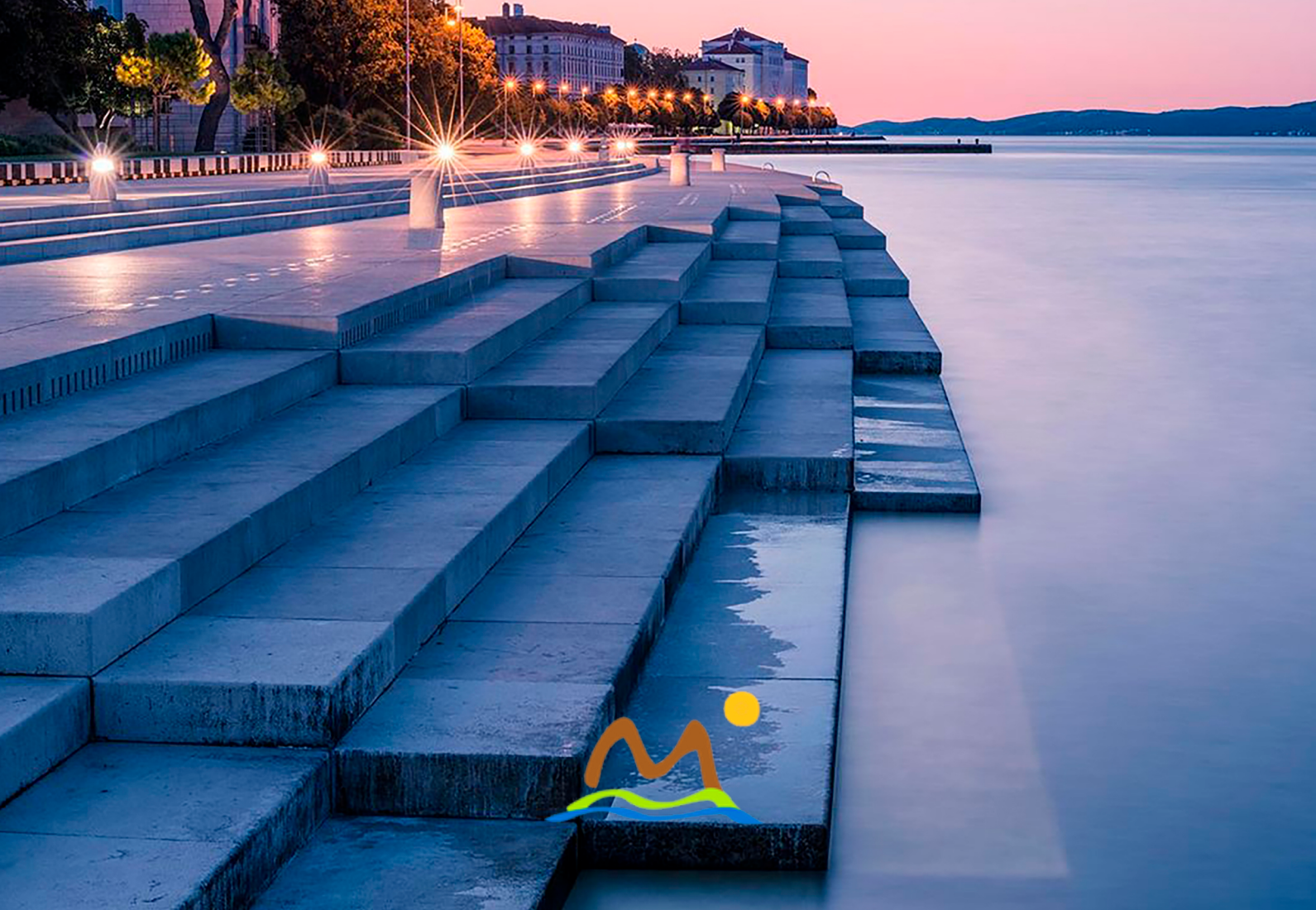




Lexicon, Lexemes, Lexicalization

MMM14 Online Proceedings

Edited by:

Marco Angster

Jenny Audring

Nikos Koutsoukos

Frane Malenica

PUBLISHED AT UNIVERSITY OF PATRAS

2026

Table of contents

Foreword	iii
Compounds or phrases: The analysis of nominal constructions with <i>à</i> in French, Pius ten Hacken, Sara Aufinger-Scheed & Maria Koliopoulou	1
Expanding the boundaries of word-formation – What German phrasal compounds tell us about the lexicon, Katrin Hein-Antonioli.....	11
Refining quantitative methods for Diachronic Construction Morphology: a study of three compound families in French, Jan Radimský & Florence Villoing.....	23
Superlative events. Intensifying verbs through the elative suffix <i>-issimo</i> , Luisa Troncone & Flavio Pisciotta.....	39
A manner of evaluating: Italian deverbal evaluative motion verbs, Luisa Corona, Luisa Brucale & Egle Mocciaro.....	57
Investigating the productivity of evaluative morphology: Diminutives of loanwords in Italian, M. Silvia Micheli & Matteo Pellegrini.....	71
A case-study on proper nouns formed by <i>-exit</i> in Contemporary Italian, Mauro Le Donne ...	85

Foreword

Ever since the first Mediterranean Morphology Meeting (MMM) was held in Mytilene, Greece, in 1997, the conference has been an event that morphologists look forward to every other year. Organized in beautiful locations around the Mediterranean and providing a focused, informal setting for presenting and discussing morphological research, MMM welcomes junior and senior scholars alike. The single-session setup facilitates interaction among researchers and gives young linguists the opportunity to present their work at a moderately sized conference.

