapable of accounting for the ambiguity of a phrase such as *criminal lawyer*, because there are only two potential structures, yet established tests show four potential meanings, (46).

(46) a. [criminal lawyer] 'the lawyer who is criminal as a person'

b. [criminal law] yer 'a person who practices law criminally'
(i.e. 'who is criminal as a lawyer')

(c) [criminal law] yer 'a person who practices criminal law', where
(i) ?the law is criminal (QAdj reading) or
(ii) the law merely pertains to crime
(RAdj reading) (Beard 1991: 201)

Beard's analysis is further buttressed by the observation that the same narrow vs. wide scope ambiguity exists in phrases such as *old friend*, which are not morphologically derived. That is, the machinery for resolution of this ambiguity must exist in the grammar independently of the requirements of bracketing paradoxes.

In his paper, Beard suggests that "in languages other than English, [such problems] seem to reduce to phonological issues unrelated to morphology" (1991: 197). In this section I argue that, on the contrary, the apparently external modifiers in Georgian described above present entirely morphological problems, that these problems reduce to bracketing paradoxes, and that these paradoxes, like the English ones discussed by Beard, are effectively analyzed by means of his proposed semantics.

Beard makes use of a simple semantics, based on Jackendoff (1983, 1987), requiring no special mechanisms, and adding only a more elaborated notion of attribute composition (1991: 205). Definitions of potential heads indicate the features "category", "function", and "properties", which are referred to by the Principle of Decompositional Composition.

(47) *The Principle of Decompositional Composition*
The semantic features of an attribute subjoin with one and only one semantic feature of its head. (Beard 1991: 208)

According to (47), the modifier ("attribute") does not combine semantically with all the features of the head on either the broad or the narrow reading; rather it combines with a single feature.

When we consider the Georgian constructions described above, we find that for each, one of the readings that is possible in English is ruled out in Georgian on grammatical (morphological) grounds. In the participial construction, we see that the structure parallel to (46a), namely (48a) below, is ruled out for (1) in Georgian by the fact that the modifier *sam* 'three' is not in the form to agree in case with the whole word *tit-močrili* '(with) finger(s) cut off'.

(48) a. *[sam tit-moč'r-il-i N-i]
[three-OBL finger-cut.off-PTCPL-NOM]
'three (men, people, statues, ...) with fingers cut off'

b. [sam tit]-moč'r-il-i N-i
[three-OBL finger]-cut.off-PTCPL- NOM
'(men, people, statues, ...) with three fingers cut off'

In Georgian, the reading in (48a) is possible only if *sam* 'three' were in the form with the suffix *-i*, agreeing with *tit-moč'rili* 'with fingers cut off' and with its head, as in (46), repeated here as (49).

(49) [sam-i tit-moč'ril-i]
three-OBL finger-cut.off-PTCPL- NOM
'three (men, people, statues, ...) with fingers cut off'

The structure in (48b), parallel to English (46b,c), with narrow semantic scope, is possible; remaining ambiguities are beyond the scope of this paper.[8] The bracketing paradox here is that semantically and in terms of case agreement *sam* 'three' modifies

[8] As far as I am aware, none of the modifiers in the participial examples has a QAdj reading (see Beard 1991: 199ff and sources cited there), but I have not done fieldwork on this specific issue. If correct, this reduces the number of readings available in Georgian.

and thus forms a constituent with *tit-* 'finger', while the two do not form a word or a syntactic constituent; rather the latter, but not the former, is part of a larger word. Applying Beard's semantics here to (1), we see that on the ordinary reading *sam* 'three' combines only with the category feature of *tit-* 'finger'.

Similarly, in the derivation of (2), the (a) reading is impossible because of lack of case agreement, as shown in §4.3.

(50) a. *[or-ze-met' marcvl-ian-i N-i]
two-on-more.OBL syllable-PROP-NOM
'more than two syllabic (words, morphemes, affixes,...)'
b. [or-ze-met' marcvl]-ian-i N-i
two-on-more.OBL syllable-PROP-NOM
'(words, morphemes, affixes,...) having more than two syllables'

The reading in (50a) is possible only if the form of the case were changed, as in (51).

(51) [or-ze met'-i marcvl-ian-i N-i]
two-on more-NOM syllable-PROP-NOM
'more than two syllabic (words, morphemes, affixes,...)'

Given the form of (2), shown also in (50), the reading with narrow scope is the only possible one in Georgian. In this instance the bracketing paradox is that *or-ze met'* 'more than two' forms a constituent with *marcv(a)l-* 'syllable' semantically and in terms of case agreement, but not in terms of word boundaries. The Principle of Decompositional Composition permits the constituent *or-ze met'* 'more than two' to combine with the category feature of *marcv(a)l-* 'syllable' alone, accounting for the narrow scope of the interpretation.

Thus, the two problems introduced at the beginning of this paper are, in fact, bracketing paradoxes. Through morphological case agreement, they avoid some of the ambiguity that is found in some bracketing paradoxes in English. The Principle of Decompositional Composition provides an effective means of accounting for such constructions.[9]

Beard (1991) has shown that even for simplex words, (external) modifiers must have access to parts, features, of a word. In this way we can explain the meanings of *old friend*. Accounting for external modifiers of complex words such as those in (1-2) does not require any additional complication of the grammar. Because external modifiers must be available for underived words, stranding does not constitute an argument for a syntactic account of noun incorporation.

[9] There remain other bracketing paradoxes in Georgian that are not effectively dealt with in this way.

6. Conclusion

It has been suggested in the linguistic literature (see §2) that part of a word cannot be modified by an element external to the word. An exception on some accounts is a modifier stranded by NI. I have shown in sections 3 and 4 that in two constructions in Georgian, the left-hand constituent of a compound or derived word can indeed be modified by a word, or an even longer string, outside.

However, Beard (1991) shows that this does not depend on morphological structure: an external element can and does ordinarily modify part of a word. In English phrases such as *old friend, prolific novelist*, the first word combines not with a particular morpheme, but with a particular element of the semantic features of the head. In these two structures in Georgian, this common semantic structure is reflected in the grammatical structure.

The putative inability of a part of a word to be modified from the outside has been interpreted by some as a basic notion of wordhood or of the Lexical Integrity Principle (see especially Bresnan and Mchombo 1995). Many kinds of exceptions have previously been noted in English, and (52) and (53) illustrate two more.

(52) small-childhood (Wharton 1998: 197)

(53) I [have] been mothered, fathered, aunt and uncled... (Simon 1965)

Wharton's (52) illustrates an external modifier (written by the author with a hyphen), while (53) could be interpreted as gapping of the participial marker *-ed* or as conjoining inside a word, with a structure like *[aunt and uncle]-d*, either of which is unexpected. Of course, in English these are unusual and not fully acceptable, while the Georgian examples discussed here are common and fully acceptable to all consulted. In other work I have challenged the accepted status of other criteria for identifying the word, including the putatively absolute inability of a word to be interrupted by a clitic or other word (Harris 2000, 2002) and the putative status of a word as an in-bound anaphoric island (Harris 2006), in languages where the phenomena at issue are neither unusual nor incompletely acceptable. I suggest that these three characteristics – inability to be interrupted, in-bound anaphoric islandhood, and inability to have modifiers of a proper subpart – while not irrelevant to wordhood, are not absolute linguistic principles, only tendencies. One possibility is that each of these is best expressed through Optimality Theory as a highly valued constraint in most languages.[10] However, before we jump to that conclusion, we should look more closely at these phenomena in other unfamiliar languages, as well as at other restrictions on the constructions studied here. For example, although both constructions included in this paper permit external modifiers of parts of words, not all modifiers are permitted.

(54) a.

es	lob-ian-i	xač'ap'ur-i
this.NOM	bean-PROP-NOM	cheese.bread-NOM

'this cheese-bread [made] with beans'

[10] Similar suggestions are made in Anderson (2005), Harris (2002: 165-166), and Rice (2004: 295).

b. *am lob-ian-i xač'ap'ur-i
this.OBL bean-PROP-NOM cheese.bread-NOM
'cheese-bread [made] with these beans'

In (54a), *es* 'this' modifies the *head,* xačap=uri, whereas in the ungrammatical (54b), it modifies the constituent *lob[io]* 'bean'. Given the facts of §4, one might expect (54b) to be grammatical, and its ungrammaticality requires further study.[11] Another example is that given the existence of in-bound anaphora in English words such as *therefore*, *therein*, *whereby,* and perhaps even *himself* and the fact that in-bound anaphora are extremely common in Georgian (Harris, 2006), we need to understand better what language-particular feature of English really blocks words such as **him-ite* and **you-less*. I suggest that much study is still needed to understand the characteristics of words cross-linguistically.

References

Anderson, Stephen R. (2005) *Aspects of the Theory of Clitics*. Oxford: OUP.

Baker, Mark C. (1988) *Incorporation: A Theory of Grammatical Function Changing*. Chicago: University of Chicago.

Baker, Mark C. (1996) *The Polysynthesis Parameter*. Oxford: Oxford University.

Baker, Mark C., Roberto Aranovich, and Lucia A. Golluscio (2005) "Two Types of Syntactic Noun Incorporation: Noun Incorporation in Mapudungun and Its Typological Implications." *Language* 81,138-176.

Beard, Robert (1991) "Decompositional Composition: The Semantics of Scope Ambiguities and 'Bracketing Paradoxes'." *Natural Language and Linguistic Theory* 9, 195-229.

Booij, Geert E. (1985) "Coordination Reduction in Complex Words: A Case for Prosodic Phonology." In H. van der Hulst and N. Smith (eds.), *Advances in Nonlinear Phonology*. Dordrecht: Foris, 143-160.

Booij, Geert (1993) "Against Split Morphology." In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology, 1993*. Dordrecht: Kluwer, 27-49.

Bresnan, Joan, and Sam A. Mchombo (1995) "The Lexical Integrity Principle: Evidence from Bantu." *Natural Language and Linguistic Theory* 13, 181-254.

Harris, Alice C. (1985) *Diachronic Syntax: The Kartvelian Case.* (Syntax and Semantics, 18.) New York: Academic Press.

Harris, Alice C. (2000) "Where in the Word is the Udi Clitic?" *Language* 76, 593-616.

Harris, Alice C. (2002) *Endoclitics and the Origins of Udi Morphosyntax*. Oxford: Oxford University Press.

Harris, Alice C. (2006) "Revisiting Anaphoric Islands." *Language*.

Jackendoff, Ray (1983) *Semantics and Cognition*. Cambridge (Mass.): MIT Press.

Jackendoff, Ray (1987) "The Status of Thematic Relations in Linguistic theory." *Linguistic Inquiry* 18, 119-126.

Mithun, Marianne (1984) "The Evolution of Noun Incorporation." *Language* 60, 847-94.

Mithun, Marianne (1986) "On the Nature of Noun Incorporation." *Language* 62, 32-37.

Nespor, Marina (1985) "The Phonological Word in Italian." In H. van der Hulst and N. Smith (eds.) *Advances in Nonlinear Phonology*. Dordrecht: Foris, 193-204.

Rice, Keren (2004) "Review of *Endoclitics and the Origins of Udi Morphosyntax*, by Alice C. Harris." *Linguistic Typology* 8, 287-298.

[11] It appears that the generalization is that words derived with *-ian-*, unlike the participial construction or the partitive, require that the left-hand element be indefinite.

Rosen, Sara Thomas (1989) "Two Types of Noun Incorporation: A Lexical analysis." *Language* 65, 294-317.
Sadock, Jerrold M. (1980) "Noun Incorporation in Greenlandic." *Language* 56, 300-319.
Sadock, Jerrold M. (1986) "Some Notes on Noun Incorporation." *Language* 62,19-31.
Šaniʒe, Ak'ak'i (1973/1953) *Kartuli enis gramat'ik'is sapuʒvlebi*. [Fundamentals of the grammar of the Georgian language.] Tbilisi: Universit'et'i.
Simpson, Jane (1991) *Warlpiri Morpho-syntax: A Lexicalist Approach*. Dordrecht: Kluwer.
Spencer, Andrew (1988) "Bracketing Paradoxes and the English Lexicon." *Language* 64, 663-682.
Spencer, Andrew (1991) *Morphological Theory*. Oxford: Blackwell.
Sproat, Richard, and Gregory Ward (1987) "Pragmatic Considerations in Anaphoric Island Phenomena." *Chicago Linguistic Society* 23, 321-335.

Sources of Examples

Gamq'reliʒe, Nana (1983) "Udiuri enis xmovanta ganac'ilebisatvis morpemis sazγvrebši." [On the distribution of Udi vowels at morpheme boundaries.] *Masalebi axalgazrda mecnier-pilologta resp'ublik'uri k'onperenciisa*. Tbilisi: University, 13-17.
Jeiranišvili, Evgeni (1971) *Udiuri ena*. [The Udi language.] Tbilisi: University.
Lomtaʒe, Amiran (1987) *Brunebis taviseburebata ist'oriisatvis megrulši*. [On the history of peculiarities of declension in Mingrelian.] Tbilisi: Universit'et'i.
Simon, Paul (1965) "A Simple Desultory Philippic (or How I was Robert McNamara'd into Submission)."
Wharton, Edith(1933/1998) *A Backward Glance: An Autobiography*. NY: Simon and Schuster.

On a Subclass of Non-Affixed Deverbal Nouns in French

Françoise Kerleroux
U. Paris 10 & CNRS 71 14
kerlerou@u-paris10.fr

Abstract

In what follows I have two limited aims: (a) draw an empirical consequence from precedent analyses on the stem allomorphy in the French verb, as regards abstract deverbal non-suffixed Nouns such as *resultat* or *corrélat* as well as *concept* or *tract,* (b) search for theoretical implications of this re-organization of the data. My communication is organised as follows: after summing up the main results of two precedent analyses on the stem allomorphy in verbal inflection and in deverbal derivation (§1), I'll show that the Extra Stem, specialised for derivation, hidden to Inflection, appears in converted deverbal lexemes as well as in suffixed ones (§ 2). After drawing some consequences (§ 3), I'll examine the status of these newly recuperated, morphologically converted deverbal N: they are formally distinct from converted deverbal N previously recensed, but are they semantically distinct (§4) ?

> "By contrast, the notion of lexeme […] is entirely independent of morphological invariance: some (probably most) lexemes have a unique form associated with them ; others do not" (Aronoff, 1994, 9)

1. Stem allomorphy in the French Verb

We speak of stem allomorphy when a single lexeme uses more than one stem for its inflexion, or when complex lexemes which are constructed on the same lexeme base do not use the same stem.

1.1.

To account for the stem allomorphy in inflection, I refer to Bonami & Boyé 2002, whose main results are:

- **Multiple stems**: lexemes may come equipped with a collection of phonologically distinct stems (cf. Lieber, 1981, Aronoff, 1994, Pirrelli & Battista 2000, Stump 2001)

- **Stem space:** lexemes come equipped with a structure for storing stems, in as much as these stems are indexed with certain grouped parts of the paradigm.

G. Booij, *et al.* (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5*)
Fréjus 15-18 September 2005, University of Bologna, 2007.
URL http://mmm.lingue.unibo.it/

Stem 1	PRST Sg
Stem 2	PRST 3 Pl
Stem 3	IMPFT & PRST1/2 PL

So there are verbs which use up to three different stems in the present (cf. *vouloir: je veux, nous voulons, ils veulent*), but all verbs use the same stem for the three singular forms, and all verbs use the same stem for the first and second plural forms.

- **Stem indexing:** the verbal stem space has a number of 12 different slots used in inflection: by default, most slots contain identical stems, so that what we call regular verbs use a single stem in the whole paradigm.

- **Morphomes**: this type of groupings across the paradigm shows that the stems involved in inflectional morphology are pure morphological objects, that is morphomes in the sense of Aronoff 1994. This stem allomorphy is not motivated, either from a phonological nor from a morphological point of view.

1.2. Distribution of derivational suffixes such as -ion*,* -eur/-rice *or* -if/-ive*.*

To account for the distribution of derivational suffixes such as *-ion*, *-eur/-rice* or *-if/-ive*, we postulated (Bonami, Boyé, Kerleroux, 2004) that French verbs possess a 13th slot, filled with an extra stem, or ES, which never surfaces in inflected verbal forms, but is visible in derived lexemes such as DÉRIVAT-ION, or SUPPLÉT-IF, or FORMAT-EUR. The default form of this ES is formed by appending *–at-* at the end of the Stem 3. This analyses relies on the classification of deverbal Nouns in *–ion* , which can be classified according to the way their form relates to that of the S3. There are 6 cases:

		Surface Relation between S3 and the nominal lexeme	examples
Class 1	Open	S3 + asjô	dérivation
Class 2	Open	S3 + kasjô	nidification
Class 3	74	S3 + jô	dispersion
Class 4	45	S3 + isjô	composition
Class 5	28	S3 + sjô	diminution
Class 6	305	X[1] + jô	abstraction

Table 1. Classes of the French deverbal N in -ion

In Table 1, one can see that the default realization of the extra stem is obtained by appending *-at* at the end of the Stem 3, as in the class 1. The class 2 collects the deverbal Nouns suffixed in *-ion* obtained from Verbs which are themselves suffixed in *-ifi*(er). As this verb class in *-ifier* is productive, the class of derived Nouns in *-ification* is altogether open and irregular. The other classes are altogether irregular and closed.

[1] In class 6, we call X a verbal stem distinct from S3, which resembles the past participle. (*abstrait/ abstract*)

2. Empirical objective

We argue here only that the extra stem ES shows up

- not only in suffixed lexemes, data which constitute the empirical basis of the hypothesis,

- but also in converted ones.

V with default ES	Suffixed Lexeme	Converted Lexeme
agglomérer	*agglomérat-ion*	*agglomérat*
attenter	*attentat-oire*	*attentat*
corréler	*corrélat-ion, corrélat-if*	*corrélat*

Table 2. Verbs with default ES

...agrégat, alternat, assassinat, attentat, condensat, corrélat, crachat, distillat, éjaculat, exsudat, filtrat, format, habitat, isolat, pissat, plagiat, postulat, prédicat, réduplicat, résultat, troncat[2]*...*

The identification of this extra stem, hidden to inflexion, is the result of a displacement of allomorphy: instead of describing phenomena of allomorphic suffixes, one describes allomorphic stems in the verbal lexeme. A considerable advantage of this analysis is that we do not need to predict parallel allomorphies for the other suffixes *-eur, -if/ive, -oire, -ure* (cf. *modificatoire, compositeur, reducteur, descriptif*, *courbature*.)

As the hidden stem has not always the default form (cf Table 1), we have to analyse other nominal lexemes as converted ones to , constructed on the Extra Stems which are irregular ones,as in Table 3.

V with irregular ES	Suffixed Lex	converted Lex
Insérer	*insert-ion*	*insert*
Substituer	*substitut-ion, substitut-if*	*substitut*
Concevoir	*concept-ion, concept-eur*	*concept*
Ouïr	*audit-ion*	*audit*
Requérir	*réquisit-ion*	*réquisit*
Tirer	*tract-ion, tract-eur*	*tract*

Table 3. Verbs with irregular ES

....institut, attribut, affect, percept, abstract, tact,

[2] We recensed 67 deverbal converted Nouns on a list of 430 Nouns ending in –at, which was provided by our colleague N.Hathout (ERSS & Toulouse) using his informatic tool called Webaffix (cf..Hathout & Tanguy, 2001)

Verbes	N affixed in -ion	converted N on ES	deverbal N / A on S3
plagier	Ø	plagi-at	plagiaire
assassiner	Ø	assassin-at	assassin
cracher	Ø	crach-at	cracheur
pisser	Ø	piss-at	pisseur, pisse

Table 4. Verbs whose ES appears exclusively in converted N.

the V *plagier, assassiner, cracher, pisser* do not construct suffixed lexemes on their ES. Their converted N are the exclusive instance of their ES, as they construct other deverbal nouns on S3 as in *plagiaire, cracheur, pisseur* or *assassin (*plagiataire, *crachateur, *pissateur, *assassinateur*).

With this hypothesis on the stem allomorphy of the V and the default form of the ES, we expect to observe new formations using it; and we did find new attestations, such as *narrat*, which the French writer Antoine Volodine gave as subtitle to his recent novel *Les anges mineurs* (Le Seuil, 2001)

"Ecoutons Volodine expliquer lui-même ce qu'est un narrat" (Internet)

We found also *defecat, replicat, comparat,* on the Net.

3. Consequences

3.1. One single nominal derivational suffix –at left

One consequence of this analysis is to eliminate one of the two processes of nominal suffixation in *–at* usually attested in morphological inventories in French: once the form X*at* has been re-identified as one in the collection of the stems of the verb, French morphology is left with only one suffixation process, showing up in:

(4) *épiscopat, cardinalat, mécénat, baronnat, kamikazat*[3]

i.e. lexemes referring to the social position hold by the referent of the lexeme simplex or to the function[4] of this role.

[3] *"Y a-t-il crise de la vocation dans le kamikazat ?"* D. Durand, *Le Canard Enchaîné*, 10/8/05.

[4] And, through metonymy, to the period or to the place where the function is exerted (Lecomte, 1997)

(i) *pendant mon professorat, pendant mon noviciat*

(ii*) dans tout le comtat, dans l'émirat, aller au commissariat*

But when *commissaire* refers to the person responsible for an exposition of fine arts, the name of the function is available:

(iii*) mon commissariat a duré trois mois.(= 'ma charge de commissaire')*

Lexeme simplex 'social role'	Suffixed lexeme in –at 'dignité de'- or function
mécène	*mécénat*
cardinal	*cardinalat*
baron	*baronnat*
vizir	*vizirat*
émir	*émirat*
professeur	*professorat*

Table 5. Suffixed denominal N in -at

3.2. Contradiction with analyses which

- either identify one single suffixe -*at*, derivational suffix supposed to apply both on verbal basis (*resultat, agrégat*) or on nominal basis (*maréchalat, artisanat*), as it is done in TLF, in Dubois, 1962.

- or which distinguish two derivational suffix, correlated with the categorial difference of verbal basis and nominal basis.(Lecomte, 1997)

- It comes also in contradiction with Huot's (2001) analysis who identify any *–at* as an aspectual suffix, not a derivational one, with value of "accompli", taken to be the continued realization of the i-e "élargissement" of the i-e root (Benveniste, 1935, chap. 9).

- But it is concordant with Corbin's (1987) analyses, who saw the segment *–at-* in *resultat* as a "segment parasite", which meant that she clearly denied the derivational status to it.

3.3.

The recognition of this particular instance of stem allomorphy implies not to confuse between extra stems in *Xat-*, and verbal stems who end in -*at* and are common to inflection and derivation, as with

(8) *combattre / combat* *appâter / appât*
ébattre / ébat *rabattre / rabat*
débattre / débat *constater / constat*
contracter / contrat (afr. *contract)*
*acheter / achat (*afr. *achapter,* XII° *achater))*

3.4.

As a consequence of this re-organisation of the data, we obtain, as morphologically derived, lexemes which were left before in isolation (from the morphological point of view), and supposed to give access only to etymological and/or to socio-historical explanations such as:

(9*)* *concept, percept*, *institut, attribut*, are adscribed to the latine supines in Petit Robert dictionary
tract (in DHLF)*, abstract*, *audit (*Gd Robert*)* are identified as anglicisms
tact, requisit, as latinisms, latin borrowings.

4. Status of these recuperated morphologically derived data

Do the deverbal nouns converted on ES form a subclass among all the converted Nouns previously recensed and analysed (Lieber, 1981, Kerleroux, 1996, Meyer- Lübke, 1974) ?

4.1.

Formally we have distinguished 4 subclasses, or twice two:

Stems common to Inflexion & Derivation		Extra Stem	
Converted N on **S1** Masc A	Converted N on **S3**: a)Masc & b)Fem B	Converted on **ES** (+ default) Masc C	Converted on **ES** (- default) Masc D
soutien *maintien* *ébat, débat,* *rabat,* *combat*	*a) change, legs, rabais, rebut,* *progrès, trot, port, accord,* *encart, don,* *abandon, pardon, regret* *b) attaque, écoute,* *annonce, commande, baisse,* *donne, relâche, nage*	*agrégat* *corrélat* *filtrat* *résultat* *troncat*	*insert* *tact* *tract* *abstract* *concept, percept* *attribut, institut* *substitut*

Table 6: Formal subclasses of converted N

N.B1: The masculine converted N on Stem 3 in the B column is no more a productive type. The feminine form is the unique productive pattern in the B case, since XVII° s.

N.B2: The masculine converted N (in every class) present the effects of some phonological regularities: the last Consonant usually falls (as in *rabais* vs *baisse*, or *maintien, combat*,); the last vocal may be nasalised (cf *don, abandon* VS *donne).* But one remarks that the converted N of class D, i.e. derived from the Extra Stem of the Verb, are not always following these phonological regularities: *attribut (attribuer*) is

similar to *rebut (rebuter), insert (insérer)* is similar to *encart (encarter)* as regards the non realization of the last consonant, which contrasts with *concept*, (VS *respect) or audit.*

4.2. Semantically: are there distinct subclasses ?

Actually we assume, using associated syntactic tests, and following Grimshaw (1990) the dual interpretation of deverbal abstract N in general in French.

We ask the question: are the deverbal N converted from the ES (Class C & D) belonging to the same class, that is, virtually semantically ambiguous, as the members of the classes A & B, which are described as apt to denote resultative objects [+R], products of the effectuation of the process, on one part, and events [+Ev] (complex or simple) on the other.

4.2.1.

The interpretation [+Ev] of the deverbal abstract N as N denoting a complex event, (that is, as argument-taking Noun), is correlated with the following syntactic constructions:

- exclusive definite determinant,
- presence of argumental complement,
- exclusive singular,
- aspectual adjectives as *constant*, *fréquent, éphémère, systématique*...

4.2.2.

The interpretation [+ R] of deverbal abstract N as denoting a resulting object, effect of the process, is correlated with the following syntactic constructions:

- all types of determinants
- no argumental complement
- variation in number
- no aspectual modifiers

These syntactic criteria permit us to see that the converted deverbal Nouns, pertaining to the class A & B, in Table 6, i.e. deverbal N converted from the verbal Stems S3 or S1, common to Inflexion and to Derivation, possess both the [+Ev] and the [+R] interpretations, as it is manifested respectively in examples (10) & (11), and as the suffixed deverbal N do too:

(10) *Le maintien des frontières de la République est un devoir sacré*
Le port constant du chapeau lui avait occasionné une légère tonsure
Le transport trop fréquent de ces dessins a causé de graves dommages
Le récit de ses malheurs lui avait pris deux jours
La baisse du prix du pétrole a eu lieu au plus mauvais moment
L'attaque de la diligence a eu lieu dans la forêt
L'annonce de sa mort a eu lieu hier à midi

(11) *Ces jeunes femmes ont des maintiens trop rigides*
Les ports majestueux des chênes et des hêtres étaient admirablement rendus par le peintre
Les transports en commun sont en grève
Les récits de Paul ravissent la famille
Les baisses subites des bourses de Londres et de Singapour sont inquiétantes
Aucune de ces attaques ne le fera reculer
Ces annonces de fermeture d'usines sont décourageantes.

4.3.

In contrast, the resultative interpretation appears to be the unique possible one for the converted N of the C & D classes (collected in Table 6) , as the contexts in (12) allow to see:

(12*)* a. *procéder à un audit, présenter des réquisits, un combinat, procéder à des inserts, élaborer des concepts, changer les attributs, faire un plagiat, obtenir un distillat, tirer des tracts, …*

b. ** L'audit systématique des nouvelles recrues est prévu* (VS *L'audition*)
** L'attribut d'une prime de fin d'année a été voté* (VS *L'attribution*)
** L'agglomérat des deux substances s'est produit en une heure* (VS *L'agglomération*)

4.4.

In particular, when the lexicon provides the pair of a N deverbal N in *–ion* and a converted N, both on the special stem ES, they present the repartition event/ resultative object, which seem to constitute a mini-scheme:

(12) *troncation/ troncat* *éjaculation/ éjaculat*
distillation/ distillat *filtration/ filtrat*
insertion /insert *attribution /attribut*

4.5.

An indirect proof of the main disponibility of the resultative interpretation 'product of the process' lies in the existence of lexemes which are morphologically non conform, since the verbal basis is lacking, but which are semantically conform;

(13) *alcoolat*: "médicament obtenu par distillation de l'alcool sur des substances aromatiques"
oléolat: "huile essentielle"
hydrolat: "eau chargée, par distillation, de principes végétaux volatils"
cf. also *lysat, cédrat, orangeat,* etc.

4.6.

An other proof of this proeminence of the resultative interpretation lies in the fact that denominal verbs suffixed in *–ifier*[5], which form abstract N in – *ification*, (= class 2 in Table1.), do not furnish any converted N in *Xificat*. And we observe that this is not strange at all since the resulting object of the process of *nidifier*, *nidification* is *nid,* that is, the simplex nominal basis itself, which blocks any **nidificat*.

(14)

Verb	Action	Noun
nidifier	*nidification:*	' un nid'
codifier	*codification:*	' un code'
pacifier	*pacification:*	' la/une paix'
désertifier	*désertification:*	' le désert'
planifier	*planification:*	' un plan'

4.7.

Some converted Nouns, built on Extra Stems however also exhibit the properties of argument-taking Nouns (exclusive definite determinant, presence of argumental complement, exclusive singular, aspectual adjectives as *constant*):

(15) *L'assassinat de Maskhadov par le FSB a eu lieu le 8 mars*
L'attentat contre la mairie s'est produit en plein jour
Le plagiat des auteurs du XVIII a eu lieu sur une grande échelle
Je peux en témoigner: le crachat a eu lieu sous mes yeux

We finally can conclude that N which are converted on the ES constitute a semantic subclass among converted deverbal N, with a resultative interpretation, denoting the resulting object of the process, while converted deverbal N on S3 or S1, are ambiguous, apt to be interpreted as N denoting objects or events. Actually, the data of (15) are the

[5] The only 2 forms in *–ificat* are *certificat* and *pontificat. certificat* is regularly derived from *certifier*,itself borrowed from the desadjectival latin verb *certificare* 'to attest'. And *pontificat* is the derived N denoting the function of the *pontife* (' the pope') and belongs to the serie of the Table 5, as an ironical gift of etymology.

same as those collected in the Table 4, that is, they are converted deverbal N which do not form a pair with a suffixed N in *-ion.*

5. Conclusion

Assuming that lexemes come equipped with an array of stems (Aronoff, 1994, chap. 2) allows us to include into morphology a certain amount of data that are usually left in isolation, except from an etymological viewpoint. It was not our aim, but it happens. Has this sort of overlapping (of Morphology upon Etymology) theoretical consequences?

Once we abandon the reductionist doctrine "that the essence of language lies in the psychological invariance of the morphemes – as in Jakobson's structuralism and early generative phonology" (Aronoff, 1994., 9), and define the unit of morphology, the lexeme, as a set of stems associated to a set of related meanings, we don't need the distinction between bound vs free forms:

> Le lexème étant par définition une entité abstraite, c'est-à-dire considérée hors emploi, il n'y a pas de sens à dire qu'un lexème est un élément libre (alors que par exemple les affixes seraient des éléments liés). La question de savoir si une unité linguistique est libre vs liée ne se pose que pour les expressions figurant dans les phrases. Tout ce qu'on peut dire dans l'optique développée ici, c'est que les lexèmes ont pour corrélats des expressions libres (i-e des mots-formes). (Fradin, 2003, 93)

And we remain wondering whether the well and long established contrast between *savant/ non-savant* or *savant/populaire,* which has been used in all morphological works on French since Darmesteter, should be questioned too, once the results here presented are obtained.

References

Aronoff, M. 1994, *Morphology by itself*, Cambridge, MA, The MIT Press

Benveniste, E., 1935, *Origines de la formation des noms en indo-européen*, Paris, Maisonneuve.

Boyé G., 2000. *Problèmes de morpho-phonologie verbale en français, espagnol et italien*. Ph D thesis, Université Paris 7.

Bonami, O & G. Boyé, 2002, "Suppletion and dependency in inflectional morphology",in F. Van Eynde, L. Hellan & D. Beerman eds., *Proceedings of the HPSG '01 Conference*, Stanford: CSLI Publications.

Bonami O., G. Boyé & F. Kerleroux, 2004, "Stem Selection in Lexeme Formation: Evidence from French", 11 th International Morphology MeetingVienna.

Corbin, D., 1987, *Morphologie dérivationnelle et structuration du lexique,*Tübingen*,* Niemeyer.

Di Lillo, 1983, "Morphologie des noms en (t)ion du français". *Cahiers de lexicologie*, 43, 117-135.

Dubois, J., 1962. *Etude de la dérivation suffixale en français moderne*, Paris, Larousse

Fradin, B., 2003. *Nouvelles approches en Morphologie*, Paris, PUF.

Grimshaw, J. 1990, *Argument structure*, Cambrige, MA, The MIT Press.

Hathout N. & L. Tanguy, 2002, "Webaffix: finding and validating morphological links on the WWW." In *Proceddings of the Third International Conference on Language Resources and Evaluation,* pp. 1799-1804, Las Palma de Gran Canaria: ELRA.

Huot, H., 2001*, Morphologie – Forme et sens des mots en français,* Paris, A. Colin.
Kerleroux, F., 1996. *La coupure invisible — Etudes de syntaxe et de Morphologie*. Villeneuve d'Ascq: Presses universitaires du Septentrion.
Lecomte, Elsa, 1997. "Tous les mots possibles en –ure existent-ils?", *Silexicales* n° 1, Actes du Colloque de Villeneuve d'Ascq, CNRS 382 et Lille 3, (191-200).
Meyer-Lübke, Wilhelm, 1974, *Grammaire des langues romanes*, tome II, Morphologie, Genève, Slatkine reprints.
Pirrelli, Vito & Marco Battista, 2000. "The paradigmatic Dimension of stem Allomorphy in Italian Verb Inflection", *Rivista di Linguistica*, 12.
Stump, G., 2001, *Inflectional Morphology*, Cambridge: Cambridge University Prss.

Dictionnaries

TLF: *Trésor de la Langue Française*
Petit Robert
Grand Robert
DHLF: *Dictionnaire Historique de la langue française*, sld Alain Rey, 1992, Le Robert
Courtois, B. 1987, *Dictionnaire alphabétique inverse*, Rapport Technique du LADL, CNRS /Paris 7.

Have Cutthroats Anything to Do with Tracheotomes? Distinctive Properties of VN vs. NV Compounds in French

Fiammetta Namer
Université de Nancy 2 & UMR «ATILF»
Fiammetta.Namer@univ-nancy2.fr

Florence Villoing
Université de Paris 8 & UMR «Structures formelles du langage»
FlorenceVilloing@u-paris10.fr

Abstract

The aim of this paper is to compare word formation constraints that operate during the construction of compound lexemes in French; precisely, we contrast VN-structured lexemes (*ouvre-boîte*$_N$*:* open-tin =*' tin opener'*) with NV ones (*anthropophage*$_A$:°man-eat*(er)* = *'cannibal'*). The question is whether both VN and NV are obtained through the same rule (and differ only with respect to the chosen components), or whether each type corresponds to a specific compounding process. Our study is grounded on large-scale corpora of VN compounds on one hand, and NV ones, on the other hand. In the first part of the paper we therefore motivate (1) our claim that bound roots, such as °*anthropo* and °*phage* in *anthropophage* can be assigned lexical categories (although they are no autonomous lexemes), and (2) our decision to assign precisely the verbal category to bound roots such as °*phage*. The second part of the paper is devoted to the NV versus VN properties investigation, according to three criteria: a) the distribution of the lexical category and the semantic values for VN and NV, b) the process type denoted by V, and c) the thematic roles plaid by N with respect to V. The comparison of VN and NV, according to these three criteria, leads us to suspect that NV and VN are indeed formed, in French, by two distinct morphological rules.

1. Introduction

1.1. Data

Within French lexical morphology, compounding processes form either nouns or adjectives. Either type can be obtained through a variety of rules, where two nouns (1a), two adjectives (1b), a verb and a noun (1c) are involved.

(1)	a.	[timbre$_N$-poste$_N$]$_N$:	stamp-post =	'postage stamp'
		[homme$_N$-grenouille$_N$]$_N$:	man-frog =	'frogman'
	b.	[doux$_A$-amer$_A$]$_A$:	sweet-bitter =	'bittersweet'

G. Booij, *et al.* (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)*
Fréjus 15-18 September 2005, University of Bologna, 2007.
URL http://mmm.lingue.unibo.it/

c.	[coupe$_V$-papier$_N$]$_N$:	cut-paper =	'paper knife'
	[brise$_V$-coeurs$_N$] $_N$:	break-heart=	'heartbreaker'
	[perce$_V$-oreille$_N$] $_N$:	pierce-ear=	'earwig'
	(porte) [coupe$_V$-feu$_N$]$_A$:	break-fire=	'firebreak (door)'

In parallel, the French lexicon includes lexical units such as (2), which often appear to be made up of two Combining Forms[1], originating from Latin or Greek. These combining forms appear to be two nouns (2a), an adjective and a noun (2b), two adjectives (2c) or a noun and a verb (2d)[2].

(2)	a.	[anthropo$_N$morphe$_N$]$_A$:	human-shape =	'anthropomorphous'
	b.	[érythro$_A$cyte$_N$] $_N$:	red-cell=	'erythrocyte'
	c.	[afro$_A$-cubain$_A$]$_A$:	african-cuban=	'afro-cuban'
	d.	[anthropo$_N$phage$_V$]$_A$:	human-eat=	'anthropophagous'
		[mélo$_N$mane$_V$]$_N$:	music-like=	'music lover'
		[caverni$_N$cole$_V$]$_{A/N}$:	cavern-live (ing)=	'cavernicole'
		[infanti$_N$cide$_V$]$_{A/N}$:	child-kill=	'infanticide'

Our aim is to examine compound nouns and adjectives formed according to the patterns illustrated in (1c) and (2d), in order to propose a contrastive analysis. This study defends the hypothesis that compounds obtained through Word Formation rules at play in (2) belong to the current French morphological system, despite the fact that some compounds are borrowed from ancient languages (eg. Latin and Greek), and that a large amount of them are common to several contemporary languages (mainly Romance, Germanic and Slavic ones, cf. Darmesteter(1877)). Their productivity is confirmed by the amount and variety of neologisms that are found in the media (*bibliophobe, tabacolâtre, Jocondovore*). Moreover, at least three arguments support the idea that neoclassical compounds do not come under Lexeme Formation Rules belonging to the ancient languages that provide their Combining Form components:

I. Compounds are often formed through the association of two Combining Forms of various origins, and thus not systematically coming from the same Latin or Greek language: French+Latin (*insecticide*), Greek+French (*macromolecule* 'macromolecule'), French+Greek (*bureaucrate* 'bureaucrat'), Latin+Greek (*planisphère* 'planisphere'), French Truncated Form+French, Latin or Greek (*cinéphile* 'cinephile'*, anglophone, alcootest* 'drunkometer');

II. The emergence of an improper thematic vowel is often observed between the components: for instance, the thematic vowel "o", greek-specific, is also used at component boundaries originating from French (*franco-allemand* 'Germano-French'*, anglo-saxon*) or from Latin, instead of the expected "i" vowel (*cérébro-*

[1] We borrow the term 'Combining Form' to Warren(1990), Fradin(2000) and Iacobini(1999). In the examples, Combining Forms are preceded by the '°' symbol ; on that matter, see section 2.3.
[2] A table, at section Appendix B, provides a literal definition of the mentioned neoclassical compound lexemes, according to the meaning of the Combining Forms they contain.

spinal 'cerebrospinal'*, génito-urinaire* 'genitourinary') (Darmesteter(1894: 256));

III. Infringements of both Greek and Latin compounding rules are met in the formation of many neoclassical coumpounds. For instance:

i. Benveniste(1974:163-170) investigates the genesis of the noun *microbe*, in order to demonstrate that its constructed meaning ("little life", that is "microscopic organism") cannot be accessed from a Greek compounding pattern, which would lead to build the "short life" interpretation, which is incoherent from a scientific point of view.
ii. Darmesteter (1894:253) analyses the nouns *oxygène* 'oxygen' and *hydrogène* 'hydrogen', in order to show their semantic "ill" formation, given that, unlike in French, the Greek element corresponding to *gène* does not mean "which produces" (*oxygène*: "which produces acid", *hydrogène*: "which produces water") but instead, "which is produced".

More precisely, a Noun+Verb combining process emerges from examples (2d), similar to the French Verb+Noun compounding process[3], with two apparent differences, that is that I components come from the Greek or Latin lexicon , and II component ordering is reversed (in a neoclassical compound, governing (X) components are on the right, and governed (Y) ones on the left, resulting in the YX pattern, whereas 'ordinary' compounds are in the XY form). Our study is thus in line with traditional morphological analyses which identify a so-called "neoclassical" compounding results (eg. in 2d) in contrast with "ordinary" French component-based compounding ones (the traditional distinction was already current in the 19th century, cf. for instance Hatzfeld et al. (1890)). From this hypothesis, we now investigate differences and similarities between both VNs and NVs. Our objective is to identify and compare semantic and categorial constraints on these two types of compounds.

1.2. General issue

Assuming that lexical units in (2d) come under some NV compounding rule, our purpose is to determine whether VN and NV compounds are formed by two distinct compounding rules in French, or by one and the same rule. In other words:

- do VN and NV compounding rules correspond to a unique French lexeme formation process, with identical properties, except components origin and, consequently, components ordering[4]?
- do they correspond to two different lexeme formation processes in French, which do not form the same unit type, but depend on different semantic and categorial constraints ?

[3] Verb+Noun compounding exists in each Romance language : on that matter, see Rainer & Varela (1992), Scalise (1992), Villalava (1992)

[4] As many authors have remarked (see for instance Iacobini (2004), or Corbin (2005)), components ordering in « neoclassical » compounding is borrowed to ancient Greek and Latin.

1.3. Outline

To answer these questions, we have tried here to identify differences and similarities between VN and NV compounding patterns, especially through the contrastive study of both categorial and semantic constraints that are implied in these compounds' formation.

Our study is carried out from a corpus designed according to principles presented in section 2 ; in doing so, we also explain which theorical issues we had to face during the task of identifying components as Combining Forms, in NV compounds such as (2d). Section 3 is devoted to the comparison of VN and NV compound properties. We are led, in section 4, to the conclusion that the VN and NV compounds are formed by two distinct compounding rules in French, rather than a single one.

2. Data collection

2.1. VN corpus

This corpus contains 2-3000 VN compound lexemes. They have been gathered mainly from lexicographic sources: general language dictionaries, encyclopaedias. Dictionaries were chosen so that they are typical enough of both general and technical French lexica. Among them, there are: *Trésor de la Langue Française*, *Grand Robert de la Langue Française*, *Dictionnaire de la langue française* (Littré), *Dictionnaire Général de la Langue française* (Darmesteter and Hartzfeld).

Beside dictionaries, other sources have been used to collect VNs: Part-of-speech tagged corpora, where VNs could be extracted, and the Internet open corpus[5]. Also, newspapers, magazines and catalogues have also occasionally allowed us to enrich our VN corpus.

2.2. NV corpus

NV corpus belongs to the French lexicon, without being limited to a single speciality domain: limiting oneself to a specific domain is actually the risk to avoid when examining lexeme formation rules using combining forms.

Basically, two lexicographic sources have been used to build up this corpus: the machine-readable version of *Trésor de la Langue Française*, and the *Dictionnaire des structures du vocabulaire savant*. The latter, compiled by H. Cottez, groups and defines some 2,700 components used in learned words formation, be these components combining forms or learned affixes.

The first version of the NV corpus resulted from a semantic intuition about which of the combining forms should actually belong to the Noun and Verb categories. On the basis of this first corpus version, we have applied a series of (still tentative) principles in order to confirm component categories. We give an account of this analysis in section 2.3. In all, we have gathered about 500 NV compound adjectives and nouns,

[5] Many thanks to M. Plénat, who sent us so many examples.

which satisfy category criteria detailed below.

2.3. Identifying constituents in NV compounds

Making up NV corpus raised important theoretical issues which are related to the identification of VN components. Several problems were posed:

i. How to recognize a compound lexeme among all the morphologically complex lexemes formed out of combining forms. In other words, how to be sure that their constituting parts are base-lexemes, and not affixes[6]?

ii. Consequently, how can 'verb' and 'noun' categories be identified for these base-lexemes, whose particular property is never autonomous in syntax?

2.3.1. Theoretical framework

These questions make sens within the lexeme-based theoretical framework (Anderson(1992), Aronoff&Fudeman(2004), Booij(2005), Fradin(2003)). In this framework, basic units in lexeme (cf. Matthews(1991)) formation processes belong to one of the major categories: N, V, Adj, and are clearly distinguished from affixes, which are nothing but exponents of phonological rules. Within this approach affixes are not assigned a major category, unlike other theories (cf. Williams(1981), Lieber(1981), Lüdeling et al. (2002)). Our position is to extend lexeme-based morphology principles to bound roots, that is to assume that Greek and Latin Combining Forms can be N, V, or Adj categorized lexemes, in the same way as free roots are. This point of view fundamentally differs from those which admit the existence of Combining Forms, without assigning categories to them (cf. Rey-Debove(2004), and Cottez(1988: VII-XX), who claims "learned vocabulary forming elements" to be "signs", that is morphemes, but provides them with neither of base-lexeme nor affix status.

On the other hand, we follow the theoretical line adopted (among others) by Corbin(1985), Warren(1990), Fradin(2000;2003) and Iacobini(2004), who make a distinction, among learned Combining Forms, between base-lexemes on one side and affixes on the other side. So, designing or NV corpus is based on the fact that Combining Forms have been identified as base-lexemes, in compounds words such as *anthropophage*.

2.3.2. Affix or Base-Lexeme (N, V, A)?

- Are compound components in example (2) base-lexemes (N, V, A) or affixes?

This question is relevant because, in words such as *anthropophage*, Combining Forms properties (eg. *anthrop(o)* and *phage)* make them closely related to base-lexemes as

[6] Rather than **« lexeme »**, we prefer to adopt the term of « **base-lexeme ».** According to authors, the former may include lexical units such as determiner, pronoun, preposition (cf. Aronoff & Fudeman (2004)), and these categories are irrelevant in Lexeme Formation.

well as to affixes.

Combining Forms look like affixes, because they are bound roots. They fulfill a structural position within lexical units, not within phrases. Combining Forms look like base-lexemes, because they have categorial and semantico-referential properties.

2.3.2.1. Semantic criterion

This criterion relies on the dichotomy between "lexical sense" and "grammatical sense", frequently mentioned in the literature (cf. Corbin(2001) who applies it to French morphology, and Kleiber(1999), to lexical semantics)

A base-lexeme is an abstract lexical unit, with context-independent phonological form, syntactic category and meaning. Unlike affixes, which do not refer to entities but are given a semantic instruction, Combining Forms refer in a stable way to a referential category, a linguistically encoded concept: and so do base-lexemes. In other words, CFs carry a referential meaning ; they are able to identify concepts, but they are not able to name them within utterances (Corbin 85, Iacobini 99, Fradin 2000). However, there are serious drawbacks when attempting to use this criterion:

i. identifying the component referential value depends on its translation and this translation lacks reliability. A big diversity is observed among dictionaries (and, sometimes, within a single dictionary) when looking up at a Combining Form translation: more especially, CF denoting processes (eg °*graph(o)*) are sometimes translated by a verb ("to write"), and sometimes by its corresponding deverbal noun ("writing"). And as far as the Robert Brio is concerned, for instance, both translations are given.

ii. identifying the component referential value also depends on the morphologically constructed meaning of the compound they are part of, and on the semantic relation between the compound constituents. Now, semantic features provided by dictionaries are often unsufficient: either they are not precise enough, or they describe referents rather than meanings. The lexicographic definitions of *pleuronecte*, for instance (« type of flat fish such as sole, turbot, halibut…") and *bathyergue* 'bathyergus' («rodent mammal from Africa, whose behaviour is close to mole") prevent us from access to the morphologically constructed meaning, because they have to do only with the referent description. Actually, *pleuronecte* and *bathyergue* are named according to their behaviours: a *pleuronecte* swims (°*necte*$_V$) on one side (°*pleur(o)*$_N$); a bathyergue works (°*ergue*$_V$), that is, "digs", in depth (°*bathy* $_N$).

Moreover, this semantic criterion is hard to use when base-lexemes have undergone a grammaticalization process: in that case, they behave as affixes; and their semantic value vacillates between referential pole and instructional pole (cf. Amiot&Dal(2005) about °*-logue*).

2.3.2.2. Categorial criterion

This criterion relies on the base-lexeme property of belonging to one of the major part-of-speech categories: noun, verb, adjective, adverb. In other words, establishing that a CF belongs to one of these categories is sufficient to guarantee its base-lexeme status. However, as CFs never have syntactic realization, none of the usual identification syntactic tests (distribution in utterances, their syntactic behaviour and their inflectional marks) can operate in order to decide whether they are base-lexemes or not. To meet this specificity, other means are required. Four of them are presented here

- **Means 1**: CF can be assign a part-of-speech category according to its translation or interpretation in French. It simply gets the same category as its translation. But this solution is rather limited, due to translation uncertainty (see above).

- **Means 2**: A CF allomorph to a base-lexeme is clearly categorizable. But the allomorphy notion sometimes conceals suppletion phenomena, and relations between a base-lexeme and its hypothetical allomorph may lack transparency. We will not address this issue here (cf Haspelmath(2002:26-28) for a synthetic description of the issue, and Boyé(2006) for a presentation of the suppletion).

- **Means 3:** CFs which are selected as affixation bases are nothing but base-lexemes. From there, if an affixation process obeys constraints on its base such that only one base category is selected, then any CF used as a base for this process, automatically gets an unambiguous category (cf. Corbin (1985:62-64; 1987:182-sq)). However, only few affixes obeys such strong categorial constraints; so in general this technique cannot determine for certain the value of base categories.

- **Means 4**: CFs can also be assigned the lexeme status (and a category) according to the morphological construction in which they are found. In particular, the semantic relation type reveals its categorial counterpart. For example, in YX compounds where there is some typical argument-to-verb relation between Y and X, we can expect X to be a verb, and Y, to be a noun.

Through combination of the means just presented, we are able to identify verbal (X component) and nominal (Y component) Combining Forms in morphologically complex lexemes such as illustrated in (2d). Examples (3) and (4) give an account of the way we have proceeded to category identification.

(3) *oculogyre*$_{\text{ADJ}}$[7]

a. °*ocul*(o)

b. °*gyre*

[7] From a lexicographic point of view, *oculogyric* qualifies muscles « that enables eyballs to turn ».

The CF °*ocul* occurs as base in the adjective *oculaire*; this allows to label it with the category "noun", via Means3. The reason is that suffix *-aire* exclusively forms adjectives from nouns: *pôle*$_{\text{NOUN}}$ → *polaire*$_{\text{ADJ}}$. Given that *oculaire* is formed by means of this process, from the base °*ocul*, the latter is not but a noun.

Similarly, categorial constraints exerted by *-oire* and *-ion* French suffixes are used to provide (3b) with a grammatical category, thanks to Means3. In French, the adjective *giratoire* 'gyratory' (also found written *gyratoire*) is given the definition: "revolving around a point or axis"; and the noun *giration* 'gyration': "act or instance of turning". We will assume that the base of these lexemes, respectively formes by *-oire* and *-ion* suffixation, is the allomorphic variant °*gir(at)* of °*gyre*. Each of these suffixes preferentially selects preferentially verbal bases (*méritoire*$_{\text{A}}$ 'meritorious' < *mériter*$_{V}$; 'to deserve,merit' *agression*$_{\text{N}}$ < *agresser*$_{\text{V}}$ 'commit agression'), but not exclusively (*républicatoire*$_{\text{A}}$ 'republicatory'< *république*$_{\text{N}}$ 'republic'; *baladoire*$_{\text{A}}$ '°balladory'[8]< *balade*$_{\text{N}}$ 'ballad'; *zonation*$_{\text{N}}$ < *zone*$_{\text{N}}$). However, when the same sequence is used as base for both affixed lexemes, it is nothing but a verbal base (*revendicatoire*$_{\text{A}}$ 'revendicatory', *revendication*$_{\text{N}}$ < *revendiquer*$_{\text{V}}$ 'revendicate'). This is confirmed by a formal clue: When the base of a *-ion* suffixed noun (or that of a *-oire* suffixed adjective) ends with *-at*, then it is identified as the supine form of a Latin verb. So, as °*girat*/°*gyrat* results from a Latin verb form, it is indeed an allomorph of °*gire* or °*gyre*. And therefore, °*gyre* is a verb.

So, Means3 allows us to identify both a nominal bound constituent (°*ocul*) and a verbal one (°*gyre*) in *oculogyre*$_{\text{A}}$. In other words, using Means3 may lead us to analyse *oculogyre* as a compound NV adjective.

(4) *anthropophage*$_{\text{ADJ}}$

a. Means1

b. This hypothesis is confirmed by Means4

Means 1 makes us assume that °*anthrop* is a noun and that °*phage* is a verb; °*phage* has to be analysed as a verb, because its semantic relation with °*anthrop* is of predicate-argument type: *phage* denotes a process which applies on its proto-patient argument, realized by °*anthrop*.

Finally, Means1 and Means4 cooperate to tell us that *anthropophage* is an NV compound adjective.

3. Analysis: three criteria

Once the task of gathering NV and VN corpora has been carried out, the analysis core has been that of comparing "neoclassical" compounds NV, and French-components based compounds VN. Three criteria **(c1)** to **(c3)** served at that purpose:

(c1) category and semantic values, for VNs and NVs ;

[8] *baladoire* qualifies « undecent » dances: the adjective originates from the musical genre « ballad ».

(c2) the type of process denoted by V verbs, for each compound type;

(c3) the semantic relation that holds between V verb and N noun

By crossing these three criteria, we ask questions such as: are VN and NV properties identical, in partial overlapping or in complementary distribution ? Are there lexical gaps, corresponding to unattested **(c1)** to **(c3)** combinations ?

3.1. VN compound properties[9]

3.1.1. Criterion C1: category and semantic types of VN compounds

- **Category**: VN compounding mainly builds nouns, and rarely adjectives. The latter are illustrated in (5):

(5) (*papier*) *tue-mouches*$_A$: kill-flies (paper) = 'flypaper'
(*porte*) *coupe-feu*$_A$:: break-fire = 'firebreak (door)'

- **Compound semantic type**

Most VN nouns denote **artefacts**:

(6) *ouvre-boîte*$_N$: open-tin = 'tin opener'
coupe-papier$_N$: cut-paper = 'paper knife'

Others refer to (human or animal) **animate entities**,

- either by their usual **function** or **occupation**

(7) *garde-barrière*$_N$: guard-gate = 'level-crossing keeper',
gratte-papier$_N$: scratch-paper = 'penpusher'

- or by some **characteristic**, **salient property** or **behaviour**

(8) *trouble-fête*$_N$: disturb-party: 'killjoy'.
(*garcon*) *rabat-joie*$_A$:°reduce-joy (boy) = 'spoilsport'
perce-oreille$_N$: °pierce-ear ='earwig' *(insect)*

Finally, a few VN nouns describe either **locations**

(9) *coupe-gorge*$_N$: cut-throat = 'cut-throat alley'
garde-meuble$_N$: mind-furniture = 'store house'

- or **events**

[9] On this matter, see among others Villoing (2003) and Fradin (2005).

(10) *lèche-vitrine*$_N$*: lick-window* = 'window-shopping'
(jouer à) saute-mouton$_N$: leap-sheep = 'leapfrog' (play)

3.1.2. Criterion C2: the type of process denoted by V

In a VN compound, the process type denoted by the verb is strongly constrained ; according to Vendler's (1967) terminology, and that of his successors (among others, Dowty(1979)), it can only be **dynamic,** that is compatible with linguistic contexts such as "être en train de V [be Ving] / se mettre à V [start to V]/ s'arrêter de V [stop Ving] ", cf. (11) (for a detailed presentation, see Villoing(2003)). On the other hand, V process type is very unlikely to be **stative** (12):

(11) *coupe-papier* (°cut-paper 'paper knife')
Jean est en train de couper le papier/ s'est mis à couper le papier /s'arrête de couper le papier.
'Jean is cutting the paper/started to cut the paper/stops to cut the paper'.

(12) **sait-latin* (°know-latin)
*Jean est en train de savoir le latin/s'est mis à savoir le latin/s'arrête de savoir le latin.
*'Jean is knowing latin/started to know latin/stops to know latin'

3.1.3. Criterion C3: V to N semantic relation

There are strong restrictions on the semantic relations between V and its participants; and consequently, on roles that N may play with respect to V. According to Dowty (1991) who refutes thematic role labelling, and proposes instead a Proto-Agent/Proto-Patient continuum, N corresponds quite exclusively to V **Proto-Patient**, within a VN compound. The reason is that N satisfies "change-of-state" and "affectedness" criteria (cf. also Foley&Van Valin (1984) and Jackendoff (1990) for a definition of these criteria).

This analysis also applies for a small number of VNs, like in (13) that seem to be based on intransitive verbs and that for which N looks like the verbal Proto-Agent argument:

(13) *trotte-bébé* (°toddle-along-baby = 'baby walker') ;
pense-bête (°think-of-silly = 'reminder').

However, the semantic of the verb is causative and N is involved as a Proto-patient

(14) *trotte-bébé* 'causes/makes the baby to toddle along' ;
pense-bête 'causes/makes the silly to think'.

Moreover, we can observe other (less frequent) semantic relations between V and N in VN compounds. For instance, we find rare cases (such as (15)) where N doesn't meet

Proto-patient criteria but rather Proto-agent ones.

(15) *cuit-vapeur*: cook-steam (saucepan type)
pousse-pied: push-foot (boat that we push with foot)

Besides, there are VNs such as (16), where N can be analysed neither as a Proto-patient , nor as a Proto-agent of the verb. Rather, N seems to correspond to a temporal (16a) or locative (16b) verb modifier.

(16) a. *réveille-matin*: wake-up-morning = 'alarm clock'
grille-midi: scorch-midday (plant)
b. *croque-télé*: crunch-tv (tv tray)

3.2. NV compound properties

3.2.1 Criterion C1: Category and semantic types of NV compounds

- **Category:** generally, NV compounding produces adjectives like (17) which function equally as nouns. (Kerleroux 1991, Kerleroux 1996, Fradin 1997)

(17) *ventriloque*$_{\text{ADJ}}$ ('ventriloquist')

- **Semantic type**.

When only the nominal category is realized, NV may refer:

- mostly, to a **concrete object**:

(18) *odontoclaste*$_N$ (cell which breaks (°claste$_{\text{V}}$) teeth (°odonto$_{\text{N}}$)) ;
trachéotome$_N$ ('tracheotome': scalpel used to incise (°tome) the trachea (°trache(o)))

- to an **animate entity**,

(19) *biographe*$_N$ ('biographer')
notonecte$_{\text{N}}$ ('noctonect')

- sometimes, to an **event**:

(20) *lipolyse*$_N$ ('lypolysis')
infanticide$_{\text{N}}$ ('infanticide')

3.2.2. Criterion C2: the type of process denoted by the verb

The process denoted by the verbal combining form may be

- either **dynamic**

(21) °*ambule*$_V$ = 'to walk' (*noct<u>ambule</u>*$_A$: 'late-night/night reveller')) *;*

- or **stative**.

Among **stative** processes, most verbs describe **spatial relations** (Talmy 2000, Vandeloise 1986).

(22) °*phore*$_V$, = 'to carry' *(mélano<u>phore</u>*$_N$: 'melanophore'),
°*fère*$_V$ = 'to contain' (*carboni<u>fère</u>*$_A$: 'carboniferous').

Others are **feeling** or **emotion** verbs (according to Levin classification in (Levin 1993)):

(23) °*phile*$_V$ = 'to like' *(russo<u>phile</u>*$_A$: 'Russophil'*)*
°*mane*$_V$ = 'to like' *(mélo<u>mane</u>*$_A$ 'music lover'*)*
°*lâtre*$_V$ = 'to *adore' (wagnerolâtre*$_A$ = who adores Wagner*),*
°*phobe*$_V$= 'to hate' *(anglo<u>phobe</u>*$_A$: 'anglophobic')

Other **stative** processes are **perception** verbs

(24) °*op* $_V$ = 'to see' (*nyctalope*$_A$: 'day-blind')

or verbs of **existence** (Levin & Rappaport 1995)

(25) °*cole*$_V$ = 'to live' *(limicole*$_A$: 'limicolous'; *cavernicole*$_A$*)*
°*bie*$_V$ = 'to live' (*phyllobie*$_N$ = (organism) living on leaves: 'green leaf weevil'),

We also found one verb of **light emission (**Levin & Rappaport 1995)

(26) °*luque* = 'to shine' (*noctiluque*$_A$: 'noctilucent')

3.2.3. Criterion C3: V to N semantic relation

In NV compounds, semantic relations between N and V belong to two distinct dimensions, according to the verbal process type.

- For **dynamic verbs**, Dowty's (1991) criteria are applied (as in the case of VN compounds). As a result, we have observed that here the noun can play different semantic roles wrt the verb:

  i. N may fulfill the **Proto-agent** criterion:

(27) *psychogène*$_A$: 'psychogenic'

  ii. (more frequently) it corresponds to a **Proto-patient**

(28) *lipolyse*$_N$: 'lypolysis'

iii. It may also be the case that N **does not meet any** of the proto-role requirements

(29) *ventriloque*$_A$ ('ventriloquist') ;
noctambule$_A$ ('late-night/night reveller')
héliotrope$_A$ ('heliotrope')

For **stative** verbs, Dowty role assignment criteria are not relevant. Rather, verb-to-noun relations can be expressed by means of Talmy Figure / Ground notions (Talmy(2000)[10]). So, when V denotes a **spatial relation**, then N refers:

- mainly to the **Figure**: (ex 25),

(30) *mélanophore*$_N$ ('melanophore')
carbonifère$_A$ ('carboniferous')

- or, sometimes, the **Ground**,

(31) *vasiducte*$_N$ (= 'carrying vessel')

When V describes a **predicate expressing an emotion or a feeling**, it sets up a relationship between two participants: the experiencer and the stimulus. We propose then to identify these verbs as spatial relations: experiencers are interpreted as **grounds** and stimuli as **figures**.

As far as we know, these verbal predicates only combine with **figure** denoting nouns: (ex 27)

(32) *mélomane*$_A$ ('music lover')
anglophobe$_A$ ('anglophobic')

4. Results

By crossing criteria **c1** to **c3**, (cf. the comparative table in section Appendix A), we notice the following contrasts between VNs and NVs:

NV compound nouns and adjectives cover a much broader spectrum than VN compounds, whatever the examined criterion:

- Categorial: NVs belong equally to ADJ or NOUN category (whereas VN compounds are almost exclusively nouns);

[10] Talmy (2000) investigates non-agentive relations, between participants of a so-called **spatial** or **localisation event**. Such events involve two main roles : the Figure and the Ground ; incidentally, other roles like Manner, Cause, or Path, may intervene. The **Figure** is identified as the moving or localized object, viewes with respect to another object (the reference objet : **Ground**).

- V process-type: V components in NV compounds describe either dynamic or stative processes, unlike in VN compounds.

- N to V semantic relation: in VN compounds, the N component is typically a proto-patient; but it fulfills a wider amount of semantic roles in NV compounds: proto-patient, proto-agent, or none, with respect to dynamic V components; and either figure or ground with respect to a stative V component which expresses a spatial relation.

In other words, NVs instanciate a wide range of combinations **c1-c3**, whereas VN compounds mainly correspond to those combinations where V is dynamic and N is proto-patient.

5. Conclusion

In some cases (when the verb is dynamic) VN and NV compounds seem to share similar categorial and semantic constraints, and this would lead us to conclude that only one compounding rule is at play and selects various types of components. However, VN and NV have clearly different behaviour when V is stative. In this case, it can be selected only as NV compounds governing component. Consequently, only the NV compounding rule can build spatial relations between N and V. 'Ordinary' morphology, by means of VN compounding, is unable to construct meaning equivalent to what NV compounding offers: syntactic patterns are required. There is, for instance, no other ways to express "melomane", but with a syntagmatic expression: "music mad".

The results of this study lead us to think that two distinct compounding rules are at play in French. This conclusion raises two major questions:

- why is VN compounding such a constrained process?

- where do NV rules come from? Should we consider them as belonging to the French linguistic system, or are they borrowed from those languages where the components come from (ie Latin, Greek) ?

The latter question is part of a larger issue, in which we try to determine whether neoclassical compounding makes use of rules inherited from the very languages components originate, or not. Following comparativist authors (Darmesteter (1894)) (and unlike eg. Benveniste(1974)) the inheritance hypothesis seems to be the mainstream.

Therefore, we tried to assess this assumption, in the specific case of NV compounds, where both N and V are of Greek origin. The underlying hypothesis is the following: if either nominal or adjectival NV compounds formation were known in ancient Greek, then we would have arguments in favour of borrowing; consequently we could provide the contemporary formation of this compounds type with an historical explanation.

However, various studies focusing on compounding in ancient Greek

(Chantraine(1933), Smyth(1920), and recent studies by A. Ralli[11]) are divergent. Only Chantraine identifies an NV compounding pattern in ancient Greek. Conversely, Smyth and Ralli consider (for different reasons) that this formation did not exist in ancient Greek. This divergence in analysis show how weak is an hypothesis that assumes neoclassical compounds to be formed by means of borrowings of classical language rules.

But do Latin and Greek have no influence at all on modern compound formation? Works devoted to neoclassical formation in French all tend towards an hybrid answer, be they in favour of the above-mentioned inheritance hypothesis or not. This is shown, for instance, by the way A. Darmesteter (1894:252-256) analyses Greek-component based neoclassical compounds. A. Darmesteter show how these compounds come from some (often improper) re-analysis of ancient Greek to French, in a section where he laments the massive intrusion of new neoclassical compounds in the 19th century French lexicon, compounds that are often disrespectful to the rules of the langages they originate from. For instance, the creation of a neoclassical compound in 19th century French, such as *adénographie*$_N$ (*adenography* = 'written or drawn record of the glands') would go through the prior invention of the equivalent word the French compound would have in ancient Greek, if this equivalent existed (e.g. °αδενογραφια = *adénographia*), from components (included verbs) which were attested in ancient Greek (e.g. αδην$_N$ = *aden-* (gland), γραφω$_V$ = *-grapho* (to write)), and in compliance (real or presumed) with ancient Greek morphological rules.

The impact of Greek (as well as that of Latin, which functioned as intermediary) is thus unquestionable, but the role played by the French linguistic system is equally crucial. We can see, indeed, that neoclassical compounds formation rules have been elaborated by French speakers having an indeniable knowledge of ancient languages.
This assumption that Greek has been re-analysed in French has been shared by various linguists all the way through the 20th century. It is found, for instance, in Benveniste(1974:170) and later, in Bouffartigue et Delrieu (1996). In further research, we still have to determine which has been the influence of the VN compounding pattern, which was already very productive at the end of the 18th century, upon the formation (or the reanalysis) of French NV neoclassical compounds.

6. References

Amiot, D. & Dal, G., 2005, « Integrating Combining Forms into a Lexeme-based Morphology», in *Fifth Mediterranean Morphology Meeting*, 15-18 sept, Fréjus

Anderson, S.R., 1992, *A-morphous Morphology*, Cambridge: Cambridge University Press.

Aronoff, M., & Fudeman, K., 2005, *What is morphology?*, Blackwell Publishing, Oxford.

Benveniste, E., 1974, "Formes nouvelles de la composition nominale", *Problèmes de linguistique générale II*, Paris, Gallimard, p. 163-176.

Booij, G., 2005, *The Grammar of Words: An Introduction to Linguistic Morphology*. Oxford, Oxford University Press.

Bouffartigue, J. et Delrieu A.-M., 1996, *Les racines grecques*. Paris, Belin.

Boyé, G., (2006), "Suppletion", *Encyclopedia of Language and Linguistics, 2nd Edition*, K. Brown, London/Amsterdam, Elsevier, p. 297.

Chantraine, P., 1933 (1979), *La formation des noms en grec ancien*. Paris, Champion

[11] Personal communication

(Klincksieck).
Corbin, D., 1985, „Les bases non-autonomes en français ou comment intégrer l'exception dans le modèle lexical", *Langue française* n°66, Paris, Larousse, pp. 54-76.
Corbin, D., 1987, *Morphologie dérivationnelle et structuration du lexique.* Lille, Presses Universitaires de Lille.
Corbin, D., 2001, "Préfixes et suffixes: du sens aux catégories", *Faits de Langue*, Paris, Ophrys, 15, pp. 41-69.
Corbin, D., 2005, "French (Indo European: Romance) " in Booij, Lehmann & Mugdan (eds) Morphology An International Handbook on Inflection and Word Formation, Berlin/New York, Walter de Gruyter, vol1, art. 121.
Cottez, H., 1988, *Dictionnaire des Structures du vocabulaire savant. Eléments et modèles de Formation*, 4ème édition. Paris, Dictionnaires Le Robert.
Darmesteter, A. 1894. *Traité de la Formation des mots composés dans la langue française, (1ère édition: 1875)*. Paris: Honoré Champion.
Darmesteter, A., 1877, *De la création des mots nouveaux dans la langue française et des lois qui la régissent*. Paris, Wieveg.
Dowty, D., 1979, *Word Meaning and Montague Grammar*. Synthese Language Library, n°7, Dordrecht: Reidel.
Dowty, D., 1991, « Thematic Proto-roles and Argument Selection ». *Language*, 67(3), pp. 547-619.
Fradin, B. 1997, "Esquisse d'une sémantique de la préfixation en anti-." *Recherches Linguistique de Vincennes*, 26: pp. 87-112.
Fradin, B., 2000, "Combining forms, blends and related phenomena", in Dolsescha Ursula, Thornton Anna (eds). *Extragrammatical and Marginal Morphology*. Münich, Lincom Europa. pp 11-59.
Fradin, B., 2003, *Nouvelles approches en morphologie*, Paris, Puf.
Fradin, B. 2005,. « On a semantically grounded difference between derivation and compounding". In *Morphology and its Demarcations*, Dressler W. U., D. Katovsky & F. Rainer (eds) Amsterdam / Philadelphia: John Benjamins.
Foley, W. A., and Van Valin, R. D., 1984.*Functional syntax and universal grammar.* Cambridge: Cambridge University Press.
Haspelmath, M., 2002, *Understanding Morphology*. London, Arnold.
Hatzfeld, A., Darmesteter, A., et al, 1890, *Traité de la formation de la langue*, Livre premier, Paris, Delagrave.
Iacobini, C., 1999, "Distinguishing derivational prefixes from initial combining forms", in *Proceedings of First Mediterranean Morphology Meeting*, Mytilene (Grecia), 19-21 septembre 1997, G. Booij, A. Ralli, S. Scalise (eds.), pp. 132-140.
Iacobini, C., 2004, "Composizione con elementi neoclassici", in M.Grossmann / F. Rainer (a c. di) *La formazione delle parole in italiano*, Tübingen, Niemeyer.
Jackendoff, R., 1990, *Semantic Structures*. Cambridge Mass., MIT Press.
Kerleroux, F, 1991, "Les formes nues sont-elles simples ?", *Linx 25*, p. 21-44.
Kerleroux, F. 1996, *La coupure invisible: études de syntaxe et de morphologie*. Villeneuve d'Ascq, Presses Universitaires du Septentrion.
Kleiber, G., 1999, *Problèmes de sémantique - la polysémie en questions*. Villeneuve d'Ascq, Presses Universitaires du Septentrion.
Levin, B. 1993. *English Verbs Classes and Alternations*. Chicago, IL: University of Chicago Press.
Levin, B. & Rappaport H.M., 1995. *Unaccusativity*. Cambridge, MA: MIT Press.
Lieber, R., 1981. *On the organization of the lexicon*. Doctoral dissertation, MIT.
Lüdeling, A., T. Schmidt, et al., 2002, "Neoclassical word formation in German." *Yearbook of Morphology 2001*,pp. 253-283.
Matthews, P. H. 1991, *Morphology, An introduction to the theory of word-structure*, Cambridge University Press, Cambridge.

Rainer, F. & Varela, S., 1992, « Compounding in Spanish », *Rivista di linguistica* 4 (I), Pisa.
Rey-Lebove, J., 2004, *Le Robert brio.* Paris, Dictionnaires le Robert.
Scalise, S., 1992, « Compounding in Italian », *Rivista di linguistica* 4 (I), Pisa.
Smyth, H. W., 1920, *A Greek Grammar for Colleges* (libre accès à l'URL: http://www.ccel.org/s/smyth/grammar/html/toc.htm). New York, American Book Company.
Talmy, L. 2000, *Toward a cognitive semantics (2 volumes)*. Cambridge, MA, Massachusetts Institute of Technology.
Vandeloise, C. 1986, *L'espace en français: sémantique des prépositions spatiales*, Les Editions du Seuil, Paris.
Vendler, Z., 1967, "Verbs and Times", *Linguistics and Philosophy*, Z. Vendler, Ithaca, Cornell University Press, pp. 97-121.
Villalava, A., 1992, « Compounding in Portuguese», *Rivista di linguistica* 4 (I), Pisa.
Villoing, F., 2002, "*Les mots composés [VN]N/A du français: réflexions épistémologiques et propositions d'analyse*." Nanterre, Université de Paris 10.
Villoing, F., 2003, « Les mots composés VN du français: arguments en faveur d'une construction morphologique », *Cahiers de Grammaire 28*, « Morphologie et Lexique », pp. 183-196.
Warren, B., 1990. "The importance of combining forms" in Dressler W. U. et al. (eds.) 1990: 111-132.
Williams, E., 1981, "On the Notions `Lexically Related' and 'Head of a Word'." *Linguistic Inquiry*, 12,(2), pp. 245-274.

Dictionaries

Gaffiot, F., 2000, Le Grand Gaffiot - Dictionnaire Latin-Français. Paris, Hachette.
Bailly, A., 2000, Dictionnaire Grec-Français, Le Grand Bailly. Paris, Hachette Education.
Dictionnaire historique de la langue française, 1993, Dictionnaires le Robert, Paris.
Cottez, Henri, 1988, Dictionnaire des structures du vocabulaire Savant, Eléments et modèles de formation, Les usuels de Robert, Paris, 4e édition.
Le Grand Robert Electronique, 1994, CD-ROM.
Le Robert Brio, Analyse des mots et régularités du lexique, 2004, Dictionnaires Le Robert-VUEF, Paris.
Trésor de la Langue Française, Dictionnaire de la langue du XIXe et du XXe siècle, 16 volumes, 1971-1994, CNRS, Gallimard, Paris. (TLFi: http://atilf.atilf.fr/tlf.htm et CDROM: CNRS édition, 2004.)

Appendix A

	Noms				Adjectives
	VN/**NV** concrete +anim	VN/**NV** concrete -anim	VN/**NV** event	VN/**NV** location	
Dynamic V + patient N	trouble-fête *(killjoy)* garde-barrière *(level crossing keeper)* gratte-papier *(penpusher)* perce-oreille (earwig) **biographe** *(biographer)* **anthropophage** *(anthropophagous)*	ouvre-boîte *(tin opener)* coupe-papier (*paper knife*) trotte-bébé *(baby walker)* pense-bête *(reminder)* **odontoclaste** *(odontoclast)* **trachéotome** *(tracheotome)*	lèche-vitrine *(window-shopping)* saute-mouton (*leapfrog*) **lipolyse** *(lipolysis)* **infanticide** *(infanticide)*	coupe-gorge *(cutthroat)* garde-meuble *(store house)*	tue-mouches *(kill-flie')* rabat-joie *(spoilport)* coupe-feu *(firebreak)* **anthropophage** *(anthropophagous)* **oculogyre** *(oculogyric)*
Dynamic V + agent N		cuit-vapeur *(cook-steam)* pousse-pied [type of boat]			**psychogène** *(psychogenic)*
Dynamic V + other participant N	**notonecte** *(notonect)* **noctambule** (*night reveller*) **ventriloque** *(ventriloquist)* **pleuronecte** **bathyergue** *(bathyergus)* **xylographe** *(xylographer)*	réveille-matin *(alarm clock)* **électrographe** *(electrographer)* *grille-midi* [plant name] *croque-télé* (tv tray) **héliotrope** (*heliotrope*)			**plantigrade** *(plantigrade)* **ventriloque** *(ventriloquist)* **héliotrope** *(heliotrope)* **noctambule** *(late-night, night-time reveller)*
Stative V Spatial relation + Ground N		**vasiducte** *('carrying vessel')*			
Stative V Spatial relation + Figure		**mélanophore** *(melanophore)*			**carbonifère** *(carboniferous)*

N Stative V Feeling + Figure N					**mélomane***(music-lover)* **wagnerolâtre** *(who adores Wagner)* **russophile** *(russophil)* **anglophobe** *(anglophobic)*
Stative V – other N	**cavernicole**				**nyctalope** *(day-blind)* **cavernicole** **limicole** *(limicolous)* **phyllobie** *(living-on-leaves)* **noctiluque** *(noctilucent)*

Appendix B

anthropophage$_A$	'anthropophagous'	*Who °phage$_V$ = 'eat' °anthropo$_N$ = 'human beings'*
anglophobe$_A$	'anglophobic'	*Who °phobe$_V$ = ''hates' °anglo$_N$ = 'the British'*
biographe$_N$	'biographer'	*He who °graphe$_V$ = writes' bio$_N$ = 'life'*
bathyergue$_N$	'bathyergus'	*(Sort of mole who) °ergue$_V$ = 'works' in °bathy$_N$ = 'depth'*
carbonifère$_A$	'carboniferous'	*Which °fère$_V$ = ''carries' °carboni$_N$ = 'carbon'*
cavernicole$_A$	'cavernicole'	*Which °cole$_V$ = 'lives' in caverne$_N$ = 'caverns'*
électrographe$_N$	'electrographer'	*(Instrument that) °graphe$_V$ = 'writes' by means of °électro$_N$ = 'electricity'*
héliotrope$_A$	'heliotrope'	*Which °trope$_V$ = 'turns' towards the °hélio$_N$ = 'sun'*
infanticide$_N$	'infanticide'	*Action of °cide$_V$ = ''kill' a °infanti$_N$ = 'child'*
limicole$_A$	'limicolous'	*Which °cole$_V$ = 'lives' in °limi$_N$ = ''mud'*
lypolyse$_N$	'lypolysis'	*Action of °lyse$_V$ = 'destroy' °lypo$_N$ = 'fat'*
mélanophore$_N$	'melanophore'	*(Cell that) °phore$_V$ = 'carries' °mélano$_N$ = 'melanin'*
mélomane$_A$	'music lover'	*Who °mane$_V$ = 'likes' °mélo$_N$ = 'music'*
nyctalope$_A$	'day-blind'	*Who °ope$_V$ = 'sees' at °nyctal$_N$ = 'night'*
noctambule$_A$	'late-night, night reveller'	*Who °ambul$_V$ = 'walks' at °noct$_N$ = 'night'*
noctiluque$_A$	'noctilucent'	*Which °luque$_V$ = 'glows' at °nocti$_N$ = 'night'*

notonecte$_N$	'noctonect'	*(Insect that) °necte$_V$ = 'swims' on its °noto$_N$ = 'back'*
odontoclaste$_N$	'odontoclast'	*(Cell that) °claste$_V$ = 'breaks' °odonto$_N$ = 'teeth'*
oculogyre$_A$	'oculogyric'	*(Muscle that enables) °ocul$_N$ = 'eye(ball)' to °gyre$_V$ ='turn'*
phyllobie$_A$	'green leaf weevil'	*(organism that) °bie$_V$ = 'lives' on °phyll$_N$ = 'leaves'*
plantigrade$_A$	'plantigrade'	*Which °grade$_V$ = 'walks' on its °planti$_N$ = 'soles'*
pleuronecte$_N$	'pleuronecte'	*(Fish that) °necte$_V$ = 'swims' on one °pleuro$_N$ = 'side'*
psychogène$_A$	'psychogenic'	*which is °gène$_V$ = 'generated' by °psycho$_N$ = 'mind'*
russophile$_A$	'Russophil'	*who °phile$_V$ = 'likes' °russo $_N$ = 'Russians'*
trachéotome$_N$	'tracheotome'	*(scalpel used to) °tome$_V$ = incise °trache(o)$_N$ = 'trachea'*
vasiducte$_N$	'carrying vessel'	*°vasi$_N$ = 'vessel' which °ducte$_V$ = 'carries'*
ventriloque$_A$	'ventriloquist'	*who °loque$_V$ = 'speaks' with his/her ventri$_N$= stomach*
wagnerolâtre$_A$	'who adores Wagner'	*who °lâtre$_V$ = 'adores' Wagner$_N$*
xylographe$_N$	'xylographer'	*(He who) °graphe$_V$ = 'writes' on °xylo$_N$ = 'wood'*

Analogy and Irregularity in Romance Verbal Morphology

Roberta Maschi
Padua University
rmaschi@libero.it

Abstract

In this paper I will focus on connections between analogy and irregularity; in particular, I aim at exploring the possibility of establishing to what degree analogical processes, which are usually considered arbitrary in traditional linguistic literature, can be predicted. I am mainly going to analyse a special kind of analogy, that is the rise and spreading of a group of irregular verbs as a *class* able to attract new members.
As is well known, Romance verbal morphology is characterized by great irregularity. A diachronic perspective reveals that irregularity is not only the result of regular processes of a purely phonological nature: a significative part of irregular verbal forms are the result of analogical processes. It seems obvious that phonological changes produce some results that are "unnatural" as regarding the verbal system, because they are blind to the paradigmatic configuration; onthe contrary, it is a surprising fact that analogy can produce irregularity, since its motivation is intrinsically morphological. Since analogy can originate irregularity, then shall we conclude that morphological change is completely arbitrary?

1. Some terminology

Before entering the main issue, it is necessary to give some preliminary explanation as regards some terminology I am using. The expression "Basic Stem" (hereafter BS, S in the tables), from Aronoff 1994 and Pirrelli&Battista 2000a, will be employed to refer to the basis for the formation of the stem or stems in a verbal paradigm[1]; then a BS is *synchronically* unpredictable on the basis of another stem in the paradigm. The BS is unique in the completely regular verbs, while the BS number increases according to irregularity in the considered paradigm, as we can see in tables 1 and 2. In these tables we can observe the present of the indicative and the subjunctive of three verbs presenting different degrees of irregularity, in Castilian and Galician: the first one, *andar*, is wholly regular, with only one BS both in the indicative and in the subjunctive; on the other hand, *conocer* and *ver* show two BSs, *venir* and *facer* three. (Note that in Galician the grapheme <x> corresponds to the fricative prepalatal voiceless phoneme /ʃ/).

[1] "Hereafter, we then will use the Aronovian notion of basic stem to refer to basic stem roots rather than stems proper" (Pirrelli&Battista 2000a: 316).

G. Booij, *et al.* (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)*
Fréjus 15-18 September 2005, University of Bologna, 2007.
URL http://mmm.lingue.unibo.it/

PRESENT INDICATIVE and SUBJUNCTIVE					
ANDAR ('to walk')		CONOCER ('to know')		VENIR ('to come')	
S_1 *And-o*	S_1 *And-e*	S_2 *Conozc-o*	S_2 *Conozc-a*	S_2 *Veng-o*	S_2 *Veng-a*
S_1 *And-as*	S_1 *And-es*	S_1 *Conoc-es*	S_2 *Conozc-as*	S_3 *Vien-es*	S_2 *Veng-as*
S_1 *And-a*	S_1 *And-e*	S_1 *Conoc-e*	S_2 *Conozc-a*	S_3 *Vien-e*	S_2 *Veng-a*
S_1 *And-amos*	S_1 *And-emos*	S_1 *Conoc-emos*	S_2 *Conozc-amos*	S_1 *Ven-imos*	S_2 *Veng-amos*
S_1 *And-áis*	S_1 *And-éis*	S_1 *Conoc-éis*	S_2 *Conozc-áis*	S_1 *Ven-ís*	S_2 *Veng-áis*
S_1 *And-an*	S_1 *And-en*	S_1 *Conoc-en*	S_2 *Conozc-an*	S_3 *Vien-en*	S_2 *Veng-an*

Table 1: Castilian

PRESENT INDICATIVE and SUBJUNCTIVE					
ANDAR ('to walk')		VER ('to see')		FACER ('to do')	
S_1 *And-o*	S_1 *And-e*	S_2 *Vex-o*	S_2 *Vex-a*	S_2 *Fag-o*	S_2 *Fag -a*
S_1 *And-as*	S_1 *And-es*	S_1 *V-es*	S_2 *Vex-as*	S_3 *F-as*	S_2 *Fag -as*
S_1 *And-a*	S_1 *And-e*	S_1 *V-e*	S_2 *Vex-a*	S_3 *F-ai*	S_2 *Fag -a*
S_1 *And-amos*	S_1 *And-emos*	S_1 *V-emos*	S_2 *Vex-amos*	S_1 *Fac-emos*	S_2 *Fag -amos*
S_1 *And-ades*	S_1 *And-edes*	S_1 *V-edes*	S_2 *Vex-ades*	S_1 *Fac-edes*	S_2 *Fag -ades*
S_1 *And-an*	S_1 *And-en*	S_1 *V-en*	S_2 *Vex-an*	S_3 *F-an*	S_2 *Fag -an*

Table 2: Galician

Index 1 is assigned to the unmarked BS, that is the most widely distributed one throughout the paradigm.

Another important definition coming from Pirrelli&Battista 2000a is that of "partition class". The partition class is the set of verbal forms sharing the same BS in one paradigm. For example, as we can see in tables 1-2, the partition class of the BSs *conozc-* and *veng-* (in Castilian), or *vex-* and *fag-* (in Galician) consists of the 1st person of present indicative and all the persons in the subjunctive. We underline that this is a set of forms which are completely heterogeneous from a semantic and a morphosyntactic point of view, and that synchronically this partition has no other justification than the morphological one: that's what Aronoff 1994 calls "morphome".

Maiden's diachronic analysis shows very clearly that even though the origin of these morphomic structures, or partition classes, is phonologically motivated, their survival and persistence within the Romance verbal system go beyond the etymological reasons: after their phonological rise, they began living their own life and spreading to verbs lacking the basic phonological conditions.

Thus, analogical change gives clear proof of the existence of the morphome and of its morphological autonomy, following the patterns shaped by the partition classes. But now we'll see that, observing different types of analogical processes. Maiden classifies them under the label of coherence, convergence and attraction. We will accept and develop his classification.

2. Typologies of analogical processes

Today we're going to examine only analogical processes regarding stems, and we will leave out from our analysis the processes on inflectional endings. According to the effects and directions of analogical changes on stems, we can classify them as it follows:

- intraparadigmatic regularization, which includes
    a. partial levelling and
    b. morphomic coherence;

- interparadigmatic adaptation, or convergence;

- morphomic productivity, split up in
    c. attraction and
    d. enlargement.

In this paper we're going to deal mainly with the third type, the morphomic productivity, which is more meaningful to our analysis.

2.1. Intraparadigmatic regularization

It can be described as a tendency towards the regularization of a paradigm, by decreasing radical allomorphies or suppletion inherited from Latin. It has two mechanisms of expression: partial levelling and morphomic coherence.

2.1.1. Partial levelling

The partial levelling shows the unity between the forms of the same partial paradigm (for example, the present indicative): the substitution of a BS with another occurs within a partial paradigm, but not within the whole verbal paradigm. Some examples are given in (1), (2) and (3).
Some Romance languages remove or decrease the radical suppletivism in the inflection. Let us considere, for instance, the present indicative of 'to go':

(1) old Castilian:
present indicative*: voy, vas, va, imos, ides, van;*
subjunctive *vaya, vayas, vaya, vayamos / vamos, vayades / vades, vayan*

modern Castilian:
present indicative: *voy, vas, va, vamos, vais, van;*
subjunctive *vaya, vayas, vaya, vayamos, vayáis, vayan*

(2) Catalan:
present indicative: *vaig, vas, va, anem, aneu, van,*
(but) perfective periphrasi: *vaig, vas, va, vam, vau, van*;

(3) some Lombard dialects (cfr. AIS VIII, 1692).:
a dialect spoken near Milan: *vo, ve, va, vem, ve, van*;
Como *vu, vet, va, vem, vi, van*

2.1.2. Morphomic coherence

The regularization can operate in a subtler way, when it uniforms from a phonological point of view the forms of a partition class. This is the analogical type called coherence by Maiden. It shows that the mutual implication between the forms of the same partition class keeps on being intact (e.g. 1st person pres. ind. <=> pres. cong.), in spite of their phonological and functional difference. Practically this means that a phonological change affecting one form could equally affect all the other forms of the same partition, regardless of the different phonological conditions.

A clear example for this analogical type is found in the Ibero-Romance perfective basic stems, characterised now by a high vowel [i] or [u] in all the partition class (preterite, imperfect subjunctive and future subjunctive); but the high vowel is phonologically regular only in the 1st person of the preterite, as shown in the medieval phases. This case has been discussed by Maiden 2001; we add here some data from Galician in Table 3.

LATIN PERFECTIVE STEM	OLD GALICIAN PERFECTIVE STEM		MODERN GALICIAN UNIQUE PERFECTIVE STEM
	1p preterite (metaphonic)	other forms of the partition	
POSU-	pus	pos-	pux- (‘put’)
POTU-	pude	pod-	puid- (‘could’)
FEC-	fiz	fez-	fix- (‘did’)
SEDU-	sive	sev-	(fu-) (‘was’)
CREDU-	crive	crev-	(reg. cre-) (‘believed’)
(TENU-)	tive	tev-	tiv- (‘had’)
(STET-)[2]	estive	estev-	estiv- (‘stood’)

Table 3: morphomic coherence of the Galician perfective basic stems

2.2. Interparadigmatic adaptation

So far, we have quickly seen an analogical process with a phonologically motivated origin for every verb, at least in one form of the paradigm; but sometimes a phonological feature analogically extended acquires such a relevance that it becomes a sort of mark, and then it is taken up by paradigms completely lacking the basic phonological conditions. It is the second analogical type, the interparadigmatic adaptation, or convergence (Maiden).

[2] *Tive* and *estive* are analogical to *sive* (the regular form *estede* is rarely attested); let us notice, in the paradigm of *ser*, ‘to be’, the suppletive forms from SEDERE in medieval Galician, from which not only the old preterite derives, but the modern present subjunctive as well: SEDEA > *seja* > *sexa*.

Table 4 presents some strong preterites in Castilian: the verbs of group 3, anduve, cupe, hube, etc., present in the actual perfective stem a high vowel, which is neither etymological -as in group 1- nor the outcome of coherence -as in group 2-. This high vowel is, indeed, an effect of the adaptation to the general feature of strong perfective stems, though these verbs lack the phonological conditions for the vocalic rising. Even for the preterite of traer, traje, a stem truj- is locally attested (truje(n)..., trujera..., trujese...) e.g. in León (Paradaseca do Bierzo, ALGa maps 386, 387, 391).

LATIN PEFECTIVE STEMS	OLD CAST. PERFECTIVE STEMS	MODERN CAST. PERFECTIVE STEMS	VERBAL LEXEME
GROUP 1	S WITH AN ETYMOLOGICAL HIGH VOWEL		
CINX-	Cinx-	(Reg. > ceñí)	Ceñir, 'to encircle'
DUX-	(Con)duj-	(con)duj-	Conducir, 'conduct'
DIX-	Dij-	dij-	Decir, 'to say'
SCRIPS-	Escris-	(Reg. > escribí)	Escribir, 'to write'
FU-	Fu- (/Sov-)	fu-	Ser, 'to be'
MIS-	Mis-	(Reg. > metí)	Meter, 'to put'
VID-	Vid-	> vi-	Ver, 'to see'
VIX-	Visqu-	(Reg. > viví)	Vivir, 'to live'
GROUP 2	S WITH A METAPHONETIC HIGH VOWEL IN 1p (EXTENDED FOR COHERENCE)		
COGNOU->*CONOV-	Conuv- / conov-	(Reg. > conocí)	Conocer,'to know'
STET-	Estid- / ested-	(> estuv-)	Estar, 'to stay'
FEC-	Hiz- / hez-	> hiz-	Hacer, 'to do'
POT(U)-	Pud- / pod-	> pud-	Poder, 'can'
POS(U)-	Pus- / pos-	pus-	Poner, 'to put'
QUAES(IU)-	Quis- /ques-	quis-	Querer, 'to want'
VEN-	Vin- /ven-	vin-	Venir, 'to come'
GROUP 3	S WITH A COMPLETELY ANALOGICAL HIGH VOWEL		
(*AND-)	Andov-	anduv-	Andar, 'to walk'
*CAPUI- > *CAUP-	Cop-	> cup-	Caber, 'to go into'
(CREDID-)	Crov-	(Reg. > creí)	Creer, 'to believe'
HABU- > *HAUB-	Ov-	> hub-	Haber, 'to have'
PLACUI- > *PLAUC-	Plog-	> plug-	Placer, 'to please'
SAPU- > *SAUP-	Sop-	> sup-	Saber, 'to know'
(TENU-)	Tov-	> tuv-	Tener, 'to have got'
IACUI- > *IAUC-	Yog-	Reg. > yací	Yacer, 'to lie'
TRAX-	Traj-	traj- (truj-)	Traer, 'to bring'

Table 4: perfective stems in Castilian

We find wide and clear evidence for convergence in the Romance languages, but I will limit myself here to show some data from Galician.

In table 5 we can see the present subjunctive of a group of very irregular Galician verbs. The regular phonological evolution made these subjunctive stems highly marked, and peculiar to the subjunctive alone. These paradigms converge from a morphological point of view, because they share this peculiar distribution of the subjunctive basic stem. In the column on the left there are the standard forms, which are

in large part etymological[3]; but the microvariation throughout the Galician territory reveals some interesting outcomes splitting off from the etymological ones: we can observe such variations in the four columns on the right (the data come from ALGa, maps 230 to 389). Each variation pattern shows a phonological sequence analogically acquired from one or more "leading" verbs, e.g: by analogy with *teña* and *veña* we find

- *feña* (instead of *faga*),
- *deña*, *esteña* (instead of *dea*, *estea*),
- and *seña* (instead of *sexa*),

and so on (the geographical distribution is given in the note 4).

The relevant issue is the following one: some paradigms which already share a morphological -that is distributive- convergence, after this process has taken place, they share a phonological convergence as well. The morphomic structure becomes then more strongly compact.

VERBAL LEXEME	Standard subj.	**"analogizing" S feature**			
		-eñ+a	-x-+a	-aiC+a	-dipht.+a
dicir 'to say' facer 'to do'	DIGA FAGA	feña		faiga	faia
ir 'to go' traer 'to bring' oír 'to hear'	VAIA TRAIA OIA		vaixa oixa		VAIA TRAIA OIA
dar 'to give' estar 'to stay'	DEA ESTEA	deña esteña	estexa		deia esteia
pór 'to put' ter 'to have got' vir 'to come'	POÑA TEÑA VEÑA	POÑA TEÑA VEÑA			
haber 'to have' ser 'to be' ver 'to see'	HAXA SEXA VEXA	seña	HAXA SEXA VEXA	haixa / haiba	haia seia
poder 'can' querer 'to want' saber 'to know'	POIDA QUEIRA SAIBA			POIDA QUEIRA SAIBA	

Table 5 : irregular subjunctives in Galician[4] (<x> = / ʃ /).

[3] Lat. DEM, STEM mediev. Galician *dé*, *esté* (modern Galician adds -a); FACJAM > med. Gal. *faça*; *faga*, attested from the XVth century, is probably originated by morphomic coherence with the 1st person *fago* (analogical to *digo*); HABEAM > *haja* > *haxa*; VIDEAM > *vexa*; SEDEAM > *seja* > *sexa*, but also *sea* (se+a) is attested from the XIIIth century, and nowadays emploied in an area of Galicia; VADAM > *vaa* > *vaja* (-j- anti-hyatus); POSSAM > med. Gal. *possa*; *poida* is analogical to *queira* and *saiba* (< *QUAERIAM, SAPIAM); PONEAM, TENEAM, VENJAM > *poña, teña, veña* (cfr. Ferreiro 1999: §§ 204-223).

[4] Geographical distribution of the subjunctives in table 5:
- *feña, deña, esteña, seña*: Asturias and northern of Lugo and A Coruña provinces (*feña...* is attested only at Calvario de Salave, Asturias). S*eña* is the most widely spred; - *vaixa, oixa, estexa*: provinces of Lugo (Pobra do Brollón), A Coruña (Mesoiro), Pontevedra (Fefiñáns), but *estexa* is also a literary form; - *faiga, haiba, haixa*: *haixa* in some localities of the provinces of Pontevedra, Lugo and A Coruña (Mesoiro); *haiba* in the northern part of A Coruña; *faiga* well attested in the provinces of Lugo, A Coruña and in Asturias; - *faia, deia, esteia, haia, seia*: mainly in the province of Pontevedra.

2.3. Morphomic productivity

There's a kind of analogy in which we can observe the extension and reproduction not only of some phonological features of the morphome (as it happens in convergence), but also of its structure, which is acquired by paradigms that didn't share it before. For instance, in the Galician dialect of Asturias we find, in the paradigm of ser 'to be', in addition to the analogical subjunctive seña (see table 5), the 1st person of present indicative seño instead of son.

This is a meaningful fact, for the following reason: ser, 'to be', in Galician as in the other Romance languages, is so irregular that it escapes from any distributive generalization and has its own partition. The partition of ser lacks that correlation between the subjunctive and the 1st person, which is so systematic in other verbs. What happened? In addition to the phonological sequence -eñ- characterising the subjunctive (*teña*, *veña* > analogical *seña* replacing *sexa*), from the verbs ter, vir what spreads in this case is the partition class, that in ter and vir includes, along with the subjunctive, the 1st person of present indicative (*teño*, *veño* > *seño*). This is a case of morphomic productivity, called "attraction" by Maiden. We'll distinguish the cases of attraction stricto sensu (in § 2.3.1) from some phenomena that we define enlargement of the basic stem (in § 2.3.2): they are two processes sharing the morphomic origin, but with very different modalities and issues.

