Morphemes and Lexemes versus "Morphemes or Lexemes ?"

François Nemo
Univ. Orleans
fnemo@univ-orleans.fr

More than a century after the first linguistic definition of the notion of morpheme by Baudoin de Courtenay (1895) and Sweet (1876), an ever-lasting debate – which I shall refer here as the "Morpheme or Lexeme" (M or L) debate – on the nature of linguistic bricks is still going on.

Since a by-product of this debate is terminological confusion in the use of the four notions of *morpheme*, *lexeme*, *word* and *item*, I shall start by describing very briefly the conflicting uses of these terms and then show that to accounting for the generation of the lexicon, i.e. of both new senses and new lexical units, requires accepting the co-existence of two distinct semantic stocks and explaining how morphemes which belong to the first stock may become lexemes or be involved in lexeme-formation processes.

1. Uses of the Notions of *Morpheme*, *Word*, *Lexeme* and *Item*

1.1 Uses of the Notion of "Morpheme"

There are basically four uses of the noun *morpheme*:

- *morpheme* may be used, following Baudoin de Courtenay (1895), in order to refer to the smallest meaningful linguistic unit, a minimal sign identified as a semantic atom through a process of decomposition, regardless of its syntactic autonomy. Within such a methodology, the definition includes both a unit like the English *milk*, which cannot be decomposed in smaller elements and is syntactically autonomous, and a unit like the French *-spir-* which is the result of the decomposition of the words *re-spir-er* (to breathe), *in-spir-er* (to breathe in), *ex-spir-er* (to breathe out) but is not syntactically autonomous;

- *morpheme* is commonly used to refer to infra-lexical semantic units, typically affixes, which are not syntactically autonomous and hence are not words. Following Corbin, I shall refer to such bound morphemes as *infra-lexical units*, leaving out of this category bound morphemes which are used for flexion, since such bound morphemes do not belong to the lexicon.

- *morpheme* may be used to refer to grammatical bound morphemes only. In such a case, the notion of morpheme is associated with the notion of flexion and therefore morphemes are not lexical units at all;

- *morpheme* has been used in contemporary linguistic semantics to refer to a form/signification pair which can be isolated by considering all the uses of a single semantic unit, including categorically distinct ones. For instance, semanticians will

G. Booij, E. Guevara, A. Ralli, S. Sgroi & 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, 2005. URL http://morbo.lingue.unibo.it/mmm/

speak of a single morpheme *but* in English, encoding stable semantic indications, so as to account for all the uses of *but* in English, for instance its uses with the non connective meanings of *almost*, *without*, *except*, *only* and with the connective meanings of *but*, *rather*, etc. . This semantic definition differs from the classical one by its using a distributional methodology – i.e. considering all the uses of a given unit, so as to isolate encoded meaning – instead of the decompositional methodology advocated by the structuralists.

The aim of this paper will be to show that this last definition is of crucial importance for any morphological theory.

1.2 Uses of the Notion of "Word"

I shall not detail here all the conflicting definitions of the notion of *word* which have been proposed so far (for an overview of the problem, see Di Sciullo & Williams 1987; Dixon & Aikhenvald 2002), and I shall limit myself to the few issues directly at stake in the "M or L" debate:

- it has been repeatedly asserted that *words* and not *morphemes* are the minimal signs of a language (Aronoff 1976; Anderson 1992). The usual justification of such a view is that all the combinatory rules, whether grammatical or morphological, are word-based and not morpheme-based and that ordinary speakers have a semantic intuition about the meaning of words, but no intuition at all about the meaning of morphemes;

- it has been repeatedly asserted that infra-lexical units such as affixes have no stable meaning (Aronoff 1976) and hence are not signs in the Saussurean sense, but processes.

- it has been assumed (Di Sciullo & Williams 1987) that there are two class of words, listemes on the one hand (which have to be learnt one by one and are either semantic atoms or unpredictable complex units) and generated words on the other hand (which are the outputs of regular word-formation processes);

- it has been argued within a constructional approach to morphology that such infra-lexical units possess a meaning indeed, but that this meaning is instructional/procedural and not conceptual;

- it has been repeatedly asserted in linguistic semantics that data-based observation of the uses of "words" shows all too clearly that the actual uses of words are semantically distinct from our intuition about their meanings (e.g. that possibly 90% of the lexicalised uses of the French verb *balayer* are not predictable from intuition), and that doing semantics implies forbidding the use of intuition, adopting a clear distinction between *signification* and *sense* (Benveniste 1954; Ducrot 1987) and admitting that only form/signification pairs are signs in the Saussurean sense, and that the form/sense pairs provided by ordinary dictionaries are not linguistic signs but only local interpretations of these signs and of the constructions in which they are inserted;

1.3 Uses of the Notion of "Lexeme"

The notion of *lexeme* is frequently used as a technical and less ambiguous equivalent of the notion of *word*. It has the advantage of allowing the integration in the lexicon of many lexical units which are not *words* in the ordinary sense (for instance because they are formed of smaller units which are also *words*). *Lexemes* may thus be a cover term so as to refer to all the semantic units stored in the lexicon, and may thus refer to lexical units (e.g. *milk*), infra-lexical units (e.g. units as *micro-*, affixes – if we admit they have a signification – and bound bases) and supra-lexical units which behave like syntactic atoms (e.g. *pomme de terre – potato*, litterally *apple of ground* – or phrases like *tout à fait*, lexicalised idioms).