This volume collects contributions from the 14th MMM, held at the University of Zadar in Croatia in June 2025. It continues the series of proceedings from the thirteen earlier meetings: 1997 (Mytilene, Greece), 1999 (Lija, Malta), 2001 (Barcelona, Spain), 2003 (Catania, Sicily), 2005 (Fréjus, France), 2007 (Ithaca, Greece), 2009 (Nicosia, Cyprus), 2011 (Cagliari, Sardinia), 2013 (Dubrovnik, Croatia), 2015 (Haifa, Israel), 2017 (again Nicosia, Cyprus), 2019 (Ljubljana, Slovenia), and 2022 (Rhodes, Greece).

The Permanent Committee of MMM consists of Jenny Audring (University of Leiden), Nikos Koutsoukos (University of Patras), and Francesca Masini (University of Bologna). The team took over from the original founders and organizers Geert Booij (Leiden University), Angela Ralli (University of Patras), and Sergio Scalise (University of Bologna). However, the lion's share of the work always rests with the local organizers. MMM14 was made possible by the excellent organizing committee chaired by Mia Batinić Angster.

The thematic session of MMM14 was titled “Lexicon, Lexemes, Lexicalization”. Keynote speakers were Ewa Dąbrowska (FAU Erlangen-Nürnberg), Claudio Iacobini (Roma Tre University), and Francesco Gardani (University of Zurich). The conference was preceded by the workshop “Untangling the bundle: language contact at the word level and below”. The proceedings contain contributions from both the thematic and free sessions.

The editors of this volume wish to thank the local organizing committee, especially Mia Batinić Angster, Metka Bezljaj, Marijana Kresić Vukosav, Jakov Proroković, and Lucija Šimičić; the student volunteers who ensured the conference ran smoothly; all attendees of MMM14; and especially the contributors to these Online Proceedings.

Compounds or phrases: The analysis of nominal constructions with *à* in French

Pius ten Hacken
Universität Innsbruck
pius.ten-hacken@uibk.ac.at

Sara Aufinger-Scheed
Translator (Self-employed)
office@saraaufinger.com

Maria Koliopoulou
National and Kapodistrian University of Athens
mkoliopoulou@gs.uoa.gr

1. Introduction

Nouns in French can be modified by a prepositional phrase headed by *à*. This is a very general preposition, for which a single translation cannot be given. Some examples are given in (1).

- (1)
- | | |
|-------------------------|--|
| a. un livre à moi | ‘a book of mine’ |
| b. une manif à Paris | ‘a demo in Paris’ |
| c. une vache à lait | ‘a cow for milk’, i.e. a milk cow |
| d. une machine à écrire | ‘a machine for writing’, i.e. a typewriter |

Semantically, the examples in (1) fall into two clearly distinct classes. In (1a-b), we have syntactic constructions in which the PP-modifier adds to the description of the item referred to by the noun. In (1c-d), we have expressions that are used to name concepts. In Germanic languages, they are often translated as compounds. In (1c), the correlation with an English compound is transparent. In (1d), English has a different naming motive, but the German compound *Schreibmaschine* ‘write-machine’ corresponds more closely to the French expression. Apart from the difference in the order of the French construction and the English and German compounds, the main difference is the presence of the preposition in French.

The question we investigated is whether the similarity of expressions such as (1c-d) to compounds in English and German is strong enough for them to qualify as compounds according to the definition of compounding given by ten Hacken (2021: 10), quoted in (2).

- (2)
- A compound is a word with a structure $[X\ Y]_Z$ or $[Y\ X]_Z$, such that:
- The denotation of Z is a subset of the denotation of Y;
 - The range of possible relations between X and Y depends on the semantics of X and Y;
 - X does not have independent access to the discourse.

The definition in (2) makes a compound a binary structure. In (2a), Y is identified as the head, but as both orders are stated, Y can be the right-hand or the left-hand component. For (2b), we have to consider the relation between the head and the non-head in the meaning of the entire compound. The question is to what extent the construction contributes to the meaning of Z. In

(2c), the question of whether the non-head introduces a separate concept is the decisive factor. If it does, the non-head can be, for instance, the antecedent of a pronoun.

In considering the role of *à* in expressions such as (1c-d), the first question is what role prepositions can play in the context of the definition in (2). This question is addressed in general terms in section 2. In order to investigate the class of nominal expressions with *à* in more detail, we collected a set of expressions and analysed it. Section 3 describes how we collected the set of expressions and section 4 presents our analysis. In section 5, we turn to the question of how the conditions (2b) and (2c) can be applied to these expressions. Finally, section 6 places our conclusion in a broader theoretical environment and compares the consequences of taking (2) as the definition of compounding with other approaches.

2. Genitives, prepositions and compounds

The connection between genitives and compounds has long been observed. Paul (1886: 279) suggests that the linking element in German compounds has its origin in the genitive. In English, genitives are often used when the non-head refers to human beings. Table 1 illustrates this correlation for English, German and French.

Table 1. English, German and French equivalents for *peace treaty* and *children's book*

<i>English</i>	<i>German</i>	<i>French</i>
peace treaty	Friedensvertrag	traité de paix
children's book	Kinderbuch	livre d'enfant

In the German translation of *peace treaty*, the non-head *Friedens* is in a form that is identical to the genitive of *Frieden* 'peace'. However, the final *-s* is generally analysed as a linking element (cf. Koliopoulou 2014, 2020). In *Arbeitsamt* 'job centre', the same linking element is used, although *Arbeit* 'work' is a feminine noun that never has an *-s* in its inflectional paradigm. In *children's book*, the form is ambiguous, as illustrated in (3).

- (3) a. She wrote a successful children's book.
b. The children's book was a gift from the council.

In (3a), the combination with the indefinite article requires an interpretation of *children's book* as a compound. The adjective modifies the entire expression *children's book*. In (3b), an interpretation along the lines 'the book the children have' is plausible. In this case, a modifying adjective, e.g. *favourite*, will go between *children's* and *book*. The French equivalents in Table 1 both have the preposition *de*. In the absence of a morphological genitive, the construction with *de* can be considered a periphrastic genitive. Grevisse (1980: 1146) gives as one of its core senses "Un rapport d'appartenance, de dépendance" ('a relation of belonging, of dependence'). Ten Hacken (2013: 102) illustrates the parallelism with the genitive in (4).