2.3.1. Attraction

In the process of attraction, as we have just seen, the morphome spreads both some phonological features and its own distributional pattern. One more example: in Galician, groups of irregular verbs of 2nd and 3rd conjugationshow in their basic stem a vocalic opening alternation which is not always etymological.

Galician verbs with a mid vowel in their unmarked basic stem show an alternation between a closed vowel and an open vowel; this variation is not phonological in all the cases. The observation of this phenomenon is relevant for the present indicative and subjunctive and for the imperative, where we find an alternation between stressed and unstressed stems (elsewhere the unmarked stem is always unstressed, and the mid vowel is automatically closed, unless we have the case of analogical levelling on the basis of the open vowel).

In the 1st conjugation the alternation, if it is applied, is automatic, since it depends on stress position: mid vowels are open if stressed, closed if unstressed, as shown in (4) and (5):

(4) *levar* 'carry':
pres. ind. l[ɛ]v-o, l[ɛ]v-as, l[ɛ]v-a, l[e]v-a̲mos, l[e]v-a̲des, l[ɛ]v-an
Pres. subj. l[ɛ]v-e, l[è]v-es, l[ɛ]v-e, l[e]v-e̲mos, l[e]v-e̲des, l[ɛ]v-en

(5) *rogar* 'pray':
pres.ind. r[ɔ]g-o, r[ɔ]g-as, r[ɔ]g-a, r[o]g-a̲mos, r[o]g-a̲des, r[ɔ]g-an
Pres.subj. r[ɔ]gu-e, r[ɔ]gu-es, r[ɔ]gu-e, r[o]gu-e̲mos, r[o]gu-e̲des, r[ɔ]gu-en

But in the 2nd conjugation, except for some paradigms levelling their vocalic opening (a coherence effect), most verbs with a mid vowel in the unmarked stem present an opening variation even within root-stressed forms: let's look at the tables 6 and 7 (4th and 5th persons are colourless, having unstressed root vowels, naturally closed):

FERVERE > *ferver*, 'to boil'			
PRESENT INDICATIVE		PRESENT SUBJUNCTIVE	
FERV-EO > *ferv-o*	[e]	FERV-EAM > *ferv-a*	[e]
FERV-ES > *ferv-es*	[ɛ]	FERV-EAS > *ferv-as*	[e]
FERV-ET > *ferv-e*	[ɛ]	FERV-EAT > *ferv-a*	[e]
(FERV-EMUS > *ferv-emos*)		(FERV-EAMUS > *ferv-amos*)	
(FERV-ETIS > *ferv-edes*)		(FERV-EATIS > *ferv-ades*)	
FERV-ENT > *ferv-en*	[ɛ]	FERV-EANT > *ferv-an*	[e]

Table 6

MOVERE > *mover*, 'to move'			
PRESENT INDICATIVE		PRESENT SUBJUNCTIVE	
MOV-EO > *mov-o*	[o]	MOV-EAM > *mov-a*	[o]
MOV-ES > *mov-es*	[ɔ]	MOV-EAS > *mov-as*	[o]
MOV-ET > *mov-e*	[ɔ]	MOV-EAT > *mov-a*	[o]
(MOV-EMUS > *mov-emos*)		(MOV-EAMUS > *mov-amos*)	
(MOV-ETIS > *mov-edes*)		(MOV-EATIS > *mov-ades*)	
MOV-ENT > mov-en	[ɔ]	MOV-EANT > *mov-an*	[o]

Table 7

We can notice that some mid vowels are closed despite bearing stress, in the 1st person of present indicative and 1st, 2nd, 3rd and 6th of subjunctive; the origin of the vocalic rising is metaphony on Latin short E and O followed by a glide in the subsequent syllable. But the relevant fact is that we can find the same alternation pattern in many paradigms lacking the phonological condition (Latin short E or O followed by a glide); in (6) and (7) we have some examples. All these verbs share the alternation pattern as seen in tables 6 and 7, but their Latin vowel didn't undergoes metaphony, or there was no glide, or both of them:

(6) verbs whose Latin mid vowel could not undergo metaphony:

- mexer 'to swing': pres. ind. MISCEO... > m[e]x-o, m[ɛ]x-es, m[ɛ]x-e, (mex-emos, mex-edes), m[ɛ]x-en; pres. subj. MISCEAM... > m[e]x-a, m[e]x-as, m[e]x-a, (mex-amos, mex-ades), m[e]x-an
- temer 'to fear': pres. ind. TIMEO... > t[e]m-o, t[ɛ]m-es, t[ɛ]m-e, (tem-emos, tem-edes), t[ɛ]m-en; pres. subj. TIMEAM... > t[e]m-a, t[e]m-as, t[e]m-a, (tem-amos, tem-ades), t[e]m-an

- poñer 'to put': pres. ind. *PONEO... > p[o]ñ-o, p[ɔ]-s, p[ɔ]-n, (poñ-emos, poñ-edes), p[ɔ]ñ-en; pres. subj. *PONEAM... > p[o]ñ-a, p[o]ñ-as, p[o]ñ-a, (poñ-amos, poñ-ades), p[o]ñ-an
- beber 'to drink': pres. ind. BIBO... > b[e]b-o, b[ɛ]b-es, b[ɛ]b-e, (beb-emos, beb-edes), b[ɛ]b-en; pres. subj. BIBAM... > b[e]b-a, b[e]b-as, b[e]b-a, (beb-amos, beb-ades), b[e]b-an
- vender 'to sell': pres. ind. VENDO... > v[e]nd-o, v[ɛ]nd-es, v[ɛ]nd-e, (vend-emos, vend-edes), v[ɛ]nd-en; pres. subj. VENDAM... > v[e]nd-a, v[e]nd-as, v[e]nd-a, (vend-amos, vend-ades), v[e]nd-an
- correr 'to run': pres. ind. CORRO... > c[o]rr-o, c[ɔ]rr-es, c[ɔ]rr-e, (corr-emos, corr-edes), c[ɔ]rr-en; pres. subj. CORRAM... > c[o]rr-a, c[o]rr-as, c[o]rr-a, (corr-amos, corr-ades), c[o]rr-an

(7) verbs without a glide in the final syllable:

- perder 'to loose': pres. ind. PERDO... > p[e]rd-o, p[ɛ]rd-es, p[ɛ]rd-e, (perd-emos, perd-edes), p[ɛ]rd-en; pres. subj. PERDAM... > p[e]rd-a, p[e]rd-as, p[e]rd-a, (perd-amos, perd-ades), p[e]rd-an
- volver 'to come back': pres. ind. VOLVO... > v[o]lv-o, v[ɔ]lv-es, v[ɔ]lv-e, (volv-emos, volv-edes), v[ɔ]lv-en; pres. subj. VOLVAM... > v[o]lv-a, v[o]lv-as, v[o]lv-a, (volv-amos, volv-ades), v[o]lv-an
- coller 'to catch': pres. ind. COLLIGO... > c[o]ll-o, c[ɔ]ll-es, c[ɔ]ll-e, (coll-emos, coll-edes), c[ɔ]ll-en; pres. subj. COLLIGAM... > c[o]ll-a, c[o]ll-as, c[o]ll-a, (coll-amos, coll-ades), c[o]ll-an

Synchronically we can't explain this alternation, and probably it couldn't exist, if a morphomic structure (as we see in table 8) had not become a systematic distributional pattern, during the language evolution (2nd-3rd -6th ind. / 1st -4th -5th ind. + subj.). Its origin is phonologically motivated, but the partition classes have been fixed beyond the diachronic reason, and have been reinforced by attraction of new members.

Distributional pattern	
PRESENT INDICATIVE	PRESENT SUBJUNCTIVE
1	1
2	2
3	3
4	4
5	5
6	6

Table 8

We will not analyse now the more complex alternations in the 3rd conjugation of Galician. We only point out that in such a complex interaction of different alternations only a few ones are etymologically motivated,while the others have analogically

assumed a shape like that shown in table 8, with 2 basic stems, or like the one in table 9, with 3 BSs:

Distributional pattern	
PRESENT INDICATIVE	PRESENT SUBJUNCTIVE
1	1
2	2
3	3
4	4
5	5
6	6

Table 9

2.3.2. Enlargement of the unmarked basic stem

The analogical process defined "morphomic productivity" could operate differently from attraction, when the unmarked basic stem S1 in a verb undergoes an enlargement process. The material used for enlargement comes from marked basic stems Sx (x = index higher than 1) of other verbs (generally from light verbs), so we can classify this process an expression of morphomic productivity. An example is given in (8):

Italian regular preterite of 2nd conjugation presents a morpheme *-ett-*, together with or in place of the inflectional endings of persons 1st, 3rd, 6th:

(8) preterite of *temere* ('to fear'):

1 *tem-ei / tem-ett-i*	**2** *tem-esti*	**3** *tem-é / tem-ett-e*	**4** *tem-emmo*	**5** *tem-este*	**6** *tem-erono / tem-ett-ero*

The origin of -ett- is found in the strong preterite of *stare* 'to stay', in table 10:

preterite of *stare*
1p STETUI > stetti
2p STETUISTI > stesti
3p STETUIT > stette
4p STETUIMUS > stemmo
5p STETUISTIS > steste
6p STETUERUNT > stettero

Table 10

In many Central and Southern Italian dialects, this morpheme is more widespread, going beyond the 2nd conj., first of all extending to the 3rd, and in some dialects even to the 1st, though remainig generally limited to persons 1st, 3rd and 6th. There are some examples in (9), (10) and (11):

(9) Spoleto (Umbria): *disetti-disette-disettero* ('to say'); *fasetti-fasette-fasettero* ('to do'); Foligno *etti-ette-ettero* ('to have'); 2nd -3rd conj. Moretti 1987.

(10) Western Abruzzo and Molise: *kandette,* 'he sang' (1st conj.), *vedette,* 'he saw' (2nd), *dormette*, 'he slept' (3rd) (Giammarco 1979); all conjugations.

(11) Neapolitan: *pərdettə, fənettə, sapettə, facettə* (2nd-3rd conj.; 1st with -a- *candattə*). Rohlfs 1968: §§ 577-578.

We underline that the original morphomic structure keeps intact in its distribution, but at the same time its markedness disappears: in these cases the enlargement process has a "stabilizing" effect: from a marked basic stem of a light verb *a morphomic sequence is isolated and reanalysed as morphemic*, and then extended. Even the strong preterite of *stare*, *stette*, in light of this process, is synchronically reanalyzed as a regular preterite built on the unmarked basic stem S_1 st- + ett-, instead of marked S_X stett-. In (12) and (13) we have two schemes showing this change of the status of -ett- in the speaker's analysis:

(12) from morphomic sequence (where -ett- is a part of strong perfective BS in 1st, 3rd and 6th persons):

STETT+i = $\mathbf{S_x}$+-i (e.g. *pers*+i 'I lost', *fec*+i 'I did', *voll*+i 'I wanted'...)

strong perfect	
1p **stett**-i	Sx
2p st-esti	S1
3p **stett**-e	Sx
4p st-emmo	S1
5p st-este	S1
6p **stett**-ero	Sx

(13) to enlargement (unmarked BS + -ett- as a part of a "complex inflectional ending"):

ST+**étti** = $\mathbf{S_1}$ + -étt-i (e.g. *perd*+*étti* 'I lost')

regular perfect	
1p st-**ett**i	S1
2p st-esti	S1
3p st-**ett**e	S1
4p st-emmo	S1
5p st-este	S1
6p st-**ett**ero	S1

For "stabilizing effect" we mean two possible situations:

1. a strong preterite becoming weak (by the change Sx > S1): *pers-i > perd etti*;

2. a weak preterite, e.g. *perd-ei*, acquiring (or being replaced by) a doublet (allotropic variant), that is a second weak preterite with -ett-, phonologically more "stable" having all the inflectional endings bisyllabic.

strong preterite		etymological weak preterite		analogical weak preterite	
1p **pers**-i	Sx	1p perd-ei	S1	1p perd-**ett**i	S1
2p perd-esti	S1	2p perd-esti	S1	2p perd-esti	S1
3p **pers**-e	Sx	3p perd-é	S1	3p perd-**ett**e	S1
4p perd-emmo	S1	4p perd-emmo	S1	4p perd-emmo	S1
5p perd-este	S1	5p perd-este	S1	5p perd-este	S1
6p **pers**-ero	Sx	6p perd-erono	S1	6p perd-**ett**ero	S1

Table 11

But we can find a morphomic sequence extended to the whole partial paradigm (preterite) in all conjugations; it's a sort of ideal "completion" of the enlargementprocess, and we can call it thematization, or morphologization.

2.3.2.1. Morphologization

That's the last analogical subtype. The phonological sequence extracted from the basic stem of a verb is morphologized, that is it acquires some functional features, and becomes a suffix for the formation of derivated stems (Aronoff's *constant thematic functions*), or a distinctive mark for a partial paradigm.
Such a process involves the speaker's reanalysis of the original *morphomic* sequence (-ett-, in our example), once again as a *morphemic* one, but this time without keeping its original distribution.

It is schematically shown in (14); in table 12, we have a "virtual" representation of the reanalysis of -ett- as a suffix.

(14) morphologization: ST+**ett**+i = S_1 + **mood-tense suffix** + inflectional ending -i

regular perfect	
1p st-ett-i	S1
2p *st-ett-esti	S1
3p st-ett-e	S1
4p *st-ett-emmo	S1
5p *st-ett-este	S1
6p st-ett-ero	S1

Table 12

We find a concrete example for the generalized extension of -ett- in Engadinese, a Rhaeto-Romance dialect, where the morphemes *-et(t-)* (1st, 2nd, 3rd conj.) and *-it(t-)* (4th conj.) characterise the whole preterite in all the verbs (Stimm & Linder in LRL III; Haiman & Benincà 1992: 89-90).

Below, the preterite of *portar* 'to carry' in two different Engadinese dialects:

putér dialect:

1 *purt-et-Ø*	**2** *purt-ett-ast*	**3** *purt-et-Ø*	**4** *purt-ett-ans*	**5** *purt-ett-as*	**6** *purt-ett-an*

vallader dialect:

1 *port-et-Ø*	**2** *port-ett-ast*	**3** *port-et-Ø*	**4** *port-ett-an*	**5** *port-ett-at*	**6** *port-ett-an*

One more example: in medieval Occitan the regular preterite (e.g. *cantèi, cantèst, cantèt, cantèm, cantètz, cantèron*) originated with the influence of *estèi* < STETI and *dèi* < DEDI (Lafont in LRL V, 2). But as it happened in Engadinese, in some dialects -et- was extended to other persons, as in Auvergne:

(15) 4p *cantetem,*
5p *cantetetz*.

Differently from the cases of enlargement previously analysed, in the latter we notice the disappearance ofthe morphome even from a distributional point of view. That's an extreme effect of morphomic productivity: that is building derivated stems from basic stems.

To summarize: morphomic productivity operates in large part by reproducing morphomic structures (attraction), thus making the verbal system more marked. On the other side, it reproduces phonological sequences giving an opposite effect of "stabilization" of the verbal system (enlargement). In this case the morphomic sequence can mantain its distribution, or spread throughout the verbal system: thus it reaches the maximum degree of productivity, but at the same time the *morphome is paradoxically cancelled*, since its distribution is cancelled.

3. Conclusions

At the beginning of this paper we have formulated the question whether morphological change is completely arbitrary, considering the relation between analogy and irregularity. I hope to have shown that, even though analogical change spreads irregularity, at the same time it originates or reinforces some distributive regularities, balancing the arbitrariness of stem formation, and sometimes also some idiosyncratic phonological evolutions.

The morphomic structures that undergo analogical processes, as convergence or attraction, are reinforced so as to constitute, despite their markedness, subregularities within the verbal system, and *so as to provide some strategies in language learning and production*.

Also analogy then displays some “rules”, and these rules make analogy somehow predictable, that is to say we can risk predictions: once a protoype and the “analogizing” feature have been focused, we can formulate hypotheses both on the lexemes involved, and on the direction and extension of the change through one paradigm (by considering partition classes).

References

AIS = Jaberg, K. e J. Jud (1928-1940). *Sprach- und Sachatlas Italiens und Südschweiz*. Zofingen: Ringier.

ALGa = Instituto da Lingua Galega (1990). *Atlas Lingüístico Galego*, vol. I: *Morfoloxía verbal*. A Coruña: Fundación Pedro Barrié de la Maza, Conde de Fenosa.

Álvarez, R., X. Xove (2002). *Gramática da lingua galega*. Vigo: Editorial Galaxia.

Aronoff, M. (1994). *Morphology by itself*. Cambridge (Mass.): The MIT Press.

Bertocci, D., R. Maschi (2004). "Alcuni verbi irregolari nelle lingue romanze e indeuropee antiche: ipotesi sul costituirsi di una *classe* paradigmatica". *Quaderni Patavini di Linguistica*, 20, 3-42.

Bybee, J., C. Moder (1983). "Morphological Classes as Natural Categories". *Language*, 59: 251-270.

Elvira, J. (1998). *El cambio analógico*. Madrid: Gredos.

Fernández Rei, F. (1978-79). *O verbo. Contribución á dialectoloxía galega*, 2 voll. Doctoral Dissertation, University of Santiago de Compostela.

Fernández Rei, F. (1990). *Dialectoloxía da lingua galega*. Vigo: Xerais.

Ferreiro, M. (1999). *Gramática histórica galega*. Santiago de Compostela: Laiovento.

Haiman, J., P. Benincà (1992). *The Rhaeto-Romance languages*. London-New York: Routledge.

Kuriłowicz, J. (1949). "La nature des procès dits 'analogiques'". *Acta Linguistica* 5, 15-37.

LRL = Holtus, G., Metzelin, M., Schmitt, C. (Hrsgg.), (1988-). *Lexicon der romanistischen Linguistik*. Tübingen: Niemeyer.

Maiden, M. (1992). "Irregularity as a determinant of morphological change". *Journal of Linguistics* 28, 285-312.

Maiden, M. (1995). *A Linguistic History of Italian*. London: Longman.

Maiden, M. (2001). "A strange affinity: perfecto y tiempos afines". *Bulletin of Hispanic Studies* 58, 441-64.

Maiden, M. (2003). "Il verbo italoromanzo: verso una storia autenticamente morfologica". In M. Giacomo Marcellesi, A. Rocchetti (eds.), *Il verbo italiano*, Roma: Bulzoni, 3-21.

Maiden, M. (2003). "Morphology as 'disease'. Is the typological distinction healty?". G. Booij, A. Ralli, S. Scalise, S. Sgroi (eds.), *Proceedings of the IV Mediterranean Morphology Meeting*, University of Catania, september 2003.

Meyer-Lübke, W. (1895). *Grammaire del langues romanes*, vol. 2: *Morphologie*. Paris: H. Welter.

OVI = Opera del Vocabolario Italiano: www.csovi.fi.cnr.it

Penny, R. (1993). *Gramática histórica del español*. Barcelona: Ariel (translation by José Ignacio Pérez Pascual of *A history of the Spanish language*. Cambridge: Cambridge University Press).

Pirrelli, V., M. Battista (2000a). "The paradigmatic dimension of stem allomorphy in Italian verb inflection". *Rivista di Linguistica*, 12-2, 307-379.

Pirrelli, V., M. Battista (2000b). "On the interaction of paradigmatic and syntagmatic stem alternation in Italian conjugation". *Acta Linguistica Hungarica*, 47, 289-314.

Pirrelli, V. (2000). *Paradigmi in morfologia. Un approccio interdisciplinare alla flessione verbale dell'italiano*. Pisa-Roma: Istituti Editoriali e Poligrafici Internazionali.

Real Academia Galega, Instituto da Lingua Galega (2004). *Normas ortográficas e morfolóxicas do idioma galego*. Vigo: Artes Gráficas Vicus.

Rohlfs, G. (1968). *Grammatica storica della lingua italiana e dei suoi dialetti*, vol.II: *Morfologia*. Torino: Einaudi.

Salvi, G., L. Vanelli (2004). *Nuova grammatica italiana*. Bologna: Il Mulino.

Santamarina, A. (1974). "El verbo gallego". *Verba*, anexo 4, Santiago de Compostela.

Constructional Licensing in Morphology and Syntax*

Jenny Audring
Faculteit der Letteren, Vrije Universiteit Amsterdam
j.audring@let.vu.nl

Geert Booij
Faculteit der Letteren, Universiteit Leiden
g.e.booij@let.leidenuniv.nl

1. Introduction

In recent work by the second author it has been argued that the theoretical insights of Construction Grammar can be applied in fruitful ways to the domain of morphology, thus leading to the idea of Construction Morphology (Booij 2005a, b). The basic idea of Construction Grammar may be specified as follows:

> "[…], the grammar represents an inventory of form-meaning-function complexes, in which words are distinguished from grammatical constructions only with regard to their internal complexity. The inventory of constructions is not unstructured; it is more like a map than a shopping list. Elements in this inventory are related through inheritance hierarchies, containing more or less general patterns."
> (Michaelis and Lambrecht 1996: 216)

Similar ideas have been put forward in Goldberg (1995, 2003).

This quotation leaves open to what extent words exhibit internal complexity. In this paper, we want to defend a unitary view of complex words (of the concatenative morphology type) and phrases. Like syntactic constructions, word formation patterns can be qualified as constructions, which may have fixed slots and variables as do constructions in sentence grammar (Booij 2005a, b). For instance, de-verbal noun formation in English by means of the suffix *-er* can be represented as a constructional idiom of the form $[[x]_V$ *er*$]_N$ 'one who Vs'. A word formation pattern with a particular affix can be conceived of as a morphological construction in which it is only the affix that is specified whereas the slot for the stem is variable. That is, each affixation pattern is a *constructional idiom* (in the sense of Jackendoff 2002), a construction in which one or more slots (but not all of them) are lexically fixed.

There is another reason why the notion 'construction' plays an important role in morphology: the use of morphological processes may be restricted to certain morphological or syntactic constructions. The implication of this form of interface between morphology and syntax is that we need a similar representational format for morphological and syntactic constructions for expressing such dependencies. A number of cases of this kind of dependency in Dutch can be found in Booij (2005a).

One of these examples of interaction between morphology and syntax in Dutch is the use of the suffix *-s* in the specifier position of noun phrases. A summary of the relevant facts can be found in Booij (2002: 34-35). Dutch nouns do not exhibit morphological case marking; this system disappeared in the transition from Middle

* We would like to thank Corrien Blom for her constructive comments on earlier drafts of this paper.

Dutch to present-day Dutch. There are, however, relics of the case system; one of them is that the suffix *-s* (historically the genitive singular case marker) can be used for a number of nouns in the specifier position of a noun phrase:

(1)	*Jan-s hoed*	'John's hat'
	Amsterdam-s rijke verleden	'Amsterdam's rich history'
	vader-s fiets	'father's bicycle'
	dominee-s studeerkamer	'reverend's study'
	ieder-s huis	'everybody's house'
	iemand-s vriend	'someone's friend'
	niemand-s schuld	'nobody's fault'

These words ending in the suffix *-s* have the function of possessor. The only nouns that can be used with this kind of possessor marker are proper names, nouns that can be used as forms of address, like *vader* father', *moeder* 'mother' and *dominee* 'reverend', that is, words functioning as proper names, and quantifying personal pronouns such as *iemand* 'someone'. Since a noun like *directeur* 'director' cannot be used as a form of address in Dutch, unlike a noun such as *dominee*, the phrase **directeurs kamer* 'the director's room' is ill formed. These *s*-marked nouns cannot be preceded by an article if they are marked as a possessor by means of *-s*. A phrase like **de dominees fiets* 'the minister's bicycle' is therefore ill-formed, unlike its English gloss. The words with *-s* in (1) can only be used in pre-nominal position: a sentence like **Deze hoed is Jans* 'This hat is John's' is ungrammatical which also shows that *-s* does not function as a genitive marker. In short, this use of words ending in this suffix -s is subject to strong syntactic restrictions. This kind of grammatical pattern is therefore best qualified as a specific construction with two sub-schemas for the two types of nouns that can be used: proper names (including names of address) and quantifying personal pronouns:

(2) a. [proper name -*s*]$_{\text{Spec-NP}}$

b. [quantifying personal pronoun-*s*]$_{\text{Spec-NP}}$

It is a constructional idiom that is productive to the extent that the slot for proper names is an open one, into which all proper names can be inserted.[1] [1]

A second example of this dependency of morphological processes on syntactic configurations, also taken from Booij (2005a), is the pluralization of Dutch numerals. The use of the plural forms of most numerals is restricted to a number of specific constructions, which are exemplified in (3):

(3) a. Number of parts:
Het schip brak in drie-en
The ship broke in three-*en*
'The ship broke into three pieces'

[1] The suffix *-s* also occurs in phrases such as *jouw moeder-s kamer* 'your mother's room', in which the specifier contains a possessive pronoun as well. Hence, the relevant constructional idiom should be modified as to also include the possibility of such a pronoun, that is, it should be represented as [(possessive pronoun) + proper name -*s*]$_{\text{Spec-NP}}$.

b. Appositive collective:
wij / ons drie-en
we /us three-*en*
'the three of us (subj. / obj.)'

c. Collective adverbial:
met ons / jullie / hun drie-en
with us / you / their three-*en*
'the three of us /you / them together'

d. Collective adverbial:
met z'n drie-en
with his three-*en*
'the three of us / you / them'

Example (3d) is a prototypical case of a constructional idiom. It has the form of a PP, headed by the preposition *met*, followed by the NP [*z'n* Numeral-*en*]. The possessive pronoun has the weak form *z'n* [zən]. In this NP the slot for the possessive pronoun is fixed as *z'n* (the 3rd pers. sg. possessive pronoun), whereas the slot for the numeral is open and can be filled with all sorts of numeral. Thus we have Dutch sentences like

(4) *We komen morgen met zijn twintig-en*
We come tomorrow with his twenty-*en*
'We will come tomorrow with twenty persons'

Note the incongruence between the person and number of the subject (1st pers. pl) and that of the possessive pronoun (3rd pers. sg.). The examples in (3c) are variants in which there is agreement in person and number between the subject of the sentence and the possessive pronoun in the collective construction. So there are two different collective constructions that are identical except that the possessive pronoun can either be a variable (and thus subject to the normal agreement constraints for possessive pronouns), or a fixed possessive pronoun *z'n*.

In addition to ordinal numerals the plural quantifiers *all-en* 'all' and *beid-en* 'both' can also be used in the constructions (3b-d). In these cases, the stem of the plural form does not occur as a word by itself.

We should note that these plural numerals cannot be used as subjects (with the exception of the noun-like numerals mentioned above). Thus, a sentence like the following is ungrammatical, although there is no clear semantic explanation for this ungrammaticality:

(5) **Drie-en gingen naar huis*
Three-*en* went to home
'Three people went home'

This illustrates once more how this productive use of pluralized numerals is restricted to very specific syntactic contexts, in other words, to constructions.

In this paper, we will present some more evidence from Dutch on the central role of the notions 'construction' and 'constructional idiom' in accounting for the dependency of morphology on morphological and syntactic constructions. In section 2, the morphological construction '*uit* + past participle', as exemplified by the complex word *uitgepraat* in the sentence *Ik ben uitgepraat* 'I am done with talking' will be analyzed in section 2. It will be shown that the use of *uit* with the meaning 'done with' is licensed only by the presence of a particular morphological form, the participial adjective. We refer to this kind of dependency as 'constructional licensing', which means that the use of words with specific meanings is licensed by specific (morphological and/or syntactic) constructions.

In section 3, we will discuss the behaviour of particle verbs with the particle *aan* such as *aanlopen* 'to arrive by walking'. These particle verbs can only be used in the form of a participle or an infinitive, and only in combination with the verb *komen* 'to come'. Hence, the formation of such lexical units is constructionally restricted. In other words, this use of *aan* with the meaning 'to arrive by' is licensed by a specific construction with certain morphological and syntactic properties.

2. Participial Compounds with *uit*

In order to understand the analytic issues involved in accounting for sentences such as *Ik ben uitgepraat* 'I am done with talking' mentioned in section 1, we first introduce some background assumptions on particle verbs since *uitgepraat* looks at first sight as the past participle of the particle verb *uitpraten*. Indeed Dutch has such a particle verb *uitpraten*, but it has a different meaning, as illustrated in (6):

(6) *Wij hebben het probleem uit-ge-praat*
We have the problem out-talked
'We talked out the problem'

(Dutch past participles are marked by both a prefix *ge-* unless the stem begins with an unstressed prefix, and a suffix *t/d* or *-en*; the suffix *t/d* is not realized phonetically after a stem ending in *t/d*.) The particle verb *uitpraten* also occurs with another meaning, 'finish talking'. Interestingly, this use of *uitpraten* is dependent on the presence of the permissive verb *laten* as the verb of the main clause, as in *Jij laat me niet uitpraten* 'You do not let me finish talking'. Thus, *Ik praat uit* 'I finish talking' is not possible. This latter type of restriction is similar to the case discussed in section 3.

Particle verbs are combinations of two words, a particle and a verb that form a lexical unit. They have been the subject of detailed research and discussion (cf. Dehé et al. eds. 2002, Blom 2005 and the literature mentioned in these references). The basic insight that is presupposed in this article is that particle verbs are not words but phrasal lexical units (Booij 2002). Hence, their formation does not belong to the domain of derivational morphology. Instead, each type of particle verb should be seen as a constructional idiom with phrasal properties. For instance, the set of Dutch particle verbs with the particle *door* can be characterised by means of the following constructional idiom:

(7) $[door\ [x]_V]_{V*}$ 'to continue V-ing'

where V* stands for a minimal verbal projection (cf. Booij 2002, Blom 2005). In this template, the slot for the verb is represented as a variable. This indicates that this position in the construction is not fixed, and can be occupied by all sorts of verb. Hence, the pattern is qualified as productive. Indeed, new particle verbs with *door* can be coined very easily. Besides existing particle verbs of this type, such as *doorwerken* 'to continue working', we may coin new ones such as *doorfotograferen* 'to continue taking pictures'.

The word *door* is also used in other contexts, as an adverb or a preposition with a range of meanings. However, when used as a particle in new cases of particle verb formation, it only has the meaning 'to continue V-ing'. This is expressed exactly by the constructional idiom in (7). It is this configuration that licenses the use of *door* with a continuative meaning.

The basic reason for considering particle verbs as multi-word units is that they can be split. In Dutch main clauses the finite forms of verbs appear in second position, but the particle is stranded:

(8) a. *Wij werkten de hele nacht door*
'We continued working all night'

The particle and the verb are split by the infinitival particle *te*, and the participial prefix *ge-* appears after the particle:

(8) b. *Wij besloten de hele nacht door te werken*
'We decided to continue working all night'

c. *Wij hebben de hele nacht door-ge-werkt*
'We continued working all night'

(In Dutch orthography particle verbs are written as one word when the two parts are linearly adjacent, as is the case in sentence (8c).)

Dutch features a number of particle verbs with the particle *uit*, such as *uitpraten* (cf. 6), *uitwerken* 'to work out' and *uitkleden* 'to undress'. However, the word *uit* can also be used in combination with participles with a special meaning, as exemplified by the following examples (taken from Booij 2004: 280):

(9)

uit-gegeten	'lit. out-eaten, finished eating'
uit-geschilderd	'lit. out-painted, finished painting'
uit-gekleuterd	'lit. out-toddlered, finished being a toddler ("*totdat haar dochter uitgekleuterd was*", J. Pardoen, *Trouw*, 5 April 2002)
uit-gebodemd	'lit. out-bottomed, has reached its lowest point, said of shares on the stock market, *Trouw*, 2 Nov 1999)

These *uit*-participles have a number of special distributional, semantic, syntactic and morphological properties which will be considered in turn.

Participles such as the ones under (9) are used as predicates in combination with the copula *zijn* 'to be', or similar copulas such as *raken* 'to become':

(10) *We zijn uit-gepraat* 'we are done talking'
We raken uit-gekeken op elkaar 'we are becoming tired of each other'

They might occasionally also be used as attributive adjectives, unlike past participles of un-ergative verbs. An example from the internet is (11a); the examples in (11b) are constructed by us:

(11) a. *Een jonge zender zoals Veronica kan zich geen inspiratieloze en uitgeprate medewerkers veroorloven*
'a young broadcasting station such as Veronica cannot afford employees without inspiration and who are through with talking'

b. *de intussen uitgegeten gasten*
'the guests who have finished eating'

de al weken uitgeprate onderhandelaars
'the negotiators who are done talking'

de uitvergaderde professoren
'the professors who are done conferring'

de uitgekleuterde kinderen
'the children who have finished being toddlers'

Can the participles in (9–11) be interpreted as forms of particle verbs such as *uiteten* 'to eat out' and *uitpraten* 'to talk out'? The first thing to notice is the contrast in argument structure between the first two phrases in (11 b) and the corresponding particle verbs in (12). In (11), the head of the NP is bound to the subject argument of the verb: they refer to negotiators who are done talking, to guest who have finished their meal. In contrast, NPs such as

(12) *de uitgegeten boomtakken* 'the eaten-out branches'
het uitgeprate probleem 'the talked-out problem'

require the head of the NP to be interpreted as the Theme, and not as the Agent of the particle verbs *uiteten* 'to eat out, and *uitpraten* 'to talk out', respectively. The latter are transitive verbs, whereas the *uit*-participles do not take an object.

Also in terms of semantics, particle verbs and *uit*-participles can differ widely. The examples in (13) show that the particle verbs *uit-eten, uit-praten* and *uit-kijken* cannot be the bases of the *uit*-participles *uitgegeten* 'done eating', *uitgepraat* 'done talking' and *uitgekeken* 'tired of' because their meaning is markedly different:

(13) *uiteten* 'to eat out, to have a farewell dinner with somebody'
uitpraten 'to talk out a conflict'
uitkijken 'to watch out for, to look forward to'

As another point on the semantic side, the *uit*-participles do not carry the telicity reading normally associated with past participles. The participles freely combine with expressions such as 'for weeks', which clash with the telic semantics of the homophonous particle verb. Example (14) illustrates the contrast:

(14) *We zijn al weken uitgepraat*
'We've been done talking for weeks'

**We hebben het probleem al weken uitgepraat*
'We've talked out the problem for weeks'

As observed by Coppens (2001), the *uit*-participles express a state or property rather than an event and they can also be used in a context that refers to a future situation:

(15) *We zijn a.s. maandag uitgepraat*
'we'll be done talking next Monday'
**We hebben a.s. maandag uitgepraat*

Morgen zijn ze wel uitvergaderd
'they are sure to be done conferring tomorrow'
**Morgen hebben ze wel uitvergaderd*

The contrasted utterances show that the *uit*-participles in (14) and (15) also differ from the particle verbs in the kind of auxiliary they select, *zijn* versus *hebben*.

(16) *We hebben / *zijn het probleem uitgepraat*
*We hebben / *zijn naar jullie uitgekeken*

Syntactically, the *uit*-participles differ from particle verbs in their behaviour in verb raising contexts. Verb raising is an operation in which the verb of a lower clause is raised to a higher clause and forms a verbal complex with the verb of that higher clause. In cases of verb raising, the particle can optionally be raised to the higher clause, unlike other elements of the embedded VP except the V itself. Verb raising is disfavoured for the *uit*-participles:

(17) a. Particle verb *uitpraten* 'to talk out':
dat we het probleem *uit* willen *praten*
dat we het probleem willen *uitpraten*
'that we want to talk out the problems'

b. *Uit*-participle:
omdat we *uitgepraat* raken
*omdat we *uit* raken *gepraat*
'because we are becoming tired of talking'

hoewel we *uitgekeken* zijn op elkaar
*hoewel we *uit* zijn *gekeken* op elkaar
'although we've got tired of each other'

A corpus search reveals that sentences with split participles do occur, suggesting that some language users might reanalyse them as 'participles of a particle verb'. The following sentences illustrate this split use:

(18) *Nadat we uit waren gezoend wilden we op een bankje gaan zitten*
'After we had finished kissing, we wanted to sit down on a bench'

Tot je uit bent gehuild ...
'till you have finished crying…'

In morphological terms, words such as *uitgeschilderd*, *uitgekleuterd* of *uitgebodemd* (recall examples 9) are special in that they only occur in participle form. The corresponding verbs **uitkleuteren* and **uitbodemen* do not exist, neither do the verbs **uiteten,* uitschilderen* with the appropriate meaning. This shows that they have been formed directly, without the stem of such particle verbs being involved. Hence, what we see here is that a particular morphological form forms a direct building block of a word, without the stem of the corresponding verb forming an intermediate step in the construction.

The specific properties of the *uit*-participles reviewed above suggest that they are not verbs at all, but adjectival compounds headed by a participle. Participles are words that have both adjectival and verbal properties. This analysis is supported by the fact that compounding with adjectival heads is productive in Dutch, whereas compounding with verbal heads is not (Booij 2002). Hence, participles can appear in the head position of compounds because they have adjectival features. These adjectival compounds specify properties (whereas participles functioning as verbs specify events). There are many other types of adjectival compounds in Dutch in which participles functions as heads, for instance *tijd-gebonden* 'time-bound' and *toekomst-gericht* 'future-oriented' (Booij 2002: 76-79, 153-157).

The productivity of the *uit*-participle pattern can be substantiated by a corpus and an internet search. (19) lists examples from the Corpus Gesproken Nederlands and from Google. (Corpora are of particular value for complex words based on derived forms like participles, since - for want of an infinitival form - they will generally not appear in dictionaries.)

(19) *ik ben nou echt wel uitgeluld eigenlijk*
'actually, I'm really done gabbing (lit. out-gabbed) now'

dan bel ik daarna nog wel terug als we niet uitgekletst zijn
'then I call you back later if we're not yet done chatting (lit. out-chatted)'

Uitgeruzied met uw partner?
'finished quarrelling (lit. out-quarrelled) with your partner?'

Het heeft in totaal 10 jaar geduurd eer we uitgetwijfeld waren en dan nog wou mijn man liever geen kinderen
'it took us 10 years in all before our doubts were resolved (lit. before we were out-doubted), and even then my husband didn't want children'

voor wandelaars die in Nederland zijn uitgewandeld. Meer dan 100 Belgische wandelingen...
'for hikers who are done walking (lit. out-walked) in the Netherlands. More than a 100 Belgian hikes...'

Mijn dochter is inmiddels ruim 9 maanden, en ik ben al 3 maanden gestopt met borstvoeding, dus inmiddels moet ik toch wel uitgezwangerd zijn
'my daughter is now more than 9 months old, I stopped breast-feeding three months ago, so I should be done being pregnant (lit. uit-pregnanted) by now'

In all these cases, the particle *uit* 'out' in combination with a past participle receives the specific interpretation 'done with', while the preposition *uit* 'out', as its English counterpart, has a range of related meanings. As argued in Booij (2005b), it is a common feature of words to acquire a specific meaning that is bound to their occurrence in lexical or syntactic constructions. For instance, the Dutch word *hoofd* 'head' when used as the first part of compounds has the meaning 'main', as in *hoofd-gebouw* 'main building' and *hoofd-ingang* 'main entrance'. Similarly, the word *uit* has acquired a specific meaning 'done with' as part of the lexical units of the type under discussion here.

The participles involved thus receive an interpretation that is bound to this morphological construction and is different from the regular ones: it specifies the end-state of the Agent of the action expressed by the verbal stem. In sum, the following constructional idiom can be assumed for Dutch:

(20) $[[uit][[x]_{V\text{-}ptcp}]_A]_A$ 'done with V-ing'

(PTCP = participle). Note that the English word *done* has a similar special meaning, and selects the auxiliary *to be* for this interpretation whereas normally a form of *to have* is selected (compare *I am done* to *I have done*).

These *uit*-participles also show their participial potential with respect to word order in embedded clauses. In Dutch we find two orders for participles, before or after the finite verb; adjectives, on the other hand, can only occur before the finite verb:

(21) *omdat ik ziek ben / *ben ziek* 'because I am ill'
omdat ik gepraat heb / heb gepraat 'because I have talked'
omdat ik uitgepraat ben / ben uitgepraat 'because I'm done talking'

als je bent uitgegeten, kan je meteen op je ski's stappen (Google)
'When you've finished eating you can get on your skis directly'

The verbal stem for the participle in template (20) may be either an existing verb, or a possible verb created through conversion of nouns to verbs. For instance, in the examples *uitgekleuterd* and *uitgebodemd* in (9) we recognize the verbs **kleuter* and **bodem* which do not exist as such in the Dutch lexicon. That is, these cases of N to V conversion have been triggered by the use of this constructional idiom with *uit*. A similar case is found in (19) where the adjective *zwanger* 'pregnant' has been converted to a verb. The co-occurrence of word formation patterns is a widespread phenomenon, and can be expressed by assuming that the language user may make shortcuts while

making new multiply complex words. Formally, this can be expressed by the 'conflation' of word formation templates. For instance, the following template conflation can be assumed for words such as *uitgekleuterd*:

(22) $[[uit][[x]_{V\text{-}ptcp}]_A]_A + [[x]_N]_V = [[uit][[[x]_N]_{V\text{-}ptcp}]_A]_A$

A nice illustration of this specific use of *uit* is the following headline of a recent newspaper article about a farewell concert by the conductor Jaap van Zweden who gave up his position as conductor of the Orkest van het Oosten:

(23) *Van Zweden uitgezwaaid, maar niet uitgezwaaid* (*Trouw*, 29 October, 2005)
Van Zweden out-waved, but not out-waved
'Van Zweden has been waved good-bye, but is not tired of conducting'

In sum, the class of *uit*-participles discussed here shows that the specific meaning of *uit* 'done with V-ing' is licensed by a specific morphological construction, adjectival compounds with a participial head.

In the next section, we will discuss a case in which the use of the word *aan* with a specific meaning is licensed by a morphologically and syntactically restricted context.

3. *Aan*-participles

In this section we will focus on a particular phenomenon in relation to particle verbs, the fact that in some cases the productive formation of Dutch particle verbs is restricted to a subset of their possible morphological forms (infinitive or past participle), and to a specific syntactic construction.

Dutch features a number of particle verbs with *aan*. They can be divided into the following three semantic subclasses (De Haas & Trommelen 1993: 135-36):

(24) a. Surface contact with object:
aanrijden 'to hit by riding';

b. Durative aspect with pejorative connotations:
aanmodderen 'to muddle on';

c. Inchoative aspect:
aansnijden 'to start cutting'

The use of the particle *aan* that is the focus of this section is its use in clauses with the verb *komen*, as illustrated by the following example:

(25) *Jan kwam aan-lop-en / aan-ge-lop-en*
John came at-walk-INF / at-PTCP-walk-PTCP
'John came walking'

(INF = infinitive). As the glosses indicate, the participle has no inherent perfective meaning since the aspect of sentence (25) is non-perfective.

The verb *komen* 'to come' can combine with verbal infinitives as in

(26) *Francesca komt et-en*
Francesca comes eat-INF
'Francesca will come for dinner'

Lourens kwam een boek breng-en
Lourens came a book bring-INF
'Lourens came to bring a book'

In other words, Dutch has a constructional idiom of the form [*komen* […]$_{VP}$] in which the verb *komen* expresses a physical movement with a certain goal specified by the VP. In this construction with the bare infinitive, the lexical meaning of *komen* is still present: it expresses a spatial movement, and hence we cannot use it as a general aspectual marker, unless we add the infinitival particle *te* (and hence use a different construction):

(27) **Hij kwam overlijden / Hij kwam te overlijden*
He came die.INF / He came to die.INF
'He died'

The open VP-slot of this constructional idiom with *komen* can be filled by infinitival and participial forms of verbs that express some movement and with the direction of movement specified:

(28) *Hij komt de hoek om huppelen / gehuppeld*
He comes the corner around hop.INF / hop.PTCP
"He comes hopping around the corner'

Hij komt naar beneden huppelen / gehuppeld
He comes downwards hop.INF / hop.PTCP
'He comes hopping downwards'

This is a clear case of construction-dependent morphology (cf. Booij 2005a for other examples) since this use of the participles is dependent on their occurring in this constructional idiom with *komen*.

One of the directional expressions that can be used in this construction is the particle *aan* with the meaning 'in the direction of the speaker', as shown in (25). The important point to be noted here is that the occurrence of particle verbs such as *aanlopen* depends on their being embedded in the *komen VP* construction. Hence, these particle verbs can only appear in either the infinitival or the participial form. The word sequences *aan + V* (in infinitival or participial form) have the status of particle verb even though so far we only saw them in syntactic contexts in which they are not split. We can conclude this from the verb raising test (cf. 17). As the following sentences show, the word *aan* in the construction under discussion indeed behaves as a particle, since both word orders are possible, as is generally the case with particle verbs (the second example with the verb and the particle split comes from a Google search):

(29) a. *... dat Jan kwam aanrijden / aangereden*
… that John came at-ride-INF / at-ride-PTCP
'… that John came riding in'

... dat Jan aan kwam rijden / aan kwam gereden

b. *...dat ze aan komt gehuppeld*
…that she at comes hop-PTCP
'… that she comes hopping toward us'

Remember that *aanrijden* and *aanhuppelen* are not used as particle verbs with these meanings in all VPs, but only in this construction.

The sub-construction [*komen aan* V-INF] is a very productive one, and is illustrated by the following examples from the Corpus Gesproken Nederlands:

(30) *als hij roept komen Anouk en Brigit aangerend*
'when he calls Anouk and Brigit come running'

haar huisje is zo'n familienest waar iedereen komt aangewaaid
'her house is such a family nest where everybody comes blowing in'

de heks kwam donker en dreigend aangeslopen
'the witch came creeping up darkly and menacingly'

hij is aan komen fietsen met z'n dochtertje
'he came cycling with his daughter'

kwam een tweede toestel aanvliegen en ramde de andere toren van 't WTC.
'came a second machine flying and hit the other tower of the WTC'

The property relevant for the argumentation in this article is that these particle verbs with *aan* can only be used in this construction. They do not exist independently from this construction. For instance, we do not have sentences like the following, with finite forms of the particle verbs:

(31) **Jan loopt aan* 'John arrives by walking'
**Jenny rent aan* 'Jenny arrives by running'

Some of these particle verbs may exist, but with a different meaning.

The particle verbs with *aan* in the *komen*-construction are always intransitive verbs that express an action. As we saw, the formation of these particle verbs is dependent on the *komen*-construction. This is why we call such particle verbs constructionally dependent: their formation is dependent on a particular construction, and is not used in a context-free manner for the expansion of the fund of lexical units of Dutch. To put it differently, this use of *aan* is licensed by the presence of a specific morphological form (the infinitive) of a verb, in a specific syntactic construction (after *komen*). The same observation on constructional dependency can be made for the other variant mentioned in (25): the particle verb with *aan* with a participle (there is no

meaning difference between the infinitival and participial variants, and the variation appears to be mainly geographical: the infinitival variant is the preferred one in the Western part of the Netherlands, whereas the participial variant is preferred in Belgium, Haeseryn et al 1997: 965).

As we saw, the *komen* + participle construction imposes a non-perfective interpretation on the participle, just like the passive construction with the verb *worden*. It is enlightening to compare this construction with some other uses of the verb *komen*. This verb can also be used as a copula, with an adjective functioning as the predicate nominal. Since participles can function as adjectives, they can also be used in that way:

(32) a. *Hoe komt die broek zo vies?*
How comes those trousers so dirty?
'How did these trousers get so dirty?'

b. *Hoe komt die vaas gebroken?*
How comes that vase broken?
'How did that vase get broken?'

This use of *komen* appears to be restricted to sentences that open with *hoe* 'how', and is therefore comparable with the English idiom *how come* …. In sentence (32b), the participle *gebroken* mentions a property that is the result of an event of breaking. In the construction [*komen aan* Participle] on the other hand, there is no perfective meaning, sentences with this construction express an ongoing event.

The verb that can be used in the *komen*-constructions does not have to mention a type of motion, since the verb can also describe an accompanying aspect of the motion, as is illustrated by the following examples (Haeseryn et al 1997: 965):

(33) *Daar komt mijn neef Nurks weer aan-mopperen / aan-gemopperd*
there comes my nephew Nurks again towards.speaker-complain.INF/PTCP
'there my nephew Nurks comes complaining'

In de verte kwamen dronken matrozen aan-zwaai-en / aan-ge-zwaai-d
in the distance came drunken sailors towards.speaker-sway-INF/PTCP
'in the distance, drunken sailors came swaying towards us'

Daar kwam de auto aan-toeter-en / aan-ge-toeter-d
there came the car towards.speaker-hooting-INF/PTCP
'There the car came hooting its horn'

The directional expression requires a movement interpretation of the verb, and hence the verbs under (33) are interpreted as verbs of motion with a specific manner, a typical case of type coercion.

A special case of the use of this construction is the occurrence of the particle verb *aankakken* with the idiosyncratic, unpredictable meaning 'to show up', with a pejorative connotation (the verb *kakken* means 'to shit'). Again, this particle verb *aankakken* can only be used in combination with the verb *komen*:

(34) *Jan kwam weer aan-kakk-en / aan-ge-kak-t*
John came again towards.speaker-shit-INF/PTCP
'John showed up again'

**Jan kakte weer aan*
'John showed up again'

This means that individual instantiations of the *komen aan V* constructions have to be specified in the lexicon if they have unpredictable meaning aspects. Another example of such a verb is the particle verb *aanzetten*, as in *Hij kwam aanzetten* 'He showed up'. Both *aankakken* and *aanzetten* can only used with this meaning in the *komen aan* construction. (The particle verb *aanzetten* also occurs in other contexts, but with a completely different meaning 'to prompt, to urge' and with transitive valency.)

There is an interesting contrast between *aankakken* and *aanzetten* since the latter from cannot be used in the partcipial form in this construction: *Hij komt aanzetten / *aangezet*. This suggests that the form with the infinitive is entrenched to such a degree that the other form is blocked from being produced. Such forms of entrenchment may be expected given the highly lexicalized nature of these constructions.

To conclude, the observations presented above show that Dutch has two almost identical constructional idioms with a meaning that focuses the addressee on the manner in which the subject performs the act of arriving:

(35) [*komen* [*aan* V-INF / V-PTCP]$_{V'}$]$_{VP}$

'to arrive by means of / while V-ing'

These constructions are instantiations of a more general constructional idiom in which the verb *komen* has a complement that is a directional VP. The non-perfective interpretation of the participle is dependent on this syntactic context. By specifying (35) as as specific instantiation of the *komen*-VP construction we express that the occurrence of particle verbs with directional *aan* depends on this specific construction. In other words, the creation of particle verbs with *aan* with this particular meaning is construction-dependent, and this use of *aan* is licensed by a specific construction defined in both syntactic and morphological terms, as shown in (35).

4. Conclusions

In this article we presented two cases of constructional licensing: the occurrence of words with a particular meaning appeared to be licensed by specific morphological and/or syntactic configurations. The correct specification of the licensing environments implied that information on morphological or syntactic context may be necessary for a proper account of the formation of words and of particle verbs.

The phenomenon of constructional licensing is in line with the findings in Booij (2005a) on the dependency of morphological and lexical constructions on specific syntactic configurations. Hence, what we need in order to provide a correct and insightful account of the facts of Dutch of the sort discussed above is a "constructicon"

of constructional idioms of varying degrees of abstractness in which reference can be made to morphological properties of the words used in these constructional idioms.

References

Blom, Corrien 2005. *Complex predicates in Dutch. Synchrony and diachrony.* Ph. D. diss, Vrije Universiteit Amsterdam. Utrecht: LOT.

Booij, Geert 2002. *The morphology of Dutch*. Oxford: Oxford University Press.

Booij, Geert 2004. Constructions and the interface between morphology and syntax. In Henk Aertsen, Mike Hannay and Rod Lyall (eds.), *Words in their places. A Festschrift for J. Lachlan Mackenzie*. Amsterdam: Faculty of Arts, Vrije Universiteit Amsterdam, 275-282.

Booij, Geert 2005a. Construction-dependent morphology. *Lingue e Linguaggio* 4, 31-46.

Booij, Geert 2005b. Compounding and derivation: evidence for Construction Morphology. In Wolfgang U. Dressler, Dieter Kastovsky, Oskar Pfeiffer and Franz Rainer (eds.) *Demarcation in morphology*. Amsterdam / Philadelphia: John Benjamins, 111-139.

Coopens, Peter 2001. Al uitgegeten? *Neder-L* 41.

Dehé, Nicole, Ray Jackendoff et al (eds.) 2002 *Verb-particle explorations.* Berlin : Mouton de Gruyter.

Goldberg, Adele 1995. *Constructions. A construction-based approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, Adele 2003. Constructions; a new theoretical approach to language. *Trends in Cognitive Science* 7, 219-224.

Haas, Wim de & Mieke Trommelen 1993. *Morfologisch Handboek van het Nederlands*. Den Haag: SDU.

Haeseryn, Walter et al. 1997. *Algemene Nederlandse Spraakkunst.* Groningen: Martinus Nijhoff / Deurne: Wolters Plantyn [2nd edition].

Jackendoff, Ray S. 2002. *Foundations of Language*. Oxford: Oxford University Press.

Michaelis, Laura A., and Knud Lambrecht 1996. 'Toward a construction-based theory of language function: The case of nominal extraposition.' *Language* 72, 215-247.

Verb-particle Constructions and Prefixed Verbs in Italian: Typology, Diachrony and Semantics*

Claudio Iacobini
Università di Salerno
ciacobini@unisa.it

Francesca Masini
Università Roma Tre
fmasini@uniroma3.it

Abstract

Verb-particle constructions are generally regarded as a peculiar property of Germanic languages. In this paper we show that verb-particle constructions also exist in Italian. The presence of verb-particle constructions in Romance languages seems to contradict Talmy's generalization about "frame-based languages" and "satellite-based languages", which makes these constructions a rather interesting typological issue. Further, the typological perspective raises the question of the development of these complex verbs in the Italian language.
The paper briefly deals with these typological and diachronic issues and then focuses on the semantic properties of these constructions in present-day Italian. In particular, we will analyse the actional properties of Italian post-verbal particles. The results of the analysis allow us to outline the diachronic and synchronic relationship between verbal prefixes and post-verbal particles, which turn out to cooperate in the expression of locative and aspectual meanings.

1. Introduction

The paper deals with verb-particle constructions (hereinafter VPC), i.e. complex predicates formed by a verbal base and a modifying particle. In recent years, a lot of interest has been devoted to these constructions, and investigation has been focusing mostly on the structure of VPCs in the Germanic languages, where the pattern is very productive and widespread in use.

Recently, some studies have showed that similar constructions also exist in Italian (cf. e.g. Schwarze 1985, Venier 1996, Simone 1997, Antelmi 2002, Jezek 2002, Iacobini 2003, Jansen 2004, Masini 2006, Cini, in press). Some examples of Italian VPCs are given in (1).

* We would like to thank the organizers of MMM5 for the pleasant atmosphere and the audience in Frejus for helpful discussion. We are also grateful to Elisabetta Jezek and Anna M. Thornton for useful comments on a previous version of this paper. The article is the result of the close collaboration of both authors; however, for academic purposes, Claudio Iacobini is responsible for sections 4.1.1, 4.3, 4.3.1, 4.3.2, 4.4, 5, 6, and Francesca Masini for sections 1, 2, 3, 4, 4.1, 4.2. A revised and extended version of this paper appeared as Iacobini, C. & F. Masini (2007), "The emergence of verb-particle constructions in Italian: locative and actional meanings". In *Morphology* 16 (2) (www.springer.com)

(1)	*venire giù*	lit. come down	'to come down, to descend'
	portare via	lit. take away	'to take away'
	mettere sotto	lit. put under	'to put (something) under, to run over'

Besides confirming the existence of VPCs in this language, our contribution aims at improving our knowledge of the semantic and aspectual properties of VPCs in Italian and, possibly, in general. In particular, we will provide new data that show that Italian post-verbal particles contribute to the *Aktionsart* of VPCs, which may be regarded as a further evidence of their establishment in the system. In the light of these results, we will take into consideration the diachronic and synchronic relationship between verbal prefixes and post-verbal particles in the expression of locative and aspectual meanings.

The paper is organized as follows. Section 2 provides an overview of the phenomenon of VPCs in Italian and in general. In particular, section 2.1 summarizes the previous studies on VPCs in general and clarifies our theoretical position and goals, whereas section 2.2 outlines the typological and diachronic background of VPCs in Italian. In section 3 we offer a brief structural and semantic description of Italian VPCs. Section 4 contains the actional analysis. After introducing the framework and the criteria adopted here, we will illustrate the findings of the analysis, which was carried out on a corpus of 165 Italian VPCs. Particular attention will be devoted to the telicity feature. Finally, section 5 faces the question of the relationship between VPCs and verbal prefixation, whereas section 6 contains some conclusive remarks.

2. Verb-particle Constructions: An Overview

In this section, we will offer an overview of VPCs, from both a theoretical and a typological standpoint. In section 2.1 we will offer a brief account of previous studies on VPCs. Further, we will outline the theoretical assumptions adopted here, as well as the main goal of our contribution, i.e. the investigation of the actional properties of Italian VPCs. Section 2.2 will add some typological remarks on the distribution of VPCs among Indo-European languages and will advance some diachronic considerations about their development in Italian. This will set the discussion for the secondary goal of the paper, i.e. the comparison between VPCs and verbal prefixation (cf. section 5).

2.1. Theoretical Background

VPCs in English (also known as *phrasal verbs* or *particle verbs*), have been largely studied, suffice it to mention the contributions by Bolinger (1971), Fraser (1976), Dixon (1982), and, more lately, den Dikken (1995) and Dehé (2002). As for the other Germanic languages, cf. e.g. Booij (2002a,b) and Blom (2005) for Dutch, Stiebels & Wunderlich (1994), Lüdeling (2001) and Müller (2002) for German, Toivonen (2003) for Swedish and Jansen (2002) for Danish[1]. Further, VPCs have been also identified in

[1] Cf. Haiden (2002) for a comparative overview of the phenomenon in the various Germanic languages.

Ugro-Finnic languages such as Estonian and Hungarian (cf. e.g. Ackerman & Webelhuth 1998, Ackerman 2003).

Over the last decade, in Generative Grammar there has been an increasing interest in VPCs in the Germanic languages, which was mainly due to their ambiguous structural status between words and phrases (cf. the introductory chapter in Dehé *et alii* 2002). Indeed, the big question was: do they belong to morphology or syntax? This demarcation problem of course refers to a modular architecture of the grammar in which the components are autonomous and distinct from each other. Consequently, the different proposals to unravel the puzzle were based on a series of syntactic and semantic criteria that pointed at demonstrating the word-like or phrase-like status of these constructions. Of course, a number of different technical solutions were put forward to account for the properties of VPCs, from the Small-Clause analysis (cf., among the others, den Dikken 1995) to the "non-projecting word" proposal for particles (cf. Toivonen 2003). However, until today, generative grammarians have not reached a general agreement on the kind of structure to assign to VPCs.

A new perspective on the issue, which we will adopt here, was introduced by Booij (2002a,b). In his contributions, Booij claims that VPCs in Dutch (i.e. so-called *Separable Complex Verbs*) are a case of "periphrastic word formation", i.e. lexical items that behave functionally as complex words but display a phrasal structure. Technically speaking, these complex verbs are regarded as *constructional idioms*[2], i.e. semi-specified syntactic structures with a (partially) noncompositional meaning that are stored in the lexicon and display a limited productivity. As Booij states, his proposal is in line with the basic tenets of Construction Grammar (cf. Fillmore, Kay & O'Connor 1988, Goldberg 1995, 2003), which claims that language consists in a network of constructions, i.e. form-meaning pairings differing in size and complexity. Of course, this implies a non-modular view of language and the presence of a syntax-morphology-lexicon *continuum*[3].

This scalar vision of grammar allows us to leave the demarcation problem in the background, due to the non-strict separation between what we traditionally refer to as the modules of the grammar. Given this, VPCs are no longer an anomaly from the point of view of the structure, but rather an expected case. Of course, this does not mean that the demarcation of phenomena is not relevant. Indeed, it is important for the individuation of the links between the different constructions. However, a lot of play is also made about the construction itself, its meaning or function, and the interaction between the constituting elements.

The constructionist standpoint just envisaged encouraged us to face an important though nowadays overshadowed issue, i.e. the semantics of VPCs, and in particular their actional properties with respect to the verbal bases and the kind of particles used. Indeed, whereas earliest contributions reflected the importance of the semantic properties of VPCs (cf. e.g. Bolinger 1971, Dixon 1982, Lindner 1983, Brinton 1988), recent works have devoted less attention to semantics, with some notable exceptions (cf. Jackendoff 2002a for English, McIntyre 2001, 2002 for German, Blom 2005 for Dutch).

The literature on Germanic VPCs usually describes their semantics according to the following tripartite classification (cf. in particular Dehé *et alii* 2002):

---

[2] The notion of *constructional idiom* can be found both in Goldberg (1995) and in Jackendoff (1997, 2002b).

[3] Cf. Booij (2005a) for a constructionist approach to morphology and Booij (2005b) for considerations about the interaction and interdependency of morphology and syntax.

i. *locative meanings*, due to the fact that VPCs originate from the combination of motion verbs and locative particles;

ii. *idiomatic meanings*, due to semantic bleaching;

iii. *aspectual* and/or *actional meanings*, with particular reference to telicity and duration.

While points *i* and *ii* can be easily applied to the Italian situation (Simone 1997, Masini 2005; cf. also section 3.2), point *iii* has not been investigated yet and will be the topic of our discussion.

In conclusion, the goal of this paper is twofold. On the one hand, we aim at contributing to the study of VPCs in general by extending the research to a new language, i.e. Italian. On the other hand, we decided to focus on semantics rather than on structure, in order to investigate a so far unknown area: the *Aktionsart* of these constructions in Italian.

2.2. Typological Background and Diachronic Development

Italian VPCs constitute an interesting case also from a typological viewpoint. According to the well-known study on lexicalization patterns found in Talmy (1985), the motion event consists of four internal components (*Figure, Ground, Path* and *Motion*) and some external components such as *Manner/Cause*[4]. Talmy identifies two main lexicalization patterns for verbal roots in Indo-European languages, i.e. *Motion+Manner/Cause* and *Motion+Path*. These two patterns typically correspond to Germanic and Romance languages respectively[5]. As a consequence, Germanic languages are defined as *satellite framed languages*, as they lexicalize the Manner/Cause of the motion event and specify the directional values by means of external particles, while Romance languages would be an example of *verb framed languages*, as they lexicalize the Path and leave the Manner/Cause specification to adjuncts. This situation is illustrated in Table 1.

LANGUAGE FAMILY	THE COMPONENTS OF A MOTION EVENT TYPICALLY REPRESENTED IN THE VERB	
	Verb Root	*Satellite*
Romance languages	Motion + **Path** (e.g. Spanish *poner*, *meter, subir*)	**Ø**
Other Indo-European languages (mainly **Germanic**)	Motion + **Manner/Cause** (e.g. English *to roll, to blow, to throw*)	**Path** (e.g. English *to run <u>out</u>*)

Table 1. Typology of verbs of motion and satellites in Indo-European languages (adapted from Talmy 1985)

[4] In the terminology of Talmy (2000b), these external components are defined as *co-events*.
[5] Talmy uses English and Spanish as sample languages for the Germanic and Romance types.

However, if we take into consideration Italian VPCs such as those exemplified in (1), it will be clear that present-day Italian differs from other major Romance languages, and in particular from Spanish, as regards Talmy's generalization. Indeed, the primary function of Italian post-verbal particles seems to be the addition of directional values to the verbal root. Hence, they function as true satellites, just as in the Germanic languages. Of course, this does not mean that Italian lacks verbal roots of the "Romance type". Rather, we would say that they are no longer the *only* or the *privileged* way of realizing the Path feature in Italian[6]. In fact Table 2, which contains some examples of English VPCs with the verbal base *to go* and their Italian counterparts, illustrates that, in current use, Italian can employ both the "Romance type" (central column) and the "Germanic type" (right column).

English *to go*	**Italian** *andare*	
VERB ROOT + SATELLITE	VERB ROOT	VERB ROOT + SATELLITE
to go after	*seguire*	*andare/correre dietro*
to go ahead	*procedere/continuare*	*andare avanti*
to go away	*andarsene*	*andare via*
to go back	*(ri)tornare*	*andare/tornare indietro*
to go down	*scendere*	*andare giù*
to go for	*avventarsi*	*andare/lanciarsi contro*
to go in	*entrare*	*andare dentro*
to go on	*continuare*	*andare avanti*
to go out	*uscire*	*andare fuori*
to go (a)round	*girare*	*andare attorno*
to go up	*salire*	*andare su*

Table 2. Some English VPCs with to go compared with their Italian counterparts

It is worth noting that also Latin belongs to the satellite framed family. In fact, Latin had a very productive system of verbal prefixes that functioned as satellites. This is quite interesting in diachronic terms, as it means that, in the passage from Latin to Italian, there was a change in the kind of satellites used for the expression of the Path in verbal constructions (from prefixes to particles). This induces us to advance some considerations concerning the development of VPCs in Italian (cf. Iacobini 2003, Masini 2005, 2006).

In our view, three major factors can be identified that contributed to the development of Italian VPCs. First of all, the passage to a more diagrammatic technique of overt locative marking, due to the morphosemantic bleaching of the Latin prefixed motion verbs in the Romance languages (cf. Schwarze 1985). As you can see from Table 3, Italian has both synthetic forms, which derive directly from Latin and are no

[6] For typological considerations about the way Italian lexicalizes Path and Manner verbs cf. Wienold & Schwarze (2002).

longer morphologically analysable, and analytic forms, which can be considered as a true Italian formation[7].

LATIN transparent locative prefixation		**ITALIAN synthetic forms inherited from Latin**		**ITALIAN analytic forms (VPCs)**	
ascendere 'to ascend'	*discendere* 'to descend'	*salire*	*scendere*	*andare su*	*andare giù*
inire 'to enter'	*exire* 'to go away'	*entrare*	*uscire*	*andare dentro*	*andare fuori*

Table 3. Synthetic and analytic verbs of motion in Latin and Italian

Secondly, a crucial role was played by the weakening of Italian verbal prefixation to express locative meanings (cf. Iacobini 2005), which also contributed to the morphosemantic bleaching of Latin prefixed verbs. Contrary to Latin, Italian verbal prefixes do not seem to be very productive for the expression of locative meanings, as Table 4 illustrates[8].

LATIN transparent locative prefixation		**ITALIAN weakness of verbal prefixation**	**ITALIAN VPC**
inicere	'to throw into'	*inbuttare	buttare dentro
eicere	'to throw out'	*sbuttare	buttare fuori
subicere	'to throw under'	*sottobuttare	buttare sotto
deicere	'to throw down'	*debuttare[9]	buttare giù

Table 4. Verbal prefixation and verb-particle formations in Latin and Italian

Thirdly, Italian presents a higher degree of analyticity with respect to Latin. In particular, it developed a quite elaborate set of prepositions that are used very efficiently. Of course, the rise of prepositions in Italian (and the other Romance languages) is related to the loss of the Latin morphological case for the expression of syntactic relations. In this respect, Jansen (2004) remarks that the widespread use of a network of different constructions built around a locative element (which the author calls *particle*) might be seen as a factor that fosters the entrenchment of the scheme for VPCs.

In conclusion, Italian VPCs are an interesting typological and diachronic issue. On the one hand, they do not conform to the typological classification in Talmy (1985,

[7] Apart from *salire* 'to ascend', which is of Indo-European origin and constitutes a good example of lexicalization of the Path into the verbal root, *scendere* 'to descend' and *uscire* 'to exit' depend on the bleaching of the prefix, whereas *entrare* 'to enter' derives from the Latin *intrare* 'to go inside', which is formed from the preposition *intra* 'inside'. This word formation pattern is no longer productive in Italian, therefore, even if *entrare* presents striking similarities with the preposition *entro* 'by', it cannot derive synchronically from the latter.

[8] We will come back to this point in section 5.

[9] Of course the verb *debuttare* exists in Italian, but it is a loan from French meaning 'to debut'.

2000b), thus setting Italian apart from the other major Romance languages. On the other hand, they allow to trace a diachronic opposition between prefixes and particles as different possible realizations of the satellite category. This of course testifies to the interaction between morphological and phrasal constructions, which is in line with and predicted by a constructionist view of language. In section 5 we will expand on this subject. In particular, we will study the interaction between particles and prefixes, or better between VPCs and prefixed verbs, in contemporary Italian. Before passing on to the analysis, we will have a closer look at the formal and semantic properties of Italian VPCs.

3. Italian Verb-particle Constructions

Traditionally, Italian VPCs are quite a neglected topic in Italian linguistics. In recent years, however, they succeeded in catching the attention of the scholars. After the seminal articles by Schwarze (1985) and Simone (1997), a number of studies emerged on the matter (cf. Venier 1996, Antelmi 2002, Jezek 2002, Iacobini 2003, Jansen 2004, Masini 2005, 2006).

However, although the attention on Italian VPCs is relatively recent, the phenomenon is not a recent innovation in this language, as it was already attested in Ancient Italian texts (cf. Jansen 2004, Masini 2005, 2006). Actually, some traces can be detected also in late Latin (e.g. *ire via*, documented by Prisciano, V/VI cent. AD). Vicario (1997), who gives an interesting diachronic account of VPCs in Friulian (a Romance variety spoken in the North-Eastern part of Italy)[10], traces back the phenomenon to the XIV century. The author studies its increasing diffusion until the present days and compares the Friulian situation with the one to be found in Standard Italian.

Nowadays, VPCs seem to be an ever more widespread lexical resource in Standard Italian, especially, though not only, in the spoken language and in less formal texts. Some VPCs have synthetic synonyms, e.g. *entrare – andare dentro* 'to enter', *introdurre/immettere –mettere dentro* 'to put in(side)', while others represent original lexicalizations of certain concepts, like e.g. *restare fuori* 'to stay outside/to be excluded'.

In this paper we will not further discuss about the diachronic origin of these constructions and we will rather concentrate on their place and role in present-day Standard Italian. In the following sections, we will give a brief description of Italian VPCs, in terms of structure (3.1) and semantics (3.2).

3.1. Structure

As Brinton (1988: 163-64) rightly notices, one might classify as VPCs quite different constructions depending on the criteria used, since particles may form more or less cohesive units with the verbal bases. In particular, VPCs are quite similar to combinations of verb plus a prepositional or adverbial phrase. Of course, distinguishing VPCs from similar constructions with prepositions and adverbs is "closely related to the

[10] Cf. Cini (2002) for an account of VPCs in some dialects spoken in the alpine valleys between Piedmont and France. Cf. also the dialectal section in Cini (ed.) (in press)

problem of classifying the particles in respect to part of speech" Brinton (1988: 165). Besides this, within the VPC itself one can recognize a series of different – though closely related – configurations.

In this paper we will refer to the minimal VPC configuration exemplified in (2). This structure consists of a simple (non-pronominal, non-reflexive) verbal base (V), which can be both intransitive (2a) and transitive (2b,c), and a post-verbal modifying particle (P), which corresponds to a locative adverb. The VPC itself may be both intransitive (2a) and transitive (2b,c)[11].

(2) $[\ [\ ]_V\ [\ \ ]_P\ ]_{VPC}$ P = LOCATIVE ADVERB

a. $[[\text{andare}]_V\ [\text{su}]_P\ ]_{VPC}$ lit. go up 'to go up, to ascend'

b. $[[\text{mettere}]_V\ [\text{giù}]_P\ ]_{VPC}$ lit. put down 'to put down'

c. $[[\text{mandare}]_V\ [\text{avanti}]_P\ ]_{VPC}$ lit. send forward 'to run' (e.g. a business)

Besides this minimal configuration, a number of other possibilities can be found. For instance, apart from reflexive forms, one may find different kinds of pronominal verbs in V position, as (3) illustrates (cf. Simone 1997)[12].

(3) a. *far-se-la sotto*
do-reflexive.PRT-pronominal.PRT under
'to quake in one's boots'

b. *ber-ci sopra*
drink-locative.PRT up
'to drink to forget something'

Moreover, the P position may be filled by elements other than simple locative elements. For instance, we can find temporal (e.g. *fare presto* lit. do early 'to hurry up') or manner (e.g. *finire male* lit. finish bad 'to come to a bad end') adverbs.

Finally, some VPCs including a locative adverb obligatorily occur with a prepositional phrase, usually introduced by the preposition *a* 'to' (cf. 4 below).

(4) a. *andare* *dietro* *a* *qualcuno*
go behind to someone
'to follow, to imitate, to like/court'

b. *passare* *sopra* *a* *qualcosa*
pass on to something
'to pass/transit, to forgive, to let something pass'

[11] The VPC does not necessarily maintain the syntactic properties of the verbal base. One of the most notable changes in this sense is the passage from a transitive and/or unergative V to an unaccusative VPC (cf. note 27).

[12] In the examples that follow we will make use of the following abbreviations (in alphabetical order): 3=third person; FUT=future; PART.PAST=past participle; PRT=particle; SG=singular.

The examples in (4) are structurally ambiguous, since they can be interpreted either as VPCs that govern a prepositional phrase, or as VPCs with complex prepositions in P position (here *dietro a* and *sopra a*).

In our analysis we limited ourselves to the minimal configuration in (2), mainly due to two reasons. First, the VPC configuration in (2) is by far the most common. Second, we wanted to carry out the analysis on a coherent corpus in terms of type of verbal bases and particles. In particular, it was important to include in the list of particles only locative adverbs, for reasons that will become clearer.

As regards their syntactic behaviour, Italian VPCs display a particular cohesion that distinguishes them from sequences formed by a verb followed by a prepositional or an adverbial phrase[13].

First of all, it seems that VPCs can be separated only by light non-argumental constituents (such as light adverbs and clitics), as (5) illustrates[14].

(5) a. *Irene ha buttato **via** la bambola*
Irene have.3SG throw.PART.PAST away the doll
'Irene threw the doll away'

b. ??*Irene ha buttato la bambola **via***
Irene have.3SG throw.PART.PAST the doll away

Secondly, Ps cannot be topicalized or left-dislocated with the construction *è... che* 'it is... that'.

(6) a. *Luigi è saltato **fuori** all'improvviso*
Luigi be.3SG jump.PART.PAST out suddenly
'Luigi suddenly popped up'

b. ***Fuori** Luigi è saltato all'improvviso*
Out Luigi be.3SG jump.PART.PAST suddenly

c. **È **fuori** che Luigi è saltato all'improvviso*
be.3SG out that Luigi be.3SG jump.PART.PAST suddenly

Thirdly, in coordinating structures VPCs behave as constituents (7a,b), contrary to verbs followed by a prepositional phrase (7c,d).

[13] For details on the syntactic diagnostics cf. also Simone (1997: 163-166), Antelmi (2002: 101-102) and Masini (2005).

[14] Actually, in the spoken language one may find occasional examples of interposition of the direct object between V and P, like in the following case: *Spero che non **mandino** le pagine **indietro*** 'I hope they won't send the pages back'. Such examples are comparable with object shift phenomena in English. As is known, object shift was a later innovation in English with respect to the rise of post-verbal particles, which were originally more bound to the verb. The choice of the particle position in current English is highly influenced by information structure (cf. Dehé 2002). For an analysis of the phenomenon in Italian cf. Masini (in press).

(7) a. *Max porterà **su** la scacchiera e Yuri ___ i pezzi*
Max bring.FUT up the chessboard and Yuri ___ the pieces
'Max will bring the chessboard and Yuri the pieces'

b. **Max porterà **su** la scacchiera e Yuri **su** i pezzi*
Max bring.FUT up the chessboard and Yuri up the pieces

c. *Max gioca **sulla** scacchiera nuova e Yuri **su** quella vecchia*
Max play.3SG on.the chessboard new and Yuri on that old
'Max plays on the new chessboard and Yuri on the old one'

d. **Max gioca **sulla** scacchiera nuova e Yuri ___ quella vecchia*
Max play.3SG on.the chessboard new and Yuri ___ that old

Of course, there is a gradience in the acceptability of the examples, mostly depending on the transparent vs. opaque semantics of the VPC. However, we can say that, generally speaking, Italian VPCs display a peculiar syntactic behaviour that sets them apart from other free syntactic structures.

3.2. The Semantics of Italian VPCs

From a semantic point of view, we may distinguish two main scenarios, in which the VPCs have either a **locative** or an **idiomatic** meaning.

In the first case, the particle may function as a direction marker, as in (8):

(8)	*andare dentro*	lit. go in	'to enter'
	saltare fuori	lit. jump out	'to jump out, to pop up'

Further, it may strengthen the locative information already present in the verb root, as in (9):

(9)	*entrare dentro*	lit. enter in	'to enter'
	uscire fuori	lit. exit out	'to exit'

Besides these transparent cases, Italian VPCs also display more idiomatic meanings, like those exemplified in (10):

(10)	*buttare giù*	lit. throw down	'to throw down, to write down, to demoralize'
	fare fuori	lit. do out	'to kill'

Examples like those in (10) are to be regarded as individual, non-systematic cases of semantic bleaching. These of course testify to the high degree of establishment of the construction in Italian. However, it should be pointed out that the original and primary function of post-verbal particles remains the indication of locative meanings.
In this paper we aim to demonstrate that Italian VPCs underwent a further semantic development. In particular, we would like to show that some post-verbal particles,

besides maintaining their basic locative function, contribute to the *Aktionsart* of VPCs, adding actional information to the whole construction, like e.g. in (11).

(11) a. *lavare* 'to wash' [±telic] vs. *lavare via* 'to clean off' [+telic]

b. *portare* 'to take' [±telic] vs. *portare appresso* 'to take with oneself' [-telic]

In section 4 below, we will provide evidence for this assumption.

4. In Search of Actional Traces

This section deals with the actional analysis of Italian VPCs. First, we will outline the view of *Aktionsart* adopted in the analysis. Second, we will pass on to a brief description of the corpus and, finally, to the illustration of the results of the analysis.

4.1. Aktionsart, Aspect and Telicity: Assumptions and Methodology

It is well-known that aspectuality is a hotly debated domain. In fact, the numerous current theoretical proposals are quite conflicting with each other: they radically differ with respect to both basic issues and terminological choices[15]. Our study of the actional properties of Italian VPCs does not want to add to this theoretical debate. Rather, it is meant as an empirical contribution. The main goal is to show that, with respect to verbal bases, VPCs display not only different locative values, but also different actional properties. This notwithstanding, we will spend some words to spell out the conceptual framework and the criteria adopted in the analysis (largely based on Bertinetto 1986, 1997), as well as our terminological choices, for the sake of explicitness and comprehension:

- Bidimensional approach: the bidimensional approach implies a strict distinction between aspect and *Aktionsart*. Generally speaking, **aspect** is a matter of viewpoint distinctions (of the perfective/imperfective type) on an event on behalf of the speaker. The latter may choose to portray an event as ongoing (imperfective aspect) or completed (perfective aspect). Structurally, aspect is normally expressed morphologically by means of verbal inflection. To the contrary, ***Aktionsart*** is essentially rooted in the lexical semantics of verbs and concerns the intrinsic temporal nature of the event according to a limited number of relevant binary features: *telic* vs. *atelic, durative* vs. *non-durative, static* vs. *dynamic* events. In sum, whereas *Aktionsart* expresses inherent characteristics, the chief task of aspect is to outline the contextual reclassification of the event (e.g. the expression of the attainment of the goal in telic events). However, even though aspect and *Aktionsart* are independent and express distinct semantic characteristics, their intersection nevertheless contributes to determine the general aspectual properties of a sentence.

[15] For a comprehensive review cf. Sasse (2002), who points out that the conflicting positions on almost any of the basic issues have as a consequence a tremendous gap between descriptive and theoretical works.

- Non-holistic representation of the event: the representation of the event is set up in terms of Vendler's time-schemata. The classification adopted here, which is the refinement of Vendler's (1967) classification put forward by Bertinetto (1986), lays on underlying binary semantic features such as [±durative], [±telic], [±dynamic], and distinguishes five classes[16] (cf. Table 5).

Vendler (1967)	Bertinetto (1986)	**durative**	**telic**	**dynamic**
ACCOMPLISHMENT	RISULTATIVO	+	+	+
ACTIVITY	CONTINUATIVO	+	-	+
ACHIEVEMENT	TRASFORMATIVO	-	+	+
ACHIEVEMENT	PUNTUALE	-	-	-
STATES	STATIVO	+	-	-

Table 5. Actional classifications by Vendler (1967) and Bertinetto (1986)

- Diagnostic tests: in order to assign each verb to one of the five classes above mentioned, we used diagnostics tests based on both the compatibility with different kinds of adverbials (e.g., *in X time* or *for X time*), and on semantic compatibility, such as the ones worked out by Klein (1969)[17].

- Terminology: from a terminological point of view, and in accordance with the bidimensional approach, we distinguish between *aspect* and *Aktionsart*. In order to make reference to the aspectual domain in general (aspect and *Aktionsart* together) we use *aspectual* and *aspectuality*, whereas *actional* refers to *Aktionsart* only. Finally, we chose *event* as a cover term to encompass both dynamic and static delimitations in the aspectual domain (other authors use *state-of-affair*, *situation*, etc.).

In what follows, we will add some considerations about the compositional nature of *Aktionsart*, and in particular about the telicity feature, which is especially relevant for the discussion that follows.

[16] The choice to adopt Bertinetto's classification was due to two basic reasons. First, it was elaborated on Italian data. Second, it proposes to split the Achievement class on the basis of the telicity (and dynamicity) feature, which is the most relevant in our analysis. Starting from the three binary oppositions, other combinatorial possibilities are allowed. For example, Smith (1991) identifies a class with the features [–durative] [–telic] [+dynamic], while other scholars advocate for the adoption of other, new features. We agree with Boogaart's (2004: 1169) view that "[e]xactly how many, and which, *Aktionsart* classes one distinguishes is determined partly by the particular language one is dealing with, as well as by the specific linguistic phenomenon one is investigating".

[17] For a detailed survey of aspectual tests cf. Dowty (1979). An enlightening critical analysis of these tests can be found in Behrens (1998: 289-302).

4.1.1. A Compositional View of Aktionsart

In current aspectual studies there is a general agreement that actional classifications do not exclusively depend on verb semantics. Although there might be some differences regarding the mechanisms of interrelation between the actional value of the verb lexeme and the other elements in the verbal phrase, the scholars agree in saying that *Aktionsart* results from many interacting factors at both the lexical and the clausal level[18]. Of course, this picture implies some difficulties at a descriptive and theoretical level, especially with respect to the reciprocal role among the elements contributing to the compositional aspectual configuration. Incidentally, the twilight of the traditional notion of *Aktionsart* as a purely lexical property of verbs and the success of the compositional view among contemporary approaches to aspectuality lead to another important theoretical consequence. Indeed, the compositionality of *Aktionsart* undermines the strict distinction between lexicon and phrasal units, and therefore between the lexical/morphological level and the syntactic level.

A typical example of recategorization of the actional value of a verb lexeme is the passage from Activity to Accomplishment by the addition of a nominal phrase in direct object position. For instance, while *to draw* is categorized as an Activity, *to draw a circle* is an Accomplishment. In these cases, time-schemata are not expressed by a verb lexeme, but rather by "abstract verb phrases or constructions […] called "terms" by Vendler" (Sasse 2002: 216). In our view, these "terms" could be easily interpreted as abstract *constructions* (in the specific sense of Construction Grammar) in which the actional interpretation is not to be ascribed to the presence of a direct object by itself, nor to the denotative meaning of the word *circle*, but rather to the value of the element in direct object position in terms of features such as [±determinate] and [±singular] (cf. Bertinetto 2001: 182). In this view, the interpretation of a phrase like *to draw a circle* as an Accomplishment does not depend on the presence of the direct object *a circle*, but on the value of the direct object itself, i.e. [+determinate] and [+singular]. This is supported by the fact that a phrase like *draw circles* is an Activity, despite the presence of the direct object. This is due to the properties of *circles*: [-determinate] and [-singular]. In sum, it seems that arguments which express a quantified reference (e.g. count nouns with a specifier) induce a telic reading, while arguments with cumulative reference (e.g. bare plurals and mass nouns) induce a durative reading.

Among the underlying semantic features that concur to determine actional classes, **telicity** is one of the most sensitive to the context of occurrence. For example, quite often the presence of an object may contribute to make the predicate telic by indicating the endpoint of the activity. The very same function can be played by prefixes (e.g. Dutch *schrijven* 'to write' [–telic], *op-schrijven* lit. up-write 'to write down' [+telic], *over-schrijven* lit. over-write 'to copy' [+telic]) or post-verbal particles (e.g. English *to write* [–telic], *to write down* [+telic], *to write up* [+telic])[19]. As observed by Boogaart (2004: 1172), "none of the prefixes or particles mentioned marks telicity *per se*", since *Aktionsart* is a property that refers to the whole construction and therefore

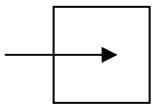
[18] In some cases, it is not only the verbal phrase that is involved, but the whole argument structure of the verb. For example, in constructions with unaccusative verbs, also the subject may affect the *Aktionsart*. Therefore, at least in these cases, *Aktionsart* is a property to be assigned at a clause rather than at a phrase level.

[19] Examples taken from Boogaart (2004: 1172).

results from the interaction between verb and prefix/particle. On the other hand, as we will see, the meaning of the particle plays an important role.

Following Brinton (1988: 26), we define telic a situation which has a necessary endpoint, "which necessarily includes a goal, aim, or conclusion. The goal is an inherent part of the situation". In this perspective, the semantic feature that may contribute to the telic reading of the verb is the indication of the endpoint of an event, which can be easily conveyed by locative particles. Actually, the locative particles that indicate movement oriented towards a specific goal may come to imply attainment of the goal (telic events), whereas particles that express stasis, location or a movement without a specific endpoint contribute to indicate atelic events (cf. Figure 1)[20].

andare dentro lit. go in 'to enter'



andare attorno lit. go around 'to wander about'

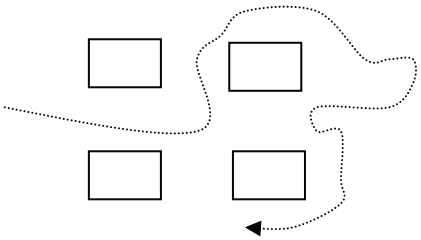


Figure 1. From locative to aspectual meanings

The actional values expressed by particles might be explained in terms of metonymic extension (cf. Brinton 1988: 191-199). Differently from interpretations based on metaphorical semantic shift, the metonymic explanation accounts for the simultaneous presence of locative and actional meanings in one and the same particle or VPC. The actional change is motivated by an iconic principle, i.e. a structural analogy between two similarly structured and conceptually related domains: spatial movement and event structure[21]. In this sense, we may speak of *tendentially telic particles* and *tendentially atelic particles*, as a consequence of their bounded or unbounded spatial meaning.

In the following pages we will present the results of our analysis on Italian VPCs. As we will see, the expectations about the parallels between the semantics of particles and the kind of telicity changes involved will be matched.

[20] This is at the basis of the distinction between *goal* and *direction*. For instance, an expression such as *moving towards the North Pole* indicates a direction, but not a necessary goal (cf. Brinton 1988: 26). Brinton also notes that Declerck (1977: 320) argues for a contrast between 'goal' expressions (e.g. *walk into the house*) and 'directional' (e.g. *walk toward the house*) or 'locative' expressions (*be in the house*); therefore, it is "crucial to distinguish purely directional expressions, which are Activities, from goal expressions, which are Accomplishments" (Brinton 1988: 278).

[21] In this respect cf. also Talmy (2000b: 231), that identifies a conceptual correlation between motion events and other kinds of events such as the temporal one: "This conceptual analogy motivates a syntactic and lexical analogy: to a great extent in a language, aspect is expressed in the same constituent type as Path (+Ground), and often by homophonous forms. Thus, in accordance with the general typology, the core schema of an event of temporal contouring appears in the main verb in verb-framed languages, while it appears in the satellite in satellite-framed languages".

4.2. The Corpus

The corpus on which we carried out the analysis consists in a group of 165 VPCs listed in two major Italian dictionaries, i.e. GRADIT and DISC. We chose to base our observations on a dictionary corpus because we wanted to take into account only acknowledged items. In fact, despite the recent interest of scholars, Italian VPCs are still a rather neglected topic in lexicographical practice, since Italian dictionaries, apart from some notable exceptions, are not very ready to register multi-word expressions. In this sense, those that are actually listed in the dictionaries are likely to be among the most established in current use.

The verbal bases in the corpus amount to 54. Some of them occur with only one particle (e.g. *entrare* 'to enter'), while others combine with ten or even more particles (e.g. *andare* 'to go', *mettere* 'to put'). Most verbal bases are verbs of motion or location, though not all of them can be ascribed to this macro-class (cf. section 4.3 for more details). The particles involved, which correspond to the requirements mentioned in section 3.1, amount to 19. A complete list is provided in (12).

(12) *accanto* 'next to', *addosso* 'on', *appresso* 'nearby', *attorno* 'around', *avanti* 'forward', *contro* 'against', *dentro* 'in(side)', *dietro* 'behind', *fuori* 'out(side)', *giù* 'down', *indietro* 'back(wards)', *intorno* 'around', *lontano* 'far away', *oltre* 'beyond', *sopra* 'on', *sotto* 'under', *su* 'up', *via* 'away', *vicino* 'near'

The whole corpus was analysed according to the basic assumptions outlined in the previous section. In what follows one may find the results of our investigation.

4.3. Results

The main results of the analysis are reported in Table 6, which illustrates the percentages concerning the kinds of telicity shifts that occur in the passage from verbal bases to VPCs. The Table also details the numbers of VPCs involved in each kind of shift according to the semantic class of the verbal base.

Following the typology of motion events in Talmy (1985, 2000b), we distinguished between verbs expressing Location (BE_L) (e.g. *essere* 'to be', *stare* 'to stay') and verbs expressing Motion. The latter are further divided into three subgroups, all of whom contain both Non-Agentive (NA) and Agentive (A) verbs[22]:

- Path verbs (e.g. *entrare* 'to enter', *uscire* 'to exit'): these form the most homogeneous class; only a limited number of Italian Path verbs are involved in VPCs and, as we will see, most of them combine with only one particle (which strengthens the meaning of the verbal root);

- Manner/Cause verbs (e.g. *correre* 'to run', *saltare* 'to jump'): Manner verbs are a more composite class; apart from Non-Agentive motion verbs like *correre* 'to run' or *volare* 'to fly', it includes two sub-groups of Agentive verbs, i.e. verbs of

[22] Talmy (2000a) defines agentivity in terms of causation and intentionality. With respect to motion events, this means that something/someone causes the Figure to move. In actual fact, in most cases the Agentive/Non-Agentive distinction corresponds to the transitive/intransitive distinction.

throwing (e.g. *gettare* 'to throw', *buttare* 'to throw') and verbs of removing (e.g. *tagliare* 'to cut', *grattare* 'to scrape');

- Generic verbs (e.g. *andare* 'to go', *mettere* 'to put')[23]: this class presents only two Non-Agentive verbs (*andare* 'to go' and *venire* 'to come'), which anyway combine with quite a number of particles, and a series of Agentive verbs including, among the others, verbs of putting (e.g. *mettere* 'to put', *porre* 'to put'; cf. also footnote 23), and verbs of sending and carrying (e.g. *mandare* 'to send', *portare* 'to bring/take').

In addition to Motion and Location verbs, one may find verbs belonging to the category Other. This includes non-motion verbs of various kinds and is therefore the most heterogeneous class.

In section 4.3.1 we will discuss the telicity changes in VPCs in more detail. In section 4.3.2 we will focus on the relationship between telicity changes and semantic classes.

V → VPC	***%***		*Motion and Location verbs*				***Other***	***Total***	
			Path	***Manner***	***Generic***	***BE_L***			
+TEL → +TEL	42.8%	**no changes 71.1%**	10	12	30	-	13	*65*	*108*
–TEL → –TEL	28.3%		1	2	1	26	13	*43*	
–TEL → +TEL	11.2%	**telicization 21%**	4	8	3	-	2	*17*	*32*
±TEL → +TEL	9.8%		-	2	13	-	-	*15*	
±TEL → –TEL	6.6%	**detelicization 7.9%**	-	-	5	-	5	*10*	*12*
+TEL → –TEL	1.3%		-	-	-	-	2	*2*	
Total	*100%*		*15*	*24*	*52*	*26*	*35*	*152*	

Table 6. Telicity changes

4.3.1. Telicity Changes from V to VPCs

The first column in Table 6 illustrates the telicity changes in the passage from the verbal base to the VPC. We assigned three values to both verbal bases and VPCs: +TEL, -TEL, ±TEL. The presence of the latter value is in line with the "aspectual multivalence" proposed by Brinton (1988: 31), i.e. "the ability of a single lexical verb to name different situation types depending upon the structures with which it combines"[24]. In

[23] Since Talmy (1985, 2000b) focuses on the kinds of lexicalization patterns for verbal roots, he does not explicitly speak of Generic verbs. However, this category is quite implicit in his work. Indeed, he speaks of "generic verbs" with reference to the English verbs *to put* and *to go* (cf. Talmy 2000b: 284) and defines the English verbs *to put* and *to take* as "suppletive forms of a single more general and non-directional 'putting' notion, where the specific form that is to appear at the surface is determined completely by the particular Path particle and/or preposition present" (Talmy 1985: 71).

[24] Cf. also Bertinetto (2001: 182): "[...] most predicates may have more than one actional classification".

order to make the results clearer, we excluded the cases in which the VPC was a ±TEL item (which is why the total number of VPCs in Table 6 is 152 instead of 165). The data missing, however, would not change the overall picture. Given this, we could identify three main types of telicity changes:

- *absence of telicity change*
- *telicization*
- *detelicization*

As one may notice from the data in the percentage column, in the great majority of cases **telicity does not change** (71.1%). Over 40% of VPCs (out of the total) have telic bases that remain telic, whereas atelic bases that remain atelic are almost 30%.

Within the +TEL→+TEL group most verbal bases are Motion verbs (Path/Manner/Generic). Here we can distinguish two main functions of the post-verbal particles with respect to the semantics of the verbal base, namely:

- *directional marking*: mainly with Manner and Generic verbs, e.g. *buttare via* 'to throw away', *mettere su* 'to put on';
- *explicitation of the telos*: with Path verbs, in which the directional information is already encoded in the verbal base (e.g. *salire su* lit. ascend up 'to ascend', *scendere giù* lit. descend down 'to descend').

This last strengthening operation might be due to the opacity of the base (which is no longer morphologically analysable) or to some communicative need of expliciteness[25].
As easily expected, in the -TEL→-TEL group almost all verbal bases belong to the BE_L group, which tipically contains stative verbs. Here the particles do not affect telicity, rather their main function is to specify the location of the event, like in *essere via* 'to be away, out of town'[26].

In the light of these first data, we might be induced to say that, generally speaking, the presence of particles does not affect the verbal bases in a systematic way, since the vast majority of VPCs display no telicity changes. However, there are also quite a number of verbs that do change their telicity. Here we have both telicizing and detelicizing cases.

As for the **telicization** cases (21%), the verbal bases mainly belong to Manner and Generic motion verbs. This seems to suggest that the particle here may function both as a direction (or Path) marker and as a telos indicator. See e.g. the VPCs in (13)[27]:

---

[25] For similar considerations, see Traugott (1982: 252), who suggests that the particle serves to make a "covert" endpoint "overt", and Lindner (1983: 169 ff.), who says that the particle serves to "profile" the goal. Antelmi (2002: 107, footnote 14) speaks of "rideterminazione [*redetermination*]".

[26] Further, many of these verbs have metaphorical meanings, e.g. *essere giù* lit. be down 'to be depressed'.

[27] A quite convincing syntactic clue of this telicization process is the fact that some verbs, after turning into VPCs, become unaccusative (cf. the example below). In fact, many scholars argue for a connection between unaccusativity and telicity.

(i) *volare* (intransitive, aux. *avere* 'have') → *volare via* (intransitive, aux. *essere* 'be')
  a. L'uccello ha volato per due ore 'The bird flied for two hours' (-TEL)

(13) *tirare* 'to pull' (Generic, Agentive) → *tirare fuori* 'to take out'
andare 'to go' (Generic, Non-Agentive) → *andare via* 'to go away'
sbattere 'to dash/throw' (Manner, Agentive) → *sbattere fuori* 'to throw out'
saltare 'to jump' (Manner, Non-Agentive) → *saltare giù* 'to jump down'

Almost all of these VPCs contain *telic particles* (cf. section 4.1.1), i.e. particles that inherently refer to a specific spatial endpoint and thus contribute to the overall telic meaning of the VPC. Here follow the most represented *telic particles* in our corpus (in decreasing order): *via* 'away', *dentro* 'in(side)', *fuori* 'out(side)', *su* 'up'. This seems to confirm the expectations outlined in 4.1.1: telicization cases do actually present *telic particles*.

Finally, we may find also few **detelicization** cases (7.9%). Interestingly enough, most of the verbs involved in this process are non-motion verbs (they are included in the class Other) and have metaphorical or idiomatic meanings: *crescere dentro* 'to grow as a person'*, dare giù* 'to beat'*, ridare fuori* 'to vomit'. However, there are also few cases with Generic verbal bases. Here, the particles used are actually of the *atelic* type (cf. section 4.1.1), i.e. particles that denote a direction without specifying any endpoint, e.g. *addosso* 'on', *appresso* 'nearby', *attorno* 'around'. However, the examples of this kind are too few to draw any serious generalization regarding the interaction between the detelicizing process and the type of particles involved[28].

4.3.2. Telicity Changes and Semantic Classes

In the previous section, we described the types of telicity changes that can be observed in our corpus and the overall role and presence of the different semantic classes in these changes. Now we will focus our attention on the telicity changes occurring within each single class, in order to understand better the role of the semantics of verbal bases in these changes. Table 7 shows the percentages of changes within each semantic class.