1.4 Uses of the Notion of "Item"

Given the "L or M" debate, and the typological differences between agglutinative and polysynthetic languages and isolating ones for instance, the term item has often been used to avoid the more controversial terms morpheme or word. Item is thus compatible with both a concatenative and a non-concatenative view of morphology, and also with a sign-based and process-based view of morphology. Semanticians do not use it at all.

2. The "Morpheme or Lexeme" Debate.

From structural linguistics to contemporary morphology, it has been widely assumed that a choice had to be made between morpheme-based models and lexeme – (i.e. words) based ones. In order to understand why we should rather consider morphemes and lexemes as two kinds of linguistic and semantic units which co-exist (and are complementary) and hence shouldn't be opposed, what must be remarked is that according to the classical definition within structural linguistics morphemes were simultaneously: i) the basic **semantic units** of a language; ii) the basic **combinatorial units** of a language. Within such a view the basic **semantic units** and the basic **combinatorial units** of a language were assumed to be the same thing. For instance the unit *table* is considered at the same time as an atomic semantic unit and a noun, i.e. as a syntactically defined unit.

The problem with such a view is that it leaves no choice but to **list** a huge part of the lexicon (e.g. words like *rétablir*, *tabler*, *tableur*) and to postulate endless sense enumerative lexicons (with as many entries for *table* as senses that the noun may have in its different uses, and with as many different units *table* as needed to explain the existence of bound bases in words like *se rétablir, rétablir, tableur*[1]).

Fortunately, since accounting for the diversity of uses (and hence senses) of a semantic unit is precisely the aim of semantics – according to the linguistic semantics framework shared by a large part of contemporary semantics (Benveniste 1954; Ducrot

[1] One needs to add to the questions of polysemy, "semantic drift" or "semantic bleaching" the question of polycategoriality: since a unit like *timap* in Palikur (Arawakan) "means" simultaneously *to hear*, *to shout, echo*, *loudly*, etc., depending on the way it is used, we can either adopt mere degrouping homonymy and have as many lexical units *timap* as there are ways to use it, or refuse this "solution" (Pustejovsky 1995) and admit that a distinction must be made between a non-categorial semantic unit *timap* and the categorial (and contextual) interpretations it receives in each of its uses.

1987, Bouchard 1995, Pustejovsky 1995, Nemo 1998, 1999, 2002a, to appear) – recent developments have enabled a clear understanding of the nature of the problem and, more importantly, of the nature of its solution, advocating for the co-existence of morphemes and lexemes in a language.

2.1 The Signification / Sense Distinction.

As mentioned earlier, the distinction between (lexical) senses and (encoded) signification is the founding postulate of all Linguistic Semantics models and descriptions. It is associated with the ideas that:

- (encoded) linguistic signification is accessible to the linguist only by considering the variety of uses of a semantic item;

- (encoded) linguistic signification must explain these uses just as rules explain sentences;

- (encoded) linguistic signification is usually not intuitive.

- significations are to senses what the equations of functions are to the points created by these equations;

- significations and senses are different in nature (Ducrot 1987; Nemo 2001c) since signification is neither some kind of a very abstract sense, nor the common denominator of these senses ;

2.2 The Lexicon as a Memory of Interpretations

Within such a view, which has proved to be extremely efficient to account for polysemy and polycategoriality, our understanding of the nature of the lexicon itself is deeply transformed. For most linguists outside of semantics, the received idea has long been that languages were formed of a lexical stock of combinatorial units on the one hand and of a set of combinatorial rules on the other, and that sentences (or words) are the outputs of a generative process whose inputs are the lexicon and the combinatorial rules.

Within linguistic semantics on the contrary, it is more and more widely acknowledged that the role of semantics is to account both for the generation of new meanings (and thus for polysemy) and for the generation of new lexical units (and thus of the lexicon itself). Within such a view, it is not legitimate to take the existence of the lexicon for granted, and the lexicon itself is what has to be explained and is therefore the **output** of a process which has to be described and whose inputs, as we shall see, are morphemes and constructions.