- (4) a. car factory
b. usine d'automobiles 'factory of cars'
c. fabryka samochodów 'factory cars_{GEN}'

The equivalents of the English compound in (4a) are the French *de* expression in (4b) and the Polish morphological genitive construction in (4c).

The question that needs to be asked at this point is to what extent the behaviour of French *de* is typical of French prepositions in general. In this context, it is worth looking at (5), an example from COCA, in which three compounds are underlined.

- (5) Diane Sawyer mentioned the immigration vote in Arizona, the abortion vote in Mississippi, and the collective bargaining vote in Ohio but neglected to mention the vote on Obamacare in Ohio.

For the French translation of the underlined compounds in (5), the expressions in (6) can be used.

- | | | |
|-----|---------------------------------------|--|
| (6) | a. vote sur l'immigration | 'vote on (the) immigration' |
| | b. vote sur l'avortement | 'vote on (the) abortion' |
| | c. vote sur la négociation collective | 'vote on (the) collective negotiation' |

In the translations in (6), the article is in brackets, because it is part of the French expression but cannot be used in the English one. In French, it marks the generic interpretation of the noun it specifies. Although both (4b) and (6) are translations of English compounds, there is a significant difference between them. In (4b), the combination of the nouns *usine* 'factory' and *automobile* 'car' is used to name a concept. The role of the preposition *de* is limited to connecting these nouns. The relation between *usine* and *automobile* can be inferred from the meanings of these nouns and from world knowledge. The correct interpretation depends on the concept that is named by the expression. This is the same as for compounds such as (4a). In (6), however, the preposition *sur* specifies the relation between the head noun *vote* and the noun designating the theme of the vote. Although the expressions in (6) can be used in the translation of the English compounds underlined in (5), they do not get their meaning in the same way. The relation between the nouns is expressed by the preposition in (6), but understood without being expressed in (5). Therefore, the expressions in (6) do not fulfill the condition in (2b) and are not compounds according to the definition in (2).

We can now reformulate our original question about the status of naming expressions with *à* such as (1c-d) as the question whether *à* patterns with *de* or with *sur*. In order to answer this, we have to look at a larger set of expressions with *à* and consider the relationship between the nouns it connects.

3. Compilation of a sample

In order to investigate the class of naming expressions of the type illustrated by (1c) and (1d), we needed a set of examples. The question of representativeness does not play a role here, because representativeness is a relationship between sets. We can investigate the electoral preferences by means of a representative poll because the electorate is a well-defined set of people. For each person it can be unambiguously determined whether they are in the electorate or not and the size of the electorate is known exactly. For French nominal expressions with a modifier introduced by *à*, we cannot do this. It is not possible to enumerate all relevant expressions and there is no empirical way to determine whether a particular expression is (correct) French. Therefore, the main criterion we used as a starting point is to collect a large set of examples.

We started with a set of nominal multi-word entries retrieved from the French *Wiktionnaire*. This is a set of 51,613 items. From this set, we extracted expressions in which *à* occurred as a word. We also included contractions with the definite article, *au* in the masculine singular and *aux* in the plural. In this way we obtained a set of 3,765 expressions. This is the base set which we analysed by formal means. From this set, we extracted a random sample of 350 expressions for a more detailed semantic analysis.

Taking the *Wiktionnaire* as a starting point means that the expressions were judged worth a description in a dictionary. The nature of the resource does not guarantee that such decisions

were based on professional lexicographic expertise, but users contributing to the content apparently perceived these expressions as names for concepts. Therefore, the data set does not command any immediate authority, but it tends to reflect intuitions by users that are keen enough to volunteer a contribution.

4. Analysis of the sample

In a first step, we carried out a formal analysis, classifying our set of 3,765 expressions with nouns modified by means of *à* according to the three criteria in (7).

- (7) a. Is *à* followed by a definite article?
- b. Does the head consist of more than one word?
- c. Does the modifier consist of more than one content word?

Examples such as (1c-d) yield negative answers to all questions in (7). The examples in (8) illustrate the kind of expressions that result in positive answers.

- (8) a. marché aux poissons ‘market with the fish_{PLUR}’, i.e. fish market
- b. chemin de fer à crémaillère ‘road of iron with cog’, i.e. cog railway
- c. chapeau à trois cornes ‘hat with three horns’, i.e. tricorne

In (8a), the definite article *les* is contracted with the preposition *à*. Articles occurred in this pattern in 21.0 % of the expressions in our sample. In these cases, the definite article is used to mark a generic interpretation of the following noun. A generic interpretation is also typical of the non-head of compounds in English, as the translation *fish market* shows. There is no obvious semantic difference between the French expressions with and without an article.

In (8b), the head noun is *chemin de fer* ‘railway’, itself a strongly lexicalized multi-word expression that qualifies as a compound according to the definition in (2). It has its own entry in dictionaries such as Académie française (1992-2024). Cases such as (8b) are relatively rare. They account for 6.5 % of our data set. As only formal criteria were used at this stage, we also found cases such as (9).

- (9) bâton de rouge à lèvres ‘stick of red for lips’, i.e. lipstick

In (9), the string that precedes *à* is not a constituent. The head noun is *bâton* ‘stick’ and the modifier is *rouge à lèvres*. This means that (9) should actually be excluded from our data set, but this is only possible by a semantic analysis, which we only carried out for our sample of 350 expressions. In this sample, we excluded 7 expressions (2 %) for this reason.

In (8c), it is the constituent following *à* that consists of more than one word. Such cases are much more common than the ones of the type in (8b). They make up 34.1 % of our data set. Although it is in principle possible that some of these should be excluded for the same reason as (9), we did not find any such examples in our sample of 350. What we did find is that many of these expressions refer to species of animals and plants. Some examples are given in (10).