First of all, as the percentages clearly show, most **Path** verbs are telic and remain telic after they combine with the particle. As we already noticed in the previous section, in these cases the particle has the function to make the telos overt (e.g. *fuggire via* lit. escape away 'to escape'). There is also a significant percentage (26.7%) of telicization cases. However, this percentage is overrated, since it regards four VPCs with the same verbal base, i.e. *passare* 'to pass' (e.g. *passare via* 'to fade away').

---

b. L'uccello è volato via 'The bird flied away' (+TEL)
Our corpus displays a number of cases like the one in (i).

[28] Here we will limit ourselves to note that the supposed atelic character of some particles seems to hold also with non-motion verbs. Consider for example the following set of VPCs formed with the base *guardare* 'to look, see, watch': *guardare avanti* 'to look forward', *guardare indietro* 'to look backwards', *guardare lontano* 'to foresee'. Here we have a verbal base (*guardare*) that can express both an Activity (e.g. *guardare la TV* 'to watch TV') and an Accomplishment when accompanied by a specific, bound object (e.g. *guardare un film* 'to watch a movie'). Since the particles *avanti/indietro/lontano* only denote a direction, and not an endpoint, they cannot be interpreted as a specific, bound object to look at and, consequently, cannot stress the potential telicity of the verbal base.

Telicity changes ***V → VPC***		*Motion and Location verbs*				***Other***
		Path	***Manner***	***Generic***	***BE$_L$***	
no changes	+TEL → +TEL	**66.6%**	**50%**	**57.7%**	-	37.2%
	–TEL → –TEL	6.7%	8.4%	1.9%	**100%**	37.2%
telicization	–TEL/±TEL → +TEL	26.7%	**41.6%**	**30.8%**	-	5.7%
detelicization	+TEL/±TEL → –TEL	-	-	9.6%	-	19.9%
Total		*100%*	*100%*	*100%*	*100%*	*100%*

Table 7. Semantic classes and telicity changes

Manner verbs are more interesting: in percentage terms, they telicize more than others. Indeed, Manner verbs constitute an optimal base for the creation of new VPCs, as they denote events that might require the specification of a Path (e.g. *buttare fuori* 'to throw out') and/or endpoint (e.g. *grattare via* 'to scrape away'). In some cases particles may have a double function (direction markers and telic markers together), such as e.g. in *volare via* 'to fly away'*, correre via* 'to run away'*, saltare fuori* 'to jump out'.

In order to confirm the hypothesis that Manner verbs are especially prone to combine with post-verbal particles, and since VPCs originating from Manner verbs are highly underrepresented in our corpus, we carried out an informal Google search. Here follow some examples we found that are not included in the corpus.

(14) a. <u>Non-Agentive</u>

gocciolare giù	'to drip down'	("*..il suo sangue sarebbe* ***gocciolato giù*** *fino a sporcare le tende del panificio...")*
sgorgare fuori	'to gush out'	("*Con le lacrime che colmavano tutta la parte inferiore dei miei occhi in attesa di* ***sgorgare fuori*** *a mo' di fontana")*
scivolare via	'to slide off'	(*"****Scivolò via*** *nel corridoio e scomparve")*
rotolare giù	'to roll down'	(*"Mentre procedeva in bicicletta sull'argine del fiume Runco, perdeva il controllo e* ***rotolava giù****")*
strisciare via	'to crawl away'	(*"..riuscì a* ***strisciare via*** *di soppiatto e a salvarsi")*

b. <u>Agentive</u>

trascinare giù	'to drag down'	(*"Letizia era svenuta e l'ho* ***trascinata giù****")*;
trascinare via	'to pull away by dragging'	(*"L'ho* ***trascinata via*** *per i pantaloni")*
spostare via	'to shift away'	(*"sarà mica mio padre che si è messo a remare per* ***spostare via*** *la Sicilia per non farmi tornare sul continente")*

Further comes the class of **Generic** verbs of motion, which includes all verbs of motion that do not lexicalize the Path, nor any *co-event* such as Manner or Cause (cf. footnotes 4 and 23). Particles seem to be especially prone to combine with this class. Indeed, it contains some of the verbal bases that combine with the highest number of particles, i.e. *andare* 'to go', *venire* 'to come', *mettere* 'to put', *portare* 'to take/bring'. As with Manner verbs, particles function either as pure direction markers when the verbal base is telic (e.g. *porre giù* 'to put down'), or as a direction and/or telicity marker when the base is ±TEL (e.g. *andare su* 'to go up(wards), *portare via* 'to take away').

As already mentioned above, Location verbs (**BE$_L$**) appear to be rather insensitive to the presence of particles. This is of course due to the fact that Location verbs are stative and therefore cannot delineate a process. Within this group, all outputs are atelic VPCs, which often have metaphorical meanings, e.g. *essere giù* 'to be depressed' and *stare/essere fuori* 'to be mad'.

Finally, the corpus displays a rather conspicuous number of VPCs with bases belonging to the **Other** class, i.e. to non-motion verbs. The combination of particles with non-motion verbs implies that the construction extended beyond the domain of spatiality, and therefore testifies to the productivity and pervasiveness of the construction in present-day Italian. However, it appears to be difficult to identify any noticeable regularity within this heterogeneous group, apart from the fact that, like for other classes, most examples do not display any telicity changes and that, as already noticed, many have non-literal meanings, e.g. *mangiare fuori* 'to have a meal out', *vedere lontano* 'to foresee'. They also represent the majority of the detelicizing cases.

In the following section we will try to draw some generalizations from the results presented here.

4.4. The Moral of the Story

The most important fact that emerged from the above investigation seems to be that most VPCs do not change telicity with respect to their verbal bases. This leads us to think that Italian particles, though clearly playing a role in determining the *Aktionsart* of the VPC (with particular reference to the telicity feature) are **not** proper aspectual markers, or rather not yet[29].

However, if we exclude the unvaried telicity group, we have quite a number of telicity changes, most of which figure a passage towards telicity. Telicization cases depend largely on *telic particles*, which add an endpoint to the event. To a much lesser extent, also *atelic particles* seem to play a role in detelicization cases. Therefore, we can identify two ways in which particles may contribute to the *Aktionsart* of the verbal bases (cf. Table 8). The main actional effect is the indication of telicity performed by *telic particles* such as *fuori* 'out', *giù* 'down', *su* 'up', *via* 'away/off'. In a minority of cases, and in a rather unsystematic way, *atelic particles* may also contribute to convey atelicity (e.g. *appresso* 'along/nearby', *indietro* 'backwards', *attorno* 'around').

[29] However, there seem to be some traces of subregularities. For instance, all verbs of removing (such as *tagliare, grattare, strappare, lavare, raschiare*) may combine with the particle *via* 'away' (cf. also Masini 2005), thus making the event telic.

Telicization	Generic	*tirare* 'to pull'	-TEL	*tirare* ***fuori*** 'to pull out'	+TEL
		tirare 'to pull'	-TEL	*tirare* ***giù*** 'to pull down'	+TEL
	Manner	*saltare* 'to jump'	-TEL	*saltare* ***dentro*** 'to jump in'	+TEL
		volare 'to fly'	-TEL	*volare* ***via*** 'to fly away'	+TEL
	Path	*passare* 'to pass'	-TEL	*passare* ***via*** 'to fade away'	+TEL
		passare 'to pass'	-TEL	*passare* ***su*** 'to drop by'	+TEL
Detelicization	Generic	*portare* 'to bring'	±TEL	*portare* ***appresso*** 'to take with one'	-TEL
	Generic	*andare* 'to go'	-TEL	*andare* ***attorno*** 'to wander about'	-TEL
	Manner	*correre* 'to run'	-TEL	*correre* ***indietro*** 'to come back by running'	-TEL

Table 8. Actional contribution of particles

A second relevant generalization can be drawn from the data presented in section 4.3.2. It is quite clear that the Italian particle system seems to be particularly productive with Manner and Generic verbs of motion (both Agentive and Non-Agentive). Also changes in telicity are mostly connected with these two classes (as well as with non-motion verbal bases). Of course, particles also occur with Path and BE_L verbs, though telicity changes are less frequent (or even absent) here.

When added to telic bases, particles either denote direction (with non-Path verbs), or make the telos explicit (with Path verbs). In these cases, particles generally do not affect the telicity of the verbal base. When they combine with atelic bases, they may either indicate the endpoint of the event, or function as direction markers.

The fact that Italian VPCs mostly involve Manner and Generic verbs of motion (as well as BE_L verbs) stresses the primary locative function of particles. Of course, we also have a lot of metaphorical examples, but the original spatial motivation behind the formation of VPCs is still totally productive (cf. section 2.2). We will come back to this issue in the next section, in which we will deal with the comparison between particles and prefixes in Italian.

5. VPCs and Verbal Prefixation

It is well known that both post-verbal particles and verbal prefixes may express locative meanings and may function as satellites in motion events. Here we will deal with the relationship between particles and prefixes in Italian. In particular, we would like to answer the following question: what kind of interaction does it exist between VPCs and prefixed verbs in contemporary Italian from a semantic point of view? Do they cooperate or compete?

Let us start from some diachronic considerations. As already mentioned in section 2.2, Italian verbal prefixes with locative meanings are rather limited in number compared to Latin. Table 9 puts side by side the rich system of verbal prefixes with

locative meanings in Latin (exemplified by the derivatives of the verbal bases *duco* 'to pull' and *mitto* 'to send') and the Italian prefixed derivatives of the verb *portare* 'to bring/take'.

Prefixed		**Latin**		**Italian**
		duco 'to pull'	***mitto*** 'to send'	***portare*** 'to bring/take' (XII c.)
ab-	'away'	*abduco*	*amitto*	*asportare* 'to remove' (XIV c.)
ad-	'to, toward'	*adduco*	*admitto*	*apportare* 'to produce' (XIII c.)
ante-	'ahead, forward'	-	*antemitto*	-
circum-	'around, on all sides'	*circumduco*	*circummitto*	-
de-	'from, down'	*deduco*	*demitto*	*deportare* 'to deport' (XIV c.)
dis-	'apart'	*disduco*	-	-
ex-	'out'	*educo*	*emitto*	*esportare* 'to export' (XV c.)
in-	'in, on, against'	*induco*	*immitto*	*importare* 'to import' 'to be important' (XIV c.)
inter-	'between'	-	*intermitto*	-
intro-	'internally'	*introduco*	*intromitto*	-
ob-	'toward, against'	*obduco*	*omitto*	-
per-	'through, thoroughly'	*perduco*	*permitto*	-
prae-	'before'	-	*praemitto*	-
praeter-	'beyond'	*praeterduco*	*praetermitto*	-
pro-	'in front of, forth'	*produco*	*promitto*	-
retro-	'backwards'	*retroduco*	-	-
se-	'aside'	*seduco*	-	-
sub-	'under'	*subduco*	*submitto*	*sopportare* 'to tolerate' (XIII c.)
super-	'above'	*superduco*	-	
trans-	'across'	*traduco*	*transmitto*	*trasportare* 'to transport' (XIV c.)

Table 9. Latin and Italian verbs with locative prefixes

On the one hand, it is interesting to see that the number of empty cells in the *portare* column is rather high. This testifies to the unproductivity of such prefixes. At the same time it is interesting to point out that most unattested prefixed verbs could be interpreted as semantically transparent possible words. Moreover, all derivatives from *portare* are of Latin origin and their first recordings in Italian date back to several centuries ago, which is also testified by their largely non compositional meanings. Finally, none of the prefixes in combination with *portare* are productively used in preverbal position in the Italian language.

On the other hand, it is worth noting that the verb *portare* is involved in a series of VPCs (e.g., from our corpus, *portare addosso* 'to wear', *portare appresso* 'to take with one', *portare avanti* 'to further/bring upfront', *portare giù* 'to bring down(stairs)', *portare indietro* 'to bring back', *portare sopra* 'to put up', *portare sotto* 'to put down',

portare su 'to bring up(stairs)', *portare via* 'to take away'), that incidentally fill some of the empty spaces in Table 9.

The decline of both the number of verbal prefixes and their possible meanings in the passage from Latin to Romance languages (and until present days; cf. Lüdtke 1996) has been in part overshadowed by the high frequency of many prefixed verbs of Latin origin that play an important role in the basic Italian vocabulary. However, in a recent study on the productivity of verbal prefixes in Italian, Iacobini (2005) shows that about the 70% of existing prefixed verbs that are attested for the first time in the XX century are derived by means of only four prefixes, i.e. *ri-, de-, dis-* and *s-*. These convey iterative (*ri-*), opposite (*dis-*), privative and reversative (*de-, dis-, s-*) meanings. What's more, Iacobini (2005) shows that only the 8% of prefixed verbs coined in the XX century are formed by prefixes that express locative meanings.

Table 10 lists the locative meanings that can be conveyed by verbal prefixes and by the nineteen particles of our corpus when used in combination with verbs of motion.

Prefixes and prefixed verbs	**Locative meanings**	**Particles and VPCs**
ante-, pre-, pro- *anteporre, premettere, progredire*	ANTERIOR, BEFORE	avanti *andare a.*
retro-, re-/ri- (?) *retrocedere, rifluire*	BACK	indietro *andare i.*
	BEHIND	appresso, dietro *andare a./d.*
contra-/contro-, ob- *contrapporre, occludere*	OPPOSITE SIDE, AGAINST	addosso, contro *andare a./c.*
giusta- *giustapporre*	NEAR	accanto, vicino *andare a. /v.*
	FAR	lontano *andare l.*
fra-, infra-, inter-, intro-, tra- *frammischiare, inframmettere, interporre, introdurre, trascegliere*	BETWEEN, INWARDS	dentro *andare d.*
ab-, de-(?), dis-, e-/es-, estra-, estro-(?), s-, se- *abdurre, deportare, disperdere, emergere, espatriare, estrapolare, estromettere, sbarcare, separare*	OUTSIDE, AWAY	fuori, via *andare f./v.*
sopra-/sovra-, sor- *sopraelevare, sovrapporre, sorpassare*	ON, ABOVE, UP, OVER	sopra, su *andare s./s.*
sotto-(?) *sottoscrivere*	UNDER, BELOW	sotto, giù *andare s./g.*
per-, trans- *trasferire, perforare*	ACROSS, BEYOND	oltre *andare o.*
circum-/circom- *circumnavigare, circondare*	AROUND	intorno, attorno *andare i./a.*

Table 10. Locative meanings expressed by verbal prefixes and post-verbal particles[30]

The first observation that can be drawn from Table 10 is that all the locative meanings that can be expressed by verbal prefixes can also be expressed by post-verbal particles. On the contrary, post-verbal particles can convey some additional meanings

[30] Table 10 distinguishes between productive and unproductive prefixes: underscored prefixes are the productive ones. Question marks (?) indicate doubtfully productive prefixes.

with respect to prefixes, i.e. 'behind' and 'far'. Moreover, half of the meanings (those in grey cells) are no longer productively expressed by verbal prefixation, and only approximately one third of prefixes (signaled through underscore) currently used in Italian complex verbs can be employed in productive word formation processes. Further, for some prefixes the locative meaning is not the only – and in some cases not even the main – value: e.g. *sopra-/sovra-* and *sotto-* may also express evaluation, *re-/ri-* almost exclusively convey iteration, and finally *de-, dis-, s-* are mainly employed with privative and reversative meanings.

As regards the comparison between prefixes and particles, we will focus on three observations. First of all, many Italian motion verbs already present a large family of prefixed verbs of Latin origin (e.g. the *mettere* 'to put' family, that counts, among others: *ammettere* 'to admit'*, immettere* 'to put in'*, sottomettere* 'to subdue', etc.). This represents a strong **restriction on available bases** for prefixes, since (differently from what happened in Late Latin) it is exceptional for Italian to have two spatial prefixes on the same verb. Instead, post-verbal particles are much freer to combine with prefixed verbs.

Secondly, post-verbal particles are preferred to the few left productive prefixes because they constitute a more articulated and **systematic** way of expressing spatial indications.

Thirdly, there seems to be a diaphasic difference between prefixed verbs and VPCs. In fact, many prefixed verbs developed a non-literal meaning (that became the most common in use), while the original locative meaning came to be confined to technical or bureaucratic registers. For instance, the verb *introdurre* is commonly used in the meaning 'to bring something into use for the first time'. Of course, it could also be employed with the original meaning 'to insert', but this is stylistically marked. Surely, nobody would use *introdurre* to say 'to put the car into the garage'. Rather, anyone would make use of the VPC *mettere dentro* 'to put inside'.

To sum up, nowadays particles fill most of the gaps left by verbal prefixes with respect to the expression of locative meanings. This was made possible thanks to both the range of meanings they can convey (as well as the systematic way in which they are conveyed), and to their less unmarked register.

Since now, we discussed the relationship between verbal prefixes and post-verbal particles with reference to the expression of locative meanings. But what about the expression of *Aktionsart*? It seems to us that, in this respect, the relationship is even clearer: verbal prefixes do not constitute a means for actional marking[31].

Diachronically, the rich and complex system of prefixes used in Early and Classical Latin to render verbs telic broke down already in the Latin language, in the very first centuries of the Christian era[32]. This can be taken as a crucial difference

[31] Ingressive change of state is expressed in Italian by parasynthetic verb formation (cf. Iacobini 2004). It is important to note that the two prefixes that take part to this process (*ad-* and *in-*) cannot be preposed to verbs, and therefore cannot be considered preverbal prefixes. Further, they have lost their original locative meaning from which they developed the aspectual one. Egressive meaning may be expressed by the prefixes *de-, dis-, s-* (which are normally used with privative and reversative values) through a reinterpretation of their ablative meaning. Some linguists include the iterative meaning among the aspectual ones. In Italian, this meaning can be productively expressed by the preverbal prefix *ri-*, and not by VPCs.

[32] Classical Latin did not employ aspectual particles: preverbs were joined to verbs as prefixes (cf. Vincent 1999). Haverling (2003) studies the role of prefixes in *Aktionsart* changes in the Latin verbal

between the development of actional values in Italian VPCs and what happened in Germanic languages. In Germanic languages there was a long period of overlapping (and thus competition) between the fading system of native prefixes (that expressed both locative and aspectual meanings), and the new particle system, which, starting from locative meanings in combination with motion verbs, also came to express *Aktionsart*, even with non-motion verbs (cf. Brinton 1988, Hiltunen 1983 for the English language). A significant temporal gap separates the emergence of VPCs in Italian from the collapse of Latin aspectual prefixes. As a consequence, Italian VPCs cannot be considered a way to replace prefixes in the expression of *Aktionsart*. Rather, the overt (though not – or not yet – consistent) expression of *Aktionsart* by means of VPCs, which develops quite naturally from a reinterpretation of the locative meanings of particles, is an Italian innovative feature. This lead us to think that the establishment of VPCs in Italian depends on a general typological restructuring of the Italian language that implies a greater degree of analyticity and a tendency towards post-modification.

In conclusion, we can say that there is no competition between VPCs and prefixed verbs in contemporary Italian. As regards the expression of locative meanings, VPCs, which form a coherent system of spatial indication, compete with prefixed verbs, but not with verbal prefixation, which is mainly used to express iterative and negative/privative meanings. Therefore, VPCs and verbal prefixation perform different functions within the language. What's more, the actional analysis carried out above showed that post-verbal particles are on the right track to become actional markers, thereby displaying a dynamism that was not to be found among Italian verbal prefixes.

6. Conclusions

The original drive behind our research was to contribute to the study of Italian VPCs, and in particular to their semantics. Traditional difficulties concerning the intermediate structure between syntax and the lexicon, the compositional view of *Aktionsart* and phenomena of semi-productivity were overcome by adopting a constructionist approach. Further, we pointed out that Italian VPCs are also an interesting typological issue, as they do not fit the Germanic vs. Romance opposition about the realization of motion events to be found in Talmy (1985, 2000b).

The main goal of the paper was to look for possible actional properties of particles within VPCs. In order to pursue this goal, we carried out an investigation on the *Aktionsart* of Italian VPCs with respect to their verbal bases and found out that, besides specifying locative information, Italian post-verbal particles do contribute to the *Aktionsart* of verbal bases by virtue of a metonymic re-interpretation of their locative meaning. In particular, there seems to be a set of *telic particles* whose actional value is rather coherent (we never find a *telic* particle that detelicizes a verbal base), but not (or not yet) so systematic as to be referred to as proper *Aktionsart* markers.

Lastly, we compared VPCs to verbal prefixation. The comparison showed that there is no competition between these two types of constructions. Rather, today VPCs in Italian *cooperate* with verbal prefixation by performing different functions in the language.

---

system and gives an account of their decline. Cf. also Romagno (2003) for the interplay of actional prefixes and thematic structure in Latin.

References

Ackerman, F. 2003. "Aspectual Contrasts and Lexeme Derivation in Estonian: A Realization-Based Morphological Perspective". In G. Booij & J. van Marle (eds.) *Yearbook of Morphology 2003*, Dordrecht: Kluwer, 13-60.

Ackerman, F. & G. Webelhuth 1998. *A Theory of Predicates*. Stanford: CSLI Publications.

Antelmi, D. 2002. "Il verbo senza significato: possibilità di slittamento del contenuto lessicale su elementi di tipo nominale". Rivista italiana di linguistica e di dialettologia 4, 97-117.

Behrens, L. 1998. *Ambiguität und Alternation: Methodologie und Theoriebildung in der Lexikonforschung*. Habilitationsschrift, Ludwig Maximilians Universität, München.

Bertinetto, P. M. 1986. *Tempo, aspetto e azione nel verbo italiano. Il sistema dell'indicativo*. Firenze: Accademia della Crusca.

Bertinetto P. M. 1997. *Il dominio tempo-aspettuale: demarcazioni, intersezioni, contrasti*. Torino: Rosenberg & Sellier.

Bertinetto, P. M. 2001. "On a Frequent Misunderstanding in the Temporal-Aspectual Domain: The 'Perfective-Telic Confusion'". In C. Cecchetto, G. Chierchia & M. T. Guasti (eds.) *Semantic Interfaces: Reference, Anaphora and Aspect*, Stanford: CSLI Publications, 177-210.

Blom, C. 2005. *Complex Predicates in Dutch. Synchrony and Diachrony*. PhD dissertation, Vrije Universiteit Amsterdam (LOT Publications Series 111).

Bolinger, D. 1971. *The Phrasal Verb in English*. Cambridge: Harvard University Press.

Boogaart, R. 2004 "Aspect and Aktionsart". In G. Booij, Ch. Lehmann & J. Mugdan (eds.) *Morphologie/Morphology: Ein internationals Handbuch zur Flexion und Wortbildung/An international Handbook on Inflection and Word-Formation* Vol. 2. Berlin and New York: Mouton de Gruyter, 1165-1180.

Booij, G. 2002a. "Constructional Idioms, Morphology and the Dutch Lexicon". *Journal of Germanic Linguistics* 14 (4), 301-329.

Booij, G. 2002b. "Separable Complex Verbs in Dutch: A Case of Periphrastic Word Formation". In N. Dehé *et alii* (eds.), 21-41.

Booij, G. 2005a. "Construction Morphology" Vrije Universiteit Amsterdam, Ms.

Booij, G. 2005b. "Construction-Dependent Morphology". *Lingue e linguaggio* 4 (2), 163-178.

Booij, G. & A. van Kemenade 2003. " Preverbs: An Introduction". In G. Booij & J. van Marle (eds.) *Yearbook of Morphology 2003*, Dordrecht: Kluwer, 1-12.

Brinton, L. J. 1988. *The Development of English Aspectual Systems. Aspectualizers and Post-Verbal Particles*. Cambridge: Cambridge University Press.

Cini, M. (2002), "I verbi sintagmatici negli etnotesti dell'ALEPO". In G. Marcato (ed.) *La dialettologia oltre il 2001* (*Quaderni di dialettologia* 6), Padova: Unipress, 143-150.

Cini, M. (in press) (ed.), *I verbi sintagmatici in italiano e nelle varietà dialettali. Stato dell'arte e prospettive di ricerca*, Frankfurt am Main: Peter Lang.

Declerck, R. 1977. "Some Arguments in Favor of a Generative Semantics Analysis of Sentences with an Adverbial Particle or a Prepositional Phrase of Goal". *Orbis* 26, 297-340.

Dehé, N. 2002. *Particle Verbs in English. Syntax, Information Structure and Intonation.* Amsterdam/Philadelphia: John Benjamins.

Dehé, N. *et alii* (eds.) 2002. *Verb-Particle Explorations*. Berlin/New York: Mouton de Gruyter.

Dikken, M. den 1995 *Particles: On the Syntax of Verb-Particle, Triadic, and Causative Constructions*. Oxford: Oxford University Press.

DISC = *Dizionario italiano Sabatini Coletti*. Firenze: Giunti Multimedia 1997.

Dixon, R. 1982. "The Grammar of English Phrasal Verbs". *Australian Journal of Linguistics* 2 (1), 1-42.

Dowty, D. R. 1979, *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Semantics and in Montague's PTQ*. Dordrecht: Reidel.

Fillmore, Ch. J., P. Kay & M. C. O'Connor 1988. "Regularity and Idiomaticity in Grammatical Constructions: the Case of Let Alone". *Language* 64 (3), 501-538.

Fraser, B. 1976. *The Verb-Particle Combination in English*. New York/San Francisco/London: Academic Press.

Goldberg, A. 1995. *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, A. 2003. "Constructions: A New Theoretical Approach to Language". *Trends in Cognitive Studies* 7 (5), 219-224.

GRADIT = *Grande dizionario italiano dell'uso*. Tullio De Mauro editor in chief, Torino: UTET 1999.

Haiden, M. 2002. "Verb Particle Constructions". Aston University, Ms. (available at http://www.univ-lille3.fr/silex/equipe/haiden/particle/case_117_vepa.htm).

Haverling, G. 2003. "On Prefixes and Actionality in Classical and Late Latin". *Acta Linguistica Hungarica* 50 (1-2), 113-135.

Hiltunen, R. 1983. *The Decline of the Prefixes and the Beginnings of the English Phrasal Verb. The Evidence from some Old and Early Middle English Texts*. Turku:Turun Yliopisto.

Iacobini, C. 2003. "L'indicazione di valori locativi a partire da basi verbali". Seminar held at the University Roma Tre (Department of Linguistics), 13 November 2003.

Iacobini, C. 2004. "Parasintesi". In M. Grossmann & F. Rainer (eds.) *La formazione delle parole in italiano.* Tübingen: Niemeyer, 165-188.

Iacobini, C. 2005. "I verbi italiani come base di derivazione prefissale". In M. Grossmann & A. M. Thornton (eds.) *La formazione delle parole.* Atti del XXXVII Congresso Internazionale di Studi della SLI. Roma: Bulzoni; 289-307.

Jackendoff, R. 1997. *The Architecture of the Language Faculty*. Cambridge: The MIT Press.

Jackendoff, Ray 2002a. "English particle constructions, the lexicon, and the autonomy of syntax". In N. Dehé *et alii* (eds.), 67-94.

Jackendoff, R. 2002b. *Foundations of Language. Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jansen, H. 2002. "Spatialpartikler. Forstudier om brugen af præpositioner og lokative adverbier på italiensk og dansk". *Ny forskning i grammatik, fællespublikation* 9, 121-140.

Jansen, H. 2004. "La "particella spaziale" e il suo combinarsi con verbi di movimento nell'italiano contemporaneo". In P. D'Achille (ed.) *Generi, architetture e forme testuali.* Atti del VII Convegno SILFI . Firenze: Franco Cesati editore, 129-144.

Jezek, E. 2002. "Lo sfondamento di un confine tipologico. Il caso dei verbi complessi nell'italiano". In P. Cordin, R. Franceschini & G. Held (eds.) *Parallela 8. Atti dell'ottavo incontro italo-austriaco dei linguisti. Lingue di confine, confini di fenomeni linguistici*. Roma: Bulzoni, 289-308.

Klein, H. G. 1969. *Das Verhalten der telischen Verben in den romanischen Sprachen erörtert an der Interferens von Aspekt und Aktionsart.* Dissertation, Johann Wolfgang Goethe-Universität, Frankfurt am Main.

Lindner, S. J. 1983. *A Lexico-Semantic Analysis of English Verb Particle Constructions*. Bloomington: Indiana University Linguistics Club.

Lüdeling, A. 2001, *On Particle Verbs and Similar Constructions in German*. Stanford: CSLI Publications.

Lüdtke, J. 1996. "Gemeinromanische Tendenzen IV. Wortbildungslehre". In G. Holtus *et alii* (eds.) *Lexicon der Romanistischen Linguistik.* Tübingen: Niemeyer, vol. II, 1: 235-272.

Masini, F. 2005. "Multi-Word Expressions between Syntax and the Lexicon: The Case of Italian Verb-Particle Constructions". In *SKY Journal of Linguistics* 18:2005, 145-173.

Masini, F. 2006. "Diacronia dei verbi sintagmatici in italiano". In *Archivio Glottologico Italiano* XCI (1), 67-105

Masini, M. (in press), "Verbi sintagmatici e ordine delle parole", In M. Cini (ed.) *I verbi sintagmatici in italiano e nelle varietà dialettali. Stato dell'arte e prospettive di ricerca*, Frankfurt am Main: Peter Lang.

McIntyre, A. 2001. *German Double Particles as Preverbs: Morphology and Conceptual Semantics.* Tübingen: Stauffenburg.

McIntyre, A. 2002. "Idiosyncrasy in Particle Verbs". In N.Dehé *et alii* (eds.), 95-118.

Müller, S. 2002. *Complex Predicates: Verbal Complexes, Resultative Constructions, and Particle Verbs in German*. Stanford: CSLI Publications.

Romagno, D. 2003. "Azionalità e transitività: il caso dei preverbi latini". *Archivio Glottologico Italiano* 88 – 2, 156-170.

Sasse, H.-J. 2002. "Recent Activity in the Theory of Aspect: Accomplishments, Achievements, or just Non-Progressive State?". *Linguistic Typology* 6, 199-271.

Schwarze, Ch. 1985. ""Uscire" e "andare fuori": struttura sintattica e semantica lessicale". In A. Franchi de Bellis & L. M. Savoia (eds.) *Sintassi e morfologia della lingua italiana d'uso. Teorie e applicazioni descrittive.* Atti del XXIV Congresso Internazionale SLI. Roma: Bulzoni, 355-371.

Simone, R. 1997. "Esistono verbi sintagmatici in italiano?". In T. De Mauro & V. Lo Cascio (eds.) *Lessico e grammatica. Teorie linguistiche e applicazioni lessicografiche*. Roma: Bulzoni, 155-170.

Smith, C. 1991. *The Parameter of Aspect*. Dordrecht: Kluwer.

Stiebels, B. & D. Wunderlich 1994. "Morphology Feeds Syntax". *Linguistics* 32, 919-968.

Talmy, L. 1985. "Lexicalization Patterns: Semantic Structure in Lexical Forms". In T. Shopen (eds.) *Language Typology and Syntactic Description, Volume III. Grammatical Categories and the Lexicon.* Cambridge: Cambridge University Press, 57-149.

Talmy, L. 2000a. *Toward a Cognitive Semantics: Concept Structuring Systems, Vol. 1*. Cambridge: MIT Press.

Talmy, L. 2000b. *Toward a Cognitive Semantics: Typology and Process in Concept Structuring, Vol. 2*. Cambridge: MIT Press.

Toivonen, I. 2003. *Non-Projecting Words. A Case Study of Swedish Particles*. Dordrecht: Kluwer.

Traugott, E. C. 1982. *From Propositional to Textual and Expressive Meanings: Some Semantic-Pragmatic Aspects of Grammaticalization*. In W. P. Lehmann & Y. Malkiel (eds.) *Perspectives on Historical Linguistics.* Amsterdam: Benjamins, 245-271.

Vendler, Z. 1967. "Verbs and Times". In Id. *Linguistics in Philosophy*. Ithaca N.Y.: Cornell University Press, 97-121.

Venier, F. 1996. "I verbi sintagmatici". In P. Blumenthal, G. Rovere & Ch. Schwarze (eds.) *Lexikalische Analyse Romanischer Sprachen.* Tübingen: Max Niemeyer Verlag, 149-156.

Vicario, F. 1997. "I verbi analitici in friulano". Milano: Franco Angeli.

Vincent, N. 1999. "The Evolution of C-Structure: Prepositions and PPs from Indo-European to Romance" *Linguistics* 37-6, 1111-1153.

Wienold, G. & C. Schwarze, 2002. "The lexicalization of movement concepts in French, Italian, Japanese and Korean: Towards a realistic typology". Arbeitspapier 112, Fachbereich Sprachwissenschaft Universität Konstanz.

Constraining Gender Assignment Rules

Anna M. Thornton
Università dell'Aquila
thornton@cc.univaq.it

Abstract

The paper deals with gender assignment, i.e., the process by which speakers are able to assign a gender feature to a lexical item not yet bearing such a feature. The need for gender assignment arises mainly in the case of headless neologisms and loanwords. The paper draws evidence mainly from loanwords into Italian, a language which has a two-gender system. The first part of the paper tests several hypotheses about the existence of dominance relations between two kinds of gender assignment criteria, formal and semantic ones, against Italian data. Italian data seem best compatible with theories that allow for the possibility that semantic rules dominate over formal rules in gender assignment. In the second part of the paper, a constraint on possible semantic gender assignment rules is proposed, the *Basic Level Hyperonym Constraint,* stating that to be able to assign gender to its hyponyms, a hyperonym must be a *basic level term.*

1. Introduction: What is gender assignment?

This paper deals with gender assignment. There are actually two views of what gender assignment is in the literature. According to view (1a), gender assignment is a process by which speakers classify the nouns of their language into gender classes; genders are conceptualized as containers, in which nouns can be stored. Another view (1b) considers gender assignment as a process by which speakers assign a gender feature to a given noun: in this view, genders are seen as features necessary for the nouns to function syntactically.

(1) Two different views on gender assignment:

a.	assign nouns to genders	=	Genders as containers
b.	assign gender to nouns	=	Genders as features

Both views are expressed and normally go undistinguished – or the difference goes unnoticed – in the literature. In (2) I list a few quotations from the bible of gender studies, Corbett (1991), which show that the author switches between the two views within the same chapter:

G. Booij, et al. (eds.), On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5) Fréjus 15-18 September 2005, University of Bologna, 2007.
URL http://mmm.lingue.unibo.it/

(2) The two views of gender assignment in Corbett (1991), chapter 2

a. native speakers allocate nouns to genders (3)
native speakers assign nouns to genders (8)
nouns are assigned to gender according to their meaning (8)
loanwords were assigned […] to the same gender by different speakers (16)
two different criteria may assign nouns to the same gender (32)

b. words borrowed from other languages acquire a gender (7)
there is a mechanism for assigning […] gender (7)
speakers give them [i.e., invented words] a gender (7)
Not only can speakers assign gender to English words […] (23)
the referent is sufficient to enable gender to be assigned (23)

Similar oscillations between the two different views of what gender assignment is are found in other important studies on the topic, such as Doleschal (2004) and Poplack, Pousada and Sankoff (1982).

The first point of view arises out of interest in the cognitive classification of experience that gender systems reflect, while the second one arises out of interest in the way gender functions in grammar.

In the present paper I will adopt view (1b), concentrating on the way a gender feature is assigned to a noun which must function as controller of gender agreement in a syntactic construction.

2. Who assigns gender to what?

The first question I want to address, then, is the following: Who assigns gender to what? Several answers are possible. I will list the main ones in (3a-d):

(3)

	Who assigns gender?		*To what?*
a.	child learning L1	→	to all nouns in L1
b.	speaker learning L2	→	to all nouns in L2
c.	linguist writing a computational grammar of Lx	→	to all nouns in Lx
d.	adult speaker of L1	→	to nouns in L1 that control gender agreement on targets but do not (yet) have a gender feature in their lexical entry

The next question is: Are the assignment criteria / rules used the same in all the cases listed in (3)?

We don't have a clear answer yet, but I believe that possibly the answer is "no". I believe there is a fundamental difference between case (3c) and cases (3a, b, d). The rules used in the latter cases require a cognitive basis, a psychological reality, that rules used for case (3c) do not need. Problem (3c) can be solved with regard to criteria of pure economy, while solutions to the problems in (3a, b, d) are constrained by cognitive principles.

Unfortunately, the literature on gender assignment often fails to distinguish between the different cases in (3), and the rules that are proposed are meant to work for all these different purposes.[1]

In this paper, I will not be concerned with cases (3a-c), and I will concentrate on case (3d), drawing my data mostly from the Italian language. I believe that the proposals that I will put forward could be extended to case (3a) (very likely, with extra provisos taking care of constraints related to children's cognitive capacities), and maybe in part also to case (3b) (with extra provisos for the specificity of L2 learning), while they have virtually no bearing on case (3c), which I consider a completely different domain of research, within the scope of computational linguistics and not of cognitive linguistics.
We can now move on to the next question: What sorts of entries do not already have a gender feature in the mental lexicon of adult speakers of a L1?
A few answers are listed in (4):

(4) a. loanwords;

b. toponyms (e.g., names of cities);

c. certain kinds of neologisms, such as:
exocentric compounds;
compounds whose head is not a noun;
nouns converted from verbs (but <u>not</u> derived nouns headed by suffixes which have an inherent gender)

d. non-prototypical controllers (e.g., clauses, cfr. Corbett 1991: 204)

e. …

2.1. Case study: names of cities in Italian

To illustrate the kind of problem I want to address, I will start with an example, that of names of cities (4b). Names of cities can go genderless in the mental lexicon of speakers of Italian for a long time. The most frequent contexts of occurrence of names of cities are the ones in (5), where the city name is governed by a preposition, and there is no gender agreement target in the context.

[1] For example, the paper by Fraser and Corbett (1995) on gender assignment in Russian adopts the point of view in (3c), with the consequence that the system of rules they propose does not account for the gender of some recent loanwords, which must be manually specified in the corresponding lexical entries (Fraser and Corbett 1995: 133): the system is thus inadequate for case (3d).

(5) *sono di Roma, vengo da Roma, vado a Roma, A Roma…*
"I am from Rome, I come from Rome, I am going to Rome, In Rome…"

But a minority of contexts in which names of cities do control gender agreement targets nevertheless exists: cfr. the data and examples in (6-8):

(6) Percentage of names of cities occurring in contexts in which they control gender agreement targets in three corpora of Italian (data from Nitrola 1998)[2]

Decameron	= 2,4%
Promessi Sposi	= 3,4%
LIP	= 4,2%

(7) Masculine gender in names of cities in nineteenth and early twentieth century written Italian (data collected by Nitrola 1998)

quel maledetto Casale	"that-m damned-m Casale"
quel povero Casale	"that-m poor-m Casale"
un Milano	"a-m Milan"
mezzo Milano	"half-m Milan"
in quel Milano	"in that-m Milan"
Milano si trova ormai in tale stato, da non vedere cosa giovasse guardarlo	"By now, Milan is in such a state that no one can understand the value of looking at it-m"
in questo Milano	"in this-m Milan"
com'è conciato Milano	"the state to which Milan is reduced-m" (Promessi Sposi)
Parigi sbastigliato	"un-bastilled-m Paris" (Alfieri)
Urbino ventoso	"windy-m Urbino" (Pascoli)
bel mi' Firenze	"beautiful-m my-m Florence" (Moretti)

(8) Feminine gender in names of cities in contemporary spoken Italian (*LIP* corpus)

una Napoli capitale	"a-f capital city Naples"
questa grande Napoli	"this-f big Naples"
una Napoli popolare	"a-f popular Naples"
la nostra bella Napoli	"the-f our-f beautiful-f Naples"
una Parigi abbastanza opaca	"a-f quite gloomy-f Paris"
Torino si è fermata	"Turin stopped-f"

[2] *Decameron* is a collection of short stories written by Giovanni Boccaccio in the 14th century; *Promessi Sposi* is a novel written by Alesssandro Manzoni in the first half of the 19th century; *LIP* is a 500.000-token corpus of spoken Italian collected in the years 1990-1992 (cfr. De Mauro, Mancini, Vedovelli and Voghera 1993).

The data in (6) to (8) show that speakers of Italian must have in their competence a system allowing them to assign gender to names of cities, when the need arises. Moreover, a comparison between the data in (7) and those in (8) shows that the criterion for gender assignment to names of cities must have changed in diachrony, as (at least some) names of cities were masculine in nineteenth century Italian, while all names of cities are feminine in contemporary Italian. Apparently, the criterion by which gender is assigned to names of cities has shifted in the history of Italian: up to about a century ago it used to be a phonological criterion, by which names of cities ending in *-o* and *-i* (such as *Milano* "Milan" and *Parigi* "Paris") were assigned masculine gender, while nowadays it seems to be a semantic criterion, by which all names of cities are feminine, regardless of their final vowel, as the examples in (8) show.

3. Criteria for gender assignment

The discussion about the gender of names of cities in Italian has served to illustrate the problem of gender assignment, and has introduced the concept of gender assignment criteria.

We will now concentrate on the main class of nouns that can for some time exist in the mental lexicon of an adult speaker without a gender feature, but must at some point acquire this gender feature in order to function syntactically, as controllers of gender agreement targets. These are neologisms and loanwords (or borrowings). To quote the universally acknowledged master of gender studies, Greville G. Corbett, «Borrowings of nouns into languages with gender systems […] are like a continuously running experiment, which allows us to verify the assignment system in the languages in question» (Corbett 1991: 71).

As Audring (2004) observes, investigating gender assignment to loanwords and neologisms ensures that we will uncover psychologically real and productive criteria that speakers exploit in "on-the-spot" gender assignment, rather than just "postfactum rationalizations", as Comrie (1999. 461) dubs some of the gender assignment criteria that have been proposed in the literature.

Corbett (1991) has provided us with a typology of gender assignment systems and rules. There are semantic rules, by which gender is assigned on the base of a noun's meaning, and formal rules, by which gender is assigned on the base of a noun's shape or of its forms. Formal rules are of two kinds: phonological and morphological. A typical phonological rule is a rule like "Nouns ending in *-a* are feminine", while a typical morphological rule is a rule like "Nouns of inflectional class 2 are feminine". The difference between phonological and morphological rules lies in the fact that phonological rules refer to a single form of the noun (typically, the citation form), while morphological rules refer to more than one form in a noun's paradigm (as many forms as are necessary to establish the inflectional class to which the noun belongs). When no specific semantic or formal rule applies to a noun, the default gender is assigned ("normal case default" in Fraser and Corbett 1997, Corbett and Fraser 1999).

Corbett's typology is not meant to account only for gender assignment in case (3d), but for all cases in (3), as well as for a further possible case: in my understanding at least, this is also a typology of redundancy rules that account for the gender of nouns which have been attested in the lexicon of a language for a long time, and to which no given speaker has ever needed to assign a gender feature "on the spot" (although

children must have acquired the gender feature together with the rest of the lexical information connected to each noun). That Corbett has in mind a typology that accounts for all instances of noun-gender pairings, and not just the ones that must be effected "on the spot" (i.e., when a speaker must produce gender agreement with a noun whose lexical entry was previously lacking a gender feature), is shown by the fact that most of Corbett's examples are regular nouns of the languages investigated, not just loanwords or other nouns of the kinds listed in (4) above.

As Corbett himself comments on his typology, «there will always be semantic assignment rules […] since no language has a purely formal assignment system» (Corbett and Fraser 2000: 297).

This brings us to the two main topics to be addressed in the present paper, presented in the form of questions in (9a-b):

(9) a. If a language has both semantic and formal gender assignment rules, is there a hierarchy between the two sorts of rules? i.e., if a noun falls within the scope of two rules which would assign different genders, namely, a semantic rule and a formal rule, how is the conflict resolved? what (kind of) rule prevails? And, if there is a principled order of application of different kinds of rules, is this order universal or language-specific?

b. Given that all languages have some semantic assignment rules, is it possible to find constraints on the kind of semantic rules that exist?

We will address both these questions in the rest of this paper. Section 4 will be devoted to discussing the questions in (9a), while section 5 will address the question in (9b).

4. Is there a hierachy of dominance between semantic and formal gender assignment criteria?

Different authors have given different answers to the questions listed in (9a), which are currently a topic of heated debate in gender assignment studies (see above all Rice 2004, Rice 2005, but also Corbett 1991, Corbett and Fraser 2000, Doleschal 1999, Doleschal 2004, and a forthcoming special issue of *Lingua* edited by Enger, Nesset and Rice).

Corbett and Fraser (2000) maintain that, universally, semantic criteria dominate formal criteria:

(10) *Universally, semantic >> formal*
«As is universally the case, the formal gender assignment rules are dominated by the semantic gender assignment rules» (Corbett and Fraser 2000:321)

Rice (2004), on the contrary, building on previous work by Steinmetz (e.g., Steinmetz 1986), maintains that universally, gender assignment criteria and language-specific gender assignment rules or constraints are crucially non-ranked with respect to each other; only the genders are ranked in a markedness hierarchy in each language; the language-specific gender assignment criteria are all non-ranked with respect to each

other but form a block that, as a whole, is ranked above the language's gender markedness hierarchy; when a conflict arises, because there is a tie between specific criteria, the noun is assigned the least marked gender of the language:

(11) *Optimal Gender Assignment Theory (OGAT)*:
universally, «crucial non-ranking»
«in all languages, all constraints referring to gender-relevant features are equally ranked»
«conflicts between them are decided by ranked constraints implementing a markedness hierarchy: when two constraints are in conflict, the noun is assigned to the least marked of the conflicting categories» (Rice 2004)

Even authors who defend the idea that there is some ranking between different gender assignment rules are divided as to whether this ranking is universal or language-specific.

Nesset (2004) maintains that there is at least one universally dominating criterion, that he calls *The Core Semantic Override Principle (CSOP),* stated in (12):

(12) *The Core Semantic Override Principle (CSOP)*
«Rules referring to biological sex take precedence in gender assignment» (Nesset 2004)

Audring (2004), on the contrary, in a study framed in OT concerning the assignment of gender to English loanwords in several European languages, maintains that the ranking of different formal and semantic criteria can vary from language to language.

4.1. Case study: masculine nouns ending in -a *in Russian and in Italian*

Now that we have reviewed the different positions found in the literature on question (9a), we can proceed to look at how the different systems explain the gender assignment that occurs in a few specific cases. The most favoured example in the literature is the Russian word for "uncle", *djadja.* An overview of the relevant rules/constraints in Russian is given in (13).

(13) *Russian*

a. 3 genders: masculine, feminine, neuter

b. Markedness hierarchy: *n >> *f >> *m
(i.e., masculine is less marked than feminine, feminine is less marked than neuter)

c. Semantic rule/constraint:
nouns denoting biological males are masculine
= *[+ MALE] ⇒ F, N

d. Formal rule/constraint:
nouns ending in +*a* are feminine
= *+A# ⇒ M, N

Djadja should be feminine by formal criteria, because it ends in -*a* (and/or maybe because it belongs to declensional class II),[3] but should be masculine by semantic criteria, because it denotes a male. *Djadja* is indeed masculine, but the authors disagree on why it is masculine. For Corbett it is masculine because semantic rules (of any kind) prevail over formal rules universally; for Nesset, it is masculine because rules referring to biological sex prevail on other (both formal and semantic) rules, again universally; for Audring, it is masculine because in Russian (but not necessarily in other languages) the rule that assigns gender according to the sex of the referent dominates the rule that assigns gender according to the phonological shape of the signifier (SEX >> PHON, in her notation); for Rice, finally, *djadja* is masculine because masculine is the least marked gender in Russian.

We can see how Rice's system works by looking at the tableau in (14):

(14) Gender assignment to Russian *djadja* "uncle" in OGAT (Rice 2004, Rice 2005: Table 4)

djadja "uncle"	GENDER FEATURES *[+ MALE] ⇒ F, N	*+A# ⇒ M, N	*N	*F	*M
☞ a. djadj+a, m.		*			*
b. djadj+a, f.	*			*!	
c. djadj+a, n.	*	*!	*		

Rice (2005) illustrates the situation depicted in (14) in the following way: «The first constraint is violated by candidates (b) and (c), since the noun denotes a male. The second constraint is violated by candidates (a) and (c). Since candidate (c) violates both of the equally ranked constraints while candidates (a) and (b) each violate just one, candidate (c) is ruled out at this point, as indicated by the exclamation point. Candidates (a) and (b) are distinguished by the markedness hierarchy. Specifically, candidate (b) is ruled out by the relatively highly ranked constraint *FEMININE, leaving candidate (a) as optimal».

As in Rice's system the semantic rule in (13c) and the phonological rule in (13d) (expressed as prohibitions rather that as positive rules, as usual in OT) are crucially unranked with respect to each other, the situation represents a "balanced conflict" between masculine and feminine gender (neuter is ruled out, as we have seen, because assigning neuter would violate both the semantic and the formal constraint). The conflict is resolved by the language-specific markedness hierarchy in (13b), by which masculine is the least marked gender in Russian.

It must be observed that *djadja* is not the kind of noun I am most interested in, since it is neither a neologism nor a recent loanword to which Russian speakers had to

[3] Rice's formulation is that "nouns ending in the segmentable morpheme" -*a* are feminine, and this formulation is claimed to be "a notational variant of the the claim that nouns of the 2nd declension are feminine" (Rice 2004, footnote 6). The precise formulation of this rule/constraint is irrelevant at the moment; what matters for present purposes is that this is a formal rather than a semantic gender assignment criterion.

assign gender on the spot. In any case, all the observations made above for *djadja* would extend to recent loanwords. In the literature, recent loanwords into Russian ending in *-a* and denoting males are not discussed much[4], but fortunately the Russian situation is entirely parallel to the Italian one, as shown in (15).

(15) *Italian*

a. 2 genders: masculine, feminine

b. Markedness hierarchy: *f >> *m
(i.e., masculine is less marked than feminine)

c. Semantic rules/constraints:
nouns denoting biological males are masculine;
nouns denoting biological females are feminine

d. Phonological rules/constraints:
nouns ending in *-a* are feminine;
nouns ending in *-o* are masculine

All the observations made above about the treatment of *djadja* in the different models extend completely to the treatment of loanwords into Italian denoting males and ending in *-a,* like the ones in (16):

(16) Nouns denoting males and ending in *-a* (loanwords into Italian)

lama	"Tibetan monk",
Dalai lama	"Id.",
Sherpa	"Tibetan baggage carrier",
Ulema	"Islamic doctor in theology and law",
ustascia	"Croatian nationalist soldier",
peshmerga	"Kurdish partisan soldier"…

The tableau in (17) shows how a balanced conflict arises in Italian for nouns denoting males and ending in *-a,* and how it is solved in Rice's system.

(17) Gender assignment to Italian *sherpa* in OGAT

sherpa "Tibetan baggage carrier"	GENDER FEATURES *[+ MALE] ⇒ F *-*a*# ⇒ M	*F	*M
☞ a. sherpa m.	*		*
b. sherpa f.	*	*!	

Sherpa should be masculine because it denotes a male, but it should be feminine because it ends in *-a*; these two constraints, if unranked with respect to each other, yield

[4] A few words in this category, however, exist: one can cite at least *mulla* "mullah", *nindzja* "ninja warrior", *maharadža* "maharajah", *lama* "Tibetan monk" and *Neruda* "Id.". Thanks to Ursula Doleschal and Tore Nesset for pointing out these examples to me.

a balanced conflict; according to OGAT, the noun is then assigned masculine gender because this is the least marked gender in Italian.

Rice (2004) rightly observes that in order to assess whether there is a universal dominance of semantic criteria over formal ones, as Corbett maintains (recall (10) above), «we must find cases of balanced conflict with a mismatch between feature type (meaning or shape) and category markedness. Specifically, we must find cases in which the shape correlates with a less marked category while the meaning correlates with a more marked category».

In two of his papers, Rice (2004, 2005) looks for such cases in Russian, but concludes that this language does not offer appropriate examples. Italian, on the contrary, offers a very interesting case of the kind we are looking for, that I will illustrate in the next section.

4.2. Case study: Italian feminine nouns ending in -o

Italian, besides having masculine nouns ending in *-a*, that should be feminine because of their phonology but are masculine because of their semantics, also has feminine nouns ending in *-o*, which should be masculine because of their phonology but should be feminine because of their semantics (cfr. (15c-d) above). This is a clear case where phonology would assign the least marked gender, while semantics would assign a more marked gender. Examples of Italian feminine nouns in *-o* are given in (18).

(18) Some Italian feminine nouns ending in *-o*

biro	"ballpoint pen",
cabrio	"kind of car",
lampo	"zip",
merino	"merino sheep",
polo	"polo shirt",
sdraio	"deck chair",
soprano	"id.",
squillo	"call girl",
torpedo	"kind of car",
virago	"id."

Three of these nouns denote women: *soprano, virago, squillo*. Each one of the three nouns presents some complications, but as these are the best examples I have been able to come up with so far of nouns that would get a less marked gender by formal criteria and a more marked gender by semantic criteria, I will exploit them as much as possible. Before testing OGAT on nouns of this type, however, I will give a little more information on the data.

Soprano actually has variable gender: a web search conducted by means of Google in August 2005 for the strings *la soprano / il soprano* "the-f soprano / the-m soprano" yielded a proportion of masculine to feminine of about 4:1 (16800 m vs. 4310 f). The contexts of usage are completely parallel for the two genders, as examples (19a-d vs. 19 e-h) show, with the well known awkward agreement facts that arise when the noun is considered masculine (masculine agreement within the NP, feminine agreement outside it, cfr. (19 g-h)).

(19) *Soprano* in context (data from Google)

a. il regista Costa Gravas girerà un film sull'amore tra Onassis e la soprano.
"Director Costa Gravas will shoot a film about the love story between Onassis and the-f soprano"

b. ... nella villa di Sirmione (dove la soprano era solita rilassarsi con lui).
"in the villa in Sirmione (where the-f soprano used to relax with him)"

c. La soprano Barbara Frittoli durante l'esecuzione dell'aria "Signore ascolta" dalla Turandot di Puccini
"The-f soprano Barbara Frittoli while singing "Signore ascolta" from Puccini's Turandot"

d. È morta la soprano Renata Tebaldi.
"The-f soprano Renata Tebaldi died"

e. Il soprano Roberta Frameglia esegue *Frontiere Borders Fronteras* ...
"The-m soprano Roberta Frameglia sings *Frontiere Borders Fronteras*…"

f. La "fantasia popolare" vede il soprano come una donna brutta, grassa, antipatica ..."Popular lore sees the-m soprano as an ugly, fat, disagreeable woman…"

g. Bravissimo anche il soprano Maria Grazia Schiavo, già ammirata in altre occasioni …
"Also very good-m was the-m soprano Maria Grazia Schiavo, who had already been admired-f in other circumstances…"

h. La più interessante di questi è stata il soprano americano Michèle Crider
"The-f most interesting of these-m was-f the-m American-m soprano Michèle Crider"

The reason why *soprano* (first attested in Italian in the fifteenth century) is used as a masculine noun even though it (nowadays) refers to women is well known: the word was originally an adjective, meaning "upper", and was used in the technical language of music in the phrase *registro soprano* "upper register", referring to the highest voice register, obtainable by women, children and castrated men. In old times, sopranos were in fact castrated men, so the masculine gender of the noun matched not only the phonology but also the semantic rule by which words referring to men are masculine; besides, the noun originated from an adjective getting its gender by agreement with the masculine head noun *registro,* through ellipsis.

Squillo "call girl", on the other hand, is consistently used as a feminine noun, notwithstanding the fact that *squillo* in the meaning of "ring (of a telephone)" is a regular masculine noun. *Squillo* is a rendering of English "call girl" by the juxtaposition of the two words *ragazza* "girl" and *squillo* "ring (of a telephone)", later reduced to *squillo* by ellipsis of the head noun.

Virago, a cultivated loanword from Latin which was feminine already in Latin, has never shown any tendency to be used as masculine.

All these nouns are somewhat flawed as examples of nouns ending in *-o* (an ending which in Italian is typical of masculine nouns, cfr. (15d) above) but receiving feminine gender for semantic reasons. *Soprano*'s obvious flaw lies in the fact that most speakers use this word as a masculine rather than a feminine noun; *squillo* on the contrary is always used as a feminine noun, but one could maintain that its gender is inherited from a deleted head *ragazza* "girl", and not directly assigned to the noun. *Virago* is the least flawed example: even if it is a cultivated loanword, and therefore one could maintain that, as is common with cultivated loanwords, it keeps the gender the word had in the donor language, it can be compared with other cultivated loanwords from Latin ending in *-o* that did not retain their feminine gender, such as *prefazio* "preface (in Mass)" (< Lat. *praefatĭo*), *dazio* "custom tax" (< Lat. *datĭo*). These nouns had no semantic reason for being feminine, and therefore the phonological assignment criterion took over; *virago*, on the contrary, was not affected by the phonological assignment criterion; therefore we can suppose that *virago* remained feminine due to its semantics, thus representing an acceptable example of a noun which receives a more marked gender through semantic criteria even if it could receive a less marked gender by formal criteria.

Let's now see in (20) how the gender of *virago* would be assigned by Rice's OGAT:

(20) Gender assignment to Italian *virago* in OGAT:

virago "id."	GENDER FEATURES *[+ FEMALE] ⇒ M	*-*o*# ⇒ F	*F	*M
☜a. virago m.	*			*
b. virago f.		*	*!	

Each of the two candidates violates one of the crucially unranked constraints: therefore, a balanced conflict arises, and it must be solved by the language-specific markedness hierarchy in favour of the least marked gender, which is masculine in Italian (cfr. Thornton 2003a). As the hand pointing to the left shows, OGAT predicts incorrect gender assignment – it predicts that Italian *virago* "id." will be masculine, while it is feminine –, at least if we accept that the constraints used in (20) are both real ones.
Should we then abandon OGAT altogether?

Before doing so, let's see whether there is a way to reconcile OGAT's strong claims with the Italian data on feminine nouns in *-o*. As is often the case in OT, a lot can change if we change the formulation of the constraints, without changing the spirit of the proposal. It must be observed that Rice formulates the relevant constraints in a negative fashion: the fact that nouns denoting women are feminine is not expressed in the positive way of (21a), but in the "negative" way of (21c); the same is true in the case of the rule stating that nouns denoting males are masculine (21b vs. 21d).

(21) a. [+ FEMALE] ⇒ F

b. [+ MALE] ⇒ M

c. *[+ FEMALE] ⇒ M

d. *[+ MALE] ⇒ F

The reasons for choosing negative rather than positive formulations are completely theory-internal to OT, and we cannot go into them now. But let us assume, contrary to Rice, that the right formulation of the constraints or rules is (21a-b) rather than (21c-d). In this spirit, even the phonological rules could be formulated differently, as (22a-b), rather than as (22c-d):

(22) a. *-a#* ⇒ F

b. *-o#* ⇒ M

c. **-a#* ⇒ M

d. **-o#* ⇒ F

If the relevant rules / constraints are the "positive" ones in (21a-b) and (22a-b), a further observation can be made: these positive formulas include rules, such as (21b) and (22b), that, on the basis of a given semantic or phonological feature, assign masculine gender, i.e., the unmarked gender of the language. It could be contended that such rules are redundant, and shouldn't exist, as the unmarked gender of a language would eventually be assigned to nouns anyway, even though no rule existed to assign it, because the nouns would be assigned that gender by default (in Corbett's terms) or by the markedness hierarchy (in Rice's terms).[5]

What would happen in OGAT if the rules were the "positive" ones we have just discussed, and specific rules assigning masculine didn't exist at all, leaving the markedness hierarchy to do the job of assigning masculine gender? (23) shows how *virago* would be treated in this case, and (24) shows the treatment of *sherpa* (to be compared with (17) above).

(23) Gender assignment to Italian *virago* in OGAT: version II

virago	GENDER FEATURES [+ FEMALE] ⇒ F	*F	*M
a. virago m.	*!		*
☞ b. virago f.		*	

(24) Gender assignment to Italian *sherpa* in OGAT: version II

sherpa	GENDER FEATURES *-a#* ⇒ F	*F	*M
a. sherpa m.	*!		*
☜ b. sherpa f.		*	

[5] See Thornton (2003a, 2003b) for further speculation on this topic.

If we remove the rules that would assign the unmarked gender, we get correct results for *virago*: only the semantic rule based on sex (21a) applies to it, and feminine gender is correctly assigned; but, by the same reasoning, this time we get wrong assignment for *sherpa*: if only the phonological rule (22a) applies to it, feminine gender is incorrectly assigned.

4.3. Provisional conclusion

The paradox illustrated in the preceding section seems to show that the conclusions stated in (25) should be drawn:

(25) a. even language-specific rules assigning the unmarked gender of the language are necessary – otherwise, we cannot assign masculine gender to *sherpa* (cfr. (24));

b. semantic rules assigning gender on the base of the sex of the referent dominate over phonological rules – otherwise, we cannot assign feminine gender to *virago*, cfr. (20).

At this point, we are in a position to reject Rice's OGAT in its current formulation; but we still don't know which of the claims about the dominance of semantic rules over phonological ones is right: so far, we haven't tested semantic rules other than those based on sex, so it could be that Corbett's claim (cfr. (10) above) that all semantic rules dominate formal rules is too strong, and perhaps all we need is Nesset's *Core Semantic Override Principle* (cfr. (12) above).

Therefore, we must test what we have developed so far with nouns that could receive gender by semantic rules of a different kind than the ones based on the sex of the referent.

4.4. Case study: names of cars in Italian

A semantic gender assignment rule not based on sex that has been shown to be quite robust for Italian (cfr. Thornton 2003a, 2003b) is the rule in (26), assigning feminine gender to nouns denoting cars. Some relevant data are given in (27):

(26) CAR ⇒ F (Thornton 2003a)

(27) Some data on feminine nouns denoting cars in Italian

a. una Fiat, una Ford, una Mercedes, una Ferrari, una Maserati, una Lamborghini
"a-f Fiat, etc."

b. la Cinquecento, la Mondeo, la Clio
"the-f Cinquecento, etc."

c. la Uno, la Tipo, la Ritmo, la Panda, la Bravo, la Tango "the-f Uno, etc." vs. *uno* m art/adj "one", *il tipo* m "the-m type", *il ritmo* m "the-m rhythm", *il panda* m "the-m panda.", *bravo* m adj "good-m", *il tango* m "the-m tango."

d. jeep, roadster, spider, torpedo, cabrio, citycar, station wagon

(27a) shows that names of cars are feminine even though they end in a C or in /i/, two phonological shapes which would assign masculine gender (cfr. Thornton 2003a); (27b) shows that names of cars are feminine even when they end in /o/, the typical masculine ending; (27c) shows that names of cars ending in /o/ are feminine even when they are homophonous with nouns that have other meanings in which they are regularly masculine, or with masculine forms of articles and adjectives; (27d) lists loanwords from English denoting cars, which are feminine even though they end in a C or in /o/ (cfr. Thornton 2003b). All the data in (27) establish that Italian has a semantic rule like (26).

Let's see now how gender would be assigned to the loanword *torpedo* by OGAT. The tableau is given in (28). Incidentally, how to formulate the constraints, whether in a positive or a negative fashion, is now irrelevant. I will reinstate the negative formulations (although I find them quite counter-intuitive) to remain faithful to Rice's original proposals.

(28) Gender assignment to Italian *torpedo* in OGAT

torpedo	GENDER FEATURES *CAR ⇒ M	*-*o*# ⇒ F	*F	*M
☜a. torpedo m.	*			*
b. torpedo f.		*	*!	

As we can see from the tableau in (28), we get the wrong predictions in this case too. The semantic rule by which nouns denoting cars are feminine and the formal rule by which nouns ending in -*o* are masculine tie. If the conflict were resolved by the gender markedness hierarchy, we would predict assignment of the masculine gender, contrary to the attested facts. *Torpedo* is feminine, according to the semantic rule that assigns feminine gender to nouns denoting cars.

The case of Italian feminine nouns ending in -*o*, therefore, seems to constitute evidence in favour of theories such as Corbett's, that maintain that semantic rules dominate over formal rules in gender assignment, or at least of theories such as Audring's, that maintain that ranking is not universal but language-specific (and in this case Italian would rank semantic constraints over formal ones), and against OGAT. The correct tableau for *torpedo* should look like the one in (29):

(29) Gender assignment to Italian *torpedo* in OT, with semantic constraints ranked above formal ones

torpedo	*CAR ⇒ M	*-*o*# ⇒ F	*F	*M
a. torpedo m.	*!			*
☞ b. torpedo f.			*	

4.5. Conclusion on dominance hierarchies between assignment criteria

There is still an alternative to save OGAT at this point: one could contend that only semantic rules assigning the unmarked gender of a language are permitted, while formal ones are not. This is the move that Rice himself would make (p.c., email of September 6, 2005, to the author). He observes that OGAT would yield the correct results in Tableaux (20) and (28) if the rule that prohibits assignment of feminine gender to nouns ending in *-o*, and therefore assigns them masculine gender, didn't exist. Then the semantic rules assigning feminine gender to nouns denoting females or cars would be the only rules ranked above the markedness hierarchy, ensuring assignment of feminine gender to *virago* and *torpedo*.

But I don't see a compelling reason for ruling out only formal rules that would assign the unmarked gender, while allowing semantic rules that do so, unless, of course, one is ready to admit that in some sense semantic rules are more important than formal rules, i.e., they dominate over them, which is exactly what OGAT denies.

5. Constraints on semantic gender assignment rules

Once we have established the relevance of semantic rules for correctly assigning gender to recent loanwords into Italian, the question arises whether it is possible to constrain the kinds of semantic gender assignment rules that we employ. We must at all cost avoid positing uncontrolled ad-hoc semantic rules just to counterbalance an unwelcome gender assignment.

No one has ever questioned the validity of semantic rules that assign gender on the basis of the sex of the referent. They come up over and over in the literature about any language which has a masculine and a feminine gender.

But what about other semantic rules? For instance, what about rule (26), which assigns feminine gender to nouns of cars in Italian? Is it grounded on any principle, or is it just my personal escape gate to account for the gender of the words in (27)?

I believe any answer to this question must be framed in a general outline of the kinds of semantic principles one is ready to allow as valid criteria for gender assignment. Existing literature on gender assignment to loanwords (too extensive to be listed) suggests that at least the three kinds of principles stated in (30) must be recognized to have effect in gender assignment to loanwords.

(30) Three criteria for gender assignment to loanwords

a. ASSOCIATE: the loanword copies the gender of a specific L1 noun to which it is associated (examples in Thornton 2003a)

b. EQUIVALENT: the loanword copies the gender of a specific L1 noun which is its translation equivalent (examples in Thornton 2003b)

c. HYPERONYM: the loanword inherits the gender of a specific L1 noun which is its hyperonym.

Effects of the ASSOCIATE criterion are often quoted in the literature. I will refer only to an example quoted by Corbett (1991):

(31) ASSOCIATE
«English *mud* has become Polish *mada* 'mud, silt' (feminine). There is no obvious reason why it should not have been *mad* (masculine); indeed, 88 per cent of the 681 loans investigated became masculine. Fisiak […] claims that the form *mada* is due to Polish *gleba* 'soil'.» (Corbett 1991: 76)

Effects of the EQUIVALENT criterion have also been observed: the classical study by Poplack, Pousada and Sankoff (1982) reports the examples in (32) in the Spanish spoken by Puertoricans in New York:

(32) EQUIVALENT
la butterfly "the-m butterfly" ← *la mariposa* "the butterfly f"
el building "the-m building" ← *el edificio* "the building m"
(Poplack, Pousada and Sankoff 1982: 11)

Hereafter, I will concentrate on some aspects of the assignment criterion called HYPERONYM in (30c).

5.1. The Basic Level Hyperonym Constraint

There is a strict test for the Hyponymy/Hyperonymy relation, stated in (33):

(33) X is a hyponym of Y iff 'all Xs are Ys' and not 'all Ys are Xs'
(Berruto 1976: 63)

According to (33), *car* is a hyponym of *vehicle*: all cars are vehicles, and not all vehicles are cars. But a Hyponymy/Hyperonymy relation isn't always exploited in gender assignment, as the Italian data in (34) show:

(34) a. *veicolo* m. "vehicle"/ *mezzo di trasporto* m. "transportation means"

b. *bicicletta* f. "bicycle", *macchina* f. "car", *motocicletta* f. "motorbike", *camion* m. "truck", …

The Italian masculine nouns in (34a), meaning "vehicle", are hyperonyms of the ones in (34b), but these do not inherit the gender of their hyperonym.

On the other hand, the nouns in (35) meaning "car" are hyperonyms of the nouns in (27) above, and these do inherit their hyperonym's gender:

(35) *macchina* f "car" / *automobile* f "car"

I think that the contrast between the lack of hyperonym effect in (34a-b) and the strong hyperonym effect in (35)-(27) can be explained if we entertain the hypothesis stated in (36), which I will call the *Basic Level Hyperonym Constraint*:

(36) *The Basic Level Hyperonym Constraint*
To be able to assign gender to its hyponyms, a hyperonym must be a *basic level term*

The concept of basic level term has been developed in cognitive psychology, in research on universal principles of categorization, notably in the work of Eleanor Rosch and colleagues (Rosch *et al.* 1976, Rosch 1978). A question asked in this kind of research was "How shall a thing be called?". This was the title of a seminal paper by Roger Brown, published in 1958 in the *Psychological Review* (quoted in Tversky 1986: 63). Barbara Tversky (1986:63) gives the following answer:

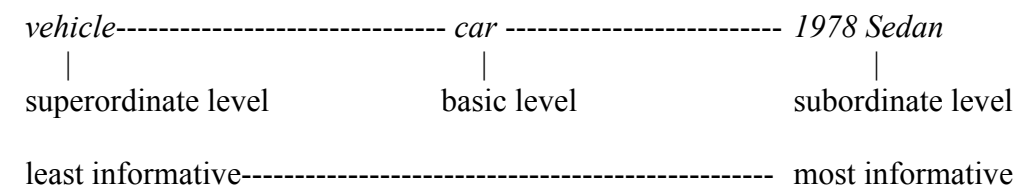
> «Most things have many names. This is a *table*, a *conference table*, a *brown wooden table*, a *piece of furniture*, and so on. Yet, when in the serious business of labelling the world to teach children how to talk, we all agree, and the child, too, that *table* is the proper name.» (Tversky 1986:63)

That means that there is a preferred level at which objects are called and recognized and classified; this level is considered the most useful level at which objects can be cognitively manipulated.

According to Rosch, «The task of category systems is to provide maximum information with the least cognitive effort» (Rosch 1978: 28), and it has been experimentally proved that «There is a trade-off between the informativeness of a category, and the number of categories or distinctions that we have to deal with» (Tversky 1986: 64).

Take the universe of vehicles as an example: the same object can be categorized at least in the three ways shown in (37): it can be seen as a *vehicle* at the superordinate level, as a *car* at the basic level, and as a *1978 Sedan* at a subordinate level.

(37) Three levels of categorization



Tversky observes that «Categorizing […] vehicles […] by make, model, and year certainly provides much more information than merely categorizing by *car, truck, motorcycle*, and so on, and that provides more information than categorizing simply as *vehicle* […]. This information, however, comes at the cost of the cognitive burden of remembering and distinguishing many different categories», and she wonders whether we can identifiy «a level in taxonomies where the benefits of information balance the costs of number of categories» (Tversky 1986: 64).

The studies by Rosch and colleagues have shown that «increasing specificity from the superordinate to the basic level leads to a large gain in informativeness, but further increases in specificity do not increase informativeness, but do increase the

mental burden of categories and distinctions» (Tversky 1986: 65). Therefore, objects are first seen or recognized as members of their basic category (Rosch 1978: 34-5). The basic level is the cognitively optimal level for categorizing new items, such as recent borrowings and neologisms. Tversky concludes that the answer to Roger Brown's question is the following:

> «Things are called at a level that is informative without imposing a burden of too many distinctions [i.e., the Basic Level]. Moreover, that level of categorization is preferred not just for naming, but in an impressive variety of tasks reflecting many aspects of human cognition» (Tversky 1986: 66).[6]

I propose that one of the tasks for which basic level categories are preferred is that of gender assignment.

The taxonomical hierarchies of physical objects investigated by Rosch and colleagues are the ones listed in (38):[7]

(38) Hierarchies of physical objects investigated by Rosch *et alii* (1976)

Superordinate Level	**Basic Level**	**Subordinate Level**
MUSICAL INSTRUMENT	guitar, piano, drum…	classical guitar, upright piano…
TOOL	hammer, saw, screwdriver…	claw hammer, cross-cutting hand saw…
CLOTHING	pants, socks, shirt…	Levis, knee socks…
FURNITURE	table, lamp, chair ….	kitchen table, desk lamp…
VEHICLE	car, bus, truck…	sports car, city bus…

The hierarchies comprise three levels: a Superordinate Level such as VEHICLE, a Basic Level such as *car, bicycle*… and a Subordinate Level which comprises subsets of the entities denoted by the Basic Level Terms.

It is interesting to observe that most of the Italian feminine nouns ending in *-o* listed in (18) denote subordinate entities, many from within the very same taxonomies investigated by Rosch and colleagues, and their feminine gender can be explained by appealing to inheritance of the gender of the noun denoting the respective basic level category. Some examples are given in (39):

[6] «Basic objects [have been] shown to be the most inclusive category for which a concrete image of the category as a whole can be formed, to be the first categorizations made during perception of the environment, to be the earliest categories sorted and earliest named by children, and to be the categories most codable, most coded, and most necessary in language» (Rosch *et al*. 1976: 382).

[7] I leave aside here the biological taxonomies, which raise problems that deserve a separate study in connection with gender assignment.

(39) Italian feminine nouns in *-o* and their Basic Level Hyperonyms

Basic Level Category	**Italian Basic Level noun with gender**	**Subordinate feminine nouns (gender assignees)**
car	*macchina* f	*cabrio* "kind of car", *torpedo* "kind of car"
shirt	*maglietta* f / *camicia* f	*polo* "polo shirt"
chair	*sedia* f	*sdraio* "deck chair"

I assume that *biro* "ballpoint pen" can be explained in a similar way, as in (40b), by appeal to the taxonomy in (40a):

(40) a. Taxonomy for writing instruments

Superordinate	**Basic Level Term**	**Subordinate**
WRITING INSTRUMENT	pen, pencil, crayon…	ballpoint pen, fountain pen …

b. Gender assignment for Italian *biro* "ballpoint pen"

Basic Level Category	**Italian Basic Level noun with gender**	**Subordinate nouns (gender assignees)**
pen	*penna* f	*biro* "ballpoint pen"

A similar explanation is available for *merino* "merino sheep", which very likely gets its feminine gender from the hyperonym *pecora* f "sheep".[8] For *lampo*, the remaining Italian feminine noun in *-o* from the list in (18), I believe that the most economical explanation is to assume inheritance of the gender of a deleted head noun. *Chiusura lampo* "zipper" is a N+N compound with the structure in (41):

(41) $[[\text{chiusura}]_N\ [\text{lampo}]_N]_N$
"zipper", lit. $[[\text{fastener}]_N\ [\text{flash}]_N]_N$

The compound regularly inherits the feminine gender of its head, *chiusura* "fastener"; when the head is deleted through ellipsis, the modifier *lampo* "flash" takes up the meaning of the whole compound and its gender.

Basic level term effects in gender assignment in the taxonomy of vehicles had already been reported in the literature, even though they had not been interpreted in the same way I am proposing. Corbett (1991: 76-77) reports data from Gouffé's (1971) study of French borrowings into Hausa. Hausa's relevant facts are summarized in (42).

[8] I will not discuss this case further because it involves discussion of the problems raised by biological taxonomies.

(42) Hausa

a.	2 genders:	masculine, feminine
b.	Semantic rules:	nouns denoting males are masculine, nouns denoting females are feminine
c.	Phonological rules:	nouns ending in *-aa* are feminine; nouns with other endings are masculine

Hausa has two genders, masculine and feminine, and the usual sex-based semantic rules; besides, it has a phonological rule by which nouns ending in *-aa* are feminine. But some borrowings are assigned feminine gender, even though they do not end in *-aa*. An examples is the one in (43):

(43)

Basic Level Category	**Basic Level noun with gender**	**Subordinate noun (gender assignee)**
car	*mootàa* f "car" (< English *motor*) [feminine by phonological rule]	*tàkàsiì* "taxi" (< *taxi*) f

As we can see from the data in (43), the phonological rule that would assign masculine gender is dominated by a semantic rule that assigns feminine gender to the borrowing meaning "taxi" by inheritance of the gender of its basic level hyperonym meaning "car". Particularly interesting for us is the minimal pair in (44):

(44) Hausa *pèežoò* < French *Peugeot*

pèežoò f "Peugeot car" vs. *pèežoò* m "Peugeot bicycle"
(data from Corbett 1991: 77)

When *pèežoò* is used to denote a car, it is feminine; when the same word is used to denote a bicycle, it is masculine, because in this case it doesn't fall within the scope of the semantic rule assigning feminine gender to nouns denoting cars.
I hope to have shown that inheritance of the gender from a hyperonym that is a Basic Level Term is a real process at work in gender assignment (at least in regard to borrowings).

A further observation I can make is that, while Basic Level Terms function as gender assigners to their hyponyms, Superordinate terms do not function as gender assigners to their basic level hyponyms. Robust evidence for this has been collected by Zubin and Köpcke (1986) for German (even though they did not limit their investigation to borrowings or neologisms), as we can see in (45), and the same happens in Italian, as we can see in (46).

(45) Superordinates do not assign gender to Basic Level terms in German (Zubin and Köpcke 1986)

Superordinate	**German Superordinate with Gender (shown by article)**	**German Basic Level Terms with Gender (shown by article)**
MUSICAL INSTRUMENT	*das Musikinstrument*	*die Harfe, das Klavier…*
TOOL	*das Werkzeug*	*der Hammer, die Hacke, das Skalpell...*
CLOTHING	*das Kleid / das Kleidungsstück*	*der Mantel, die Hose, das Hemd...*
FURNITURE	*das Möbel / das Möbelstück*	*der Tisch, die Lampe*
VEHICLE	*das Fahrzeug*	*der Wagen, der Bus...*

(46) Superordinates do not assign gender to Basic Level terms in Italian

Superordinate	**Italian Superordinate with Gender**	**Italian Basic Level Terms**
MUSICAL INSTRUMENT	*strumento* m	*chitarra* f "guitar", *pianoforte* m "piano"... *balalaika* f (from Russian)
TOOL	*strumento* m / *attrezzo* m	*martello* m "hammer", *sega* f "saw"…
CLOTHING	*capo d'abbigliamento* m	*camicia* f "shirt", *pantaloni* m "pants"…
FURNITURE	*mobile* m	*tavolo* m "table", *sedia* f "chair"… *consolle* f (from French)
VEHICLE	*veicolo* m / *mezzo di trasporto* m	*macchina* f "car", *autobus* m "bus"…

For Italian, where possible I have tried to exemplify the categories in (46) with data concerning borrowings; I did not find many suitable examples, very likely because Basic Level Terms are more commonly traditional nouns belonging to the native vocabulary, and very few recent borrowings gain Basic Level Term status, at least in a relatively stable language such as Italian. Most borrowed nouns denote very specific and peculiar entities, which have rather subordinate status: it makes sense to have a gender assignment constraint that makes them inherit gender from their Basic Level Hyperonym.

A proposed constraint on the format of possible semantic gender assignment rules is therefore the one in (47):

(47) A proposed constraint on the format of possible semantic gender assignment rules

Nouns denoting Subordinates may inherit gender from their Basic Level Hyperonym, but Basic Level Terms do not inherit gender from their Superordinate Hyperonym.

6. Conclusion

In this paper, I have investigated whether there is a dominance hierarchy between formal and semantic gender assignment rules / constraints, on the basis of Italian data. Gender assignment to loanwords and neologisms in contemporary Italian shows that when a conflict between a semantic gender assignment rule and a formal one arises, semantic rules prevail, contrary to what would be predicted according to the Optimal Gender Assignment Theory of Rice (2004, 2005). The paper has then proceeded to investigate what kinds of semantic rules exist, besides the well-known rule assigning masculine and feminine gender to nouns denoting male and female humans respectively. On the basis of Italian data, I have proposed that a constraint on semantic gender assignment rules is the "Basic Level Hyperonym Constraint", stating that rules that assign gender to a noun through inheritance of the gender of one of its hyperonyms exist only between Basic Level Hyperonyms and Subordinate nouns (and not, for example, between Superordinate nouns and Basic Level nouns, as evidence collected for German by Zubin and Köpcke 1986 had already shown).

References

Audring, Jenny, 2004, *Constraint Competition in Gender Assignment to Loanwords*, ms, Vrije Universiteit Amsterdam.

Berruto, Gaetano, 1976, *La semantica*, Bologna, Zanichelli.

Comrie, Bernard, 1999, "Grammatical Gender Systems: A Linguist's Assessment", *Journal of Psycholinguistic research* 28, 457-466.

Corbett , Greville, 1991, *Gender*, Cambridge, Cambridge University Press.

Corbett, Greville G. and Norman M. Fraser, 1999, "Default genders" in Barbara Unterbeck, Matti Rissanen (eds), *Gender in Grammar and Cognition: I. Approaches to Gender, II. Manifestations of Gender* (Trends in Linguistics. Studies and Monographs), Berlin-New York: Mouton de Gruyter, 55- 97.

Corbett, Greville G. and Norman M. Fraser, 2000, "Gender assignment: a typology and a model", in Gunter Senft (ed.), *Systems of nominal classification*, Cambridge, Cambridge University Press, 293-325.

De Mauro, Tullio, Mancini, Federico, Vedovelli, Massimo, and Voghera, Miriam, 1993*, Lessico di frequenza dell'italiano parlato* (LIP), Milano, Etaslibri.

Doleschal, Ursula, 1999, "Gender Assignment Revisited", in Barbara Unterbeck, Matti Rissanen (eds), *Gender in Grammar and Cognition: I. Approaches to Gender, II. Manifestations of Gender* (Trends in Linguistics. Studies and Monographs), Berlin-New York: Mouton de Gruyter, 117-165.

Doleschal, Ursula, 2004, *Genus als grammatische und textlinguistische Kategorie,* Muenchen, Lincom Europa.

Enger, Hans-Olav, Nesset, Tore and Rice, Curt (eds.), in press, *Lingua* special issue: Proceedings of the Oslo Workshop on Gender.

Fraser, Norman M. and Corbett, Greville G., 1995, "Gender, Animacy and Declensional Class Assignment: A Unified Account for Russian", *Yearbook of Morphology 1994*, 123-150.

Fraser, Norman M. and Corbett, Greville G., 1997, "Defaults in Arapesh", *Lingua* 103, 25-57.

LIP = De Mauro *et al.* 1993

Nesset, Tore, 2004, "Gender meets the Usage-Based Model: Four Principles of Rule Interaction in Gender Assignment", ms, CASTL-University of Tromsø (in press in Enger *et al*. (eds.))

Nitrola, Cristina, 1998, *Morfologia dei toponimi italiani*, tesi di laurea, Università dell'Aquila.

Poplack, Shana, Alicia Pousada and David Sankoff, 1982, "Competing influences on gender assignment: variable processes, stable outcome", *Lingua* 57, 1-28.

Rice, Curt, 2004, "Optimizing gender", ms, University of Tromsø (in press in Enger *et al*. (eds.))

Rice, Curt, 2005, "Optimizing Russian Gender: A Preliminary Analysis", in *Formal Approaches to Slavic Linguistics 13: The South Carolina Meeting*, ed. by Steven Franks, Frank Y. Gladnev and Mila Tasseva-Kurktchieva, Ann Arbor MI, Michigan Slavic Publications, 265-275.

Rosch, Eleanor, Mervis Carolyn B., Gray Wayne D., Johnson David M., Boyes-Braem Penny, 1976, "Basic objects in natural categories", *Cognitive Psychology* 8, 382-439.

Rosch, Eleanor, 1978, "Principles of Categorization", in *Cognition and Categorization*, ed. by Eleanor Rosch and Barbara B. Lloyd, Hillsdale NJ, Erlbaum, 27-48.

Steinmetz, Donald, 1986, "Two principles and some rules for gender in German: Inanimate nouns", *Word* 37, 189-217.

Thornton, Anna M., 2003a, "L'assegnazione del genere in italiano", in *Actas del XXIII Congreso Internacional de Lingüística y Filología Románica*, ed. by Fernando Sánchez Miret, vol. I, Tübingen, Niemeyer, 467-481.

Thornton, Anna M., 2003b, "L'assegnazione del genere ai prestiti inglesi in italiano" in *Italiano e inglese a confronto*, ed. by Anna-Vera Sullam Calimani, Firenze, Cesati, 57-86.

Tversky, Barbara, 1986, "Components and categorization", in *Noun classes and categorization*, ed. by Colette Craig, Amsterdam / Philadelphia, Benjamins, 63-75.

Zubin, David A. and Klaus-Michael Köpcke, 1986, "Gender and folk taxonomy: the indexical relation between grammatical and lexical categorization", in *Noun classes and categorization*, ed. by Colette Craig, Amsterdam / Philadelphia, Benjamins, 139-180.

Morphological Word Structure in English and Swedish: the Evidence from Prosody [1]

Renate Raffelsiefen
IDS Mannheim
raffelsiefen@ids-mannheim.de

Abstract

Trubetzkoy's recognition of a delimitative function of phonology, serving to signal boundaries between morphological units, is expressed in terms of alignment constraints in Optimality Theory, where the relevant constraints require specific morphological boundaries to coincide with phonological structure (Trubetzkoy 1936, 1939, McCarthy & Prince 1993). The approach pursued in the present article is to investigate the distribution of phonological boundary signals to gain insight into the criteria underlying morphological analysis. The evidence from English and Swedish suggests that necessary and sufficient conditions for word-internal morphological analysis concern the recognizability of head constituents, which include the rightmost members of compounds and head affixes. The claim is that the stability of word-internal boundary effects in historical perspective cannot in general be sufficiently explained in terms of memorization and imitation of phonological word form. Rather, these effects indicate a morphological parsing mechanism based on the recognition of word-internal head constituents.

Head affixes can be shown to contrast systematically with modifying affixes with respect to syntactic function, semantic content, and prosodic properties. That is, head affixes, which cannot be omitted, often lack inherent meaning and have relatively unmarked boundaries, which can be obscured entirely under specific phonological conditions. By contrast, modifying affixes, which can be omitted, consistently have inherent meaning and have stronger boundaries, which resist prosodic fusion in all phonological contexts. While these correlations are hardly specific to English and Swedish it remains to be investigated to which extent they hold cross-linguistically.

The observation that some of the constituents identified on the basis of prosodic evidence lack inherent meaning raises the issue of compositionality. I will argue that certain systematic aspects of word meaning cannot be captured with reference to the syntagmatic level, but require reference to the paradigmatic level instead. The assumption is then that there are two dimensions of morphological analysis: syntagmatic analysis, which centers on the criteria for decomposing words in terms of labelled constituents, and paradigmatic analysis, which centers on the criteria for establishing relations among (whole) words in the mental lexicon. While meaning is intrinsically connected with paradigmatic analysis (e.g. base relations, oppositeness) it is not essential to syntagmatic analysis.

[1] Parts of this material were presented at the workshop on word structure in Leipzig (April 2004), at the MMM5 in Fréjus (September 2005), at the universities in Tübingen (February 2005), Stony Brook, Princeton (March 2006) and the Institut für Deutsche Sprache in Mannheim (May 2006). I thank the audiences for valuable comments and criticism, in particular I wish to acknowledge Joachim Ballweg, Stig Eliasson, Christiane Fellbaum, Lutz Gunkel, Alice Harris, Robert Hoberman, Daniel Osherson, Hubert Truckenbrodt, Bernd Wiese, and Gisela Zifonun. Thanks to Roger Schwarzschild for discussing the ideas in section 4 with me. All errors are mine.

1. Introduction

The question of whether and how to divide words into morphological constituents is controversial. Some consider the word the smallest unit of grammatical analysis and reject segmentation altogether, in accordance with the traditional position in western linguistics (cf. Robins 1967:25). On such a view, morphological analysis amounts to recognizing relations among whole words in the mental lexicon, based on similarities in meaning and form (cf. Paul 1880). Once these relations are established, learners may detect recurrent patterns, allowing for the abstraction of schemas or 'correlative patterns' (Marchand 1969:) of the types illustrated in (1).

(1) a. [ri:zʌ́mpʃən]$_N$ 'resumption' - [ri:zú:m]$_V$ 'resume'
[əsʌ́mpʃən]$_N$ 'assumption' - [əsú:m]$_V$ 'assume' [Xʌ́mpʃən]$_N$ - [Xú:m]$_V$
[kənsʌ́mpʃən]$_N$ 'consumption' - [kənsú:m]$_V$ 'consume'

b. [ɔ:dǽsəti]$_N$ 'audacity' - [ɔ:déɪʃəs]$_A$ 'audacious'
[tɪnǽsəti]$_N$ 'tenacity' - [tɪnéɪʃəs]$_A$ 'tenacious' [Xǽsəti]$_N$ - [XéɪʃəS]$_A$
[mɛnd´æsəti]$_N$ 'mendacity' - [mɛndéɪʃəs]$_A$ 'mendacious'

c. [ʃáɪnəs]$_N$ 'shyness' - [ʃáɪ]$_A$ 'shy'
[káɪndnəs]$_N$ 'kindness' - [káɪnd]$_A$ 'kind' [Xnəs]$_N$ - [X]$_A$
[i:vənnəs]$_N$ 'evenness' - [i:vən]$_A$ 'even'

The knowledge of schemas such as [xʌ́mpʃən]$_N$ – [xú:m]$_V$ is prone to enhance the recognizability of relatedness between additional pairs (e.g. *subsumption* - *subsume*)and could be used productively to coin new words (e.g. ? [ɪgzʌ́mpʃən] '*exhumption'* based on knowing the verb [ɪgzú:m] *exhume*[2]).

While it would be possible to enrich the schemas by inserting word-internal (labeled) brackets, there is no clear motivation for doing so. Such labels would contribute neither to clarifying the conditions under which relations between existing words are recognized, nor to adequately restricting the conditions under which new words are created.

The position that only paradigmatic relations between whole words should be recognized is supported by the failure of attempts to provide clear criteria for identifying word-internal constituents (morphemes). Invocation of meaning is problematic as some apparently morphologically complex words cannot be divided into meaningful parts. Consider the verbs *undergó*, *undertáke* and *undermíne*, where final main stress indicates a bimorphemic structure (cf. sections 3 and 4), yet their meanings cannot be computed compositionally. Aronoff's (1976) proposal to define morphemes as crucially involving an arbitrary relation, not necessarily between form and meaning, but also between two forms allows for the identification of a morpheme *go* in *undergo* (cf. *undergo* - *underwent*) and a morpheme *take* in *undertake* (cf. *undertake* - *undertook*). Yet, this criterion does not allow for the identification of the remaining

[2] The coinage *exhumption* used in the context of cemeteries or graves is indeed attested in Google.

parts (e.g. *under*, *mine* in *undermine*), which do not exhibit irregular alternations with other strings (cf. Bochner 1993:30).[3]

Despite the absence of a coherent set of criteria determining the identification of morphemes it appears that reference to word-internal constituents in linguistic descriptions is characterized by certain tacit conventions. For English, these include the following:

(2) a. Recurring sound strings exhibiting recurring alternations and/or a common etymological origin are treated in a uniform manner (e.g. [su:m] (from Latin *su:mere* 'to take up') in the verbs *assume*, *consume* which alternate with [zu:m] in *resume*, *presume* and [sʌmp] in -*assumption, consumption*).

b. Stem constituents which correspond to independent words are distinguished from those which do not correspond to independent words (e.g. *fate* in *fateful* is associated with a boundary/category distinct from the one associated with *grate* in *grateful*)

c. Constituents which assimilate are distinguished from those which do not assimilate (e.g. *im* in *impolite* is associated with a boundary/category distinct from the one associated with *un* in *unpleasant*)

Apart from the suspect adherence to etymology in synchronic description there are grounds for questioning the relevance of any of the properties addressed in (2) for the analysis of word-internal morphological structure. Specifically, the evidence from boundary signals presented below suggests that morphological segmentation is determined by head recognition, which means that only the properties of heads can be relevant to segmentation. The central concern of this paper is then to argue for a non-uniform treatment of the case in (2a), where a head affix is recognized in some but not all words, as opposed to a uniform treatment of the cases in (2b,c), where head constituents, are recognized in all words.

The article is organized as follows. Section 2 describes the sort of phenomenon intended by the term "boundary effect". In section 3 I explore the distribution of these effects as a window on the morphological structure of English, concluding that there are two types of structure. Additional correlations pertaining to these two types, including semantic properties, are discussed in section 4. In section 5 I discuss some supporting evidence from Swedish. In section 6 I confront the findings of this study with previous results of psycholinguistic work (Hay 2001, 2002).

[3] An important empirical argument supporting the non-existence of word-internal constituent structure concerns the (alleged) invisibility of such structure to morphosyntactic processes. Invisibility has led researchers to postulate a number of constraints (cf. the “Bracket Erasure Convention” (Pesetsky 1979), “Lexical Integrity Hypothesis” (Lapointe 1981), “Atom Condition” (Williams 1981), “Morphological Island Constraint” (Botha 1981)). As noted by Anderson (1992) these conventions could be dispensed with if internal structure were not recognized to begin with.

2. Boundary effects

The notion of boundary effect used here is restricted to those deviations from canonical sound patterns which involve coinciding morphological and prosodic boundaries. Compare the form of the adjective in (3a), which represents regular sound patterns of English, with the form in (3b), which exhibits an internal boundary effect:

(3) a. [sə.bɔ́r.də.nət] 'subordinate'
b. [sʌ**b.**ɔ́r.bə.təl] 'suborbital'

The syllabification of the boldfaced consonant in (3b) is 'deviant' in that it is syllabified as a syllable coda rather than an onset, despite preceding a stressed vowel (cf. the regular syllabification of a consonant before a stressed vowel in (3a)). This deviation indicates a division of the word in two separate prosodic domains, which coincide with separate morphological constituents. It is because of this coincidence that the deviation in question qualifies as a boundary effect. Throughout this paper I will represent prosodic boundaries with round brackets and morphological boundaries with square brackets:

(4)	a.	(subordinate)	b.	(sub)(orbital)
		[subordinate]		[sub][orbital]

The occurrence of coinciding prosodic and morphological boundaries as in (4) indicates the satisfaction of alignment constraints of the type given in (5), where the (left or right) boundaries of some morphological category *GCat* must align with the (left or right) boundaries of some prosodic boundaries *PCat*. The letter "E" in (5) is used as a variable ranging over left and right boundaries:

(5) Align (GCat, E; PCat, E)

GCat stands for all grammatical (morphosyntactic) categories including word-internal categories such as root, stem and affix. PCat includes prosodic categories such as syllable, foot and pword (phonological word), as well as prosodic features. Alignment of the type described in (5) is henceforth referred to as GP-alignment.

The first task is then to analyse observable prosodic effects as in (3) by identifying the relevant alignment constraints, in order to arrive at the "underlying" GCat.[4] Before tackling this task a general remark is in order concerning the diagnostic value of using prosodic boundary effects as a window on morphological structure. That is, while the presence of prosodic boundary effects reliably indicates the presence of morphological boundaries, the absence of prosodic boundary effects does not necessarily indicate the absence of morphological boundaries. This is because alignment constraints can be crucially dominated by other constraints. For instance a high-ranking markedness constraint ONSET, which prohibits syllables without an onset, results in the

[4] The ultimate task is to identify the criteria for morphological analysis which yield the respective morphological boundaries. The question is whether the structure in (3b) is determined by the recognition of the prefix *sub-*, by the recognition of the stem *-orbital*, or by the recognition of both parts. That task is pursued in section 3.

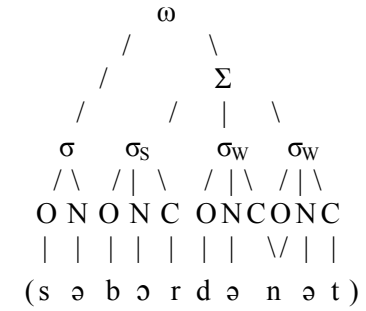
absence of boundary effects in cases where a consonant-final morpheme is followed by a vowel-initial morpheme. This sort of constraint domination can be illustrated by comparing French *subalpin* in (6a), which forms a single domain of syllabification, with English *subalpine* in (6b), which consists of two separate domains.

(6) a. s[y.ba]lpin 'subalpin' b. s[ʌb.æ]lpine 'sub-alpine'

The general absence of word-internal boundary effects in French in cases where a consonant-final morpheme is followed by a vowel-initial morpheme does accordingly not indicate that such combinations lack morphological structure. Rather, this absence indicates that alignment constraints, too, are violable[5] (cf. Prince & Smolensky 1993). While it is possible then that the string spelled *sub* in English *subordinate* is also a morpheme it cannot be the same type of morpheme as *sub* in *suborbital*. This is because whatever constraints dominate the relevant alignment constraint in English causing a fused prosodic structure in *subordinate* should have the same effect in *suborbital*. What can be said then is that the prosodic boundary effects observed in (3) clearly indicate some internal morphological boundary in (3b), where an analogous structure is ruled out in (3a).

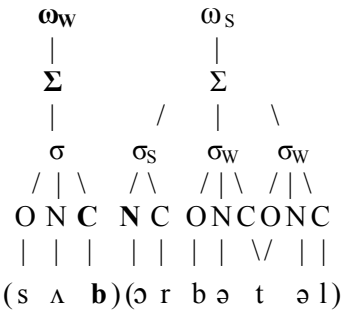
Turning now to the question of how to represent the prosodic contrast in (3) there is evidence for the representation in (7), where *sub* in *suborbital* constitutes a separate phonological word ω = pword, Σ = foot σ = syllable):

(7) a.



'subordinate'

b.