Understanding the co-existence of morphemes and lexemes in that perspective requires only to understand that morphemes, which encode significations, are the inputs of a process in which:

- each time a morpheme is used, it is inserted in a construction and in a context and it receives a constructional and contextual local interpretation;

- if the same use is repeated, i.e. if the same morpheme is used in the same construction and the same context, the interpretation process is not repeated and the interpretation becomes memorised, a process called conventionalisation or lexicalisation.

Thus, within linguistic semantics, the lexicon is only a memory of the interpretations of morphemes in their different uses. As a result, lexemes are not the basic semantic units and all languages have two stocks of semantic units, a stock of linguistic units on the one hand (that we shall call morphemes from now on) and a stock of lexical units on the other hand (called lexemes and including lexical, supra-lexical and infra-lexical units).

And finally the whole picture consists in a triple (and parallel) distinction between:i) signification and sense; ii) morphemes which encode signification and lexemes which have senses; iii) the linguistic stock consisting of morphemes and the lexical stock consisting of lexemes.

2.3 The Signification / Sense Distinction and Morphology

As for morphology, the distinction between signification and senses has far-reaching consequences:

- the notorious instability of the form/sense relationship, which has led many linguists to consider meaning as irrelevant for morphological theory (Aronoff 1976) has misled them about the semantics of morphemes. Morphemes have a very stable meaning in all their uses, and are also very stable in diachrony (i.e. much more stable than grammatical or morphological structures);

- the importance of listemes in the lexicon, possibly 40% of the word-forms found in corpus-based studies and probably up to 80–90 % of the senses of apparently well-formed words, can be accounted for only by using the morpheme/lexeme distinction.

2.3.1 Accounting for Semantic Instability

Within contemporary linguistic semantics, i.e. by adopting the signification/sense distinction, it has become possible to show that the diversity of uses of a semantic unit was compatible with the fact that this unit encodes a very stable signification.

So that for instance, the English semantic unit *but* does not encode a connective and pragmatic sense (whose equivalents would be the German *aber* or *sondern*, the Spanish *pero* or *sino*, or the French *mais*) on the one hand and have unpredictable non connective and non pragmatic uses on the other (with the meanings of *almost*, *without*, *only*, *except*, etc..). Instead, what we have is:

- a single morpheme *but*, encoding the indication that "something had (could have, should have, etc.) been stopped", and which may be inserted in different constructions and positions (for instance in connective and non-connective positions) where it receives a local (constructional and contextual) interpretation;

- different lexemes *but* (with the lexicalised meanings of *aber*, *sondern*, *almost*, *without*, *only*, *except*, etc.) with their own polysemy.

Within such a view (Nemo 2002a, to appear), it becomes indeed possible to account for the various interpretations of *but* in the three following utterances: (1*) The price is interesting but I have no money*; (2) *But for Peter, I would be dead*; (3) *This specie has but disappeared*; since despite constructional differences, it describes in the three cases the fact that a process is not completed, with "having no money" being the blocking factor in (1), Peter's intervention being the blocking factor in (2), and the not fully completed disappearance accounting for the "almost" interpretation in (3).
Within such a view: i) morphemes are semantic units (i.e form meaning pairs) but not syntactic units (they provide no combinatory information); ii) lexemes are syntactic-semantic units; iii) morphemic meaning can be identified only by taking into consideration all the uses of the morpheme regardless of its syntactic status; iv) morphemic meaning is indicational and lexical meaning is a conventionalization of a morphemic/constructional/contextual complex; v) morphemic meaning is encoded, lexical meaning is memorized; vi) lexical meaning is not the starting point of semantic analysis but an intermediate level.

A semantic account which can be formalised (Gasiglia, Nemo & Cadiot 2001) by saying that the senses *s* of both lexemes and non memorised uses are only **the results** of a function *f*, which may be described as:

f(morpheme, construction, context) = s

having as a result the necessity for the linguist to admit the existence of three semantic stocks, namely:

- morphemes, which are form/signification pairs that exist independently of the construction and context in which they are inserted;

- constructions which are form/interpretation pairs that exit independently of the morphemes used (Goldberg 1995);

- lexemes which minimally are morpheme/construction pairs and are associated with lexicalised meanings (i.e. senses),

and to avoid any methodology taking for granted that lexemes are the inputs of linguistic processes. Otherwise, as we shall see now, one has no other choice but to **list** whatever is not predictable from these units and their intuitive senses, i.e. to list most of the lexicon.

2.3.2 The Origin of Listemes

Within contemporary morphology, words are indeed to be listed if they are *somehow* irregular, i.e. whenever:

- the "input" is problematic;

- there are no rules to account for the observed pattern;

- the semantic output is unpredictable from the semantic input.

As a result, most of the lexicon has to be listed, even though:

- whenever the "input" is problematic it may be observed that i) the frequency of problematic inputs may be as high (see French *en-*) as to be considered "regular"; ii) problematic inputs often produce unproblematic "outputs", with meanings very similar to those of non-listed words (e.g. *s'enticher* and *s'enamourer*); iii) bound bases are simply not studied before listing is decided (e.g. *re-tali-ation*).