- (10) a. inule à feuilles en glaive ‘inula with leaves in sword’, i.e. sword-leaved inula
- b. canard à bec jaune ‘duck with bill yellow’, i.e. yellow-billed duck

In both examples in (10), the head designates a group of species and the modifier narrows this down. In (10a), *inule* ‘inula’ is a genus of plants with about 80 species. In (10b), *canard* ‘duck’ is a non-monophyletic group, i.e. it cuts across the taxonomy reflecting the evolutionary

relationships between species. In these names for species, the modifier is headed by a noun which is itself modified typically by a prepositional phrase, as in (10a), or an adjective, as in (10b). Table 2 summarizes the relation between the frequency of the appearance of a modified noun after *à* and the use for designating a species of animals or plants in our sample of 350.

Table 2. Species names and the complexity of modifiers with *à*

<i>Semantic class</i>	<i>Simple noun</i>	<i>Noun with modifier</i>	<i>Total</i>
Animal	19	54	73
Plant	19	20	39
Other	166	44	210
Total	204	118	322

Table 2 shows both the remarkable frequency of species names, amounting to 32 % of the sample, and the preference for complex modifiers, resulting in 62.7 % of the expressions with a complex modifier being species names.

The numbers in Table 2 do not include the expressions in which the modifier is not headed by a noun. This explains that the total number of expressions is less than 350. Examples with verbs and adjectives are given in (11).

- (11) a. table à repasser ‘table for ironing’, i.e. ironing board
 b. démarrage à froid ‘moving_off with cold’, i.e. cold start

Examples of the types in (11) are relatively rare. In our sample of 350, there were 16 expressions with an infinitive verb (4.6%) as in (11a) and 12 with an adjective (3.4%) as in (11b). As they require a separate analysis but the number of items in the sample is too small to serve as a basis for this, we excluded them from our analysis here. We plan to return to these constructions in a future study.

We can summarize the analysis so far as follows. The naming construction with *à* introducing a modifier of a noun is quite flexible. The preposition may or may not be followed by a definite article. The modifier can be a noun, a noun with an adjective or prepositional phrase, a verb, or an adjective. We only consider cases in which the modifier is headed by a noun. The noun modified by *à* may also be complex, but this is relatively rare and in some of these cases, the structure is different, so that the expression has to be excluded.

5. Application of the definition

On the basis of the data we collected, we can now apply the conditions in the definition of compounding in (2). As the headedness in (2a) has already been used in the selection of the data, the focus is on the conditions (2b) and (2c).

The central question in the application of (2b) is the degree to which the preposition *à* contributes to the meaning of the entire expression. Two examples illustrating different relationships between the nouns connected by *à* are given in (12).

- (12) a. moulin à légumes ‘mill for vegetables’, i.e. food mill
 b. moulin à billes ‘mill with balls’, i.e. ball mill

Both (12a) and (12b) designate instruments, but the role of the referent of the modifier in the use of the instrument is quite different. In (12a), we have a hand-operated kitchen utensil for puréeing food, especially vegetables. This means that the item is a *moulin à légumes* whether or not the *légumes* ‘vegetables’ are in or near it. The *légumes* only indicate the purpose of the

instrument, which is why we translated *à* as ‘for’ in (12a). In (12b), we have a professional instrument for grinding dry substances into powder by means of small balls. It is typically used, for instance for grinding cacao in chocolate manufacturing. The *billes* ‘balls’ are a crucial component of the instrument and without them something cannot be a *moulin à billes*. Therefore, we translated *à* as ‘with’ in (12b).

The two relationships in (12) are clearly different. At the same time, they are quite vague, as can be seen when considering the examples in (13).

- (13) a. moulin à papier ‘mill for paper’, i.e. paper mill
 b. moulin à vent ‘mill for wind’, i.e. windmill

The examples in (13) correspond to the well-known English pair used by Allen (1978) to illustrate the Variable R Condition. Although *for* can be used in the description, the relation between the *moulin* and the modifier is quite different to the one in (12a). In order to specify the relation, we need to add a further predicate. In (12a), we can say that the instrument is ‘for processing’ *légumes*, i.e. the *légumes* are the input. In (13a), we have an instrument ‘for producing’ *papier* ‘paper’, i.e. the *papier* is the output. In (13b), the *vent* ‘wind’ is the source of energy. We might describe it as ‘for being driven by’ *vent*. This shows that classifying the expressions as involving *for* or *with* in the relation does not fully determine the nature of the relation.

When we quantify the use of the *for* and *with* relations, we have to take into account the particular frequency of the names for species of animals and plants as noted in section 4. Table 3 therefore gives separate figures for the full set and for the set without the species names.

Table 3. Frequency of *for* and *with* relations

<i>Type of relation</i>	<i>All N à N</i>		<i>Excluding species names</i>	
for	62	19.7 %	56	27.6 %
for or with	8	2.5 %	8	3.9 %
with	178	56.5 %	80	39.4 %
other relations	67	21.3 %	59	29.1 %

The interpretation of the relation between the two components of a compound depends on the type of object that is named by it. In the case of species of animals or plants, a frequent naming strategy is to choose a hyperonym and add information about a particular feature of the body that is different from other species covered by this hyperonym. The examples in (10) illustrate this. This naming strategy explains why the *à* construction is often used in this context. It also explains why the preposition can often be paraphrased as ‘with’. Therefore, the fact that more than half of the expressions in our sample can be described as having ‘with’ as the type of relation, as seen in Table 3, should not be taken as crucial for the analysis of the *à* construction. If species names are excluded, the different relation types are much closer in frequency, as shown in the right-hand column. Even without excluding them, however, it is clear that it is not just the preposition that determines the relationship.

In a few cases, both the ‘for’ and the ‘with’ relation are applicable. Two examples are given in (14).