'suborbital'

According to the theory of Prosodic Phonology, pwords, feet and syllables are part of a hierarchy of prosodic constituents such that pwords rank immediately above feet, which in turn rank immediately above syllables (cf. Selkirk 1981, 1995, Nespor & Vogel 1986). Pwords differ from lower prosodic constituents in that they necessarily align with morphological constituents, which makes their proper identification especially relevant for the task at hand. Assuming the structures in (7), not only the "deviant" syllabification but also the "deviant" pretonic stress in *suborbital* can be explained in terms of general constraints on the Prosodic Hierarchy. Specifically, the coda syllabification of the prevocalic consonant satisfies the constraint Containment in (8a) and pretonic stress satisfies Headedness in (8b).

[5] This is not to deny a potential functional difference between the two cases illustrated in (6). Plausibly, in English *sub-alpine*, compared to French *subalpin*, access to morphological structure is facilitated by the prosodic signaling of the morphological boundaries, making it easier for the hearer to recognize the constituents in question. The impact of prosodic fusion on the recognition of word-internal constituents is addressed repeatedly below.

(8) a. Containment
A unit of a given level is exhaustively contained in the superordinate unit of which it is a part. (e.g. syllables are properly contained within feet)

b. Headedness
A given non-terminal unit is composed of one or more units of the immediately lower category. (e.g. a pword dominates at least one foot)

One type of evidence to support alignment of the prefix boundaries with pword boundaries (rather than just foot boundaries) concerns general constraints on syllable rhymes and the special status of pword-final consonants. In English, non-final rhymes contain no more than a single coda consonant preceded by a short vowel as in (9a), unless that coda consonant is a sonorant or *s* followed by a coronal or homorganic voiced obstruent in onset position (e.g. *shoulder*, *chamber*, *rooster*). Closed syllables with either a complex nucleus as in (9b) or a complex coda as in (9c) do not occur:

(9)	a.		b.		c.	
		h[ɛl].met 'helmet'		*h[i:l].met		*h[ɛlk].met
		[æt].las 'atlas'		*[eɪt].las		*[æst].las
		gr[ʌm].py 'grumpy'		*gr[oʊm].py		*gr[ʌlm].py

While being absent pword-internally such rhymes occur freely in pword-final position.

(10)		
	(h[i:**l**])$_{\omega}$ 'heal'	m[ɪl**k**])$_{\omega}$ 'milk'
	(w[eɪ**t**])$_{\omega}$ 'wait'	l[ɪs**t**])$_{\omega}$ 'list'
	(f[oʊ**m**])$_{\omega}$ 'foam'	[ɛl**m**])$_{\omega}$ 'elm'

A possible analysis of these patterns is that pword-final consonants are only phonetically, but not structurally, part of the syllable coda (cf. also section 5).[6] Significantly, the occurrence of such "extrasyllabic" consonants signals right pword boundaries, which necessarily align with morphological boundaries. Below I illustrate the occurrence of "Final-C effects" as signals of internal compound boundaries in (11a), of the boundary between a stem and a suffix in (11b), and of the boundary between a prefix and a stem in (11c).

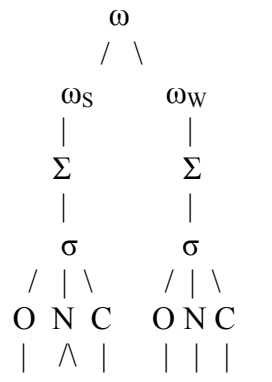
[6] Cf. Piggott (1999) and Harris and Gussman (2003), who also inform about the general acceptance of the 'final-onset view' in traditional non-western linguistics.

(11)

a.	b.	c.
(p[eɪ**n**])$_{\omega}$(pill)$_{\omega}$ 'pain pill' (m[ɪl**k**])$_{\omega}$(bar)$_{\omega}$ 'milk bar'	(p[eɪ**n**])$_{\omega}$(ful)$_{\sigma}$ 'painful' (m[ɪl**k**])$_{\omega}$(man)$_{\sigma}$ 'milkman'	(p[oʊ**st**])$_{\omega}$(doctoral)$_{\omega}$ 'postdoctoral' ([æn**t**])$_{\omega}$(arctic)$_{\omega}$ 'antarctic'

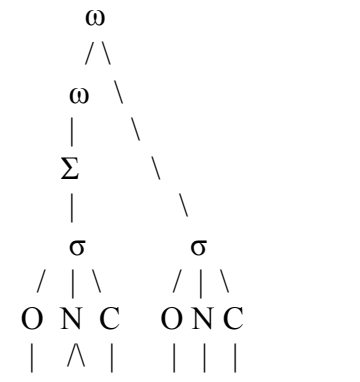
The data in (11) illustrate the significance of prosodic boundaries for signaling not only the presence of morphological boundaries, but for indicating the sort of morphological category involved. Whereas both *pill* in *pain pill* and *ful* in *painful* function as heads in that they determine the category of the respective complex word the latter has historically developed into an affix, which can no longer stand for the whole word. Such constituents are henceforth referred to as 'head affixes', whereas the term 'head' is reserved for constituents having the same category as the complex word. Significantly, prosodic structure reflects this syntactic contrast in English in that a head, but not a head affix, forms a separate pword as is shown in (12). [7] The labeling of the respective left constituents in (12) will be discussed in section 3.1.

(12) a.



[[pain][pill]$_{HEAD}$]$_{N}$
'pain pill'

b.

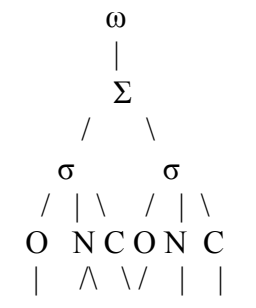


[[pain][ful]$_{HEAD\ AFFIX}$]$_{ADJ}$
'painful'

Although the Final-C effect indicates the presence of a word-internal pword boundary in (12b) it is often assumed that the affix is integrated, forming a trochaic foot together with the stem (cf. Burzio 1993). However, there is a systematic correlation between Final-C effects and Containment effects, which supports the non-integration of suffixes. Containment of syllable-structure is indicated by the phonetic contrasts between simplexes and suffixed words indicated in (13). A consonant between a stressed and an unstressed nucleus is regularly weak, which indicates foot-internal ambisyllabicity as in (13a). The relative strength of the intervocalic consonant marked by the superscript "f" (fortition) in (13b) indicates that it is strictly syllable-initial, which indicates satisfaction of Containment (cf. Umeda & Coker 1974). This analysis is corroborated by the "deviant" vocalic length in (13b), which signals foot-final (rather than foot-internal) position.

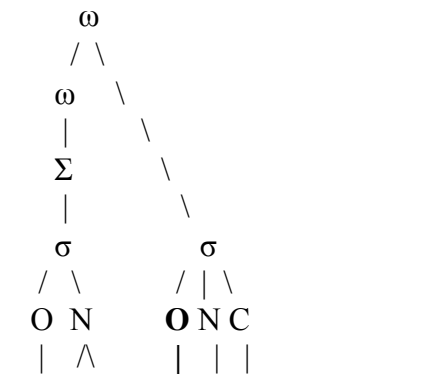
[7] Regular GP-alignment constraints yield the structure in (12a), where the pwords in a compound are dominated by an additional pword. However, given the lack of independent prosodic motivation for this topmost pword node (e.g. reference to this node is neither sufficient nor necessary for stating relative prominence) I will assume a convention by which multiple adjacent pword boundaries are reduced to single boundaries (e.g. $((X)_{\omega}(Y)_{\omega})_{\omega}$ => $(X)_{\omega}(Y)_{\omega}$).

(13) a.



[minus] ‘minus’

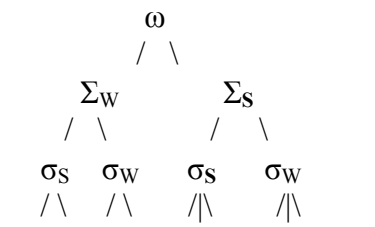
b.



[[shy][ness]$_{\text{HEAD AFFIX}}$] ‘shyness’

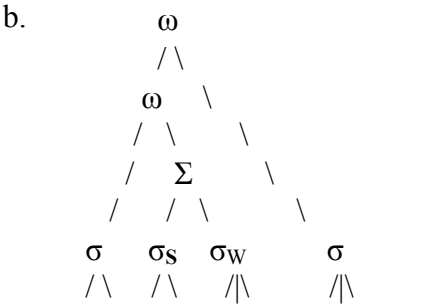
More salient Containment effects concerning foot structure can be observed in polysyllabic words. The simplex in (14a) illustrates regular foot structure in English, with stress on a closed penultimate syllable. The deviant stress pattern in (14b) indicates the presence of an internal pword boundary, which confines the domain of foot construction. Again, the prosody signals not only the presence of a morphological boundary but also indicates the types of constituents involved (i.e. stem plus head suffix):

(14) a.



[rhododendron]
‘rhododendron’

b.



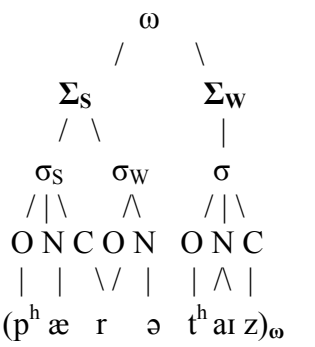
[[environ][ment]$_{\text{HEAD AFFIX}}$]
‘environment’

Assuming the adequacy of the representations in (12) - (14), the non-integration of head suffixes could be described in terms of a separate GP-alignment constraint, aligning the boundaries of head suffixes with syllable boundaries. However, positing such a constraint misses the generalization that the prosodic organization of English head affixes consistently reflects independently motivated constraints on parsing segmental material into prosodic constituents. That is, strings of segments are parsed into syllables, depending on the sonority structure of the string. For instance, the head affixes illustrated in (12) - (14), all of which consist of CVC(C) strings, are parsed into single syllables. Not being dominated by a separate pword, these syllables are unfooted, as reflected in the reduced vowels. Single syllables are footed only under specific (segmental) conditions, including the occurrence of the fricative [h] in onset position (cf. the suffix -*hood*). Here the foot is stabilized by a constraint aligning [h] with foot-initial position (cf. Davis and Cho 2003) in conjunction with a constraint prohibiting the deletion of segments. Given that in English both the syllabification and the pedification

of head affixes are determined by segmental (consonantal) structure there is a further generalization that the only prosodic constituent involved in English GP-alignment is the pword.

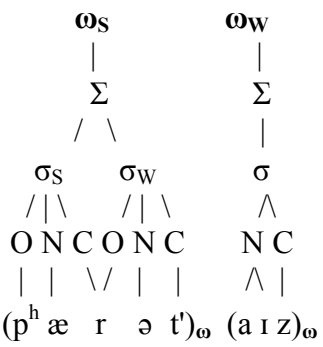
The distinct treatment of heads and head affixes proposed here (i.e. alignment of head boundaries, but not head affix boundaries, with prosodic boundaries) is supported by the striking contrast in the syllabification of stem-final *t* in (15a) versus (15b). Whereas aspiration (strengthening) in (15a) indicates regular onset syllabification before a stressed nucleus the glottalization (weakening) in (15b) in the same segmental environment indicates the presence of a following boundary:

(15) a.



[[parrot][ize]HEAD AFFIX]
'parrotize'

b.



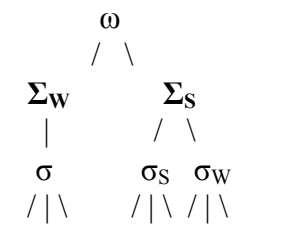
[[parrot][eyes]HEAD]
'parrot eyes'

The generalization emerging from a comparison of the pairs in (12) and (15) is that both members of a compound, including the relatively weak head, form separate pwords regardless of their segmental structure. By contrast, head suffixes are separate from the pword of the stem only if they begin with a consonant-vowel sequence. Vowel-initial suffixes like *-ize* or consonantal suffixes like *-th* are integrated into the pword of the stem, presumably to satisfy phonological markedness constraints (cf. the discussion of French *subalpin*).[8]

The distinct prosodic organization of identical segmental material in (15) accounts not only for the aspiration versus glottalization of stem-final *t* but also indicates distinct sources for the prominence on the initial syllable. In (15a) initial main stress is attributed to the strength of the initial foot within the pword whereas in (15b) initial main stress is attributed to the strength of the initial pword within compounds. Evidence for this distinction comes from cases where the final foot is trochaic. Here the rule that the initial pword in a compound is strongest still holds, regardless of the pword-internal foot structure (cf. 16b). By contrast, *within* pwords a final foot consisting of more than one syllable attracts main stress as in (16a).

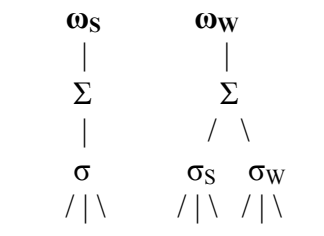
[8] The integration of consonantal suffixes into the pword of the stem can be inferred from the fact that suffixed words like *truth* rhyme perfectly with simplexes like *tooth*. The relevant phonological constraints dominating alignment (thereby causing the absence of boundary effects) concern the requirement that all segments must be parsed into syllables along with a constraint on minimal sonority of syllable nuclei.

(16) a.



[[gest][ation]$_{\text{HEADAFFIX}}$]
'gestation'

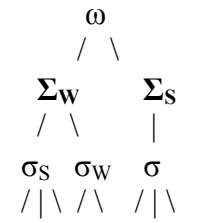
b.



[[gas][station]$_{\text{HEAD}}$]
'gas station'

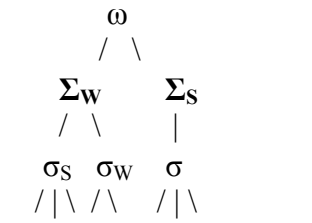
As a result of the phonologically conditioned fusion, the noun *gestation*, which contains a head suffix, is prosodically indistinguishable from a simplex like *dalmation*. Similarly, the noun *cameleer*, which also contains a head suffix, is prosodically indistinguishable from the simplex *cavalier*: in both nouns the final foot attracts main stress because it contains a high tense vowel. Weak stress on the final foot in *camel ear* is then again a boundary signal, indicating that relative prominence follows the compound rule.

(17) a.



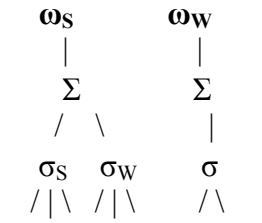
[cavalier]
'cavalier'

b.



[[camel][eer]$_{\text{HEADAFFIX}}$]
'cameleer'

c.



[[camel][ear]$_{\text{HEAD}}$]
'camel ear'

The data in (15) to (17) indicate that the integration of head affixes due to higher-ranking phonological constraints is complete. That is, apart from possible paradigm uniformity effects, which are independently motivated (cf. the end of this section), words with integrated head affixes are precisely like simplexes.

Prosodic structure offers cues not only to the morphological contrast between affixes and non-affixes but also indicates distinctions among affixes. Consider the contrast in the pronunciation of the word-initial syllables as transcribed by Wells (2000). Simplexes with pretonic vowels spelled <e> are consistently represented with two variants, one with schwa and one with a short raised vowel as in (18a). Significantly, there are two distinct patterns of deviation from that structure, both of which qualify as boundary effects. One type, illustrated in (18b), contains a prefix transcribed with a stressed long vowel and is consistently represented with a single form. This type matches the examples *suborbital*, *postdoctoral*, and *antarctic*, where the prefix forms a separate pword. The other type, illustrated in (18c), is consistently represented with three variants, two of which match the simplex patterns. The third variant, boldfaced in (18c), deviates from the simplex patterns in that it is transcribed with a tense, long vowel in prestress position, a structure henceforth referred to as "Final Nucleus Enhancement". The relevant form differs from the case illustrated in

(18b) not only in its association with variants, but also in that the vowel, though long and tense, is not marked for stress. The absence of stress indicates that the prefix does not form a separate pword in (18c), thereby contrasting with the type in (18b).

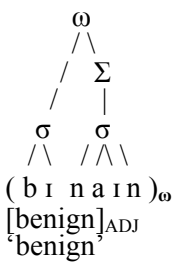
(18)	Phonetic transcriptions (Wells 1990):	Prosodic representations:	
a.	[bənáɪn]$_A$, [bɪnáɪn]$_A$ [rəg´ætə]$_N$, [rɪg´ætə]$_N$ [bəlú:gə]$_N$ [bɪlú:gə]$_N$	(benígn)$_ω$ (regátta)$_ω$ (belúga)$_ω$	‘benign’ ‘regatta’ ‘beluga’
b.	**[rì:bɝ́:θ]**$_N$ **[prì:kǽn^{t}sərəs]**$_A$ **[dì:m´ɪstɪfàɪ]**$_V$	(rè)$_ω$(birth)$_ω$ (prè)$_ω$(cáncerous)$_ω$ (dè)$_ω$(mýstify)$_ω$	‘rebirth’ ‘precancerous’ 'demystify'
c.	[bəgét]$_V$, [bɪg´ɛt]$_V$, **[bi:gɛ́t]**$_V$ [dədú:s]$_V$, [dɪdú:s]$_V$, **[di:dú:s]**$_V$ [rənú:]$_V$, [rɪnú:]$_V$, **[ri:nú:]**$_V$ [prəzú:m]$_V$, [prɪzú:m]$_V$, **[pri:zú:m]**$_V$	((be)$_σ$(gét)$_ω$)$_ω$ ((de)$_σ$(dúce)$_ω$)$_ω$ ((re)$_σ$(new)$_ω$)$_ω$ ((pre)$_σ$(súme)$_ω$)$_ω$	‘beget’ ‘deduce' ‘renew’ ‘presume’

The boundary effect in question is most salient in words which include an intervocalic *s*C-cluster. Such clusters are regularly heterosyllabic as in (19a), where the initial syllable is closed, but they are syllable-initial when preceded by a prefix forming a separate pword as in (19b). In (19c), "Final Nucleus Enhancement" correlates with the tautosyllabic syllabification of the cluster to indicate hat the prefix is not integrated, although it does not form a separate pword.

(19)				
	a.	s[əs.p]éct]$_V$	(suspect)$_ω$	'suspect'
	b.	r[ì:sp]éll]$_V$	(re)$_ω$(spell)$_ω$	'respell'
	c.	r[i: sp]éct]$_V$	((re)$_σ$(spect)$_ω$)$_ω$	'respect'

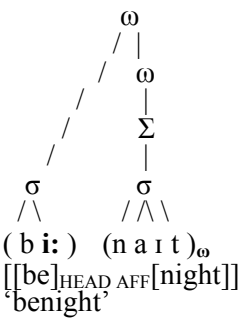
"Final Nucleus Enhancement", like all other effects discussed here, is sensitive to pword structure in that it occurs immediately before a pword boundary (cf. (20)):[9]

(20) a.



[benign]$_{ADJ}$
‘benign’

b.



[[be]$_{HEAD}$ $_{AFF}$[night]]
‘benight’

[9] Reference to a following pword boundary is also essential to vowel tenseness observed in the compound *béll*[i]*bànd* 'bellyband' or the affixation *béll*[i]*ful* 'bellyful', as opposed to lack of tenseness in *béll*[i]*còse* 'bellicose'.

Turning now to the question of what the contrast in the prosodic structure of the prefixes reveals about the underlying morphological structures, we find a correlation between prosodic and morphosyntactic properties. Specifically, the prefixes in (18b), (19b), which are stressed and form separate pwords, never affect the combinatory properties of the complex word and can accordingly be omitted without affecting grammaticality (e.g. *(re)birth of a nation, (pre)cancerous lesions*). They are henceforth referred to as modifying prefixes. By contrast, the prefixes in (18c) and (19c), which are unstressed and marked by Final Nucleus Enhancement, cannot be omitted, a property shared with head suffixes. Additional motivation for the analysis of these prefixes as head affixes concerns their association with specific syntactic categories. In English, such prefixes are primarily associated with verbs, for which the boundary effects in question are consistently marked in Wells (2000)[10], but also with abstract nouns and prepositions.

Recall that the prosodic contrast between heads and head suffixes observed in pairs like *pain pill* versus *painful* is accompanied by the fact that head boundaries are always signaled whereas the boundaries of head suffixes are signaled only when the suffix begins with a consonant. Otherwise the suffix is integrated into the pword of the stem. The question then arises of whether the prosodic contrast between modifying prefixes and head prefixes observed in (18) and (19) correlates with a similar difference in phonological sensitivity. In fact, there are phonological conditions under which fusion is systematic for head prefixes, whereas modifying prefixes consistently form separate pwords. First, head prefixes integrate when preceding an unstressed syllable, forming a trochaic foot together with that syllable as is shown in (21a). Phonologically, such fused structures become indistinguishable from simplexes. Consonants are ambisyllabic when preceded by a stressed vowel and followed by an unstressed vowel (i.e. in foot-internal position). Vowels are lax when followed by two or more syllables, the first of which is unstressed ("Trisyllabic Laxing" as in *p*[ˊɛ]*lican* (**p*[íː]*lican*) 'pelican', [ˊæ]*necdòte* (*[éɪ]*necdòte*) 'anecdote). "Final Nucleus Enhancement" observed in (18c) and (20b) is accordingly restricted to pretonic position.

(21)	a.	[[re]$_{\text{HEADPREF}}$[concíle]]$_{\text{VERB}}$	=>	(r[ɛ́]concìle)$_{\omega}$	'reconcile'
		[[de]$_{\text{HEADPREF}}$[legáte]]$_{\text{VERB}}$	=>	(d[ɛ́]legàte)$_{\omega}$	'delegate'
		[[pre]$_{\text{HEADPREF}}$[dicáte]]$_{\text{VERB}}$	=>	(pr[ɛ́]dicàte)$_{\omega}$	'predicate'
	b.	[[be]$_{\text{HEADPREF}}$[líttle]]$_{\text{VERB}}$	=>	((b[iː])$_{\sigma}$(líttle)$_{\omega}$)$_{\omega}$	'belittle'
		[[de]$_{\text{HEADPREF}}$[líver]]$_{\text{VERB}}$	=>	((d[iː])$_{\sigma}$(líver)$_{\omega}$)$_{\omega}$	'deliver '
		[[re]$_{\text{HEADPREF}}$[cóver]]$_{\text{VERB}}$	=>	((r[iː])$_{\sigma}$(cóver)$_{\omega}$)$_{\omega}$	'recover'
		[[pre]$_{\text{HEADPREF}}$[váricàte]]$_{\text{VERB}}$	=>	((pr[iː])$_{\sigma}$(váricàte)$_{\omega}$)$_{\omega}$	'prevaricate'

The representations in (21) are intended to show that given identical morphological structures consisting of a uniform head prefix and a root, the independent contrast in the stress pattern of the root could account for fusion in (21a) vis-à-vis the occurrence of

[10]That is, when occurring in verbs, abstract nouns or prepositions, the prefixes *be-*, *re-*, *de-*, and *pre-*, are consistently transcribed with a tense vowel in Wells (2000) (e.g. *prepare*, *decubitus*, *behind*). This does not hold for corresponding initial strings in words belonging to other categories (e.g. *benign, beluga, regatta*).

boundary effects in (21b).[11] The plausibility of this analysis lies in the observation that there is an independently motivated phonological markedness constraint (Foot Binarity), whose ranking above the relevant GP-alignment constraint would cause fusion in (21a), but not in (21b).[12] While this analysis may reflect the (historical) cause of the restricted occurrence of boundary effects in (21) and may explain the restriction of native head prefixation to stress-initial base words it is questionable that the morphological structures in (21a) and (21b) are indeed identical in the minds of speakers. Rather, it is likely that (historical) prosodic fusion in (21a) affects the morphological analyzability of the verbs, in particular, the recognizability of the prefix. The point here is to demonstrate a contrast between head prefixes, which exhibit boundary effects only under specific phonological conditions (cf. (21b) versus (21a)), and modifying prefixes, which consistently form separate pwords, regardless of the phonological structure of the stem as is shown in (22):

(22)	$[[re]_{MODPREF}[combíne]_X]_{VERB}$	=>	$(r[ì:])_\omega(combíne)_\omega$	'recombine '
	$[[de]_{MODPREF}[compóse]_X]_{VERB}$	=>	$(d[ì:])_\omega(compóse)_\omega$	'decompose '
	$[[pre]_{MODPREF}[concéive]_X]_{VERB}$	=>	$(pr[i:])_\omega(concéive)_\omega$	'preconceive '

The phonological evidence for the distinct prosodic organizations in (21) and (22) is supported by relative prominence patterns. Recall that a final monosyllabic foot within a polysyllabic pword is usually weak. This rule also applies to verbs, except that a final monosyllabic foot is strong if the verb ends in a consonant cluster (e.g. *àpprehénd*, *rèsurréct*). Final main stress in verbs ending in a single consonant (or none) as in (22) indicates then that relative prominence is determined not with reference to pword-internal foot structure, but follows the rule in (23) (cf. also previous examples like *sùbálpine, àntárctic, pòstdóctoral*):

(23)

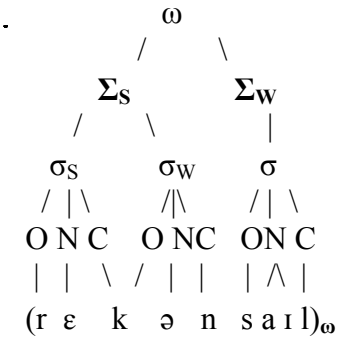
If:		Then:	
ω ω		$\omega_{\mathbf{W}}$ $\omega_{\mathbf{S}}$	
| |		| |	
$([X]_{MODPREF})_\omega([Y]_{HEAD})_\omega$		$([X]_{MODPREF})_\omega([Y]_{HEAD})_\omega$	
where head = verb, adjective			

The rule in (23) is typical of relative prominence rules in English in that it refers both to prosodic structure and to morphological categories. Significant for the purposes of this paper is the reference to word-internal pwords, which indicates the presence of complex morphological structure. Specifically, the stress pattern (final main stress in a polysyllabic verb) indicates the presence of a modifying prefix in (24b), in contrast to the verb in (24a), which exhibits the regular relative prominence relations within pwords.

[11]The stress pattern is independent in that it can be predicted on the basis of the total number of syllables in the root and the weight of the final syllable.

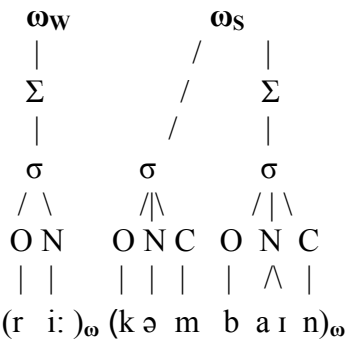
[12]In (21b), a trochaic foot involving the prefix could be formed only if the root-initial foot were deleted. This, however, would violate higher-ranking PARSE-constraints and is therefore ruled out. As a result, GP-alignment prevails in (21b).

(24) a.



[[re]$_{\text{HEADPREFIX}}$[concíle]]
'reconcile'

b.

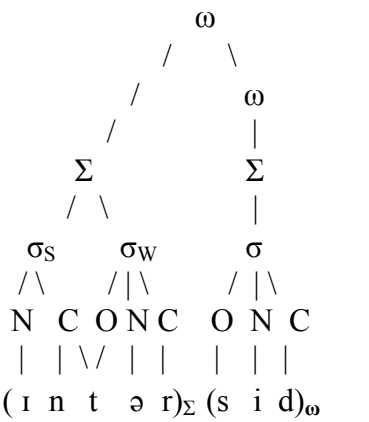


[[re]$_{\text{MODPREFIX}}$[combíne]]
'recombine'

It is, however, not the case that final main stress in a polysyllabic verb (ending in maximally one consonant) necessarily indicates the presence of a modifying prefix. Final main stress is also quite regular in verbs with a disyllabic head prefix followed by a monosyllabic root as illustrated in (25a) (cf. also *overcóme, undermíne*).[13] The representations in (25) are based on the assumptions that modifying prefixes regularly form separate pwords due to GP-alignment (i.e. Align (MODPREF,E;ω,E) whereas the prosodic form of head prefixes is determined by segmental structure alone. That is, whereas the presence of the initial foot in (25b) is an independent Headedness effect it results from the occurrence of two sonority peaks in the relevant segment string in (25a).

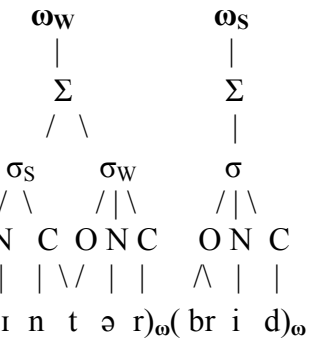
Word-final main stress in (25a) would then be accounted for under the further assumptions that roots (i.e. the sister constituents of head affixes within words) form separate pwords and that for any combination of distinct prosodic constituents, the higher-ranking constituent is more prominent. By contrast, in (25b) final main stress is determined by rule (23), which yields weak prominence on the modifying prefix.[14]

(25) a.



[[inter]$_{\text{HEADPREFIX}}$[cede]$_{\text{ROOT}}$]
'intercede'

b.



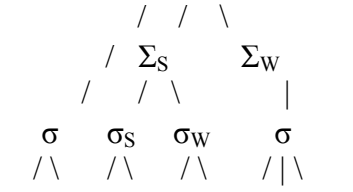
[[inter]$_{\text{MODPREFIX}}$[breed]]
'interbreed'

[13]Exceptions to this generalization concern verbs which relate to nouns with regular initial main stress and can accordingly be analysed as paradigm uniformity effects (e.g. *súpervìse - súpervìsor*).

[14]The verb *persevére*, which includes neither a modifying prefix nor a recognizable head prefix, illustrates a third source for final main stress. Recall that the last foot in a polysyllabic words is strong if it dominates a syllable with a high tense nucleus (e.g. *cavalíer*).

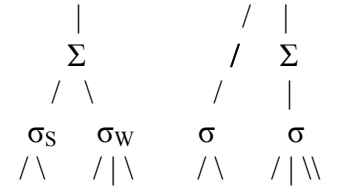
The distinction between these structures is again supported by cases of phonologically conditioned integration. Recall that head prefixes, but not modifying prefixes, integrate when preceding an unstressed syllable. As a result, we find a clear contrast in the stress patterns of the verbs in (26). In (26a), relative prominence is determined by the regular rule applying within pwords, which says that a monosyllabic final foot is weak. In (26b), relative prominence is determined by rule (23), which yields weak prominence on a modifying prefix.

(26) a.



(ɪ n t ɝ p ə l eɪ t)$_{\omega}$
[[inter]$_{\text{HEADPREFIX}}$[pelláte]]
'interpellate'

b.



(ɪ n t ə r)$_{\omega}$(k ə n ɛ k t)$_{\omega}$
[[inter]$_{\text{MODPREFIX}}$[connect]]
'interconnect'

Additional examples are given in (27). In (27a), the bisyllabic head prefix is not integrated because it precedes a stressed syllable. In (27b), integration before a stressless syllable is regular. In (27c) we see that modifying prefixes consistently form separate pwords, regardless of phonological factors:

(27) a.	[[inter]$_{\text{HEADPREFIX}}$[díct]$_{\text{ROOT}}$]$_{\text{WORD}}$ =>	((ìnter)$_{\Sigma}$(díct)$_{\omega}$)$_{\omega}$	'interdict'[15]
	[[inter]$_{\text{HEADPREFIX}}$[véne]$_{\text{ROOT}}$]$_{\text{WORD}}$ =>	((ìnter)$_{\Sigma}$(véne)$_{\omega}$)$_{\omega}$	'intervene'
	[[inter]$_{\text{HEADPREFIX}}$[fére]$_{\text{ROOT}}$]$_{\text{WORD}}$ =>	((ìnter)$_{\Sigma}$(fére)$_{\omega}$)$_{\omega}$	'interfere'
b.	[[inter]$_{\text{HEADPREFIX}}$[rogáte]$_{\text{ROOT}}$]$_{\text{WORD}}$ =>	(intérrogàte)$_{\omega}$	'interrogate'
	[[inter]$_{\text{HEADPREFIX}}$[poláte]$_{\text{ROOT}}$]$_{\text{WORD}}$ =>	(intérpolàte)$_{\omega}$	'interpolate'
	[[inter]$_{\text{HEADPREFIX}}$[caláte]$_{\text{ROOT}}$]$_{\text{WORD}}$ =>	(intércalàte)$_{\omega}$	'intercalate'
c.	[[inter]$_{\text{MODPREFIX}}$[depénd]$_{\text{HEAD}}$]$_{\text{WORD}}$ =>	(ìnter)$_{\omega}$(depénd)$_{\omega}$	'interdepend'
	[[inter]$_{\text{MODPREFIX}}$[diffúse]$_{\text{HEAD}}$]$_{\text{WORD}}$ =>	(ìnter)$_{\omega}$(diffúse)$_{\omega}$	'interdiffuse'
	[[inter]$_{\text{MODPREFIX}}$[reláte]$_{\text{HEAD}}$]$_{\text{WORD}}$ =>	(ìnter)$_{\omega}$(reláte)$_{\omega}$	'interrelate'

An additional context for the integration of head prefixes concerns S-V junctures, where S is any segment and V is a vowel. In (28a) it is shown that (for some speakers) C-V contexts cause fusion for head prefixation, but not for modifying prefixation (cf. MacCarthy 1945). In (28b) I illustrate the analogous contrast for V-V contexts:

(28) a.	[[en]$_{\text{HEADPREFIX}}$[áble]$_{\text{ROOT}}$]$_{\text{VERB}}$ =>	(e[.]nable)$_{\omega}$	'enable'
	[[un]$_{\text{MODPREFIX}}$[áble]$_{\text{HEAD}}$]$_{\text{WORD}}$ =>	(ùn)$_{\omega}$(áble)$_{\omega}$	'unable'
b.	[[re]$_{\text{HEADPREFIX}}$[áct]$_{\text{ROOT}}$]$_{\text{VERB}}$ =>	(r[i]áct)$_{\omega}$	'react'
	[[re]$_{\text{MODPREFIX}}$[áctivàte]$_{\text{HEAD}}$]$_{\text{WORD}}$ =>	(r[ì:])$_{\omega}$(áctivàte)$_{\omega}$	'reactivate'

[15] This verb is relevant here when pronounced [ɪntərdaɪt], with a single final consonant.

As a result of fusion, *enable* is prosodically on a par with simplexes like *finagle*. Similarly, the hiatus in *react*, transcribed only with a tense, short [i] in Wells (2000), is indistinguishable from the hiatus in simplexes like *m*[i]*ánder* 'meander'.
The effects of fusion for head prefixation can generally be demonstrated only with loanwords. For native head prefixation, contexts which would induce fusion usually yield gaps. There are accordingly no cases of native head prefixation based on a word beginning with a vowel or with a stressless syllable. The results of this section are summarized below:

- There is a systematic correlation between Containment effects, Final-C effects, Final Nucleus Enhancement, Headedness effects and Relative Prominence effects in English. All effects refer to pword boundaries, which necessarily coincide with morphological boundaries.

- Each member of a compound and modifying prefixes form separate pwords in English, regardless of phonological contexts. As a result each of these constituents exhibit consistent boundary effects.

- Unlike modifying affixes, head affixes do not form separate pwords and integrate into the pword of the stem under specific phonological conditions. The segments of non-integrated head affixes are parsed into syllables and possibly feet, in accordance with general rules of prosodic parsing. Integration results in prosodic identity with the structure of simplexes.

The summary calls for a few clarifying remarks. The claim that certain types of morphological structures exhibit consistent boundary effects does not rule out the possibility that individual words may fuse prosodically into a single pword. Such fusion is illustrated in (29a) vis-à-vis the unfused structures in (29b):

(29)	a.		b.	
		[kʌ́bərd] 'cupboard'		[klíp.bɔ̀:rd] 'clipboard'
		[də.síntəgreɪt] 'disintegrate'		[dɪ̀s.íntrəst] 'disinterest'
		[ná:nsəns] 'nonsense'		[nɑ̀:n.stɑ́:p] 'nonstop

The phenomenon illustrated in (29a) will be referred to as "High Frequency Fusion" because high token frequency appears to be a necessary (but by no means sufficient) prerequisite for the (historical) loss of boundary signals. It is important to properly identify cases of High Frequency Fusion, to distinguish such cases from cases of regular fusion resulting from the domination of GP-alignment constraints by phonological markedness constraints.

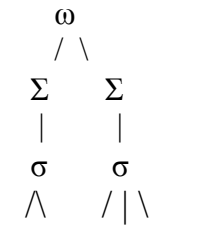
The claim that certain constituents exhibit consistent boundary effects does not mean that the effects are equally salient. The Containment effect in the compound *night rate*[16] vis-à-vis the simplex *nitrate* is perhaps always easily perceived, regardless of register, whereas the phonetic contrast between the compound *oxe-eyed* and the simplex *oxide* all but vanishes in fast speech (cf. Jones 1956:102). This difference in perception is due to the salient allophony characteristic of *t* (e.g. *nigh*[t']*rate* 'night rate' with

[16]The effect is that *t* preceding the *r* is contained within the initial constituent *night* and hence syllabified in coda position, rather than syllabified in onset position to form a cluster *tr*.

glottalized *t* indicative of syllable-final position versus *ni*[t^h]*rate* 'nitrate' with aspirated *t* indicative of syllable- and foot-initial position) compared to the lack of special allophones associated with the cluster [ks]. Such differences in salience are irrelevant to the argumentation, which rests on the demonstration that there are some (phonological) contexts and some register, presumably careful though not hyperarticulated speech, where boundary effects exist.

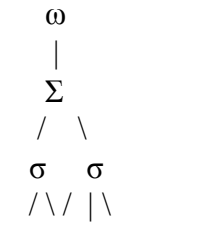
In general, careful investigation of the phonological context is essential for the proper identification of boundary effects. For instance, foot stability may qualify as a Headedness effect, indicative of the presence of pword boundaries but may also be contextually determined. Compare the lack of stress of the final syllable in the nouns in (30a), as opposed to the stress on the corresponding syllable in (30b), which could be diagnosed as a Headedness effect indicative of the structures $(\text{hém})_{\omega}(\text{lòck})_{\omega}$ versus $(\text{hámmock})_{\omega}$. However, the relevant contrast in stress could also be attributed to the contrast in syllable structure, demonstrated with the well-known variants in the pronunciation of Arab in (30). In (30a), the heaviness of the initial syllable allows for the following consonant to be syllabified in strictly syllable-initial position, which in turn is necessary for forming a separate foot (cf. the constraint Containment in (8a)). When ending in a non-coronal obstruent such feet appear to be fairly stable (cf. Fidelholtz 1967). In (30b), the intervocalic consonant necessarily closes the initial stressed syllable to ensure bimoraicity. Being ambisyllabic, that consonant can occur only foot-internally, but not foot-initially, which indicates the representation in (30b):

(30) a. hémlòck



'Arab'

b. hámmock



'Arab'

The evidence from stress consequently does not motivate internal pword boundaries for words like *hémlòck*, *shámròck* or *wédlòck*, where the initial syllable is closed. Similarly, word-final stress in (31a) does not motivate the presence of internal pword boundaries but is sufficiently motivated by the presence of [h] in onset position. The necessary alignment of [h] with foot-initial position mentioned above apparently stabilizes the word-final foot in (31a). The connection between stress and the presence of [h] is again supported by the specific variants for *mayhem* in (31):

(31) a. cóhòrt
[méɪhèm] 'mayhem'

b. yógurt
[méɪəm] 'mayhem'

These examples demonstrate the need for careful study of the (syntagmatic) phonological context before concluding that a specific sound pattern qualifies as a boundary effect indicative of complex morphological structure. Equally important for the evaluation of potential deviations from canonical phonology is the study of the

relevant paradigmatic context. Consider again the occurrence of tense [i:] in pretonic position in the verbs in (32a) (as opposed to the absence of such a vowel in the corresponding position in the non-verbs *benign* or *belúga*) which has been analysed as a boundary effect indicative of complex morphological structure ("Final Nucleus Enhancement"). Phonetically similar violations are seen in the words in (32b), which are paradigmatically related to the respective words to their right:

(32)	a.	b[i:]cóme 'become'	b.	l[i:]gálity 'legality' -> l[í:]gal 'legal'
		b[i:]líeve 'believe'		d[i:]mónic 'demonic' -> d[í:]mon 'demon'
		b[i:]gín 'begin'		[i:]g´yptian 'Egyptian' -> [í:]gypt 'Egypt'

The deviations from simplex phonology in (32b) cannot be analysed as boundary effects because they do not involve coinciding morphological and prosodic boundaries. Instead these deviations appear to be "licensed" by the corresponding vowels in the respective base words. "Licensing" means that the phonological feature identified as deviation in one word (e.g. the occurrence of a long, tense vowel in unstressed position in *l*[i:]*gálity*) appears in a different phonological context in a related word, such that that context sanctions the feature in question (e.g. the occurrence of a long, tense vowel in *stressed* position in *l*[i:]*gal*). The "transfer" of the feature from the regular context (i.e. stress) to the irregular context (i.e. lack of stress) is then motivated by a constraint on paradigm uniformity, which requires identity of corresponding phonological structure in paradigmatically related words.

Given this analysis the occurrences of the pretonic tense vowels in (32a) versus (32b) are entirely distinct phenomena. In (32a), this deviation from regular phonology serves as a boundary signal indicating the presence of a head prefix which functions as an indicator of syntactic category. In (32b), the deviation in question signals the existence of a paradigmatically related word in the lexicon, which licenses the deviation. The latter function has nothing to do with morphological complexity as it can also be detected in words like (33a), which do not have internal morphological structure. The noun in (33a) has not (?yet) developed initial stress, unlike the nouns with comparable syllable structure in (33b). The stability of final stress in the noun in (33a) is presumably a PU-effect (paradigm uniformity effect), to secure sameness of stress with respect to the base verb, where final stress is regular.[17]

(33)	a.	succéss -> succéed	b.	áccèss, príncèss, récèss, ábscèss

Whereas morphological complexity is irrelevant to the occurrence of PU-effects it is essential for the occurrence of boundary effects. By contrast, the existence of paradigmatically related words is essential for the occurrence of PU-effects but not to the occurrence of boundary effects (cf. the examples in (32a)).[18] For the purpose of this paper it is only important to be aware of PU-effects as a possible source of "deviant" sound patterns which, unlike boundary effects, do no reflect on word-internal morphology.

[17]Several of the nouns in (33b) are also etymologically related to iambic verbs (e.g. *accéde, recéde*), but unlike in the case of *success - succeed*, there are no close semantic relations.

[18] Additional differences between boundary effects and PU-effects are discussed in Raffelsiefen 2005.

3. Boundary effects as a window on morphological structure

Assuming now that phonological boundaries signal the presence of "underlying" morphological structure it suggests itself to analyse the distribution of such signals to gain information about morphology, including the recognition and labeling of morphological structure. For instance, given the boundary effects (Containment and Headedness) in *suborbital*, the question arises of what motivates the underlying morphological structure: the fact that *orbital* matches an independent word, the fact that *sub-* recurs in other words like *subalpine* or *subtropical*, or the fact that *sub-* has meaning? Similarly, is the boundary effect in *repel* (Head-Final Enhancement) due to the recurrence of *–pel* in other verbs (e.g. *compel*, *impel*, *expel*), the recurrence of *re-* (*reject*, *relent*) or the existence of the near-homophonous productive modifying prefix *re-*? Should *pain* in *painful* be categorized as a noun, a word, a stem or a root? How about the categorization of *orbital* in *suborbital*, *little* in *belittle* or *-pel* in *repel*? The answers to these questions based on the evidence from boundary signals are presented below.

3.1. The results

The evidence from word-internal boundary effects in English supports two basic morphological structures. One type consists of a modifier or a modifying prefix followed by a head whereas the other type consists of a head-affix and a root arranged in either order. Moreover, boundary effects indicate that the recognition of the head, or head affix respectively, determines the overall structure. The relevant evidence for the two types of structures is presented in section 3.1.1 and section 3.1.2.

3.1.1. Modifier-head structures

The evidence from boundary effects indicates an asymmetry regarding the status of the components in compounds. The basic generalization is that boundary effects in compounds may persist for as long as the rightmost member corresponds to an independent word. If that word becomes obsolete, however, fusion results as in (34a). By contrast, if the word corresponding to the lefthand member becomes obsolete prosodic boundaries may persist as is shown in (34b). The restriction "may" is added to account for the fact that "High Frequency Fusion" is always possible, regardless of the status of the rightmost member (cf. possible fusion in all compounds ending in *-berry*, e.g. [rɑ:zbəri] 'raspberry', [blu:bəri] 'blueberry').

(34)	a.	√ice-†ickel (c.f. O.E. gićel 'glacier')	(ícicle)$_{\omega}$
		√nose-†thirl (c.f. O.E. þȳrel 'hole')	(nóstril)$_{\omega}$
	b.	†luke-√warm (c.f. O.E. hlēow 'warm')	(lúke)$_{\omega}$(wàrm)$_{\omega}$
		†step-√child (cf. O.E. ste:op 'bereaved')	(stép)$_{\omega}$(chìld)$_{\omega}$

Assuming that the examples in (34) are representative they indicate that the recognition of a head constituent, which in English is rightmost within the word, is crucial to morphological analysis. Specifically, given an input word with a certain category as in (35a) and the recognition of a constituent in rightmost position corresponding to an independent word with that same category as in (35b), the constituent is labeled as head of the input word as in (35c). Paradigmatic knowledge is accordingly essential to the recognition of head constituents in complex words.

(35) a. $[\text{lúkewàrm}]_{\text{ADJ}}$

b. $[\text{lúke}[\text{wàrm}]]_{\text{ADJ}} = [\text{wárm}]_{\text{ADJ}}$

c. $[\text{lúke}[\text{wàrm}]_{\text{HEAD}}]_{\text{ADJ}}$

Since any two items with the same category as in (35b) necessarily commute, the rest of the word could be omitted without affecting grammaticality. As a result that rest is classified as a modifier, regardless of its properties as in (36a). The boundaries of both heads and modifiers are aligned with pword boundaries as in (36b), giving rise to boundary effects.

(36) a. $[[\text{lúke}]_{\text{MOD}}[\text{wàrm}]_{\text{HEAD}}]_{\text{ADJ}}$

b. $[([\text{lúke}]_{\text{MOD}})_{\omega}([\text{wàrm}]_{\text{HEAD}})_{\omega}]_{\text{ADJ}}$

Given the prosodic structure in (36b) the occurrence of the word in actual speech will presumably satisfy potential prosodic requirements for head recognition, thereby ensuring the stability of the boundary effects in historical perspective.
If no head is recognized as in the somewhat hypothetical form in (37a), the entire word is mapped into a single pword with the (eventual) result that the phonological structure matches that of simplexes as in (37b). [19]

(37) a. $[\text{nósethìrl}]_{\text{NOUN}}$

b. $([\text{nóstril}]_{\text{NOUN}})_{\omega}$

It seems that once prosodic fusion has affected the phonological form of the (former) compound (e.g. loss of the weak foot, assimilation), head recognition is ruled out, even if the word corresponding to the original head were to reappear in the language. If this is correct, the recognition process modeled in (35)-(36) is also sensitive to (surface) prosodic structure, represented by the stress marks in the input in (35a). The recognition procedure outlined above aims accordingly not for maximal parsimony in lexical entries. Rather, the aim is to capture the conditions necessary for transfering language

[19]Plausibly not only the complete loss of the relevant word from the language, but a low token frequency relative to the compound, may suffice to cause fusion. This is because lower relative frequency implies that the word is likely not to be known by the time the compound is acquired, and consequently cannot be recognized. This may be the cause of prosodic fusion in *há*[ŋ]*kerchief* 'handkerchief', which is far more common than the historically related noun *kérchief* (cf. section 6).

structures from speakers to learners, thereby accounting for potential stability of morphological and prosodic structure in historical perspective.

The analysis illustrated in (35)-(36) extends to the cases in (38), which are usually considered part of derivational morphology. Specifically, the presence of two separate pwords can generally be attributed to head recognition as described above. That is, given an input word such as the adjective *subalpine* in (38a) and the recognition of a constituent in rightmost position corresponding to an independent adjective as in (38b), that constituent is labeled as head of the input word as in (38c).

(38) a. $[sùbálpìne]_{ADJ}$

b. $[sùb[álpìne]]_{ADJ} = [álpìne]_{ADJ}$

c. $[sùb[álpìne]_{HEAD}]_{ADJ}$

The rest of the word is again classified as a modifier, specifically a modifying prefix, marked by the subscript "MODP". The classification of a given "rest"-constituent as a modifying prefix as opposed to a simple modifier appears to be determined by the semantic relationship obtaining between the input word and its head, which supports the relevance of paradigmatic knowledge to syntagmatic analysis involving heads (as opposed to head affixes, cf. section 3.1.2. below).[20] The boundaries of both heads and modifying prefixes are aligned with pword boundaries as in (39b), giving rise to boundary effects.

(39) a. $[[sùb]_{MODP}[álpìne]_{HEAD}]_{ADJ}$

b. $[([sùb]_{MODP})_{\omega}(álpìne]_{HEAD})_{\omega}]_{ADJ}$

The distinction between modifiers and modifying prefixes is motivated by relative prominence patterns: *sub-* in (39), although forming a separate pword like *luke-* in (36), has weak prominence. This distinction is systematic only for adjectives and verbs. All modifiers, including modifying prefixes, tend to have main stress in nouns.

The irrelevance of the inherent properties of modifiers is demonstrated by the occurrence of stable boundary effects in words with unique modifying prefixes. The modifying prefixes in (40a) do not recur, yet their stable stress in pretonic position qualifies as a Headedness effect, indicative of their status as separate pwords. The analysis of stress in (40a) as a Headedness effect is based on the prosodic contrast with the words in (40b), which do not allow for the recognition of a head.

(40)	a.	$([æb])_{\omega}(nórmal)_{\omega}$ 'abnormal'	b.	[əb]nóxious 'obnoxiuos'
		$([ɪg])_{\omega}(nóble)_{\omega}$ 'ignoble'		[ɪg]nóre 'ignore'
		$([ænt])_{\omega}(árctic)_{\omega}$ 'antarctic'		[ən]ténna 'antenna'

[20] Typical semantic relations that motivate the classification of a modifier as a prefix ("MODP") are non-gradient, including privative relations (*asymmetry - symmetry, nontoxic - toxic*), contrary relations (*unfriendly - friendly*, *impolite - polite*), and spacial or temporal relations (e.g. *precook - cook*, *postdate - date*).

Given their non-recurrence, the modifying prefixes in (40a) cannot be learned and recognized independently. Instead, their prosodic form as separate pwords derives from their role as "rests", which remain as parts of words after the respective heads have been recognized and bracketed as shown in (41):

(41)	Input:	1. Head recognition	2. "Rest" => modifier
	[àbnórmal]$_{ADJ}$	[àb[nórmal]$_{HEAD}$]$_{ADJ}$	[[àb]$_{MODP}$[nórmal]$_{HEAD}$]$_{ADJ}$
	[ìgnóble]$_{ADJ}$	[ìg[nóble]$_{HEAD}$]$_{ADJ}$	[[ìg]$_{MODP}$[nóble]$_{HEAD}$]$_{ADJ}$
	[àntárctic]$_{ADJ}$	[ànt[árctic]$_{HEAD}$]$_{ADJ}$	[[ànt]$_{MODP}$[árctic]$_{HEAD}$]$_{ADJ}$

On this analysis, morphophonological properties of modifying prefixes are expected to be likewise irrelevant for the prosodic organization of words. In fact, the prosodic evidence shows that the modifying prefix *iN*-, which exhibits regular phonologically conditioned allomorphy, forms a separate pword on a par with non-varying modifying prefixes such as *non-, pan-,* and *un-*. That is, all modifying prefixes are stressed to satisfy Headedness, yielding the (potential) contrast in pretonic initial stress seen in (42b) vs. (42c) (cf. the stress marks in Wells 2000, Webster's 2000).

(42)	a.	b.	c.
	(nò**n)**$_{ω}$(mémber**)**$_{ω}$	(ì**m)**$_{ω}$(móral**)**$_{ω}$	(immédiate**)**$_{ω}$
	(ù**n)**$_{ω}$(líke**)**$_{ω}$	(ì**l)**$_{ω}$(légal**)**$_{ω}$	(illúsion**)**$_{ω}$
	(pà**n)**$_{ω}$(Búddhism**)**$_{ω}$	(ì**r)**$_{ω}$(régular**)**$_{ω}$	(iráscible**)**$_{ω}$

Others have proposed to represent the words in (42b) as single pwords, arguing that "assimilation" in (42b), as opposed to (42a), indicates prosodic fusion (cf. Szpyra 1989). However, in contrast to the low-level allophonic effects discussed in section 2 the sort of variation seen in (42b) is hardly a consequence of prosodic organization. Instead, the variation in the form of the negative prefix indicates phonologically conditioned allomorph selection with no obvious reference to suprasegmental structure.

Although not a consequence of prosodic structure, the occurrence of adjacent identical sonorants in (42b) conceivably enhances the likelihood of prosodic fusion in casual or fast speech. This would account for the fact that Wells lists at least three variants for words with modifying *in-, im-, il-, ir-,* as illustrated in (43), but not for words with modifying *non-, un-,* or *pan-*. Wells uses the diacritic ['] to indicate that the following syllable has main stress, [ˌ] indicates secondary stress.

(43)		
	[ˌɪm'mɑrːəl], [ˌɪ'mɑːrəl], [ɪ'mɑːrəl]	'immoral'
	[ˌɪl'liːgəl], [ˌɪ'liːgəl], [ɪ'liːgəl]	'illegal'
	[ˌɪr'regjʊlər], [ˌɪ'regjʊlər], [ɪ'regjʊlər]	'irregular'

A greater tendency for phonetic fusion in (42b), as opposed to (42a), could in addition be due to the lesser intensity of the prefix vowel [**ɪ**], compared to the vowels [å], [ʌ], and [æ]. Significantly, both these differences in vowel quality and the relevant phonotactic differences (i.e. adjacency of more similar consonants across morpheme boundaries for *iN*-prefixations compared to *un-, non-*, and *pan*-prefixations) concern segmental structure and its potential effect on the salience of boundary marking (cf. the

discussion of *night rate* versus *oxe-eyed* in section 2).[21] These differences accordingly do not argue against representing the modifying prefixes in the words in both (42a) and (42b) as separate pwords.

Representing all modifying prefixes as separate pwords is motivated not only by the (potential) contrast in pretonic stress as in (42a,b) vis-à-vis (42c) [22] but by the more stable and salient contrasts in relative prominence. The generalization in nouns and adjectives is that a branching foot is strong relative to a following non-branching foot resulting in initial main stress as in (44a). Since modifying prefixes are always weak in adjectives there are clear contrasts in relative prominence as shown in (44a) versus (44b): [23]

(44)	a.		b.	
		(rétrogràde)$_\omega$		(ùn)$_\omega$(afráid)$_\omega$
		(tácitùrn)$_\omega$		(ùn)$_\omega$(concérn)$_\omega$
		(circumspect)$_\omega$		(ìn)$_\omega$(diréct)$_\omega$
		(érudìte)$_\omega$		(ìm)$_\omega$(políte)$_\omega$

Both the Headedness effects observed in (42) (i.e. the presence of pretonic stress in (42a,b) as opposed to (42c)) and the contrasts in relative prominence illustrated in (44) support the parsing mechanism outlined above. That is, the deviations from the canonical stress patterns observed in (42a,b) and (44b) follow if morphological parsing is determined by head recognition as in (45), such that the inherent properties of the modifier are irrelevant. Basing alignment on the morphological structures inferred in (45) will yield the prosodic structures illustrated in (42) and (44).

(45)	Input:	1. Head recognition	2. "Rest" => modifier
	[ìmpolíte]$_{ADJ}$	[ìm[políte]$_{HEAD}$]$_{ADJ}$	[[ìm]$_{MODP}$[políte]$_{HEAD}$]$_{ADJ}$
	[ìllégal]$_{ADJ}$	[ìl[légal]$_{HEAD}$]$_{ADJ}$	[[ìl]$_{MODP}$[légal]$_{HEAD}$]$_{ADJ}$
	[ùnlíke]$_{ADJ}$	[ùn[líke]$_{HEAD}$]$_{ADJ}$	[[ùn]$_{MODP}$[líke]$_{HEAD}$]$_{ADJ}$
	[nònmémber]$_N$	[nòn[mémber]$_{HEAD}$]$_N$	[[nòn]$_{MODP}$[mémber]$_{HEAD}$]$_N$

A further prediction of the parsing mechanism outlined in (41) and (45) is that variation in prosodic structure arises whenever the head is somewhat obscure, known to some but not to others. The variation observed in (46) can be analyzed as a direct consequence of head recognition.

[21] The relevance of these differences in segmental structure for the perception of boundaries and the occurrence of prosodic fusion (with concomitant reduction) could be tested by comparing combinations such as *Tim Miller* versus *Don Miller*.

[22] Recall that words like *immediate* or *illusion*, which do not include a modifying prefix, never have initial stress.

[23] The evidence from relative prominence in support of analysing all modifying prefixes as separate pwords, regardless of allomorphy, is particular strong in Swedish or German, where modifying prefixes have main stress. That is, there is a striking contrast between Latinate words with regular main stress on the final syllable (e.g. *legál, radikál, fundamentál*) and a word with a modifying prefix such as *íllegàl* with initial main stress. For a detailed review of the evidence supporting the analysis of *iN-* as a separate pword, see Raffelsiefen 1999 and 2004.

(46)	Input:	Head recognition	Fully parsed structure
	[inclément]$_{ADJ}$	-	[inclément]$_{ADJ}$ ~ [ínclement]$_{ADJ}$
		[in[clément]$_{HEAD}$]$_{ADJ}$	[[ìn]$_{MOD}$[clément]$_{HEAD}$]$_{ADJ}$
	[acéphalous]$_{ADJ}$	-	[[ə]céphalous]$_{ADJ}$
		[a[céphalous]$_{HEAD}$]$_{ADJ}$	[èɪ]$_{MOD}$[céphalous]$_{HEAD}$]$_{ADJ}$

The irrelevance of the modifier is supported by the observation that even productive modifiers are not parsed as separate pwords when occurring in a word lacking a recognizable head as in (47).

(47)	Input:	Head recognition	Alignment
	[nónchalant]$_{ADJ}$	-	([nónchalant]$_{ADJ}$)$_{ω}$
	[míschievous]$_{ADJ}$	-	([míschievous]$_{ADJ}$)$_{ω}$
	[pandémic]$_{ADJ}$	-	([pandémic]$_{ADJ}$)$_{ω}$
	[discrépant]$_{ADJ}$	-	([discrépant]$_{ADJ}$)$_{ω}$
	[subtráction]$_{N}$	-	([subtráction]$_{N}$)$_{ω}$

Alternatively, the cause of failed morphological analysis (as reflected by the absence of boundary signals) could relate to the fact that the words in (47) are loanwords. However, word prosody offers clear evidence for the morphological analysis of loanwords as long as there is a recognizable head. Some examples are given in (48):

(48)	Input:	1. Head recognition	2. "Rest" => modifier
	[dishónest]$_{ADJ}$	[dis[hónest]$_{HEAD}$]$_{ADJ}$	[[dis]$_{MOD}$[hónest]$_{HEAD}$]$_{ADJ}$
	[malconténт]$_{ADJ}$	[mal[conténт]$_{HEAD}$]$_{ADJ}$	[[mal]$_{MOD}$[conténт]$_{HEAD}$]$_{ADJ}$
	[archbíshop]$_{N}$	[arch[bíshop]$_{HEAD}$]$_{N}$	[[arch]$_{MOD}$[bíshop]$_{HEAD}$]$_{ADJ}$
	[panóptical]$_{ADJ}$	[pan[óptical]$_{HEAD}$]$_{ADJ}$	[[pan]$_{MOD}$[óptical]$_{HEAD}$]$_{ADJ}$
	[misadvénture]$_{N}$	[mis[advénture]$_{HEAD}$]$_{N}$	[[mis]$_{MOD}$[advénture]$_{HEAD}$]$_{ADJ}$
	[asýmmetry]$_{N}$	[a[sýmmetry]$_{HEAD}$]$_{N}$	[[a]$_{MOD}$[sýmmetry]$_{HEAD}$]$_{ADJ}$

Aligning the morphological structures in the righthand column with pword boundaries yields the familiar correlation of boundary effects including relative prominence effects (i.e. weak-strong), Containment effects (e.g. di[s.ɑ]nest 'dishonest'), and Headedness effects (pretonic stress with concomitant stability of vowels as in [è] sýmmetry (*[ə]sýmmetry) 'asymmetry').

To summarize, while the presence of frequent modifiers like *non-*, *dis-*, *re-*, or *pre-* plausibly influences the morphological parsing of a word the prosodic evidence suggests that the recognition of such a modifier is neither a necessary condition for morphological analysis (cf. the data in (40)), nor a sufficient condition (cf. the data in (47)). What is both necessary and sufficient for the analysis of a modifier-head structure is the recognition of a head constituent.[24]

[24]Potential counter-examples include the adjectives *uncouth* and *unkempt*, which suggest that the presence of specific modifiers (possibly only *un-* in English) can be sufficient for morphological parsing.

3.1.2. Head-affix root structures

The evidence from boundary effects in words (historically) derived by head affixation supports the notion of asymmetry in morphological parsing. Here the basic generalization is that boundary effects may persist for as long as there is a recognizable head affix. In the examples in (49) the Final-C and Containment effects indicate the existence of word-internal pword-boundaries despite the absence (or extremely low frequency) of the respective base words.

(49)		
	†ruth-√less (cf. M.E. *ruthe* 'pity')	$(\text{rúth})_{\omega}(\text{less})_{\sigma}$
	†gorm-√less (cf. M.E. *gome* 'attention')	$(\text{gorm})_{\omega}(\text{less})_{\sigma}$
	†feck-√less (cf. Scott. *feck* 'efficacy')	$(\text{féck})_{\omega}(\text{less})_{\sigma}$
	†grate-√ful (cf. M.E. *grate* 'agreeable')	$(\text{gráte})_{\omega}(\text{ful})_{\sigma}$
	†wist-√ful (cf. M.E. *wistly* 'intently')	$(\text{wíst})_{\omega}(\text{ful})_{\sigma}$
	†dole-√ful (cf. M.E. *dol* 'pain, grief')	$(\text{dóle})_{\omega}(\text{ful})_{\sigma}$
	†bale-√ful (cf. M.E. *bale* 'evil influence; anguish')	$(\text{bále})_{\omega}(\text{ful})_{\sigma}$
	†rue[25]-√ful (cf. M.E. *rue* 'sorrow; regret')	$(\text{rúe})_{\omega}(\text{ful})_{\sigma}$
	†environ-√ment (cf. M.E. *envirounen* 'to encircle')	$(\text{envíron})_{\omega}(\text{ment})_{\sigma}$
	†oint-√ment (cf. M.E. *oint* 'to anoint')	$(\text{óint})_{\omega}(\text{ment})_{\sigma}$

In fact, there are no cases where the low frequency or loss of a base word has affected the prosodic structure of derived words (i.e. the presence of internal pword boundaries) as long as there has been a recognizable head affix. This generalization indicates the parsing mechanism illustrated in (50), which is determined by the recognition of a head affix.

(50)	Input:	Head affix recognition	"Rest" => root
	$[\text{gráteful}]_{\text{ADJ}}$	$[\text{gráte}[\text{ful}]_{\text{H-AFF}}]_{\text{ADJ}}$	$[[\text{gráte}]_{\text{ROOT}}[\text{ful}]_{\text{H-AFF}}]_{\text{ADJ}}$
	$[\text{fáteful}]_{\text{ADJ}}$	$[\text{fáte}[\text{ful}]_{\text{H-AFF}}]_{\text{ADJ}}$	$[[\text{fáte}]_{\text{ROOT}}[\text{ful}]_{\text{H-AFF}}]_{\text{ADJ}}$
	$[\text{rúthless}]_{\text{ADJ}}$	$[\text{rúth}[\text{less}]_{\text{H-AFF}}]_{\text{ADJ}}$	$[[\text{rúth}]_{\text{ROOT}}[\text{less}]_{\text{H-AFF}}]_{\text{ADJ}}$
	$[\text{tóothless}]_{\text{ADJ}}$	$[\text{tóoth}[\text{less}]_{\text{H-AFF}}]_{\text{ADJ}}$	$[[\text{tóoth}]_{\text{ROOT}}[\text{less}]_{\text{H-AFF}}]_{\text{ADJ}}$

The indiscriminate labelling of all "rests" in (50) as *roots*, regardless of whether or not these rests correspond to independent words, is based on the evidence from prosody. Specifically, the right boundaries of all roots preceding non-integrated head affixes exhibit the properties characteristic of right pword boundaries, including Final-C effects and Containment effects. This observation is accounted for by aligning all roots with pword boundaries. In addition, the outer word boundaries are aligned with pword boundaries which leaves the segments of the head affixes to be parsed "bottom-up", resulting in (unfooted) syllables.

[25]In addition to the noun *rue*, which is the historical base of this adjective and became obsolete, there is also a verb *rue*, meaning 'regret'. However that verb is far less common than the adjective *rueful* and also differs semantically. According to the American Heritage Dictionary of the English Language (4th edition, 2000) the main meaning of the adjective is 'inspiring pity or compassion'.

(51)

Alignment	Output
$([([\text{gráte}]_{\text{ROOT}})_{\omega}[\text{ful}]_{\text{H-AFF}}]_{\text{ADJ}})_{\omega}$	$((\text{greɪt})_{\omega}(\text{fəl})_{\sigma})_{\omega}$
$([([\text{fáte}]_{\text{ROOT}})_{\omega}[\text{ful}]_{\text{H-AFF}}]_{\text{ADJ}})_{\omega}$	$((\text{feɪt})_{\omega}(\text{fəl})_{\sigma})_{\omega}$
$([([\text{rúth}]_{\text{ROOT}})_{\omega}[\text{less}]_{\text{H-AFF}}]_{\text{ADJ}})_{\omega}$	$((\text{ru:}\theta)_{\omega}(\text{ləs})_{\sigma})_{\omega}$
$([([\text{tóoth}]_{\text{ROOT}})_{\omega}[\text{less}]_{\text{H-AFF}}]_{\text{ADJ}})_{\omega}$	$((\text{tu: }\theta)_{\omega}(\text{ləs})_{\sigma})_{\omega}$

The key question raised by the analysis illustrated in (50) is what determines the recognition of the suffixes. Because of the phonologically conditioned integration of all vowel-initial and consonantal suffixes into the pword of the root the prosodic evidence offers potential insight only for words with consonant-initial syllabic suffixes. Here the generalization emerges that productivity is the decisive factor for recognition. Rare cases of apparent High Frequency Fusion aside (e.g. *business, beautiful*), there is a clear tendency for productive suffixes, in particular *-ness, -less,* and *-ful*, but also *-ment*, *-ship*, and *-hood*, for which productivity is confined to base words with specific morphological or semantic properties[26], to be associated with stable word-internal boundary effects.

What complicates the evaluation of the role of productivity for affix recognition are potential PU-effects. Consider the adjectives in (52a), which include an unproductive suffix and yet deviate from canonical phonological form :

(52)

a.	[loʊðsəm] 'loathsome' [loʊnsəm] 'lonesome'	
b.	cf. [loʊð] 'loathe' cf. [loʊn] 'lone'	
c.	$(\text{bʌksəm})_{\omega}$ 'buxom'	< M.E. buhsum (cf. O.E. bugan 'to bend')
	$(\text{lɪsəm})_{\omega}$ 'lissom'	< M.E. l[i]thsom (cf. l[i:]the 'gentle')
	$(\text{gru:səm})_{\omega}$ 'gruesome'	< dial. gruesome/grewsome (†grue 'to shiver')
	$(\text{nɔɪsəm})_{\omega}$ 'noisome'	< M.E. noyesum (†noy 'to trouble, vex')
	$(\text{hænsəm})_{\omega}$ 'handsome'	< M.E. handsom *-> hand

Evidence for the analysis of the non-canonical sound patterns in (52a) as PU-effects rather than boundary effects pertains to the observation that such deviations from simplex structure are consistently licensed by a base word (cf. (52b)). When there is no recognizable base, either because the historical base became obsolete (e.g. (†*grue* 'to shiver' in *gruesome*), dissociated by sound change (e.g. *buxom* (< O.E. *bu:hsum*) - *bow* (< O.E. *bu:gan*), or because of a concrete meaning[27] (e.g. *hand* in *handsome*) we find

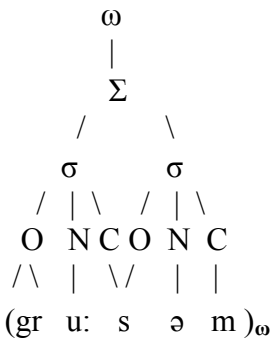
[26] The suffix *-ment* combines productively with *be-* or *en*-prefixations (e.g. *besmirchment*, *endearment*, cf. Marchand 1969:332), the suffix *-hood* combines productively with relational nouns and age-related nouns referring to humans (e.g. *sisterhood, adulthood*, cf. Marchand 1969:293), the suffix *-ship* combines productively with terms referring to ranks (e.g. *kingship, dictatorship*, cf. Marchand 1969:346).

[27] The meaning of concrete base nouns is prone to be reflected less and less in the meaning of derived adjectives over (historical) time. Compare the meanings of *handy, fishy, hairy,* which are based on concrete nouns, with the meaning of adjectives based on abstract nouns such as *hungry, wealthy, greedy*.

that (historical) -*some* derivations exhibit no deviation from simplex phonology as is shown in (52c).

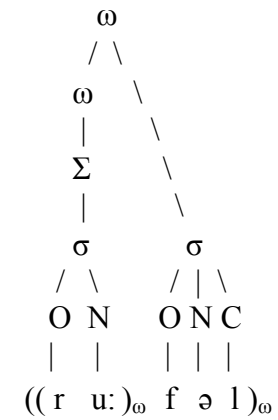
The effects of affix recognition on the prosodic organization of words are seen in the subtle difference between *gruesome*, where the stressed vowel is foot-internal in a closed syllable, and *rueful*, where the stressed vowel is foot-final and therefore lengthened (cf. the transcriptions in Wells 2000). For some speakers schwa can delete foot-internally, as in *gruesome*, but not outside the foot (hence the contrast between *rueful*, transcribed with schwa, and *rifle*, transcribed with a syllabic sonorant in Kenyon & Knott (1953)). The contrasts in foot and syllable structure seen in (53) correspond accordingly precisely to the contrasts between the simplex *minus* and suffixed *shyness* represented in (13).

(53) a.



'gruesome'

b.



'rueful'

The morphological parsing mechanisms indicated by the prosodic structures in (53) are given in (54):

(54)	Input:	Head affix recognition	"Rest" => root	Output:
	[grúesome]$_{ADJ}$	-	-	see (53a)
	[rúeful]$_{ADJ}$	[rúe[ful]$_{H\text{-}AFF}$]$_{ADJ}$	[[rúe]$_{ROOT}$[ful]$_{H\text{-}AFF}$]$_{ADJ}$	see (53b)

The claim that affix recognition is essential only to boundary effects, but not to PU-effects, is supported by the clear cases of PU-based phonological changes in (55):

(55)		
	comp[æ]rison > comp[ɛ]rison 'comparison'	(cf. -> comp[ɛ]r 'compare')
	cons[ɑ]latory > cons[oʊ]latory 'consolatory'	(cf. -> cons[oʊ]le 'console')
	p[æ]tronage > p[eɪ]tronage 'patronage'	(cf. -> p[eɪ]tron 'patron')
	sph[ɛ]rical > sph[ɪ]rical 'spherical'	(cf. -> sph[ɪ]re 'sphere')

All suffixes in (55) are vowel-initial, hence necessarily fused with the root into a single pword, and none is productive in English. Prosodic fusion and non-productivity of affixes are consequently consistent with the occurrence of PU-effects. The only condition for the occurrence of the PU-effects is the recognition of relatedness between words, which can be based entirely on phonological and semantic similarities between the relevant words. Given that the occurrence of PU-effects is consistent with lack of

internal morphological structure and given that PU-effects tend to be sporadic, eliminating alternations in some, but not all, related words, the English data confirm the relevance of affix productivity to word-internal morphological structure. Specifically, these data show that words which include productive suffixes have stable boundary effects indicative of internal pword boundaries, whether or not the root corresponds to an independent word. These data further show that words with unproductive suffixes consistently have simplex structure when there is no recognizable base (because PU-effects are ruled out then) and often have simplex structure even when there is a recognizable base (because PU-effects are sporadic and do not necessarily cause deviations from simplex structure). Examples are the words *laughter* and *knowledge*, which include non-recurring suffixes and are phonologically indistinct from simplexes like *after* and *college*:

(56)	Input:	Head affix recognition	Alignment	Output:
	[láughter]$_{ADJ}$	-	([láughter]$_{ADJ}$)$_{\omega}$	(læftər)$_{\omega}$ - (æftər)$_{\omega}$ 'after'
	[knówledge]$_{N}$	-	([knówledge]$_{N}$)$_{\omega}$	(nɑ:lɪdʒ)$_{\omega}$ - (kɑ:lɪdʒ)$_{\omega}$ 'college'

To summarize, the investigation of the prosodic evidence as a window on morphological structure indicates the crucial importance of suffix recognition, yielding the results in (57) for English. Recall that the prosodic evidence can be explored for non-cohering (i.e. consonant-initial, syllabic) suffixes only:

(57)	Recognized suffixes		Unrecognized suffixes
	<----------		---------->
	-ness, -less, -ment, -man, -ful -hood, -ship, -ling	?-dom	-some, -ter, -ledge

Suffix-recognition for *-ness, -less, -ment, -man* and *-ful* can be related to productivity, which in turn may be enhanced by phonological structure, possibly the combination of a salient (sonorant or non-coronal) onset and a coronal coda. The suffixes *-hood, -ling,* and *-ship*, which are also associated with stable boundaries, are less productive but their recognition may be secured by the combination of recurrence and phonological salience. Most notable here is the presence of full vowels, which in turn results from the consonantal structure of these suffixes. The full vowel in *-hood* owes its presence to the stability of the foot, which is secured by the restriction of [h] to foot-initial position in English (cf. 31). The combination of a strictly syllable-initial (i.e. non-ambisyllabic) onset and the occurrence of a non-coronal obstruent in coda position secures the stability of the foot and the concomitant stability of the full vowel in *-ship* (cf. 30). Finally, the postvocalic velar nasal ensures the stability of the feature [+high] in the vowel in *-ling*.

The significance attached to the recognizability of the suffixes for word-internal morphological structure is supported by the evidence from head prefixation. The results are presented in (58):

(58)	Recognized head prefixes		Unrecognized head prefixes
	<----------		---------->
a.	re-, de-, pre-, be-	? un-, ?in-/en, ?im-/em	se-, e-, per-, for-, ab-, ob-, neg-, con-, com-, col-, cor- sub-, suf-, sur-, dis-, ex-
b.	inter-, super-, over- under-		circum-, retro-, extra-

The classification of the monosyllabic prefixes in (58a) is based on the observation that only verbs including the prefixes *re-, de-, pre-*, or *be-* exhibit clear and consistent boundary effects. That is, apart from phonologically conditioned cohesion described in section 2 these prefixes never form a single domain of syllabification with the root and, for many speakers, are consistently distinguished by 'Head Final Enhancement'. The exclusive relevance of the identity of the head prefixes is demonstrated by the data in (59). All verbs in (59a) involve non-recurring roots, yet they exhibit consistent boundary effects, clearly because of the presence of the respective prefixes. By contrast, the verbs in (59b) include the recurring and hence in principle learnable (etymological) root *-lect* (From Latin *legere* 'to gather, choose'), yet these verbs are phonologically indistinguishable from simplexes. Tensing or lengthening of the prefix vowel in (59b) is ungrammatical, despite its occurrence in syllable-final position and its orthographic representation with <e>.

(59)	Input:	Head affix recognition	"Rest" => root	
a.	[relént]$_V$	[[re]$_{H\text{-}AFF}$lént]$_V$	[[re]$_{H\text{-}AFF}$[lént]$_{ROOT}$]$_V$	'relent'
	[desíre]$_V$	[[de]$_{H\text{-}AFF}$síre]$_V$	[[de]$_{H\text{-}AFF}$[síre]$_{ROOT}$]$_V$	'desire'
	[prepáre]$_V$	[[pre]$_{H\text{-}AFF}$páre]$_V$	[[pre]$_{H\text{-}AFF}$[páre]$_{ROOT}$]$_V$	'prepare'
	[begín]$_V$	[[be]$_{H\text{-}AFF}$gín]$_V$	[[be]$_{H\text{-}AFF}$[gín]$_{ROOT}$]$_V$	'begin'
b.	[se**léct**]$_V$	-	-	'select'
	[e**léct**]$_V$	-	-	'elect'
	[neg**léct**]$_V$	-	-	'neglect'

The parsing mechanism in (59) yields the structure in (60), which serves as a basis for alignment. The correct output forms presuppose alignment of both word and root boundaries with pword boundaries. By contrast, there is no motivation for invoking alignment when parsing the segments of head prefixes, which form monosyllabic and hence unfooted syllables. "Final Nucleus Enhancement" applies before pword boundaries as in (60a), but not before foot boundaries as in (60b):

(60)	Morphologically parsed structures:	Alignment:	Output:
a.	[[re]$_{H\text{-}AFF}$[lént]$_{ROOT}$]$_V$	([[re]$_{H\text{-}AFF}$([lént]$_{ROOT}$)$_\omega$]$_V$)$_\omega$	((r[i:])$_\sigma$(lént)$_\omega$)$_\omega$
	[[de]$_{H\text{-}AFF}$[síre]$_{ROOT}$]$_V$	([[de]$_{H\text{-}AFF}$([síre]$_{ROOT}$)$_\omega$]$_V$)$_\omega$	((d[i:])$_\sigma$(sire)$_\omega$)$_\omega$
	[[pre]$_{H\text{-}AFF}$[páre]$_{ROOT}$]$_V$	([[pre]$_{H\text{-}AFF}$([páre]$_{ROOT}$)$_\omega$]$_V$)$_\omega$	((pr[i:])$_\sigma$(páre)$_\omega$)$_\omega$
	[[be]$_{H\text{-}AFF}$[gín]$_{ROOT}$]$_V$	([[be]$_{H\text{-}AFF}$([gín]$_{ROOT}$)$_\omega$]$_V$)$_\omega$	((b[i:])$_\sigma$(gín)$_\omega$)$_\omega$
b.	[seléct]$_V$	([seléct]$_V$)$_\omega$	(s[ɪ]léct)$_\omega$
	[eléct]$_V$	([eléct]$_V$)$_\omega$	([ɪ]léct)$_\omega$
	[negléct]$_V$	([negléct]$_V$)$_\omega$	(n[ɪ]gléct)$_\omega$

The irrelevance of (etymological) root recurrence for prosodic structure (and, presumably for morphological parsing), can be further demonstrated with the near-minimal pairs in (61). These data support the claim that the occurrence of boundary effects, in particular the tensing and lengthening of the prefix vowel, is exclusively determined by the recognizability of a head prefix:[28]

(61)	Input:	Head affix recognition:	"Rest": root	Alignment:
	$[redúce]_V$	$[[re]_{H\text{-}AFF}dúce]_V$	$[[re]_{H\text{-}AFF}[dúce]_{ROOT}]_V$	$[[re]_{H\text{-}AFF}([dúce]_{ROOT})_\omega]_V$
	$[sedúce]_V$	-	-	$([sedúce]_V)_\omega$
	$[recéde]_V$	$[[re]_{H\text{-}AFF}céde]_V$	$[[re]_{H\text{-}AFF}[céde]_{ROOT}]_V$	$[[re]_{H\text{-}AFF}([céde]_{ROOT})_\omega]_V$
	$[secéde]_V$	-	-	$([secéde]_V)_\omega$

The indiscriminate labeling of "rests" as roots, regardless of whether or not those rests correspond to independent words, is supported by prosodic structure. In the verbs presented in (62), all etymological roots correspond to independent words, thereby differing from the verbs in (59). Yet, prosodically all of these verbs are on a par: there are systematic boundary effects for *re-, de-, pre-*, and *be*-prefixation. There are no boundary effects elsewhere:

(62)	Input:	Head affix recognition	"Rest" => root	Output:
	$[renéw]_V$	$[[re]_{H\text{-}AFF}néw]_V$	$[[re]_{H\text{-}AFF}[néw]_{ROOT}]_V$	$(r[i:])_\sigma(néw)_\omega$
	$[decéase]_V$	$[[de]_{H\text{-}AFF}céase]_V$	$[[de]_{H\text{-}AFF}[céase]_{ROOT}]_V$	$(d[i:]_\sigma(céase)_\omega$
	$[prescríbe]_V$	$[[pre]_{H\text{-}AFF}scríbe]_V$	$[[pre]_{H\text{-}AFF}[scríbe]_{ROOT}]_V$	$(pr[i:])_\sigma(scríbe)_\omega$
	$[bewítch]_V$	$[[be]_{H\text{-}AFF}wítch]_V$	$[[be]_{H\text{-}AFF}[wítch]_{ROOT}]_V$	$(b[i:])_\sigma(wítch)_\omega$
	$[secúre]_V$	-	-	$(s[ɪ]cúre)_\omega$
	$[condénse]_V$	-	-	$(c[ə]ndénse)_\omega$
	$[submérge]_V$	-	-	$(s[ə]bmérge)_\omega$
	$[abúse]_V$	-	-	$([ə]búse)$

The generalization, already established for English head suffixation, is simply that the morphosyntactic status of the "rest" is irrelevant to prosody.

Considering now the question of what makes the head prefixes *re-, de-, pre-*, and *be-* recognizable, as opposed to the other historical monosyllabic prefixes in (58a), a possible generalization concerns the existence of the near-homophonous modifiers *re-, de-,* and *pre-*. Specifically the fact that these modifiers combine productively with verbs in native word formation (e.g. *rewrite, demystify, precook*), might be relevant. This proposal raises the question of why precisely these prefixes, which historically emerged from head prefixes in Latinate loan verbs, became productive. Possibly the productivity

[28] The representation of the string *-duce* in *reduce*, but not in *seduce*, as a root could be criticized because of the identical alternations observed in *reduce- reduction* and *seduce - seduction* (cf. Aronoff 1976). However, this correspondence is arguably significant from a paradigmatic perspective only, as illustrated in (1), but does not reflect on word-*internal* structure.

of these modifying prefixes has been adopted from French. Perhaps, there is something to be said about phonological form here as well. The assumption that the combination of a salient onset (i.e. non-coronal and/or voiced consonants) and an open syllable makes head prefixes more recognizable accounts not only for the recognizability of *re-, de-,* and *pre-* in Latinate loan verbs (and possibly their rise to productivity as a modifying prefix in native word formation) but also explains the fact that *be-* is the only surviving head prefix from Germanic (as opposed to *†for-, †to:-, †a:-, †of-, †on- †oθ-, †ymb- †at-, †ed-, †with-* etc.).

The relevance of productivity and phonological form for head prefix recognition is supported by the bisyllabic prefixes in (63a). Here the prosodic evidence, specifically the evidence from relative prominence, indicates that the prefixes *inter- super-*, *under-*, and *over-* are recognized, whereas other verb-initial iambic feet are not recognized, as shown in (63b). Again, root-inherent properties are irrelevant:

(63)		Input:	Head affix recognition	"Rest" => root
	a.	[ìntermít]$_V$	[[ìnter]$_{H\text{-}AFF}$mít]$_V$	[[ìnter]$_{H\text{-}AFF}$[mít]$_{ROOT}$]$_V$
		[sùperscríbe]$_V$	[[sùper]$_{H\text{-}AFF}$scríbe]$_V$	[[sùper]$_{H\text{-}AFF}$[scríbe]$_{ROOT}$]$_V$
		[ùndermíne]$_V$	[[ùnder]$_{H\text{-}AFF}$míne]$_V$	[[ùnder]$_{H\text{-}AFF}$[míne]$_{ROOT}$]$_V$
		[òvercóme]$_V$	[[òver]$_{H\text{-}AFF}$cóme]$_V$	[[òver]$_{H\text{-}AFF}$[cóme]$_{ROOT}$]$_V$
	b.	[pérsecùte]$_V$	-	-
		[rétrogràde]$_V$	-	-
		[círcumcìse]$_V$	-	-
		[éxtradìte]$_V$	-	-

Aligning all word- and root boundaries with pword boundaries yields the structures in (64a). The prosodic parsing of the segments of the head prefixes yields disyllabic and hence trochaic feet as in (64b):

(64)	a. Alignment:	b. Output:
	([[ìnter]$_{H\text{-}AFF}$([mít]$_{ROOT}$)$_\omega$]$_V$)$_\omega$	((ìnter)$_\Sigma$(mít)$_\omega$)$_\omega$
	([[sùper]$_{H\text{-}AFF}$([scríbe]$_{ROOT}$)$_\omega$]$_V$)$_\omega$	((sùper)$_\Sigma$(scríbe)$_\omega$)$_\omega$
	([[ùnder]$_{H\text{-}AFF}$([míne]$_{ROOT}$)$_\omega$]$_V$)$_\omega$	((ùnder)$_\Sigma$(míne)$_\omega$)$_\omega$
	([[òver]$_{H\text{-}AFF}$([cóme]$_{ROOT}$)$_\omega$]$_V$)$_\omega$	((òver)$_\Sigma$(cóme)$_\omega$)$_\omega$
	([pérsecùte]$_V$)$_\omega$	(pérsecùte)$_\omega$
	([rétrogràde]$_V$)$_\omega$	(rétrogràde)$_\omega$
	([círcumcìse]$_V$)$_\omega$	(círcumcìse)$_\omega$
	([éxtradìte]$_V$)$_\omega$	(éxtradìte)$_\omega$

Recall that combining a foot with a pword yields weak-strong prominence, thus yielding boundary effects indicative of internal morphological structure for all verbs ending in maximally one consonant (cf. the first four examples in (64)).[29] The regular stress for such verbs is seen in the last four examples in (64), where the combination of a trochaic and a monosyllabic foot yields strong-weak prominence.

[29]I assume that initial main stress in the verb *supervise* is due to paradigm uniformity with the noun *supervisor*, where initial main stress is regular.

Recognition of the prefixes can in some instances be related to the existence of homophonous modifiers which productively combine with verbs. This holds in particular for *over*- and *under*- (e.g. typical pairs like *overcharge* - *undercharge*, *overspend* - *underspend*, *overfeed* - *underfeed*), less so for *super*-, which however is productive with nouns and adjectives, and least for *inter*-. The property shared by the prefixes in question is again phonological. All prefixes are disyllabic trochees ending in *-er*.

The relevance of morphosyntactic category for the significance of affix productivity for morphological parsing is supported by the phonology of verbs historically derived by ex-prefixation. As a modifier, *ex*- is highly productive with nouns (e.g. *ex-husband, ex-cop*), but not with verbs. The claim that this prefix fails to be recognized in Latinate loan verbs is supported by the cases in (65), all of which involve a (historical) root spelled with initial <h>. Such verbs are typically marked by prosodic fusion, as is indicated by the correlating 'silence' of the <h> and the voicing of the prefix-final cluster.

(65)	e[gz]áust	'exhaust'
	e[gz]ílaràte	'exhilarate'
	e[gz]órt	'exhort'
	e[gz]íbit	'exhibit'

Pronunciation of the root-initial [h] and concomitant voicelessness of the syllable-final cluster [ks] in (66) is not a boundary effect indicative of head prefix recognition but rather a paradigm effect. This is because such structure presupposes the existence of an independent, semantically related word, in which <h> occurs in a context where pronunciation is ensured (e.g. after sonorant). The words in (66a,b) are directional opposites, where the variation in (66b) plausibly relates to the very low token frequency of *inhume* compared to *exhume*.[30] The relevance of meaning relations for the occurrence of the phonological effect in question is demonstrated by the examples in (66c,d). Here the <h>-initial root recurs in independent words, which however fail to be semantically related to the historic *ex*-prefixations. As a result, the latter conform entirely to canonical phonological patterns.

(66)	a.	e[ks.h]ále	'exhale'	<=> in[h]ále 'inhale'
	b.	e[ks.h]úme ~ e[gz]úme	'exhume'	<=> † in[h]úme 'inhume'
	c.	e[gz]íbit	'exhibit'	in[h]íbit 'inhibit'
	d.	e[gz]órt	'exhort'	co[h]ort 'cohort'

The examples in (66) versus (59) to (64) support the need to distinguish between boundary effects and paradigm uniformity effects, both of which involve deviations from the sound structure of simplexes. Whereas boundary effects presuppose a recognizable head affix in a word with a specific category, regardless of the existence of other words, the occurrence of paradigm uniformity effects as in (66) presupposes the

[30] A well-known example for a paradigm uniformity effect based on semantic oppositeness is the change in the pronunciation of the vowel in English *female* (i.e. *femelle* > *female*), clearly in analogy to the vowel in *male*.

recognition of a semantic relation between whole words (cf. also the analogous cases in (52) versus (49) to (51)).

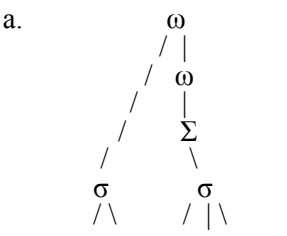
3.2. Summary and discussion

In this section I have presented a morphological parsing mechanism based on the recognition of heads and head affixes, respectively. The (non-cohering) head affixes which are systematically recognized are listed in tables (57) and (58). The observation that these affixes are amenable to a characterization in phonological terms suggests that phonological form might be essential to the recognition of head affixes. In contrast to the analysis of modifier-head structures, the analysis of words derived by head affixation does not appear to involve paradigmatic knowledge.

The parsing rules given here are motivated by the prosodic structure of words. Specifically, these rules yield word-internal morphological structures which serve as a basis for alignment with pword boundaries. Such alignment is manifested by a range of systematically correlating deviations from the sound structure of simplexes including Containment effects, Headedness effects, Final-C effects, Relative Prominence effects and Final Nucleus Enhancement. The claim is that the historical stability of these effects, some of which are quite subtle (e.g. *shyness - minus*), indicates an acquisition process involving the parsing procedure for word-internal morphology outlined here. To demonstrate this claim it is adequate, in fact appropriate in view of the actual conditions for language learning, to represent input forms with surface phonological structure, including stress. It may be significant, however, that the correct output forms would result even if most of the stress marks, including all relative prominence marking, were eliminated from input forms.

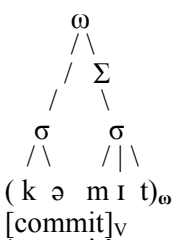
A possible objection to the analysis of the morphology-phonology interface presented here is that it lacks comprehensiveness. Chomsky and Halle (1968) analyse stress in verbs such as *commít* versus *vómit* in terms of morphological complexity, arguing that final stress indicates the structure *com+mit*, consisting of a prefix and a root, compared to the simplex *vomit*. However, unlike the verb *remit*, which is marked by Final Nucleus Enhancement, the phonological form of *commit* does not exhibit any deviation from the sound patterns of simplex verbs. Both patterns in (67b) and (67c) are equally regular and stable for verbs (cf. final stress in *caréss, haráss, ignóre, avér*).

(67) a.



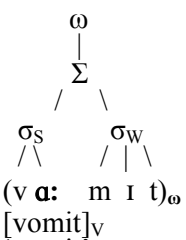
[[re]$_{\text{H-AFF}}$[mit]]$_V$
'remit'

b.



[commit]$_V$
'commit'

c.



[vomit]$_V$
'vomit'

As has been demonstrated above, the complex prosodic structure in *remit* is due to the recognition of the head prefix *re-*. The cause for the distinct stress pattern in *commít* versus *vómit* is historical, specifically the antepenultimate stress pattern in the donor language Latin (cf. Latin *commítere* (Fr. *cométre*) > Middle English *commítten* > New

English *commít* versus Latin *vómere* (past part. *vómitus*) > Middle English *vómiten* > New English *vómit*).

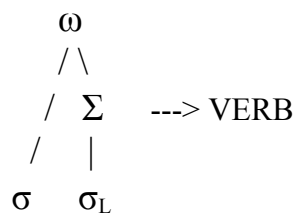
The claim that foot structure in English loan verbs is determined by the position of stress in the etymological source along with historical stability is further illustrated by the verbs in (68), all of which involve the same historical root *-jur* 'right, law'. Final stress in (68a) is due to the stress in the donor languages, presumably based on the (boldfaced) inflected present tense singular form, rather than a form with suffixal stress. Initial stress in (68b) is due to the origin of the words as back-formations from the respective nouns, for which initial stress is regular.[31] Synchronically these are paradigm uniformity effects, supported by the close meaning relations.[32]

(68) a. abjúre < M.E. abjúren < Fr. **abjúr**/abjurér
adjúre < M.E. adjúren < Fr. **adjúr**/adjurér

b. ínjure < ínjury]$_N$
pérjure < pérjury]$_N$

It appears then that the stress patterns in none of the verbs in (67b,c) or (68a,b) indicate word-internal morphological analysis. Instead, these data demonstrate the historical stability of stress, regardless of syllable weight, in English verbs[33], as opposed to the tendency to leftward stress shift in nouns (e.g. *perfúme* > *pérfume*).[34] The generalization that final main stress, especially on a light syllable, (weakly) indicates verbhood is expressed in (69) ("σ_L" means 'light syllable')[35].

(69)



The intuition that learners know something about words like *commít* or *haráss*, namely that these words are most likely verbs, can be captured as in (69) without referring to word-internal morphological structure.

A second case where morphological complexity may seem to be indicated by phonological structure concerns word-internal phonotactics. Trubetzkoy (1958) remarks that certain types of clusters can function as boundary signals, indicative of an internal morphological boundary (cf. 1958:247). For German he lists various consonant clusters, including clusters consisting of a consonant and [h]. Similar cases of unusual clusters are found in English, as is illustrated in (70):

[31] In addition, there are variants *cónjure* versus *conjúre*, which according to the OED go back to a "stress mutation" in Old French.

[32]Paradigm uniformity effects might also reinforce the stability of stress in *commít* (cf. the nouns *commítal, commítment*) and *vómit* (cf. the noun *vómit*).

[33]The only cases of stress instability in disyllabic verbs are those which include a recognizable head suffix like *-ize* or *-ate* (e.g. *chastíze ~ chástize, capsíze ~ cápsize, rotáte ~ rótate, donáte ~ dónate*).

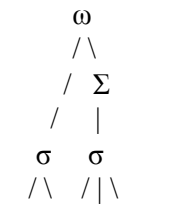
[34] Cases where stress has seemingly shifted in verbs, like the variant *cóntrast*, invariably involve a noun that has undergone prior stress shift (i.e. contrást]$_N$ > cóntrast]$_N$) and are best analysed as conversions exhibiting PU-effects.

[35] I ignore for now the existence of adjectives, which pattern partially with verbs (e.g. the stress in *absúrd*), partially with nouns (e.g. the stress in *séparate*).

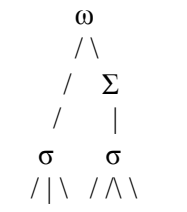
(70) a[**bh**]ór 'abhor'
co[**nf**]íne 'confine' (cf. triu[mf] *triu[nf] 'triumph')
o[**bt**]áin 'obtain'

ùHowever, while the clusters may indicate that these words are originally adopted from Latin or French, they do not indicate morphological complexity. All relevant segments are easily parsed into pword-internal prosodic constituents, including the [h] in (71a), which occurs foot-initially:

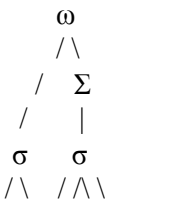
(71) a.



b.

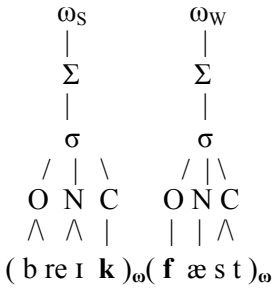


c.

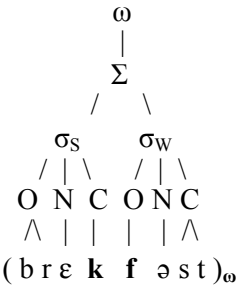


The claim is then that the occurrence of unusual clusters does not qualify as boundary signal if the relevant segments can be parsed into well-formed pword-internal prosodic constituents.[36] The insignificance of mere rareness of consonant combinations compared to the boundary effects reviewed above is revealed by the outcome of historical prosodic fusion. In (72) it is shown that HFF (High Frequency Fusion) results in the correlating loss of the Headedness effect (loss of the weak foot no longer dominated by a pword) and the loss of the Final-C effect (shortening of the superheavy rhyme no longer in pword-final position). By contrast, the cluster [kf] emerges unscathed, even though it is the only such consonant combination within an English word.[37]

(72) a.



b.



While the phonological form in the examples in (71) does not indicate word-internal morphological structure it does again signal word class membership. Specifically, the lack of stress on the initial syllable despite syllable closure indicates that the word is a

[36] For instance, the cluster [bh] in *abhor* is not a boundary signal since [h] is parsed in foot-initial position. By contrast, intervocalic [h] in *pro*[h]*awaiian* cannot be parsed in foot-initial position, thereby signaling pword-initial position indicative of a morphological boundary (i.e. (pro)ω(Hawaiian)ω).