- whenever there are no rules allowing prediction, the existence of categorially problematic listemes is as high as to be considered "regular" (e.g. *buteur*, *footballeur*), the listed "outputs" often share the same global meaning as non-listed one (*pétrolier, chimiquier*), the possibility of exocentric derivation is not tested.

- whenever *semantic drift* or semantic incoherence is postulated, its existence is based on the hypothesis of the existence of a primary meaning directly accessible through intuition and familiarity, no study of polysemy is ever made and the "input"/"output" semantic relationship is believed to be stable.

On the contrary, within the Linguistic Semantic framework presented here, instead of the classical view according to which lexemes are the inputs of morphology:

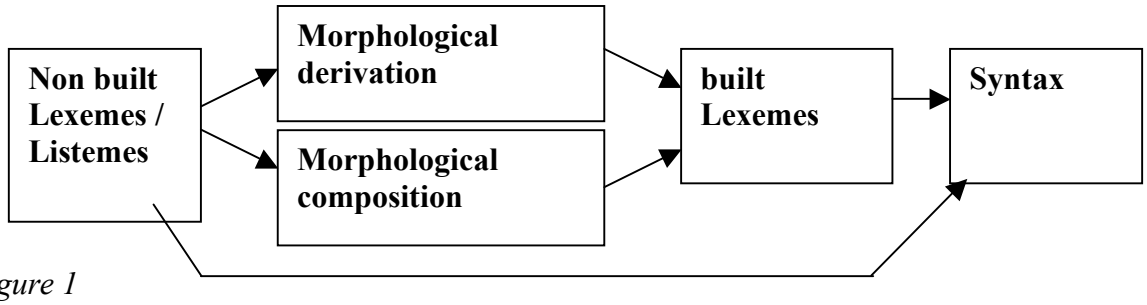

Figure 1

what we have is a morpheme/lexeme/construction distinction:

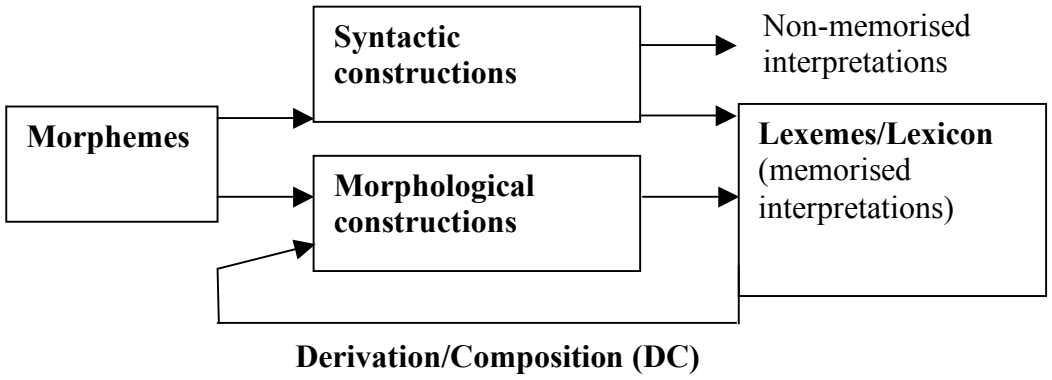

Figure 2

according to which morphemes and constructions are the inputs of the process which must be described and the lexicon its output.

2.3.2 Listing and Falsification

The main difference between the two models contrasted here consists in the predictions they make about the existence or inexistence of listemes: since listemes (which are not basic words) are those lexical units whose existence or meaning cannot be predicted as a result of lexeme-based word-formation process, it is easy to see that what figure 1 describes is in fact the DC re-entering arrow of figure 2, and that whatever is directly generated by inserting morphemes into constructions will have to be listed.

One of the main problems associated with the view presented in figure 1 is that even though much time has been dedicated to a theorisation of what listemes are (Di Sciullo & Williams 1987), no quantification of listing has ever been made on real data. So that even if criteria have been defined so as to decide whether a given word (or meaning of a word) has to be listed, the fact that using these criteria often leads to the listing of a large number of word-forms and most word meanings is not acknowledged at all. Tenants of the classical view never indeed acknowledge (or simply mention) that applying the criteria defined in 2.3.1. to the diversity of uses of morphemes like *table* or *coll* leads to the listing of most words-forms in which these morphemes are inserted – namely words such as *collecte*, *collection*, *collision*, *collusion*, *accolade* – and also to the listing of a large part of the meanings of the remaining word-forms, such as the *taking off* sense of *décoller* for a plane, . It should on the contrary be remarked that given the semantic methodology adopted by tenants of the classical view, which combines intuition and prototypicality to isolate the supposedly basic/true meaning of a word, almost all the lexicon should be listed, since the average frequency of these intuitive meanings is never higher than 10–20 % of the uses of a word (see Gasiglia, Nemo & Cadiot for an illustration of this about the French word *balayer*).