- (14) a. pot à lait ‘jar with/for milk’, i.e. milk jug
 b. capsule à crayon ‘capsule with/for pencil’, i.e. fuel rod case

In both examples in (14), a container is marked by what it contains or is meant to contain. In (14a), it is a household item. Whether it is filled with milk or not, it is a *pot à lait*. However,

also a different *pot* can be referred to by (14a) if it contains *lait*. Therefore, as long as one of ‘for’ or ‘with’ applies, (14a) can be used as a designation. In (14b), we have a specialized item in the domain of nuclear energy (FranceTerme, 2016: *capsule à crayon*) which has the same dual use. Here, *crayon* ‘pencil’ is used metaphorically for the fuel rod of a nuclear reactor.

For the purpose of determining the range of possible relations, the last category in Table 3 is particularly relevant. Two examples of this category are given in (15).

- (15) a. *pâte à sucre* ‘dough from sugar’, i.e. sugar paste
 b. *nid à rats* ‘nest of rats’, i.e. rats’ nest

In (15a), the *pâte* ‘dough’ is made out of *sucre* ‘sugar’. Here, ‘for’ does not apply, but it would also be wrong to characterize the relation as ‘with’. In order to create the *pâte*, the *sucre* is transformed, so that there is no *sucre* as a substance in the *pâte*. In (15b), the proper characterization is that the *rats* have built the *nid* ‘nest’. There is no implication of ‘with’, because it remains a *nid à rats* when there are no *rats* in it. It would also be wrong to characterize the relationship as ‘for’, because this would imply that someone might have built it for use by rats, as in *cowshed*. In fact, (15b) is used mainly metaphorically. However, the metaphor is based on the shape and structure of what rats build, not on the purpose of the result.

Taking together the information about the relation between the two components in the *N à N* expressions, we can say that the contribution of the preposition *à* to the meaning is much more limited than in the cases with *sur* in (10). There is a certain bias towards certain relations, but, as shown in (12), it does not go into a uniform direction. With ‘for’ and ‘with’, there are two poles of attraction. Moreover, as Table 3 shows, there are quite a few examples that do not go into either direction. Therefore, we argue that condition (2b) is met.

Let us now turn to (2c). Ten Hacken (2021: 11) uses the pronominal reference test for this condition, which was introduced originally in ten Hacken (1994: 74-77). In (16), this test is applied to the expressions in (1c) and (12a).

- (16) a. Jean a une vache à lait_i. Il le_{*i} boit chaque matin.
 ‘Jean has a cow for milk. He it drinks every morning.’
 i.e. Jean has a milk cow. He drinks it (the milk) every morning.
 b. Marie a un moulin à légumes_i. Elle les_{*i} achète au marché.
 ‘Marie has a mill for vegetables. She them buys at the market.’
 i.e. Marie has a food mill. She buys them (the vegetables) at the market.

In (16a), *lait* ‘milk’ is not introduced into the discourse, so that it is not available for pronominal reference. This effect is the same as for the corresponding English compound. In (16b), *légumes* ‘vegetables’ is in a similar position. In the English translation, there is no direct equivalent to *légumes*, because the English compound has a different non-head. As a result, *them* in the translation cannot be interpreted at all. However, if *them* is replaced by *it* to make the pronoun compatible with the English non-head *food*, the same effect is obtained. It is interesting to note that the French expressions with an article, e.g. (8a), behave in the same way, as illustrated in (17).

- (17) Au marché aux poissons_i, ils_{*i} sont toujours frais.
 ‘At the market with the fish_{PLUR}, they_{PLUR.MASC} are always fresh.’
 i.e. At the fish market, it (the fish) is always fresh.

In (17), *poissons* ‘fish_{PLUR}’ is not introduced into the discourse. The pronoun *ils* could in an appropriate context be used to refer to a kind of fish, e.g. *saumons* ‘salmons’, if a suitable form

had been introduced in the preceding context, but not to *poissons* as it appears in (17). This is exactly the same for *it* in the English translation. It should also be noted that this pronominal reference effect depends on the preposition. For the examples in (10), pronominal reference is possible, as shown in (18).

- (18) Après le vote sur l'immigration_i, celle-ci_i a diminué sensiblement.
 'After the vote on the immigration, it has diminished markedly.'
 i.e. After the immigration vote, it (the immigration) diminished markedly.

The French coreference in (18) is entirely natural. In English, it is natural in the gloss, but ungrammatical in the translation in which a compound is used. In view of (17) this cannot be attributed to the definite article. The form of the pronoun *celle-ci* deserves a comment. It is a demonstrative rather than a personal pronoun. Although it is more natural in (18) than *elle*, the latter is also possible. Replacing the personal pronouns by the corresponding forms of the demonstrative in (16) and (17) does not improve the grammaticality. In (17), using *ceux-ci* instead of *ils* results in an ungrammatical sentence.

On the basis of the examples (16) to (18), we can conclude that the N *à* N construction fulfills condition (2c) of the definition of compound. As the fulfillment of the other conditions was shown before, this means that the N *à* N construction as illustrated in (1c), (8a) and (12) to (15) is a compounding construction according to the definition in (2).

6. Discussion

In the literature on French word formation, there is no general consensus on what constitutes a compound. Grevisse (1980: 118) proposes the definition in (19).

- (19) Un mot, quoique formé d'éléments graphiquement indépendants, est *composé* dès le moment où il évoque dans l'esprit, non les images distinctes répondant à chacun des mots composants, mais une image unique.
 'A word, even if formed from orthographically independent elements, is a compound as soon as in the mind it evokes a single image rather than the distinct images corresponding to each of the component words.'

The definition in (19) identifies a much broader set of items than the one in (2). Among the examples he gives are the expressions in (20).