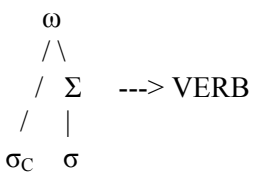
[37] It remains to be investigated whether there are other combinations, in particular certain violations of constraints on syllable contacts, combinations of obstruents differing only in voicedness, combinations of nasals differing only in place of articulation, etc. which, unlike the cluster [kf], are affected by prosodic fusion and consequently do signal boundaries. My point is that rareness or even uniqueness of clusters in historically complex words in itself is insignificant.

verb, rather than a noun. In bisyllabic nouns, closed pretonic syllables are regularly stressed as is demonstrated by the contrasts in (73a,b):[38]

(73) a. c[ə]n.táin]$_V$ 'contain'
c[ə]m.páre]$_V$ 'compare'
c[ə]n.trást]$_V$ 'contrast'
[ə]b.séss]$_V$ 'obsess'
s[ə]s.péct]$_V$ 'suspect'

b. c[æ]n.téen]$_N$ (*c[ə]ntéen]$_N$) 'canteen'
c[æ]m.páign]$_N$ (*c[ə]mpáign]$_N$) 'campaign'
p[ɑ]n.tóon]$_N$ (*p[ə]ntóon]$_N$) 'pontoon'
[æ]b.sínth]$_N$ (*[ə]bsínth]$_N$) 'absínth '
s[ɛ̀]s.tét]$_N$ (*s[ə]stét]$_N$) 'sestet')

The rule in (74) expresses the knowledge that the words in (71) and (73a) are verbs ("σ_C" means 'closed syllable'). Reference to word-internal structure is again unmotivated.

(74)


To summarize, I argue that word-internal boundary effects, which crucially involve reference to pword boundaries, should be distinguished from other peculiarities of sound structure including rare phonotactics and word class specific phonological patterns. Assuming then that English verb stress is adequately described by the structures in (67) and (71) both the acquisition of the sound patterns in cases like (68) and (70) and their concomitant stability in historical time can be explained without referring to word-internal morphological structure. By contrast, the acquisition and stability of the sound patterns referred to as boundary effects above does require reference to morphological structure. This structure reflects the parsing mechanism based on head recognition outlined in section 3.1.

4. A case for universality: optionality and the form-meaning parallelism

Up to this point the discussion has been focused on the morphology-prosody interface. Specifically, a parsing mechanism has been introduced which yields morphological structures based on the recognition of heads or head affixes respectively. For instance, assuming that *able* is recognized as a head in the adjective *unable* and that *en-* is recognized as a head affix in the verb *enable* this parsing mechanism yields the output in (75a)

(75)				
a.	Output of parsing:	[[un]$_{\text{MOD-AFF}}$[able]$_{\text{HEAD}}$]$_{\text{WORD}}$	[[en]$_{\text{H-AFF}}$[able]$_{\text{ROOT}}$]$_{\text{WORD}}$
b.	Correlating affix properties:	paradigmatic variability syntagmatic autonomy	no paradigmatic variability less syntagmatic autonomy
	Prosodic properties:	=> (un)$_\omega$(able)$_\omega$ ("top-down"-parsing, "crisp boundaries")	=> (enable)$_\omega$ ("bottom-up"-parsing, fusion)
	Semantic properties:	necessarily inherent meaning (not')	possibly inferred meaning (cause to become')

[38] The fact that pretonic destressing of closed syllables is ungrammatical only in disyllabic nouns, but not in longer nouns (cf. c[`æ]n.téen]$_N$ (*c[ə]ntéen]$_N$) 'canteen' but [`æ]nténna ~ [ə]nténna 'antenna') shows that foot stability is a minimality effect here.

As was noted above, the definition of heads based on commutation with the input word implies the optionality of the modifying affix. This is what is meant by paradigmatic variability in (75b). Head affixes, which are recognized based on their occurrence within a word with a specific category, are inherently not optional, implying a lack of paradigmatic variability. Below I will discuss a possibly universal correlation between paradigmatic variability and syntagmatic autonomy, both with respect to prosody and meaning.

A close connection between paradigmatic variability and syntagmatic autonomy in relation to prosodic structure is demonstrated in section 3. Specifically, it has been shown that modifying prefixes, which can be omitted, are mapped into separate pwords, necessarily resulting in stress (because of Headedness) and "crisp boundaries" (because of Containment). This sort of prosodic parsing of segments dominated by a separate pword is "top-down" in the sense that a pword necessarily dominates a foot, regardless of the segmental structure involved.

By contrast, head affixes, which cannot be omitted, are not parsed into separate pwords. Rather, the segments of head affixes are parsed into syllables and feet in accordance with general constraints on prosodic structure. This sort of parsing has been characterized as "bottom-up", because of the dependence of foot licensing on the segmental material (e.g. the presence of a foot in the suffix *-hood*, but not in the suffix *-ness*).

The fact that head affixes do not form separate pwords affects not only stress, but also the potential of phonologically conditioned fusion. The verb *enable* in (75) illustrates fusion of a head prefix with a vowel-initial root. The phonological conditioning indicates domination of the relevant GP-alignment constraints by other constraints. Significantly, this sort of fusion (or lack of syntagmatic autonomy) affects only (obligatory) head affixes, not (optional) modifying affixes.

In addition to correlating with differences in prosodic structure, the optionality of affixes also correlates with semantic differences. Specifically, modifying affixes always have inherent meaning, which cannot be inferred from either syntagmatic or paradigmatic structure. By contrast, the meaning associated with head affixes can often be inferred on the basis of syntagmatic context and paradigmatic relations. For instance, the meaning "causative" associated with the verb *enable* must not be analysed as an inherent meaning component of the head prefix *en-* because all transitive verbs for which an adjectival base can be recognized have a causative meaning. A statement of the relevant rule from an analytic perspective (i.e. from the perspective of the hearer/learner) is given in (76).[39] For illustration see the examples in (76b) ("X => Y" reads 'For word X, word Y is recognized as the base'; M(X) means 'meaning of X'):

(76) a. If: $[X]_{TV}$ => $[Y]_A$
Then: M(X): cause to become Y'

b.

[**en**large]$_{TV}$ => [large]$_A$	'cause to become large'
[humid**ify**]$_{TV}$ => [humid]$_A$	'cause to become humid'
[legal**ize**]$_{TV}$ => [legal]$_A$	'cause to become 'legal'
[wid**en**]$_{TV}$ => [wide]$_A$	'cause to become wide'
[corrupt]$_{TV}$ => [corrupt]$_A$	'cause to become corrupt'

[39]For an alternative approach from a synthetic perspective (i.e. the perspective of the speaker) see Beard (1995:177ff), who refers to Szymanek (1988). For criticism of Beard's analysis, which would not apply to the analytic approach presented here, see Plag 1999:237ff).

Transitivity of verbs can be inferred from the syntagmatic context (i.e. the utterance in which the verbs appear). The paradigmatic relation to the relevant adjectives, which must be assumed to be stored in the mental lexicon, is recognized on the basis of sound-meaning correspondences (cf. Raffelsiefen 1998). It would be inappropriate to treat the boldfaced head affixes, including the "zero" affix in the verb *corrupt*, as Saussurean signs which yield compositional meanings in combination with the respective roots. Rather the predictability of the meanings on the basis of the independently given syntagmatic relations (transitivity) and paradigmatic relations (base relations) shows that the affixes in question have no meaning. Their sole function is to indicate word class, i.e. the membership of the relevant word to the class of verbs. This morphosyntactic function is also fulfilled in the cases in (77), where a noun could be recognized as a base (77a) or no word at all (77b). Significantly, the alleged semantic function, to signal causativeness, is no longer manifest.[40] This demonstrates that causativeness is a not a property of specific affixes, but rather derives from the paradigmatic relation between a transitive verb and a recognized base adjective.

(77) a. [**en**courage]$_{TV}$ => [courage]$_{N}$
[burglar**ize**]$_{TV}$ => [burglar]$_{N}$
[person**ify**]$_{TV}$ => ?[person]$_{N}$
[threat**en**]$_{TV}$ => [threat]$_{N}$

b. [**en**chant]$_{TV}$ => Ø
[ostrac**ize**]$_{TV}$ => Ø
[rat**ify**]$_{TV}$ => Ø

An additional rule of semantic interpretation crucially involving paradigmatic knowledge is stated in (78a). This rule describes the meaning assignment to abstract nouns ("AN") for which an adjectival base is recognized. The classification "abstract" is here used with reference to syntax, meaning that the nouns can appear without a determiner. The rule is illustrated in (78b):

(78) a. If: [X]$_{AN}$ => [Y]$_{A}$
Then: M(X): condition/property/state of being Y'

b.

[kind**ness**]$_{N}$ => [kind]$_{A}$	'condition/property/state of being kind'
[obes**ity**]$_{N}$ => [obése]$_{A}$	'condition/property/state of being obese'
[silen**ce**]$_{N}$ => [sílent]$_{A}$	'condition/property/state of being silent'
[just**ice**]$_{N}$ => [just]$_{A}$	'condition/property/state of being just'
[warm**th**]$_{N}$ => [warm]$_{A}$	'condition/property/state of being warm'
[height]$_{N}$ => [high]$_{A}$	'condition/property/state of being high'
[moist**ure**]$_{N}$ => [moist]$_{A}$	'condition/property/state of being moist'
[modest**y**]$_{N}$ => [modest]$_{A}$	'condition/property/state of being modest'
[cold]$_{N}$ => [cold]$_{A}$	'condition/property/state of being cold'

[40]Deciding whether or not causation is involved is sometimes difficult since many transitive verbs involving volition lend themselves to a paraphrase with a causative element (e.g. to eat X: 'to cause X to go down one's esophagus'). For some discussion see Comrie (1985:332ff). Certainly causation can persist (temporarily) in a verb after the relation to the etymological base adjective is obscured by sound change, as is shown in (ia,b) (cf. Raffelsiefen 1998). Only the example in (ic) shows clear loss of a causative meaning:

(i) a. blea[tʃ]]$_{TV}$ 'bleach' *=> blea[k]]$_{A}$ 'bleak' 'cause to become *bleak]$_{A}$
b. clo[z]]$_{TV}$ 'close' *=> clo[s]]$_{A}$ 'close' 'cause to become *close]$_{A}$
c. loa[ð]]$_{TV}$ 'loathe' *=> loa[θ]]$_{A}$ 'loath' *'cause to become loath]$_{A}$

Again, given the generality of the rule in (78a), it would be inappropriate to assign the meaning 'condition/property/state of being' to the individual affixes, including the 'zero affix', and computing the meaning of the nouns in a compositional fashion. The claim that the meaning is not an inherent component of the affixes, but rather results from knowledge of syntagmatic and paradigmatic structure, is supported by the fact that the homophonous affixes, boldfaced in (79), are associated with a deverbal meaning (act of Y'ing) when a verb is recognized as a base:

(79)	[forgive**ness**]$_N$ => [forgive]$_V$	'act/process/result of forgiving'
	[serv**ice**]$_N$ => [serve]$_V$	'act/process/result of serving'
	[grow**th**]$_N$ => [grow]$_V$	'act/process/result of growing'
	[pursui**t**]$_N$ => [pursue]$_V$	'act/process/result of pursuing'
	[fail**ure**]$_N$ => [fail]$_A$	'act/process/result of failing'
	[talk]$_N$ => [talk]$_V$	'act/process/result of talking'

An argument against the paradigmatic rule stated in (78a) and for inherent affix meanings concerns the semantic differences below, first observed by Riddle (1985:438):

> (3) Her **ethnicity** was not a factor in the hiring decision. We are an equal opportunity employer.
> (4) Her **ethnicness** was certainly a big factor in the director's decision. He wanted someone who personified his conception of the prototypical Greek to play the part.

Riddle (1985:438) comments as follows:

> In (3), ethnicity refers to nationality or race, an abstract entity, while in (4), ethnicness refers to an embodied trait involving personal characteristics.

The semantic difference in question concerns the full range of possible values of the spectrum associated with the adjective *ethnic* (i.e. *ethnicity* 'the question of which ethnic group someone belongs to', where everyone is assumed to belong to some group), as opposed to a positive value on that spectrum, in this case asserting that the referent ranks highly on the scale of 'Greekness'. This difference is indeed systematic, but does not, contra Riddle and Plag (2003:66ff), prove that *-ness* and *-ity* have distinct meanings. It can be shown that the "full-spectrum" reading is not inherently associated with *-ity*, but with any established abstract noun for which an adjective can be recognized. The "positive-factual" reading, on the other hand, is consistently obtained for non-established *-ness*-coinages based on that same adjective. We accordingly get an analogous contrast by inserting the noun *age*, which is the established abstract noun based on the adjective *old*, as opposed to the non-established coinage *oldness*, in identical sentence frames:

(80) *Her **age** was a factor in the hiring decision* => No implicature: referent could be young or old
*Her **oldness** was a factor in the hiring decision*. => Necessarily positive value on the scale: assertion that referent is old

The notion of establishedness is reflected in token frequency, with *ethnicity* and *age* vastly outnumbering *ethnicness* and *oldness*, respectively.[41] The relevant contrasts emerge most clearly for adjectives which are semantically unmarked (cf. Lyons 1977).[42] Some additional examples attested in Google are given in (81):

(81) *...request clarification on the **frequency** with which safety showers must be tested...* => no implicature regarding the rate or occurrence of testing
*...the **frequentness** with which her family moved...* => necessarily positive value on the scale: assertion that referent moved frequently.

*...I felt that the **length** of the book was sufficient...* => no implicature regarding the question of whether the referent is considered long or short.
*…I liked alot except the **longness** of the book. It got kind of boring...*
=> Necessarily positive value on the scale: the referent is considered long.

The generalization emerging from these examples is stated below:

> Given an *established* abstract noun X, for which a relation to an unmarked adjective Y is recognized, where unmarkedness implies denotation of the set of all values of the relevant dimension/spectrum, X also includes the set of all values. By contrast, the use of a *non-established* abstract noun Z based on Y, formed by applying a productive morphological rule, yields the positive value of the relevant dimension, resulting in a factual reading.

The generalization suggests that the relevant contrasts in meaning are not inherently associated with the relevant suffixes, but are pragmatic effects predictable on the basis of the respective paradigmatic relations formed in a hearer's mental lexicon. That is, a noun regularly 'inherits' the full spectrum of readings from its (unmarked) adjectival base. The positive-factual reading is a special effect resulting from the use of a non-established noun licensed by a highly productive morphological rule, which would normally be blocked by the existence of the established word (so-called "synonymy-blocking"). This special effect is accordingly pragmatic in nature, resulting from a violation of the 'maxim' to use established words ("Talk like the others").
This interpretation of the semantic contrast in question entails two predictions. First, *ness*-derivatives should yield positive-factual readings only when they are 'blocked' by established nouns. The examples in (82) show that full-spectrum readings are indeed available for 'unblocked' *ness*-derivatives:

(82) *... The dots can't vary in **darkness** or size ... The usual result is that the print is either too dark or too light...* => no implicature regarding the question of whether the dots are dark or not.

*Humans and animals sense a wide range of sound amplitude, volume or **loudness**--from the very quiet to the extremely loud ...* => no implicature regarding the question of whether the sensations are loud or not.

[41] For instance, in February 2006, the number of hits in Google for *ethnicity* versus *ethnicness* were as follows: *ethnicity*: 41.600.000; *ethnicness*: 181.
[42] Unmarked adjectives have both a specific meaning and a general meaning, relating to the whole dimension in question. They appear in neutral questions as in *How old is she*?, as opposed to the question *How young is she*?, which presupposes that the referent is young. For discussion of semantic markedness, see Cruse 2000:172ff.

Assuming that the effect is pragmatic it should not be language-specific. The second prediction is accordingly that the same effect should be found in other languages, whenever there are both 'lexical' abstract nouns for which an adjectival base can be recognized and a highly productive rule for coining nouns based on the relevant adjectives. This prediction is born out by the Swedish example in (83a) and the German example in (83b).[43] To avoid transfer of semantic properties from the corresponding English etablished versus non-etablished nouns I translate the relevant nouns by giving the relevant base relations (e.g. '[N-> long]' means a noun for which *long* is recognized as a base).

(83) a. *Bokens **längd*** [established] *är ca 200 sidor vilket gjorde den ganska snabbläst*. ' The [N-> long] of the book is roughly 200 pages, which made it a rather quick read '=> no implicature regarding the question of whether the referent is considered long or not.
*Även den onödiga **långheten*** [non-established noun formed by productive *-het*-suffixation] *är något som till slut gynnar boken*.
'Even the unnecessary [N-> long] is something which in the end serves the book well' => Necessarily positive value on the scale: the referent is considered long.

b. ...dein **alter** [established] interessiert niemanden, beurteilt wird nach leistung ... 'your [N -> old] does not interest anyone, people are judged by their performance' => no implicature regarding the question of whether the referent is considered old or not
*Deine **Altheit*** [non-established noun formed by productive *-heit*-suffixation] *widert mich an.* 'Your [N -> old] disgusts me'. => Necessarily positive value on the scale: the referent is considered old.

Given that suffixes like English *-ness* or *-ity* are entirely meaningless their sole function is to signal that the words in question are nouns. The table in (75) can thus be continued as follows:

(84)

Output of parsing:	$[[un]_{\textbf{MOD-AFF}}[able]_{HEAD}]_{WORD}$	$[[en]_{\textbf{H-AFF}}[able]_{ROOT}]_{WORD}$
Affix function:	Semantic modification	Word class marking

The claim is then that only modifying affixes, which have inherent meaning, contribute to the meaning of the complex word in a compositional fashion. The main function of head affixes, on the other hand, is to signal word class, in particular the opposition verbs versus non-verbs.[44] For English it holds that head prefixes, which in most instances are non-cohering, signal verbhood (e.g. *be(gin)*, *re(fute), en(large)*).[45] By contrast, all non-

[43]Like all examples in this section these examples were found by using Google.
[44] It is true that there are also head affixes which do seem to have semantic content. For instance the contrast in meaning between the English nouns *employer* and *employee* appears to be associated with the suffixes. But even in these cases there is a question to what extent the meaning contrast is inherent in the suffixes. Perhaps it is not a coincidence that the canonical sound shape for nouns (i.e. lack of stress on the final syllable) in *employer* is associated with the less marked active meaning as opposed to the association of the non-canonical sound shape with final main stress in *employee* with the more marked passive meaning.
[45] The occurrence of non-cohering head prefixes, rather than suffixes, is in accordance with the overall preference for word-final stress in verbs.

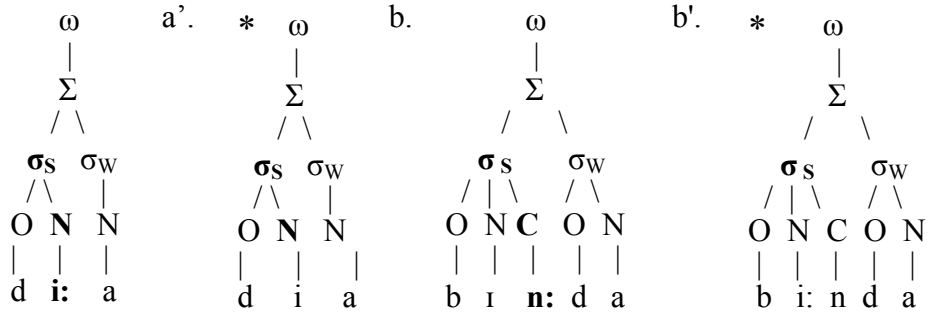
cohering suffixes signal that the word in question is not a verb (e.g. (*shy)ness*, (*meaning)ful*, (*reck)less*).

5. Swedish

The phonology of Swedish supports the relevance of head affix recognition for the prosodic organization of words. That is, the occurrence of boundary effects can be related to the presence of specific affixes, rather than to the properties of stems (e.g. frequency, relatedness to independent words). The boundary effects in question largely correspond to the effects observed in English (cf. section 2). Here I will focus primarily on Final-C effects and on relative prominence effects.

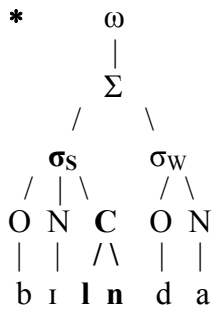
One manifestation of Final-C effects in Swedish concern violations of regular quantity patterns observed in simplexes, which are characterized by so-called complementary length. The notion "complementary" refers to the fact that each stressed syllable includes either a long vowel or a long postvocalic consonant. Significantly, the site of length is predictable in many cases such that a stressed vowel in word-final or prevocalic position is necessarily long, whereas a stressed vowel preceding a cluster which does not occur word-initially (with the exception of obstruent-liquid clusters) is necessarily short. These particular constraints on the distribution of length appear to indicate determination by syllable structure: vowels are lengthened in open syllables as in (85a), ruling out the structure in (85a') whereas the postvocalic consonant is lengthened in a closed syllable as in (85b), ruling out the pattern in (85b')[46]. Clusters such as [nd], which cannot occur in onset position, are henceforth referred to as "closure clusters".

(85) a. a'. b. b'.

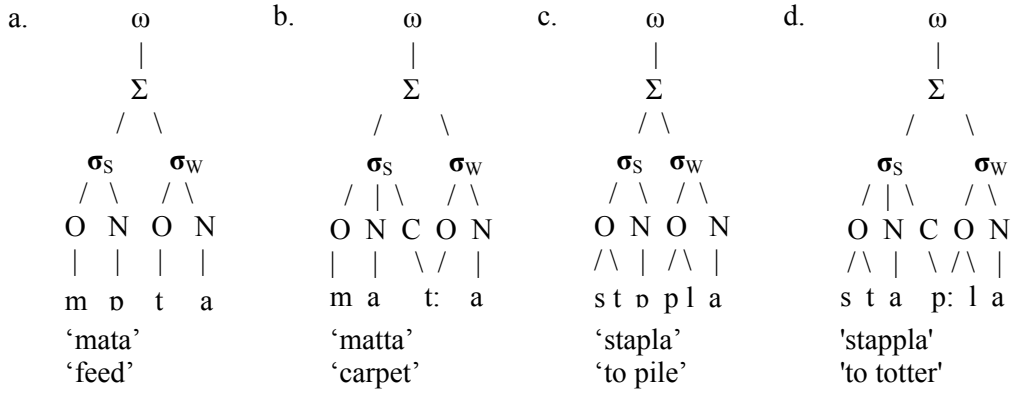


As in English, consonant clusters are not permitted in coda position, ruling out the structure illustrated in (86):

[46] According to the transcriptions in Hedelin's pronunciation dictionary (1997) postvocalic consonants are lengthened only in pword/foot-*final* position. By contrast, in Svenska Ord (1992), all such consonants are transcribed as long in stressed syllables. I follow the latter convention here.

(86) 

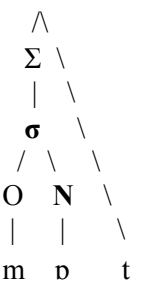
The description of length as a function of syllable structure can also be extended to cases where a vowel is followed by a single consonant (cf. (87a,b)) or a cluster with increasing sonority (cf. (87c,d)). In such cases the relevant length contrasts can be related to ambisyllabic versus non-ambisyllabic structure of the postvocalic consonant as is shown below:

(87) 

Occurrences of long vowels before a consonant in word-final position are analyzable as Final-C effects, assuming that the distribution of length concerns the phonological and not just the phonetic level.[47] On this view, vowel length in pword-final syllables results from the non-association of the pword-final consonant with the coda position. I will not discuss the various proposals of how to represent the special status of pword-final consonants[48] but tentatively associate such consonants directly with the pword-node as in (88). As was noted above, the special status of pword-final consonants also explains the occurrence of word-final clusters. Given the representation in (88c) there is no complex coda since the relevant consonants are not jointly associated with the coda position:

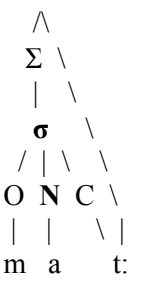
[47] Recall that the notion "Final-C effect" refers to the observation that pword-final consonants exhibit only the phonetic, but not the distributional, properties of coda segments.
[48] For discussion, see Hall 2002.

(88) a.



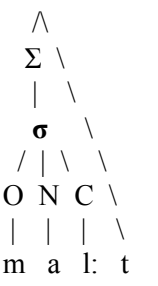
'mat' 'food'

b.



'matt' 'faint'

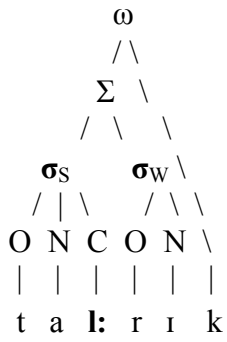
c.



'malt' 'malt'

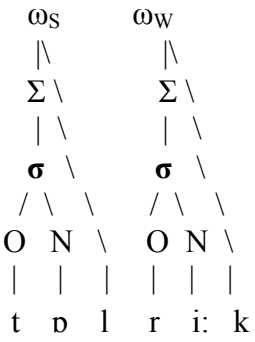
Word-internal Final-C effects indicative of complex morphological structure are demonstrated in (89b). Note that [lr] is a closure cluster inducing regular length of [l], rather than the preceding vowel, as in (89a). Vocalic length in (89b) qualifies accordingly as a Final-C effect, indicating the presence of a pword boundary, and hence a morphological boundary, after [l]. The word *talrik* is indeed a compound, consisting of the constituents *tal* 'number' and *rik* 'rich':[49]

(89) a.



'tallrik' 'plate'

b.



'talrik' 'numerous (number-rich)'

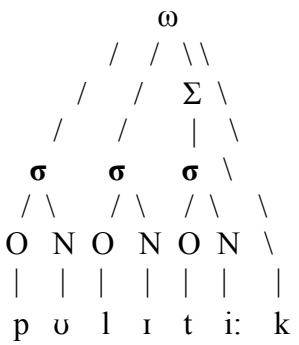
Turning now to the second type of boundary effect investigated here, relative prominence effects, we find that in Swedish the relation among two pwords within a morphological word is always strong-weak, regardless of the morphological or semantic properties of the word-internal constituents:

(90) If: $[\omega\omega]_{WORD}$ Then: $[\omega_S\ \omega_W]_{WORD}$

Relative prominence effects are easily detected because a pword usually contains only one foot, which comprises the rightmost syllable(s). The syllable heading that foot has main stress within the pword and undergoes lengthening as is shown in (91a). It is accordingly both the presence of two feet, manifested in two lengthening sites, and the weak prominence of the word-final foot, which show that *talrik* repeated in (91b) consists of two pwords. As in English, the evidence for word-internal pwords from relative prominence and from Final-C effects correlate systematically.

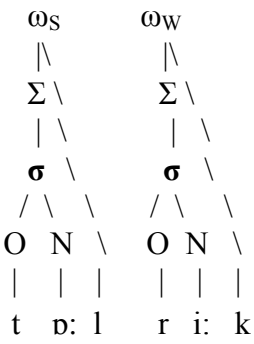
[49]According to the transcriptions in Hedelin, only lax vowels occur in unstressed syllables as in the final syllable in *tallrik*.

(91) a.



'politik' 'politics'

b.



'talrik' 'numerous (number-rich)'

Investigating now the phonological evidence for word-internal morphological structure in Swedish derivational morphology we again find the occurrence of boundary effects to be tied to the presence of specific non-cohering head affixes. All relevant suffixes are consonant-initial and include those which form a separate foot (and presumably a separate pword) listed in (92a), henceforth referred to as H-AFF-1, and those which lack stress and begin with a coronal sonorant in the left column in (92b). These suffixes are referred to as "recognized" suffixes below. Words with other suffixes exhibit no Final-C effects, which may indicate lack of suffix recognition:

(92)

		Recognized head suffixes	Unrecognized head suffixes
		<----------	---------->
a.	H-AFF-1	-bar, -lös, -full, -lek, -mål -sam, -het, -dom, -skap	
b.	H-AFF-2	-na, -nad, -ning, -lig, -ling	-ma, -ja, -ga, -ka, -ska, -sa, -sel

The subclassification of head affixes into those which form separate feet/pwords (H-AFF-1) and those which lack stress is supported by the evidence from lengthening in (93). In the words in (93a) each suffix includes either a long postvocalic consonant or a long vowel. By contrast, none of the (historical) suffixes in the words in (93b) is subject to lengthening (cf. *Svenska Ord* 1992).

(93) a. H-AFF-1

vérk[sàm:] 'verksam' 'effective'
sjúk[dὸm:] 'sjukdom' 'illness'
kráft[fɵl:] 'kraftfull' 'powerful'
ˈöm[sὶn:t] 'ömsint' 'tender'
rét[bˈɒ:r] 'retbar' 'irritable'
slˈäkt[skˈɒ:p] 'släktskap' 'kinship'
kˈär[lὲ:k] 'kärlek' 'love'
klár[hὲ:t] 'klarhet' 'clarity'
tíd[lˈö:s] 'tidlös' 'timeless'
slágs[mὸ:l] 'slagsmål' 'fight'
frˈåg[vὶ:s] 'frågvis' 'inquisitive'
ˈönsk[vˈɒ:rd] 'önskvärd' 'desirable'

b. H-AFF-2

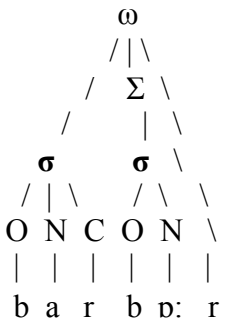
týst[nad] 'tystnad' 'silence'
ˈöv[nɪŋ] 'övning' 'practice'
sˈär[lɪŋ] 'särling' 'individualist'
lˈämp[lɪ(g)] 'lämplig' 'suitable'
kˈän[səl] 'känsel' 'perception of touch'
fét[na] 'fetna' 'to become fat'
fét[ma] 'fetma' 'fatness'
skíl[ja] 'skilja' 'to distinguish'
stád[ga] 'stadga' 'to consolidate'
jˈäm[ka] 'jämka' 'to adjust'
hˈäl[sa] 'hälsa' 'health'
vˈät[ska] 'vätska' 'fluid'

The question of whether or not a suffix forms a separate pword appears to be largely conditioned by segmental form and historical origin. All stressed suffixes in (93a)

consist of (s)CVC(C) sequences and typically originate from word-final compound members.[50] By contrast, (s)CV-suffixes are always stressless. The CVC-suffixes in (93b) differ from those in (93a) in that they are (historically) vowelless (i.e. *-sel* < *-sl* cf. Tamm 1897:79) or bimorphemic, consisting of *-n/-ing*, *-l/-ing* and *-na*/*-d*, respectively. Synchronically, these etymological facts are reflected in the presence of reduced vowels: epenthetic schwa in *-sel* and the corresponding raised allophone [ɪ] before high consonants in *-ning* and *-ling*.[51]

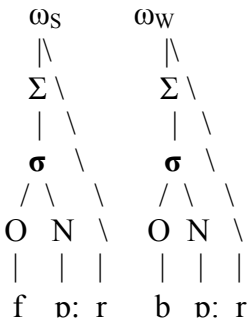
As in English, stress on a suffix appears in itself to ensure recognition, even if the suffix in question is unproductive. Recognition of the stressed suffixes in (93a) is demonstrated by regular relative prominence effects illustrated in (94). Specifically weak stress on the word-final foot in (94b) indicates that relative prominence is determined by the rule in (90).[52] The evidence from relative prominence correlates with the presence of two feet, manifested in two lengthening sites. By comparison, the word in (94a) illustrates the regular phonological form characteristic of simplexes. This word consists of a single pword dominating a single foot, which consists of the word-final syllable. This syllable alone is subject to lengthening.

(94) a.



[barbar]$_N$ 'barbarian'

b.



[[far]$_{ROOT}$[bar]$_{HEADSUFF}$]$_{ADJ}$ 'passable'

Additional examples are given in (95). The words in (95a) illustrate the regular prosody expected in simplexes or in words with cohering (i.e. vowel-initial) suffixes, respectively.[53] These words consist of single pwords dominating a single foot and consequently contain a single long segment. By contrast, each word in (95b) contains one of the head suffixes in (93a), which form separate pwords.[54]

---

[50]The boundary between compound members and affixes is notoriously fuzzy.

[51] Perhaps the reduction of *-lig* also involves a bimorphemic analysis, based on the independent suffix *-ig*.

[52]Alternatively, one could assume that the suffixes form a separate foot not integrated into the pword of the stem (e.g. (far)$_\omega$(bar)$_\Sigma$), where weak prominence would follow from the rule that a constituent occupying a lower position in the prosodic hierarchy has less prominence than a higher constituent. Crucially, this assumption, too, implies a word- internal pword boundary indicative of complex morphological structure.

[53] In (95) I have analysed as simplexes some words where others might posit morphologically complex structures (and possibly vice versa). The question of whether or not for instance the noun *kastrull* contains a suffix *-ull* (in analogy to nouns like *schatull*, *ampull*) is irrelevant as long as the suffix is vowel-initial. This is because vowel-initial suffixes are cohering with the result that the structure of the relevant words corresponds to the structure of simplexes. What matters is that none of the words in (95a) is a compound or includes a recognizable consonant-initial suffix.

[54] In the prosodic representations, indicated by parentheses, I have ignored the process of supradentalization characteristic of standard Swedish, whereby [r] followed by a dental consonant is

(95) a. (skandalǿ:s)$_{\omega}$ 'skandalös'
[[skandal]$_{ROOT}$[øs]$_{HEADSUFF}$]$_{ADJ}$
'scandalous'

(sʊlɪdɪté:t)$_{\omega}$ 'soliditet'
[[sʊlid]$_{ROOT}$[itet]$_{HEADSUFF}$]$_{N}$
'solidity'

(ɛsk͜ʉlʊ́:p)$_{\omega}$ 'eskulap'
[eskulap]$_{N}$
'medical doctor' (humorous)

(kʊpé:k)$_{\omega}$ 'kopek'
[kʊpek]$_{N}$
'kopeck'

(kɔlɛstərǿ:l)$_{\omega}$ 'kolesterol'
[kolestərol]$_{N}$
'cholesterol'

(dɪrɛktrí:s)$_{\omega}$ 'direktris'
[[direktr]$_{ROOT}$[is]$_{HEADSUFF}$]$_{N}$
'woman manager'

(mansʊ́:rd)$_{\omega}$ 'mansard'
[mansard]$_{N}$
'attic'

(teləgrám:)$_{\omega}$ 'telegram'
[telegram]$_{N}$
'telegram'

(kastrǿl:)$_{\omega}$ 'kastrull'
[kastrul]$_{N}$
'saucepan'

(labʏrín:t)$_{\omega}$ 'labyrint'
[labyrint]$_{N}$
'labyrinth'

b. (mʊ́ːka)$_{\omega}$(lʊ̀:s)$_{\omega}$ 'makalös'
[[maka]$_{ROOT}$[løs]$_{HEADSUFF}$]$_{ADJ}$
'matchless'

(ɧʏ́l:dɪ)$_{\omega}$(hè:t)$_{\omega}$ 'skyldighet'
[[ɧyldi]$_{ROOT}$[het]$_{HEADSUFF}$]$_{N}$
'duty, obligation'

(bʉ́:r)$_{\omega}$(skʊ̀p)$_{\omega}$ 'burskap'
[[bur]$_{ROOT}$[skap]$_{HEADSUFF}$]$_{N}$
'burgership'

(stʊ́:r)(lè:k) 'storlek'
[[stʊr]$_{ROOT}$[lek]$_{HEADSUFF}$]$_{N}$
'size'

(slák:s)$_{\omega}$(mò:l)$_{\omega}$ 'slagsmål'
[[slaks]$_{ROOT}$[mol]$_{HEADSUFF}$]$_{N}$
'fight'

(dɪstrík:t)(vì:s) 'distriktvis'
[[distrikt]$_{ROOT}$[vis]$_{HEADSUFF}$]$_{N}$
'districtwise'

(ɛ́l:sk)$_{\omega}$(væ:rd)$_{\omega}$
[[elsk]$_{ROOT}$[verd]$_{HEADSUFF}$]$_{N}$
'lovable'

(sé:də)$_{\omega}$(sàm:)$_{\omega}$ 'sedesam'
[[sedə]$_{ROOT}$[sam]$_{HEADSUFF}$]$_{ADJ}$
'modest, decent'

(lás:t)$_{\omega}$(fˆəl:) 'lastfull'
[[last]$_{ROOT}$[ful]$_{HEADSUFF}$]$_{ADJ}$
'depraved'

(ló:g)$_{\omega}$(sìn:t)$_{\omega}$ 'lågsint'
[[log]$_{ROOT}$[sint]$_{HEADSUFF}$]$_{ADJ}$
'mean'

The prosodic structures in (95) can accordingly be taken to indicate the recognition mechanism in (96), which has been established for English in section 3:

---

merged into a single supradental consonant. In the morphological representations, indicated by square brackets, I have omitted quantity and quality distinctions between vowels, assuming that all such distinctions are determined by syllable and foot structure.

(96)	Input:	Head affix recognition	"Rest" => root
	[barbár]$_{N}$	-	-
	[fárbàr]$_{ADJ}$	[fár[bàr]$_{H\text{-}AFF\text{-}1}$]$_{ADJ}$	[[fár]$_{ROOT}$[bàr]$_{H\text{-}AFF\text{-}1}$]$_{ADJ}$

Aligning the boundaries of both roots and affixes classified as H-AFF-1 with pword boundaries will then yield the correct output forms as is shown in (97).

(97)	Alignment	Output
	([barbár]$_{N}$)$_{\omega}$	(barbɒ́ːr)$_{\omega}$
	[([fár]$_{ROOT}$)$_{\omega}$([bàr]$_{H\text{-}AFF\text{-}1}$)$_{\omega}$]$_{ADJ}$	(fɒ́ːr)$_{\omega}$(bðːr)$_{\omega}$

Importantly, there is no alternative explanation for the correlating contrasts between the two phonological structures in (95a) and (95b). Specifically, paradigm uniformity constraints, which could in principle be invoked to account for vocalic length in *f*[ɒ́ː]*rbar* (cf. *f*[ɒ́ː]*ra* 'fara' to go'), do not explain the contrast in relative prominence between (95a) and (95b). Moreover, similar Final-C effects (i.e. vowel lengthening before a cluster with decreasing sonority) also occur in cases where no semantically related word exists which could potentially license vocalic length. For instance, vowel length in *b*[ʉ̶ː]*rskap* 'burgership' cannot be a PU-effect, since the only potential source for such an effect, the word *b*[ʉ̶ː]*r* 'cage', is unrelated.

Turning now to the suffixes classified as H-AFF-2 in (92), which do not form a separate foot, the bipartition into recognized versus unrecognized affixes is based on Final-C effects alone. Specifically, it can be shown that complex coda clusters or long vowels before closure clusters are stable when preceding a suffix with an initial coronal sonorant, but not before other unstressed suffixes. The examples in (98) illustrate the occurrence of Final-C effects in cases where "irregular" vowel length or coda clusters could not potentially constitute paradigm uniformity effects. This is because the relevant roots fail to correspond to semantically related words.

(98)	[fɒ́ːl+na] 'falna'	'to die down, fade'
	[vóːl+nad] 'vålnad'	'apparition, ghost'
	[ansέːn+lig] 'ansenlig'	'considerable, large'
	[hέmːp+ling] 'hämpling'	'linnet'

The analysis of "deviant" phonological structure in (98) as boundary effects based on the recognition of a head affix is illustrated in (99). The claim is again that the existence of base words which could potentially license vocalic length such as [steːl] 'stel' 'stiff' is not crucial for the presence of quantity 'violations' in the derived words. This is because the same sort of 'violation' is also seen in cases like *falna*, where a semantically related base word is lacking.

(99)	Input:	Head affix recognition	"Rest" => root
	[sté:lna]$_{V}$	[sté:l[na]$_{H\text{-}AFF\text{-}2}$]$_{V}$	[[sté:l]$_{ROOT}$[na]$_{H\text{-}AFF\text{-}2}$]$_{V}$
	[fɒ́ːlna]$_{V}$	[fɒ́ːl[na]$_{H\text{-}AFF\text{-}2}$]$_{V}$	[[fɒ́ːl]$_{ROOT}$[na]$_{H\text{-}AFF\text{-}2}$]$_{V}$

The word-internal morphological structure in (99) serves as the basis for alignment, as is shown in (100). As a result, the deviant vowel length is expected to be stable in historical perspective, regardless of potential PU-effects.

(100)	Alignment	Output	
	$([([\text{sté:l}]_{\text{ROOT}})_{\omega}[\text{na}]_{\text{H-AFF-2}}]_{\text{ADJ}})_{\omega}$	$((\text{sté:l})_{\omega}(\text{na})_{\sigma})_{\omega}$	'to stiffen'
	$([([\text{fɒ́:l}]_{\text{ROOT}})_{\omega}[\text{na}]_{\text{H-AFF-2}}]_{\text{ADJ}})_{\omega}$	$((\text{fɒ́:l})_{\omega}(\text{na})_{\sigma})_{\omega}$	'to die down, fade'

The crucial role attributed to suffix recognition in (99) is supported by the fact that verbs containing any of the (historical) unstressed suffixes not beginning with a coronal sonorant conform to the regular patterns occurring in simplexes. That is, vowels in closed syllables are short and coda clusters are absent, regardless of the phonological structure of the corresponding historical base words. Vowel length alternations resulting from the absence of word-internal boundaries in such cases are illustrated in (101). In the right column I list the etymological base words.

(101) a. -ma

$(\text{fét:ma})_{\omega}$ 'fetma' 'fatness'	cf. [fe:t] 'fet' 'fat'
$(\text{s´øt:ma})_{\omega}$ 'sötma' 'sweetness'	cf. [sø:t] 'söt' 'sweet'

b. -ga

$(\text{víd:ga})_{\omega}$ 'vidga' 'to widen'	cf. [vi:d] 'vid' 'wide'
$(\text{glød:ga})_{\omega}$ 'glödga' 'to make red-hot'	cf. [gløːd] 'glöd' 'live coal'
$(\text{nød:ga})_{\omega}$ 'nödga' 'to force'	cf. [nø:d] 'nöd' 'need'
$(\text{stád:ga})_{\omega}$ 'stadga' 'to consolidate, steady'	cf. [stɒ́:d] 'stad' 'stead'

c. -ja

$(\text{smør:ja})_{\omega}$ 'smörja' 'to smear'	cf. [smø:r] 'smör' 'butter'
$(\text{fér:ja})_{\omega}$ 'färja' 'ferry'	cf. [fɒ́:ra] 'fara' 'to go'
$(\text{vél:ja})_{\omega}$ 'välja' 'to choose'	cf. [vɒ:l] 'val' 'choice'
$(\text{kvél:ja})_{\omega}$ 'kvälja' 'to nauseate'	cf. [kvɒ:l] 'kval' 'pain, torture'
$(\text{tém:ja})_{\omega}$ 'tämja' 'to tame'	cf. [tɒ:m] 'tam' 'tame'
$(\text{vén:ja})_{\omega}$ 'vänja' 'to get used to'	cf. [vɒ:n] 'van' 'experienced'

d. -sa

$(\text{ɕøk:sa})_{\omega}$ 'köksa' 'kitchen-maid'	cf. [ɕǿ:k] 'kök' 'kitchen'
$(\text{h´ɛl:sa})_{\omega}$ 'hälsa' 'health'	cf. [he:l] 'whole'

e. -ska

$(\text{grøn:ska})_{\omega}$ 'grönska' 'verdure'	cf. [grø:n] 'grön' 'green'
$(\text{brøs:ka})_{\omega}$ 'brådska' 'hurry'	cf. [bro:d] 'bråd' 'hasty, busy'
$(\text{vɛ́t:ska})_{\omega}$ 'vätska' 'fluid'	cf. [vo:t] 'våt' 'wet'

f. -ka

(hál:ka)$_{\omega}$ ‘halka’ ‘to slip (and fall)’	cf. [hɒ:l] 'hal' ‘slippery, evasive’
(svál:ka)$_{\omega}$ ‘svalka’ ‘coolness’	cf. [svɒ:l] ‘sval’ ‘cool’
(dʏr:ka)$_{\omega}$ ‘dyrka’ ‘to adore’	cf. [dy:r] ‘dyr’ ‘dear, valuable’

g. -sel

(hœ́r:səl)$_{\omega}$ ‘hörsel' 'hearing'	cf. [hœ́:ra] 'höra' 'to hear'
(ɕœ́r:səl)$_{\omega}$ ‘körsel' 'transport (with horse and carriage)'	cf. [ɕœ́:ra] 'köra' 'to drive'
(stʏ́r:səl)$_{\omega}$ ‘styrsel' 'steering, control'	cf. [sty:ra] 'styra' 'to steer'
(ʏ́r:səl)$_{\omega}$ ‘yrsel' 'dizziness'	cf. [y:ra] 'yra' 'dizzy'
(vár:səl)$_{\omega}$ ‘varsel' 'foreboding, warning'	cf. [vɒ:r] 'var' 'cautious'
(jǿs:əl)$_{\omega}$ ‘gödsel’ ‘manure, fertilizer’	cf. [jǿ:da]$_{A}$ ‘göda’ ‘to feed up’
(bét:səl)$_{\omega}$ ‘betsel’ ‘bridle’	cf. [bí:ta] ‘to bite’

Given the lack of evidence for any sort of internal pword boundary the (historically) derived words in (101) are represented as single pwords.
The relevance of suffix recognition for the phonological form of words is especially striking in the word pairs in (102a-d), which were historically derived from identical bases (i.e. [fe:t 'fet' 'fat', [svɒ:l] 'sval' 'cool', [grø:n] 'grön' 'green, and [glø:d] 'glöd' 'live coal', respectively). In addition to relating to the same etymological base the word pairs in (102) exhibit comparable postvocalic consonant clusters, all of which qualify as closure clusters.

(102)

a.	((f[**e:**]t)$_{\omega}$na)$_{\omega}$	'to fatten'	(f´ɛ**t:**ma)$_{\omega}$	'fatness'
b.	((sv[**ɒ:**]l)$_{\omega}$na)$_{\omega}$	'to cool down'	(svá**l:**ka)$_{\omega}$	'coolness'
c.	((gr[**ø:**]n)$_{\omega}$ling)$_{\omega}$	'kind of carp'	(grø**n:**ska)$_{\omega}$	'verdure, green foliage'
d.	((gl[**ø:**]d)$_{\omega}$ning)$_{\omega}$	'glow, embers'	(glø**d:**ga)$_{\omega}$	'make red-hot'

The words in the right column (102) match the canonical forms of simplexes, where the coda of a stressed syllable contains a single long consonant, and are consequently represented as single pwords. The central empirical claim is that long vowels in such a phonotactic environment (i.e. before a 'closure cluster') can persist only if the postvocalic consonant is immediately followed by a pword boundary, as in the left column in (102).[55] Crucially, the presence of the pword boundary presupposes the parsing process, specifically the process of head affix recognition illustrated in (103). The subscript 'R' stands for the category 'ROOT'. The input words are represented orthographically to demonstrate that lack of suffix recognition will result in short root vowels in these cases, regardless of the quantity patterns in the input words. The lengthening site in the output forms is boldfaced in (103)

[55]The notion 'persistence' concerns primarily the process of language acquisition, referring to the likelihood that a child encountering a given output feature (e.g. a long vowel) will replicate that feature in her own speech. The impact of persistence is seen most clearly in historical perspective (cf. the discussion of English head prefixation above).

(103)	Input:	Head affix recognition	"Rest" => root	Alignment
a.	$[fetna]_V$	$[fet[na]_{H\text{-}AFF\text{-}2}]_V$	$[[fet]_R[na]_{H\text{-}AFF\text{-}2}]_V$	$([([\mathbf{fet}]_R)_\omega[na]_{H\text{-}AFF\text{-}2}]_V)_\omega$
	$[fetma]_N$	-	-	$([\mathbf{fet}ma]_N)_\omega$
b.	$[svalna]_V$	$[sval[na]_{H\text{-}AFF\text{-}2}]_V$	$[[sval]_R[na]_{H\text{-}AFF\text{-}2}]_V$	$([([\mathrm{sv}\mathbf{al}]_R)_\omega[na]_{H\text{-}AFF\text{-}2}]_V)_\omega$
	$[svalka]_N$	-	-	$([\mathrm{sval}\mathbf{k}a]_N)_\omega$
c.	$[grönling]_N$	$[grön[ling]_{H\text{-}AFF\text{-}2}]_N$	$[[grön]_R[ling]_{H\text{-}AFF\text{-}2}]_V$	$([([\mathrm{grö}\mathbf{n}]_R)_\omega[ling]_{H\text{-}AFF\text{-}2}]_N)_\omega$
	$[grönska]_N$			$([\mathrm{grön}\mathbf{s}ka]_N)_\omega$
d.	$[glödning]_N$	$[glöd[ning]_{H\text{-}AFF\text{-}2}]_N$	$[[glöd]_R[ning]_{H\text{-}AFF\text{-}2}]_N$	$([([\mathrm{glö}\mathbf{d}]_R)_\omega[ning]_{H\text{-}AFF\text{-}2}]_N)_\omega$
	$[glödga]_V$			$([\mathrm{glö}\mathbf{d}ga]_V)_\omega$

Similar types of prosodic contrasts can be illustrated with pairs consisting of a word with a stressed H-AFF-1-suffix and a word with an 'unrecognizable' suffix. The words in (104) are etymologically related to the simplexes [vɒ:r] 'var' 'cautious', [dy:r] 'dyr' 'dear, expensive', [jɛm:n] 'jämn' 'even' and [nʊ:d] 'ond' 'evil, sore', respectively.

(104)					
a.	$(\mathrm{vɒ}\mathbf{:r})_\omega$(sàm:$)_\omega$	'cautious'	(vá$\mathbf{r:}$səl$)_\omega$	'foreboding, warning'
b.	$(\mathrm{dý}\mathbf{:r})_\omega$(bɒ́:r$)_\omega$	'costly'	(d$\mathbf{Ɏr:}$ka$)_\omega$	'to adore'
c.	(jɛ́$\mathbf{m:n})_\omega$(hè:t$)_\omega$	'evenness'	(jɛ́$\mathbf{m:}$ka$)_\omega$	'to adjust'
d.	(ʊ́$\mathbf{n:d})_\omega$(sìn:t$)_\omega$	'malevolent'	(ʊ́$\mathbf{n:}$ska$)_\omega$	'malice'

Again, the words in the right column (104), which include a single long coda consonant, match the canonical forms of simplexes and are therefore represented as single pwords. The 'anomalies' observed in the left column in (104), including long vowels before sonorant-obstruent combinations, complex coda clusters, and weak word-final feet, indicate the presence of internal pword boundaries. The contrast in the prosodic structures of these cognates thus supports the key role attributed to suffix recognition shown in (105).

(105)	Input:	Head affix recognition	"Rest" => root	Alignment
a.	$[varsam]_A$	$[var[sam]_{H\text{-}AFF\text{-}1}]_A$	$[[var]_R[sam]_{H\text{-}AFF\text{-}1}]_A$	$[([\mathrm{v}\mathbf{ar}]_R)_\omega([sam]_{H\text{-}AFF\text{-}1})_\omega]_A$
	$[varsel]_N$	-	-	$([\mathrm{var}\mathbf{s}el]_N)_\omega$
b.	$[dyrbar]_A$	$[dyr[bar]_{H\text{-}AFF\text{-}1}]_A$	$[[dyr]_R[bar]_{H\text{-}AFF\text{-}1}]_A$	$[([\mathrm{d}\mathbf{y}\mathrm{r}]_R)_\omega([bar]_{H\text{-}AFF\text{-}1})_\omega]_A$
	$[dyrka]_V$	-	-	$([\mathrm{dy}\mathbf{r}ka]_V)_\omega$
c.	$[jämnhet]_N$	$[jämn[het]_{H\text{-}AFF\text{-}1}]_N$	$[[jämn]_R[het]_{H\text{-}AFF\text{-}1}]_N$	$[([\mathrm{jä}\mathbf{mn}]_R)_\omega([het]_{H\text{-}AFF\text{-}1})_\omega]_N$
	$[jämka]_V$	-	-	$([\mathrm{jä}\mathbf{m}ka]_V)_\omega$
d.	$[ondsint]_A$	$[ond[sint]_{H\text{-}AFF\text{-}1}]_A$	$[[ond]_R[sint]_{H\text{-}AFF\text{-}1}]_A$	$[([\mathrm{o}\mathbf{nd}]_R)_\omega([sint]_{H\text{-}AFF\text{-}1})_\omega]_A$
	$[ondska]_N$	-	-	$([\mathrm{o}\mathbf{n}dska]_N)_\omega$

The claim is again that suffix recognition is the crucial prerequisite for the persistence of the boundary effects. That is, the word-internal morphological structure in (105) serves as the basis for the alignment of morphological and prosodic boundaries, which is crucial for the stability of both the 'deviant' vocalic length and the 'deviant' consonant clusters in historical perspective.

Unlike the data examined so far, where the presence of boundary effects correlates systematically with the presence of specific head affixes, there are some cases of phonological anomalies in words containing an "unrecognized" suffix. Characteristic of these cases, boldfaced below, is both the coalescence of a voiced and a voiceless obstruent and the existence of variants with canonical sound patterns:

(106) ?[**bli:dka**] ~ (blíd:ka)$_{\omega}$ /(blít:ka)$_{\omega}$ 'blidka' 'to appease'
[**klé:dsəl**] ~ (klét:səl)$_{\omega}$ 'klädsel' 'dress, attire'
[**ví:gsəl**] ~ (vík:səl)$_{\omega}$ 'vigsel' 'wedding'
[**blý:gsəl**] ~ (blÝk:səl)$_{\omega}$ 'blygsel' 'shame'
[**drý:gsəl**] ~ (drÝk:səl)$_{\omega}$ 'drygsel' 'extensiveness'
[**trí:vsəl**] ~ (tríf:səl)$_{\omega}$ 'trivsel' 'well-being'

The intervocalic consonant clusters tend to undergo regressive voicing assimilation to conform to the canonical sound patterns of Swedish, in which case quantity patterns conform as well. For literate speakers the "anomalous" clusters have an independent source in spelling pronunciations which also accounts for the association of the boldfaced variants with careful, perhaps slightly hyperarticulated speech. Assuming that these clusters play a crucial role for the irregular quantity patterns observed in the boldfaced variants in (106) two explanations come to mind. First, the presence of such clusters could enhance suffix recognition, resulting in boundary effects (Final-C effects) in the relevant words. Second, the presence of such clusters could trigger recognition of relatedness to other words in the paradigm, giving rise to paradigm uniformity effects. The latter analysis is supported by the fact that each occurrence of 'anomalous' quantity patterns in (106) correlates consistently with the existence of a semantically related word, which could potentially license vocalic length. The relevant base words are listed in (107b):

(107)

a.		b.	
[bli:dka] 'blidka' 'to appease'		[bli:d] 'blid' 'mild'	
[klé:dsəl] 'klädsel' 'dress, attire'		[klé:da] 'kläda' 'to dress'	
[ví:gsəl] 'vigsel' 'wedding'		[vi:ga] 'viga' 'to wed'	
[blý:gsəl] 'blygsel' 'shame'		[blý:g] 'blyg' 'shy'	
[drý:gsəl] 'drygsel' 'extensiveness'		[drý:g] 'dryg' 'lasting, ample'	
[trí:vsəl] 'trivsel' 'well-being'		[trí:vas] 'trivas' 'to get on well'	

I tentatively conclude then that the phonology of the boldfaced variants in (106) is best analysed as a spelling pronunciation (of the intervocalic consonant cluster) in combination with paradigm uniformity effects (affecting quantity patterns). On either analysis the cases in (106) conform to the basic generalizations suggested by the Swedish data, which are consistent with the conclusions based on English:

- Swedish words exhibit correlating deviations from canonical sound patterns which match the boundary effects established for English (e.g.

Final-C effects, relative prominence effects) and indicate the presence of word-internal pwords.

- Pword boundaries align systematically with the boundaries of morphological constituents, determined by head recognition.

- A prerequisite for the recognition of the relevant head affixes concerns phonological form. In Swedish, both stress (H-AFF-1-suffixes) and initial coronal sonorants (H-AFF-2-suffixes) fulfill the condition for the recognition of head suffixes. Other (consonant-initial) suffixes are not recognized.

- Roots have no status other than 'rests', which remain after head affixes have been identified.

As in English, there is also a correlation between phonological factors for recognizing unstressed suffixes (i.e. the presence of initial sonorants) and productivity. The similarity between English and Swedish further extends to the correlations concerning head affixation versus modifier-head structures discussed at the end of section 4. In (108) I illustrate the relevant correlations by comparing the adjective *urusel* 'extremely bad', which consists of the modifier *ur-* and the head *-usel*, with the verb *erövra* 'conquer', which consist of the head prefix *er-* followed by the root *-övra*:

(108)	Output	$[[ur]_{MOD\text{-}AFF}[usel]_{HEAD}]_{WORD}$	$[[er]_{H\text{-}AFF\text{-}1}[övra]_{ROOT}]_{WORD}$
	Correlating affix properties:	paradigmatic variability syntagmatic autonomy	No paradigmatic variability less syntagmatic autonomy
	Prosodic properties:	$(\text{ʉ́}\text{:r})_{\omega}(\text{ʉ̀}\text{:səl})_{\omega}$ ("crisp boundaries")	$(\text{ǽ}\text{:r})_{\omega}(\text{ə̀ : vra})_{\omega} \sim (\text{ǽ}\text{:r}\text{ə̀ : vra})_{\omega}$ (possible fusion)
	Semantic properties:	necessarily inherent meaning	possibly no inherent meaning
	Affix function:	Semantic modification	Word class marking (verb)

Head prefixes and head suffixes in Swedish are alike in that some form a separate pword (e.g. *er-*) whereas others do not (e.g. *för-*, *be-*). This motivates the subclassification of the prefix *er-* in (108) as H-AFF-1. The classification into modifier-head structures versus head affix-root structures is based on the parsing rules introduced above:

(110)	Input:	Recognition of - head - head affix	2. "Rest" => - modifier - root
	$[urusel]_{ADJ}$ 'extremely bad'	$[ur[usel]_{HEAD}]_{ADJ}$	$[[ur]_{MODP}[usel]_{HEAD}]_{ADJ}$
	$[erövra]_{V}$ 'to conquer'	$[[er]_{H\text{-}AFF\text{-}1}övra]_{V}$	$[[er]_{H\text{-}AFF\text{-}1}[övra]_{ROOT}]_{V}$

Modifying prefixes can be freely omitted ('paradigmatic variability') whereas head affixes cannot be omitted. Paradigmatic variability correlates with 'syntagmatic autonomy', as is shown by the semantics and the prosody of the complex words. It is true that *ur-* means 'extremely' in combination with adjectives (e.g. *urfånig* 'extremely silly', *urgammal* 'extremely old') but 'initial, original' in combination with nouns (e.g. *urskog* 'virgin forest', *urinvånare* 'original inhabitant'). However the meanings in question, whether treated as cases of homonymy or polysemy, cannot be inferred on the basis of syntagmatic structure alone but must be considered inherent properties of the relevant modifying prefixes. By contrast, no clear meaning can be associated with the head prefix *er-* in *erövra*, despite the etymological relatedness to the modifying prefix *ur-*. Relative syntagmatic autonomy of the modifying prefix vis-à-vis the head prefix is also supported by prosodic evidence. That is, while both type of prefixes can form separate pwords the head affix tends to fuse with the root, forming a single domain of syllabification.[56] By contrast, the boundaries of modifying prefixes are consistently crisp, such that the prefix-final consonant cannot be syllabified as the onset of the following vowel. As in English, the relevant contrasts correlate with distinct affix functions. Whereas modifying prefixes contribute to the meaning of the complex word in a compositional fashion the primary function of head affixes is to mark membership in a word class.

6. Summary and discussion of related psycholinguistic work

In the present article I have discussed some implications of strictly prosodic evidence for underlying morphological structure and concomitant parsing procedures. It has been demonstrated that certain word-internal phonological boundary signals co-occur systematically and indicate the presence of coinciding morphological and prosodic boundaries. Specifically the presence of pword boundaries has been shown to indicate the relevance of head recognition in compounds (e.g. recognition of *warm* in *lukewarm*) and words derived by modifying prefixation (e.g. recognition of *polite* in *impolite*),as opposed to the relevance of affix recognition in words derived by head affixation (e.g. the recognition of *be-* in *begin*). Whereas head recognition involves paradigmatic knowledge, i.e. knowledge of (meaningful) relation to other words in the mental lexicon, the recognition of head affixes concerns the syntagmatic level only. In general, meaning plays a minor, if any, role for the recognition of head affixes but phonological form may be relevant.

The specific parsing mechanism indicated by word prosody is not necessarily to be understood as modeling the "online" processing of speech. Instead, this mechanism might affect the (initial) analysis of words and the prosodic form in which these words are subsequently stored, with no concomitant claim that such words are decomposed each time they are encountered in speech.

The question of what factors are relevant for the morphological analysis of words has also been addressed in psycholinguistic work. Below I will discuss some

[56] Prosodic fusion does not result in conformity to the structure of simplexes since the accent structure resulting from the previous complex pword structure (initial main stress, weak stress on the word final foot, cf. the rule in (90)) is retained. The resulting highly irregular prosodic structure is presumably unstable.

conclusions by Hay (2001),(2002) focusing on those which are inconsistent with the conclusion reached by interpreting the evidence from word prosody.

Hay's main claim is that "decomposability" of words is determined by relative frequency, meaning that a word is likely to be decomposed if the base is more frequent than the complex form. Hay's notion of decomposition conflates syntagmatic and paradigmatic analysis as she considers complex words to be items which can be "broken down" into components consisting of base words and affixes. By contrast, I have argued for a distinction between 'base words', which are paradigmatically related to but distinct from complex words in the mental lexicon, and notions such as 'roots' and 'affixes', which are components contained within complex words. Since frequency is a property of words, and not of roots or affixes, it can be relevant only to paradigmatic analysis. Indeed, the prosodic evidence does not support Hay's claims regarding the decomposability of complex words based on relative frequency. Compare the examples in the right column in (111), which according to Hay favor decomposition because the frequency of the base exceeds the frequency of the complex word as opposed to the examples in the left column in (111), which according to Hay favor non-decomposition because the frequency of the base is less than the frequency of the complex word (cf. Hay 2001:1048). The frequency data are based on the CELEX corpus.

(111)

word A	freq.	Base freq.	analysis:	word B	freq.	Base freq.	analysis:
abasement	6	2	[abasement]	enticement	3	64	[entice][ment]
alignment	57	44	[alignment]	adornment	41	75	[adorn][ment]
rueful	14	9	[rueful]	woeful	14	68	[woe][ful]
hapless	22	13	[hapless]	topless	27	3089	[top][less]
listless	42	19	[listless]	tasteless	30	402	[taste][less]

The claims regarding the analysis of the complex words given in (111) are based on speaker intuitions of perceived 'complexity', based on an experiment conducted in writing. The notion of complexity conveyed to the subjects was semantically based, involving the potential breakdown of words into "smaller, meaningful units" (Hay 2001:1048). However, there is no evidence that semantics or relative frequency play a crucial role in the decomposition of words derived by head affixation into constituent parts. Given that *-ment, -ful*, and *-less* are among the 'recognized' English head suffixes, I predict that *all* words in (111) are analysed as consisting of a root and a head affix as is illustrated with some of the relevant pairs in (112):

(112)

Input:	Head affix recognition	"Rest" => root
$[abasement]_{N}$	$[abase[ment]_{H\text{-}AFF}]_{N}$	$[[abase]_{R}[ment]_{H\text{-}AFF}]_{N}$
$[enticement]_{N}$	$[entice[ment]_{H\text{-}AFF}]_{N}$	$[[entice]_{R}[ment]_{H\text{-}AFF}]_{N}$
$[rueful]_{ADJ}$	$[rue[ful]_{H\text{-}AFF}]_{ADJ}$	$[[rue]_{R}[ful]_{H\text{-}AFF}]_{ADJ}$
$[woeful]_{ADJ}$	$[woe[ful]_{H\text{-}AFF}]_{ADJ}$	$[[woe]_{R}[ful]_{H\text{-}AFF}]_{ADJ}$
$[listless]_{ADJ}$	$[list[less]_{H\text{-}AFF}]_{ADJ}$	$[[list]_{R}[less]_{H\text{-}AFF}]_{ADJ}$
$[tasteless]_{ADJ}$	$[taste[less]_{H\text{-}AFF}]_{ADJ}$	$[[taste]_{R}[less]_{H\text{-}AFF}]_{ADJ}$

Aligning the outputs of the morphological parsing with prosodic boundaries yields internal pword boundaries, resulting in the output forms below (cf. the transcriptions in Wells (2000)):

(113)	Output		Output	
	((əbéɪs)$_{ω}$mənt)$_{ω}$	'abasement'	((ɪntáɪs)$_{ω}$mənt)$_{ω}$	'enticement'
	((əláɪn)$_{ω}$mənt)$_{ω}$	'alignment'	((əd´ɔ:rn)$_{ω}$mənt)$_{ω}$	'adornment'
	((rú:)$_{ω}$fəl)$_{ω}$	'rueful'	((wóʊ)$_{ω}$fəl)$_{ω}$	'woeful'
	((h´æp)$_{ω}$ləs)$_{ω}$	'hapless'	((t´ɑ:p)$_{ω}$ləs)$_{ω}$	'topless'
	((l´ɪst)$_{ω}$ləs)$_{ω}$	'listless'	((téɪst)$_{ω}$ləs)$_{ω}$	'tasteless'

The presence of internal pword boundaries in *all* words in (113) is supported not only by Final-C effects (cf. *abasement, alignment, listless*) but also by Containment effects (cf. *rueful*, discussed in (53),(54) and *hapless*, where the glottalization of [p] indicates strict coda syllabification, despite the following liquid). The parallel prosodic structures in *abasement* and *enticement* are also supported by the evidence from regressive voicing assimilation. Within pwords there is a strong tendency for the coronal fricative to be voiced before voiced consonants (cf. *pla*[zm]*a* 'plasma', *co*[zm]*opolitan* 'cosmopolitan' *a*[zm]*a* 'asthma'), but this tendency does not affect the root-final [s] in *abasement* and *enticement*, due to the intervening pword boundary.

The prosodic evidence thus clearly supports the analogous representations in (112). These findings are not surprising as the irrelevance of frequency to the syntagmatic analysis of words derived by head affixation has been demonstrated extensively above. Recall the presence of word-internal boundary effects in English *begin, relent, desire*, which cannot be associated with any base at all, as opposed to the absence of boundary effects in words such as *laughter* and *knowledge*, which relate to base words with far higher relative frequency but lack recognizable head affixes.

The most significant structural evidence cited by Hay in support of the relevance of relative frequency to morphological analysis concerns the noun *government* (cf. Hay 2002:542ff). This noun, being more frequent than its base *govern*, is assumed to be treated as a simplex. In support of this analysis Hay cites the formation *governmental*, as opposed to ungrammatical formations like **employmental* or **eagernessal*, the latter of which are claimed to be ruled out by a restriction of *-al*-suffixation to simplex words. Hay's analysis of the relevant restrictions in *-al*-formations in terms of morphological complexity is inconsistent with the prosodic evidence, which clearly indicates the complexity of the noun *government*. Specifically, the lack of stress on the closed penultimate syllable in conjunction with the Final-C effect indicate the presence of a word-internal pword boundary (i.e. ((góvern)$_{ω}$ment)$_{ω}$). This structure in turn supports the recognition of the suffix *-ment* described below:

(114)	Input:	Head affix recognition	"Rest" => root
	[góvernment]$_{N}$	[góvern[ment]$_{H\text{-}AFF}$]$_{N}$	[[góvern]$_{ROOT}$[ment]$_{H\text{-}AFF}$]$_{N}$

Given the analysis in (114) the restriction on *-al*-suffixation observed by Hay must have other explanations. The illformedness of **eagernessal* is in accordance with the general ungrammaticality of attaching a cohering (vowel-initial) suffix to a non-cohering (consonant-initial) suffix. Additional examples, some of which further demonstrate the irrelevance of relative frequency to the restriction in question, are given in (115):

(115) *(kind)$_{\omega}$ness+y
*(grate)$_{\omega}$ful+ize
*(fear)$_{\omega}$less+ish
*(reck)$_{\omega}$less+ity
*(hard)$_{\omega}$ship+ish

The existence of *-al*-suffixations based on *-ment*-derivations such as *governmental, developmental* or *argumental* is accordingly exceptional, perhaps explained by the fact that the suffix combination *-mental* is independently licensed in English (e.g. pairs of loanwords like *instrumental - instrument*, *ornamental - ornament*). The contrast in acceptability between **eagernessal* along with the other cases listed in (115) vis-à-vis *governmental* is accordingly explained by the absence of words which would license the relevant suffix combinations.

Consider next the contrast in acceptability between *governmental* and *employmental*, which Hay again explains in terms of relative frequency. Hay points out that the wellformedness of *governmental* versus **employmental* is not sufficiently explained by the resulting stress clash in **emplòyméntal.* This is because there are additional unattested words where stress clash would not be a problem (e.g. **nourishmental*, **managemental*) (cf. Hay 2002:544). However, the fact that there are many *-ment*-formations with root-final stress (e.g. *amázement*, *endéarment*) may play a role in stifling the productivity of *-mental* formations in general. Indeed *-al*-suffixation is also unacceptable based on words such as *testament* or *armament*, which would be simplexes on Hay's analysis.[57]

While there is no evidence for the relevance of relative frequency for the syntagmatic analysis of words derived by head affixation, relative frequency can be expected to be relevant for the analysis of compounds and words derived by modifying prefixation. This is because the analysis of such words crucially involves the recognition of a head corresponding to a paradigmatically related word. The evidence from word prosody indeed indicates a fundamental difference between the data in (111) involving head affixation, where differences in relative frequency play no role, and the data in (116) adopted from Hay (2001:1047), which involve modifying prefixation and the differences in relative frequency are relevant.

(116)

word A	freq.	Base freq.	analysis:	word B	freq.	Base freq.	analysis:
incongruous	55	3	[incongruous]	invulnerable	23	400	[in][vulnerable]
impatient	227	114	[impatient]	imperfect	50	1131	[im][perfect]
inanimate	34	4	[inanimate]	inaccurate	53	377	[in][accurate]
immobile	55	11	[immobile]	immodest	13	521	[im][modest]
immutable	40	4	[immutable]	immoderate	6	223	[im][moderate]

[57] The phenomenon that potential stress clashes in a fair subset of relevant coinages can stifle the productivity of suffix-combinations in general is well-attested in English. For instance, the lack of productivity of *-ity* with regard to adjectives in *-ive* can often be related to potential stress clashes (e.g. **abùsívity, *attràctívity, *constrùctívity*), a cause which may explain the general unacceptability of native *-ív-ity*-suffixation (e.g. **prìmitívity, *pòsitívity, *lùcratívity*). Similarly, the suffix combinations *-íst-ic*, *-ál-ity*, or *-ós-ity*, all of which have initial stress, lack full productivity. By contrast, the combinations *-abíl-ity, -ific-átion*, or *-iz-átion*, which never involve stress clashes, are entirely productive.

The relevance of frequency is reflected in the transcriptions given in Wells (2000), who lists a variant with stress on the prefix for all adjectives in the right column in (116), where the base is more frequent than the complex word, but, with the exception of *immutable*, does not list such a variant for the adjectives in the left column. These transcriptions indicate the prosodic structures in (117) (cf. the discussion of *iN*-prefixation above):

(117)

Output		Output	
(ɪŋkɑ:ŋgruəs)$_{\omega}$	'incongruous'	(ɪn)$_{\omega}$(vʌlnərəbəl)$_{\omega}$	'invulnerable'
(ɪmpeɪʃənt)$_{\omega}$	'impatient'	(ɪm)$_{\omega}$(pɜ˞fɪkt)$_{\omega}$	'imperfect'
(ɪnænəmət)$_{\omega}$	'inanimate'	(ɪn)$_{\omega}$(ækjərət)$_{\omega}$	'inaccurate'
(ɪmoʊbəl)$_{\omega}$	'immobile'	(ɪm)$_{\omega}$(mɑ:dɪst)$_{\omega}$	'immodest'
(ɪm)$_{\omega}$(mju:təbəl)$_{\omega}$	'immutable'	(ɪm)$_{\omega}$(mɑ:dərət)$_{\omega}$	'immoderate'

The structures in (117) indicate that heads are consistently recognized only if they occur more frequently than the complex word. This is plausible because the head in a complex word can be recognized only if it is known to the hearer, and relatively higher frequency would enhance the probability that the relevant word is established in the mental lexicon.[58] The data hence support the recognition process in (118):

(118)

Input:	Head recognition	"Rest" => modifier
[incongruous]$_{N}$	-	-
[invulnerable]$_{N}$	[in[vulnerable]$_{HEAD}$]$_{N}$	[[in]$_{MOD}$[vulnerable]$_{HEAD}$]$_{N}$

The fact that the stress in *immutable* indicates complexity may relate to the productivity of *-able*-suffixation, which possibly suffices to secure knowledge of the head *mutable* in the mind of the hearer.

To summarize, the parsing mechanism presented here has clear implications for the relevance of word frequency. In particular, it is predicted that relative frequency should be crucial for morphological analyses involving knowledge of paradigmatic relations, including the recognition of heads as described in 3.1.1. This is because the existence of a (paradigmatic) base can affect recognition only if the relevant word is known by the hearer. By contrast, the frequency of the (historical) base relative to the frequency of the complex word is predicted to be irrelevant in the case of head affixation as described in section 3.1.2. This is because head affixations are analysed on the basis of affix recognition, without recourse to paradigmatic knowledge.

[58] Cf. the example *handkerchief* discussed in footnote 19 and the examples involving modifying prefixes in (41).

References:

Anderson, Stephen R. (1992). *A-morphous Morphology*. Cambridge University Press. Cambridge.

Aronoff, Mark (1976). *Word Formation in Generative Grammar*. MIT Press.Cambridge, MA.

Beard, Robert (1995). *Morpheme-lexeme base morphology*. State University of New York Press. Albany.

Bochner, Harry (1993). *Simplicity in Generative Morphology.* Berlin: de Gruyter.

Botha, Rudolf P. (1981). "A base rule theory of Afrikaans synthetic compounding". In *The Scope of Lexical Rules,* M. Moortgat, H. van der Hulst, and T. Hoekstra (eds.), 1-78. Dordrecht. Foris Publications.

Burzio, Luigi (1994). *Principles of English stress*. Cambridge University Press. Cambridge.

Chomsky, Noam & Morris, Halle (1968). *The sound pattern of English*. New York:Harper & Row.

Comrie, Bernard (1985). Causative verb formation and other verb-deriving morphology. In *Language typology and syntactic description. Volume III.* T. Shopen (ed.) Cambridge University Press. Cambridge.

Cruse, Alan (2000). Meaning in language. Oxford University Press. Davis, Stuart and Cho, Mi-Hui (2003). The distribution of aspirated stops and /h/ in American English and Korean: an alignment approach with typological implications', *Linguistics* 41:607-52.

Fidelholtz, James L. (1967). *English vowel reduction*. Ms, MIT.

Hall, T.A. (2002). Against extrasyllabic consonants in German and English.*Phonology* 19: 33-75.

Harris, John & Edmund Gussmann (2003). *Word-final onsets*. Ms [Available on Rutgers Optimality Archive]

Hay, Jennifer (2001). Lexical Frequency in Morphology: Is Everything Relative? *Linguistics* 39:1041-1070.

- (2002). From Speech Perception to Morphology: Affix-ordering Revisited. *Language* 78:527-555.

Hedelin, Per (1997). Norstedts svenska uttalslexikon. Norstedts Förlag AB.

Jones, Daniel (1956). "The hyphen as a phonetic sign". *Zeitschrift für Phonetik und Allgemeine Sprachwissenschaft* 9:99-107.

Kenyon, John S. & Thomas A. Knott (1953). A pronouncing dictionary of American English. Springfield, Mass: Meriam.

Lapointe, Steven (1981). "A Lexical Analysis of the English Auxiliary Verb System". In *The Scope of Lexical Rules,* M. Moortgat, H. van der Hulst, and T. Hoekstra (eds.), 1-78. Dordrecht. Foris Publications.

Lexin (1992). Svenska ord - med uttal och förklaringar. Norstedts Förlag.

Lyons, John (1977). *Semantics*. Cambridge: Cambridge University Press.

Marchand, Hans (1969). *The categories and types of Present-day English word formation: a synchronic-diachronic approach*. 2nd edn. München: C.H. Beck'sche Verlagsbuchhandlung.

McCarthy, John and Alan Prince. 1993. "Generalized Alignment". In *Yearbook of Morphology* G. Booij and J. v. Marle (eds.) 79-153. Dordrecht: Kluwer.

Nespor, Marina & Irene Vogel (1986). *Prosodic Phonology*. Dordrecht: Foris.

Paul, Hermann (1880). *Prinzipien der Sprachgeschichte*. Tübingen. Niemeyer.

Pesetsky, David (1979). *Russian Morphology and Lexical Theory*" Ms MIT.

Piggott, Glyne (1999). At the right edge of words. The Linguistic Review 16.2: 143-185.

Plag, Ingo (1999). *Morphological productivity*. Berlin. Mouton de Gruyter.

- (2003). *Word-formation in English*. Cambridge University Press. Cambridge.

Prince, Alan & Paul Smolensky (1993). *Optimality Theory: Constraint Interaction in generative grammar*. Ms. Rutgers University and University of Colorado.

Raffelsiefen, Renate (1998). "Semantic stability in derivationally related words". In *Historical Linguistics 1995*, Volume 2. : R. M. Hogg & L. van Bergen (eds.) 247-268 John Benjamins Publishing Company. Amsterdam/Philadelphia.

- (1999). Diagnostics for prosodic words revisited: the case of historically prefixed words in English. In *Studies of the phonological word* T. A. Hall & U. Kleinhenz (eds.), 133-201. John Benjamins Publishing Company. Amsterdam/Philadelphia.

- (2004). *Phonological effects in word formation*. Habilitation thesis. Freie Universität Berlin.

- (2005). Paradigm uniformity effects versus boundary effects. In *Paradigms in Phonological Theory,* L. Downing, T.A. Hall und R. Raffelsiefen (eds), 211-262. Oxford University Press. Oxford.

Riddle, Elisabeth (1985). A historical perspective on the productivity of the suffixes *-ness* and *-ity*, In *Historical semantics, historical word-formation,* Fisik (ed.), 435-461. Berlin: Mouton de Gruyter.

Robins, R.H. (1967). *A short history of linguistics*. Longmans.

Selkirk, Elisabeth O. (1980). "Prosodic domains in phonology: Sanskrit revisited". In *Juncture: a collection of original papers*, Aronoff, and M.-L. Kean (eds.), 107-129. Saratoga, CA: Anma Libri.

- (1995).“The prosodic structure of function words”. In *Papers in Optimality Theory*, J.N. Beckmann S. Urbanczyck and L. Walsh (eds.), 439-469. University of Massachusetts Occasional Papers 18. Amherst.

Szpyra, Jolanta (1989). The morphology-phonology interface: cycles, levels and words. London and New York. Routledge.

Szymanek, Bogdan (1988). Categories and categorization in morphology. Lublin: Catholic University Press.

Tamm, Fredr. (1897). Om avledningsändelser hos svenska substantiv. Almqvist & Wiksells Boktryckeri-aktiebolag. Uppsala.

The American Heritage Dictionary (1985) 2nd ed. Houghton Mifflin Company. Boston.

Trubetzkoy, N. S. (1936). "Die phonologischen Grenzsignale". *Proceedings of the 2nd International Congress of the Phonetic Sciences.* Cambridge. 45-49.

- (1958). *Grundzüge der Phonologie*. Vandenhoeck & Ruprecht. Göttingen. Reprinted from the edition (1939) in Travaux du Cercle Linguistique de Prague 7.

Umeda, Noriko & C.H. Coker (1974). “Allophonic variation in American English,” *Journal of Phonetics* 2:1-5.

Webster's Ninth New Collegiate Dictionary. (1990) Merriam-Webster Inc., Springfield, Massachusetts.

Wells, J. C. (2000). (revised edition) *Longman Pronunciation dictionary*. Harlow: Addison Wesley Longman.

Williams, Edwin (1981). “On the notions ‘Lexically Related’ and ‘Head of a Word’”.*Linguistic Inquiry* 12, 245-274.

Learning Morphology by Itself[1]

Vito Pirrelli
Istituto di Linguistica Computazionale, Pisa, Italy
vito.pirrelli@ilc.cnr.it

Ivan Herreros
Univesitat Pompeu Fabra, Barcelona, Spain
ivan.herreros@ilc.cnr.it

Abstract

The paper reports on a few experimental results of a computer simulation of learning the verb morphology of Italian, English and Arabic with the same type of neural architecture based on Kohonen's self-organizing maps. Issues of the mental organization of the resulting morphological lexica are explored in some detail and discussed in the light of the differential distribution of regular and irregular inflections in the three languages. It is shown that typologically diverse, non trivial aspects of the underlying paradigmatic structure of the three verb systems effectively emerge through sheer exposure to realistic distributions of verb forms devoid of morpho-syntactic content. We argue that these results go a long way towards explaining how global organization effects in the mental morphological lexicon may eventually result from local word processing steps.

1.1 Introduction

The developmental acquisition of the inflectional system of a language requires the fundamental ability to identify, on the basis of a child's exposure to its unanalysed parental input, a repertoire of formal means of marking morphological contrast. In a deliberately simplified version of this task, the child's input can be assumed to be an unstructured list of independent word forms, already properly segmented out of their embedding phonetic stream, and perceived by a learner according to a certain probability distribution. Although this helps to focus on issues of word internal structure only, the task of morphological marker identification remains, for a number of reasons, a considerably hard one.

First, morphological markers are known to wildly vary cross-linguistically (Bybee 1985, Anderson 1992, Croft 2001, Stump 2001, Haspelmath 2002), thus leaving the learning child with an exceedingly unconstrained space of alternative hypotheses for word segmentation, ranging from affixation to templatic structures and reduplication. Secondly, they are poorly salient from a perceptual viewpoint, as they tend to appear in

[1] The present paper is the outcome of a joint, highly cooperative effort. However, for the specific concerns of the Italian and Spanish Academies, Ivan Herreros is responsible for sections 3, 4 and 5 and Vito Pirrelli for sections 1, 2 and 6.

G. Booij, *et al.* (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)*
Fréjus 15-18 September 2005, University of Bologna, 2007.
URL http://mmm.lingue.unibo.it/

phonologically weak, often unstressed, word boundary positions. Moreover, they convey fairly abstract and procedural semantic content (*i.e.* morpho-syntactic properties), having very few if any perceptual correlates in the grounding environment where words are uttered. Finally, when a language offers more than one realization of a given array of morpho-syntactic properties (indeed an unmarked case in the entire Indo-European family), multiple markers appear to cluster in paradigmatically-related classes, whose identification is part and parcel of the process of mastering the selection of the proper inflectional material, given a word's inflectional class. All in all, learning the inflection system of a language requires the development of a highly abstract classification system (which we may dub, in traditional linguistic terms, a "grammar") that, far from being an epiphenomenal by-product of a basically unstructured, whole-word lexicon, plays an active role in both on-line word processing and lexical access and representation (Marslen-Wilson et al. 1994).

In many respects, the learning task is reminiscent of the Harrisian goal of developing linguistic analyses (and ultimately a linguistic ontology of basic categories and atomic constituents) on the basis of purely formal, algorithmic manipulations, traditionally known as *discovery procedures*, of relatively raw language data (Harris 1951). Here (as with the problem of learning inflection broached above) a level of linguistic explanation is attained by first developing a *generalization method*, to then assess the obtained generalizations against some established theoretical background.

However, as we shall see in more detail in the following section, morphology learning cannot be simply equated with the linguist's job of establishing an ontology of morphological markers. Linguists rely on an extensive battery of *a priori* procedural knowledge (such as "morphologically complex words can be segmented exhaustively into non-overlapping constituent morphemes", "allomorphs tend to be arranged into a minimum number of disjunctive paradigm-based classes" etc.). This knowledge plays a fundamental role in ensuring convergence of Harrisian procedures on the sort of empirical generalizations aimed at by linguists. We are thus faced with the issue of whether children grow up equipped with the same battery of knowledge biases. In other words: where does all these *a priori* assumptions on word structure come to a learner from? Can we identify some basic cognitive mechanisms that are primary and foundational in the ontogenetic development of language acquisition with respect to more elaborated and specific categories of linguistic knowledge?

To address these questions, this paper presents a computer model of morphology learning that is intended to portray the learning task of marker identification as a process of *emergence* of morphological structure in the learner's mental lexicon. The approach is aimed at addressing a number of well-known aspects of cognitive development, such as the role of fluency and entrenchment in the ontogenetic development of procedural knowledge (Anderson 1993, Boyland 1996), the impact of sequential distributions on aspects of reduction in the individual articulatory gestures of word production (Bybee 2002), morphological irregularization, global effects in the morphological organization of both lexicon access and lexical representation (as opposed to whole-word models of the speaker's mental lexicon), well-known effects of local similarity in on-line morphological processing (Albright 2002), graded morphological structure (Hay and Baayen 2005). To anticipate some of the conclusions we shall draw in the final part of the paper, computer modelling of language learning, with its strong reliance on probability distributions and machine learning algorithms, may apparently bear little resemblance to traditional theoretical accounts of inflectional morphology. It may turn

out that the fine-grained levels of explanation offered by computational simulations of morphology learning are not straightforwardly amenable to traditional grammatical categories. Yet, we agree with Goldsmith and O'Brien (2006) that simulating the emergence of complex levels of morphological organization in the mental lexicon is by no means incompatible with the view that speakers internalize a complex body of abstract linguistic competence. As we will show in the following pages, such a body of abstract knowledge is more intimately related to usage-based aspects of the language input than some theoretical linguists are ready to recognize.

The paper is structured as follows. Section 2 overviews different approaches to the problem of morphology learning in the light of the above-mentioned cognitive requirements and recapitulates some of the hardest challenges in modelling what we know about human morphological behaviour. Section 3 provides an introductory description of so-called Self-organizing Maps (Kohonen 2001), a member of the family of competitive neural networks exhibiting a topological behaviour that is particularly suitable for modelling the dynamics of lexical organization and on-line morphological processing.. Sections 4 e 5 outline the neural architecture used for our experiments and review a few learning results obtained on typologically diverse training data. Finally, in section 6 we draw some conclusions, carve out our future research agenda and sketch some prospective work.

2. Background

The acquisition and mastering of productive systems of inflectional morphology in natural languages are known to be extremely difficult tasks. Most adult second language learners develop relatively fixed syntactic constructions, with words typically occurring in one morphological form only (Klein and Perdue 1997, Wilson 2003). Similarly, pidgin and Creole languages are characterised by a relatively impoverished system of inflectional morphology. Moreover, inflectional competence, in both adults and children's language behaviour, tends to be relatively brittle and break down fairly quickly under various kinds of processing pressure and language impairments (Dick et al. 2001).

In this section, we shall focus on what we take to be a logically preliminary step in morphology learning: the process of scanning a word form through, to search for its morphological formatives. In particular, we shall mainly be concerned with the issue of identifying markers of inflectional categories such as person, number, gender, tense and mood, which are known to form the grammatical backbone of conjugational paradigms and constitute a primary goal of early efforts of morphology learning in child language maturation. The problem is traditionally conceptualized as the task of splitting an inflected word form into its constituent morphemes. As the notion of morpheme in both theoretical and cognitive linguistics has been the locus of much controversy over the last thirty years or so (a debate upon which dust does not seem to have settled yet), we deliberately sidestep the issue of the ontological status of inflectional and lexical formatives, to portray the task under scrutiny as the somewhat preliminary and more fundamental goal of identifying recurrent elements of formal realization of morphological contrast. To be more concrete, we would like to focus on the ontogenetic process by which an Italian child, exposed to verb forms such as *amiamo* 'we love', *canto* 'I sing', *vengo* 'I come' etc., is able to identify the recurrent segmental units

'*-iamo*' and '*-o*' as typical (albeit not necessarily exclusive or minimal) carriers of information about person and number in the Italian present indicative sub-paradigm.

Even when stated in these simplified terms, the problem is considerably harder than expected. First, the child has no way to know, a priori, where inflectional formatives should be looked for within a verb form. Her/his search space is thus potentially very large: a huge haystack with comparatively few morphological needles. Secondly, the amount of formal redundancy exhibited by verb forms in a given language goes well beyond the limited range of recurrent morphologically relevant markers. Rhyming words, false friends, false prefixes and the like are virtually ubiquitous and tend, at least in principle, to obscure morphologically relevant analogies. We may refer to this as the background noise problem. On top of that, relevant analogies happen to be often confined to one segment only, in the perceptually weak coda of a word final syllable. Even in the same language, prefixation, suffixation and stem alternation often present themselves simultaneously in tricky combinations. Particularly in highly frequent irregular or subregular verb forms, more strategies of morphological marking appear to often be overlaid, to the point that formal discontinuity is a prominent feature even of those languages that do not exhibit non-concatenative morphology.

It is very difficult, for a non linguist, to disentangle herself/himself from such an intricate coil of input evidence. The machine learning literature, with its large array of assumptions about algorithmic searching of formal redundancies, has enormously contributed to shed light on these and related issues (Pirrelli 2003 and references therein). The apparently naïve question at the heart of our investigation is thus the following: what does it take for a child to become sensitive to few morphologically relevant formal analogies and remain blind to the very many ones bearing no or scanty relationship to abstract principles of grammatical organization in the morphological lexicon? Linguists have often confronted themselves, in either direct or indirect ways, with this puzzling question. In the following section we briefly recapitulate some of the most influential answers in the literature. This will lead us to talk about the problem of the possible sources of the knowledge required by a child to home in on the appropriate battery of language-specific markers of inflectional features.

2.1. Nativism

According to a well-established nativist position, emanating from the generative approach to adult grammar competence of Chomskyan inspiration, children are equipped with an innate set of options for acquiring distinct language types. According to this view, the extensive cross-linguistic variation exhibited by the morphology of human languages can be explained by positing language-specific ways of setting these options, called "parameters", in the grammar word module. In particular, it has been argued that the child scans her/his linguistic environment for designated structures or "cues", to be found in the mental representations which result from hearing, understanding and parsing words (Lightfoot 1999, Hyams 1986, 1996). Cues which are realized only in certain typological families of grammar constitute the parameters. For example, upon understanding a word form like *sneezing*, the child comes up with an abstract representation such as $[_{V}[sneeze]\ _{progr}[ing]]$ allowing her/him to set a word-final parameter concerning the position of inflectional markers in the English verb. Surely,

the parameter can only be set when a child has already homed in on a partial analysis which treats *sneeze* and *-ing* as separate sub-word constituents, the latter being interpreted as a marker of abstract morpho-syntactic features. Thus, the availability of valuable cues for morphology learning *presupposes* an appropriate segmentation of *sneezing* rather than providing a principled solution to the haystack search problem.

2.2. Connectionism

Over the last twenty years, connectionism has challenged the symbolic view of morphological processing dominant in the Chomskyan tradition to provide a coherent alternative approach to the issue of learning word internal constituents. One of the most articulated and full-fledged recent illustrations of this proposal (Plaut and Gonnerman, 2000) views morphology as an interface realm, emerging as a pattern of activations in the layer of hidden units mediating the relationship between lexico-semantic and phonological word representations in an artificial neural network (Figure 1).

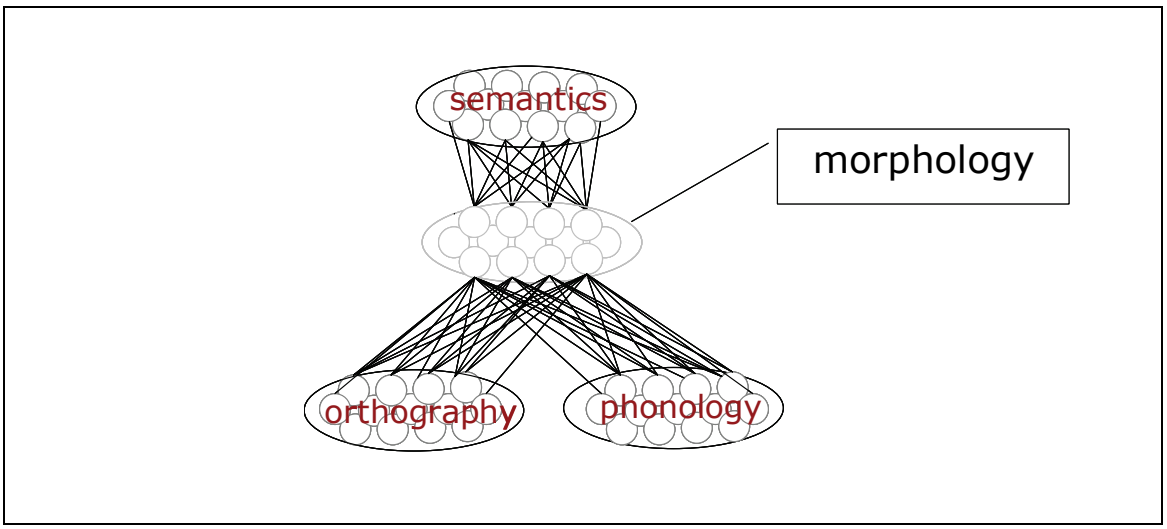


Figure 1: A connectionist framework for lexical processing (adapted from Plaut & Gonnerman, 2000)

To the extent that a particular surface pattern occurs in many words and maps consistently to certain aspects of lexical meaning, the representation conveyed through the internal (hidden) layer as an array of activation states will come to reflect this mapping, and will process it relatively independently of other parts of the word. This developmental process accounts for gradient effects of morphological structure, with intermediate degrees of morphological transparency being related to intermediate degrees of either phonological or semantic transparency (Plaut and Gonnerman 2000, Hay and Baayen 2005). For our present concerns, the interest of this proposal rests on the possibility that the child's hypothesis space be effectively constrained by relating the search for formal redundancies to the existence of shared semantic representations. This should considerably limit the combinatorial explosion of useless mappings between deceptively similar word forms (background noise), but does not seem to address our segmentation problem in a principled way, for two basic reasons. First it requires that children have access to highly complex, fully developed lexico-semantic

representations, whose early availability in the parental input to the child is somewhat moot. In fact, we have evidence that the acquisition of abstract morpho-syntactic categories and a full understanding of their role in language processing tend to occur at a comparatively late stage of language maturation, when the child has already mastered those aspects of morphological realization and marker selection we are presently concerned with (Clahsen 1989, Wilson 2003). Secondly, connectionist representations of the phonological input of inflected word forms do not offer a principled account of the word mapping problem. This point is illustrated by the input word representations used by Plunkett and Juola (1999) for experiments on learning English noun plural inflection (illustrated in Figure 2 for the words *cats* and *oxen*). Input representations are obtained by integrating phonological and morphological information into a fixed-size template-like structure (where the segment sequence /ts/ in /kAts/, for example, is, contra phonological evidence, split by an intervening empty vowel slot), with the result of enforcing a built-in alignment between input representations of words selecting different inflectional endings. The alignment has the effect of slipping in a strong language-specific bias that appears to *presuppose,* rather than explain, the problem we are presently concerned with.

C	C	C	V	V	C	C	C	V	C
#	#	k	#	A	#	#	t	#	S
#	#	#	#	O	#	k	s	E	N

Figure 2: A templatic input representation for English noun plurals (from Plunkett and Juola 1999)

2.3. Distributionalism

Harris' assumption that morphological categories can be derived mechanically from an analysis of the distributional properties of word forms in context has the potential of addressing the range of questions we are concerned with in this paper. According to Harris' view (Harris 1951), identification of relevant inflectional formatives is the final result of building a statistical model of the way overt, perceptually salient strings of phonological segments follow each other in the language input which is the ultimate object of linguistic investigation. Endorsing a somewhat radical mistrust in the role of semantic or more generally non perceptually overt knowledge in language analysis (Matthews 1993), Harris delineates a purely formal methodology whereby the only evidence available to the linguist is made up out of strings of linguistic units and their distributions. His approach, after a long-lasting obsolescence, has recently played an inspirational role for a number of machine learning approaches to unsupervised morphology acquisition (Gaussier 1999, Goldsmith 2001, Schone and Jurafsky 2000, Creutz and Lagus 2004, Wicentowsky 2004).

In a recent adaptation of Harris' ideas, John Goldsmith (2001, 2006) casts the distributional hypothesis into a powerful information theoretic framework, known as Minimum Description Length (MDL, Rissanen, 1989). Starting form the assumption that morphological information about a language can hardly be reduced to *local* information about letter bigrams or trigrams of that language, Goldsmith frames the task as a data compression problem: "find the battery of inflectional markers forming the shortest grammar that best fits training evidence", where i) a grammar is a set of paradigms defined as lists of inflectional markers applying to specific verb classes and

ii) the training evidence is a text corpus. The task is a top-down global optimization problem and boils down to a grammar evaluation procedure. Given a set of candidate markers, their probability distribution in a corpus and their partitioning into paradigms, MDL allows calculation of i) the length of the grammar (in terms of number and size of its paradigms) and ii) the length of the corpus generated by the grammar (*i.e.* the set of inflected forms licensed by the grammar according to a specific probability distribution). In MDL, the notion of length is derivative of the information theoretic notion of the number of bits required to encode linguistic units, whether they are stems, suffixes or word tokens. Intuitively, minimising the length of the corpus in bits requires that very frequent tokens should be assigned a shorter bit code than less frequent tokens. Minimising the length of the grammar, on the other hand, requires that frequently used paradigms are given preference to rarely used ones, as the cost of encoding a rare paradigm in bits is very high. Hence, a good language model is the one where the sum of the length of the grammar and the length of the corpus generated according to the probability assigned by the grammar is smallest. This policy disfavours two descriptively undesirable extremes: a corpus-photograph model, with a very long grammar where each verb form has, as it were, a paradigm of its own, such that the inflected forms generated by the grammar have the same probability distribution found in the corpus; and a very short but profligate model, with one paradigm only, where any verb combines with any marker according to the product of their independent probability distributions, thus generating many word forms that are not attested in the training corpus (including *goed* for *went*, *stricked* for *struck*, *bes* for *is* etc.).

From a cognitive perspective, Goldsmith's approach has the merit of addressing the problem of morphology learning with no recourse to prior language-particular knowledge. Furthermore, he adopts a mathematical framework where the development of morphological knowledge can be viewed as the emergent result of data compression, arising, both phylogenetically and ontogenetically, from the pressure of keeping a potentially unbounded amount of lexical knowledge in a finite memory store. We find these ideas fundamentally correct. On a less positive note, in Goldsmith's approach the issue of morpheme segmentation is kept separate from that of morpheme inventory evaluation, both logically and algorithmically. The two learning phases make no contact, so that we are left with no principled answer to the problem of the interplay between word processing and morphological organization in the speaker's mental lexicon: does morphological organization play any role in word processing?