Tenants of figure 2 on the other hand, who are today's leading semanticians, do not accept the idea of such a global frequency (80–90%) of "semantic drift", "semantic bleaching" and "idiosyncrasy". They also refuse the underlying methodology and the consequence it has on our understanding of the generation of meanings. For instance, if we consider the description of the meaning of *but* proposed by B. Fraser (1998) as an illustration of the shortcomings of such an intuitive and unexplanatory methodology, it is clear that the linguist has no choice but to pick out a prototypical meaning (on a "trust me" basis), to declare it basic *("the core meaning of* but *is to signal simple contrast, nothing more, and the speaker will select it when intending to highlight a contrast"*), to declare this description unfalsifiable even when it is directly falsified (e.g. saying that "*even if one cannot find two specific areas of contrast between the direct S2 and S1 messages, the messages may nevertheless be contrasted in one of several ways"* in order to account for uncontrastive uses such as "*Paul is brillant but so is John*"), then to add that defining definition is impossible *("I can offer no precise definition of what qualifies as a Contrastive Discourse Marker*) and finally, concerning other uses of *but* (i.e. concerning the so-called semantic drift), to declare that "*I am not treating other uses of* but *such as found in: «* All *but* one left today », « There was no doubt *but* that he won », « it has not sooner started *but* it shopped», « He was *but* a poor man », « I may be wrong *but* I think you are beautiful ». *Whether or not they could be included under my analysis is left open".* The important point is of course to understand that what Fraser is saying here about *but*, whose morphemic account was presented above, is due to the

same attitude adopted by many morphologists (especially in the generative framework), according to which: i) semantics should be intuitive and not explanatory; ii) semantics should not take more time than five minutes; iii) whatever might require more than 5 minutes should be left over and forgotten.

It is indeed quite clear that he same thing holds in morphology when we have to account for words like *décoller* (to take off, for a plane), *collecte*, *collection*, *collision*, *collusion*, *accolade*: the fact that only a minority of these words and/or a minority of the uses of these words are predictable from the ordinary meanings of the words *colle* (glue) or *coller* (stick) is not a trace of the fact that "*the lexicon is like a prison – it contains only the lawless, and the only thing that its inmates have in common is lawlessness*" nor of the fact that since it "*is simply a collection of the lawless, there neither can nor should be a theory directly about it*" (Di Sciullo & Williams 1987), but a trace of the fact **that the ordinary meanings** of the words *colle* and *coller* are **only local interpretations of the indications encoded by the morpheme** [*coll*], interpretations which, since they are not coded, are completely unable to block the generation of new interpretations in new contexts and the generation of new lexemes in new constructional positions. Listing all these uses and words in the lexicon because they cannot be predicted from these local interpretations is thus the equivalent of describing the lexicon as an endless sense enumerative lexicon criticised by Bouchard (1995) and Pustejovsky (1995), a list of leaves unrelated one to another by any branch or tree.

Thus, adopting the "Morpheme and Lexeme Hypothesis" is a way to avoid adopting the Generative Morphology's hypothesis of an "Ungenerated Lexicon":

- if the only thing morphology can say about words such as *tabler* (to bank on), *rétablir* (to restore, to re-establish, to reinstate), *se rétablir* (to recover, to return), *tableur* (spreadsheet) *tableau* (board, chart, table, instrument panel, dashboard), or idioms like *se mettre à table* (to tell everything), *dresser un tableau de la situation* (to paint the picture of the situation) – which are not compositionally predictable from the meaning of the lexeme *table* (as a *piece of furniture*) they seem to include – is that these words (and/or meanings) are listed, unrelated and have to be learnt one by one;

- if the only thing morphology can tell us about all the uses of the noun *table* (dining table, changing table, arithmetic charts, book contents, editing bench, etc.) is to describe them in terms of semantic drift, semantic bleaching or homonymic degrouping;

- if the only words predictable from the DC arrow are the words *tablée* and *s'attabler*, whose interpretation clearly presuppose the lexicalised meaning of *table* as a *piece of furniture*.

then it would have to be acknowledged that morphology and common sense have exactly the same (un)explanatory power, morphology being unable to account for anything more than what immediate intuition would.