- (20) a. hôtel de ville 'hotel of town', i.e. town hall
 b. chou-fleur 'cabbage-flower', i.e. cauliflower
 c. contredire 'against_speak', i.e. contradict

Because of the way (19) is formulated, it is not possible to derive tests from it as a way of determining the status of borderline cases. It appeals to unaided intuitive judgements. In the context of Grevisse's (1980) descriptive grammar, this is not disturbing. A descriptive grammar does not have to present a general theory of language. Examples such as (20c) show that the notion of compound is quite wide. The definition in (2) identifies in much more detail how a compound gets its meaning and restricts the set of expressions classified as compounds.

A different approach is taken by Fradin (2009). Following Corbin (1992: 50), he adopts the principle that compounds are morphological constructs and as such may not be built by syntax. He elaborates this in (21) as a set of two conditions (2009: 418).

- (21) a. The sequence of elements could have been generated by syntax.
 b. The phonology of the sequence of elements is that of a syntactic unit.

Fradin's proposal is that, if both conditions in (21) are fulfilled, we are dealing with a syntactic construct, which excludes a classification as a compound. As an illustration of the application of this principle, he gives (22).

- (22) Le cavalier mit pied à terre.
 'The horseman set foot on ground.', i.e. The horseman dismounted.

The fact that the sequence *pied à terre* occurs in a sentence such as (22) can be used to exclude the noun *pied-à-terre* in the sense of 'a flat or house where one does not live permanently, but only occasionally'. This means that the conditions in (21) take no account of the semantics of the expression. Fradin (2009: 432-433) also discusses the opposition in (23).

- (23) a. espace fumeurs 'space smokers', i.e. smoking area
 b. espace pour (les) fumeurs 'space for (the) smokers'

Expressions such as (23a) fail the tests in (21). They contrast with syntactic expressions such as (23b). The bracketed article in (23b) indicates an optional element. As generic NPs are often expressed with a definite article in French, the article is more common in French than in English. Fradin (2009: 433) states that "the expressions in (33) also exist with a preposition heading a PP". His (33) is a set of examples including our (23a). This suggests that Fradin takes (23a) and (23b) to be the same expression. We contend that this is not the case. Whereas (23a) refers to a concept by naming it, (23b) describes a particular *espace*. This is related to the intuition underlying Grevisse's definition in (19). The way (23b) gets its meaning is similar to what we saw in (6). The preposition, *pour* in (23b), *sur* in (6), determines the relation between the two nouns.

For the construction we discuss here, Fradin (2009: 419) gives the examples in (24).

- (24) a. avion à réaction 'plane with reaction', i.e. jet plane
 b. serpent à sonnettes 'snake with bells', i.e. rattlesnake

According to Fradin (2009: 419), "[b]y the same token, the nouns in [(24)] must not be considered as compounds". Here, "the same token" refers to the reasoning for *pied-à-terre* in view of (22). He refers to an earlier edition of Booij (2012: 85), who includes French N à N constructions among "lexical expressions in Romance languages [that] are sometimes incorrectly called compounds although they have in fact a phrasal form".

What is striking about the approach by Fradin (2009) and Booij (2012) is, first of all, their complete exclusion of semantic considerations. In our view, language cannot be analysed without taking into account the meaning of expressions. Secondly, they present their concept of *compound* as the only possible proper concept.

We argue for an approach that is more terminological in orientation. In terminology, a term can be defined in different ways and the choice of a definition should be based on how useful the concept delimited by the definition is in formulating theoretical insights. In our view, the definition in (2) identifies a useful concept. It is more constrained than the *composé* as defined by Grevisse (1980) in (19). By means of the additional constraints, a more precise set of expressions can be identified. These expressions share not only the nature of their place in the mental lexicon, but also the way they are associated with a meaning.

In his discussion of the borderline between compounds and phrases, Booij (2012: 84-87) does not invoke a definition, but only states for a number of constructions that they “must be considered phrases”. This includes also genitive constructions of the type illustrated in (3a). Booij (2009) argues for a distinction between naming and description that is independent from the one between words and phrases. What we aim to do is to identify a cross-linguistic class of expressions that have certain properties in common. The properties described in the definition in (2) are more important than the name *compound* assigned to this class. It remains the case, however, that prototypical N N compounds in languages such as German and English belong to this class. The selection of properties in the definition and its application to a range of constructions in different languages gives insight into the nature of naming.

For the French N *à* N construction illustrated by (1c), we have shown that it belongs to the class of constructions identified by the definition in (2). We call these compounds. They constitute a proper subset of the expressions identified as *composés* by Grevisse (1980). This result is significant from the perspective of how word formation is used as a naming device. In ten Hacken & Panocová (2024) the theoretical background for this perspective is elaborated in more detail.

References

- Académie française. 1992-2024. *Dictionnaire de l'Académie française, neuvième édition* (4 vol.). Paris: Fayard.
- Allen, Margaret Reece. 1978. *Morphological Investigations*. PhD Dissertation, Storrs CT: University of Connecticut.
- Booij, Geert. 2009. Phrasal names; A constructionist analysis. *Word Structure* 2: 219-240.
- Booij, Geert. 2012. *The Grammar of Words: An Introduction to Linguistic Morphology*, 3rd edn. Oxford: Oxford University Press.
- Corbin, Danielle. 1992. Hypothèse sur les frontières de la composition nominale. *Cahiers de grammaire* 17: 25-55.
- Fradin, Bernard. 2009. IE, Romance: French. In Lieber, Rochelle & Štekauer, Pavol (eds.). *The Oxford Handbook of Compounding*. Oxford: Oxford University Press, pp. 417-435.
- FranceTerme. 2016. *FranceTerme*, Paris: Commission d'enrichissement de la langue française, <https://www.culture.fr/FranceTerme>.
- Grevisse, Maurice. 1980. *Le bon usage: Grammaire française avec des Remarques sur la langue française d'aujourd'hui, onzième édition revue*, Paris/Gembloux: Duculot.
- ten Hacken, Pius. 1994. *Defining Morphology: A Principled Approach to Determining the Boundaries of Compounding, Derivation, and Inflection*, Hildesheim: Olms.
- ten Hacken, Pius. 2013. Compounds in English, in French, in Polish, and in General. *SKASE Journal of Theoretical Linguistics* 10. 1): 97-113.
- ten Hacken, Pius. 2021. The Nature of Compounding. *Cadernos de Linguística* 2(1): 1-21.
- ten Hacken, Pius & Panocová, Renáta. 2024. *Word Formation as a Naming Device*. Edinburgh: Edinburgh University Press.
- Koliopoulou, Maria. 2014. How close to syntax are compounds? Evidence from the linking element in German and Modern Greek compounds. *Rivista di Linguistica* 26(2): 51-70.
- Koliopoulou, Maria. 2020. Linking elements in German compounds: a morphological analysis in comparison with Greek. In Georgiafentis, Michalis; Giannouloupoulou, Giannoula; Koliopoulou, Maria & Tsokoglou, Angeliki (eds.). *Contrastive Studies in Morphology and Syntax*, London: Bloomsbury, pp. 102-118.
- Paul, Hermann. 1886. *Principien der Sprachgeschichte, zweite auflage*. Halle: Niemeyer.