Moreover, it is hard to see how a child learning morphology can possibly be engaged in such a top-down search for global minima. What we know about word processing in human subjects supports the view that speakers are extremely sensitive to *local* similarity maxima and tend to analyse and generate novel word forms predominantly (if not exclusively) by analogy to their closest cognates. For example, Italian speakers appear to be able to use fine-grained classes of verb stems to assign them the appropriate conjugation paradigm. According to Albright (2002), Italian speakers are able to assign a 0.937 conditional probability to the event that an *X*[end] verb stem is inflected for the second conjugation class. This means that when an Italian speaker is exposed to the nonce 1s present indicative form *trendo*, (s)he is almost certain that its infinitive is *trendere* (and not *trendare* or *trendire*). Most evidence for such an acute sensitivity to local similarity comes from *irregularly inflected* verb forms (but see again Albright 2002, for similar effects with regular verbs) and is often contrasted, in the psycholinguistic literature, with the somewhat opposite tendency

towards using default rules in the production of *regularly inflected* forms (Say and Clahsen 2001). According to many scholars (Pinker and Prince 1988, Prasada and Pinker 1993, Marcus et al. 1995 among others), the contrast supports a *dual route model* of word processing and learning: irregular forms are stored in full and are generalized over by local similarity, while regular forms are stored and indexed by their roots and affixes and produced by default rules of some kind. Other scholars oppose to such view and argue in favour of a unitary underlying mechanism accounting for both regular and irregular forms (Rumelhart and McClelland 1986, Plunkett and Marchman 1991, Bybee 1995, Ellis and Schmidt 1998 among others).

We have no room here to address this debate in any detail. Suffice it to point out that, for our present purposes, we are faced with an apparent paradox. We agree with Goldsmith that learning the morphology of a language can be framed, in machine learning terms, as a global optimization problem: morphologically relevant analogies (unlike local, potentially misleading similarities) emerge from a global analysis of the available input evidence. On the other hand, we are forced to reconcile this truth with the fact that speakers use local analogy-based strategies to develop morphological generalizations. The somewhat paradoxical question then is: how can a learner home in on global, paradigm-based analogies on the basis of local processing strategies?

In the remainder of this paper we intend to show that Self-Organizing Maps (SOMs), a particular family of artificial neural networks, can offer an interesting way out of this apparent paradox. As we shall see, SOMs can develop topological maps of input stimuli where the latter are organized according to global classification criteria. This is so in spite of the fact that SOMs learn and process input stimuli on the basis of principles of purely local analogy as will be shown in the following section.

3. Self-Organising Maps

3.1. Brain Maps

A Self-Organizing Map (hereafter *SOM*, Kohonen 2001) is an unsupervised machine learning algorithm drawing considerable neuro-physiological inspiration from the behaviour of so-called "brain maps". Brain maps are medium to small aggregations of neurons found on the cortical area of the brain that are involved in the specialized processing of specific classes of sensory data. Processing simply consists in the activation ("firing" in neurophysiological terms) of a certain neuron (or neuron aggregation) each time a particular stimulus is presented. The associative links between a stimulus and its firing neurons are described, in neurophysiological terms, as "mapping". A crucial feature of the sort of mapping performed by brain maps is that similar stimuli fire nearby neurons. As we shall see in more detail later on, such a local sensitivity to similarity in the presented stimuli develops inside a globally ordered topological structure. This is so because local mapping must obtain over the entire brain map area, thus enforcing an incremental principle of global organization of firing neurons. Examples of brain maps are i) the *somatotopic* map where stimuli generated in different parts of the body are mapped onto different specialised areas, ii) the *tonotopic* map where neurons respond to sound stimuli according to the frequency of the sound or, iii) the colour map on the visual area V4. The genesis of such brain maps is also

interesting for our present cognitive concerns. Although some of them can be considered as genetically pre-programmed, there is evidence that at least some aspects of such global neural organizations emerge according to the sensory experience of the subject (Jenkins et al. 1984, Kaas et al. 1983).

In 1984, Teuvo Kohonen described an iterative, unsupervised Artificial Neural Network (ANN) exhibiting some salient characteristics and behaviour of brain maps. Each unit/node of an ANN can be viewed as a receptor neuron that reacts to (or is activated by) a particular class of stimuli only. A node is an independent processing unit associated with a small memory trace that stores the stimulus the node is sensitive to. The more faithful the trace, the more sensitive the receptor. From this perspective, simulating the behaviour of a brain map is tantamount to developing an incremental ANN where similar stimuli trigger topologically neighbouring nodes.

The SOM learning algorithm is iterative. At each iteration the network is exposed to a random input stimulus. The first phase of the iteration consists in a network activation, culminating in the identification of the best matching unit on the map. The best matching unit (*BMU*) is the receptor whose memory trace happens to be closest to the current input. Memory traces are sometimes called prototype vectors (because they are represented as vectors), but they can also be referred to simply as the unit memory content. If we consider that the input is just another vector, the search for the best matching unit simply consists in finding the map node that contains the vector most similar to the input. Returning to the analogy to brain maps, the best matching unit in an ANN plays the role of a real neuron(s) being fired in a brain map.

In a SOM, the activation part of a learning iteration is followed by an updating phase. The memory contents of a certain number of map units are updated for them to look closer to the new information provided by the last input stimulus. An update consists in adjusting a number of memory traces to the input pattern just presented to the map. Using a slightly far-fetched metaphor, we can describe the neuron memory as a camera film being repeatedly exposed to an image at very short time intervals (learning iterations) for an amount of time insufficient for a clear image to imprint the film one-shot. At each exposure, the image on the film resembles more and more closely the input image the film is exposed to. It is important to appreciate that the update process is undergone neither by all neurons, neither at the same rate for all involved neurons. In fact two parameters, the neighbourhood radius and the learning rate, govern the learning process in determining, respectively, the number of units being updated at each iteration and the amount of incremental adjustment at each time tick. Both parameters play a crucial role in the dynamics of the learning process and decrease as learning progresses.

3.2. The neighbourhood radius

After the BMU is identified, a number of neurons are updated: these include the BMU itself and a set of its neighbouring units on the map, within a distance (from the BMU) defined by the neighbourhood radius. At the beginning of the learning process, the radius is long enough to guarantee that large neighbouring areas of the map are updated at each iteration. This ensures that a global order is enforced upon memory traces. Finer-grained relationships are learned at later stages, when the neighbourhood radius is progressively reduced in the course of learning. This defines a fundamental dynamics of a SOM learning trajectory to which we shall return later.

3.3. The learning rate

The learning rate defines the amount by which the memory content of each unit is modified at each iteration. At early stages of learning, the rate is kept high, thus allowing memory traces to quickly adjust to input data. As learning progresses, however, the rate decreases and memory traces gain in stability.

3.4. Plasticity

The joint effect of the dynamics of both neighbourhood radius and learning rate defines the so-called network plasticity, i.e. the capability of a map to modify its content to adapt it to input data. At early stages, the map content is extremely unstable and adaptive due to a long neighbourhood radius and a high learning rate. In the end, the map plasticity reduces considerably, thus allowing for a process of fine tuning only. As a result of this joint dynamics, a SOM can learn the global order underlying input data only at early stages, when plasticity is high and the map topology can be modified easily. By the same token, it is only when plasticity goes down and the network becomes more stable that fine-grained distinctions are acquired.

3.5. Frequency effects on a self-organising map.

SOMs are very sensitive to input frequency. To better understand this point, it is important to bear in mind that the basic task of a SOM is to accommodate input stimuli on its surface by associating them with corresponding memory traces. If there is enough room on the map, then every input stimulus will be assigned a faithful trace though learning. For lack of room on the map, on the other hand, similar input stimuli will tend to compete for the same memory traces. In this competition, both stimulus token frequency and stimulus class frequency play a key role. By their being repeatedly exposed to the map, high token frequency inputs are bound to carve out a map area where they are memorized faithfully, even if they form a class of their own: due to their high token frequency, they can in fact win the competition by themselves. On the other hand, low token frequency stimuli will leave a memory trace on the final map only if they are part of a high frequency class of stimuli, that is a class where the sum of token frequencies of its member is high.

This has a simple probabilistic interpretation. For example, in the case of the Italian past participle, if we consider the class of verb forms ending in *-ato*, its class frequency will tell us how likely we are to find a member of that class in a given corpus. This has also implications in terms of memory traces. When memory traces are exposed not to a single stimulus type, but to an entire class of similar stimuli, they will tend to reflect, for lack of sufficient room, what the class members have in common. Let us suppose we are running a toy experiment where the word *cantato* (‘sung’) leaves a memory trace on the final SOM only because it is part of a high frequency class whose other two members are *amato* (‘loved’) and *pensato* (‘thought’). As these three forms share the same memory trace, the latter will reflect the commonalities partaken by them, for example the ending *-ato*. The case illustrates a simple effect of “generalization as a shortage of memory”.

On the other hand, if a high frequency stimulus forms a class of its own, the particular memory trace (or memory area) fired by it on the map should be able to represent it faithfully. This is what happens, for example, when we find that the form *said* is fully memorized on an English past participle map. Note that this situation is the mirror image of what we found out in the previous paragraph: in fact, full storage of a very frequent input leaves no room for generalization. The natural question at this juncture is: if words with high token frequency are fully memorized, what is the relationship between them and other partially memorized, less frequent words which nonetheless belong to the same class as the former?

So far we have discussed how both token and class frequency affect a) the possibility for a stimulus to be learned by a SOM and b) the kind of memory trace the stimulus is likely to leave on the output map. In both cases we have been looking at the end result of learning. It is now time that we turn to discussing in some detail what happens *during* the learning process as such.

In the process of learning, traces memorised on a SOM slowly approximate original input representations. If an input is presented a number of times exceeding a certain threshold, the SOM will contain a memory trace with a faithful representation for that input. According to the dynamics of the learning process, however, the minimum threshold goes down as learning progresses (and the neighbourhood radius decreases). In other words, the rate at which an input is presented to the map has a direct impact on the rate at which the input is "learned". Intuitively, single tokens are going to be learned at a rate which is proportional to their frequency. Similarly, frequent stimulus classes are learned earlier than less frequent classes (if the latter are going to be learned at all).

To be more concrete let us turn to the map of Italian past participles. As *stato* ('been') is a very frequent item in the training corpus, we expect the following three consequences: a) *stato* will be memorized on the map; b) it is likely to have one or more dedicated firing neurons, that is neurons with a faithful image of the input; c) the image of *stato* will appear at an early stage of the learning process. What about less frequent forms like *amato* or *cantato*?

In a SOM, the global order of the map is fixed at early learning stages, at around the same time frequent forms are memorized in full. In fact, it turns out that high-frequency items play a crucial role in i) shaping the (high-level) topography of the resulting SOM, and ii) conditioning the topological distribution of the remaining information. In fact, very frequent input items act as prototypical representatives for their whole class over the first learning stages, thus anchoring their class representation throughout the learning process. It is only at later stages, when fine-grained information about the remaining members of the class is separated from the information about the class prototype that the whole class is fully learned. The complex paradigmatic structure of Italian past participles is particularly suitable to illustrate this kind of behaviour.

4. Experiment Architecture

The computational architecture we developed for these experiments consists of two hierarchically connected SOMs (Fig. 3), whose mode of interaction is reminiscent of Time Delay Neural Networks (Waibel et al., 1989). Input word forms are strings of alphabetic characters. At each exposure, an input verb form triggers a series of

activations on the first level map. All activations triggered by the same input form are integrated into an "activation image" that is in turn processed by the second level SOM. This way, the first level image plays the role of a short term memory temporarily storing the character-based information relative to a single input form, for this information to be processed wholesale by the second level map.

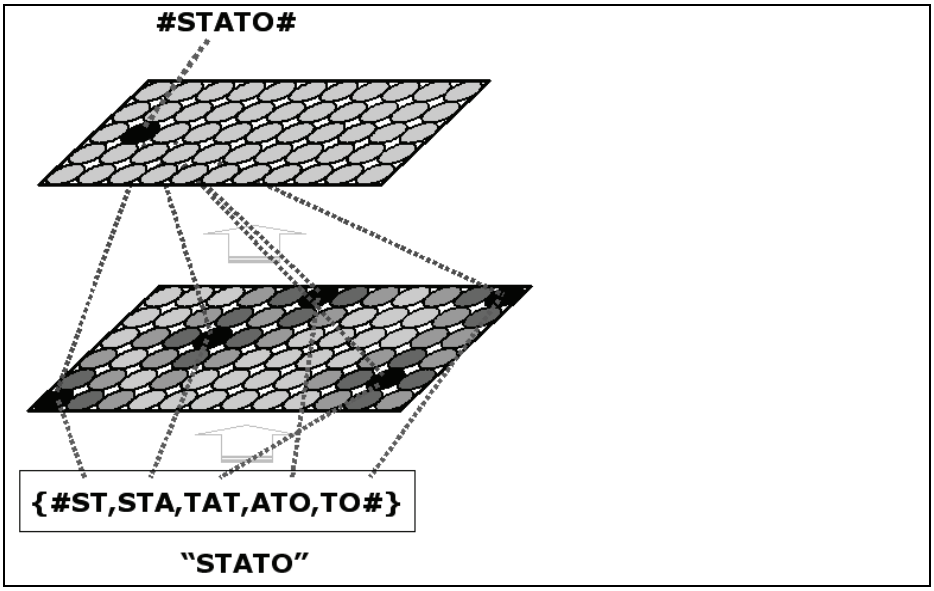


Figure 3: A two-level SOM architecture

Activations on the first level SOM are produced by consecutive time-bound scans of the input form. Each activation represents the map's response to a sub-context of the whole input, whose fixed size is measured as the number of alphabetic characters it contains. For the current set of experiments, the context size was set to 3. The second level SOM takes as input the output image of the current input form on the first level map. The upper level SOM clusters word forms according to their activations images on the lower level SOM. Another way to look at the first level SOM is as a perceptual interface between the raw character-based representation of an input form and receptors on the second level map. The fundamental benefit of this two-staged processing strategy is that the dimensionality of each "activation image" remains fixed, independently of input size.

5. Experimental Results

In this section, we report the experimental results of a few computer simulations of learning verb forms in Italian, English and Arabic. In the Italian and English experiments, we contrast the differential results obtained by training the map on two data sets for each language, one where verb forms are presented to the map according to their frequency distribution in a corpus, the other one where training data are distributed uniformly. The comparison sheds light on the role of frequency in the learning dynamics by exploring two considerably different verb systems (Italian and English) in terms of the relative prominence of sub-regular verb forms with respect to fully regular

ones. On the other hand, the Arabic results are based on one training configuration only, whereby fully vocalized verb forms are presented to the map according to their frequency distributions in the LDC Arabic corpus (Maamouri et al. 2004). For each experiment, we focused on paradigmatically homologous verb forms: past participle masculine singular forms for Italian, past tense forms for English, perfective masculine third singular forms for Arabic.

The experiments are intended to cover a fairly wide range of both typological and structural dimensions of cross-linguistic morphological variation. English and Italian verb forms are contrasted with Arabic data along the concatenative vs non-concatenative dimension of inflectional marking. From this standpoint, the crucial issue is whether it is possible for a single map to simulate the differential processes of acquisition of typologically as diverse morphological constructs as inflectional endings, continuous stems, word patterns and discontinuous stems on the basis of uniform requirements on input representation. This is not trivial, since, as we saw above, the representation requirements for English and Arabic data are potentially conflicting and generate hard alignment problems.

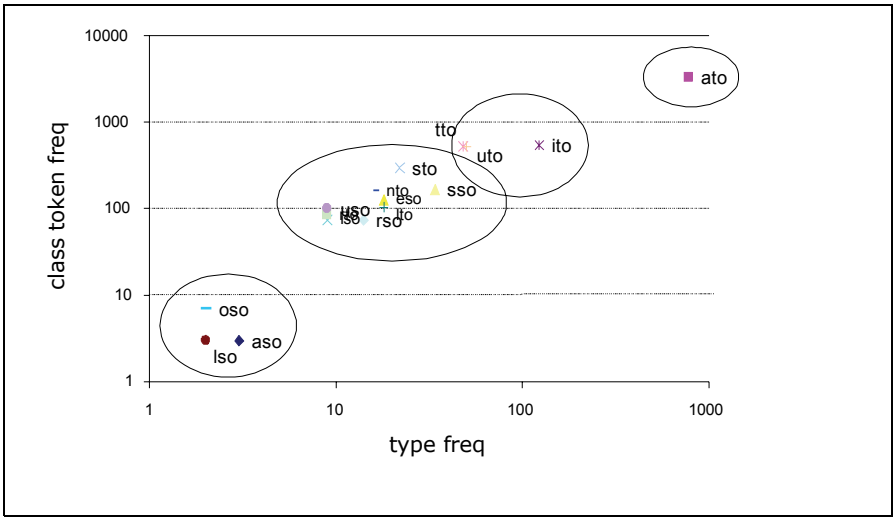


Figure 4: Type vs. class token frequency in the Italian past participle

Another important dimension of variation this set of experiments is intended to shed light on concerns the different distribution of regular and irregular inflections in languages such as English and Italian and the way such differences may impact morphology learning. Figures 4 and 5 illustrate this point in connection with the distribution of past participle and past tense forms in the two languages, by type and class token frequency plotted on a log scale. Forms are grouped according to loosely defined inflectional classes, each including inflected forms sharing the orthographic rendering of the word final syllable nucleus. Admittedly, the criterion is fairly crude and fails to cluster together forms such as *left* and *felt* whose past tense formation processes are very similar. Nonetheless, the resulting classifications retain some morphological plausibility and stake out the space of formal variation a learner is exposed to.

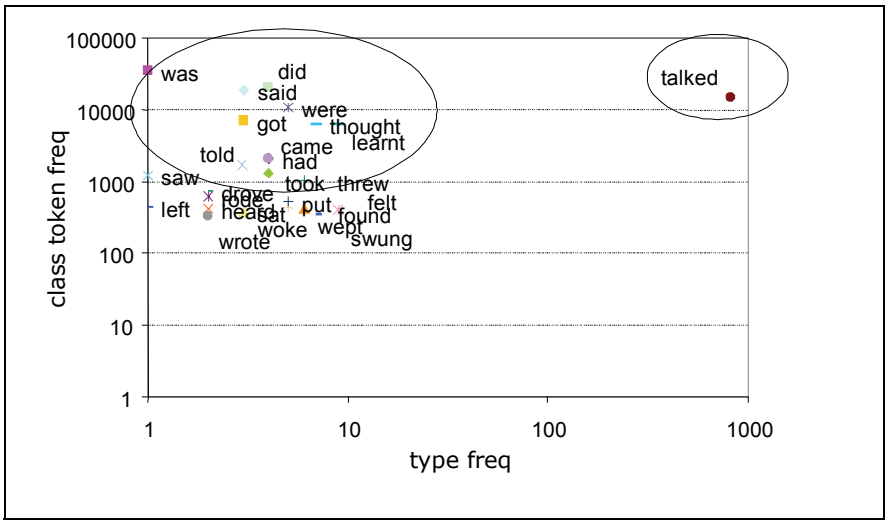


Figure 5: Type vs. class token frequency in the English past tense

The difference in distribution between Italian and English data is striking. Italian forms are scattered along a fairly uniform continuum, with subclasses of irregular forms exhibiting increasingly prominent gang effects in terms of their type cardinality as we move from left to right along the x-axis. On the other hand, English forms can sharply be divided into two groups: irregular forms on the left-hand side of our plot, forming a constellation of scantly represented morphological sub-classes (whose cardinality hardly exceeds the 10 units) and the class of regularly inflected forms on the other hand, covering the vast majority of English verb types. Besides, the Italian distribution shows a prominent log-linear correlation between type frequency and class token frequency, totally missing in the English past tense. This fact, as we shall see in a moment, has significant consequences on the learning dynamics of the two systems.

5.1. Italian past participles

Input forms consist of 470 different singular past participle forms (for a total amount of 5157 tokens) from the Italian Treebank (Montemagni et al. 2003). The highest frequent form is *stato* ('been', with 382 occurrences), followed by *fatto* ('done', 180), *detto* ('said', 131), *visto* ('seen', 77) and *avuto* ('had', 70). The least frequent forms appear only 3 times in the Treebank and cover 106 different forms, 75 of which ending in *-ato* (first conjugation), 15 in *-ito* (third conjugation) and only 2 in *-uto* (second conjugation). Of the remaining 14 form types of frequency 3, all undergoing a sub-regular past tense formation (Pirrelli, 2000), only two are verb base forms (namely *sciolto* and *stretto*) while the remaining 11 are derivatives such as *esteso*, *rimosso* and *trascorso*. As the training corpus is a collection of newspapers articles, speech report verbs such as *dichiarato* ('declared'), *aggiunto* ('added') or *annunciato* ('announced') are among the most frequent form types. Surely, these figures prevent us from taking this experiment representative of the typical input evidence an Italian child is exposed to in the course of her/his morphology maturation. Nonetheless, our word distributions, however not as realistic as we would like them to be, do reflect, to a certain degree of approximation, a general bias towards consistently sub-regular high-frequency Italian

verb forms such as *detto* ('said'), *fatto* ('done'), *visto* ('seen'), *chiesto* ('asked') etc., that happens to hold independently of variation of topic, gender and pragmatic grounding.

We simulated two different learning sessions: one where verb forms are presented to the map according to their frequency distribution in the Italian Treebank, the other where training data are assumed to be distributed uniformly. Figure 6 gives two snapshots of the state of a second level map in the two learning sessions at the same (early) stage. Grey triangles highlight map nodes that are fired when past participles of the *-sto* class (e.g. *visto* 'seen' and *chiesto* 'asked') are presented to the map. Black triangles highlight nodes that are sensitive to the *-tto* family (*fatto* 'done', *detto* 'said' etc.).

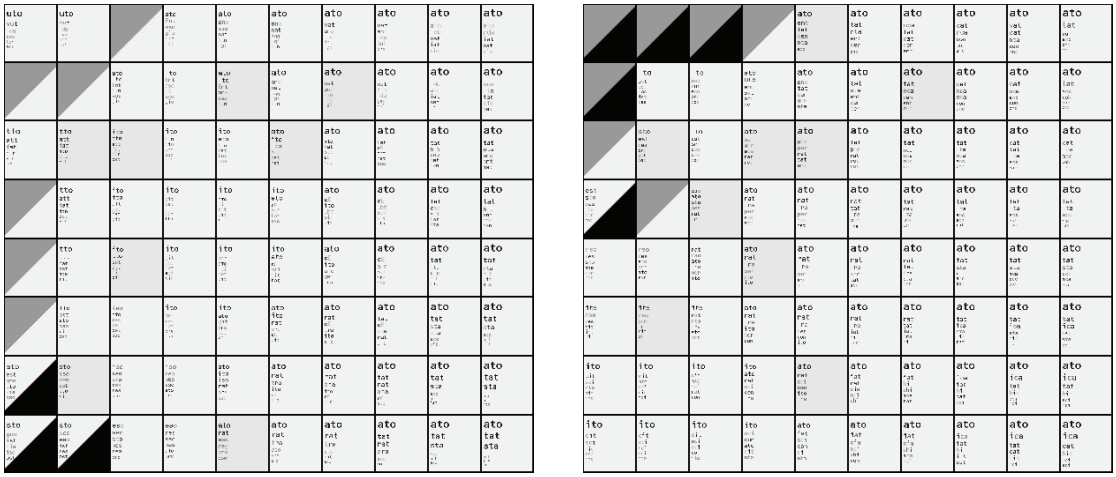


Figure 6: The Italian past participle

The important difference between the two snapshots is that the map trained on token frequencies neatly separates the two verb classes, while the other map tends to confound them: in other words, the former map develops entrenched, differential specialisation for *–tto* and *–sto* ending forms quite early on the basis of their token frequencies, while the latter SOM more reluctantly converges towards specialisation, for lack of evidence on token distribution. Later in the paper, we explain this developmental difference by arguing that very frequent forms like *visto*, *fatto* and *detto* tend to act as *prototypes* of their own class.

5.2. English past tenses

The experiment input consists of the 548 most frequent past tense forms in the British National Corpus (Leech 1992). The top-most ranked such forms are *was* (34836 occurrences), *did* (20247), *said* (18051), *were* (10570) and *had* (9573), accounting for a total of 93278 occurrences, out of all 141501 past tense forms attested in the training corpus. Like with the Italian experiment, we simulated two learning sessions, with and without token frequencies. Interim results of the two sessions, at comparable learning stages, are depicted in Figure 7, showing two dramatically different topological structures. In the right-hand map of Figure 7, grey squares mark map neurons activated

by *–ed* ending forms, while black squares are fired by *was*, *had* and *did*. In a 6x8 map grid, regular forms spread over 41 nodes, leaving only seven nodes to all remaining irregular forms. The result sets the stage for massive regularization of sub-regular forms, which are swamped by their *–ed* competitors, and seemingly lends support to the view that the two sets of regular and sub-regular past tense forms cannot possibly be learned through the same mechanism. On the other hand, the left-hand map of Figure 7 shows the results of learning by token frequencies: regular *–ed* ending forms now take up only six nodes on the map, while *was*, *had* and *did* have each a dedicated neuron. More room is left for memorizing other irregular forms, such us *said*, *paid*, etc. The evidence is in line with the intuition that irregularly inflected forms have the chance to survive the regularizing pressure of *–ed* forms, if the former are frequent enough to carve out their own dedicated area on the map by repeatedly firing a highly specialised, although comparatively circumscribed area of map nodes.

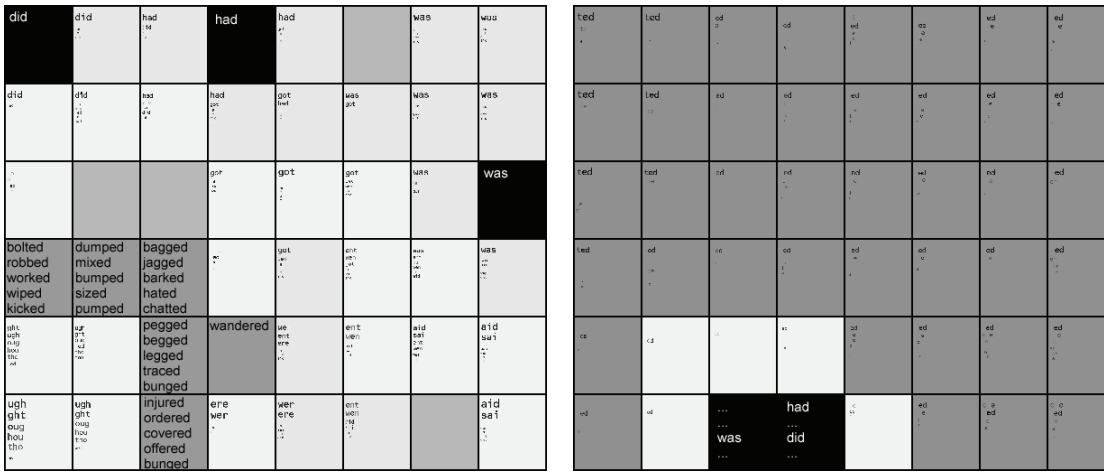


Figure 7: The English past tense

5.3 Arabic 3ps-SG perfectives

Comparable results are obtained by feeding a two-level SOM on Arabic verb forms, namely third masculine singular perfectives. This time, morphological markers do not form continuous strings of characters (as with Indo-European endings), but rather vowel patterns that are interdigitated with discontinuous roots. Since forms are presented to the map according to their corpus-based frequencies, the pattern *a_a_a*, by far the most frequent and regular one in Arabic perfective verb forms, takes over a substantial portion of the second level map, as shown in Figure 8. We have no room here to comment on the topological structure of Figure 8 in detail. Suffice it to point out at this stage that other less regular patterns of perfective formation emerge from the map, including low frequency *a_i_a* patterns. Most remarkably, high frequent forms such as *kAna* ('(s)he/it was') and *qAla* are recognised as wholes by specialized receptors (located in the top left corner of the map). Finally, it should be appreciated that the Arabic forms used for training the two-level map are given the same input representations as English and Italian forms. Nonetheless, the resulting topology consistently reflects the specific non-concatenative nature of Arabic morphology. We

take this to show that our architecture exhibits a highly adaptive and convergent topological behaviour, based on a comparatively poor battery of built-in inductive biases.

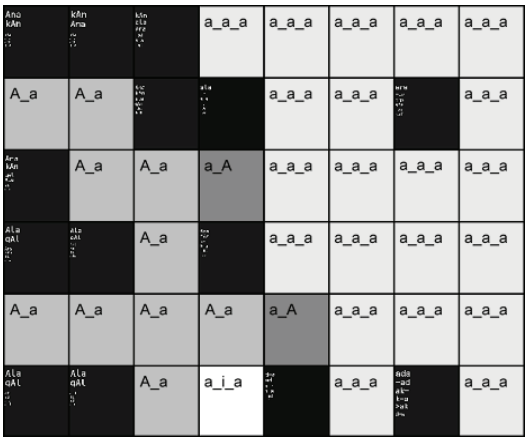


Figure 8: The Arabic perfective

6. General discussion

It is a well known fact that highly frequent forms tend to be shorter cross-linguistically, more readily accessible in the mental lexicon, independently stored as whole items (rather than being part of bigger families of paradigmatically-related forms) and thus more easily learnable and usable (Caramazza et al. 1988, Stemberger and MacWhinney 1988, Bybee 1995, Mowrey and Pagliuca 1995, Slobin 1997, Hare et al. 2001). These features make them also fairly resistant to morphological overgeneralization through time, thus establishing an interesting correlation between irregular inflected forms and frequency (Bybee 1985, 1995, Corbett et al. 2001). In the cognitive literature, it has also been shown that the existence of a type of instance that occurs with high token frequency may provide a highly relevant "cognitive anchor", serving to organise memory and reasoning about other related types (Strack and Mussweiler 1997, Goldberg et al. 2004). If we try to reconcile the latter finding with classical accounts of lexical entrenchment, we arrive at the seemingly paradoxical conclusion that irregular forms should act as "models" of the morphological organisation of the speaker's mental lexicon.

Observation of the learning behaviour of a SOM in our previous experiments can help us to understand why this paradox is only apparent. Entrenchment of a SOM memory trace is a direct function of input frequency and reflects the receptor sensitivity to input features. Similar memory traces tend to cluster in locally connected areas of the map. During training, specific, connected areas of receptors become increasingly more sensitive to specific classes of input stimuli, mimicking what we know about the functional specialization of the brain cortex. By training a SOM on a corpus-based distribution of inflected forms, then, very frequent short forms such as *is* or *did* are, at early stages of learning, the only input items to be fully memorised by receptors. These early "specialised" receptors are very likely to be fired by other, less frequent input forms (e.g. *said*, *read* or *led*) that happen to be similar to already entrenched memory

traces. As lexical stems show a greater degree of formal variability than inflectional markers, memory traces of highly frequent forms tend to be fired by similarly inflected, less frequent forms. The area of the map surrounding the *did* receptor, for example, becomes more and more sensitive to *d*-ending verbs.

To sum up, we can describe the dynamic behaviour of a SOM learning the morphology of a language along the following lines:

- highly frequent forms leave deeply entrenched and highly salient memory traces that act as standards of comparison (anchors) for other similarly inflected forms
- highly frequent forms are eventually memorised in full
- less frequent forms tend to fire connected topological areas of the map that are sensitive to shared morphological markers
- the surface of each connected area is proportional to the number of form types sharing a specific marker: regular markers are thus distributed over larger areas
- formally similar markers are memorised in contiguous areas of the map, thus developing hierarchical clusters of formally graded inflections (e.g.: *-id*, *-ed*, *-t*, *-nt* for the English past tense)
- principles of SOM specialization approximate a maximally compact arrangement of memory traces.

It is important to emphasise at this stage that these promising results are obtained through unsupervised training sessions, whereby a SOM is given no indication about the possible morpho-syntactic content associated with each form. In a sense, as adumbrated in the title of this contribution, morphological classes are learned through recourse to purely morphological information only. This is interesting, as it allows us to speculate that formal principles of the morphological organization of a language can be learned by a child through sheer exposure to plain forms and their frequencies, rather than to full-fledged sign-based word representations, coupled with form and meaning. It is tempting to suggest that a child can use acquired formal principles of paradigmatic organization to regiment the proper interpretation of the morpho-syntactic content associated with inflectional endings. This suggestion is in line with the empirical evidence that children master the morphological inflection of their own language before they can use it in the appropriate morpho-syntactic contexts (Clahsen 1989, Wilson 2003).

We would like to conclude the present contribution with a few remarks. First, the developmental interplay between type and token frequency of input items throws in sharp relief the profound interconnection between entrenchment of highly frequent items and overall effects of global organization in the topology of the mental lexicon. This is not trivial and serves to reconcile two apparently contradictory but established facts in the child learning literature: i) the first stages of language learning are best described as a process of item-based rote memorization, leading to gradual development of more and more abstract morphological schemata (Tomasello 2000, in press); ii) the most frequent evidence available to a child learning the morphology of a language is in

fact the most untypical and resistant to rule-based generalizations. We can explain away this paradoxical state of affairs by observing that the most frequent input items do in fact exemplify a wide range of processes of inflectional marking, thus contributing to shaping the overall organization of the child's morphological lexicon. In the early stages of learning, they act as powerful attractors of their own class mates and do so in a very focused and efficient way, since they are usually very short and reduced items, in which morphological marking has the upper hand, as it were, over lexical marking. Once the overall topology is established, the role of prototypes progressively shrinks, to eventually give way to a finer-grained organization of inflectional classes. It is at this stage, that regular patterns emerge. In our view, SOMs illustrate this dynamics in a very clear and intuitive way. Moreover they help us to gain insights into the haystack search problem we broached at the beginning of the paper. Being reduced forms, prototype attractors make it considerably easier to focus on the morphologically relevant bits of word forms.

Another related observation is that the existence of highly frequent prototypes also solves the paradoxical interplay between local processing and global, long-term memory structures. Word processing remains local throughout, but it gets progressively influenced by competition among different prototype attractors, each developing a local area of item-based influence. We suggest that the interplay of these two factors goes a long way towards explaining how global organization effects may eventually result from local processing steps.

The computational framework for morphology learning presented here leaves many issues open. At this juncture, we would like to only mention a couple of them. We have been using SOMs as topological metaphors of the mental lexicon, or, in more neuro-psychological terms, of long term memory structures. This is attractive but leaves us with the following problem: if morphological clusters develop through underspecified memory traces, how can a learner retrieve a fully inflected form? We emphasised that only highly frequent forms are memorised as wholes and do not participate in inflectional clusters: where are the remaining parts of a partially memorised form to be found in the lexicon? We have no room here to address these questions at the level of detail they deserve. We can only suggest that the maps shown in the present paper represent a (first) level of morphological (as opposed to lexical) organization of the space of inflected word forms. We know that word forms occupy a multidimensional linguistic space, and can thus be classified according to multiple perspectives. In this paper, we were mainly concerned with issues of morphological processing and classification, because of the peculiar and paradoxical problems they seem to raise. No doubt, a full psycho-computational account of the mental lexicon should make provision for several classificatory layers, which, in the present framework, are likely to be associated with separate, independently self-organized, but associatively-related maps.

Another interesting, related issue has to do with the classical dynamics of child morphology learning known, since the seminal work of Rumelhart and McClelland (1986), as the U-shaped curve (Plunkett and Marchman 1991), and its relationship to our computational model. In fact, it would not be too difficult to equate the first phase of SOM learning, where only very frequent items are memorised in full, with the stage of rote memorization characterizing the top left onset of a U-shaped learning curve. The intermediate, over-regularization stage may in turn correspond to a phase where bigger clusters set in, thus pushing itemized learning into the background. Eventually, the final,

mature stage may correspond to a phase of learning fine-grained morphological classes. For this picture to be put to the challenging test of a computer simulation, however, several further steps remain to be taken in the direction tentatively suggested here.

References

Albright, A., Islands of reliability for regular morphology: Evidence from Italian. *Language 78*. 684-709.

Anderson, J.R., 1993, *Rules of the Mind*. Hillsdale, NJ: Lawrence Erlbaum.

Anderson S., 1992, *A-Morphous Morphology*. Cambridge, UK: Cambridge University Press.

Aronoff M., 1994, *Morphology by itself*. Cambridge (MA): The MIT Press.

Boyland, J.T., 1996, *Morphosyntactic Change in Progress; a Psycholinguistic Approach.* Dissertation: Linguistic Department, University of California.

Bybee, J., 2002, Sequentiality as the basis of constituent structure. in T. Givón and B. Malle (eds.) *The Evolution of Language out of Pre-Language*, Amsterdam: John Benjamins. 107-132.

Bybee J., 1985, *Morphology: A study of the relation between meaning and form*, Benjamins.

Bybee J., 1995, Regular Morphology and the Lexicon, *Language and Cognitive Processes 10* (5).

Caramazza A., Laudanna A., Burani C., 1988, Lexical access and inflectional morphology. *Cognition 28*.

Clahsen, H., 1986, Verb inflections in German child language: acquisition of agreement markings and the functions they encode, *Linguistics 24*. 79-121.

Clahsen H., 1996, Introduction. In H. Clahsen (ed.) *Generative perspectives on language acquisition: Empirical findings, theoretical considerations, and cross-linguistic comparisons*. Benjamins.

Corbett, G., Hippisley, A., Brown, D., 2001, Frequency, regularity and the pardigm: A perspective from Russian on a complex relation, in J. Bybee, P. Hopper (eds.) Frequency and the emergence of linguistic structure, Benjamins.

Croft, W., 2001, Radical Construction Grammar, Oxford University Press.

Creutz, M. and Lagus, K. 2004. Induction of a Simple Morphology for Highly-Inflecting Languages. In Proceedings of the 7th Meeting of the ACL SIGPHON, Barcelona.

Dick, F., Bates, E., Wulfeck, B., Utman, J., Dronkers, N. & Gernbascher, M.A., 2001, Language deficits, localization and grammar: Evidence for a distributive model of language breakdown in aphasics and normals. *Psychological Review, 108(4), 759-788.*

Ellis, N.C., Schmidt, R., 1998, Rules or Associations in the Acquisition of Morphology? The Frequency by Regularity Interaction in Human and PDP Learning of Morphosyntax. *Language and Cognitive Processes 13 2/3*. 307-336.

Frost, R., Forster K.I., Deutsch A., 1997, What can we learn from the morphology of Hebrew? A masked-priming experiment of morphological representation, Journal of Experimental Psychology: Learning, Memory and Cognition 23(4).

Frost, R., Grainger J., Rastle K., 2005, Current issues in morphological processing, Language and cognitive processes 20(1/2).

Gaussier, E., 1999, Unsupervised Learning of Derivational Morphology from Inflectional Lexicons, in Proceedings of Unsupervised Learning in Natural Language Processing, ACL 1999 Workshop, Maryland USA.

Goldsmith, J., 2001, Unsupervised learning of the morphology of a natural language, Computational Linguistics, 27(2).

Goldsmith, J., 2006, An algorithm for the unsupervised learning of morphology. *Natural Language Engineering 12*. 1-19.

Goldsmith, J., O'Brien, J., 2006, Learning inflectional classes. MS.

Goldberg, A.E., Casenhiser, D., Sethuraman, N., 2004, Learning Argument Structure Generalizations, *Cognitive Linguistics*.

Hare M. L., Ford M., Marslen-Wilson W.D., 2001, Ambiguity and frequency effects in regular verb inflection, in J. Bybee, P. Hopper (eds.) Frequency and the emergence of linguistic structure, Benjamins.

Harris, Z., 1951, *Methods in Structural Linguistics*. Chicago: University of Chicago Press.

Haspelmath, M., 2002, *Understanding Morphology*. London: Understanding Language Series, Arnold.

Hay, J., R.H. Baayen, 2005, Shifting paradigms: gradient structure in morphology, *Trends in Cognitive Sciences*, 9(7): 342-348.

Hyams, N., 1986, *Language Acquisition and the Theory of Parameters,* Dordrecht, Reidel.

Hyams, N., 1996, The underspecification of functional categories in early grammars, in H. Clahsen (ed.), *Generative Perspectives on Language Acquisition*, Amsterdam, Benjamin.

Jenkins, W., Merzenich M., Ochs M., 1984, Behaviorally Controlled Diferential Use of Restricted Hand Surfaces Induces Changes in the Cortical Representation of the Hand in Area 3b of Adult Owl Monkeys. *Society for Neuroscience Abstracts 10*. 665.

Juola, P., Plunkett, K., 1999, A Connectionist Model of English Past Tense and Plural Morphology. *Cognitive Science 43(1)*.

Kaas J., Merzenich M., Killackey H., 1983, The Reorganization of Somatosensory Cortex Following Peripheral Nerve Damage in Adult and Developing Mammals. *Annual Review of Neuroscience 6*. 325-56.

Klein, W., C. Perdue, 1997, The Basic Variety (or: Couldn't natural languages be much simpler?), *Second Language Research 13 (4):301-347*.

Kohonen, T., 2001, *Self-Organizing Maps.* Springer.

Leech, G., 1992, 100 million words of English: the British National Corpus. *Language Research 28 1*. 1-13.

Lightfoot, D., 1999, *TheDevelopment of Language. Acquisition, Change, and Evolution*, Blackwell publishers.

Matthews, P.H., 1992, *Morphology*. (second edition) Cambridge UK: Cambridge University Press.

Matthews, P.H., 1993, *Grammatical Theory in the United States*. Cambridge UK: Cambridge University Press.

Marslen-Wilson, W.D., Tyler, L.K., Waksler, R., Older, L., 1994, Morphology and meaning in the English mental lexicon, *Psychological Review 101*. 3-33.

Maamouri, M., Bies, A., Buckwalter, T., Mekki W., 2004, The Penn Arabic Treebank: Building a Large-Scale Annotated Arabic Corpus, NEMLAR International Conference on Arabic Language Resources and Tools: Cairo.

Marcus, G.F., Brinkmann, U., Clahsen, H., Wiese, R. and Pinker, S., 1995, German Inflection: The exception that proves the rule. *Cognitive Psychology 29*. 198-256.

Montemagni, S. et al., 2003, Building the Italian Syn-tactic-Semantic Treebank, in A. Abeillé (ed.), *Building and using Parsed Corpora*. Language and Speech series, Dordrecht: Kluwer. 189-210.

Mowrey R., Pagliuca W., 1995, The reductive character of articulatory evolution, *Rivista di Linguistica 7* (1).

Pinker, S., Prince, A., 1988, On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 29*. 195-247.

Pirrelli, V., 2003, Is machine language learning yet another machine learning task? In Proceedings of *AI*IA*: Pisa.

Pirrelli, V., 2000, *Paradigmi in Morfologia: un approccio interdisciplinare alla flessione verbale dell'italiano*. Pisa-Roma: Istituti Editoriali e Poligrafici Internazionali.

Pirrelli V., Calderone B., Herreros I., Virgilio M., 2004, Non-locality all the way through: Emergent Global Constraints in the Italian Morphological Lexicon , In Proceedings of 7th Meeting of the ACL SIGPHON, Barcelona.

Plaut, D.C., Gonnerman, L.M., 2000, Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing?, Language and Cognitive Processes 15(4/5).

Plunkett, K., Marchman, V., 1991, U-shaped learning and frequency effects in a multi-layered perceptron. Implications for child language acquisition. *Cognition 38*: 3-102.

Plunkett, K., Juola, P., 1999, A Connectionist Model of English Past Tense and Plural Morphology. *Cognitive Science* 43(1).

Prasada, S., Pinker, S., 1993, Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes 8*. 1-56.

Rissanen, J., 1989, *Stochastic Complexity in Statistical Inquiry*. World Scientific Series in Computer Science. Singapore: World Scientific.

Rumelhart, D., McClelland, J., 1986, On learning the past tense of English verbs. In D.E. Rumelhart and McClelland J. (eds.) *Parallel distributed processing: Explanations in the microstructure of cognition*. Cambridge (MA): MIT Press.

Say, T., Clahsen, H., 2001, Words, Rules and Stems in the Italian Mental Lexicon. In S. Nooteboom, F. Weerman and F. Wijnen (eds.) *Storage and computation in the language faculty*. Dordrecht: Kluwer Academic Publishers. 75-108.

Schone, P., Jurafsky, D., 2000, Knowledge-free induction of morphology using Latent Semantic analysis, in Proceedings of CoNLL 2000 and LLL-2000, Lisbon.

Slobin, D. I., 1997, The Origins of Grammaticalicizable Notions: Beyond the Individual Mind, in Dan I. Slobin (ed.) The Crosslinguistic study of Language Acquisition, Lawrence Erlbaum Associates.

Stemberger J.P. MacWhinney B., 1988, "Are inflected forms stored in the lexicon?" In M. Hammond and M. Noonan (eds.) *Theoretical Morphology. Approaches in modern linguistics*. Academic Press.

Stump G., 2001, *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge (UK): Cambridge University Press.

Strack F., Mussweiler T., 1997, "Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility" *Journal of Personality and Social Psychology* 73.

Tomasello, M., 2000, Do young children have adult syntactic competence? *Cognition 74*. 209-253.

Tomasello, M., in press, Acquiring linguistic constructions. To appear in R. Siegler, D. Kuhn (eds.), *Handbook of Child Psychology*.

Waibel A., Hanazawa T., Hinton G., Shikano K., Lang K., 1989, "Phoneme Recognition Using Time-Delay Neural Networks" *IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 37, No. 3*

Wicentowski R., 2004, Multilingual Noise-Robust Supervised Morphological Analysis using the WordFrame Model, In Proceedings of 7th Meeting of the ACL SIGPHON, Barcelona.

Wilson, S., 2003, Lexically specific constructions in the acquisition of inflection in English. *Journal of Child Language 30*. 75-115.

French Pronominal Clitics and the Design of Paradigm Function Morphology

Olivier Bonami
Université Paris 4 & LLF (UMR 7110)
olivier.bonami@paris4.sorbonne.fr

Gilles Boyé
ERSS (UMR 5610) & Université Nancy 2
gilles.boye@univ-nancy2.fr

Realizational approaches to inflectional morphology (e.g. Matthews, 1972; Anderson, 1992; Zwicky, 1992; Aronoff, 1994; Stump, 2001) generally presuppose that the realization relation is a function: for each cell in a lexeme's paradigm, the morphology provides a single realization. Although there are well known exceptions to this generalizations (e.g. individual lexemes with two distinct paradigms) these are usually assumed to be very local and are taken to be exceptional irregularities.[1]

In this paper we exhibit an extended example of an inflectional system that is highly non-functional, the system of French Pronominal Clitics (henceforth FPCs). Although many authors have argued that FPCs should be treated within inflectional morphology, the fact that they violate functionality has not been discussed previously. After reviewing the relevant data we provide an analysis of the FPC system within Paradigm Function Morphology (Stump, 2001), and propose a modification of the framework to allow for a single feature combination to give rise to more than one realization.

1. Features of French Pronominal Clitics

Starting with Miller (1992), a number of authors have argued that French pronominal clitics are best analyzed as inflectional affixes, rather than syntactic atoms.[2] Arguments in favour of this position fall in three broad classes. First, FPCs give rise to positional effects which are reminiscent of templatic inflectional morphology, but are quite alien in ordinary syntax (§1.1). Second, in informal varieties, clitic-host combinations give rise to idiosyncratic phonological realizations that are not attributable to regular phonology, and are typical of morphological combinations (§1.2). Third, there are

[1] This paper was presented as a poster at the Fifth Mediterranean Morphology Meeting, and orally at a meeting of the GDR 2220 *Description et modélisation en morphologie*. We thank the audiences of both events for their questions, comments and criticisms, and in particular Anne Abeillé, Denis Creissels, Georgette Dal, Bernard Fradin, Ana Luís, Philip Miller and Andrew Spencer.

[2] An early precursor is (Stump, 1981), which proposes an inflectional analysis of clitics but without much empirical justification. Note that adopting an inflectional analysis for FPCs does not commit one to saying that all traditional clitics should receive this type of treatment, as Miller emphasizes. Even for Romance weak pronominal forms, it is not clear that a uniform treatment is called for (see e.g. Crysmann, 2002, on Portuguese).

morphotactic restrictions on clitic combinations that escape a syntactic explanation (§1.3).[3]

1.1. Position Class Effects

The order of realization of FPCs does not correlate with the morphosyntactic properties they express; for instance dative clitics may precede (1) or follow (2) accusatives depending on person and reflexive status.[4]

(1) a. Paul me la présentera.
Paul DAT.1S ACC.3FS present-FUT.3S
'Paul will present her to me.'

b. *Paul la me présentera.
Paul ACC.3FS DAT.1S present-FUT.3S

(2) a. Paul la lui présentera.
Paul ACC.3FS DAT.3S present-FUT.3S
'Paul will present her to him.'

b. *Paul lui la présentera.
Paul DAT.3S ACC.3FS present-FUT.3S

In addition, some FPCs are in complementary distribution, despite the fact that they express compatible syntactic information. For instance, *se* and *te* cannot co-occur (4a) despite the fact that they can express different arguments of a single verb (3). In such a situation, the only solution is to express the dative argument as a full pronoun, an option which is normally blocked by the possibility of using an FPC.

---

[3] Note that all the data discussed in this section is well-known from traditional descriptions, although no precise analysis accounts for all of it, as far as we know. We do not discuss here possible counter-arguments to a morphological treatment of FPCs, but see e.g. Delais-Roussarie (2001).

[4] The data is trickier for enclitics in imperatives, where alternate orders are possible, at least in some varieties. For instance the realizations in (i) and (ii) are both possible in at least some varieties. The sociolinguistic and geographical determinants of the variation between (i) and (ii) are ill understood, but both possibilities are available in the speech of the authors. By contrast, (iv) is completely excluded.

(i) Donne le moi !
give.imp acc.3ms dat.1s
'Give it to me!'

(ii) %Donne moi le !
give.imp dat.1s acc.3ms

(iii) Donne le lui !
give.imp acc.3ms dat.3s
'Give it to him!'

(iv) *Donne lui le !
give.imp dat.3s acc.3ms

Note that robust data on enclitic complement FPCs is generally lacking, since (i) these only occur in imperatives, and are thus quite rare in existing (oral and written) corpora, and (ii) intuitions are not very reliable because of strong sociolinguistic effects associated with some attested combinations such as (ii).

(3) a. Paul te présentera Jean.
Paul DAT.2S present-FUT.3S Jean
'Paul will present Jean to you.'

b. Paul se présentera à Marie.
Paul REFL.3 present-FUT.3S to Marie
'Paul will present himself to Marie.'

(4) a. *Paul se te présentera.
Paul REFL.3 DAT.2S present-FUT.3S

b. *Paul te se présentera.
Paul DAT.2S REFL.3 present-FUT.3S

c. Paul se présentera à toi.
Paul REFL.3 present-FUT.3S to PRO.2S
'Paul will present himself to you.'

These properties are easily captured by treating the FPC system as an instance of position class morphology, where (more or less arbitrary) collections of items compete for realization in a single position. Concretely, it is usually assumed that the proclitic system can be captured using a series of seven position classes, as indicated in table 1.

In this table, the morphosyntactic information expressed by the clitic is shown between brackets. Note that the negative marker *ne* is the only non-pronominal element of the template. [5]

1	2	3	4	5	6	7
[*1s,nom*]	*negation*:	[*1s,acc/dat*]:	[*3ms,acc,nonrefl*]:	[*3s,dat,nonrefl*]:	[*loc*]:	[*de*]:
je	**ne**	**me**	**le**	**lui**	**y**	**en**
[*2s,nom*]		[*2s,acc/dat*]:	[*predicative*]:	[*3p,dat,nonrefl*]:		
tu		**te**	**le**	**leur**		
[*3ms,nom*]		[*3,acc/dat,refl*]:	[*3fs,acc,nonrefl*]:			
il		**se**	**la**			
[*3fs,nom*]		[*1p,acc/dat*]:	[*3p,acc,nonrefl*]:			
elle		**nous**	**les**			
...		[*2p,acc/dat*]:				
		vous				

Table 1: The traditional description of the proclitic system

1.2. Morphophonological Idiosyncrasies

There are a number of phonological reduction phenomena involving FPCs that cannot be accounted by regular phonology, since they are sensitive to properties inaccessible to

[5] For present purposes we ignore 'ethical' datives, but see note 21 below.

phonology. On the other hand these properties resemble the morphophonological idiosyncrasies typical of affix-host combinations.

First, there are cases of clitic-host fusion. For instance, the nominative 1s clitic *je* has a special realization when followed by the form *suis* of the lexeme ÊTRE 'be' (5a). Notice that such a realization is excluded with the otherwise homophonous form of the lexeme *suivre* 'follow' (5b), barring any hope of a phonological conditioning of the reduction.

(5) a. ʒə sɥi yn fij
je suis une fille
'I am a girl.'/'I follow a girl.'

b. ʃɥiynfij
'I am a girl.' / *'I follow a girl.'

Second, we also find cases of clitic-clitic fusion. The most well established of these concerns the clitic sequence *je lui*, which gives rise to various reductions that are not available in phonologically similar contexts. So for instance the reduced realizations found in (6a) are not possible with the present 1s form of the lexeme *luire* 'glow', despite its being homophonous with the clitic *lui* (6b).

(6) a. Je lui dirai.
NOM.1S DAT.3S tell-FUT.1S
'I will tell him.'
[ʒlɥidiʁɛ]/[ʒɥidiʁɛ]/[ʒidiʁɛ]

b. Je luis dans le noir.
NOM.1S glow-PRST.1S in DEF black
'I glow in the dark.'
[ʒlɥidɑ̃lnwaʁ]/*[ʒɥidɑ̃lnwaʁ]/*[ʒidɑ̃lnwaʁ]

Third, FPCs are among the items giving rise to 'elision' in French, the non-realization of a word-final vowel before a vowel-initial word. Note that despite a long tradition of treating elision as a general, phonological phenomenon, it concerns exactly four items in French: the definite article *la*, the complementizer *si*, and two FPCs, the nominative 2s *tu* and the accusative 3fs *la*. Moreover there are really three different behaviours at hand : *si* elides optionally, and only before the subject clitic *il(s)*, the socially preferred realization being elision. *Tu* also elides optionally, but before any vowel, and the socially preferred realization is non-elision, elision being somewhat informal. Both the article and the FPC *la* give rise to obligatory elision in all varieties.

(7) a. % si il vient
if he comes

b. s'il vient

c. si Isabelle vient
if Isabelle comes

d. * s'Isabelle vient

(8) a. Tu iras à Paris.
NOM.2S go-FUT.2S at Paris
'You will go to Paris.'

b. %T'iras à Paris.

(9) a. l' épouse
def. wife

b. *la épouse

(10) a. Je l' épouse.
NOM.1S ACC.3FS marry-PRST.3S
'I marry her.'

b. *Je la épouse.

This data points to the observation that elision is not at all a general, homogeneous phenomenon in French, but is best treated as a case of local, lexically specified, phonologically-conditioned allomorphy. Although we do not want to prejudge the analysis of the article *la* and the complementizer *si*, it is notable that this type of allomorphy is very common in morphological combinations, but is quite rare in syntactic combinations.

1.3. Morphotactics

Finally, there are a number of strange restrictions on the cooccurrence of FPCs. First, there are a number of arbitrary gaps : some clitic pairs can never be realized, such as pairs consisting of a clitic from the series *me*, *te*, *se*, *nous*, *vous* followed by a clitic in the series *lui*, *leur* (8a). Once again, this opens up the possibility of using a full pronoun for the dative (8b).

(8) a. *Paul se lui présentera.
Paul REFL.3 DAT.3S present-FUT.3S

b. Paul se présentera à elle.
Paul REFL.3 present-FUT.3S to PRO.3FS
'Paul will present himself to her.'

Second, there is at least one well documented case of clitic drop:[6] in informal varieties, FPCs from the series *le*, *la*, *les* can be dropped before FPCs in the series *lui*, *leur* (10a). That this is a robust observation is shown by the contrast in (9)-(10) : the lexeme *apporter* 'bring' is one of the few verbs which take an obligatory direct object, as illustrated in (10)b. Thus (10)a cannot be treated as a simple case of direct object drop : the possibility of not realizing the accusative complement is correlated with the presence of the clitic *lui*. This is accounted for if we assume that *lui* counts as a realization of both the accusative and the dative in this context, i.e (10)a is an alternate realization of the morphosyntactic information expressed by(9)a.[7]

(9) a. Paul la lui apportera.
Paul ACC.3FS DAT.3S bring-FUT.3S
'Paul will bring it to her.'

b. Paul l' apportera à Marie.
Paul ACC.3FS bring-FUT.3S to Marie
'Paul will bring it to Marie.'

(10) a. Paul lui apportera.
Paul DAT.3S bring-FUT.3S
'Paul will bring it to her.'

b. *Paul apportera à Marie.
Paul bring-FUT.3S to Marie

Once again, this type of data is familiar from inflectional (and in particular templatic) morphology, but quite alien for syntactic combinations.

[6] Other cases discussed by Miller (1992) include haplology data, such as the possibility of dropping the locative clitic *y* before the future forms of the lexeme *aller* (i). Note that a locative complement is otherwise obligatory with *aller* (ii).

(i) Paul ira.
Paul go-FUT.3S
'Paul will go there.'

(ii) *Paul est allé.
Paul is go.PASTP.MS
'Paul went.'

[7] It is sometimes suggested that (10)a is simply the effect of a phonological reduction of [la] to [l] followed by a degemination from [llɥi] to [lɥi]. However this would be the only case where [la] reduces to [l] before a consonant. Notice in particular that the reduction is not available before the verb *leurre* (i-ii), which is homophonous to the clitic *leur* before which reduction is possible (iii).

(i) Paul la leurre admirablement.
Paul ACC.3FS lure-PRST.3S wonderfully
'Paul wonderfully lures her.'

(ii) #Paul leurre admirablement. (not compatible with a specific interpretation of the object)

(iii) Paul leur apportera.
Paul dat.3p bring-fut.3s
'Paul will bring it to them'

2. Previous Approaches to FPCs

We follow Miller (1992) in concluding from the previous data that FPCs are better treated as an instance of inflectional morphology that as an instance of syntax.[8] In this section we review briefly two existing formal analyses in this tradition.

2.1. Miller and Sag (1997)

Miller and Sag propose an HPSG (Pollard and Sag, 1994) analysis of FPCs. Although most of the paper deals with issues in the syntax of FPCs, it includes a realizational, template-based analysis of the morphology of clitics.

The realization of clitics is governed by a function F_{PRAF}, which taxes as input (i) an inflected form, (ii) the HEAD value of the verb (specifying its morphosyntactic features), and (iii) the verb's ARGUMENT-STRUCTURE list (specifying the list of its arguments). The function outputs a specification of a clitic template, including information as to whether the template is enclitic or proclitic. The template consists of seven slots ; it includes a specification of specific enclitic forms (slot 5) which were not taken into account in table 1. Negative *ne* is not taken into account.

Table 2 summarises the effect of F_{PRAF} in the case of proclitics.[9] The first column corresponds to the HEAD argument of the function. Column 2 indicates the morphosyntactic properties associated with a member of the ARG-ST list which is to be realized as a clitic. Column 3 specifies in which slot it is realized and column 4 under what form.

---

[8] As is well known, noninverted subject clitics have some properties that are less affix-like than those of the rest of FPCs (Couquaux, 1986). In particular they may take scope over a coordination of VPs, at least in formal varieties (i)—this is impossible for object clitics in all varieties (ii). This property points to an analysis of subject clitics as syntactic atoms. However treating subject clitics as syntactic atoms makes it hard, if not impossible, to account for the clitic-host and clitic-clitic fusions observed in section 1. There are two possible routes here : either subject clitics should be treated as "true clitics" in the sense of Zwicky (1977), that is, syntactic atoms with the morphophonological properties of word-internal morphs—the problem with such a hypothesis is that we lack a formally precise proposal for accounting for such true clitics. Or we could follow Miller (1992) in assuming that there are really two competing systems for subject clitics in contemporary French, one where they are syntactic atoms, and one where they are affixes. One argument in favor of this second solution is the fact that wide scope over coordination and clitic-host fusion seem to be in complementary distribution (iii).

(i)	Il	lira	ce	livre	et	le	critiquera.
	he	read-FUT.3S	this	book	and	it	criticize-FUT3S
	'He will read this book and criticize it.'						
(ii)	*Il	le lira	aujourd'hui	et	critiquera		demain.
	he	it read-FUT.3S	today	and	criticize-FUT.3S		tomorrow.
	'He will read it today and criticize it tommorrow.'						
(iii)	Je	suis	et	resterai	content.		
	I	be-PRST.1S	and	stay-FUT.1S	happy		
	'I'm happy and will stay that way.'						

[ʒəsɥiɛʁɛstəʁɛkɔ̃tã] / *[ʃɥiɛʁɛstəʁɛkɔ̃tã]

[9] Miller and Sag's analysis also applies to the case of enclitics, but we do not discuss that part of the analysis since (i) it would take us too far afield and (ii) the variety described in Miller and Sag differs strongly from our own as far as enclitics are concerned, making comparisons difficult.

HEAD	**ARG-ST element**	**slot**	**form**
tensed	[*1s*, *nom*]	SL-1	je
tensed	[*2s*, *nom*]	SL-1	tu
tensed	[*3ms*, *nom*]	SL-1	il
tensed	[*3fs*, *nom*]	SL-1	elle
…	…	…	…
	[*1s*, *obj*]	SL-2	me
	[*2s*, *obj*]	SL-2	te
	[*3s*, *obj*, *refl*]	SL-2	se
	[*1p*, *obj*]	SL-2	nous
	[*2p*, *obj*]	SL-2	vous
	[*3ms*, *acc*]	SL-3	le
	[*3ms*, *predicative*]	SL-3	le
	[*3fs*, *acc*]	SL-3	la
	[*3p*, *acc*]	SL-3	les
	[*3s*, *à*$_1$]	SL-4	lui
	[*3p*, *à*$_1$]	SL-4	leur
	[*à*$_2$]	SL-6	y
	[de]	SL-7	en

Table 2—Miller and Sag's (1997) realization function in the enclitic case

Table 2 encodes a number of auxiliary hypotheses on French morphosyntax. The forms *à* and *de* are taken to be case markers rather than prepositions ; two distinct case values can be realized as *à*, corresponding to the clitics *lui*/*leur* and *y* respectively. The value *obj* is an underspecified case value corresponding to a choice between accusative and *à*$_1$.

In addition to the specification of F_{PRAF}, Miller and Sag propose that the template is subject to structural constraints such as the following, which account for some morphotactic idiosyncrasies :

(11) If slot 2 is nonempty, then slot 4 is empty.

This constraint makes sure that clitics in the series *me/te/se/nous/vous* cannot cooccur with clitics in the series *lui/leur*, as observed in §1.3.

Although the syntactic part of Miller and Sag's analysis is impressively detailed, there are problems with the morphological proposal, which is clearly underdeveloped. First, the proposal deals nicely with position class effects, but has nothing to say on morphophonological idiosyncrasies.[10] The treatment of morphotactics based on constraints on the templates is not formalized and not very promising of an elegant formalization. Second and more importantly, the system is set up so that each feature combination may give rise to only one phonological realization. This is highly problematic, since most of the alternative realizations discussed in section 1 are optional : fused forms, *tu* elision and clitic drop are only options which coexist with

[10] Remember from the discussion in section 1 that these idiosyncrasies cannot be attributed to phonology, and should be treated within the morphological component.

other possible realizations in the grammar. Thus the problem is not so much that Miller and Sag's analysis describes a standardized variety, but that their formal setup cannot be adapted to the description of a variety where alternative realizations are allowed.

2.2. Monachesi (1999, 2005)

Monachesi (1999) develops an analysis of Italian pronominal clitics based on data analogous to Miller's, and that may easily be mimicked in an analysis of French. Monachesi (2005) extends the approach to Romance in general with special reference to Rumanian.

Monachesi's proposal is based on the postulation of implicational constraints linking the value of a verb's CLITICS list, representing the arguments of the verb that need to be realized as clitics, and the morphological features STEM and AFFIX. Each constraints states globally a possible morphological realization for a full CLITICS list—the constraints deal directly with clitic sequences rather than individual clitics. Thus (12)a is the constraint realizing a dative 1s and a locative as *mi ci*. This way of dealing with clitic sequences allows Monachesi to accommodate morphophonological idiosyncrasies directly: for instance (12)b states an idiosyncratic realization for the sequence of a dative 1s and an accusative 3ms (the realization is idiosyncratic because *mi lo* would be expected instead of *me lo*). Finally optionality is not an issue, because disjunctions can be included in the constraints. For instance, (12)c states two alternative (elided and non-elided) realizations for a pre-vocalic accusative 3ms.

(12) a. $\left[\begin{array}{l} complex-word \\ \text{CLITICS } \langle \text{NP}[dat,1sg],\text{NP}[loc] \rangle \end{array}\right] \rightarrow \left[\text{AFFIX } \langle mici \rangle\right]$

b. $\left[\begin{array}{l} complex-word \\ \text{CLITICS } \langle \text{NP}[dat,1sg],\text{NP}[acc,3ms] \rangle \end{array}\right] \rightarrow \left[\text{AFFIX } \langle melo \rangle\right]$

c. $\left[\begin{array}{l} complex-word \\ \text{CLITICS } \langle \text{NP}[acc,3ms] \rangle \end{array}\right] \rightarrow \left[\begin{array}{ll} \text{STEM} & \langle vowel,\ldots \rangle \\ \text{AFFIX} & \langle l \rangle \vee \langle lo \rangle \end{array}\right]$

Although Monachesi's setup can deal with the very issues that were problematic for Miller and Sag, it seems to miss important generalizations, at least for French. As we just saw, Monachesi states a different constraint for each clitic sequence that has to be generated. This might be a reasonable move for Italian, where the number of possible clitic clusters is low (56 according to Monachesi's 1999 data), and thus the proportion of idiosyncratic clusters is quite high. But this will not do for French, where we have to deal with subject clitics, subject clitic inversion, and the fact that enclitic clusters differ from proclitic clusters. A simple (but fastidious) calculation shows that even in conservative varieties such as that described by Miller and Sag, there are 1909 different cases to consider. It is clearly not satisfying to need nearly 2000 rules to deal just with the standard variety.

2.3. Discussion

The two analyses we just discussed have complementary advantages and complementary drawbacks: Miller and Sag's analysis is elegantly simple and general, at the price of coverage—ironically, this analysis gives no account for the very data Miller (1992) collected to argue in favour of a morphological view of PFCs. Monachesi's analysis can deal with the full set of data, but at the price of denying any structure to the clitic sequence, which gives rise to an explosion of the number of rules. Clearly, what is needed is a way of treating the idiosyncrasies without denying that in simple cases the clitic sequence is analyzable.

Paradigm Function Morphology (PFM; Stump 2001) seems like a good candidate for this task: PFM has been applied in detail to cases of position class morphology, and is known to be able to deal with cases where the position class system seems to collapse, because of affix fusions, position reversals, etc.

3. A PFM Analysis of French Pronominal Clitics

3.1. A Sketch of Paradigm Function Morphology

In this section we provide a sketch of Paradigm Function Morphology. The presentation is based on a PFM analysis of French conjugation that is detailed in appendix A.

Paradigm Function Morphology is an explicit theory of inflectional morphology that is both inferential and realizational: First, affixes are not treated as signs, but as the result of the application of a rule relating morphosyntactic features to a phonological function modifying a base. Second, inflectional rules do not consume features, but merely express (realize) them. Thus there is no requirement that a feature must be expressed exactly once.

In PFM, the inflectional system of a language is modelled by a *paradigm function*. Paradigm functions take as input a root and a feature set, and return a phonological form.

Although many approaches to morphology can be defined within this setup, PFM assumes that paradigm functions are further characterized by using a system of *realization rule blocks*. Each rule is assigned a block index, and the paradigm function defines the order in which the blocks must be traversed.

Realization rules come in two varieties. *Rules of exponence* simply associate a phonological modification to the expression of a given feature set. For instance, (13) is the rule used for regular first person plural inflection on verbs in French:[11] the rule states that to express the feature set σ on a base X of category V, if the feature set contains the specification that the person is 1 and the number is plural, then /ɔ̃/ is suffixed to the base.

(13) X_{V},σ: {PER *1*, NB *pl*} → X⊕ɔ̃

[11] Note that we use the notation proposed by Ackerman and Stump (2004) rather than that of (Stump, 2001). We also adopt the HPSG practice of typesetting attribute names in SMALL CAPITALS and atomic values in *italics*.

Rules of referral do not express an exponent directly but refer the realization of a feature set in a given block to that of a possibly different feature set in a possibly different block. For example, in French conjugation, the forms used in first person plural and second person plural are always identical, in all tenses, except for the different person endings (*-ons*/*-mes* and *-ez*/*-tes*).[12] This can be captured by the rule in (14). This rule presupposes an analysis of French conjugation with four rule blocks, where the last block 4 contains rules such as (13) introducing the person endings. Rule (14) belongs to the preceding block 3. The notation $\langle X,\sigma\rangle : \beta$ denotes the result of submitting the form X with feature set σ to the rules of block β. So in words, (14) states that the expression of second person plural on form X is identical to the expression of first person plural on the same form X in block 3.

(14) X_V,σ: {PER *2*, NB *pl*} → ⟨X, σ / {PER *1*}⟩ : 3

Finally the system accommodates the notion of a *portmanteau rule*, a rule corresponding to the traversal of a block sequence instead of a single block. This provides a simple way of accommodating both affix fusion and full form suppletion. In affix fusion, a single rule of realization is used instead of a succession of rules in consecutive blocks. In full form suppletion, a single rule specifies directly the full form associated with some feature bundle for a given lexeme, bypassing the whole block system. An example from appendix A is rule (16) accounting for the use of the suppletive form *sommes* for the first person plural present form of *être*, instead of the expected *étons*.

(15) a. Sample block 1 rule:[13]
X_V,σ: {TENSE *prst*} → Y, where Y is X's basic stem.

b. Sample block 4 rule:
X_V,σ: {PER *1*, NB *pl*} → X⊕ɔ̃

(16) Portmanteau 1-4 rule:
$X_{être},\sigma$: {TREL *eq*, TREF *deictic*, PER *1*, NB *pl*} → sɔm

Note that despite the fact that the inflectional system as a whole is postulated to be a function, nothing precludes multiple rules to be appropriate for the expression of the same feature set in the same block. Thus special care must be taken to make sure that a single result always prevails. In PFM this is done by assuming a specificity ordering on the rules. First, no system can contain two rules in the same block such that both rules may apply to the same input feature set and neither of the two rules is more specific than the other. Second, when facing a choice between two rules within a block, the more specific rule always prevails: this is what Stump calls Panini's principle.[14] These two

[12] The only exceptions to this generalization are the two verbs with a suppletive 2P form in the present indicative, *faire* and *dire*. These are best treated as special cases (see Bonami and Boyé 2002 for discussion) and can be dealt with in PFM using portmanteau rules specific to a single lexeme.

[13] The notation from appendix A has been simplified for ease of exposition.

[14] A special case must be made for portmanteau rules, because these cannot be compared directly to ordinary rules. Stump's assumption, embedded in the Function Composition Default (Stump, 2001: 142)

assumption are sufficient to make sure that the block system indeed produces at most one output for any given input.

3.2. *The Standardized Subsystem*

We start by providing a simple PFM analysis of what we call the "standardized subsystem" of proclitics, that is, the part of the Standard French system modelled by (Miller and Sag, 1997). For simplicity, we follow Monachesi and assume that verbs carry a list-valued CLTS feature expressing what clitics must be realized in a given syntactic context; we leave out the specification of the syntactic constraints determining what ends up on the CLTS list, to avoid the issue of integrating PFM with an explicit syntactic framework.[15]

For the standardized subsystem we rely on a sequence of seven blocks, corresponding to the position classes in table 1. The block system is depicted informally in figure 1, where each block is represented as a transition between two states and each rule as an arrow relating these two states.[16]

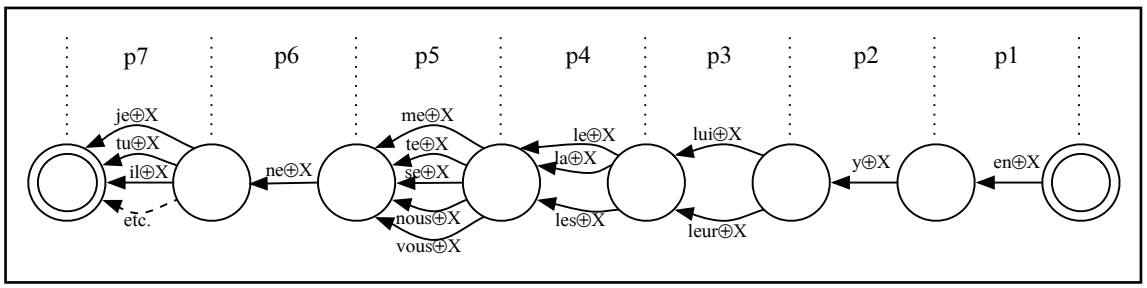


Figure 1: The standardized subsystem

To express the rules, we assume a case/marking system identifying different types of arguments of verbs. The basic cases are *nominative*, *accusative*, *dative* for *à*-marked phrases alternating with the clitics *lui*/*leur*, *locative* for static locative expressions in the sense of (Bonami, 1999), and *de* for phrases marked with *de* in general, be they NPs, PPs or VPs (see Abeillé et al. 2004, 2005 for relevant data and analysis on *de* phrases).[17] We also adopt Miller and Sag's (1998) underspecified *obj* case/marking value which generalizes over *accusative* and *dative*; this is the case value assigned to FPCs in slot p3.

---

is that portmanteau rules always win competition against non-portmanteau rules, whatever the specificity of these.

[15] However it is a simple exercise in typed feature structure modelling to embed a version of PFM in an HPSG grammar; this allows one to use directly Miller and Sag's syntactic analysis in combination with the current morphological analysis.

[16] The similarity to the standard depiction of finite-state automata is not completely fortuitous—see Karttunen (2003) and Malouf (2005) for tentative formulations of PFM as an instance of finite state morphology.

[17] Note that *dative* corresponds to Miller and Sag's $\grave{a}_1$, and *locative* corresponds to Miller and Sag's $\grave{a}_2$.

Rule (17) simply states that a clitic corresponding to a *de*-marked phrase can be realized by prefixing ɑ̃(n) to a verb form in block p1.[18] Parentheses in the phonological representations note a *latent* segment, that is, a segment that is only realized if followed by a segment of the opposite category : latent consonants surface only before vowels, and latent vowels realize only before consonants.[19] Rule (18) indicates that a clitic corresponding to a locative phrase is realized by prefixing [i] in block p2. Rule (19) allows two realizations for a clitic corresponding to a dative nonreflexive in block p3 (dative reflexives can only be realized in block p5).

(17) Block p1
X_V,σ : {CLTS ⟨…,{*de*},…⟩} → ɑ̃(n) ⊕ X

(18) Block p2
X_V,σ : {CLTS ⟨…,{*loc*},…⟩} → i ⊕ X

(19) Block p3
a. X_V,σ : {CLTS ⟨…,{*dat*,*3s*,*nonrefl*},…⟩} → lɥi ⊕ X

b. X_V,σ : {CLTS ⟨…,{*dat*,*3p*,*nonrefl*},…⟩} → lœʁ ⊕ X

Block p4 includes the expected rules realizing a clitic corresponding to an accusative nonreflexive argument. Notice that final [ə]s are not treated as latent vowels, since [ə]s are generally subject to deletion rules in French. A further rule is used to realize clitics corresponding to predicative arguments (adjectival or nominal complements of the copula), whose form is not dependent on gender.

(20) Block p4
a. X_V,σ : {CLTS ⟨…,{*acc*,*3ms*,*nonrefl*},…⟩} → lə ⊕ X

b. X_V,σ : {CLTS ⟨…,{*acc*,*3fs*,*nonrefl*},…⟩} → l(a) ⊕ X

c. X_V,σ : {CLTS ⟨…,{*acc*,*3p*,*nonrefl*},…⟩} → le(z) ⊕ X

d. X_V,σ : {CLTS ⟨…,{PRED +},…⟩} → lə ⊕ X

Block p5 realizes clitics corresponding to reflexive and/or fist/second person accusative or dative arguments.

[18] In the interest of readability we abbreviate feature descriptions where possible. Thus *acc* stands for {CASE *accusative*}, 3ms for {PERSON *3,* GENDER *masculine*, NUMBER *singular*}, etc.

[19] The use of latent segments is more or less standard in descriptions of French phonology. Although they can certainly be dispensed with, their inclusion simplifies greatly the description of morphophonology. See Bonami, Boyé and Tseng (2004) for an explicit formalization of the notion of latent segment in an HPSG grammar of French.

(21) Block p5

a. X_V,σ: {CLTS ⟨…,{*obj*,*1s*},…⟩} → mə ⊕ X

b. X_V,σ: {CLTS ⟨…,{*obj*,*2s*},…⟩} → tə ⊕ X

c. X_V,σ: {CLTS ⟨…,{*obj*,*3*,*refl*},…⟩} → sə ⊕ X

d. X_V,σ: {CLTS ⟨…,{*obj*,*1p*},…⟩} → nu(z) ⊕ X

e. X_V,σ: {CLTS ⟨…,{*obj*,*2p*},…⟩} → vu(z) ⊕ X

The sole rule in block p6 allows for the realization of negative *ne*. In Contemporary French *ne* does not express negation by itself, but acts as a scope-marker indicating where negative words (adverbs such as *pas* 'not', quantifiers such as *personne* 'nobody', adverbs of quantification such as *jamais* 'never') take scope (Kayne, 1984). Following Godard (2004) we assume that the relation between the scope of negation and the realization of the form *ne* is mediated by a morphosyntactic feature ±NE ; thus *ne* is the realization of a {NE +} specification.

(22) Block p6

a. X_V,σ: {NE +} → nə ⊕ X

Block p7 accounts for the realization of nominative pronominal arguments as proclitics. The rules are sensitive to a feature INV which serves as a trigger for subject clitic inversion : contexts licensing a clitic inversion,[20] such as root interrogatives, introduce an {INV +} specification. Here since we are dealing with proclitic realizations the rule is sensitive to an {INV –} specification.

(23) Block p7

a. X_V,σ: {INV –, CLTS ⟨…,{*nom*,*1s*},…⟩} → ʒə ⊕ X

b. X_V,σ: {INV –, CLTS ⟨…,{*nom*,*2s*},…⟩} → ty ⊕ X

c. X_V,σ: {INV –, CLTS ⟨…,{*nom*,*3ms*},…⟩} → il ⊕ X

d. etc.

Finally, note that the rule system above does not block the realization of sequences such as *se lui* (8). We propose to treat this restriction not as part of the rule block system, but as a feature cooccurrence restriction indicating that a reflexive or nonthird person object and a dative cannot co-occur on the CLTS list. A direct advantage of this move is that it accounts directly for the fact that the data in (8) has direct

[20] As is known since Kayne (1972), *clitic subject inversion* should not be confused with *stylistic inversion* of full NPs, which is triggered in a different set of contexts (mostly, extraction contexts) and has different formal properties.

parallels in the enclitic system, despite the fact that different clitics (and thus different position classes) are used (25). Thus we capture a generalization that cannot be captured if the restriction is treated as a restriction on the block system.

(24) Feature cooccurrence restriction
$\{\text{CLTS}\ \langle \ldots, x : \{obj\}, \ldots, \{dat\}, \ldots \rangle\} \Rightarrow x : \{3, nonrefl\}$

(25) a. *Présente lui moi !
present-IMP.2S DAT.S OBJ.1S

b. Présente moi à elle !
present-IMP.2S OBJ.1S to PRO.3FS
'Present me to her!'

3.3. Some Common Variants

In this section we show how the account of the standardized subsystem can be extended to account for the data discussed in section 1. Our basic strategy is to add more rules, many of which are portmanteau rules. The overall layout of the system is as depicted in figure 2.[21]

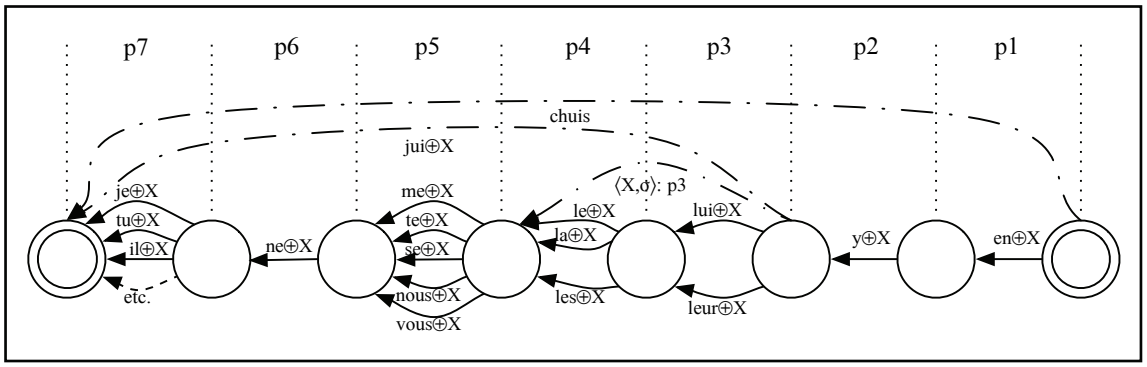


Figure 2: Some common variants

[21] A further complication that should be taken into account is the distribution of so-called 'ethical dative' clitics, which are forms in the series *me*/*te*/*nous*/*vous* which are used in colloquial spoken French to mark emphatically the surprise of the speaker at the reported state of affairs (see Leclère, 1976, for extended discussion). As (i–ii) shows, ethical datives can cooccur with a clitic belonging to block p5; thus they should be treated by adding a further block p5' between p5 and p6. Rule (iii) could be used to this effect, assuming that {SURPRISE +} is a placeholder for whatever is the correct characterization of the pragmatic effect of the ethical dative.

(i) Il t' a fait un de ces boucans !
he ETH.DAT has done one of these noises
'He did so much noise!'

(ii) Il te nous a passé un de ces savons !
he ETH.DAT us has passed one of these soaps
'He gave us an incredible telling-off!

(iii) Block p5′
X_V,σ: {SURPRISE +} → ⟨X, {CLTS ⟨…,{*dat* ,*nonrefl*},…⟩}⟩ : p3

To account for elided realizations of *tu*, we include an extra rule in block p7 stating that *tu* can be generated with a final 'latent' vowel, which surfaces only if followed by a consonant. Note that since the original *tu* rule is not suppressed, we end up with optional rather than obligatory elision.

(26) Block p7

$X_{V,\sigma}$: {INV –, CLTS ⟨…,{*nom*,*2s*},…⟩} → t(y) ⊕ X

We account for the possibility of realizing *je suis* as [ʃɥi] by stating a portmanteau rule covering the whole block system. This rule states that if we are dealing with a first person singular present indicative form of the lexeme *être*, and that a subject clitic must be realized, then [ʃɥi] is a possible realization. Note that the rule states that the subject clitic must be the sole member of the CLTS list, thus correctly barring [ʃɥi] from being a realization of e.g. *je le suis*. On the other hand, [ʃɥi] can be used in negative contexts, as a variant of *je ne suis*, since the rule does not exclude the possibility of a {NE +}) specification.

(27) Portmanteau p1–p7 rule

$X_{\text{être},\sigma}$: {INV –, CLTS ⟨{*nom*,*1s*}⟩,MODE *indic*, TENSE *prst*} → ʃɥi

A similar, but different rule is used to account for reduced realizations of *je lui* sequences. We posit a portmanteau rule covering blocks p3 to p7 and stating that both nominative 1s and a dative 3s clitics can be realized by the single sequence [ʒɥi]. Note that the rule completely bypasses the expression of features that could be realized by blocks p4, p5, p6. As a matter of fact this desirable: (i) realization of *ne* in p6 is optional in standard French, so that the [ʒɥi] reduction is possible even in negative (NE +) contexts; (ii) the restriction in (24) excludes the possibility that an input to rule (28) also asks for a clitic to be realized in p5; and finally the realization of block p5 clitics is optional before block p4 clitics, as already noted above and accounted for directly below in the general case, so that [ʒɥi] is indeed a possible realization for a nominative-accusative-dative clitic sequence. Finally, the rule correctly authorizes the realization of clitics from blocks p1 and p2, accounting for examples such as *je lui en donnerai* [ʒɥiɑ̃dɔnʁɛ].

(28) Portmanteau p3–p7 rule

$X_{V,\sigma}$: {INV –, CLTS ⟨…,{*nom*,*1s*},…,{*dat*,*3s*,*nonrefl*},…⟩} → ʒɥi ⊕ X

To account for clitic drop we use a portmanteau rule which is also a rule of referral. This rule basically states that when realizing both a nonreflexive third person accusative and a nonreflexive third person dative, one option is to refer to the realization of the same feature set in block p3 while ignoring block p4. As the reader can check this captures our generalization correctly.

(29) Portmanteau p3–p4 rule
X_V,σ : {CLTS ⟨…,{*acc*,*3s*,*nonrefl*},{*dat*,*3s*,*nonrefl*},…⟩} → ⟨X, σ⟩ : p3

4. A Relational Version of PFM

4.1. The Issue

The rule system defined in section 3 seems to account adequately for all the properties of FPCs discussed in section 1, with one big caveat. In Paradigm Function Morphology, as the name indicates, inflectional morphology takes the form of a function—for each possible input the system provides no more than one output. To make sure that the rule block system indeed provides a single output for each input, PFM relies on Panini's principle: for every possible input there must be a most specific rule in each block compatible with the feature set to be expressed, which is chosen as the output of that block.

As it stands, the rule system defined above both violates Panini's principle and produces inadequate results. First, block p7 contains two rules applicable to the exact same feature set, for elided and non-elided *tu*. Second, many rules turn out to be applied obligatorily because they are specific enough, blocking the application of another possible rule. For instance rule (27) is very specific, and blocks the possibility of realizing *je suis* as [ʒəsɥi], which is empirically incorrect. The same holds for the two other portmanteau rules in (28) and (29).

Thus it seems that PFM has one design property that is at odds with the empirical characteristics of the FPC system: the FPC system is not functional, but relational. On the other hand, the rule formats authorized by PFM have been shown to provide adequate means of modelling some of the peculiar properties of FPCs. Thus an adequate analysis can be provided by modifying PFM so that the functional requirement is dropped.

Formulating a relational version of PFM is a somewhat tedious task ; a proposal can be found in appendix B. Intuitively, what needs to be done is to define traversing a block of rules as a process which may give rise to different outputs, according to the rules that have been chosen. Likewise, the notation ⟨*X*,σ⟩ : β should be interpreted not as denoting the result of applying the narrowest rule in block β to ⟨X,σ⟩, but as describing nondeterministically any of the different ways for block β to provide an output for the input ⟨X,σ⟩.[22] One small complication is the status of the Identity Default condition, the condition that when no rule in a block can apply, then the block produces an output identical to its input. In standard PFM, this is dealt with by assuming that each block contains a special, identity rule of exponence. In the relational version, this won't work, because it would amount to making all realization rules optional (since every rule can now apply irrespective of specificity). The solution is to embed the Identity Default condition in the definition of rule blocks, and to state that a block may be traversed with no modification if and only if no rule within that block may be applied to the input.

[22] Note that this means that realization rules themselves are relations and not functions, since rules of referral must now be able to refer to any of the possible realizations of a form/feature set pair within a given block.

4.2. Are Inflectional Systems Functional?

We have shown that it is possible to define a relational version of PFM that can accommodate the type of inflectional behaviour exhibited by FPCs. What remains to be addressed is an empirical issue: typical inflectional systems seem to be functional in Stump's sense. Thus what is it about FPCs that triggers the use of a relational system?

We think that this question cannot be answered directly, and that a broad empirical overview is needed: it might well be the case that relational inflectional systems are more common than (Stump, 2001) seems to assume. Here are a few examples collected at random:

(30) In Spanish, every verb has two sets of forms for the subjunctive imperfective, a set in *-ra* and a set in *-se* (Alcoba, 1999). Each set provides one form for each of the six persons. For example, with the verb *amar* 'love', the following forms of the subjunctive imperfective alternate freely: *amara/amase, amaras/amases, amara/amase, amáramos/amásemos, amarais/amaseis, amaran/amasen*.

(31) In Nepali, negative polarity is morphologically marked in the conjugation (Adhikari 1993). For the present tense, there are two inflectional forms for each person of the negative paradigm: a long form and a short form (except for verb stems ending in -VC which only have long forms). These negative forms are in free variation: *birsanna*/*birsãdinã* 'I don't forget', *birsannas/birsãdejnas* 'you don't forget', etc.

(32) In Finnish, there are two inflectional forms for the genitive plural: a weak form and a strong form (Anttila 1997). For words of one or two syllables, the two forms are in complementary distribution based on phonological criteria. For longer words, free variation arises : *fyysikko* 'physicist', for example, has two genitive plural forms *fyysikkojen* (weak) and *fyysikoiden* (strong).

This list is definitely far from exhaustive, and might not be representative either. However it is sufficient to show that we should not presuppose as self-evident the idea that inflectional systems are functional; whether they are or not seems to be an open research question. Be that as it may, it remains that the nonfunctionality exemplified here is a lot more limited than what we have found in the FPC system. In the three cases mentioned, we could account for the data by assuming a supplementary morphosyntactic feature encoding a distinction between two subparadigms. Although such a feature would be spurious from a morphosyntactic point of view, such local inefficiencies are not unexpected in natural language grammars. By contrast, our observations suggest that the relational character of the FPC system cannot be reduced by using a simple feature distinction, since alternate realizations are distributed throughout the system. We conjecture that nonfunctionality is the gist of the difference between argument realization systems such as the FPC system and true inflection.

Appendix A: A PFM Analysis of French Conjugation

In this appendix we present succinctly a PFM analysis of French conjugation. This analysis serves as an illustration of the workings of PFM, and provides the bases for the PFM analysis of French clitics : the clitic systems takes inflected verb forms as its input.

A.1 Morphosyntactic Features of French Verbs

We follow the tradition in assuming that French verbs inflect for TENSE, MODE, NUMBER, PERSON and GENDER. The only point where we depart from this tradition is in the exact makeup of the TENSE and MODE classification.[23] We adopt Verkuyl *et al.* (2004)'s neo-Reichenbachian (Reichenbach, 1947) analysis allowing one to make sense of the role of the *imparfait* and *conditional* tenses in the French system. Verkuyl et al. propose that the indicative tense system be analyzed along two dimensions: the *temporal relation* between reference time and event time may be either precedence, equality, or succession, giving rise to past, present or future tenses. Independently, the identification of the reference time may be done on a *deictic* basis, i.e. by identifying it with the speech time ; or on an *anaphoric* basis, by identifying it with a previously established time discourse referent. The *imparfait* is the anaphoric version of the present, and the *conditional* is the anaphoric version of the future (there happens to be no anaphoric version of the past). This allows one to make sense e.g. of the distribution of tenses in conditional sentences, or in dependent clauses : in conditionals sentences, antecedents in the *imparfait* associate with consequents in the *conditional* tense, whereas antecedents in the *present* associate with consequents in the *future*. Likewise, when reporting an attitude towards the future, a *present tense* report entails the use of the future in the dependent clause, whereas an *imparfait* report entails the use of the *conditional*.

(33) a. Si Jean vient, il sera furieux.
if Jean come-PRST he be-FUT furious
'If Jean comes he will be furious.'

b. Si Jean venait, il serait furieux.
if Jean come-IMPF he be-COND furious
'If Jean came he would be furious.'

(34) a. Jean pense qu' il viendra.
Jean think-PRST that he come-FUT
'Jean thinks he will come.'

b. Jean pensait qu' il viendrait.
Jean think-IMPF that he come-COND
'Jean thinks he will come.'

[23] Note that we do not model directly periphrastic tenses, which we take to be generated by syntax. See Abeillé and Godard (2002) for a thorough syntactic analysis of French compound tenses.

The feature inventory we propose is given in (35). Table 3 makes explicit the correspondence between tense feature values and the traditional names of French tenses. Note that we treat PERSON, NUMBER and GENDER features as direct features of the verb, rather than using a embedded feature structure specifying what the verb agrees with. This is motivated by the fact that, depending on the syntactic context, French past participles may agree with the subject or the object (36). Yet there is no separate set of exponents for subject and object agreement: rather, the same exponents are used for agreement with whatever target syntax asks for. Thus we assume that the verb carries appropriate PERSON, NUMBER and GENDER features, and that syntax will determine how these features covary with features of an NP in the context.

(35) Features and values

a. MODE : *indicative*, *subjunctive*, *imperative*, *participle*

b. TREF : *anaphoric*, *deictic*

c. TREL : *precedence*, *equality*, *succession*

d. PER : *1*, *2*, *3*

e. GEN : *mas*, *fem*

f. NB : *sg*, *pl*

	{TREL *prec*}	{TREL *eq*}	{TREL *succ*}
{MODE *indic*, TREF *deictic* }	simple past	present	future
{MODE *indic*, TREF *anaphoric* }	—	*imparfait*	conditional
{MODE *subj*}	past subjunctive	present subjunctive	—
{MODE *part*}	past pasticiple	present participle	—

Table 3 – Correspondence between feature values and traditional tenses

(36) a. Auxiliary *avoir*, in-situ object : no agreement
Paul a écrit la lettre.
Paul.MS has write-PASTP.MS.SG the.FS letter.FS
'Paul wrote the letter.'

b. Auxiliary *avoir*, extracted object : object agreement
la lettre que Paul a écrite
the.FS letter.FS that Paul.MS has write-PASTP.FEM.SG
'the letter Paul wrote'

c. Auxiliary *être*: subject agreement
Marie est morte.
Marie.FS is die-PASTP.FEM.SG
'Marie died.'

Finally (37) lists the appropriate feature cooccurrence restrictions : (a) the TREF feature is appropriate only in the indicative, thus there is only one tense series in other moods ; (b) future tenses exist only in the indicative ; (c) there are no tense distinctions in the infinitive and the imperatives ; (d) there is no anaphoric past tense ; (e) only finite moods are compatible with PERSON ; (f) only past participles have GENDER ; (g) infinitives do not inflect for NUMBER ; and finally (h) imperatives have only three forms.

(37) a. $\{\text{TREF } x\} \Rightarrow \{\text{MODE } \textit{indic}\}$

b. $\{\text{TREL } \textit{succ}\} \Rightarrow \{\text{MODE } \textit{indic}\}$

c. $\{\text{TREL } x\} \Rightarrow \neg(\{\text{MODE } \textit{inf}\} \vee \{\text{MODE } \textit{imper}\})$

d. $\neg(\{\text{TREF } \textit{ana}\} \wedge \{\text{TREL } \textit{prec}\})$

e. $\{\text{PER } x\} \Rightarrow \neg(\{\text{MODE } \textit{inf}\} \vee \{\text{MODE } \textit{part}\})$

f. $\{\text{GEN } x\} \Rightarrow (\{\text{MODE } \textit{part}\} \wedge \{\text{TREL } \textit{prec}\})$

g. $\{\text{NB } y\} \Rightarrow \neg(\{\text{MODE } \textit{inf}\})$

h. $\{\text{MODE } \textit{imper}\} \Rightarrow (\{\text{PER } \textit{2}\} \vee (\{\text{PER } \textit{1}\} \wedge \{\text{NB } \textit{pl}\}))$

A.2 The Stem Space

French conjugation is characterized by a complex system of stem allomorphies. Whereas inflectional exponents are uniform for all but a handful of verbs (see table 4), various kinds of stem alternations motivate the traditional clasification of verbs in three groups. Verbs in group I have three distinct stems : basic stem X used by default, a stem Xe in the infinitive and past participle, Xa in the simple past and past subjunctive. Verbs in group II have two stems with a very different distribution : Xi in the present singular, the imperative singular, the infinitive, future, conditional, and past participle, Xis elsewhere. Group III is really a repository for all verbs that pattern neither with group I nor with group II, and a few dozen distinct patterns of stem alternation are attested there. For instance the verb *aller* ‘go’ has six distinct stems : [va] in the present and imperative singular, [ale] in the infinitive and past participle, [ala] in the indicative and subjunctive past, [iʁ] in the future and conditional, [aj] in the present subjunctive singular, [al] elsewhere.

	PER *1*, NB *sg*	PER *2*, NB *sg*	PER *3*, NB *sg*	PER *1*, NB *pl*	PER *2*, NB *pl*	PER *3*, NB *pl*
MODE *indic*, TREF *deictic*, TREL *eq*	id.	id.	id.	⊕ɔ̃	⊕e	id.
MODE *indic*, TREF *ana*, TREL *eq*	⊕ɛ	⊕ɛ	⊕ɛ	⊕jɔ̃	⊕je	⊕ɛ
MODE *indic*, TREF *deictic*, TREL *prec*	raise	id.	id.	⊕m	⊕t	raise ∘ ⊕ʁ
MODE *indic*, TREF *deictic*, TREL *succ*	⊕ʁɛ	⊕ʁa	⊕ʁa	⊕ʁɔ̃	⊕ʁe	⊕ʁɔ̃
MODE *indic*, TREF *ana*, TREL *succ*	⊕ʁɛ	⊕ʁɛ	⊕ʁɛ	⊕ʁjɔ̃	⊕ʁje	⊕ʁɛ
MODE *subj*, TREL *eq*	id.	id.	id.	⊕jɔ̃	⊕je	id.
MODE *subj*, TREL *prec*	⊕s	⊕s	id.	⊕sjɔ̃	⊕sje	⊕s
MODE *imper*	—	id.	—	⊕ɔ̃	⊕e	—

	GEN *mas*, NB *sg*	GEN *fem*, NB *sg*	GEN *mas*, NB *pl*	GEN *fem*, NB *pl*
MODE *part*, TREL *prec*	shorten	id.	shorten	id.