Consequently, it seems clear that instead of assuming that "*there neither can nor should be a theory"* about listemes, linguists should understand that (most) listemes are a direct **empirical falsification** of the classical view, and that they should be considered

as such. It may be the case that listemes are the nightmare of combinatorial/categorial morphology, but it may also be the case that listemes are in fact an open window into the reality of word-formation processes:

Limiting morphology to the DC arrow implies that describing the generation of the French noun *re-spir-ation* from the verb *re-spir-er* is possible, but that accounting for *re-spir-er* itself is impossible. It also implies that it is impossible:

- to account for the production of words/lexemes like French *rot-ation*, or *obstruction* which are listemes since their input is problematic (*roter* is not a French verb, *obst* has no syntactic autonomy);

- to account for words/lexemes like French *dé-coll-er* (to take off, for a plane) which are also listemes since their semantic output is not predictable from the sense of an existing lexeme,

- to account for the existence of listemes such as *re-cycl-er*, *but-eur* or *chimiqu-ier* which cannot be predicted by any combinatorial WFR but are the result of the general existence in French of constructions associated with a pattern of exocentric interpretation;

All of which can be predicted within the "Morpheme and Lexeme Hypothesis" (see Nemo 2001a) presented here, if a correct description of the signification of the morpheme and of the variety of possible morphological constructions is provided, i.e. if polymorphy and morphological flexibility are considered.

Polymorphy, i.e. the fact that *morph* and *form*, *coul*-er and dé-*goul*-in-er, *rot-ation* and *tor-dre* are two forms of the same morpheme, allows to account for a large part of problematic bases and for the inexistence of semantic drift: for instance even though the meaning of the word *amorphe* (inactive) in French is not compositional in a DC sense, it may be directly predicted from a lexicalised interpretation of *forme* in *être en forme, avoir la forme.* This leads to the conclusion that instead of the systematic postulation of semantic drift, we should rather consider the reality of the kinds of formal drift involved in polymophy, namely that the same signification (and sometimes meaning) can be associated with various (related) "signifiants", and study such polymorphy as a regular phenomenon[2].

Flexibility of morphological constructions is another issue, directly related to the interpretability constraint: comparing words like *chimiqu-ier* (chemical tanker)*, but-eur* (striker) and *re-cycl-er* which are all listemes because their base is syntactically distinct from what we would expect (respectively an adjective and not a noun, a noun and not a verb, a noun and not a verb), with predictable words such as *pétrolier* (tanker), *tu-eur* (killer) or *re-pousser* (to push back), allows us to understand (and therefore to predict) the possibility of generating such "listemes", either because:

- French systematically admits the possibility of an exocentric interpretation of the base of affixed words; so that instead of requiring a forcefully nominal or verbal

[2] Systematic polymorphy in French consists mainly in: i) alternating non-voiced and voiced consonants (p/b, k/g, t/d) as in *coul/goul*; ii) permutation/metathesis, as in *uple*, *plu*, *pul*, *supplément* and *plus*; iii) expansion, as in *-able* and *habile*; iv) alternating au/al, ou/ol, etc. as in *autre*, *alterner, haut, altitude*; v) combining any of the former, as in *obst/stop*.

base, it allows the nominal or verbal head to remain implicit and one of its arguments to become the base, thus replacing the "[[*pétrole*]$_N$]$_{NP}$ (oil) transported in *pétrolier* (tanker)" interpretation of *pétrolier*, by the exocentric "[[produits]$_N$[*chimiques*]$_A$]$_{NP}$ (chemicals) transported in *chimiquier"*interpretation, or replacing the [[*tuer*]$_V$]$_{VP}$ interpretation of *tu-eur* by the [[marquer]$_V$[des *buts*]$_O$]$_{VP}$ interpretation of *but-eur.*

or because:

- the meaning of the noun *cycle* actually unifies/coincides with the indications encoded by the morpheme *re*, thus allowing *recycle* to be semantically interpretable, and hence well-formed.

So that in both cases, it is possible to show that the criteria proposed in order to decide what had to be accounted for and what does not have to be accounted for, lead morphologists to overlook the existence and diversity of word-formation processes, and to ignore the fact that their model is heavily falsified.

2.3.2 Word-formation ≠ Derivation/Composition

Much more important in a certain way, the distinction proposed in figure 2, allows the linguist to draw a clear line between true **derivational and compositional processes** (represented in figure 2 by the re-entering arrow DC), which really take *lexemes* with their lexicalised meaning **s** as inputs for word-formation, and **insertional processes** which take morphemes – and the **f(m, cstr, ctxt)** functions they encode – as inputs and force a new (contextual and constructional) interpretation of the morpheme (which may become lexicalised if the use becomes a usage).

In other words, it is of considerable importance for the linguist to be able to distinguish between an horizontal relationship between two words (or meanings), such as the relation between two leaves of the same branch (polysemy or polycategoriality), and a vertical/derivational relationship, in which an indisputable transfer of meaning occurs.