Expanding the boundaries of word-formation – What German phrasal compounds tell us about the lexicon

Katrin Hein-Antonioli
Leibniz Institute for the German Language (IDS)
hein@ids-mannheim.de

1. Introduction

Phrasal compounds like *Leck-mich-am-Arsch-Tag* ‘kiss-my-ass day’ or *Beckham-im-Gedächtnis-Grätsche* ‘Beckham-in-mind straddle’ are characterized by the integration of a syntactical unit (an entire sentence or syntactic phrase) into a complex word. Even if they deviate from the prototypical pattern of compounding and (among others) violate the Principle of Lexical Integrity (cf. Spencer 1991: 42), there is strong empirical evidence that we are dealing with a productive process of German word-formation which is by no means irregular or marginal (cf. Hein 2015). The sketched-out divergence from canonical compounding can come with expressive effects (cf. Meibauer 2007) and justifies the labeling of phrasal compounds as a phenomenon of extravagant morphology in the sense of Haspelmath (1999).

By integrating a phrasal component into a word, phrasal compounds expand the boundaries of word-formation. Thus, we are dealing with an interface phenomenon between morphology and syntax, and also with a perfect example for the interplay of word-formation and multi-word expressions (cf. Steyer & Hein 2018). In the linguistic literature, phrasal compounds are part of a general debate (which is not limited to German) about “the role of phrases in the word formation process” (Radimský 2015: 33) and also provide interesting insights into the relation between creativity and routine.

Due to their status as an interface phenomenon, phrasal compounds are challenging for morphological theorizing and for an adequate conception of the lexicon. This observation is the starting point for this article. With reference to corpus data from written and spoken German¹, I will pursue the question of what phrasal compounds tell us about the lexicon.

Following Gaeta (2016), phrasal compounding is a phenomenon that can be understood by invoking “two different senses of the term lexicon”: the Bloomfieldian sense which “generally refers to the set containing any sort of entrenched or idiomatic expressions” and a second sense which “refers to the ‘infinite set of potential [...] lexemes for any given language’” (Gaeta 2016: 109; Aronoff 1994: 10). As I will argue in this article, this observation clearly holds for phrasal compounds whose non-head can be analyzed as a fixed multi-word unit, e.g. *Alles oder nichts-Devise* ‘all-or-nothing motto’, but is less convincing for formations whose non-head is formed ‘on the fly’ (e.g. *Guck-ich-hab-es-wieder-aus-dem-Bett-geschafft-Rap* ‘Look-I-made-it-out-of-bed-again rap’. The latter are clearly attested in corpus data from written and spoken German and play a significant role for the pattern of phrasal compounding, even if lexicalized non-heads dominate quantitatively (cf. Hein & Antonioli 2025; Gaeta 2016: 136). Then the

¹ More precisely, I will refer to newspaper texts from the German Reference Corpus (DeReKo; <https://www.ids-mannheim.de/en/digspra/pb-s1/projects/corpus-development/>), to song lyrics provided in the song corpus (<https://songkorpus.de/>) as well as to spoken interaction from the FOLK corpus (<https://dgd.ids-mannheim.de/DGD2Web/jsp/Welcome.jsp>).

question arises where those non-heads “made out of the blue” (Gaeta 2016: 136) are built and what consequences they bring for an adequate conception of the lexicon.

Before I turn to this question in more detail in the main part of this article (section 3), I will first provide an overview of phrasal compounding in German (section 2) sketching out the main properties of the phenomenon as well as its treatment in the literature. In the conclusion (section 4), I will also make a connection to Construction Grammar respectively Construction Morphology (Booij 2010). This theoretical framework seems to be best suited to describing an interface phenomenon like phrasal compounding.

2. Phrasal compounding in German

The aim of this section is to carve out the characteristics of phrasal compounding in German, with particular regard to the notions of creativity and extravagance.

2.1 Central properties and ignorance in the research literature

With the integration of a syntactic unit into a complex word, phrasal compounds form a special type of determinative compound; more prototypical instances are words like *Baumhaus* ‘tree house’. Nonetheless, their status as genuine compounds is uncontroversial (cf. Schlücker 2012; Hein 2011), as they are characterized by a binary structure and a classical determinans-determinatum relation between the immediate constituents. Thus, the head (*Rap* in Fig. 1) determines the main semantic class as well as the formal properties (part of speech, gender, inflection) of the corresponding complex word, while the non-head (*Ich finde du bist weak* in Fig. 1) specifies the meaning of the head semantically:

[“Ich-finde-du-bist-weak”]_{nonhead} - [Rap]_{head}
 (‘I-think-you-are-weak rap’)

Figure 1. Binary structure of phrasal compounds

Nonetheless, phrasal compounds are less frequent than prototypical N+N-compounds, and they also go against some word-specific properties like Lexical Integrity (cf. Lieber & Scalise 2006; cf. Hein 2011: 355–356 for an overview).