MODE *part*, TREL *eq*	⊕ɑ̃
MODE *inf*	⊕ʁ

Table 4 – Inflectional exponents for verbal feature combinations[24]

From a thorough investigation of the stem alternation data, Bonami and Boyé (2002, 2003) conclude that there is no hope of explaining the choice and distribution of stems on a morphosyntactic or morphophonological basis. French verbal stems are thus *morphomes* in Aronoff's (1994) sense : pure morphological objects which do not express any features and are chosen on an arbitrary basis. To model this explicitly, Bonami and Boyé (2002) propose that each verbal lexeme come equipped with a *stem space*, a data structure with 12 distinct slots that may be filled with identical, related or non-related stems. Individual inflectional rules then refer to a specific slot in the stem space to decide what endings attach to what stem. Table 5 indicates the correspondence between stem slots and inflectional forms.

[24] Exponents are notated as functions: "id." is the identity function ; for any X, "⊕X" is the function suffixing X to its input; "raise" is a function raising a low final vowel to a mid-open vowel (it has no effect on nonlow vowels), thus turning [lava] into [lave] but [fini] into [fini]; "shorten" is a function suppressing a final consonant, thus turning [ɛ̃klyz] into [ɛ̃kly] but [kɔ̃kly] into [kɔ̃kly]; "∘" is function composition. Note that we ignore the issue of *liaison* consonants throughout. Note further that we assume that first group verbs take a final [ʁ] in the infinitive, which is then suppressed by a general phonological rule deleting [ʁ] after [e] (Boyé, 2000).

slot	inflectional forms	slot	Inflectional forms
1	*imparfait*, present 1pl and 2pl	7	present subjunctive singular and 3pl
2	present 3pl	8	present subjunctive 1pl and 2pl
3	présent sg	9	infinitive
4	present participle	10	future, conditional
5	imperative singular	11	simple past, subjunctive past
6	imperative plural	12	past participle

Table 5 – The French verbal stem space

The crux of Bonami and Boyé (2002, 2003) and later research (e.g. Bonami and Boyé 2005, 2006 ; Bonami, Boyé and Kerleroux, to appear) is to determine how the stem space is structured, and how regular and irregular inflection patterns can be explicated as different ways of filling it. Here we will remain agnostic on the exact formulation of such constraints, which we take to be orthogonal to the realization of inflectional exponents. Thus we will presuppose the existence of the stem space, and use *stem selection rules* (Stump, 2001: chap. 6) in the first rule block to serve as an interface between the stem space and the paradigm function. Concretely, we will assume twelve functions Stem_1, Stem_2, …, Stem_{12}, from roots to phonological forms such that $\text{Stem}_1(X)$ returns the stem occupying slot 1 in the stem space of X, etc.

A.3 The Rules

We propose a system of four rule blocks, whose general structure is outlined in figure 3. Block 1 rules are stem selection rules. Block 2 rules add the [ʁ] realizing {TREL *succ*} in the future and conditional tenses, and the [s] expressing the past subjunctive. Block 3 accounts for the suffixes [j] and [ɛ] realizing {TREF *ana*} in the *imparfait* and conditional (note that this corresponds to the traditional notion that [j] and [ɛ] are allomorphs) and for [j] in the subjunctive. Finally block 3 accounts for the final person/number endings.

Block 1 rules are listed in (38). Since there are 12 different stem slots to choose from, we need at least 12 different rules, one for each slot. The assumption of Panini's principle helps limit the number of rules and eases their formulation by allowing quite a bit of underspecification. For instance, we need not specify overtly that Stem 1 is selected in the *imparfait* and in the present indicative 1pl and 2pl. Rather, (38)a simply specifies that stem 1 realizes {TREL *eq*}, and is overruled by (38)g and (38)h in the relevant cases. The only place where Panini's principle is not sufficient to allow a compact formulation is for stem selection for the subjunctive present: here Stem 7 and Stem 8 are used for two subparts of the paradigm neither of which can be given a nondisjunctive description. Thus we need two separate rules for the selection of stem 8, respectively (38)e and (38)f.[25]

[25] Note that we could equivalently state (38)f as a rule of referral.

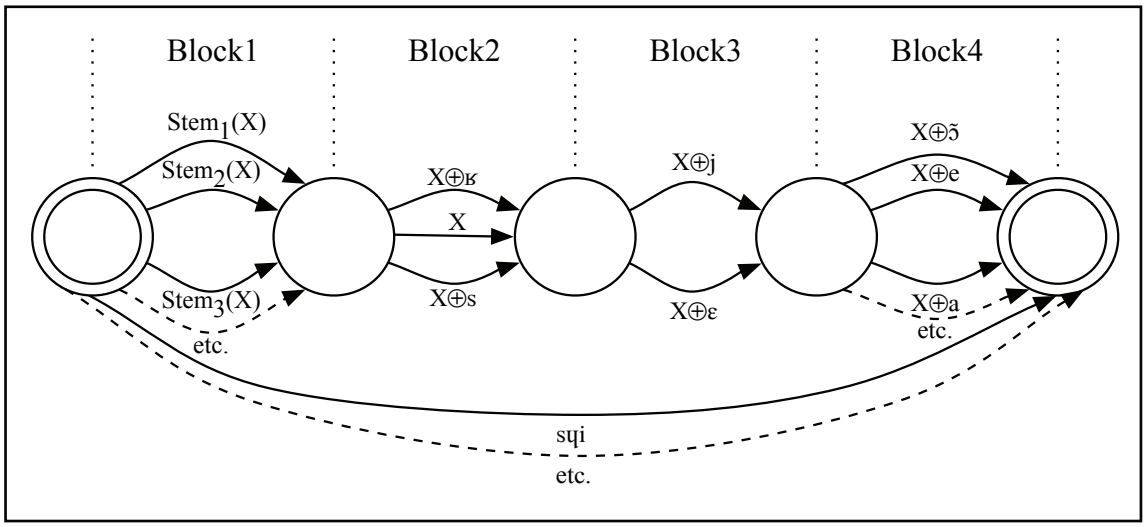


Figure 3 – The French conjugation block system

(38) Block 1 rules

a. X_V,σ: {TREL *eq*} → $Stem_1(X)$

b. X_V,σ: {TREL *prec*} → $Stem_{11}(X)$

c. X_V,σ: {TREL *succ*} → $Stem_{10}(X)$

d. X_V,σ: {TREL *eq*, MODE *subj*} → $Stem_7(X)$

e. X_V,σ: {TREL *eq*, MODE *subj*, PER *1*, NB *pl*} → $Stem_8(X)$

f. X_V,σ: {TREL *eq*, MODE *subj*, PER *2*, NB *pl*} → $Stem_8(X)$

g. X_V,σ: {TREL *eq*, TREF *deictic*, NB *sg*} → $Stem_3(X)$

h. X_V,σ: {TREL *eq*, TREF *deictic*, PER *3*, NB *pl*} → $Stem_2(X)$

i. X_V,σ: {MODE *imper* } → $Stem_6(X)$

j. X_V,σ: {MODE *imper*, NB *sg*} → $Stem_5(X)$

k. X_V,σ: {MODE *inf*} → $Stem_9(X)$

l. X_V,σ: {MODE *part*} → $Stem_4(X)$

m. X_V,σ: { TREL *prec*, MODE *part*} → $Stem_{12}(X)$

Rules in blocks 2 to 4 attempt to account for the realizations noted in table 4 in the most economic fashion possible. Block 2 (39) contains just two rules, for future/conditional [ʁ], and subjunctive past [s]. A third, somewhat unnatural-looking rule is needed to account for the absence of [s] in the subjunctive past 3sg. Note that (to the extent that they use the subjunctive past, which is by far the least used tense in contemporary French) speakers tend to realize an [s] here, suggesting that the relevant rule is indeed unnatural.

(39) Block 2 rules

a. X_V,σ: {TREL *succ*} → X⊕ʁ

b. X_V,σ: {TREL *prec*, MODE *subj*} → X⊕s

c. X_V,σ: {TREL *prec*, MODE *subj*, PER *3*, NB *sg*} → X

Block 3 (40) contains rules realizing [ɛ] in the *imparfait* and conditional, and [j] in the *imparfait*, conditional, subjunctive present and subjunctive past. We do no attempt to account for all occurrences of /j/ with a single rule, since we know of no good morphosyntactic reason to consider the *imparfait*, conditional and subjunctive to form a natural class. However we can account for the general similarity between 1pl and 2pl forms by using a rule of referral that is insensitive to tense and mode.

(40) Block 3 rules

a. X_V,σ: {TREF *ana*} → X⊕ɛ

b. X_V,σ: {TREF *ana*, PER *1*, NB *pl*} → X⊕j

c. X_V,σ: {MODE *subj*, PER *1*, NB *pl*} → X⊕j

d. X_V,σ: {PER *2*, NB *pl*} → ⟨X, σ / {PER *1*}⟩ : 3

Finally block 4 (41) accounts for the final endings.

(41) Block 4 rules

a. X_V,σ: {PER *1*, NB *pl*} → X⊕ɔ̃

b. X_V,σ: {PER *2*, NB *pl*} → X⊕e

c. X_V,σ: {TREL *succ*, TREF *deictic*, NB *sg*} → X⊕a

d. X_V,σ: {TREL *succ*, TREF *deictic*, PER *1*, NB *sg*} → X⊕ɛ

e. X_V,σ: {TREL *succ*, TREF *deictic*, PER 3, NB *pl*} → X⊕ɔ̃

f. X_V,σ: {TREL *prec*, TREF *deictic*, PER *1*, NB *sg*} → raise(X)

g. X_V,σ: {TREL *prec*, TREF *deictic*, PER *1*, NB *pl*} → X⊕m

h. X_V,σ: {TREL *prec*, TREF *deictic*, PER 2, NB *pl*} → X⊕t

i. X_V,σ: {TREL *prec*, TREF *deictic*, PER 3, NB *pl*} → raise(X)⊕ʁ

j. X_V,σ: {MODE *inf*} → X⊕ʁ

k. X_V,σ: {MODE *part*, TREL *eq*} → X⊕ɑ̃

l. X_V,σ: {MODE *part*, GEN *mas*} → shorten(X)

A.4 Suppletive Forms

We must account for the existence of a small number of *suppletive inflectional forms* : these are forms whose suppletive character is not limited to the use of a suppletive stem, but are characterized by the absence of normal exponents from block 4. Following Stump (2001) we treat these as the effect of portmanteau rules which are specific to a single lexeme.[26] Note that in a number of cases a more sophisticated analysis is feasible. For instance we could replace (42)e with a rule of referral stating that the form *dites* results from the combination of the verb's stem 3 with the exponent of 2pl normally found only in the simple past (Morin, 1987; Kilani-Schoch and Dressler, 2005). We refrain from writing such rules since (i) there is no empirical way of testing the acuity of a rule whose application is limited to a single lexeme, and (ii) they do not result in a simplification, but in a complexification, of the rule system.

(42) Portmanteau 1–4 rules

a. X_{aller},σ: {TREL *eq*, TREF *deictic*, PER *1*, NB *sg*} → vɛ

b. X_{aller},σ: {TREL *eq*, TREF *deictic*, PER *3*, NB *pl*} → vɔ̃

c. X_{avoir},σ: {TREL *eq*, TREF *deictic*, PER *1*, NB *sg*} → ɛ

d. X_{avoir},σ: {TREL *eq*, TREF *deictic*, PER *3*, NB *pl*} → ɔ̃

e. X_{dire},σ: {TREL *eq*, TREF *deictic*, PER 2, NB *pl*} → dit

[26] Note that technically, we could equivalently put these rules in block 4: since the ouput of each rule is insensitive to the input form X, there is no harm in saying that blocks 1 to 3 have been traversed normally before the suppletive form is selected.

f. $X_{dire,\sigma}$: {MODE *imper*, PER 2, NB *pl*} → dit

g. $X_{être,\sigma}$: {TREL *eq*, TREF *deictic*, PER *1*, NB *sg*} → sɥi

h. $X_{être,\sigma}$: {TREL *eq*, TREF *deictic*, PER *1*, NB *pl*} → sɔm

i. $X_{être,\sigma}$: {TREL *eq*, TREF *deictic*, PER *2*, NB *pl*} → ɛt

j. $X_{être,\sigma}$: {TREL *eq*, TREF *deictic*, PER *3*, NB *pl*} → sɔ̃

k. $X_{faire,\sigma}$: {TREL *eq*, TREF *deictic*, PER 2, NB *pl*} → fɛt

l. $X_{faire,\sigma}$: {MODE *imper*, PER 2, NB *pl*} → fɛt

k. $X_{faire,\sigma}$: {TREL *eq*, TREF *deictic*, PER *3*, NB *pl*} → fɔ̃

Appendix B: A Relational Version of PFM

In this appendix we formulate a relational version of PFM, that is, a system of realizational morphology that is intended to be similar to PFM except for the fact that there may be more than one output for every input. In specifying the system, we attempt to stay as close as possible to the original design of PFM as specified in Stump (2001).

One design decision that we will not follow is the idea that realization rules take as input and output *indexed forms*. Here we simply assume that realization rules are applied to a triplet consisting of a phonological form, a lexeme, and a feature set. Representing lexemes explicitly in the input of rules allows for an easier formalization of stem selection rules.

(43) Let A and V be two sets. The set P of *attribute-value pairs* over A and V and the set S of *feature structures* over A and V are the two smallest sets such that :

a. If $a \in A$ and $v \in V$, then $\langle a,v\rangle \in P$.

b. If $s \subseteq P$ and $\forall a \forall v[\langle a,v\rangle \in s \rightarrow \forall v'[\langle a,v'\rangle \in s \rightarrow v = v']]$, then $s \in S$.

c. If $a \in A$ and $s \in S$, then $\langle a,s\rangle \in P$.

(44) A *feature system* is a triple $\langle A, V, \Lambda\rangle$ where

a. A is a set of atoms, the *attributes*.

b. V is a set of atoms, the *atomic values*.

c. Λ is a subset of S, the set of feature structures over A and V.

Λ corresponds to the set of licit feature structures licensed by the feature system. Such a set is usually characterized by a combination of feature cooccurrence restrictions; here we will remain vague as to the exact formulation of a language for writing such restrictions, and just assume Λ as given.27

(45) Let Φ be the set of phonological sequences. We call any binary relation on Φ a *morphophonological relation*

(46) A *block index system* is a linear order $B=\langle I,<\rangle$ on a finite set I.

(47) A *lexeme system* is a set of object representing lexemes.

The makeup of a lexeme object will depend on the inflectional system to be analyzed. At the very least, each lexeme should specify some phonological sequence which will serve as the base for the application of rules, the lexeme's *root.*

(48) Given a feature system F, a block index system B and a set L of lexemes, a *realization rule* is a 4-tuple $R=\langle i,c,t,\rho\rangle$, where

a. i is a set of adjacent members of B, the block sequence the rule applies in.

b. c is a subset of L, the set of lexemes the rule applies to.

c. t is a feature structure of F, the feature structure realized by the rule.

d. ρ is a morphophonological relation, the change effected by the rule.

(49) Given a lexeme l, a feature set σ, a sequence x, and a rule $r = \langle i,c,t,\rho\rangle$, y is an *output* of rule r for the input triple $\langle\sigma,l,x\rangle$ just in case :

a. $l \in c$;

b. $t \subseteq \sigma$;

c. $\langle x,y\rangle \in \rho$.

(50) A *realization system* is a 4-tuple $\Sigma=\langle B,F,L,R\rangle$, where B is an block index system, F is a feature system, L is a lexeme system, and R is a set of realization rules over B, F, and L.

[27] Using the language of Gazdar *et al.* (1985) is an obvious choice here, since it is the choice Stump (2001) makes. However this language is not sufficient to formulate the FCR in (24). An alternative would be to use a much richer language for feature structure description, such as RSRL (Richter, 2000). This is probably a reasonable choice if one wants to embed the current analysis in a general grammar, but seems overly complex for our current purposes.

(51) Given a sequence of block indices *i*, a realization system Σ is said to associate *x* to *y* through block sequence *i* for feature set σ and lexeme *l* if and only if :

a. either *R* contains a rule $r = \langle i,c,\tau,\rho\rangle$ such that *y* is an output of rule *r* for input $\langle\sigma,l,x\rangle$;

b. or there is no such rule, *i* is a singleton and $x = y$.

(52) A realization system Σ licenses *x* as a realization of feature set σ for lexeme *l* if and only if there is a partition of *B* in block sequences $i_1, \ldots, i_n$ such that

a. Σ associates the root of *l* to a sequence x_1 through block sequence i_1 ; and

b. Σ associates x_1 to a sequence x_2 through block sequence i_2 ; and

c. …

d. Σ associates x_{n-1} to *x* through block sequence i_n.

Note that nothing so far constrains the nature of the morphophonological relations embedded in the realization rules : any set of associations between input and output is a possible effect of a rule. This is a design feature, since we do not want to constrain overly the types of rules we authorize (and Stump 2001 itself is less than explicit on an exact inventory). However we can define two familiar rule types :

(53) A realization rule $r = \langle i,c,\tau,\rho\rangle$ is a *rule of exponence* if and only if there is a morphophonologically natural[28] function $f: \Phi \rightarrow \Phi$ such that for any *x* and *y*, $\langle x,y\rangle \in \rho$ just in case $f(x)=y$.

In such a case, we write the effect of *r* as : X_c, $\sigma: \tau \rightarrow f(X)$

(54) Given a realization system $\Sigma=\langle B,F,L,R\rangle$, a rule $r = \langle i,c,\tau,\rho\rangle$ in Σ is a *rule of referral* if and only if there is a block sequence i' of *B* and a function $\delta : \Lambda \rightarrow \Lambda$ such that for every lexeme *l* and feature set σ, *y* is a realization of $\langle\sigma,l,x\rangle$ by rule *r* just in case $\Sigma'=\langle B',F,\{l'\},R\rangle$ licenses *y* as a realization of feature set $\delta(\sigma)$ for lexeme l', where l' is an abstract lexeme with root *x* and $B'=\langle i',<'\rangle$, where $<'$ is the restriction of < to i'.

In such a case, we write the effect of *r* as : X_c, $\sigma: \tau \rightarrow \langle X, \delta(\sigma)\rangle{:}i'$

[28] Of course this definition presupposes a definition of morphophonologically natural functions. We may safely assume that e.g. affixation functions are to be taken as natural.

References

Abeillé, Anne, Olivier Bonami, Danièle Godard, and Jesse Tseng (2004). "The syntax of french *de*-N′ phrases". In S. Müller (ed.), *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar*. Stanford (Cal.): CSLI Publications, 6-26.

Abeillé, Anne, Olivier Bonami, Danièle Godard, and Jesse Tseng (2005). "The syntax of french *à* and *de*: an HPSG analysis. In P. Saint-Dizier (ed.), *Dimensions of the Syntax and Semantics of Prepositions*. Berlin: Springer.

Abeillé, Anne, and Danièle Godard (2002). "The syntactic structure of French auxiliaries". *Language*, 78, 404-452.

Ackerman, Farrell and Gregory T. Stump (2004). "Paradigms and periphrastic expression". In L. Sadler and A. Spencer (eds.), *Projecting Morphology*. Stanford (Cal.): CSLI Publications, 111-157.

Adhikari, Hemang Raj (1993). *Samsamayik Nepali Vyakaran*. Kunjal Prakashan, Nepal.

Alcoba, Santiago (1999). "La flexión verbal". In I. Bosque and V. Demonte (eds.), *Gramática Descriptiva de la Lengua Española*. Madrid: Espasa Calpe, 4915-4991.

Anderson, Stephen R. (1992). *A-Morphous Morphology*. Cambridge (Mass.): Cambridge University Press.

Anttila, Aarto (1997). "Deriving variation from grammar". In F. Hinskens, R. van Hout, and W. L. Wetzels (eds.), *Variation, Change and Phonological Theory*. Amsterdam: John Benjamins, 35-68.

Aronoff, Mark (1994). *Morphology by itself*. Cambridge (Mass.): MIT Press.

Auger, Julie (1995). "Les clitiques pronominaux en français parlé informel : une approche morphologique". *Revue Québécoise de Linguistique*, 24, 21-60.

Bonami, Olivier (1999). *Les constructions du verbe : le cas des groupes prépositionnels argumentaux*. Doctoral dissertation. Université Paris 7.

Bonami, Olivier and Gilles Boyé (2002). "Suppletion and dependency in inflectional morphology". In F. Van Eynde, L. Hellan and D. Beerman (eds.), T*he Proceedings of the HPSG'01 Conference*. Stanford : CSLI Publications.

Bonami, Olivier and Gilles Boyé (2003). "Supplétion et classes flexionnelles dans la conjugaison du français". *Langages,* 152, 102-126.

Bonami, Olivier and Gilles Boyé (2005). "Remarques sur les bases de la conjugaison". Manuscript.

Bonami, Olivier and Gilles Boyé (2006). "Subregular defaults in French conjugation". Paper read at the 12th International Morphology Meeting, Budapest, May 2006.

Bonami, Olivier, Gilles Boyé, and Françoise Kerleroux (to appear). "L'allomorphie radicale et la relation flexion-construction". In B. Fradin, F. Kerleroux and M. Plénat (eds.), *Aperçus de morphologie*. Saint Denis: Presses Universitaires de Vincennes.

Bonami, Olivier, Gilles Boyé and Jesse Tseng (2004). An integrated analysis of French liaison. In G. Jaeger (ed.), *Formal Grammar Preproceedings*, Nancy, France.

Boyé, Gilles (2000). *Problèmes de morpho-phonologie verbale en français, espagnol et italien*. Doctoral dissertation. Université Paris 7.

Couquaux, Denis (1986). "Les pronoms faibles sujet comme groupes nominaux". In M. Ronat and D. Couquaux (eds.), *La grammaire modulaire*. Paris : Editions de Minuit, 25-46.

Crysmann, Berthold (2002). *Constraint-based Coanalysis*. Doctoral dissertation. Universität des Saarlandes.

Delais-Roussarie, Elisabeth (2001). "Prosodie des clitiques en Français". In C. Muller, P. de Carvalho, L. Labrune, F. Lambert, and K. Ploog (eds.), *Clitiques et cliticisation, Actes du colloque de Bordeaux, octobre 1998*, Paris: Honoré Champion.

Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum and Ivan A. Sag (1985). *Generalized Phrase Structure Grammar*. Cambridge: Harvard University Press.

Godard, Danièle (2004). "French negative dependency". In F. Corblin and H. de Swart (eds.), *Handbook of French Semantics*. Stanford (Cal.): CSLI Publications, 351-389.

Hirschbühler, Pierre and Marie Labelle (2003). "Residual tobler-mussafia effects in French dialects". In A. T. Perez-Leroux and Y. Roberge (eds.), *Romance Linguistics: Theory and Acquisition : Selected Papers from the 32nd Linguistic Symposium on Romance Languages*, Amsterdam: John Benjamins, 149-164.

Karttunen, Lauri (2003). "Computing with realizational morphology". In A. Gebulkh (ed.), *Computational Linguistics and Intelligent Text Processing*, Heidelberg: Springer Verlag, 205-216.

Kayne, Richard S. (1972). "Subject inversion in French interrogatives". In J. Casagrande and B. Saciuk, (eds.) *Generative studies in Romance languages*. New York : Newbury House, 70-126.

Kayne, Richard S. (1984). *Connectedness and Binary Branching*. Dordrecht: Foris Publications.

Kilani-Schoch, Marianne and Wolfgang U. Dressler (2005). *Morphologie naturelle et flexion du verbe français*. Tübingen: Gunter Narr Verlag.

Leclère, Christian (1976). "Datifs syntaxiques et datifs éthiques". In J.-C. Chevalier and M. Gross (eds.), *Méthodes en grammaire française*. Paris: Klincksieck, 73-96.

Matthews, P. H. (1972). *Inflectional Morphology. A Theoretical Study Based on Aspects of Latin Verb Conjugation*. Cambridge: Cambridge University Press.

Malouf, Robert (2005). "Disjunctive rule ordering in finite state morphology". Paper presented at the 41st Meeting of the Chicago Linguistics Society.

Miller, Philip (1992). *Clitics and constituents in Phrase Structure Grammar*. New York: Garland.

Miller, Philip and Paola Monachesi (2003). "Les pronoms clitiques dans les langues romanes". In D. Godard, (ed.), *Les langues romanes. Problèmes de la phrase simple.* Paris: CNRS Éditions, 67-123.

Miller, Philip and Ivan A. Sag (1997). "French clitic movement without clitics or movement". *Natural Language and Linguistic Theory*, 15, 573-639.

Monachesi, Paola (1999). *A lexical approach to Italian cliticization*. Stanford (Cal.): CSLI Publications.

Monachesi, Paola (2005). *The Verbal Complex in Romance*. Oxford: Oxford University Press.

Morin, Yves-Charles (1979). "La morphophonologie des pronoms clitiques en français populaire". *Cahiers de linguistique*, 9, 1-36.

Morin, Yves-Charles (1987). "Remarques sur l'organisation de la flexion des verbes français". *ITL Review of Linguistics*, 77-78, 13-91.

Pollard, Carl and Ivan A. Sag (1994). *Head-driven Phrase Structure Grammar*. Chicago and Stanford: the University of Chicago Press and CSLI Publications.

Reichenbach, Hans (1947). *Elements of Symbolic Logic*. New York: The Macmillan Company.

Richter, Frank (2000). *A Mathematical Formalism for Linguistic Theories with an Application in Head-Driven Phrase Structure Grammar*. Doctoral Dissertation. Universität Tübingen.

Stump, Gregory T. (1981). “An infectional approach to French clitics”. In A. Zwicky (ed.), *Ohio State University Working Papers in Linguistics, No 24: Clitics and Ellipsis*. Columbus (Ohio): Ohio State University, 1-54. Ohio State University, Columbus.

Stump, Gregory T. (2001). *Inflectional Morphology. A Theory of Paradigm Structure*. Cambrige: Cambridge University Press.

Verkuyl, Henk, Co Vet, Andrée Borillo, Myriam Bras, Anne Le Draoulec, Arie Molendijk, Henriëtte de Swart, Carl Vetters, and Laure Vieu (2004). “Tense and aspect in sentences”. In F. Corblin and H. de Swart (eds.), *Handbook of French Semantics*. Stanford: CSLI Publications, 233-270.

Zwicky, Arnold. M. (1977). *On clitics*. Indiana University Linguistics Club.

Zwicky, Arnold. M. (1992). “Some choices in the theory of morphology”. In R. D. Levine (ed.), *Formal Grammar. Theory and Implementation*.Oxford: Oxford University Press, 327-371.

Integrating Neoclassical Combining Forms into a Lexeme-Based Morphology

Dany Amiot
Université d'Artois, Arras & Grammatica

Georgette Dal
UMR 8163 STL ("Savoirs, Textes, Langage"), CNRS & Universities of Lille 3 and Lille 1

1. Introduction

The aim of this paper is to account for a part of the so-called combining forms (CFs) – such as *anthrop-, -logue, lud-* in ANTHROPOLOGUE 'anthropologist' and LUDOTHÈQUE 'game library' –, in present-day French in the context of a lexeme-based morphology. We will consider only one type of CFs, the so-called neoclassical CFs (Bauer 1983, Lüdeling & *al*. 2002) or classical CFs (Fradin 2000), which can be characterised by the following four properties:

- Their lexematicity in source languages: in Latin or Greek, they were usually lexemes with grammatical words associated (e.g. LUDUS, from Thomas d'Aquin: "*Ludus est necessarium ad conversationem humanae vitae*").

- The absence of syntactic realisation in the target language: in French (English, German, and so on), neoclassical CFs can only appear as bound constituents of lexemes, without receiving associated grammatical words in the target language. E.g. in French: **J'ai vu un **anthrop(e)** avec un chapeau* (*I saw an anthrop with a hat); **Les enfants aiment les **lud(e)s*** (*Children love luds).

- The kind of vocabulary they serve to form: generally, the complexes in which they appear belong to the learned vocabulary of scientific or technical fields: medicine (LAPAROSCOPIE 'laparoscopy'), biology (BACTÉRIOLOGIE 'bacteriology'), physics (LITHOSPHÈRE 'lithosphere'), technology (CHRONOGRAPHE 'chronograph'), etc.

- The presence of a linking vowel (*o* or *i*) between the two constituents in the phonological context /… C_fC_i…/ where C_f and C_i are consonants in, respectively, final position of the first constituent and initial position of the second constituent: for French, it is generally claimed that *-o-* appears when at least one of the constituents has a Greek origin – *ludothèque, cassettothèque* –, and *i* when at least one has a Latin origin – *omnivore, herbicide*. But *-o-* is more common than *-i-* because it also appears when only one constituent is of Greek origin (such as in LUDOÉDUCATIF 'edutainment', CYTOCHIMIE 'cytochemistry') or with non-neoclassical CFs (as in AFRO-CUBAIN 'Afro-Cuban', ELECTROAIMANT 'electromagnet').

G. Booij, *et al.* (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)*
Fréjus 15-18 September 2005, University of Bologna, 2007.
URL http://mmm.lingue.unibo.it/

This set of properties seems to establish a well-defined class of homogeneous elements, but we show that they are not as homogeneous as they may appear. For example, not all the neoclassical CFs have the same positional constraints: some of them can appear in initial or final position (*anthrop* in **ANTHROP**OPHAGE 'anthropophagite, cannibal' *vs.* AFRICAN**ANTHROPE** 'Fossil of pre-hominian discovered in Eastern Africa'); some only in initial position (*micro-* in **MICRO**ORGANISME 'micro-organism' or **MICRO**AMPERE 'microampere'; and others only in final position (*vore* in OMNI**VORE** 'omnivore' or PUBLI**VORE** 'fond of publicity / publicity hound').

Our examination of previous studies on neoclassical CFs (§ 2.) indicates that such analyses are anything but homogeneous, even though the authors have basically tried to prove the existence of a special category for these elements. Our perspective is slightly different from these approaches: categorization of CFs will be not our main aim; what we will do is to examine whether the basic units of lexeme-based morphology, lexemes and exponents of lexeme construction rules (affixes, non-segmental or supra-segmental phenomena) can account for neoclassical CFs, or whether a new type of unit is required.

The assumptions we defend here are (i) neoclassical CFs are not a homogeneous category, they are not all of the same type and cannot be analysed in the same way; (ii) the notions provided by lexeme-based morphology are sufficient to analyse neoclassical CFs. As a result, we pose the following questions: Do neoclassical CFs belong to the French (English, German, etc.) language or to Latin and/or Greek? Are they part of an international stock common to most Indo-european languages?

2. Previous Approaches

Previous analyses of neoclassical CFs vary according to the criteria taken into account: the bound nature of these elements, whether their position is fixed or not, their semantic nature (lexical or grammatical), their phonological properties, etc. Another important factor is whether these criteria are considered separately or in conjunction. Overall, these analyses have led to four main results: neoclassical CFs are (i) affixes, (ii) roots/stems[1], (iii) roots/stems in some cases and affixes in others, (iv) neither affixes nor roots.

(i) Neoclassical CFs are considered affixes when the only criterion taken into account is their boundness; Williams (1981) or Bauer (1979) propose this kind of analysis. For Bauer (1979) for example, *anglo-, bio-* or *electro-* are prefixes in, respectively, *Anglo-Indian, biochemistry* and *electrocardiogram* and *-crat,* and *-phile* are suffixes in *bureaucrat* and *audiophile*. So, words like *biocrat* or *electrophile* appear "to be made up of a prefix and a suffix, but ha[ve] no root" (*op. cit.*, 509).

[1] In the presentation of the previous approaches, we do not distinguish between the two terms, which are often used as equivalent by the authors we refer to.

Analyses of this sort have been criticized, for example by Scalise (1984: 75-76) and by L. Bauer himself[2], who challenge the idea that a complex could be formed by joining a prefix to a suffix.

When scholars take into account other criteria, such as semantic and positional criteria, the results of the analysis are different, but not necessarily homogeneous:

(ii) Booij (1992) for example considers that neoclassical CFs are not affixes but "non-native roots" following two basic patterns: "root + X" (where X is an existing non-native word, such as *fisica* in *astrofisica*) and "root + root" (e.g. *psychografie*). Plag (2003) adopts a similar analysis: for him, even when a CF has a fixed position, it is never an affix, but a bound root.

(iii) S. Scalise or D. Corbin both make a clear distinction between affixes on the one hand and neoclassical CFs (stems for Scalise 1984, archeoconstituants for Corbin 2001) on the other hand, even for elements with a Greek (or Latin) origin that do not constitute syntactic units in French (English, etc.). In Corbin (2001), the distinction is based on a semantic criterion: for her, affixes have an instructional meaning (they are operators), while archeoconstituants have a descriptive meaning[3]. She uses this criterion to distinguish between *archi-*, *pré-* or *iso-*, affixes, and *phile*, *graph*, *aero*, *gastro*, archeoconstituants. The same criterion allows her to distinguish between two *micro-* in French: the prefix, when *micro-* is joined to a measurement noun (indicating that the measurement is divided by 106, such as in *microseconde*) and the archeoconstituant with an adjectival meaning in other cases (e.g. in *micro-organisme* which refers to a 'very small organism'). The only difference between the prefix and the archeconstituant is the kind of meaning they have: instructional or lexical.
In a similar vein, Iacobini (2004) analyses Italian neoclassical CFs. He, too, considers them heterogeneous, but he distinguishes between three subcategories: neoclassical CFs with a lexical meaning, prefixes with a classical origin, and a third type, which is not labelled, like *-crate*, *-voro* or *-fero*, on the borderline between lexemes and suffixes. Like suffixes, this third type of CFs occur only in a final position and they are productive. Like lexemes, they have a lexical content. For example, in *carnivoro* 'meat eater', *calorifero* 'heater', *-voro* and *-fero* can be regarded as verbs with an argument corresponding to the left

[2] Bauer (1983: 214) notes that the "notion of a prefix and a suffix occurring with no root thus leads to a contradiction".

[3] Corbin (2001: 44): "Cette façon de décrire le sens des affixes et leur intervention dans la construction du sens des unités construites permet de les différencier des autres unités infralexicales entrant dans la construction des mots que sont ce que j'appelle les *archéoconstituants*, c'est-à-dire les constituants empruntés au latin et au grec (ex. *brachy-* 'court', *anthropo-* 'homme', *-cide* 'qui tue'), et les fractoconstituants, c'est-à-dire les représentants tronqués et lexicalisés comme tels d'unités française (ex. *euro-* = *Europe*) dans *eurocorps, eurodéputé, euromissile*, etc.) : archéoconstituants et fractoconstituants ont un sens de nature descriptive et non instructionnelle".
'Describing in this way the meaning of the affixes and their semantic role in the construction of complex units allows them to be distinguished from the other infralexical units involved in word-building and that I call *archeoconstituents*, *i.e.* constituents borrowed from Latin or Greek (e.g. *brachy-* 'short', *anthropo-* 'man', *-cide* 'that kills') and *fractoconstituents*, *i.e.* constituents that have been shortened and lexicalised directly from French lexical units (e.g. *euro-* = *Europe*) in *eurocorps* 'euro-body'*, eurodéputé* 'euro-MP'*, euromissile* '*ibid*', etc.): archeoconstituents and fractoconstituents have a descriptive and not an instructional meaning.'

constituent. The complexes in which they appear resemble the synthetic compounds of Germanic languages (*meat-eater* / *carnivoro*) and the V+N compounds of romance languages (*portacenere* 'ashtray' / *calorifero*). For Iacobini, this type of CFs form compounds, but they do not seem to be roots.

(iv) Warren (1990) proposes a different analysis. While both roots and CFs have a lexical content, CFs are not roots because they do not correspond to a free form. Nor can they be considered affixes, because they belong to open classes. Moreover, they are different from suffixes because they have a lexical meaning, and from prefixes because (a) they don't need "have productive force" (p. 123), (b) they do not have the same phonetic shape, and (c) they correspond to a model-word (e.g. *phyto-* / Gk: *phyton*). So, neoclassical CFs are irreducible to the categories of the other systems of word formation.

When we look at conceptualisations of the relationship between neoclassical compounding and "ordinary" compounding, we see that these are no more homogeneous than the analyses of neoclassical CFs. Two attitudes can be distinguished:

- the differences between the two types of formation are emphasized; cf. e.g. Warren (1990) or Plag (2003). Plag, for example, stresses the fact that neoclassical compounds have formal properties (combinatory and phonological properties, the presence of the linking vowel between the two elements) "that distinguish them from the other types of compounds" (*op. cit.*: 159).

- their similarities are emphasized, in two different ways:

    - Some scholars like Booij (1992) or Scalise (1984) stress the fact that neoclassical compounds share some fundamental properties with other compounds. For example, Booij (1992: 56) claims that the non-native compound "conforms to the general Dutch pattern of compounding in that the second constituent is the head".

    - Another way to bring out similarities in the two types of compounding is to show that neoclassical CFs share properties with other elements, not only with affixes or lexemes, but also with other phenomena such as blending, clipping or secretion. The works of Bauer (1998), Iacobini (2004) or Lüdeling & *al*. (2002) follow this approach. For Bauer (1998: 419-420) for example, "neoclassical compounding is a name for a relatively but not completely arbitrary subdivision of word-creation space and should be read as being a prototype rather than a clear-cut category". For the author, word creation is conceived as a three-dimensional space (whose three parameters are: belonging to the patrimonial lexicon (native *vs.* foreign), formation type (simplex / derivative / compound); degree of shortening) and *neoclassical compound* is used to label a part of this space.

Lüdeling & *al.* (2002: 253) have very different theoretical presuppositions (in their opinion there is no difference between stems and affixes) but their results are very

similar to those of Bauer (at least for German): "no clear-cut principled difference can be found" between neoclassical and native word formation, because neither phonological properties nor differences in the combinability or in the productivity of these elements allow them to be distinguished from native elements. According to these authors, the relationship between neoclassical compounding and native compounding has to be conceived as a continuum.

3. The Tools of Lexematic Morphology

In this section, we will examine whether lexematic morphology is equipped to account for neoclassical CF's. First, we will define successively the notions of lexeme and of affix in the framework of lexematic morphology.

3.1. Lexemes

Since Matthew (1974), the lexeme is generally considered to be an abstract lexical unit that possesses the following properties:

(i) it belongs to an open list and is a member of a major lexical category, *i.e.* it is a noun, a verb or an adjective. Sometimes, the category of adverb is added, especially complex adverbs based on adjectives; e.g., in French, the class of adverbs suffixed by *-ment* (cf. Fradin 2003: 73[4]).

(ii) semantically, a lexeme has a constant and entirely specified meaning (cf. Fradin & Kerleroux forthcoming).

(iii) it has a phonological representation.

While properties (ii) and (iii) do not raise problems, property (i) does: How can a syntactic category be attributed to an element that does not correspond to a grammatical word? It is possible to answer this question when a CF can be used as a base for suffixation, such as *hydr* or *phob* in the adjectives *hydrique* 'hydric'*, phobique* 'phobic': since in French the suffix *-ique* is used to form adjectives on nominal bases (e.g. *colère / colérique* 'anger/quick tempered'*, scène / scénique* 'stage$_{N/A}$'), *hydr* and *phob* can be analysed as nouns. But not all neoclassical CFs serve as a base for derivation, cf. e.g. *micro-* or *-cide*. We will return to this question, § 4.2.

We should also point out, and it will be useful for some of our analysis, that a lexeme can have one or several roots / stems (hence *radicals*), some of them not being visible to the syntax: in French, for example, a verb such as *démontrer* 'demonstrate' has two radicals: *démontr-* and *démonstr-*; the peculiarity of the latter being that it never has syntactic realisations, and only appears in morphologically complex lexemes, such as ***démonstr****ation* 'demonstration' or ***démonstr****atif* 'demonstrative'.

[4] If the suffixation in *-ment* is derivational (see Dal 2007).

3.2. Affix

Within this theoretical framework, an affix is not the same sort of element as a lexeme. Morphological objects are not the result of the concatenation of morphemes but the result of the application of a rule to lexemes. Affixes are therefore the exponents of rules – such as reduplication, apophony, and so on –, which can be characterised as realising phonetically and graphically a semantic function. Thus, "affix" is a simpler way of saying 'exponent of lexeme construction rules' (LCR), and LCRs can be regarded as generalizations between two sets of lexemes, one more complex than the other.

4. Neoclassical CFs in Lexeme-based Morphology

Taking the two basic notions of lexeme and exponent of rule / affix, we investigate whether it is possible to account for the different sorts of neoclassical CFs; to do so, we analyse four different CFs: *lud, anthrop, micro* and *logue*, which will serve to illustrate the different kinds of analyses it is possible to propose within the framework of lexematic morphology.

4.1. "Lud": Radical B of a Lexeme with Multiple Radicals

Lud- has at least two special characteristics: it always appears in initial position and means 'jeu' ('game/play') in the complex lexemes in which it appears: **LUD**IQUE 'relating to a game, ludic, playful', **LUD**OTHÈQUE 'game library', **LUD**ICIEL 'game software'[5].

Following Fradin (2003), who defines the lexeme as a multistratal entity including five types of information (graphemic (G), phonological (F), syntactical (SX), morphological (M) and semantic (S)), each independent of the others, we shall consider that *lud* is one of the radicals of the lexeme JEU. This analysis is supported by the fact that the forms *jeu* and *lud* appear in complementary distribution in complex lexemes: *jeu* appears in final position (ANTI**JEU** 'a game which is the antithesis of what is normally considered a game', INTER**JEU** **'**interplay**'**, CONTRE**-JEU** **'**play-back'**)**, and *lud* in initial position (**LUD**IQUE, **LUD**OTHÈQUE, **LUD**ICIEL).

Our hypothesis is that the choice of *jeu* or *lud* has a phonological motive and is part of a strategy to find the correct form of the output in lexical construction. The radical *jeu*, with its phonological pattern CV, is not a good input, especially because most French adjectival suffixes begin by a vowel (*-ique* [ik] but also *-aire* [ɛR]*, -al* [ɑl]*, -eux* [ø]*, -el* [ɛl], etc.). To avoid forms such as **jeuique* or **jeuthèque*, the suppletive form *lud* (whose phonological pattern is CVC) is preferred.

In this account, *jeu* and *lud* differ only in their graphemic and phonological forms, but, since they belong to the same lexeme, the question of the categorical identity of *lud* (is it a noun or not?) simply does not arise.

[5] We except here PRÉ**LUDE** (from Latin PRAELUDIM), INTER**LUDE** and POST**LUDE**, which form a set, and PRÉ**LUDER** and E**LUDER**, inherited from Latin PRAELUDARE and ELUDARE.

Figure 1 proposes a representation of this analysis. In Figure 1, according to Fradin (2003), "#' indicates the citational form of the radical; the circle prefixing *lud* indicates that this form does not appear independently in syntax; 'res.: init.', that *lud* is reserved for initial position:

JEU		A	B
	(G)	jeu$^{\#}$	°lud$^{\#}$
	(F)	[ʒø]	[lyd]
	(SX)	← c a t : n →	
	(M)		res:init.
	(S)	← j e u ' →	

Figure 1.

The same treatment is proposed for:

(i) other neoclassical CFs such as *pyr*, interpreted as a B radical of the lexeme FEU ([fø]), cf. ALLUME-FEU 'fire-lighter', COUPE-FEU 'firebreak', COUVRE-FEU 'curfew', CONTRE-FEU 'fireback', PARE-FEU 'firebreak / fireguard' *vs* PYROGÈNE 'pyrogenic', PYROGRAVURE 'poker work', PYROMANE 'pyromaniac', PYROTECHNIE 'pyrotechnics', PYROLYSE 'pyrolisis'.

(ii) other types of CFs, termed "fractomorphèmes" by Tournier (1985), "fractoconstituants" by Corbin & Paul (2000), "fractoformants" by Fradin (2000), *i.e.* constituents such as *pétro-* in PETRODOLLAR 'petrodollar / arabodollar': *pétro* is a B radical of the lexeme PÉTROLE.

4.2. "Anthrop": Radical B of a Lexeme with Multiple Radicals or Sole Radical of a Lexeme with no Associated Grammatical Word?

The case of *anthrop* is less straightforward. Though semantically it can be considered a suppletive radical for the lexeme HOMME, the argument based on a positional complementary distribution, used in the case of *lud* vs *jeu*, does not hold, since *anthrop* and *homme* can appear in initial position (ANTHROPOMÉTRIE 'anthropometry', HOMME-SANDWICH 'sandwich man') or in final position (MISANTHROPE 'misanthrope', SURHOMME 'superman'). On the other hand, considering it as the only graphemic and phonological form of a lexeme ANTHROP leads to difficulties with the definition of the lexeme as the result of abstracting away inflectional marking (cf. the definitions of Fradin 2003: 102), since, by definition, a combining form never appears in syntax. Thus, this solution requires at least a revision of the notion of lexeme, such as that proposed by Booij (2002:141) or in Fradin and Kerleroux (forthcoming), who define the lexeme as "l'entité linguistique qui sert de base aux RCL", ("the linguistic entity on which LCRs are based") independently of any syntactic realization. From this point of view, *anthrop*, which can be suffixed by *-ique* (cf. ANTHROPIQUE 'anthropic') is a noun, since *-ique* coins relational adjectives on nominal bases (e.g. COLÉRIQUE$_A$ 'quick-tempered' < COLÈRE$_N$ 'anger'; ALGÉBRIQUE$_A$ 'algebraic' < ALGÈBRE$_N$ 'algebra').

Yet considering *anthrop* as a B radical of a lexeme is, in our opinion, a better solution, since this allows us to deal with another suppletive form of HOMME, *homin-*. We believe that *anthrop-* and *homin-* can be considered, respectively, B and C radical of HOMME, the choice depending largely on the specialist language register and possibly even on the origin of the constituents: *homin-* tends to be used in biology (HOMINAL '*ibid.*', HOMINICOLE 'living in the human body') or in zoology (HOMINIDÉS 'hominidae', HOMINIENS 'hominidae'), in conjunction with constituents of Latin origin, whereas *-anthrop-* appears in a wider variety of fields – geography (ANTHROPOGÉOGRAPHIE 'anthropogeography'), esotericism (ANTHROPOSOPHIE 'anthroposophy'), anthropology (ANTHROPOGÉNÈSE 'anthropogeny'), and so on – in conjunction with constituents of Greek origin. As for HOMME, it appears in native compounds.

The representation of such an analysis for *-anthrop-* and *homin-* appears in fig.2.

HOMME		A	B	C
	(G)	homme#	°anthrop#	°homin#
	(F)	[ɔm]	[ɑ̃ntʀɔp]	[ɔmin]
	(SX)		← c a t : n →	
	(M)		res:gr.	res:lat. res: biology/zoology
	(S)	←homme' (as representative of mankind)→		

Figure 2.

However, we also propose (fig. 3) a representation of the other type of analysis, which consists in considering *-anthrop-* a lexeme without a grammatical word:

ANTHROP	
(G)	°anthrop#
(F)	[ɑ̃ntʀɔp]
(SX)	c a t : n
(M)	
(S)	anthrop'

Figure 3.

4.3. "Micro": The Exponent of an LCR

Regarding *micro-* as the exponent of an LCR, *i.e.* as a prefix, implies that this constituent has undergone a process of grammaticalization as defined by Olsen (2000: 901):

> An originally free word that has entered into a compound can serve as the basis for an entire pattern of like compounds. Once such a pattern takes hold and becomes productive, the original constituent may begin to deviate from its free equivalent in form or meaning and develop into an affix-like element.

In fact, the evolution of *micro-* meets the criteria of grammaticalization (cf. e.g. Heine et al. 1991, Hopper and Traugott 1993, Lehmann 1995):

(i) it originates from a Greek adjective, MIKROS, whose meaning was 'small, short'

(ii) it has lost its syntactic autonomy in French

(iii) its meaning – or, more precisely, the meaning of the LCR it is associated with – is partly different from that of MIKROS in Greek.

Micro- always has a quantitative role:

- Most of the time, it applies to the entirety of what the noun located on its right denotes and indicates that its referent is smaller than the standard, such as in MICROFILM 'microfilm' or MICRO-ORGANISME 'micro-organism', which mean, respectively, 'film / molecule of a very small size'. However, sometimes *micro-* does not apply to the entirety of the referent but only to one of its dimensions, such as in MICROCHIRUGIE 'microsurgery' or MICROÉCONOMIE 'micro-economics'. MICROCHIRUGIE does not denote 'minor surgery' but the 'branch of surgery concerned with very small living structures, done with very small instruments, often under a microscope'. As for MICROÉCONOMIE, it is the 'branch of economics that studies limited economic phenomena'.

- When *micro-* is connected to a noun of measurement, cf. e.g. in MICROFARAD '*ibid'* or MICROSECONDE 'microsecond', it means one millionth of the unit denoted by the noun, FARAD or SECONDE.

Since these dual interpretations are possible, Corbin (1992) proposes a dual analysis of *micro-*: *micro-* is an adjectival archeoconstituant (more or less a bound root) in the first interpretation (cases of MICROFILM or MICROCHIRURGIE) and it is a prefix in the second interpretation (case of MICROSECONDE). Yet Corbin (1992; cf. also 2001) considers that, at a more abstract level, it is fundamentally the same element, whose behaviour and semantic role vary according to the type of noun it is connected to.

Although this analysis is interesting, we believe that it is possible to account for the semantic behaviour of *micro-* in a simpler way, insofar as *micro-* basically behaves in French like *-et*, that is an exponent of an LCR. This is very clear if we compare complexes such as: CLOCHETTE ('small bell') and MICRO-ORGANISME ('very small organism'); RÉFORMETTE ('reform that concerns only limited problems') and MICROÉCONOMIE ('branch of economics that studies limited economic phenomena'). The meanings of the complexes formed by *-et* and *micro-* are very similar. Why then consider *-et* as an affix and not *micro-*? *Micro-* holds, in fact, as an affix:

- in all interpretations it has a quantitative meaning;

- variations in interpretation are due to the meaning of the base: whether it is a noun of measurement or not; whether it is a noun that refers to an entity with a spatial extension or not, etc. It does not differ in this way from a suffix like *-et*.

Possible arguments against such an analysis of *micro-* appear to be based on several factors: our etymological knowledge, the presence of the vowel *-o* at the end of *micro-* and the fact that the lexemes in *micro-* often belong to a learned lexicon, although this is not always the case, cf. *microdécision, micro-ordinateur,* etc. Yet if we want to account for the competence of a run-of-the-mill speaker, it seems better to consider *micro-* as a prefix in modern French.

For French, other constituents can be analysed in the same way: *macro-, mini-, mega-, maxi-*, etc. It is also possible to extend the analysis to foreign constituents such as (Eng.) *free*, (Germ.) *frei,* (Dutch) *vrij*, in, respectively TAX-FREE*,* FEHRLERFREI 'lit. without error, perfect'*,* AUTOVRIJ lit. 'without car, 'no car traffic': these constituents, which were originally adjectives, have also been grammaticalized and hold as exponents of an LCR that has formed adjectives whose meaning is characterised by the absence of what denotes the lexeme-base.

4.4. -logue in the names of specialists: a suffix?

The behaviour of *-log-* ([lɔg]) is complex in modern French. It holds in two distinct ways:

(i) It appears in lexemes such as LOGOPATHIE 'impairment of speech, of language faculty', LOGOMACHIE 'quarrel on words', DIALOGUE 'dialogue' or MISOLOGUE 'somebody who hates argumentation, an enemy of the scientific method'. *-Log-* in this case appears in either initial (**LOGO**PATHIE, **LOGO**MACHIE) or final position (DIA**LOGUE**, MISO**LOGUE**) and means 'speaking, discourse', *i.e.* it has kept the meaning the deverbal noun *logós* had in old Greek.
In modern French, this *-log-* is still used to form learned complex lexemes, but only in initial position: LOGOPATHIE but also LOGOPHASIE 'logophasia', LOGOPHILE 'logophile', LOGOMORPHISME 'logomorphism', LOGOSPHÈRE 'logosphere', etc. All the lexemes in which it appears in final position (DIALOGUE, MISOLOGUE) are borrowed from Greek.

(ii) It also appears, only in final position, in nouns for specialists, such as PSYCHOLOGUE 'psychologist'*,* SISMOLOGUE 'seismologist', DERMATOLOGUE 'dermatologist' EGYPTOLOGUE 'egyptologist', DÉCLINOLOGUE 'specialist in decline'[6], etc. This *-logue* does not mean 'speaking, discourse", appears in the structure of a great number of lexemes and is currently being used to coin many neologisms like DÉCLINOLOGUE, FUTUROLOGUE 'futurologist' or BOBOLOGUE[7].

We propose to consider these two *log* as different: $[lɔg]_1$,(= *-log-* < *logós*) holds more or less like *-anthrop-*, *i.e.* is a B radical of a lexeme, and we do not propose to

[6] This word is a neologism often used in the French media today in an ironic sense to refer to the "specialists / experts" who consider France as a land in decline and who are always negative in their judgments about it.

[7] *Bobo* in *bobologue* is a kind of acronym for *BOurgeois* 'middle-class' *BOhême* 'bohemian / unconventional' and denotes a social group. A *bobologue* is a "specialist" on the people in this social group.

study it in this work. We will focus on $[lɔg]_2$ (= *-logue* < *lógos*), which does not work in the same way as $[lɔg]_1$ but whose status has not yet been established.

In old Greek, *lógos,* which *-logue* comes from, was a bound form connected to the verb *légo* 'speak, say', and appeared in lexemes such as *άστρολόγος* (> fr. ASTROLOGUE 'astrologist'), *μυθολόγος* (> fr. MYTHOLOGUE 'mythologist'); these lexemes received an agentive interpretation (*άστρολόγος* 'a man who speaks about stars', *μυθολόγος* 'somebody who composes myths / legends'). This type of noun was therefore considered to have an NV structure.

In modern French, such an analysis is no longer valid, at least when attempting to account for the competence of a run-of-the-mill speaker, frequently non-grecophone: today, nouns such as PSYCHOLOGUE, DÉCLINOLOGUE or FUTUROLOGUE are not perceived as having a verb in their structure, nor do they have an agentive interpretation. The behaviour of *-logue* seems to be very different from that of constituents like *-cide, -vore, -phile*, *-fere* etc. in, for example, insecticide 'insecticide', publivore 'publicity hound', CINÉPHILE 'cinema lover, film buff' or FLORIFÈRE 'flowering'. In modern French, these constituents keep the verbal interpretation they had in old Greek / Latin and have a predicate-argument relationship with the constituent on their left. They correspond (cf. Iacobini (2004), quoted § 2) to the VN structure of native compounds (*florifère / porte-plume* 'penholder'; *insecticide / tue-mouche* 'fly swatter'). But, whereas the native compounds are in their great majority nouns[8], neoclassical compounds are uniformly A compounds[9].

If *-logue* is not a verb in nouns for specialists, what is its status: the radical of a lexeme or an exponent of an LCR?

The first assumption raises problems because in old Greek, *logos* was a bound form**:** thus, we have to hypothesize that there was a process of reanalysis in which, little by little, in the X*logue* lexemes, the verbal bound form was reinterpreted as a nominal form, whose meaning was first 'speaker', then 'specialist'. From this perspective, the complex lexemes X*logue* have an N_2N_1 structure, in which N_1 is the determined and N_2 the determiner; the complex denoting a hyponym of N_1. Yet a problem still remains: is the noun *-logue* the stem of a lexeme that does not have a syntactic realisation or a B stem of SPÉCIALISTE?

The second assumption is easier to support: from a synchronic point of view, *-logue* is the exponent of an LCR that forms nouns for specialists, such as *-iste* in, for example PIANISTE 'pianist', DENTISTE 'dentist' or CHIMISTE 'chemist'. From a certain point of view, it is possible to say that, in these lexemes, *-iste* also means 'specialist', which does not prevent us from considering it a suffix. Given that *-logue* and *-iste* hold in the same way, we consider *-logue* to be a suffix, like *-iste*, and in figure 4 we propose a representation of this analysis:

	SIMPLE	COMPLEX
(G)	…	…logue
(F)	[…]	[…lɔg]
(SX)	cat:n	cat::n
(S)	object	Specialist of the object denoted by the simple

Figure 4.

[8] On VN compounds in French, cf. Villoing (2002).

[9] For a comparison between VN and NV compounding in French, cf. Namer & Villoing (forthcoming).

6. Conclusion

The analysis proposed in this paper allows us to shed light on certain crucial points:

- Not all neoclassical CFs can be analysed in the same way; on this point, we are in agreement with Iacobini (2004).

- The tools of lexematic morphology are both sufficient and suitable for analysis of neoclassical CFs (at least those examined herein): these elements can be exponents of LCRs (*micro-* and *-logue*) or suppletive stems of a lexeme, used in constrained contexts (*lud-, -anthrop-*).

- Neoclassical CFs, although much is made of their heterolexicality, are easily integrated into the patrimonial lexicon.

However, our analysis also reveals (at least) two weaknesses: (i) in the current state of our knowledge, a distinction cannot be made between allomorphy and suppletion (in analysis of this type, everything is suppletion), and (ii) the problem of the difference in order between neoclassical compounding (determiner + determined) and native compounding (determined + determiner), in French as well as in other romance languages, remains to be solved, especially since we claim that neoclassical CFs are integrated into the patrimonial lexicon. We will deal with (ii) in Amiot and Dal (forthcoming)

There is nevertheless one factor that lexematic morphology cannot account for: the transitional stages that lead from a lexeme to an affix in diachronic evolution, because the lexemes and the exponents of rules (affixes) are conceived to be of a different nature. In the cases of *micro-* and *-logue*, the analysis did not raise any particular problems because these two elements are well grammaticalized in exponents of LCRs, but this is not always the case: e.g. CFs such as *-cide, -vore* or *-phage*, which share some characteristics with lexemes and others with affixes (cf. Iacobini 2004). The theory will certainly need to be refined if problems of this sort are to be solved.

References

Bauer, Laurie (1979). "Against Word-Based Morphology." *Linguistic Inquiry* 10/3, 508-509.

Bauer, Laurie (1983). *English Word-Formation*. Cambridge: Cambridge University Press.

Bauer, Laurie (1998). "Is there a class of neoclassical compounds, and if so, is it productive?" *Linguistics* 36/3, 403-422.

Booij, Geert (1992). "Compounding in Dutch." *Rivista di Linguistica* 4/1, 37-59.

Booij, Geert (2002). *The Morphology of Dutch*. Oxford: Oxford University Press.

Corbin, Danielle (1992). "Hypothèses sur les frontières de la composition nominale." *Cahiers de grammaire* 17, 26-55.

Corbin, Danielle (2001). "Préfixes et suffixes : du sens aux catégories." *Journal of French Language Studies* 11/1, 41-69.

Corbin, Danielle (2004). "French (Indo-european: Romance)." In G. Booij, Ch. Lehmann, J. Mugdan and S. Skopeteas (eds), *Morphologie / Morphology. Ein internationales Handbuch zur Flexion und Wortbildung / An International Handbook on Inflection and Word-Formation*. vol. 2, Berlin / New York: Walter de Gruyter, [Article 121], 1285-1299.

Corbin, Danielle and Paul, Jérôme (2000). "Aperçus sur la créativité morphologique dans la terminologie de la chimie." *La banque des mots* 60, 51-68.

Dal, Georgette (2007). "Les adverbes en *-ment* du français : flexion ou dérivation ?" In N. Hathout and F. Montermini (eds), *Morphologie à Toulouse. Actes du colloque international de morphologie 4e Décembrettes*. München: Lincom Europa, 121-147.

Fradin, Bernard (2000). "Combining Forms, Blends and Related Phenomena." In U. Doleschal and A. Thornton (eds), *Extragrammatical and Marginal Morphology*. München: Lincolm Europa, 11-59.

Fradin, Bernard (2003). *Nouvelles approches en morphologie*. Paris: Presses Universitaires de France.

Fradin, Bernard and Kerleroux, Françoise (forthcoming). "L'identité lexémique." In B. Fradin, F. Kerleroux and M. Plénat (éds.), *Aperçus de morphologie du français*. Paris: Presses Universitaires de Vincennes.

Heine, Bernd, Claudi, Ulrike and Hünnemeyer, Friederike (1991). *Grammaticalization*. Chicago: Chicago University Press.

Hopper, Paul and Traugott, Elizabeth C. (1993). *Grammaticalization*. Cambridge: Cambridge University Press.

Iacobini, Claudio (2004). "Composizione con elementi neoclassici." In M. Grossmann and F. Rainer (eds), *La formazione delle parole in italiano*. Tübingen: Max Niemeyer Verlag, 69-95.

Kerleroux, Françoise (2006). "Les théories morphologiques à la fin du XXe siècle." In S. Auroux, K. Koerner, H.-J. Niederehe and K. Versteegh (eds), *Histoire des sciences du langage*. Berlin / New-York: de Gruyter, [article 242], 2313-2324.

Lehmann, Christian (1995). *Thoughts on Grammaticalization*. München / Newcastle: Lincom Europa.

Lüdeling, Anke, Schmid, Tanja and Kiokpasoglou, Sawwas (2002). "Neoclassical Word Formation in German." *Yearbook of Morphology 2001*, 253-283.

Matthews, Peter H. (1974). *Morphology. An Introduction to the Theory of Word-Structure*. Cambridge: Cambridge University Press.

Namer, Fiammetta and Villoing, Florence (forthcoming). "Have *Cutthroats* to Do with *Tracheotomes*? Distinctive Properties of VN *vs* NV Compounds in French." *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5) Fréjus 15-18 September 2005*. University of Bologna.

Olsen, Susan (2000). "Composition." In G. Booij and *al.* (eds), *Morphologie / Morphology*. Berlin: Mouton de Gruyter, [article 87], 897-916.

Plag, Ingo (2003). *Word-Formation in English*. Cambridge: Cambridge University Press.

Scalise, Sergio (1984). *Generative Morphology*. Dordrecht (Holland) / Cinnaminson (U.S.A.): Foris Publications.

Tournier, Jean (1985). *Introduction descriptive à la lexicogénétique de l'anglais contemporain*. Paris / Genève: Champion / Slatkine.

Villoing, Florence (2002). *Les mots composés $[VN]_{N/A}$ du français: réflexions épistémologiques et propositions d'analyse*. Thèse de doctorat. Université de Paris X-Nanterre.

Warren, Beatrice (1990). "The Importance of Combining Forms." In W. Dressler *et al.* (eds), *Contemporary Morphology*. Berlin / New York: Mouton de Gruyter, 111-132.

Williams, Edwin (1981). "On the Notions 'Lexically Related' and 'Head of a Word'." *Linguistic Inquiry* 12, 245-274.

The Mirror Principle and the Order of Verbal Extensions: Evidence from Pular

Federico Damonte
Università di Padova
federico.damonte@gmail.com

Abstract

As an empirical generalization, the Mirror Principle (MP, Baker, 1985) says that there is a close parallelism between morphology and syntax, but it does not specify according to which general principles are affixes merged into syntactic structure. Following Cinque (1999, 2006), I assume that affixes are merged in a fine-grained hierarchy of functional projections to check the corresponding features. In particular, I will claim that argument structure changing affixes in Pular are merged in a fixed hierarchy of theta-related functional heads and that the complements they introduce are merged in the specifiers of these functional projections. As evidence, I will show that verbal affixes in Pular occur in a fixed order which is not based on semantic scope and that the order of the affixes matches the underlying order of their complements.

0. Introduction: two approaches to morphology[+]

The traditional view on morphology is that word-formation takes place in the lexicon, and that morphological rules are different in nature and operate on different primitive elements than syntactic rules: morphology operates on stems and morphemes to produce words, while syntax operates on words to produce phrases and sentences. This view has been formalized as the *lexical integrity principle* (LI)[1], which makes sure that syntactic rules cannot operate on word parts, so that, for instance, affixes cannot be detached from a word by syntactic rules. An alternative approach has been initiated by the seminal work of Baker (1988): syntax operates on both words and morphemes, and a complex word can be formed by syntactic rules, and more specifically head movement, through *incorporation* of a lexical root to a morpheme. This approach can account for generalizations that cover both morphological and syntactic elements, the best known one being the *mirror principle* (MP, Baker (1985)), which states that morphological derivation reflects syntactic derivation (and viceversa). If the morphological structure of a complex word is derived through head-movement of the lexical root to the heads where the morphemes are base-generated, the MP follows straightforwardly: "the order of morphemes in a complex word reflects the natural syntactic embedding of the heads that correspond to those morphemes" (Baker (2002, 326))[2]. Notice that this approach also captures LI effects, since the result of the incorporation process is still a word-level (i.e a X°) category. This sets the incorporation model apart from other syntactic

[+] Thanks are due to Guglielmo Cinque, the supervisor of my thesis, Damonte (2004), on which this work is based. Heartfelt thanks also to my long-suffering informant, Rabiatou Diallo, who provided judgments on all the examples in this paper and many more. As usual, all mistakes are my own responsibility.

[1] Cfr. Di Sciullo and Williams (1987) and Bresnan and Mchombo (1995), among many others.

[2] Note that this only holds if incorporation is an adjunction process and adjunction is only to the left of the category being adjoined to. In the antisymmetric framework of Kayne (1994), adopted here, no such assumption is necessary, as right-adjunction is impossible.

approaches to morphology, where it is assumed that a complex word is formed of distinct lexical words in syntax, and these only form a word at the phonological level[3]. In this case, LI effects need to receive an independent explanation.

This paper is organized as follows: in section 1 I propose a formulation of the MP based on syntactic features; in section 2 I propose that theta roles are also syntactic features; in section 3 the system of argument structure changing verbal extensions found in Pular, an agglutinative Atlantic language, is briefly introduced; section 4 describes the order of these affixes and section 5 the order of the corresponding complements. In this section a syntactic test is used to show that the *underlying* order of the complements matches the order of verbal extensions. Finally section 6 provides the conclusions.

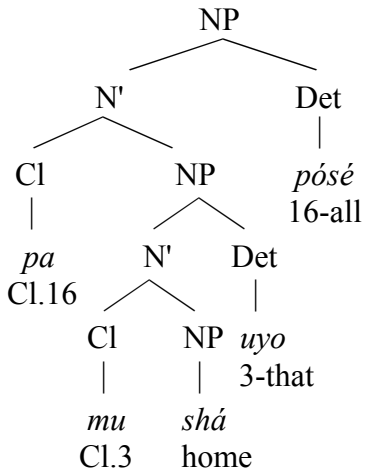
1. The Mirror Principle and syntactic features

LI per se does not prevent the features of a word to be visible to syntax. A noun, for example, has person, number and gender features, and these must be made available to syntax for agreement to take place between the noun and an adjective or verb. The LI though, at least in the strong formulation of Di Sciullo and Williams', does prevent the *relationship* between the features of a word and its internal structure to be relevant to syntax (Di Sciullo and Williams (1987, 49)). The LI is thus incompatible with the syntactic approach to morphology but also with empirical generalizations such as the MP. More generally, the two approaches make quite different predictions about the relationship between syntax and morphology: if the syntactic approach is correct, there should be a systematic parallelism between the order of morphemes and the order of the corresponding phrases, while if the LI (in its strong form) is correct, there should be no such parallelism. A neglected question, and one that will be the focus of this article, is which morphemes and phrases are expected to show this parallelism and which ones are not. A relevant example is discussed by Bresnan and Mchombo (1995, 216 – 217). They quote the absence of agreement between phrasal modifiers and some class prefixes in Chichewa as evidence that class-marked nouns are generated in the lexicon. Their argument is based on Myers' (1987) analysis of nouns with two class prefixes in Shona: he proposes that each prefix heads its own syntactic projection, and can therefore agree with a modifier in its specifier:

(1) *Alternative concord in Shona* (Myers (1987, 104))
a. *Pa-mu-shá uyo p-ósé p-a-káchén-a*
Cl.16-Cl.3-home (Cl.3)that Cl.16-all Cl.16-white
"At that whole white house"

[3] This is the case of Myers' (1987) analysis of class prefixes in Shona, on which see the next section. A similar syntactic analyses of prefixes has been proposed by Julien ((2002).

b.



Myers' analysis correctly captures the agreement properties of the class prefixes that can occur together on the same noun: the external one agrees with the last modifier, and the internal one agrees with the modifier following the noun (the so-called "alternative concord"), as shown in (1b)[4]. Bresnan and Mchombo (1995, 201) then claim that the impossibility of agreement between the internal class prefix and a modifier in Chichewa shows that the class prefixes that can appear noun-internally are merged with the noun in the lexicon, not in the syntax[5]. For the purposes of our discussion, the relevant point is that the authors' argument is based on the assumption that class-agreement morphology represents a case of (necessary) parallelism between morphology and syntax: an affix which (class-)agrees with a modifier belongs to the same projection of the modifier. But class agreement in Bantu languages seems to extend over a wider domain than "agreement" in languages without a class system. In particular, note that the external prefix in (1) is a locative class prefix meaning "at" and "agrees" with the adjectival modifier *all* to the exclusion of the noun. On the basis of standard theories of agreement this is unexpected, and putting the locative prefix and the adjective together in the same functional projection, as proposed by Myers, and accepted by Bresnan and Mchombo, does not shed any light as to why these two elements show class-agreement in the first place. It thus seems that class agreement in Bantu languages spells out different syntactic operations than standard "agreement"[6]. From the viewpoint of the incorporation approach to morphology, not discussed by Bresnan and Mchombo, this means that class-agreement prefixes in Bantu languages are not expected to (uniformly) fall under the MP, as they do not correspond to a single syntactic operation..

Going back to the question of which morphemes and phrases are expected to obey the MP and which ones are not, it thus seems that the question should be rephrased

[4] Note that in this analysis no syntactic mechanism makes sure that the prefix and the noun form a word. In Myers' account, this only happens in the phonological component (Myers (1987, 12)).

[5] Actually, in the case in which the external prefix is one of the locative class-markers, as in the Shona example (1a), alternative concord is possible in Chichewa as well. Bresnan and Mchombo (1995, 201) conclude that locative class prefixes are generated in the syntax in both languages. In Chichewa, if the external prefix is not locative, then all modifiers must agree with the external prefix (Bresnan and Mchombo (1995, 198 – 199)). See also below in this section and the following footnote.

[6] This conclusion holds even if locative class markers are not analyzed as prepositions, an option refuted by Bresnan and Mchombo (1995, 208 – 213). Remember that this type of "outer agreement" between the most external prefix and the modifier(s) of the noun is the only possible one in Chichewa when the external prefix is not locative, see preceding footnote.

as follows: which *syntactic operations* are mirrored in the order of morphemes? An answer to this question is provided by the work of Cinque (1999). On the basis of a vast cross-linguistic survey, he reports that the order of temporal, modal and aspectual (TMA) verbal suffixes in agglutinating languages is fixed and matches (modulo the MP) the order of the corresponding adverbs in non-agglutinative languages. Crucially, this correspondence can be established only with a very fine-grained classification of adverbs and TMA suffixes in narrow semantic classes, such as *past*, *epistemic* or *completive*; a simpler classification in temporal, modal and aspectual modifiers would not have been able to establish such a generalization. Cinque further proposes that this specific semantic classes are represented in the grammar as syntactic features, which in turn project their own projection, such as PastPhrase, EpistemicPhrase, CompletivePhrase and so forth. The crucial point is that all formatives which carry a specific syntactic feature are supposed to be base-generated in the syntactic projection corresponding to that feature, independently of their categorial status. Thus, the PastPhrase projection will host past tense adverbs and suffixes, but also all formatives with that meaning, including prefixes, auxiliaries, functional particles, PPs and so forth. If this is correct, then the MP can be rephrased at a more abstract level as establishing a correspondence between abstract *syntactic features* and *syntactic positions* rather than morphemes and phrases:

(2) *A feature-based Mirror Principle*[7]
All exponents of the same syntactic feature are associated with the same syntactic position

Returning to the "alternative concord" case discussed by Bresnan and Mchombo (1995), this version of the MP does *not* predict that all word-internal class prefixes should be able to have an agreeing modifier, since class agreement in Bantu languages does not seem to depend on a syntactic feature carried by either the modifier or the prefix, as clearly shown by example (1), where a locative prefix meaning "at" agrees in class with the adjective *all*. More generally, a formulation of the MP based on syntactic features allows us to *motivate* the association between morphology and syntax, and therefore to make more precise hypotheses about which morphemes and phrases are associated and which ones are not.

2. Argument structure changing morphology

The feature-based version of the MP in (2) is both weaker and stronger than the original version: it is weaker, in that it does not assume that all syntactic operations are reflected in morphology; and it is also much stronger, in that it covers all specifier and head material carrying a given feature, suffixes and DPs being just one case of a much wider correspondence between syntax and morphology[8]. In this paper I will not explore the

[7] The hypothesis put forward in (2) is implicit in much work on functional projections, and has been proposed explicitly by Cinque (2006, 44), but I remain solely responsible for the way it is formulated and used in this paper.

[8] And between different types of phrases and different types of heads. According to (2), a temporal PP like *nella scorsa settimana* "in the last week" should have the same underlying syntactic position of a DP with the same meaning, such as *la settimana scorsa* "last week". Likewise, the past suffix *-ed* in English

consequences of the MP as defined in (2) for other types of phrases and heads, referring the reader to Cinque (2006) for an extension of the hypothesis to "restructuring" predicates and to Schweikert (2005) for a discussion of adverbial PPs within the same framework. I will instead focus on the consequences of a feature-based MP for the analysis of the empirical domain on which the MP was originally formulated by Baker (1985), namely "grammatical function changing", or "argument structure changing" verbal morphemes. More precisely, I will study whether there is a parallelism between the order of argument structure changing morphemes (so called "verbal extensions") and the complements[9] associated with these morphemes in Pular[10] (Atlantic, Niger Congo), an agglutinating language with a vast array of verbal extensions. The following example illustrates the comitative extension:

(3) *O habh-id-ii e Aboubakar*
he fight-Com-Past with Aboubakar
"He fought with Aboubakar"

As (3) shows, the comitative meaning of the complement of the verb is marked twice: by the affix *-(i)d-* on the verb and by the preposition *e*. A feature-based MP then predicts that both the complement and the affix have the same order with respect to other complements and affixes, respectively. I will argue that even if the surface order of verbal extensions and complements sometimes do not match, an independent syntactic test will show that the *underlying* order of the complements is indeed the same as that of the affixes. Note that the feature which associates the affix and the complement is related to a thematic notion, namely "comitative", as this seems to be the relevant meaning that both the affix and the phrase share. I will therefore assume the following hypothesis:

(4) *Thematic Functional Projections Hypothesis*
The functional structure of the clause contains a fixed hierarchy of labeled functional projections that introduce the complements (both arguments and adjuncts) of the predicate, in different positions according to their thematic relationship[11].

The hypothesis in (4) says that the interpretations usually associated with theta roles (i.e "comitative", "instrumental", "benefactive" etc.) are represented in syntax through

---

should be merged in the same head as the auxiliary *did*.

[9] The phrases associated with a given verbal extension may be an argument of the verb or an adjunct. Since the argument or adjunct status of these phrases will not be discussed in this paper, I will refer to them with the neutral term "complement".

[10] The informant I worked with, Rabiatou Diallo, was born in Lab é, in the Fuuta Jaloo region of the Republic of Guinea. For typographical reasons, I will write the implosive consonants as <bh, dh, yh> and the palatal nasal consonant as <ny>.The language is spoken all over west Africa and has different names in different regions: in eastern dialects the name of the language is *Fulfulde*, in Senegal *Pulaar*, and in Guinea *Pular*. Other names of the language used by European scholars include *Peul*, *Fula* and *Ful*. I will refer to the language as "Pular", and when reference will be made to other dialects, they will be called with the name of the language in the dialect. Diallo (2000) is the only modern grammar of Fuuta Jaloo Pular.

[11] Cfr. Damonte (2004). The hypothesis is actually a formalization of those theories that postulate theta-related functional heads and it is implicit in the works of Cinque (2006) and Schweikert (2005), where functional projections such as "BenefactivePhrase" are proposed.

syntactic features. These features are identical to the TMA ones studied by Cinque (1999): they head their own projections and occupy a fixed position with respect to other functional projections in the structure of the clause, the only difference being that they also introduce a complement. A feature-based MP will in turn force all the lexical formatives that are associated with a specific theta role to be merged in the corresponding "thematic" projection. The complement will thus be base-generated in the specifier and the verbal extension in the head of the relevant thematic projection. The (partial) structure of (3) would then be the following:

(4) $[_{\text{ComitativeP}} [_{\text{PP}}$ *e Aboubakar*] *-id-* $[_{\text{VP}}$ ... *habh-* ...]]

In this paper I will try to show that the formulation of the MP proposed in (2) together with the hypothesis in (4) that there is a fixed hierarchy of theta-related functional projections can account for the close parallelism found between the order of argument structure changing verbal extensions and their corresponding complements in Pular, thus providing clear evidence in favour of the incorporation approach to morphology.

3. Verbal extensions in Pular

"Verbal extension" is the traditional label used for those verbal affixes that "extend" or change the lexical meaning of the verb, as opposed to TMA affixes, which do not change the basic meaning of the verb[12]. For this reason verbal extensions are usually considered derivational affixes, and rules that extend verbs with these affixes are supposed to take place in the lexicon. Pular has a vast array of such affixes, as shown in the following table[13], and there is already a relevant theoretical literature about them[14].

[12] The names of the extensions are the ones used in the literature on Pular and go back at least to Arnott (1970). Note that the term "extension" can be used ambiguously to refer to either the affix (such as "the extension *-(i)t-*") or the meaning (such as "the reversive extension").

[13] Some verbal extensions were omitted from the table, namely several rare unproductive extensions, whose precise meaning is sometimes not easy to establish, on which see Breedveld (1995, 151 – 164); and the "celerative" and "simulative" extensions. The first means "to do X quickly" and the simulative "to pretend to X", where X is the base verb.

[14] The most important works are Sow (1966), Arnott (1970), de Wolf (1986), Gottschligg (1992), Fagerli (1994),Breedveld (1995) and Paster (2005), among others.

Meaning	Affix	Examples
Reversive Repetitive Reflexive Intensive	-(i)t-	*udd-* "close", *udd-it-* "open", *yah-* "go", *yah-it-* "go again" *war-*"kill", *war-t-* "kill oneself" *hel-* "break", *hel-t-* "smash"
Comitative, Completive	-(i)d-	*yah-* "go", *yaa-d-* "go with someone", *heew-* "be full", *heew-d-* "be completely full"
Causative	-(i)n-	*and-* "know", *and-in-* "inform"
Modal, Locative, Instrumental	-(i)r-	*yah- "go", yaa-r*"go in a certain manner" *art-* "return", *art-ir-* "return from some place" *tayh-* "cut", *tayh-ir-* "cut with something"
Benefactive	-an-	*yah-* "go", *yah-an-* "go for someone"
Reciprocal	-indir- -ondir	*and-* "know", *and-indir-* "know each other" *wall-* "help", *wall-ondir-* "help each other"
Distantive	-oy-	*sood-* "buy", *sood-oy-* "go and buy"

Table 1:Pular verbal extensions

In this article I will focus on those meanings which are associated with a complement. These uses are more extensively illustrated in (5) below:

(5) a. *Mi okk-in-ii Buuba baaba maako kaalis* (Causative)
I give-Caus-Past Buuba father his money
"I made Buuba give his father money"

b. *O habh-id-ii e Aboubakar* (Comitative)
he fight-Com-Past with Aboubakar
"He fought with Aboubakar"

c. *O wupp-ir-ii bagi on (e) saabunde* (Instrument)
he wash-Ins-Past cloth Det. with soap
"He washed the cloth with soap"

d. *O art-ir-ii Conakry* (Locative)
hereturn-Loc-Past Conakry
"He returned from Conakry"

e. *Rabiatou def-an-ii Mamadou* (Benefactive)
Rabiatou cook-Ben-Past Mamadou
"Rabiatou has cooked for Mamadou"

Verbal extensions that do not introduce a complement, such as the reflexive and reciprocal extension, as well as passive and middle voice, will not be discussed here, as the point under investigation is whether there is a parallelism in the order of verbal extensions and the order of their complements. Likewise, I will not consider the aspectual meanings of these affixes, even though the fact that many affixes have both an

aspectual and argument-structure changing meaning calls for a principled explanation[15]. Before proceeding to examine the order of verbal extensions, and then compare it to the order of their complements, certain characteristics of these affixes have to be pointed out, and in particular the differences between them and the better known applicative morpheme, with which they are often classified.

First, verbal extensions are associated with far more specific meanings than the applicative morpheme. So for instance, even if we consider only argument-structure changing meanings, the Swahili applied object has several different interpretations that require different verbal extensions in Pular[16]:

(6) *Interpretations of the applied object in Swahili*

a. Benefactive (Ngonyani, (1996, 4))
b. Malefactive
c. Goal
d. Instrumental
e. Motive
f. Locative
g. Reason

Second, verbal extensions are not agreeing markers, as they do not agree with the noun they introduce in number or class. Finally, not all verbal extensions can be analyzed as case markers or transitivizers, since some of them introduce PPs, and not DPs, see the cases of the comitative and instrumental extension illustrated in (5a) and (5c)[17]. This stands in contrast with the applicative morpheme, which only introduces DPs. This conclusion is further confirmed by the fact that some adverbial elements, which do not require case, also trigger the presence of a verbal extension:

(7) *Non hokk-ir-dhaa-mi-nga* (McIntosh (1984, 71))
thus give-Ins-you-me-it
"That's how you gave it to me"

In the preceding example *non* "thus" is clearly not a nominal element, since it lacks the class markers that all nouns bear in this language. If the extension *-ir* were a case assigner, there would be an undischarged case in (14), and the sentence would expected to be ungrammatical, contrary to fact. Furthermore, as both McIntosh (1984, 71) and Breedveld (1995, 178) point out, it is only manner adverbials that require the instrumental manner extension. This seems to show that the extension is sensible only to the interpretation of the complement or modifier it introduces, but not to its case requirements. I will therefore conclude that verbal extensions in Pular are purely argument structure changing devices, and that the semantic modification they cause can be described in terms of specific thematic relationships.

[15] In particular, it is a significant fact that the same combination of aspectual and argument changing meanings is found in verbal affixes in other languages: compare the Italian prefix *co-*, which can have both a comitative (*coprodurre*, "co-produce") and a completive (*cospargere*, "spread on") usage.

[16] The reader is referred to Ngonuani (1996) for extensive exemplification. For a unified analysis of the Swahili applicative and Pular verbal extensions, see Damonte (2004).

[17] See Gottschligg (1992) for an in-depth discussion of case and grammatical relations in Pular.

4. Order of verbal extensions in Pular

Another crucial difference between verbal extensions and the applicative suffix is that verbal extensions can be stacked and quite complex verbal forms can be derived, as in the following example:

(8) *Debbo labbh-in-ir-an-i mo bee buurdhi* (Fagerli (1994, 51))
woman clean-Caus-Ins-Ben-Past him with brush
"The woman cleaned for him with a brush"

The restrictions on the order of these suffixes in agglutinating languages has attracted considerable attention (see Paster (2005), for an overview). The hypotheses put forward to account for the attested orders can be roughly divided in two classes: those that assume that the order is determined by a morphological template, such as the one proposed by Hyman (2002) for the whole Bantu language family; and those that claim that the order of affixes is determined by semantic scope, the most coherent proposal in this sense being that of Rice (2000)[18]. The semantic scope hypothesis seems to better capture those cases in which argument structure changing affixes are not rigidly ordered, and alternative orders are possible. These cases are problematic both for the templatic approach and the hypothesis adopted here, in that affixes are expected to mirror the order of the corresponding functional projections, and functional projections are not supposed to freely recur or be freely ordered. In this section I will then try to show that there is no conclusive empirical support for the semantic scope approach in Pular[19], and more precisely that not all possible scope orderings of verbal extensions are attested, and that when they are, they actually correspond to two different meanings, not two different scope positions of the same meaning. Before proceeding, though, we have to clarify the relationship between the (original) MP and the semantic scope hypothesis, as well as that between the templatic approach and the morphological component.

An often mentioned shortcoming of the MP, as applied to verbal argument structure changing morphology, is the "mobility" of some of these suffixes, meaning that even within a given language they can appear in different combinations. Baker (1988, 373) proposed to account for this variation by postulating different underlying orders of the non-incorporated lexical items[20] corresponding to the different ordering possibilities of the incorporated morphemes. The theory then predicts that different orders of the morphemes have different scope interpretations, with the most external suffixes always scoping over the internal ones, and the individual suffixes retaining their meanings. This in turn has led many to assume that the MP and the semantic scope hypothesis are two sides of the same coin and that one implies the other. This conclusion does not seem to be correct, though: the MP per se does not put any constraint on how the formatives corresponding to incorporated morphemes are ordered in the syntax. The MP only makes sure that the surface order of the morphemes in a

[18] See also Paster (2005), who claims that the order of (consonantal) affixes in Fuuta Tooro Pulaar is largely determined by semantic scope with some cases having a fixed templatic order. Another approach says that the order is determined by phonological constraints, such as the sonority scale. I will not discuss this hypothesis here: see Fagerli (1994, chp. 5) on Pular and Paster (2005) for a general overview.

[19] A conclusion also reached by Fagerli (1994, chp. 3).

[20] Recall that in Baker's (1988) original analysis, verbal argument structure changing affixes correspond to lexical items such as verbs and prepositions, which in turn are merged in lexical, not functional, projections.

complex word will be the "mirror" of the underlying syntactic order, whatever that order is. It is then up to our view of the argument structure changing operations to determine how are the formatives corresponding to surface morphemes merged into syntactic structure. If morphemes are syntactic affixes merged in the heads of the corresponding functional projections, and if these projections are rigidly ordered and are not free to recur or occupy different scope positions[21], then the (original) MP predicts that affixes will be rigidly ordered as well. Thus, the MP approach can be made compatible with templatic hypotheses about the order of affixes, showing that there is no inherent connection between the morphological component and templates or the MP and semantic scope.

This conclusion, though, seems to make the MP unable to deal with those cases in which affixes are not rigidly ordered, as discussed in the preceding paragraph. But while different orderings are a problem for the original MP, they can be successfully accounted for by the feature-based MP proposed in (2). The crucial point is that, given the formulation of the MP in (2), the parallelism between syntactic projections and affixes is indirect, because it is mediated by specific syntactic features. Thus, different orderings of affixes are not per se evidence against (2), *as far as the same affix has different meanings in different positions*[22]. To illustrate, let us examine the extension *-(i)t* in Pular. As shown in table 1, it has several meanings, including repetitive and reversive. The two meanings, though, are associated with two different positions of the affix:

(9) a. *Debbo on sow-it-id-ii bagiiji ndin fow*
woman Det. fold-Rev-Comp-Past cloths Det. all
"The woman unfolded all the cloths"

b. *Debbo on sow-id-it-ii bagiiji ndin fow*
woman Det. fold-Comp-Rep-Past cloths Det. all
*"The woman unfolded all the cloths"
"The woman folded all the clothes again"

As (10) shows, the suffix *-(i)t-* can convey the reversive meaning only if it occurs immediately after the verb root (10a), otherwise it can only be interpreted as the *repetitive* extension (10b). This shows that even if the *suffixes* are not rigidly ordered, the *meanings* they convey are[23]. Note that the same argument holds for phrases: an adverb, for instance, may have different meanings, but then, according to (2), it will only have one possible interpretation in a given syntactic position[24]. On the other hand, the same syntactic feature could be associated with different types of exponents in different contexts. For instance, past tense could be marked by an affix in some contexts and by a root change in others. The MP as formulated in (2) predicts that all these

[21] But Cinque (1999, 91) proposes that some aspectual features are associated with two different positions, with different scope readings.
[22] See Cinque (2006, chp. 7) for extensive exemplification.
[23] Paster (2005) reports that in Fuuta Tooro Pulaar the reversive extension can occur before and after the "comprehensive" extension *-(i)d*. But examples like (9b) in which the reversive follows the comprehensive extension were not accepted by my informant, and are not reported in any source I could check.
[24] For discussion of some Italian examples, see Cinque (2006, 125 – 126). For an extensive argument that this is also the case for adverbial PPs in German, see Schweikert (2005).

exponents should occupy the same syntactic position.

With these clarifications out of the way, let us examine the order of those Pular verbal extensions that introduce a complement. The following examples show that most pair wise combinations of these affixes are indeed rigidly ordered:

(10) Causative < Comitative

a. ??*O goll-in-d-ii-lan e Rabiatou*[25]
He work-Caus-Com-Past-me with Rabiatou
"He made me work with Rabiatou"

b. **O goll-id-in-ii-lan e Rabiatou*

(11) Causative < Instrumental

a. *Mi labbh-in-ir-ii oto on saabunde*
I clean-Caus-Ins-Past car Det. soap
"I cleaned the car with soap"

b. **Mi labbh-ir-in-ii oto on saabunde*

(12) Comitative < Instrumental

a. *Mi def-id-ir-ii e Rabiatou uurere nden*
I cook-Com-Ins-Past with Rabiatou pot Det.
"I cooked together with Rabiatou with a pot"

b. **Mi def-ir-id-ii e Rabiatou uurere nden*

(13) Causative < Benefactive

a. *Debbo on labbh-in-an-ii Mamadou oto on*
woman Det. Clean-Caus-Ben-Past Mamadou car Det.
"The woman cleaned the car for Mamadou"

b. **Debbo on labbh-an-in-ii Mamadou oto on*

(14) Comitative < Benefactive

a. *Mi def-id-an-ii Mamadou teewu on e Rabiatou*
I cook-Com-Ben-Past Mamadou meat Det.with Rabiatou
"I cooked the meat with Rabiatou for Mamadou"

b. **Mi def-an-id-ii Mamadou teewu on e Rabi*

(15) Instrumental < Benefactive

a. *Rabiatou labbh-in-ir-an-ii Mamadou oto on saabunde*
Rabiatou clean-Caus-Ins-Ben-Past Mamadou car Det. soap
"Rabiatou cleaned the car for Mamadou with soap"

b. **Rabiatou labbh-in-an-ir-ii Mamadou oto on saabunde*

[25]For the fully grammatical version of this sentence, see below example (17b) and related discussion.

As examples (10-15) show, there seems to be only one overall order for the verbal extensions examined here, namely (16):

(16) Causative > Comitative > Instrumental > Benefactive

The order in (16) is the same as that reported by Diallo (2000, 150), and is largely compatible with the one given by Arnott (1970) and Fagerli (1994), the only difference being a lower position of the causative extension, which Arnott (1970) and Fagerli (1994) put above the comitative extension. As discussed by Damonte (2004), though, this might be due to the fact that there seem to be two causative heads, with different uses and positions: the lower one introduces the external argument of the verb (cfr. Kratzer (1994)[26]; the higher causative head, instead, introduces a non-argumental causer. Evidence in favour of the hypothesis that the low causative head does not introduce an (external) causer comes from the fact that the causative extension is quite productively used in Pular to add an external argument to an unaccusative predicate, as illustrated by the extended root *labbh-in-* "clean", which is derived from the stative predicate *laabh-* "be clean". Evidence that the causative head which introduces causers lies in a higher position comes from the following examples:

(17) a. ??*O goll-in-d-ii-lan e Rabiatou*
He work-Caus-Com-Past-me with Rabiatou
"He made me work with Rabiatou"

b. *O goll-in-d-in-ii-lan e Rabiatou*
He work-Caus-Com-Caus-Past-me with Rabiatou
"He made me work with Rabiatou"

By hypothesis, the causative extension *-(i)n* in (17a) is in the low causative head and introduces the external argument of the verb, while the causer is introduced by a higher causative head, which is not spelled out in (17a). This analysis seems to be confirmed by the fact that although the variant (17a) is judged acceptable, the preferred form, and the one spontaneously produced by my informant, is (17b). In this example there are two causative extensions, but the sentence does not have a double causative meaning, that is, it does not mean "He made (someone) make me work with Rabiatou"[27]. This can be accounted for if the lower causative extension is assumed to introduce the causee *-lan* "me", while the higher one introduces the causer *o* "he". This analysis predicts that when the base lexical root is a stative predicate, the double causative form should again lack a double causative meaning, since stative predicates do not have external arguments. The hypothesis seems to be borne out:

[26] For concreteness, I will refer to this extension as "Agentiviser".
[27] According to my informant, the unmarked way to express a double causative form in Pular is by means of the auxiliary verb *wadh-* "make" and a causative-derived verb.

(19) *Men hey-dh-in-t-in-ii aadi men ndin*[28]
We new-Dev-Caus-Rep-Caus-Past decision our Det.
"We renovated our decision"

Note that the two different positions of the causative extension do not correspond to two different scope positions: if the interpretation of causative affixes were based on scope a double causative form should have different interpretations with different types of predicates: predicates without an external argument should not trigger a double causative interpretation, while predicates with their own external argument should. This is not the case, which shows that the higher and lower causative extension introduce different types of (agentive) arguments. Finally, note that there seems to be a morphological difference between the two causative extensions as well: as pointed out by Fagerli (1994, 68), verbal roots with a CVVC shape change to CVCC when causativised, but only if the verb is unaccusative[29]: *laabh-* "be clean" > *labbh-in-* "clean", but *dhaan-* "sleep" > *dhaan-in-* "make sleep". While the productivity and interpretation of double causative forms in Pular remains to be fully investigated, the hypothesis proposed here could account for the peculiar variability in the order of the causative and comitative extensions reported by Paster (2005, 179 – 180) and Fagerli (1994, 63 - 65). According to Fagerli, only the order comitative < causative is possible in Adamawa Fulfulde, but is compatible with both scope readings. For Fuuta Tooro Pulaar Paster reports that both orderings of the affixes are possible, with either scope reading. These facts can receive a unified explanation under the current hypothesis, if we assume that both causative heads are activated in (single) causative constructions, but only in Fuuta Jaloo Pular they can be both spelled out at the same time[30]. Different dialects would then vary in the way they spell out the two causative heads: Adamawa Fulfulde seems to be able to spell out only the higher one, Fuuta Jaloo Pular only the lower one[31], and Fuuta Tooro Pulaar both. The variation in the order of the causative and comitative affixes would then be only apparent, in that the underlying structure would be the same for all dialects of the language. Crucially, this variation is not directly linked to semantic scope, as both orderings of affixes allow either reading. Strikingly, the order of the affixes does not seem to be based on semantic scope even when only one scope reading is possible: both Paster and Fagerli provide unambiguous examples, but even in these cases the order of the affixes is fixed in Adamawa Fulfulde (namely *(i)n* < *(i)d*, cfr. Fagerli (1994, 64)), and free in Fuuta Tooro Pulaar (Paster (2005, 180)[32]. It thus seems that the current approach can capture the variability in the

[28] The verbal form in (19) is found in Diallo (2000, 147). Other double causative forms without double causative meanings are reported by Diallo (2000, 147) and Fagerli (1994, 42). Fagerli also reports that the causative extension is the most easily doubled.

[29] Actually, Fagerli says that the verb has to be intransitive, but this seems to be incorrect, in the light of cases like *dhaan- **in-*** "make sleep".

[30] And then only in special circumstances: nearly all the examples with double causative extensions reported in the literature have another extension intervening between the two causative affixes. My informant finds that (17b) sounds odd without the comitative extension.

[31] That is, if only one causative head is spelled out, cfr. (10).

[32] There is one exception, though: when the scope is unambiguously causative < comitative, as in the sentence "Together, they taught him", only the order *(i)n* < *(i)d* is possible in Fuuta Tooro Pulaar, the order expected by semantic scope (Paster (2005, 179)). While this fact needs an explanation, the crucial point is that potentially ambiguous cases seem to allow both scope readings in Fuuta Tooro Pulaar as well, independently of the order of the affixes (Paster (2005, 180)). If affix order were determined by semantic scope, there should no such ambiguity. Note that in Adamawa Fulfulde even in this case the

order of the comitative and causative extensions better than the semantic scope approach, which would presumably conclude that the order of these affixes is scope-based in one dialect (Fuuta Tooro Pulaar) and templatic in another (Adamawa Fulfulde)[33].

To conclude this brief discussion of alternative orders of extensions in Pular, Paster (2005) reports that the order of the comitative and instrumental extensions is scope-based in Fuuta Tooro Pulaar. She provides the following examples:

(20) a. *Mi sok-r-id-ii baafe ɗe coktirgal goɗngal*
I lock-Ins-Compl-Past doors Det. key different
'I locked each of the doors with a different key'

b. *Mi sok-d-ir-ii baafe ɗe coktirgal*
I lock-Compl-Ins-Past doors Det. key
'I locked all of the doors with a key' (the same key)

These examples do not seem to be relevant, though, as the *-(i)d-* affix in this case is actually the aspectual completive extension, not the argument-structure changing comitative. If the extension has the comitative meaning, my informant considers only the order Comitative > Instrumental grammatical in Fuuta Jaloo Pular.

Finally, clear evidence in favour of the feature-based MP approach adopted here comes from the reduplicated extensions. Given the assumption that functional projections cannot freely recur, the original MP bans the same affix to occur twice on the same verbal form. But again, a feature-based MP allows this if the two extensions have different meanings. This is indeed the case with the benefactive and instrumental extensions:

(21) a. *Gujjo wujj-an-an-ii Mamadou Abubakar kaalis*
thief steal-Ben-Ben-Past Mamadou aboubacar money
"The thief stole some money for Mamadou from aboubacar"

b. *O art-ir-ir-ii Conakry oto on*
He return-Ins-Ins-Past Conakry car Det.
"He returned from Conakry by car"

Crucially, the reduplicated extensions in (21) are associated with two different types of complements in each case and *cannot* refer to two benefactive or locative complements, respectively. Note that these examples remain unexplained under a semantic scope approach, since there is no semantic reason why a single predicate should not have two different benefactive complements.

In view of the preceding arguments we must therefore conclude that the order of verbal extensions is not determined by scope, since it appears to be more rigid than what a scope based theory would predict. The overall order of argument-introducing

---

order of the affixes is *(i)d < (i)n* (Fagerli (1994, 64).