As we have seen there is only a horizontal relationship between the different lexemes *but* in English, and none of the *almost, without, only,* etc. meanings associated with these lexemes may be said to be derived from a "basic" connective meaning nor be the result of any bleaching of this supposedly basic meaning. The same thing holds in morphology when we have to account for words like *décoller* (to take off, for a plane), *collecte*, *collection*, *collision*, *collusion*, *accolade*: the fact that only a minority of these words and/or a minority of the uses of these words are predictable from the ordinary meanings of the words *colle* (glue) or *coller* (stick) is not a trace of the fact that "*the lexicon is like a prison – it contains only the lawless, and the only thing that its inmates have in common is lawlessness*" nor of the fact that since it "*is simply a collection of the lawless, there neither can nor should be a theory directly about it*", but a trace of the fact **that the ordinary meanings** of the words *colle* and *coller* are **only local interpretations of the indications encoded by the morpheme** [*coll*], interpretations which, since they are not coded, are completely unable to block the generation of new interpretations in new contexts and the generation of new lexemes in new positions. Listing all these uses and words in the lexicon because they cannot be predicted from

these local interpretations is thus the equivalent of describing the lexicon as an endless sense enumerative lexicon (see Bouchard 1995; Pustejovsky 1995).

All this will lead us to a single conclusion: if we are to account for the existence of so-called listemes, we need to understand that: i) it is always possible for a speaker to use a morpheme in a new construction or a new context, thus creating new interpretations and freeing him/her from conforming with the senses associated with previous uses; ii) the well-formedness of a new lexeme is not a matter of applying existing rules to an existing lexical stock, as in figure 1 above, but mainly a matter of **interpretability**.

If such is the case, and if indeed most listemes (such as *multiple*, *rotation*, *rétablir*) are semantically and constructionaly interpretable despite their not being formed by the kind of combinatorial mechanisms (WFRs) morphologists were looking for (i.e. the **DC** arrow of figure 2), then it means that understanding what interpretation is about, how contextual unification works and what the relationship between non-categorial morphemes and categorially defined lexemes is, should be a central issue in morphology.

A word like *re-tali-ation* in English is not well-formed because it can be produced by general combinatorial rules, but only, as the French word *re-cycle*, because the signification of *re-* is to indicate the existence of two anti-oriented processes p1 and p2, and because the meaning of the base (*Talion*), as opaque as it may seem, does unify with these indications (losing an eye as a p2 punishment for the p1 crime of making somebody lose an eye, etc.). Word-formation and word-construction, it seems, is thus cemented by interpretation.

This conclusion directly falsifies one of the founding postulates of the Chomskian approach to linguistics, according to which: (i) a linguistic theory should describe the combinatorial mechanisms which allow the generation of new sentences or new words; (ii) there is no way semantic considerations could help us explain the combinations that are acceptable and the ones that are not. It seems quite clear on the opposite, that it is impossible to account for the generation of the lexicon, i.e. of a large part of the first task, without taking into account the fact that, ultimately, the cement of word-formation is interpretation, i.e. without dropping the second assumption.

What we need hence in order to be able to account for the generation of the lexicon, to avoid listing most of it and to integrate the repeated demonstration, within Linguistic Semantics, of the fact that no combinatorial information is attached to (encoded by) the basic semantic units of a language (i.e. morphemes), is to understand that **well-formedness** in morphology **is not a matter of grammaticality but a matter of interpretability**. And thus that what we need is a theory of interpretation consistent with the empirical observations of data-based studies, and a methodology which strictly forbids the use of introspection and intuition in the definition of what has to be accounted for and of how to account for it. Ultimately, the choice is not between doing morphology with or without semantics, as Chomsky seemed to suggest, but between doing it with bad or good semantics. Any semantic model whose ambition (and result) is not to account for the generation of the lexicon, i.e. of new senses and new lexemes, should be abandoned in morphology if morphology wants to be something else than a formalisation of the shortcomings of common sense.

References

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

Anshen, F. & M. Aronoff (1988), *Producing Morphologically Complex Words*, "Linguistics" 26, 641–655.

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

Aronoff, M. (1993), *Morphology by Itself: Stems and Inflectional Classes*, Cambridge (Mass.), The MIT Press.

Aronoff, M. & F. Anshen (1998), *Morphology and the Lexicon*, in A. Spencer & A. Zwicky (eds.), *The Handbook of Morphology*, Oxford, Blackwell, 248–271.

Baayen, H. (1991), *Quantitative Aspects of Morphological Productivity*, in G. Booij & J. van Marle (eds.), "Yearbook of Morphology" 1991, Dordrecht, Kluwer, 109–149.

Baayen, Harald (1993), *On Frequency, Transparency and Productivity*, in G. Booij & J. van Marle (eds.), "Yearbook of Morphology" 1992, Dordrecht, Kluwer, 181–208.

Baudouin de Courtenay, J. (1895), *Versuch einer Theorie phonetischer Alternationen: Ein Kapitel aus der Psychophonetik*, Strassburg/Crakow.

Benveniste, E. (1954), *Problèmes de reconstruction en sémantique*, in *Problèmes de linguistique générale*, 2, Paris, Gallimard.

Bouchard, D. (1995), *The Semantics of Syntax*, Chicago, Chicago University Press.

Broschart, J. (1997), *Why Tongan does it differently?: categorial distinctions in a language without nouns and verbs*, "Linguistic typology" 1, 123–165.

Bybee, J.L. (1988), *Morphology as Lexical Organization*, in M. Hammond & M. Noonan (eds.), *Theoretical Morphology: Approaches in Modern Linguistics*, San Diego, Academic Press.

Bybee, J.L. & P. Hopper (2001)(eds.), *Frequency and the Emergence of Linguistic Structure*, Amsterdam, John Benjamins Publisher.

Corbin, D. (1987), *Morphologie dérivationnelle et structuration du lexique*, Max Niemeyer Verlag, Tübingen.

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

Dixon, R & A. Aikhenvald (2002), *Word: A Cross-linguistic Typology*, Cambridge, Cambridge University Press.

Ducrot, O, (1987), *L'interprétation comme point de départ imaginaire de la sémantique*, in Dire et ne pas dire ([1972], 1991), Paris, Hermann.

Fradin, B. (1993), *La théorie morphologique face à ses choix*, "Cahier de Lexicologie", 1993/2, 63, 5–42.

Fradin, B. (1999), *Syntaxe et morphologie*, "Histoire Epistémologie et Langage" XXI, 2, 7–44.

Gasiglia, N., F. Nemo & P. Cadiot (2001), *Meaning and the Generation of Reference*, in P. Bouillon (ed.), *Generative Approaches to the Lexicon*, Université de Genève.

Fraser, B. (1998), *Contrastive Discourse Markers in English*, in A.H. Jucker & Y. Ziv (eds.), *Discourse Markers: Descriptions and Theory*, Amsterdam, John Benjamins.

Gil, D. (2000), *Syntactic categories, cross-linguistic variation and universal grammar*, in P. Vogel & B. Comrie (eds), 173–216.

Goldberg, A. (1995), *Constructions*, Chicago, University of Chicago Press.

Jalenques, P. (2000), *Contribution à l'étude du préfixe RE en français contemporain: pour une analyse compositionnelle du verbe regarder*, PhD Thesis, Paris 7.

Levin, B. & M. Rappaport Hovav (1998), *Morphology and Lexical Semantics*, in A. Spencer & A. Zwicky (eds.), *The Handbook of Morphology*, Oxford, Blackwell, 248–271

Lieber, R. (1992), *Deconstructing Morphology: Word Formation in Syntactic Theory*, Chicago, University of Chicago Press.

Matthews, P.H. (1991), *Morphology: An Introduction to the Theory of Word-Structure*, Cambridge, Cambridge University Press.

Nemo, F. (1998), Enfin, encore, toujours entre indexicalité et emplois, in *Actes du Colloque de Bruxelles. Sens et fonctions*, XIV, Tübingen, Max Niemeyer Verlag.

Nemo, F. (1999), *The Pragmatics of Signs, The Semantics of Relevance, and The Semantic/Pargmatic Interface*, in *The Semantics-Pragmatics Interface from Different points of View*, CRiSPI Series, Amsterdam, Elsevier Science, 343–417.

Nemo, F. (2001a), *Morpheme Semantics and the Autonomy of Morphology. The Stable Semantics of (Apparently) Unstable Constructions*, in M. Andronis et al. (eds.), *CLS 37: The Panels. Papers from the 37th Meeting of the Chicago Linguistic Society*, vol. 2. Chicago, Chicago Linguistic Society

Nemo, F. (2001c), *Pour une approche indexicale (et non procédurale) des instructions sémantiques*, "Revue de Sémantique et de Pragmatique" 9/10, 195–218

Nemo, F. (2002a), *But (and mais) as morpheme(s)*, "Delta" (Sao Paulo), n°18–2.

Nemo F, (to appear), *Discourse Particles as Morphemes and as Constructions*, in K. Fischer (ed.), *Approaches to Discourse Particles*, Amsterdam, Elsevier.

Plag, I. (1998), *The Polysemy of -ize Derivatives. The Role of Semantics in Word Formation*, in G. Booij & J. van Marle (eds.), "Yearbook of Morphology" 1997, Dordrecht, Kluwer, 219–242.

Pustejovsky, J. (1995), *The Generative Lexicon*, Cambridge (Mass.), The MIT press.

Spencer, A. (1991), *Morphological Theory: An Introduction to Word Structure in Generative Grammar*, Blackwell Publisher.

Sweet, H. (1876), *Words, logic and grammar*, Transactions of the Philological Society, in W.F. Bolton & D. Crystal (eds.)(1969), *The English language: essays by linguists and men of letters*, vol. 2, Cambridge, Cambridge University Press

Vogel, P. & B. Comrie (eds)(2000), *Approaches to the Typology of Word Classes*, Berlin-New York, Mouton de Gruyter.