[33] This analysis could perhaps be extended to the variable order between the causative and the instrumental extensions reported by Paster (2005, 182) for Fuuta Tooro Pulaar, even if this would force us to assume that the high causative head is higher than the instrumental head, which is not compatible with the fixed order causative < instrumental reported for all other dialects of the language.

verbal extensions in Pular would then be:

(22) Agentiviser < Comitative < Causative < Instrumental < Benefactive

5. Order of complements

While there is a sizable literature on the order of verbal extension in Pular, not much has been written on the order of the complements introduced by these affixes, with the notable exception of Gottschligg (1992). Grammars of the language report a fixed order of these phrases with respect to the object of the verb, and this is confirmed bu my informant:

(23) a. *Rabiatou def-an-ii Mamadou teewu on*
Rabiatou cook-Ben-Past Mamadou meat Det.
"Rabiatou has cooked the meat for Mamadou"

b. **Rabiatou def-an-ii teewu on Mamadou*

(24) a. *Mi def-ir-ii nyiiri ndin e kuddu*
I cook-Ins-Past rice Det with spoon
"I cooked the rice with a spoon"

b. ?**Mi def-ir-ii e kuddu nyiiri ndin*

As examples (23 - 24) show, the benefactive complement occurs immediately after the verb, before the direct object of the verb, while the instrumental complement follows the object, in its unmarked position. If the order of complements is rigid, given our hypothesis (2) we would then expect it to be the mirror of order of the verbal extensions that introduce them. No exhaustive research has been carried out on the order of all possible combinations of complements introduced by verbal extensions in Pular, but there is at least one case in which the unmarked order of the complements is not the mirror image of the order of the affixes:

(25) a. *Mi def-id-ir-ii e Rabiatou uurere nden*
I cook-Com-Ins-Past with Rabiatou pot Det.
"I cooked together with Rabiatou with a pot"

b. ?**Mi def-id-ir-ii uurere nden e Rabiatou*

It seems therefore that the order of complements in Pular does not confirm the hypothesis in (2): there is no parallelism between the order of the affixes and the order of the complements. I would like to argue, though, that the *underlying* order of the complements is the expected one, more precisely, I will propose that that the surface positions of the complements in (25a) does not correspond to the position where they are merged into the syntactic structure, and that the order of these latter positions does indeed correspond to the mirror order of the corresponding verbal extensions. In order to do this, I will use a syntactic test drawn from Ngonyani's (1996) work on Swahili.

This is an ellipsis test on the possibility of deleting the complements of the verb:

(26) *Mi sood-an-ii Mamadou mango e hay Fatou sood-an-ii*
I buy -Ben-Past Mamadou mango and also Fatou buy-Ben-Past
"I bought Mamadou some mango and also Fatou bought
(Mamadou some mango)"

That the construction in (26) is indeed a case of ellipsis is shown in (27):

(27) *Rabiatou ne'-ii paykoy makko koy*
Rabiatou educate-Past children her Det
e hay Fatou ne'-ii ~~paykoy makko koy~~
and also Fatou educate-Past children her Det
a. "Rabiatou educated her children and Fatou also educated her own children"
b. "Rabiatou educated her children and Fatou also educated them (= Rabiatou's)"

As Ngonyani (2000) observes, the possibility of sloppy reading in (27a) shows that the construction in (26) is indeed a case of ellipsis. Now, the ellipsis test in (26) shows that if the verb has two complements, there is an asymmetry: the complement introduced by the benefactive verbal extension can be deleted only together with the object of the verb, but it cannot be deleted alone:

(28) a. *Mi sood-an-ii Mamadou mango ...* Benefactive > Direct Object
I buy -Ben-Past Mamadou mango
"I bought Mamadou some mango ..."

b. *e Fatou sood-an-ii Abou* ~~mango~~.
and Fatou buy-Ben-Past Abou mango
"and Fatou bought Abou (some mango)"

c. **e Fatou sood-an-ii* ~~*Mamadou*~~ *teewu.*
and Fatou buy-Ben-Past Mamadou meat
"and Fatou bought (Mamadou) meat"

d. *e hay Fatou sood-an-ii* ~~*Mamadou mango*~~
and also Fatou buy-Ben-Past Mamadou mango
"and also Fatou bought (Mamadou some mango)

These data can be accounted for if we assume that the benefactive complement is merged in a higher position than the direct object in the functional structure of the clause, as argued by Ngonyani (1996) for Swahili:

(29) a. *sood*$_i$-*an*$_k$-*ii* [BenefactiveP *Mamadou* t$_k$ [VP t$_i$ *mango*]]

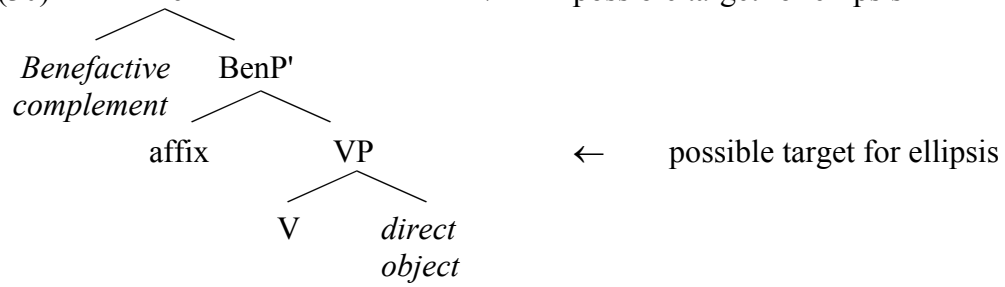
b. *sood*$_i$-*an*$_k$-*ii* [BenefactiveP *Abou* t$_k$ ~~[VP t$_i$ *mango*~~]]

c. **sood*$_i$*-an*$_k$*-ii* ~~[BenefactiveP *Mamadou* t$_k$~~ [VP t$_i$ *teewu*]]

d. *soodi-an*$_k$*-ii* ~~[BenefactiveP *Mamadou* t$_k$ [VP t$_i$ *mango*]]~~

In (29b) the VP, which contains the direct object, has been deleted, and the benefactive complement occupies a higher position. In (29c), on the contrary, the lower direct object is spelled out, while the higher benefactive complement is deleted. In this case deletion has not targeted a node of the structure, leading to ungrammaticality. Finally, in (29d) both the benefactive complement and the direct object are deleted, and the result is grammatical[34]. The relevant nodes that can be grammatically deleted are shown in (30):

(30)



The reader is referred to Ngonyani (2000) for a fuller discussion of this test and its validity for Bantu languages. Granted that this test successfully probes the underlying positions of complements in Pular, let us apply it to other complements introduced by verbal extensions. As expected, the same asymmetry shown in (28) between the benefactive complement and the object of the verb holds for other complements as well:

(31) a. *Mamadou def-in-ii-lan nyiiri ndin...* Causative > Direct Object
Mamadou cook-Caus-Past-me rice Det.
"Mamadou made me cook the rice ..."

b. *e Aboubacar def-in-ii-mo* ~~*nyiiri ndin*~~.
and Aboubacar cook-Caus-Past-him rice Det.
"and Aboubacar made him cook (the rice)"

c. **e Aboubacar def-in-ii-*~~*lan*~~ *teewu on*
and Aboubacar cook-Caus-Past-me meat Det.
"and Aboubacar made (me) cook the meat"

d. *e hay Aboubacar def-in-ii-*~~*lan nyiir ndin*~~
and also Aboubacar cook-Caus-Past-me rice Det.
"and also Aboubacar made cook (me the rice)"

[34]Note that this analysis implies that the verb has moved out of the VP. Since the verb in Pular carries TMA morphology, this seems correct.

(32) a. *Mi def-id-ii mango e Rabiatou ...* Comitative > Direct Object
I cook-Com-Past mango with Rabiatou
"I cooked mango with Rabiatou ..."

b. *e o def-id-ii ~~mango~~ e Fatou*
and he cook-Com-Past mango with Fatou
"and he cooked (mango) with Fatou"

c. **e o def-id-ii nyiiri ~~e Rabiatou~~*
and he cook-Com-Past rice with Rabiatou
"and he cooked rice (with Rabiatou)"

d. *e hay o def-id-ii ~~mango e Rabiatou~~*
and also he cook-Com-Past mango with Rabiatou
"and he also cooked (mango with Rabiatou)"

(33) a. *Mi def-ir-ii nyiiri ndin e kuddu ...* Instrumental >Direct Object
I cook-Ins-Past rice Det with spoon
"I cooked the rice with a spoon ..."

b. *e o def-ir-ii ~~nyiiri ndin~~ e ndihal*
and he cook-Ins-Past rice Det with water
"and he cooked (the rice) with water"

c. **e o def-ir-ii fonnye ~~e kuddu~~*
and he cook-Ins-Past with spoon fonio
"and he cooked fonio (with a spoon)"

d. *e hay o def-ir-ii ~~nyiiri ndin e kuddu~~*
and also he cook-Ins-Past rice Det. with spoon
"and he also cooked (the rice with a spoon)"

Note that the asymmetry between the complement and the direct object of the verb holds even if the complement follows the direct object, as in examples (32 – 33). More importantly, there seems to be an asymmetry between two complements as well:

(34) a. *Mamadou goll-in-d-ii-lan e Fatou...* Causative > Comitative
Mamadou work-Caus-Com-Past-me with Fatou
"Mamadou made me work with Fatou"

b. *e Aboubacar goll-in-d-ii-~~lan~~ e Rabiatou*
and Aboubacar work-Caus-Com-Past-me with Rabiatou
"and Aboubacar made (me) work with Rabiatou"

c. ?**e Aboubacar goll-in-d-ii-mo ~~e Fatou~~*[35]
and Aboubacar work-Caus-Com-Past-him with Fatou
"and Aboubacar made him work (with Fatou)"

[35]The sentence is grammatical without the comitative extension on the verb.

d. *e hay Aboubacar goll-in-d-ii ~~lan e Fatou~~*
and also Aboubacar work-Caus-Com-Past
"and also Aboubacar made work (me with Fatou)"

(35) a *Fatou loot-id-ir-ii oto on e Rabiatou saabunde* Instrumental > Comitative
Fatou wash-Com-Ins-Past car Det. with Rabiatou soap
"Fatou washed the card with Rabiatou with soap ..."

b. *e Abou loot-id-ir-ii ~~oto on e Rabiatou~~fittirgol*
and Abou wash-Com-Ins-Past car Det. with Rabiatou brush
"and Abou washed with a brush (the car, with Rabiatou)"

c. ??*e Abou loot-id-ir-ii ~~oto on~~e Mamadou ~~saabunde~~*
and Abou wash-Com-Ins-Past car Det. with Mamdou soap
"and Abou washed with Mamadou (the car, with soap)"

d. *e hay Abou loot-id-ir-ii ~~oto on e Rabiatou saabunde~~*
Fatou wash-Com-Ins-Past car Det. with Rabiatou soap
"and also Abou washed (the car, with Rabiatou, with soap)"

(36) a. *Mi def-id-an-ii Abubakar e Rabiatou ...* Benefactive>Comitative
I cook-Com-Ben-Past Abubakar with Rabiatou ...
"I cooked with Rabiatou for Aboubakar ..."

b. *e o def-id-an-ii Mamadou ~~e Rabiatou~~*
and he cook-Com-Ben-Past Mamadou with Rabiatou
"and he cooked for Mamadou (with Rabiatou)"

c **e o def-id-an-ii ~~Abubakar~~ e Fatou*
and he cook-Com-Ben-Past for Aboubakar with Fatou
"... and he cooked (for Abubakar) with Fatou"

d. *e hay o def-id-an-ii ~~Abubakar~~ ~~e Rabiatou~~*
and he cook-Com-Ben-Past Abubakar with Rabiatou
"and he also cooked (for Abubakar with Rabiatou)"

(37) a. *Mi def-ir-an-ii Abou kuddu on ...* Benefactive>Instrumental
I cook-Ins-Ben-Past Abou spoon Det.
"I cooked for Abou with the spoon ..."

b. *e o def-ir-an-ii Rabiatou ~~kuddu on~~*.
and he cook-Ins-Ben-Past Rabiatou spoon Det.
"and he cooked for Rabiatou (with the spoon)"

c. **e o def-ir-an-ii ~~Abou~~ uurere nden*
and he cook-Ins-Ben-Past pot Det.
"and he cooked (for Abou) with the pot"

d. *e hay o def-ir-an-ii ~~Abou kuddu on~~*.
and also he cook-Ins-Ben-Past Abou
"and he also cooked (for Abou with the spoon)"

While the judgments are sometimes not so crisp as in examples (28) and (31 – 33)[36], there seems to be a clear asymmetry in the grammaticality of the deletion of the complements introduced by verbal extensions. Benefactives, for instance, can never be deleted in isolation, while a different complement occurring together with a benefactive can (36 –37)[37]. This is unexpected under a theory in which these complements are adverbial modifiers adjoined to the VP. Since adjunction is free and unordered, it should be possible to delete any of the complements in a sentence, contrary to fact. While more research is definitely needed, on the basis of the preliminary results in (34 – 37) it is possible to establish an overall order of the complements in Pular, namely:

(38) Causative < Benefactive < Instrumental < Comitative

This order matches the one established for verbal extensions, and thus show clear evidence in favour of (2).

6. Conclusions

The idea behind this study is that the question "how are syntax and morphology connected?" can be answered (at least in part) by looking at which *features* are visible on both affixes and phrases. Since even strong lexicalist approaches like Di Sciullo and Williams' (1987) have to concede that some lexical features are made visible to syntax, we have then to make specific hypothesis on how syntax can access these features. The hypothesis adopted in this paper is that the connection takes place because features project their own functional projections, and all lexical items carrying that feature must be merged in that projection. This hypothesis has already been successfully applied to TMA modifications by Cinque (1999, 2006) and Schweikert (2005) and I tried to show in this study that argument structure changing affixes are also amenable to be analyzed in this way. I would like to underline that while the high number of functional projections that goes along with this hypothesis may look like a heavy enrichment of the theory, the conceptual tools being used are actually quite few: features and projections are an inevitable part of any theory of phrase structure, and also the minimalist program assumes that lexical items have to check their features in the corresponding projections. The question is then to find which features are syntactically relevant. If the analysis of argument structure changing affixes proposed in this study is correct, thematic relationships are represented in syntax as features. This hypothesis represents a rather dramatic depart from current assumptions about theta-roles, that would rather place them outside syntax proper, but the precise correspondence between the order of verbal

[36]In particular, several pairs of sentences containing both an instrumental and comitative complement, but no direct object (as opposed to example (35)) were judged equally good when either complement was deleted.

[37]Note that this cannot be due to the the fact that the other complement is a DP, as in the case of the instrumental complement in (37), since the comitative complement in (36) is a PP, and it still blocks deletion of the benefactive complement (36c).

extensions and the (underlying) order of their complements in Pular shows that this hypothesis might be correct. If so, future research will have to locate these theta-related projections in the overall hierarchy of functional projections.

References

Arnott David (1970) *The nominal and verbal systems of Fula,* Oxford, Clarendon Press.

Baker, Mark. (1985) "The Mirror Principle and morphosyntactic explanation", *Linguistic Inquiry* 16: pp. 373-415

Baker, Mark. (1988) *Incorporation*, Chicago, Chicago University Press.

Baker, Mark (2002) "Building and merging, not checking: the nonexistence of (Aux)-SVO languages", *Linguistic Inquiry* 33: 321-328.

Breedveld, Johanna (1995) *Form and Meaning in Fulfulde: a Morphophonological Study of Maasinankoore*, Ph.D Thesis, University of Leiden.

Bresnan, Joan & Sam Mchombo (1995) "The lexical integrity principle: evidence from Bantu", *Natural Language and Linguistic Theory*, 13: pp. 181-254.

Damonte, Federico (2004) *The Thematic Field: the syntax of valency-enhancing morphology*, Ph.D Thesis, University of Padua.

Cinque, Guglielmo (1999) *Adverbs and Functional Heads*, Oxford, Oxford University Press.

(2006) *Restructuring and Functional heads. The Cartography of Syntactic Structure, Vol. 4*, Oxford, Oxford University Press.

Diallo, Abdourahmane (2000), *Grammaire descriptive du pular du Fuuta Jaloo (Guinée)*, Frankfurt, Peter Lang.

Di Sciullo, Anna Maria & Edwin Williams (1987) *On the definition of word*, Cambridge (Mass.), The MIT Press.

Fagerli, Ole (1993) *Verbal Derivations in Fulfulde*, Cand. Philol. Thesis, University of Trondheim.

Gottschligg, Peter (1992) *Verbale Valenz und Kasus im Ful*, Vienna, VWGÖ.

Hyman, Larry (2002) "Suffix Ordering in Bantu: A Morphocentric Approach", Yearbook of Morphology, pp.

Julien, Marit (2002) *Syntactic heads and word formation*, Oxford, Oxford University Press.

Kayne, Richard (1994) *The Antisymmetry of Syntax*, Cambridge (Mass.) The MIT Press.

Kratzer, Angelika (1996) "Severing the External Argument from its Verb", in J. Rooryck & L. Zaring (eds.), *Phrase Structure and the Lexicon*, Dordrecht, Kluwer, pp. 109 - 137.

McIntosh, Mary (1984) *Fulfulde Syntax and Verbal Morphology*, Boston, Routledge and Kegan Paul.

Myers, Scott P. (1987), *Tone and the Structure of Words in Shona*, Ph.D. Dissertation, The University Of Massachusetts, Amherst.

Ngonyani, Deogratias (1996) *The Morphosyntax of Applicatives*, Ph.D. Thesis, University of California Los Angeles.

Ngonyani, Deogratias (2000) "The constituent structure of Kindendeule applicatives", in Vicky Carstens and Frederick Parkinson (eds.), *Advances in African Linguistics*, Trenton, Africa World Press.

Paster, Mary (2005) "Pulaar verbal extensions and phonologically driven affix order", in Geert Booij and Jaap van marle (eds), *Yearbook of Morphology 2005*, pp. 155 – 199.

Rice, Keren (2000) *Morpheme Order and Semantic Scope: Word Formation in the Athapaskan Verb*, Cambridge, Cambridge University Press.

Schweikert, Walter (2005) *The Order of Prepositional Phrases in the Structure of the Clause*, Amsterdam, Benjamins.

Sow, Âlfa Ibrâhîm (1966) Remarques sur les Infixes de Derivations dans le Fulfulde du Foûta Djalon (Guinée) , The Journal of West African Languages, Vol. III, n. 1, 13-21.

de Wolf, Paul P. (1986) "Verbal extensions in Noŋaare Fulani: causative, dative, celerative, distantive and simulative", *Afrika und Übersee*, 70, pp. 61 – 72.

Agent Nouns, Productivity and Diachrony: An Analysis of [VN/A]$_{N/A}$ Compounds and *-eur* Derivations in French

Maria Rosenberg
Department of French, Italian and Classical Languages,
Stockholm University
maria.rosenberg@fraita.su.se

Abstract

This study addresses the semantic structure of two types of French agent nouns, [VN/A]$_{N/A}$ compounds (*porte-drapeau*) and *-eur* derivations (*porteur*) from a diachronic perspective. The focus is on established words listed in the *TLFi*. The main objective is to question Dressler's (1986) hierarchy of Agent: Human Agent > Animal Agent > Plant > Impersonal Agent > Instrument > Locative, which proposes that meaning extension diachronically follows this direction. My previous research results show that, synchronically, the Instrument is the most productive/profitable meaning for the [VN/A]$_{N/A}$ formation. The diachronic results of this study likewise fail to confirm the hierarchy. In particular, the semantic structure of the [VN/A]$_{N/A}$ violates the direction implied by the hierarchy. Hence, I reach the conclusion that the different meanings in the polysemy of Agent can be more or less central for different types of agent formations, i.e. both derivations and compounds; the Agent is not necessarily always the primary meaning from which all others originate. This hypothesis will be further examined in future studies that take into account other agentive formations in French and contrast them with similar formations in Swedish.

1. Introduction

The objective is to examine the semantic structure in diachrony of French [VN/A]$_{N/A}$ compounds and *-eur* derivations. These two productive formation types principally give rise to Agents such as *brise-cœur* ('heartbreaker') and Instruments such as *tâteur* ('false key'). The analysis adopts a semasiological perspective, going from form ([VN/A], *-eur*) to meaning (agentive, instrumental, etc.). Synchronically, my previous results show that in Modern French, Instrument is the most productive meaning, in the profitable sense (see 2 below), for the [VN/A]$_{N/A}$ compound, while Agent is the most profitable one for the *-eur* derivations (see Rosenberg (in preparation)). The synchronic results thus contradict Dressler's (1986) Agent hierarchy. In this study, I will question in particular the theoretical relevance of the Agent hierarchy proposed by Dressler (1986) in diachrony as well. In broader terms, I am questioning the seemingly self-evident view that meaning extension is directional, with one primary meaning being the origin of all others.

The internal structure of the [VN/A]$_{N/A}$ compound reflects the syntactic/semantic relation between a verb/predicate and its direct object/internal argument[1]. The second

[1] N.B. Compounds from the 16th century that have an external argument with an Experiencer role, and not an Agent role, are attested, such as *aime-bal* 'ball lover' (Wooldridge 1998:217). There are also some compounds where the second element is instead a subject/external argument to the verb, e.g. *cauchemar* 'nightmare'. These phenomena will be further examined in Rosenberg (in preparation).

G. Booij, *et al.* (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)* Fréjus 15-18 September 2005, University of Bologna, 2007.
URL http://mmm.lingue.unibo.it/

element occasionally has an adverbial function (e.g. *traîne-nuit* 'a bum that hangs out all night'). The $[VN/A]_{N/A}$ compound can be classified as exocentric, i.e. lacking a head; neither the gender nor the number of one of its elements spreads to the compound as a whole. Instead, almost all of these compounds are masculine, with the plural marker in final position. Their meaning is sometimes opaque, such as *accroche-cœur*, 'curl', but more often transparent, such as *ouvre-boîte*, 'tin opener'.

The corpus consists of 1,023 $[VN/A]_{N/A}$ compounds listed in *Le Trésor de la Langue Française informatisé*, *TLFi*. *Le Trésor de la Langue Française* is a dictionary of the French language of the 19th and 20th centuries, and the *TLFi* is the free, online computertised version, consisting of 100,000 words and their history. There are a further 144 *-eur* derivations included in the corpus that are listed in the *TLFi* and based on a verb that is also found in one of the $[VN/A]_{N/A}$ compounds[2]. In all, there are 162 verb types that are part of the 1,023 compounds, but 18 of them do not have an attested *-eur* derivation in the *TLFi*.

Finally, it should be noted that this study makes no difference between nouns and adjectives[3], meaning that an adjective can be classified as Instrument (e.g. *gilet pare-balles*, 'bullet-proof vest') or Agent (e.g. *domestique casseuse*, 'careless servant girl') in the same way as a noun. This decision is based on the fact that there are few adjectives in the corpus, and many of them are nominalised. However, I do not deny the importance of this distinction, which will be taken into account in a future study.

2. Productivity delimiting the object of study

The $[VN/A]_{N/A}$ composition and the *-eur* derivation are two productive formations in Modern French (see e.g. Sleeman & Verheugd (2004:142) regarding *-eur* derivations, and Picone (1992:192-193) regarding $[VN]_N$ compounds). It is therefore important to define the complex notion of productivity. A classic definition is posited by Schultink (1961), who takes productivity to be a morphological phenomenon that occurs unconsciously and gives rise to an infinite number of formations. I have no objections to this definition, but I find Teleman's (1970:18-19) definition theoretically more precise, and it can, in fact, apply to the $[VN/A]_{N/A}$ and the *-eur* type. Given this definition, it is possible to (i) semantically and syntactically describe the elements of a productive formation, and (ii) predict its meaning. Moreover, the two notions of Corbin (1987:42, 177) involving productivity are highly relevant in delimiting my corpus: "*rentabilité*", the number of attested occurrences formed by a process (a quantitative approach), and "*disponibilité*", the capacity to form new words (a qualitative approach). Bauer's definition of productivity makes use of these notions as well:

> 'Productivity' deals with the number of new words that can be coined using a particular morphological process, and is ambiguous between the sense 'availability' [*disponibilité*] and the sense 'profitability' [*rentabilité*]. (Bauer 2001:211)

[2] The grounds for this narrow selection of *-eur* derivations are based on the aim of investigating, in a future study, the potential competition between the semantic structure of the compound and that of the derivation based on the same verb.

[3] Chomsky (1981) assumes both categories possess the feature [+N]: N = [+N, –V] and A = [+N, +V].

These two notions thus delimit the object of study: profitability refers to established words, but availability concerns existing words[4]. Following Bauer (2001:36), a word comes into existence from the moment it is first coined, and becomes established when it takes part in the norm (e.g. enters a dictionary). Although this study focuses on the established words in the *TLFi*, the qualitative aspect will not be neglected.

3. The polysemy of Agent

Fillmore assumes the agentive to be "the case of the typically animate perceived instigator of the action identified by the verb" (1968:24), whereas the instrumental is inanimate. Booij (1986:509), however, proposes an intermediate meaning between the Personal Agent and the Instrument, namely the Impersonal Agent. Gross (1990:84) as well notes that a compound such as *coupe-papier* ('paper cutter') is ambiguous between a machine that cuts paper (i.e. an Agent), and an instrument used for cutting paper. Ulland (1993:20-22), relying on Cruse (1973:21), likewise includes in the concept of Agents various machines and inanimate Agents, as long as they use their proper force to perform the action.

Regarding the French *-eur* derivation, Winther (1975:59) considers context to play a central role in disambiguating the Agent from the Instrument. Villoing (2002:276, note 29) likewise notes that the Agent and the Instrument are conceptually close. She (2002:290) recalls the hypothesis of Corbin & Temple (1994), according to which [VN]$_{N/A}$ compounds are not specified for the opposition person/object, which instead depends on pragmatic, socio-cultural and other factors[5].

Dressler (1986:526), who adheres to the Natural Morphology[6] approach, assumes the polysemous concept of Agent to manifest the hierarchical structure: Agent > Instrument > Locative or Source/Origin. The last two meanings in the hierarchy, Source/Origin, are exemplified by *London-er* or *foreign-er*, following Dressler (1986:525, 527). Furthermore:

> This hierarchy is reflected not only in the relative frequency of these meanings but also in the primacy of agentivity in language acquisition […] and in Breton language decay […], and in diachrony meaning extension seems to go in the same direction […]. (Dressler 1986:526)

Dressler argues that the "agent hierarchy seems to correspond to the animacy hierarchy" (1986:527) (for the animacy hierarchy, see e.g. Comrie (1981)). So a more elaborated structure is yielded when the two hierarchies are combined: Human Agent > Animal Agent > Plant Agent > Impersonal Agent > Instrument > Locative or Source/Origin, which includes several types of Agents, the Impersonal one as well, thus following Booij's (1986) extended scheme. Henceforth, when referring to the Agent hierarchy (of Dressler (1986)), I mean this more elaborate hierarchy.

Sleeman & Verheugd (2004:145-149), in line with Dressler (1986), claim that verb-based *-eur* derivations show a gradual deverbalisation during their nominalisation, which follows the semantic structure: Agent [+human] > Instrument [−human] > Product

[4] Following Dal (2003:13), existing words are seen here as a subgroup of possible words, but they are not necessarily already formed. Also included here are hapaxes (which can result from an error) and neologisms.
[5] N.B. This hypothesis reflects Benveniste's (1975) assumption about the two meanings of the French nouns in -(*t*)*eur*.
[6] For a description of this theoretical approach, see e.g. Dressler (1977, 1986).

[−human]. However, in my opinion, this structure is far too simple. In particular, the features [+/−human] are not sufficient to account for Animal and Plant Agents. According to their structure, these latter meanings should be classified as Instruments. Nor do Sleeman & Verheugd explicitly state what feature they use to separate the Instrument from the Product (a term that they (ibid.) recycle from Winther (1975) and exemplify by *durcisseur d'ongles*).

With a similar purpose, Devos & Taeldeman (2004) posit the Agentivity hypothesis, according to which the formation of deverbal nouns in Dutch and other Germanic and Romance languages is supposed to be governed by the following structure: **Agent Noun**[7] [+agentive] [+animate] > Instrumental$_2$ [+agentive] [–animate] > **Action Noun** [–agentive] [–animate] > Instrumental$_1$ [–agentive] [–animate]. The instrumental$_2$ nouns denote machines and devices (i.e. mostly Impersonal Agents in accordance with the terminology of this study), whereas the instrumental$_1$ nouns refer to substances and products (Devos & Taeldeman 2004:158). This structure is also problematic in my opinion, given its circular use of redundant features that appear simultaneously on two different levels (as a label and as a distinctive feature). The Agents are (needless to say) always [+agentive], but the Instruments can be either [+agentive] or [−agentive] with quite a fuzzy boundary between them. Furthermore, Devos & Taeldeman claim:

> The (just as) frequently used type 'stem + noun' (for example *brise-soleil* 'sun-blind'[8] […]) derives deverbal nouns in the agentive field, both real 'nomina agentis' and instrumental$_2$ nouns. Just like derivations with *-oire*(*e*) they never cross their semantic field. The same holds for equivalent structures in English, for example *killjoy*, *pickpocket* (agentive nouns) and Italian, for example *tagliaborse* 'pickpocket' […] (agentives) next to *tagliacarte* 'letter-opener' […] (instrumental$_2$ nouns). (2004:161-162)

The statement above is strongly contradicted by my synchronic results (cf. 5.1 below) as well as those of the present study (cf. 6 below), since the Instrument is by far the most profitable meaning for this compound type. Devos & Taeldeman remark, however, that the [VN/A]$_{N/A}$ compound "can have a (additional) locative interpretation, too (for example *garde-robe* 'wardrobe')" (2004:165), and they also admit other exceptions to the Agentivity hypothesis, such as the instrumental$_1$ extension of the French nouns in *-eur* and *-ant* (e.g. *autobronzant* 'self tanner').

In light of what is shown above, I emphasise that, according to Devos & Taeldeman (2004), the [+agentive] meanings (agent noun or instrumental$_2$) are historically primary, while the locative [-agentive] meaning is only additional. Consequently, Dressler (1986), Sleeman & Verheugd (2004) and Devos & Taeldeman all claim that semantic extension necessarily follows a diachronic direction: an idea to which I will return and strongly question in this paper (see especially 7 and 8 below).

[7] The bold characters indicate the respective prototypes for the agentive and non agentive poles.
[8] The only attested meaning for *brise-soleil* in the *TLFi* is 'sun protection': "*archit. Dispositif formé soit d'un cadre muni de lamelles métalliques, soit d'éléments en béton avançant sur la façade d'un bâtiment pour protéger des rayons du soleil les baies vitrées*."

4. The classification terms

My semantic classification of the corpus will make use of the following meanings, presented without any hierarchical order: Human Agent – Animal Agent – Plant Agent – Impersonal Agent – Other Agent – Instrument – Locative – Action – Result. Since the Impersonal Agent is taken into account, the distinction between Agent and Instrument based on the features [+/–animate] is not valid. I will instead use Gross (1990:84), who observes the ambiguity between Agent and Instrument, in order to come up with two criteria to distinguish Impersonal Agent from Instrument[9]: (i) a V-*eur*/VN that V/V (det) N → Impersonal Agent; (ii) a V-*eur*/VN with what one V/V (det) N → Instrument. The compound *gobe-sous* ('money-burning machine') is an example of an Impersonal Agent, and *tueuse*[10] ('sword'), of an Instrument. I admit that these two criteria are not watertight, and can surely be improved, but this is not my objective here. Apart form the Impersonal Agent, the Agents need to be divided into different subtypes: Human Agent, e.g. *pleure-misère* ('person that always complains of being out of money'), Agent Animal, e.g. *taille-vent* ('seabird'), and Plant Agent, e.g. *teint-vin* ('lingonberry'). The Locative meaning can be exemplified by *coupe-gorge* ('a dangerous, desert place where one risks getting rubbed or murdered'). Rainer (2005:21) notes that Meyer-Lübke (1890) observed the conceptual ambiguity of recipients between Instruments and Locatives.

In my classification, recipients are included under Locative, e.g. *cuiseur*, 'boiling pot'[11]. The two meanings Source/Origin are not attested, and are thus excluded from my classification. In addition, I have included three meanings in my classification that are not explicitly present in Dressler's (1986) hierarchy: Action, Result and Other Agent. An example of an Action is *frotte-nombril* ('rubbing one's nose against another person's nose'). The Result meaning[12] is the result of the action expressed by the verb, such as *chauffe-double* ('spirits heated twice, the second time with new wine added'). The label Other Agent includes five compounds, e.g. *croque-mitaine* ('imaginary monster') and *trousse-galant* ('disease', especially 'cholera', caused by a micro-organism, hence, agentive).

[9] Sometimes, the explications in *TLFi* are helpful because they use such words as "*appareil*", "*machine*", etc. (= Impersonal Agents) vs. "*instrument*", "*outil*", etc. (= Instrument).

[10] Zwanenburg (1983:138) quotes Dubois (1962:44), claiming that the feminine form *-euse* was one way to refer to the machine/Instrument, in contrast to the masculine *-eur,* which referred to the man/Agent, but as the machine became more frequent in industry, this distinction lost its importance. However, Spence (1990:34) questions Dubois' (1962) claim of an autonomous instrumental *-euse* suffix, given that instrumental *-eur* derivations are much more frequent than the instrumental *-euse* derivation (see also 7.2 below). Moreover, Dressler (1986:526) signals that if Agents, Instruments and Locatives are expressed by different, but parental, affixes, the agentive affixes are the least marked, whereas the locative ones are the most marked. He (ibid.) notes further that, in several languages, the instrumental and locative suffixes are identical to the feminine suffixes (longer than the masculine ones). In line with this reasoning, the French instrumental [VN] formation can, in my opinion, be seen as more marked than the agentive *-eur* derivation, given the more complex structure of the first formation type.

[11] Dressler (1986:526) mentions recipient as one meaning in the Agent polysemy, but he does not indicate whether it is instrumental or locative. N.B. Dressler (1986) mentions no criteria whatsoever as to separate the different meanings from each other.

[12] My term seems to correspond to "effected objects" in Devos & Taeldeman (2004:158), i.e. "what comes into existence by V-*ing* ".

5. Previous studies

5.1. Synchronic results in brief

Synchronically, the French $[VN/A]_{N/A}$ formation matches the following productivity hierarchy: Instrument > Human Agent > Impersonal Agent > Action > Locative > Animal Agent > Plant Agent > Other Agent > Result (Rosenberg in preparation). Clashing with the Agent hierarchy, Instrument has become the core meaning (580, or 45%, of the 1,286 attested meanings are Instruments, whereas only 22 % are Human Agents), a phenomenon that leads to two hypotheses: (i) Agent and Instrument constitute two conceptual categories: **homonymy**, thus a view differing from e.g. Booij (1986) and Dressler (1986)[13]; (ii) Agent and Instrument are one conceptual category: **polysemy.** According to Kastovsky, "we should not talk about the productivity of a morphological pattern as a whole, but rather about the productivity of a morphological-semantic type" (1986:596-597). In Modern French, the *-eur* derivation is more profitable for the Agent meaning, whereas the $[VN/A]_{N/A}$ composition is more profitable for the Instrument. Moreover, Štekauer remarks that productivity should concern whole concept clusters, such as Agent:

> While admitting competition between synonymous suffixes he [Bauer (2001)] – like other morphologists before him – disregards the much wider competition, including various word-formation processes and types. (2003:699)

So there may seem to be a case of competition as far as productivity is concerned between the different meanings in the Agent polysemy of derivations and those of compounds. Hence, this constitutes the hypothesis adopted in this study, and which, in fact, I claim to be the only possible one, given the existence of several polysemous $[VN/A]_{N/A}$ compounds and *-eur* derivations, which cannot be attributed to mere coincidence (or homonymy).

5.2. Previous diachronic studies bearing on the semantics of $[VN/A]_{N/A}$ compounds

Lloyd (1966:158), referring to Heinimann (1949) and Spitzer (1951, 1952), claims that the $[VN/A]_{N/A}$ compound was used from the beginning exclusively as a proper name to refer to Human Agents, often in an ironic and playful way, and that the fact that this compound belonged to the lower classes would explain the low frequency of attested examples before the 11th and 12th centuries. During the transformation of proper names into common nouns, these compounds often retained some original feature, such as the omission of the determiner before the second noun in the compound (Lloyd 1966:259). Yet, Rohlfs (1954:229) gives the example *vinceluna* ('new moon') from the 8th century, and the locative noun *Tenegaudia* as well as the proper name *Zeccadenario* from the 9th century; Kreutzer (1967:184) place names such as *Tosabarba* from 723, and *Tenegaudia* from 739. Otherwise, this type is used especially to denote plants, animals and insects, as well as other objects often in the technical domain (Giurescu 1975:68-69).

Bork's (1990) study examines the Romance compounds from a diachronic perspective, and is interested in such aspects as their functions. Before presenting my own results (in 6 below), I will give those of Bork (1990) regarding Old and Middle French:

[13] In order to avoid fuzzy boundaries, the distinctive features [+/−animate] can be used to separate Agent from Instrument.

Century/Meaning	**11th/12th**	**13th**	**14th**	**15th**	**Total**
Human Agent	5	18	20	28	**71**
Animal, Zoological Agent	1	5	-	1	**7**
Plant Agent	-	2		4	**6**
Instrument	1	6	19	21	**47**
Locative/Place name (+Recipient)	1	-	8	6	**15**
Clothes, Armament, Substance, etc.	2	1	10	8	**21**
Action (+Result), Game	5	7	7	7	**26**
Adverbial	-	1	2	4	**7**
Other[14]	3	7	6	12	**28**
Total	**18**	**47**	**72**	**91**	**228**

Table 1: Bork's (1990) diachronic results for French [VN/A]$_{N/A}$ compounds

I draw attention to the fact that Bork's (1990:94) results only go from the 11th to the 15th century, as opposed to my results below. Moreover, Bork's (1990:71) classification contains some meanings, e.g. the group of clothes, armament and different substances, that I have in most cases classified as Instruments. Note also that adverbials, excluded from my results, are included in Bork's results. Furthermore, there are two other noticeable differences between Bork's (1990) results and my own: firstly, that he does not include Impersonal Agents, and secondly, Bork's (1990) results only give a single meaning for each compound, i.e. the very first attestation, whilst my results also account for the polysemy manifested by single compounds.

To sum up, Bork's (1990) results show that nearly all the different meanings of the Agent polysemy manifested by the [VN/A]$_{N/A}$ compounds are attested from the very beginning. Hence, they can all be seen as primary, and the Agent is not necessarily the one from which all the others originate. Consequently, I claim that Dressler's (1986) Agent hierarchy is not confirmed by Bork's (1990) results. Nevertheless, as shown by table 1 above, Human Agent is the most profitable meaning from a diachronic perspective. However, amongst the first 18 attestations, Action is as profitable as the Human Agent. In addition, table 1 shows that the Instrument is almost as profitable as the Human Agent as early as in the 14th and 15th century.

6. Diachronic results for the polysemy of Agent

Recall that the objective of this diachronic study, ranging from the first attestations in the 12th century to attestations in the 20th century, is to examine the relevance of Dressler's (1986) Agent hierarchy for the semantic structure of [VN/A]$_{N/A}$ compounds and *-eur* derivations. Remember as well that my results are based on information from the *TLFi*[15].

[14] In this group, Bork (1990:83-92) includes among others nine compounds with *passe-*, e.g. *passefelon*, 'someone that is the most unfaithful' (a Human Agent according to my classification), or *passe-mervoille* 'thing … more than marvellous' (in my opinion, an Instrument). Other compounds included here are *picavet* 'sort of faggot' and *rompetout* 'an impediment to action' (two Instruments following my classification). *Passetemps*, 'joy, satisfaction' is classified as an Action by Bork, in line with my classification, whereas Bork classifies *passejoie* 'extreme joy' as Other.

[15] Given the vast quantity of [VN/A]$_{N/A}$ compounds used in my study, a consultation of e.g. *FEW* (Wartburg 1922-) would not have been very fruitful.

Century/ Meaning	**12th**	**13th**	**14th**	**15th**	**16th**	**17th**	**18th**	**19th**	**20th**	**Total**
Instrument		8	8	11	21	29	19	183	227	**506**
Human Ag	1	9[16]	9[17]	7[18]	22[19]	16	20	94[20]	71	**249**
Impers Ag				4	1	1	5	33	92	**136**
Action	1	1	4	2	7	7	6	28	25	**81**
Locative	2[21]	1	2	1	2	3	6	24	21	**62**
Animal Ag		1			4	5	9	19	7	**45**
Plant Ag	1				9	3	7	16	5	**41**
Other Ag					2			2	1	**5**
Result		1			1	1		1	1	**5**
Total	**5**	**21**	**23**	**25**	**69**	**65**	**72**	**400**	**450**	**1130 (1023)**[22]

Table 2: Diachronic results for the [VN/A]$_{N/A}$ compounds

This table shows that already the first attestations manifest polysemy. Hence, my results confirm those of Bork's (1990) study (see 5.2 above), that is to say that Dressler's (1986) hierarchy does not fit the semantic structure of the [VN/A]$_{N/A}$ compounds. They also suggest that the Instrument is just as profitable as the Human Agent for the period from the 13th to the 18th century, but during the 19th and the 20th centuries, the Instrument meaning dominates. Another interesting fact is that the Impersonal Agent appears for the first time in the 15th century, but does not become profitable before the 19th century (see also 6.1 below). Note also that the Result meaning is marginal throughout the period. However, the diachronic results for the semantics of the *-eur* derivations do not strictly follow the results above

Century/ Meaning	**11th**	**12th**	**13th**	**14th**	**15th**	**16th**	**17th**	**18th**	**19th**	**20th**	**Total**
Human Ag	1	18	34	20	4	32	10	1	20	2	**142**
Impers Ag								3	23	16	**42**
Instrument				3	1	2			16	8	**30**
Animal Ag			1			2	3	3	6	1	**16**
Plant Ag								2		1	**3**
Locative								1	2	1	**4**
Result										1	**1**
Action											**0**
Total	**1**	**18**	**35**	**23**	**5**	**36**	**13**	**10**	**67**	**30**	**238 (144)**

Table 3: Diachronic results for the -eur *derivations*

[16] Including two proper names.
[17] Including five proper names.
[18] Including one proper name.
[19] Including one proper name.
[20] Including one proper name.
[21] Including one place name.
[22] In tables 2 and 3, the number in parentheses corresponds to the total number of compounds/derivations in my corpus, whilst the other number corresponds to the total number of meanings.

Table 3 shows that the Agent meaning is by far the most dominant one. The Impersonal Agents and the Instruments do not become profitable until the 19th and 20th century, and the Plant Agent cannot be considered a profitable meaning for the derivations. Furthermore, Action and Result are rare meanings for the *-eur* derivations.

To conclude, the semantic structure of the *-eur* derivations does not deviate as much from Dressler's Agent hierarchy as do the results of the [VN/A]$_{N/A}$ compounds. Nevertheless, in contrast to Dressler's hierarchy, but similar to the results of the compounds in table 2 above, table 3 shows that the Impersonal Agent does not find any high position in the hierarchy until the 19th century. This observation thus calls for a closer look at the distribution between Instrument and Impersonal Agent.

6.1. Distribution between Impersonal Agent and Instrument

In his study of the Spanish suffix *-dor,* Rainer (2004) observes that the Impersonal Agent is not attested before the 19th century, that is to say, before the industrial revolution. Diachronically, Rainer (2005:30) as well as my results in tables 2 and 3 above show that the Impersonal Agent is quite a modern phenomenon, dating from the 19th and 20th century. In other words, Impersonal Agents are not sanctioned diachronically, but synchronically, they are justified, given their high frequency in the semantic structure of agent nouns. My intention in presenting the two tables below is to highlight the impact the introduction of the Impersonal Agent has on the semantic structure of agent nouns, especially in the sense that it favours Agents at the expense of Instruments:

Century/ Meaning	**12th**	**13th**	**14th**	**15th**	**16th**	**17th**	**18th**	**19th**	**20th**	**Total**
Instrument		8	8	11	21	29	19	183	227	506
Impers Ag				4	1	1	5	33	92	136
Total		**8**	**8**	**15**	**22**	**30**	**24**	**216**	**319**	**642**
Instrument=**Total**		**8**	**8**	**15**	**22**	**30**	**24**	**211**	**311**	**629**[23]

Table 4: Diachronic results of the distribution between Instrument and Impersonal Agent for the [VN/A]$_{N/A}$ compounds

Century/ Meaning	**12th**	**13th**	**14th**	**15th**	**16th**	**17th**	**18th**	**19th**	**20th**	**Total**
Impers Ag							3	23	16	42
Instrument			3	1	2			16	8	30
Total			**3**	**1**	**2**		**3**	**39**	**24**	**72**
Instrument=**Total**			**3**	**1**	**2**		**3**	**37**	**22**	**68**

*Table 5: Diachronic results of the distribution between Instrument and Impersonal Agent for the-*eur *derivations*

In comparing the two tables above, the difference between the compounds and the derivations is clear: when the Impersonal Agent is introduced, the first ones are still predominantly instrumental (79 % are Instruments), whereas the second ones become

[23] N.B. The difference as to the total number of the two classification models (Imp Ag or Instr vs. Instr) in tables 4 and 5 depends on the fact that some compounds/derivations are polysemous, thus referring to both an Impersonal Agent and an Instrument.

predominantly agentive (58 % are Impersonal Agents). In my opinion, there is no reason to abandon the non fuzzy criterion based on the distinctive features [+animate] (=Agent) and [–animate] (=Instrument) of Fillmore (1968). Moreover, only this criterion can be diachronically motivated. In future works, I will therefore claim that Impersonal Agents should be excluded from the polysemy of Agent.

6.2. Semantic extension of the polysemous compounds and derivations

In this section, the aim is to follow the semantic extension for each polysemous word in order to see how well it matches with the direction stipulated by Dressler's (1986) hierarchy. In the corpus, there are 96 $[VN/A]_{N/A}$ compounds (9 %) and 72 *-eur* derivations (50 %) that are polysemous, and hence, manifest an internal hierarchy. The fact that the compounds are less polysemous than the derivations is, in my opinion, due to their more complex structure, which gives rise to more specialised and restricted meanings, which are not as easily extendable. Let us first look at the polysemous compounds in table 6.

Table 6 shows that of the 96 polysemous $[VN/A]_{N/A}$ compounds, 57, i.e. 59 %, follow the direction of Dressler's (1986) Agent hierarchy; in contrast, 39 compounds, i.e. 41 %, do not (those marked in grey). The number of violations here is considerable, and I claim therefore that Dressler's hierarchy does not seem valid in accounting for the semantic extension of the polysemous compounds. Moreover, in most cases, the first attested meaning is instrumental. For example, the compound *casse-gueule* manifests a semantic extension clashing with the direction of Dressler's hierarchy. The first attestation is Locative ('a dark and dangerous place'), the second is Instrument ('spirits'), and the third is Action ('a risky mission').

[VN/A]$_{N/A}$ Compounds	**Hum Ag**	**Anim Ag**	**Plant Ag**	**Imp Ag**	**Instr**	**Loc**	**Act**	**Res**	
abat-jour		18th			17th				
accroche-cœur	1635			1874	1842				
allume-feu				19th	19th				
attrape-mouche(s)		1752	1700		19th				
boit-tout	19th				18th				
bouche-trou	17th				18th				
boute-feu	15th				14th				
boute-hors					17th		14th		
bouteroue	13th (PrN)				17th				
brise-glace(s)				20th	18th				
brise-mèche				20th	20th				
casse-cou	1798					1718	1808		
casse-croûte					1803		1898		
casse-cul	20th						18th		
casse-gueule					1866	1808	1914		
casse-mèche				20th	20th				
casse-museau					15th		17th		
casse-noisettes	19th				17th				
casse-noix		16th			17th				
casse-pierre			19th		20th				
casse-tête					17th		18th		
cauchemar	18th						14th		
chasse-marée	13th			15th					
chasse-neige				19th	19th		20th		
chausse-trap(p)e			12th		13th				
claque-dent	15th (PrN)					19th			
cloche-pied	15th				19th				
coupe-choux	14th (PrN)				19th				
coupe-racines				19th	19th				
coupe-tête	14th						17th		
engoulevent	13th (PrN)	18th							
essuie-glace(s)				19th	19th				
fouille-merde	20th	19th							
gagne-pain	13th						16th		
garde-côte	12th			15th					
garde-note(s)	16th				19th				
garde-robes						12th[24]		16th	
gobe-mouche(s)	16th	17th	18th						
gratte-ciel	1915				1911				
happe-lopin	19th	19th							
happelourde	1532				1564				
hausse-pied		19th			14th				
hoche-pied		19th			14th				
lève-nez	19th				20th				
mange-tout	19th		18th	16th					
marchepied					13th	14th			
monte-lettres				20th	20th				
monte-livres				20th	20th				
monte-paquets				20th	20th				

[24] *Garde-robe* has in fact several locative meanings: (i) 1190 'room where one keeps the clothes'; (ii) 13th century 'chest' or 'closet'; (iii) 1314 'toilet'. Note further the metonymical relation CONTAINER for CONTENTS for the meaning (iv) 1540 'all the clothes of someone' (cf. Panther & Thornburg (2002:283)).

[VN/A]$_{N/A}$ Compounds	**Hum Ag**	**Anim Ag**	**Plant Ag**	**Imp Ag**	**Instr**	**Loc**	**Act**	**Res**	
passe-colère	20th				20th				
passe-lacet	1880				1827				
passe-pied						17th	16th		
pince-fesses						1949	1931		
pince-maille	14th (PrN)				19th				
pique-bœuf	16th	18th			17th				
porte-aiguille					18th	19th			
porte-bagages					19th	20th			
porte-bonheur	18th				19th				
porte-bouteilles	16th				18th	19th			
porte-chance	20th				20th				
porte-cigares					1845	1837			
porte-cigarettes					1886	1857			
porte-clef(s)	1571				1581				
porte-crosse	17th				19th				
porte-épée	1552				1581				
porte-étendard	17th				18th				
porte-fort	1951						1936		
porte-glaive	18th				20th				
porte-guigne	20th				20th				
porte-jupe	17th				20th				
porte-malheur	17th	18th							
portemanteau	1507				1640	1547			
porte-montre					1975	1908			
porte-parole	16th				19th				
porte-plume		19th			18th				
porte-queue	15th	18th							
porte-respect	18th				17th				
porte-veine	20th				20th				
porte-voix	18th				17th				
rince-bouches						19th	20th		
rince-bouteilles				20th	19th				
serre-file	17th			19th					
serre-papier(s)					1766	1720			
souffre-douleur	1662	1678			1607				
taille-douce							16th	17th	
taille-vent		19th			20th				
tape-cul					15th		19th		
tâte-vin	15th				16th				
tire-ligne	19th				17th				
tirelire	17th					13th			
tourne-bride						18th	17th		
tourne-broche	15th	1678			1663				
tranche-caillé				20th	20th				
trotte-menu	13th	17th							
tue-mouche(s)			1823		1872				
vide-bouteille	16th				19th	18th			
Total									**96**

Table 6: Diachronic results for the polysemous [VN/A]$_{N/A}$ compounds

The polysemous -*eur* derivations are shown in table 7 below:

***-eur* Derivations**	**Hum Ag**	**Anim Ag**	**Plant Ag**	**Imp Ag**	**Instr**	**Loc**	**Act**	**Res**	
abaisseur	1564				19th				
accrocheur	1635			1874	1842				
allumeur	1374			1890					
arracheur	13th				1866				
batteur	1204				1877				
becqueteur	1883	1882							
boucheur	1550			19th					
branleur	1690			1930					
briseur	12th			20th					
brûleur	13th			1853					
chauffeur	1680					1830			
chercheur	1538			20th					
colleur	1544			20th					
compteur	1213			1752					
coucheur	1534			20th					
coupeur	1230	1805		20th					
cuiseur	1270			1928		1929			
écorcheur	13th				19th				
écraseur	1571				1857				
essuyeur	19th				1377				
étouffeur	19th	1775							
forceur	1507	20th							
fouilleur	1511				20th				
frotteur	1372				20th				
fumeur	1690					1868			
gobeur	1524	1679							
gratteur	14th			1829					
guérisseur	14th			19th					
hacheur	14th[25]			20th					
haleur	1680				20th				
lanceur	13th			1864					
laveur	1390			1867					
marcheur	1500	1791		1773					
mireur	1872			1840					
mouilleur	1576			1831					
ouvreur	1210			1877					
pêcheur	1140				19th				
peigneur	1243			1812					
peleur	20e				1861				
perceur	15th			1894					
peseur	1250			1949					
piqueur	1387				19th				
pisseur	1464				1963				
pleureur	1050		20th		19th				

[25] The Human Agent meaning is only attested in Littré (1957): 'chaser' or 'engraver' in the 14th century. This is the only instance in this study where a dictionary other than the *TLFi* has been consulted.

-eur Derivations	Hum Ag	Anim Ag	Plant Ag	Imp Ag	Instr	Loc	Act	Res	
porteur	12th	1680		1869					
pousseur	1690			1959	1903				
presseur	1384				19th				
protecteur[26]	1234				1869				
purgeur	1531			1869					
rabatteur	1585	1850			1904				
racleur	1576				1896				
ramasseur	1509			1867					
ratisseur	1532				1530				
releveur	1200			1877	1865				
repousseur	1611			20th					
rinceur	1561			1904					
rogneur	1354			1875					
rongeur	1530	1800			1314				
rouleur	1284	1734		1725					
sauteur	1380	1526			1875	1828			
sécheur	1611			1874					
suceur	1564	1809		1948	1964				
suiveur	1200			20th					
tailleur	1165							1905	
tâteur	1372			1961	1833				
tordeur	14th	1803		1872					
traceur	1558			1877					
traîneur	1330	1694							
trancheur	1208			19th					
trompeur	1390				1557				
trotteur	15th	1215		1894					
tueur	1200				14th				
Total									**72**

Table 7: Diachronic results for the polysemous -eur *derivations*

It follows from table 7 that of the 72 polysemous derivations, 16, i.e. 22%, have a semantic extension that violates the direction of Dressler's Agent hierarchy (these are marked in grey). In contrast, the extension of 56 derivations, i.e. 78 %, matches with the direction of the hierarchy. *Rongeur* is one of the rare derivations with an instrumental meaning attested first ('rasp'); the second attested meaning is Human Agent, followed by Animal Agent. Once again, we see that the *-eur* derivations follow the direction of Dressler's hierarchy much more closely than do the compounds, presumably because they are predominantly agentive. Following Rainer (2005:30), the suffix *-dor* seems to manifest a complementary distribution of agentive vs. instrumental at least in Modern Spansih. However, my results do not favour the account of complementary distribution as to the *-eur* derivations, since there is only one derivation in the corpus, *rompeur*, that has only the instrumental meaning attested. Consequently, I cannot see how the complementary distribution could be a sufficient reason to reject the idea of semantic extension.

[26] N.B. The suffix here is *-teur*. In the same way as *amateur*, this derivation differs from the other *-eur* derivations in the corpus. Nevertheless, I decided to include these sorts of derivations as well.

To sum up the diachronic results, the Agent hierarchy as proposed by Dressler (1986) finds no support where the two French agentive formations studied here are concerned. This hierarchy seems to be just a model, logical and theoretically motivated, but that does not reflect the structure of the Agent polysemy such as it is expressed by different word formations types. It is the semantic structure of the [VN/A]$_{N/A}$ compound, in particular, that prompts me to reach this conclusion.

7. Explanations of semantic extension

7.1. Three explanations of a diachronic nature

Rainer (2005:26, 29) observes that given a diachronic or a synchronic perspective, semantic extension will be explained differently; it is therefore dangerous to study semantic extension by extrapolating diachronic evidence from synchronic evidence. He (2005:26-29) identifies three diachronic origins of the instrumental and locative uses of the deverbal agentive suffixes in Romance languages that have nothing to do with semantics or cognition: ellipsis, homonymisation and borrowing.

The ellipsis explanation is proposed by Darmesteter (1972) [1877], according to Rainer (2005:28), and stipulates that the instrumental use of the *-eur* derivation, dating from the 19th century, is caused by an ellipsis of the head noun in the NP[27]. I note that Winther (1975:78) too assumes the *-eur* suffix to be fundamentally adjectival. Given that the instrumental use precedes the adjectival one, Spence (1990:33) thinks that it is absurd to consider the adjectival nominalization to be the only source of the instrumental noun in *-eur*. In light of my results, it is obvious that a vast majority of the first instrumental attestations do not issue from ellipsis (e.g. *essuyeur* ('towel'), from the 14th century). As to the instrumental use of the [VN/A]$_{N/A}$ compound, often prior to the agentive one, it can rarely be assigned an elliptic origin.

As to the explanation of homonymisation, i.e. a collision of two different suffixes mediated by a phonetic change, it surely cannot account for the [VN/A]$_{N/A}$ compound, even though it might have some relevance for the *-eur* derivation (cf. note 5 above)[28].

The borrowing explanation is, according to Rainer (2005:28), already noted by Meyer-Lübke (1966:§66) [1921] for some instrumental and locative formations in Old French. Rainer (2005:32) also points to the possibility that Romance influence could be the cause of the non agentive uses of Germanic *-er* derivations (cf. *counter* vs. *compteur*). Agent nouns in Proto-Germanic seem to have lacked the non agentive use, but within a few centuries during the Middle Ages, all European languages seem to have acquired such a use (Rainer 2005:33). As to my results, it is clear that this last explanation can hardly account for the polysemy of the [VN/A]$_{N/A}$ compound, which is already present in the first attestations. Moreover, the verification or not of one of these explanations does not constitute a central objective of this study. I will instead go for a semantic/cognitive explanation, which has the power to account for the synchronic results as well.

[27] There are two possibilities: *appareil* + A in *-eur* → *releveur* and *machine* + A in *-euse* → *peleuse*.
[28] Rainer (2005:28) notes that in Provencal, in Catalan and in some Italian dialects, the instrumental Latin suffix, *-torium*, has become identical to the agentive Latin suffix, *-torem*, which explains the resultant polysemy. However, this polysemy is also found in Spanish, but here the two suffixes are still separate (-d*ero* and *-dor*), thus contradicting the homonymisation explanation (Rainer:ibid.).

7.2. Semantic/cognitive explanations

Apart from those discussed in the previous section, Rainer (2005:22-23) mentions some semantic/cognitive explanations for the extension of the Agent polysemy. According to Panagl (1977), the instrumental use is the result of an extension of the agentive use by metaphor, an idea expressed early on by Meyer-Lübke (1890), or by metonymy. This explanation, however, is problematic since a primary agentive formation does not always exist (Panagl (1977:13) cited by Rainer (2005:22)). Thus, in order to keep the explanation involving metaphor or metonymy, Rainer firstly posits reinterpretation of the agentive use as instrumental mediated by metaphor or metonymy (Rainer 2005:22-23) as one possible mechanism. The metaphorical or metonymical extensions explain the diachronic origin of the instrumental use, which can later, by reinterpretation, become the basis for other instrumental uses. Another possibility is the mechanism of approximation: the instrumental use can appear by metaphor or metonymy in using an agentive pattern (not an individual word as in the case of reinterpretation) in an approximate way (Rainer 2005:23). According to Rainer (2005:26), metaphoric approximation is the most probable explanation. He (2005:24-25) finds support for this in the first instrumental attestations of the Spanish suffix *-dor*, since several of them lack a primary agentive formation; instead, it seems to be a question of complementary distribution. However, my results for the *-eur* derivations do not show much evidence of complementary distribution (cf. 6.2 above). Nor can the Agent always be considered as primary; 91 % of the $[VN/A]_{N/A}$ compounds are monosemous and their meaning is for the most part non agentive. In other words, none of these explanations is satisfactory according to my results.

Blocking, i.e. competing instrumental patterns that block the instrumental extension of a certain word formation process, constitutes another explanation (Rainer 2005:30). Using this explanation, Spence (1990) puts forward the hypothesis that the instrumental extension of *-eur* in French depends on a loss of productivity of the (competing) instrumental suffixes *-oir* and *-oire* (< Latin *-orium*/*-oriam*)[29]. Spence (1990) questions the hypothesis of Dubois (1962) that *-eur* replaced *-oir* during industrialisation, when machine replaced man, because he claims that the instrumental use of *-eur* is attested even before the industrial revolution[30]. Beard (1990:118), however, questions the whole blocking explanation, because in Serbo-Croatian, the existence of a productive instrumental suffix does not block the instrumental use of another agentive suffix. Nor do my results lend support to this hypothesis; given the strong productivity of the Instrument meaning of the $[VN/A]_{N/A}$ compound, that certainly does not block the instrumental use of the *-eur* derivation. I will thus once again emphasise the importance of taking into account the competition between different word formation processes, i.e. both derivations and compounds, and not just limiting the study to different derivation types.

Another explanation mentioned by Rainer (2005:31-33) is the cognitive approach by Ryder (1991), who studies the passage from Agent to Instrument in terms of prototype analysis. Diachronically, this explanation is just as problematic as that of Booij (1986) – Rainer (2005:30) criticises Booij's (1986) semantic extension scheme (Pers Ag > Impers Ag > Instr) predicting that the agentive interpretation, even if not attested, should always be possible – since the first instrumental attestations do not correspond to Impersonal Agents; that would constitute the natural transition between Agent and Instrument (Rainer

[29] Furthermore, Spence (1990:29) mentions that some *-oir* derivations express the Locative rather than the Instrument, and in English, the *-ory* suffix, with the same Latin origin, forms Locative nouns (ibid.:32).
[30] N.B. This is the same criticism as that put forward by Spence (1990) concerning the ellipsis explanation (cf. 7.1 above).

2005:31). Furthermore, Ryder (1991:299), referring to Kastovsky (1971), claims that, in Old English, the *-er* suffix is restricted to Human Agents (e.g. *bæcere*, *writere*), but that:

> Extensions to other referent types found in modern *-er* [here and below, my italics] forms are the result of shifts in construal of the defining episode, with resultant changes in the importance of each of the characteristics of the referents of originally agentive *-er* forms. (1991:303)

This quotation clearly indicates that Ryder considers Agent to be the primary meaning.

To summarise, my objection to all the explanations given here lies precisely in the fact that they all view the Agent meaning as primary, a view that has been contradicted by my results, especially those for the [VN/A]$_{N/A}$ compound.

8. Conclusion

This study, which addresses the Agent polysemy in diachrony as manifested by French [VN/A]$_{N/A}$ compounds and *-eur* derivations, questions the relevance of Dressler's (1986) Agent hierarchy. My results show that the direction indicated by this hierarchy is not confirmed. In particular, the semantic structure of the [VN/A]$_{N/A}$ compound, with the Instrument as its central, and most profitable, meaning, goes against the direction of this hierarchy. In contrast, the semantic structure of *-eur* derivations, with the Agent as the most profitable meaning, follows the hierarchy's direction quite closely, except for the low profitability of the Animal and Plant Agents. This difference between compounds, which adhere to the instrumental pole, and derivations, which adhere to the agentive pole, has a great impact on the distribution of Instruments vs. Impersonal Agents as well: the first favour the Instruments, while the second favour the Impersonal Agents. Moreover, I would like to emphasise that as a consequence of the introduction of Impersonal Agents in the polysemy of Agent, the total number of agentive meanings increases, hollowing out the instrumental ones. As to the semantic extension of polysemous words in my corpus, it does not strictly follow the direction of Dressler's (1986) Agent hierarchy. Once again, the extension of the compounds in particular goes against the hierarchy, given that almost all of the different meanings of the Agent polysemy can be classified as primary. Furthermore, after having discussed several diachronic and semantic/cognitive explanations, I reach the conclusion that none of them is satisfactory in accounting for my results, because each departs from the hypothesis that the Agent is necessarily the one and only primary meaning.

Instead, I claim that all the different meanings in the Agent polysemy can be qualified as primary, but that they are more or less central, and more or less profitable, for different word formation types. More precisely, I object to the idea of a (diachronic) passage from Agent to Instrument[31]. My future studies will further explore the agentive field in French by extending it to two other Agentive types, i.e. the *-oir*(*e*) derivations (one hypothesis is that this type is predominantly Locative), and those with *-ant*, and will also take a contrastive perspective with Swedish, a Germanic language.

[31] Another of my hypotheses, in line with this reasoning, is that the figurative meaning can be primary to the literal one. This is the case for [VN/A]$_{N/A}$ compounds whose first or only attested meaning is often figurative, e.g. *tord-boyaux*, *casse-pattes*, *casse-poitrine* (all three denote 'bad and strong alcohol').

Electronic resources

Le Trésor de la Langue Française informatisé, http://atilf.atilf.fr/tlf.htm.

References

Bauer, Laurie (2001). *Morphological Productivity.* Cambridge: Cambridge University Press.

Beard, Robert (1990). "The Nature and Origins of Derivational Polysemy." *Lingua* 81, 101-140.

Benveniste, Émile (1975). *Noms d'agent et noms d'action en indo-européen.* Paris: Librairie d'Amérique et d'Orient, Adrien Maisonneuve.

Booij, Geert (1986). "Form and Meaning in Morphology: The Case of Dutch Agent Nouns." *Linguistics* 24, 503-517.

Bork, Hans Dieter (1990). *Die lateinisch-romanischen Zusammensetzungen Nomen + Verb und der Ursprung der romanischen Verb-Ergänzung-Komposita.* Bonn: Romanistischer Verlag.

Chomsky, Noam (1981). *Lectures on Government and Binding: The Pisa Lectures.* Dordrecht: Foris.

Comrie, Bernard (1981). *Language Universals and Linguistic Typology.* Oxford: Blackwell.

Corbin, Danielle (1987). *Morphologie dérivationnelle et structuration du lexique.* Tübingen: Niemeyer.

Corbin, Danielle and Temple, Martine (1994). "Le monde des mots et des sens construits : catégories sémantiques, catégories référentielles." *Cahiers de Lexicologie* 65, 213-236.

Cruse, D. Alan (1973). "Some Thoughts on Agentivity." *Journal of Linguistics* 9, 11-23.

Dal, Georgette (2003). "Productivité morphologique : définitions et notions connexes." *Langue française* 140, 3-23.

Darmesteter, Arsène (1972) [1877]. *De la création actuelle de mots nouveaux dans la langue française et des lois qui la régissent.* Geneva: Slatkine Reprints.

Devos, Filip and Taeldeman, Johan (2004). "Deverbal Nouns and the Agentive Dimension across Languages." In D. Willems, B. Defranq, T. Colleman and D. Noël (eds.), *Contrastive Analysis in Language: Identifying Linguistic Units of Comparison.* Hampshire, New York: Palgrave Macmillan, 155-171.

Dressler, Wolfgang U. (1977). "Elements of a Polycentristic Theory of Word Formation." *Wiener Linguistische Gazette* 15, 13-32.

Dressler, Wolfgang U. (1986). "Explanation in Natural Morphology, Illustrated with Comparative and Agent-Noun Formation." *Linguistics* 24, 519-548.

Dubois, Jean (1962). *Étude sur la dérivation suffixale en français moderne et contemporain.* Paris: Larousse.

Fillmore, Charles J. (1968). "The Case for Case." In R. T. Harms, E. Bach and C. J. Fillmore (eds.), *Universals in Linguistic Theory: Symposium Held at the University of Texas at Austin, 1967.* New York: Holt, Rinehart and Winston, 1-90.

Giurescu, Anca (1975). *Les mots composés dans les langues romanes.* The Hague: Mouton.

Gross, Gaston (1990). "Définition des noms composés dans une lexique-grammaire." *Langue française* 87, 84-90.

Heinimann, Siegfried (1949). "Die italienischen Imperativkomposita." *Archiv fur das Studium der neueren Sprachen und Literaturen (ASNS)* 186, 136-143.

Kastovsky, Dieter (1971). "The Old English Suffix *-er*(*e*)." *Anglia: Zeitschrift für englische Philologie* 89, 285-325.

Kastovsky, Dieter (1986). "The Problem of Productivity in Word-Formation." *Linguistics* 24, 585-600.

Kreutzer, Käthe (1967). *Der Typus Guardavalle (Schauinsland) in den romanischen Sprachen.* Tübingen: Fotodruck Präzis, Barbara v. Spangenberg KB.

Littré, Émile (1957). *Dictionnaire de la langue française.* Paris: Éditions Jean-Jacques Pauvert.

Lloyd, Paul M. (1966). "A Possible Structural Factor in the Development of Verb-Complement Compounds in the Romance Languages." *Studia Neophilologica* 38, 257-262.

Meyer-Lübke, Wilhelm (1890). *Italienische Grammatik.* Leipzig: Reisland.

Meyer-Lübke, Wilhelm (1966) [1921]. *Historische Grammatik der französischen Sprache.* Heidelberg: Winter.

Panagl, Oswald (1977). "Zum Verhältnis von Agens und Instrument in Wortbildung, Syntax und Pragmatik." *Wiener Linguistische Gazette* 16, 3-17.

Panther, Klaus-Uwe and Thornburg, Linda L (2002). "The Roles of Metaphor and Metonymy in English *-er* Nominals." In R. Dirven and R. Pörings (eds.), *Metaphor and Metonymy in Comparison and Contrast.* Berlin, New York: Mouton de Gruyter.

Picone, Michael D. (1992). "Compound Word Formation in French: One Hundred Years Later." In C. Laeufer and T. A. Morgan (eds.), *Theoretical Analyses in Romance Linguistics.* Philadelphia, Amsterdam: John Benjamins Publishing Company, 181-195.

Rainer, Franz (2004). "Del nombre de agente al nombre de instrumento y de lugar en español: ¿cuándo y cómo?" *Iberoromania* 59, 97-122.

Rainer, Franz (2005). "Typology, Diachrony, and Universals of Semantic Change in Word-Formation: A Romanist's Look at the Polysemy of Agent Nouns." In G. Booij, E. Guevara, A. Ralli, S. Sgroi and S. Scalise (eds.), *Morphology and Linguistic Typology, On-line Proceedings of the Fourth Mediterranean Morphology Meeting (MMM4) Catania, 21-23 September 2003.* University of Bologna, 21-34.

Rohlfs, Gerhard (1954). *Historische Grammatik der Italienischen Sprache und ihrer Mundarten.* Bern: A. Francke AG. Verlag.

Rosenberg, Maria (in preparation). *Les composés [VN/A]$_{N/A}$ en français : perspectives morphologique, sémantique et contrastive.* Doctoral dissertation. Stockholm University.

Ryder, Mary Ellen (1991). "Mixers, Mufflers and Mousers: The Extending of the -er Suffix as a Case of Prototype Reanalysis." *Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society*. Berkeley (Calif.), Berkeley Linguistics Society, 299-311.

Schultink, Henk (1961). "Produktiviteit als morfologisch fenomeen." *Forum der letteren* 2, 110-125.

Sleeman, Petra and Verheugd, Els (2004). "Action and Agent Nouns in French and Polysemy." In D. Willems, B. Defranq, T. Colleman and D. Noël (eds.), *Contrastive Analysis in Language: Identifying Linguistic Units of Comparison.* Hampshire, New York: Palgrave Macmillan, 137-154.

Spence, Nicol Christopher William (1990). "Instrumental *-eur*/*-ateur*/*-euse*/*-atrice*." *The Modern Language Review* 85, 29-35.

Spitzer, Leo (1951). "Sur quelques emplois métaphoriques de l'impératif. Un chapitre de syntaxe comparative. I." *Romania* 72, 433-478.

Spitzer, Leo (1952). "Sur quelques emplois métaphoriques de l'impératif. Un chapitre de syntaxe comparative. Deuxième article." *Romania* 73, 16-63.

Štekauer, Pavol (2003). "Review of Laurie Bauer. Morphological Productivity." *Lingua* 113, 697-701.

Teleman, Ulf (1970). *Om svenska ord.* Lund: Gleerups.

Ulland, Harald (1993). *Les nominalisations agentive et instrumentale en français moderne.* Berne: Peter Lang.

Wartburg, Walther von (1922-). *Französisches etymologisches Wörterbuch: eine Darstellung des galloromanischen Sprachschatzes.* Basel: Zbinden.

Villoing, Florence (2002). *Les mots composés [VN]$_{N/A}$ du français : réflexions épistémologiques et propositions d'analyse.* Doctoral dissertation. Université Paris X-Nanterre.

Winther, André (1975). "Note sur les formations déverbales en *-eur* et en *-ant*." *Cahiers de Lexicologie* 26, 56-84.

Wooldridge, Russon (1998). "Le lexique français du XVI^{e} siècle dans le GDFL et le FEW." *Zeitschrift für romanische Philologie* 114, 210-257.

Zwanenburg, Wiecher (1983). *Productivité morphologique et emprunt. Étude des dérivés déverbaux savants en français moderne*. Amsterdam, Philadelphia: John Benjamins Publishing Company

Romanian Participle: 3 Items with 1 Morphological Unit

Elena Soare
Université de Paris 8
elena@linguist.jussieu.fr

Abstract

This paper attempts to reanalyze the relationship between the Romanian Participle and the Romanian Supine, two homophonous participial forms considered as different by Romanian grammarians. On the one hand, it can be shown that the "mixed" nature of the Supine is rather contextually dependent; therefore, this element can be considered to be a neutral form, categorized by the syntactic context. This strongly suggests an analysis in terms of under-specification, in the lines of Distributed Morphology, with category-less items. On the other hand, the fact that the same morphology is used in Supine and Participle contexts cannot be a simple coincidence. The base form should be one single element, i.e. one of the forms of the verbal root in Romanian, enlarged with the participial morpheme. We will suggest that this basic unit is unique for past Participle and Supine, supporting a view in which Romanian morphology is partly based on stems[1].

1. Introduction

1.1. The facts

There are two syntactic uses of Perfect Participle crosslinguistically: a) the perfect verb formation (+HAVE): b) the passive verb formation (+BE). Romanian has a third use of the Perfect Participle: the "Supine". The form of past participle in Romanian is used in three types of syntactic environments: nominal, adjectival and verbal. There is on the one hand a past participle, with verbal and adjectival uses:

(1) a. am citit cărți
have read books
'I have read books'

b. cărțile sînt citit-e
books are read-AGR

[1] Romanian past participles are of the form: Root + Thematic Vowel [A/E/I/U] + T/S
cînta – cîntat "sing"
vedea – văzut see"
merge – mers "walk"
hotărî – hotărît "decide"
iubi – iubit "love"
In the following, we will use -AT a shortcut for the participial morpheme.

G. Booij, *et al.* (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)*
Fréjus 15-18 September 2005, University of Bologna, 2007.
URL http://mmm.lingue.unibo.it/

On the other hand, traditional grammars take some of the contexts of the Romanian participle (preceded by prepositions) to involve a distinct form of the verbal paradigm, called *supine*, considered to have a mixed [+N, +V] nature (cf. (2)-(4)). In (2), the participial form appears in a nominal structure, in combination with a definite determiner, and the complement is assigned Genitive case. In (3), the supine follows an aspectual auxiliary, expressing the completion of an activity; in this case, it does not take an article, and requires the presence of a marker, a functional "preposition", *de*. In (4), the supine is preceded by a subcategorized preposition in the complement position of a verb of motion (expressing the Goal).

A. nominal supine: D° + participle:

(2) citit-ul cărți-lor
reading-the books-GEN
'the reading of the books'

B. verbal supine: preposition + participle[2]

B_1 participle without D°, obligatorily preceded by the "preposition" *de*:

(3) am terminat de citit cărțile
have finished to read books-the
'I have finished reading the books'

B_2 participle without D°, preceded by a subcategorized preposition

(4) mergem la pescuit de scoici
go-2PL to fishing of mussels
'we are going fishing mussels'

1.2. The issue

We are faced here with a theoretical puzzle, namely how to solve a case of « grammatical homonymy », how to treat this multi-functionality of a single morphological unit getting three different syntactic uses?

In the generative literature, some verbal nouns have been analyzed as "mixed categories" (+N, +V), for instance the Arabic *masdar* (Fassi Fehri 1991), the Welsh verbal noun (Rouveret 1993) as well as the English gerund. Is the supine also a verbal noun? In section 2, we will present evidence that this is not the case.

Moreover, we are faced with the problem of giving a status to the "participial morpheme", in our case AT (see footnote 1). How should this element be treated? As we shall see in the following, it can be considered as an inflectional affix, deriving

[2]The example in (3) illustrates the supine in aspectual constructions. The same form appears in copular structures, reduced relatives, Tough constructions, i.e. in predicative contexts. Besides the adjunct position, when it corresponds to a PP, as in (4), the supine is not equivalent to a subordinated CP; the Romanian complementation uses the indicative or subjunctive forms for subordination.

participles. But it can be part of derivational formations also. Is it the same element? Should we analyze it as a default, "decategorizing" affix? Since Aronoff (1994), the existence of forms like the Latin Supine is considered as evidence in favor of a "pure morphology", with no meaning-form correspondence. Does this view extend to the Romanian Supine? Sections 3 and 4 will concentrate on these topics and try to give some insights.

2. Mixed or underspecified?

There are several theoretical possibilities to account for this kind of mismatch. The "mixed" analysis mentioned above, in the line of a long grammatical tradition, tries to capture the property of a "participle" to "participate" in the verbal and in the nominal "nature" as well. One may wish to capture this property in the lexicon or in the syntax. In the lexicalist models such as HPSG, it is natural to assume that categories are in the lexicon, and to allow the existence (formation) of the appropriate number of lexical categories. Since case is assigned by the lexical head, there will be as many categories as there are case inflections in the domain of that category.

Another option is the one adopted in the framework of Distributed Morphology (Marantz (1997), Harley and Noyer (1998)). In this view, the items listed in the vocabulary have no category, categorization being contributed by the syntactic component. The insertion of an element in the appropriate syntactic context makes it a nominal, or verbal, or adjectival element. We take this option to be more satisfactory for Semitic roots, which give rise to verb as well as noun formation.

There is a clear connection between the syntactic behavior and categorization. Lexical categories, heads of syntactic projections, determine the internal structure of the projection (selection, projection, complement licensing), and the type of position in which the corresponding phrases will be inserted, as well.

For the case of the supine, there are empiric facts supporting an "under-specification"-type analysis. A criterion for the mixed character is the existence of two kinds of properties **in the same projection** and **in the same time**. For instance, the distribution would be nominal, as for the English gerund, which can appear in contexts excluded by non-nominal projections:

(4) a. we were concerned about Pat's watching television

b. *we were concerned about that Pat was watching television

The same type of projection is characterized by internal properties specific for verbs, i.e. Accusative case assignment and adverbial modification:

(5) a. John's building a spaceship

b. I disapproved of Pat's watching television

c. Pat disapproved of my *quiet/ quietly leaving before anyone noticed

d. *Pat disapproved that leaving

In (5)a-d, we give some other examples of mixed categories, manifesting a "griffon" behavior: a verbal head with hybrid properties – nominalized infinitive in Italian ((6)a), Spanish ((6)b), Old Romanian infinitives ((6)c) and Arabic masdar ((6)d):

(6) a. il riverede un compagno d'armi
the see-again a companion of arms
'the fact of seeing again a brother in arms'

b. el haber-me-lo dicho
the have-me-it said
'the fact that he told it to me'

c. tăierea capul lui
cutting-the head-the him
'the fact of cutting his head'

d. quatl-u Zayd-in Muhammad-an
murder-NOM Zayd-GEN Muhammad-ACC
'the murder of Muhammad by Zayd'

The behavior of the Romanian supine does not respect the mixed-behavior criterion; instead of showing hybrid properties **in the same time**, its verbal / nominal nature (or "ambiguity", according to traditional grammars) manifests as contextually dependent. And indeed, we see that the supine combined with a determiner fails to assign accusative or nominative case. Its projection is completely reorganized according to the nominal pattern:

(7) a. *culesul mere
picking-the apples
'apples picking'

b. *culesul Ion
picking-the Ion
'Ion's picking'

The problem in these examples is the fact that the arguments are not assigned case, which in Romanian corresponds to the morphological case, or to the insertion of a "preposition". The problem disappears when the internal argument appears in the form of an NP in the Genitive or that of a PP adjunct:

(8) a. culesul merelor
picking-the apples-GEN
'apple's picking'

b. culesul de mere
picking-the of apples
'the picking of apples'

As expected, the supine loses its ability to assign Accusative case when combined with a D° ((9)a), and does not assign Nominative case ((9)b):

(9) a. *pescuitul scoici
fishing-the mussels
'the fishing of mussels'

b. *pescuitul Ion
fishing-the John
'the fishing of John'

In this case, the internal argument appears as a Genitive DP ((10)a) or as an adjunct PP ((10)b):

(10) a. pescuitul scoicilor
fishing-the mussels-GEN
'the fishing of mussels'

b. pescuitul de scoici
fishing-the of mussels
'the fishing of mussels'

In the case of verbal supines, the licensing of the object (Accusative Case assignment) depends on a (semi)auxiliary.

(11) a. am de cules mere
have to pick apples
'I have to pick apples'

This observation leads to the idea that the supine cannot assign itself a case to its complements. In the case of the nominal projection, this is done by the presence of the nominal determination. Indeed, in Romanian, the incorporated determiner bears the case inflection. Within the nominal projection, the supine combines with semi-auxiliaries, forming a verbal complex, which, as a whole, is responsible for case-licensing of the complements. The result is not a "mixed" projection because the different properties of the supine do not manifest in the same time, in the same projection, but in different projections. The supine changes its projection type as it changes its morphological properties.

There is an apparent exception to this generalization, represented by the supine inside a PP projection, in which the Preposition is selected by the main verb or has an autonomous lexical meaning (such as Goal), in the case of adjuncts. In traditional grammars, this prepositional supine is considered to be verbal (Accusative-Case Assigner). It is the essential argument of traditional grammars for the view that the supine keeps its verbal properties in this kind of contexts.

(12) am plecat la cules mere
have gone at picking apples
'I'm going to pick apples'

On the basis of this type of examples, traditional grammars take the view that the supine is a case-assigner when it is introduced by a Preposition. This view is nevertheless contradicted by the fact that, when the supine is preceded by lexical sub-categorized Prepositions, the prepositional accusative becomes impossible for the object of that supine. This shows that in fact the Accusative Case is not assigned by the supine:

(13) *am renunțat la invitat pe Ion / pe acest om
have renounced to invite PE-ACC Ion / PE-ACC this man
'I renounced to invite Ion / this man'

This is even more striking if we compare the supine with another non-finite form of the Romanian verbal system, the infinitive, which is perfectly compatible with prepositional Accusative Case:

(14) am renunțat la a invita pe Ion / pe acest om
have renounced to invite PE-ACC Ion / PE-ACC this man
'I renounced to invite Ion / this man'

Another important remark is that in prepositional contexts, the object is always strictly adjacent and rather non-determined, whereas in ordinary verbal constructions, the object allows determination, quantification etc, and can be separated from the verb by temporal modifiers:

(15) a. *am plecat la cules multe mere / *toate merele
have gone to pick many apples / all apples-the
'I am going to pick many apples / all the apples'

b. *am plecat la cules imediat mere
have gone to pick immediately apples
'I am going to pick immediately apples'

Therefore, it seems that a direct object in such supine constructions manifests a special behavior, to be distinguished from the regular behavior of a direct object in an ordinary verbal construction. The suggestion that we would like to make here is that this behavior is due to the fact that the structure is frozen, and relies probably on a composition of the type N-N, generated by lexical rules (as compounds). An argument in this sense is given by the existence of structures of the type in (16), appearing in enumerations or other particular contexts (like titles or labels), where the model of composition is possible with participles:

(16) n-am uitat nimic: cumpărat bilete, făcut bagaje…
not have forget anything: buying tickets, packing bags….
'I didn't forget anything: buying tickets, packing bags…'

We will then set apart the contexts with a supine introduced by a lexical Preposition (sub-categorized or semantically selected), as being particular formations, resulting from the application of a lexical rule like compounding.

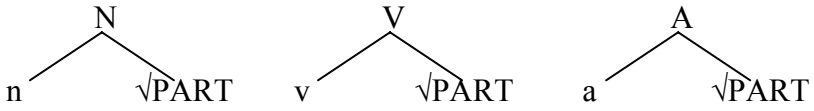
Nowadays Romanian tends to favor the construction with *de*-insertion instead of the direct one with the Accusative object. The perspective that we have, if we look at productive supine structures, is that of an element whose verbal-nominal nature is not "double" or "mixed", but clearly context-dependent, distinguishing this form from "true" verbal nouns.

3. Elements for an analysis

3.1. The main point

If this view is correct, we have to admit that the Romanian supine has a well differentiated behavior, not [+N] AND [+V], but [+N] OR [+V]. It is not a mixed category, but something that can accept to become a verb and a noun as well. An analysis that seems to impose itself is the "under-specification" analysis, proposed in the Distributed Morphology framework. The Participle fills a cell in the verbal system, at a morphological level, and has no categorial features, being categorized by the syntactic context.
In the DM framework, the categorization supposes the contribution of functional categories as n, v, a. For our purpose, all that we need to say is that by simply placing a lexical underspecified item in a typical verbal, nominal or adjectival position, this category acquires verbal, nominal or adjectival properties:

(17)



According to the view of Distributed Morphology, there are no categorial features at the "Vocabulary" level. The participle would then be a single uncategorized item competing for several syntactic contexts.

What is less clear is which label to put on the participial form itself. In DM, there are category-neutral "Roots", and there are affixes with features competing for a specific value to express. The analysis that we would like to propose is that the Participle, here above PART, is itself category-less. Or, it is not really a Root, in the sense of "simple", but it is rather constructed from a verbal root and the participial affix. We shall now try to find a solution to this puzzle.

3.2. The proper category-less level of abstraction

There are several theoretical views that are coherent with the existence of a un-categorial level of grammatical representation. Baker (2003), for instance, considers that categories are given by syntax. According to him, the categorial identification is done by the syntax in the following way:

A Noun – has a referential index
A Verb – has a specifier
An Adjective – by default: it is –N, -V

In his system, however, the Participle is not really discussed; it is considered a 'verbal adjective' without further investigation.

Schütze (2003) takes the participial suffix as being category-changing, creating Participles from Verbs. This makes them derivational affixes. As for the Participle, it is considered as not being (really) a Verb; it does not carry Voice, but only (lexical?) aspect. So, the Participle is a de-categorized Verb. This, however, takes in fact the Participle to be a distinct category.

Aronoff (1994) discusses the problem of the (English) Participle, which, according to him, illustrates the "morphomic" level; being purely morphological, this element is appropriate for the very different syntactic constructions of past and passive. Another argument for the "morphomic" level discussed by Aronoff (1994) is the existence of the Latin "third stem", realized in participle, Supine, and future active participle. In this case, a single stem, also a "morpheme", is used in various syntactic environments. The supine was a verbal noun; derived from a participial stem (from a synchronic point of view), it was an item that allowed nominal inflection (Case marking) and appeared as Goal adjunct with verbs or adjectives:

(18) a. eo lusum
'go playing'

b. mirabile visu
'wonderful to see'

The argument in favor of a verbal noun analysis comes from the possibility of case marking on the Direct Object by the supine, and the co-occurrence of the nominal inflection on it. According to Aronoff (1994), the supine should be treated as the manifestation of the same *Stem* (in the strict morphological sense, at the "morphomic" level) as the one of the Participle, even if the values (aspect, voice) of the categories derived from this stem are different (the supine does not admit the passive interpretation in Latin). The same stem is used to derive a number of deverbal nouns in Latin, such as *pictura*, derived from *pingo, pict-*. In this way, Aronoff (1994) builds an argument for the view that the morphologic level should be kept distinct from syntax, semantic or phonology; morphemes do not encode (grammatical) meaning since, in the cases illustrated, they do not always have the same value.

Indeed, the same Thematic element appears in very different formations, like the active future participle, meaning "those which will V" and in the supine, denoting the activity without further specifications, and in the past participle, denoting a (resulting) state.

Active Future Participle

(19) mor-it-uri te salutant
'die-Th-FUTP you salute'

Supine

(20) eo pisc-at-um
'go1sg fish-Th-ACC'

It could also be interesting to note that, putting aside any attempt to diachronically explain the existence of the Romanian supine (in other words, the large use of the participial stem), the Romanian supine and the Latin supine present rather similar distributions, i.e. the expression of the goal or of the point of view.

We are not able to propose, at this moment, an analysis for both Latin and Romanian, but we may simply retain Aronoff's suggestion for Romanian, that a single stem is at work in the two cases. This stem can be used as a base as well for verb as for noun formation. In sum, none of these forms is basic; they are all derived from a single "sound form" – a stem.

Therefore, we would like to apply the same view to Romanian Participle and Supine, which, as we saw, are homophonous. The advantage would be that we would unify two categories of the non-finite verbal system of Romanian that seem to have all in common. But in this case, we would rather like to say that a single morphological element, categorially neutral, is used to build a Noun (the Supine), a Verb (in combination with an Auxiliary, as seen above), or an Adjective. This is the analysis proposed above, and it goes somehow in the same direction as Aronoff's discussion.

One further question to ask is whether Romanian morphology can be considered to be based on stems, and if the Participle is a stem. Such a view could be supported by the fact that the –AT formation is also used in derivation. -AT can also attach to non-verbal roots

Root +V → V-at (categorially neutral element)

(21) Mîncat 'eating', cîntat 'singing', citit 'reading', mers 'walking'

Root +N → [Adj] (state of someone who has…)

(22) Sprîncenat 'eye-browed', migdalat 'almond-ed'

-AT could be considered also as participating in the derivation of agent Nouns. The idea that it would be the realization of the same stem is however contradicted by the variation illustrated in (23)d-e. However, the correspondence between the participle and the stem of agent Nouns is stable for the "regular" classes of verbs.

Root + t +-or → agent Nouns

(23) a. cînta 'sing'– cîntăt-or 'singer'

b. măguli 'flatter'– măgulit-or 'flatterer'

c. hotărî 'decide'– hotărît-or 'decisive'

d. vedea 'see'– văzut – văzător - *văzutor

e. merge – mers – mergător - *mersor

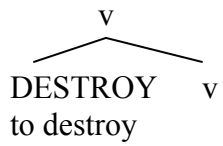
There are, we think, reasons to treat the participial form as one of the *stems* of the verb in Romanian. This could be a rather peculiar use of the notion of *stem*, which should be used for "what remains when all affixes are set apart". In our case, contrary to Latin, there are no affixes going with the nominal vs. verbal or adjectival status[3]. But we could assign the status of a stem to the part which is common to all these syntactic uses mentioned above, and it would correspond to "what remains when the categorial features are set apart".

4. What we need to add to Marantz's (1997) model

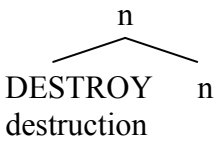
Let us sketch now a way of making sense of all the intuitions above. We will assume that the most qualified model which could account for the facts outlined in this paper is Distributed Morphology, as depicted in Marantz (1997). The only inconvenient would be that our participles are in the same time basic and constructed (see above, section 2). In order to make the machine work, let us assume that word formation starts with ROOTS, but at the level of pure forms (the "morphomic" level), stem alternation can apply. For instance, Romanian verbs are derived either from a perfective (participial) or from an imperfective stem (infinitive). In English, this alternation is manifested by the two possible forms of an abstract ROOT as DESTROY: destroy vs. destruct.

Then, neutral categories, ROOTS in Marantz's system, morphologically mapped into stems in our system, are combined with nominal, adjectival, verbal heads. In our case, that of Romanian Participle, these heads do not have phonological content. Put differently, we assume that the difference between the formation of Romanian Participle / Supine and that of English Gerund is that -ING is a nominal head, whereas -AT is not. We illustrate hereafter the different type of word-formation corresponding to the Gerund and to the Supine; all of them take place in the syntactic component, according to Distributed Morphology.

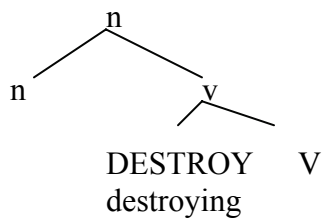
(24) a.



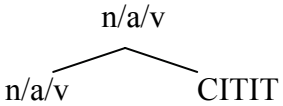
b.



c.



d.



[3] There are, of course, marks of agreement in gender and number when the Participle is in an adjectival position, but no marks of declension for the Supine use.

5. Is -AT (simply) an empty morpheme?

What -AT is, then? The answer, in Aronoff's terms, would be that -AT is an empty thematic morpheme and we have already seen the arguments (section 2). We may have some arguments for taking this morphological piece as the expression of (lexical) Aspect. Participial stems, as we will argue below, encode (telic) Aspect. If this is correct, the view of Aronoff (1994) about the complete absence of semantic-grammatical value for the participial stem could be challenged, at least for Romanian Participle. The thematic affix -AT seems to keep a certain value in Romanian, which we take to be an aspectual one.

-AT encodes a [+ Telic] feature in Perfect Participle and Supine, in periphrastic (aspectual) constructions, but also in event nominalizations.
In some cases, the aspectual value of the participial stem can be changed, for instance in some periphrases with supine expressing completion, a value that has also to do with telicity. See for example (25), where the action of reading has to be completed, or the movement to reach its goal:

(25) a. am de citit acest articol pînă mîine
have to read this article till tomorrow
'I have to read this article till tomorrow'

b. am de mers la piață
have to go to market
'I have to go to the market place'

There are also event nominalizations in which the Telicity is encoded:

(26) a. cititul ziarului de dimineață
reading-the newspaper-GEN of morning
'the reading of the newspaper in the morning

b. ?cititul de dimineață
reading-the of morning
'reading in the morning'

The Supine nominal, however, can be atelic, in examples like the following:

(27) cîntatul este un dar
singing-the is a gift
'singing is a gift'

Those are contexts with a generic reading, the supine denotes a generic event, and the aspectual value is shifted to the iterative-habitual reading.

In other derived nouns, the Participial stem is associated with a state or with a result reading. An apparent counterexample to the idea that –AT encodes telicity could be seen in -tor derived agent Nouns, which are active, non telic. In those examples too, there is an aspectual component which is habitual. Take for instance mâncător 'eat-er', mergător 'walker', dansator 'dancer'; it is reasonable to say that you have to do some dance (to have some dancing experience) in order to be a 'dancer'.

We may assume, then, that the Supine/Participle stem encodes Aspect; its basic value could be considered to be telicity. Some contexts, however, may involve shifting to an iterative-habitual reading.

As for the Voice value, this stem is considered to be Voice-ambiguous in the Romanian literature. This may go in the sense of Aronoff (1994)'s discussion referred to above. But in fact, Supine's properties lead to think that it is rather a non-active form (maybe a middle). The active reading is not possible unless the supine has nominal properties, i.e. in the prepositional context discussed above. In the other cases, if it does not have a clear passive reading (which holds for the supine reduced relatives), the active reading is associated with an arbitrary reading of the subject.

However, for the topics addressed in this section, further research is needed.

6. Conclusion

In the present paper, we proposed an analysis according to which the participle, ROOT+AT *by itself* is not [+N], [+V]. We have outlined the fact that participles need syntactic supporters – functional elements, i.e. auxiliaries or determiners, in order to receive a category. The (Stative and Resultative) Participle, the verbal "Supine", "Supine" Event Nouns in Romanian are all syntactic realizations of an aronoffian "3rd stem".

The AT-Stem (or PART, or the 3rd stem) simply combines with different functional layers attributing categories: a, n or v. We do not assume that AT is itself a, n or v.

Stems are categorially neutral and accessible to inflection and to derivation. This leads to the triple use of the Romanian Participle known as the Past Participle / Supine parallelism in Romanian. We think that the view of the grammar which could fit the facts discussed in this paper is the one outlined in Distributed Morphology, where the notion of "mixed categories" is not needed. In such a framework, it is possible to have an analysis in which a single morphological piece corresponds to three linguistic units. The only device that we would have to add is that sound forms of ROOTS, i.e., stems, are categorially neutral and represent starting points in word formation.

References

Abney, P.S., 1987, *English NP in its Sentential Aspect*, MIT Dissertation, Cambridge, Mass.

Aronoff, M., 1994, *Morphology by itself. Stems and Inflectional classes,* MIT Press

Baker, Mark C., 2003, *Lexical Categories. Verbs, Nouns, Adjectives*, Cambridge University Press

Booij, G, Guevara, E, Ralli, A, Sgroi, S & Scalise, S (eds) *Morphology and Linguistic Typology, On-line Proceedings of the Fourth Mediterranean Morphology Meeting (MMM4) Catania 21-23 September 2003,* University of Bologna, 2005, URL http://morbo.lingue.unibo.it/mmm/

Don, J, 2005, "Roots, Deverbal Nouns and Denominal Verbs", in Booij, G, Guevara, E, Ralli, A, Sgroi, S & Scalise, S (eds)

Embick, D., 2004, "On the Structure of Resultative Participles in English", LI 35, 3, 355-392

Fassi Fehri, A., 1991, *Arabic clauses and words*. Dordrecht: Kluwer

Haspelmath, M., 2002, *Understanding Morphology,* "Understanding Language Series", Arnold, London

Harley, H. and Noyer, R., 1998, "Mixed nominalizations, short verb movement and object shift in English", *NELS* 28, 143-157.

Malouf, R., 1998, *Mixed Categories in the Hierarchical Lexicon*, PhD Dissertation, Stanford University

Marantz, A., 1997, "No Escape from Syntax: Don't try morphological analysis in the privacy of your own Lexicon", in A. Dimitriadis et al. (eds), *Proceedings of the 21st annual Penn Linguistics Colloquium*, "Upenn WPL", 4, 201-225

Rouveret, A.,1993, *Syntaxe du Gallois. Principes généraux et typologie*, Editions du CNRS

Schütze, C.T., 2003, "When is a Verb not a Verb?", *University of Tromsø working papers on language and linguistics*, Vol.31, Nr. 2

Soare, E., 2002, *Le supin roumain et la théorie des catégories mixtes*, thèse de doctorat, Université de Paris VII.