According to the research literature, phrasal compounding is mainly used to form nouns (cf. Lawrenz 2006), i.e., nominal phrasal compounds are the most productive subtype of phrasal compounding. Also, the only empirical study available for German to date focuses on nominal phrasal compounds (cf. Hein 2015). However, it is claimed in the research literature that adjectival phrasal compounds can also be formed (cf. Günther, Kotowski & Plag 2020). This aligns with the attestation, though very rarely, of adjectival phrasal compounds in the data of my recent investigation of word-formation products in German song lyrics (cf. Hein 2023), e.g. *nach Oma-Schlüpfer-riechendes (Miststück)* ‘smelling-like-granny-panties (bitch)’ (Corpus of Song Lyrics: Die Ärzte/MISC).

Please note that the phenomenon is not limited to expressive text types like song lyrics. Rather, it is also widespread in less marked text types like newspaper texts. Moreover, I want to highlight that not all phrasal compounds are as expressive as the previous example. Rather, with the help of corpus data one can observe that the expressivity (and also the extent of creativity) of phrasal compounding varies gradually. On the one hand, there are completely lexicalized phrasal compounds like *Armeleuteessen* ‘poor-people meal’, which have their own lexical entry and thus are not very remarkable. The same holds for compounds containing a modifier with numerals (e.g. *1000-Stunden-Jahr* ‘1000 hours year’, a very productive subtype of phrasal compounding. On the other hand, there is a variety of phrasal compound whose modifier is a complex sentence formed on the fly. Corresponding formations like *Guck-ich-*

hab-es-wieder-aus-dem-Bett-geschafft-Rap ‘Look-I-made-it-out-of-bed-again rap’ are very expressive and display a high degree of creativity. As those examples indicate, expressivity does not only depend on lexicalization, but is also influenced by the syntactic complexity of the first constituent.

Accordingly, the construction grammatical modelling carved out in Hein (2015) claims that phrasal compounding can be traced back to two main sub-patterns: One relatively unmarked construction which has (almost) no expressive effects and one marked construction whose instantiations are quite expressive and can fulfill specific communicative functions (cf. also Hein & Antonioli 2025: 68–69).

In the past, the distinctive features of phrasal compounding were often equated with irregularity. Accordingly, phrasal compounds were long considered an unproductive, marginal phenomenon that violates central principles of word-formation. As a consequence, they have been marginalized in the research literature (cf. Meibauer 2003). Even worse, their existence was denied in some models of generative grammar (cf. Hein 2011).

However, even if phrasal compounds violate the Principle of Lexical Integrity and deviate from the prototypical pattern of compounding, the first broad corpus-based study of nominal phrasal compounding in newspaper texts (Hein 2015, cf. 2017 for a shorter version in English) has shown that we are dealing with a productive process of German word-formation which is by no means irregular or marginal. Follow-up studies on song lyrics (Hein 2023) and spoken German (Hein & Antonioli 2025) confirm those findings.

Recently, there is a growing (and theoretically motivated) interest in phrasal compounding which is not limited to German, cf. Goldberg & Shirtz (2025), Hein & Antonioli (2025), Gutzmann & Turgay (2024), Günther, Kotowski & Plag (2020), Hein (2017, 2015, 2011), Gaeta (2016), Meibauer (2015, 2007, 2003), Trips & Kornfilt (2017), Lieber & Scalise (2006), etc.

Due to their status as an interface phenomenon between morphology and syntax, phrasal compounds are challenging for morphological theorizing. For example, they pose a severe problem for models of generative grammar that assume a linear interaction between morphology and syntax. As I have shown in previous work (Hein 2015; cf. also Goldberg & Shirtz 2025: 307), a construction-grammar approach is fruitful and suitable for capturing a phenomenon like phrasal compounds (cf. section 4).

2.2 Creativity & extravagance

As stated above, phrasal compounds display a classical determinans-determinatum relation between their immediate constituents which suggests that they are indeed compounds. Nonetheless, due to the integration of a syntactical unit into a complex word we are dealing with a phenomenon that somehow diverges from the ‘routine in compounding’, i.e. from canonical compounds like *Baumhaus*. As already mentioned above, this can cause expressive effects (cf. Meibauer 2007) and justifies the labeling of phrasal compounds as a phenomenon of ‘extravagant morphology’ (cf. Haspelmath 1999; Eitelmann & Haumann 2022a). Accordingly, the use of phrasal compounds may be in part explained by Haspelmath’s (1999: 1063) ‘maxim of extravagance’ (“Talk in such a way that you are noticed”).

„The maxim of extravagance seeks to grasp the observation that communicative behaviour does not boil down to achieving the goal of being understood [...] In order to be noticed, language users would also overstep boundaries and ‘choose new ways of saying old things’.” (Eitelmann & Haumann 2022b: 2)

In a recent publication by Goldberg and Shirtz (2025: 313), the witty and sarcastic potential of phrasal compounds is confirmed by various large-scale surveys:

“Sentences with PALs [Phrase-As-Lemma; K.H.] are also judged to be more witty [...] and more sarcastic [...]. Remarkably, these factors are evident even in extremely frequent PAL tokens when compared to paraphrases [...]. To be sure, we are not claiming that all or most PAL tokens imply a close intimacy with the comprehender, are laugh-out-loud funny, or are witheringly sarcastic. But naive English speakers recognize that the construction adds a dash of these rhetorical flourishes.”

This is in line with the claim mentioned above that the degree of expressivity in phrasal compounds is gradual (cf. section 2.1).

As the notion of creativity used above is quite fuzzy, I would like to clarify it with the help of Schmid (2025) who addresses the question of “What is Creative to Whom and Why?” and makes a crucial reference to Sampson’s (2016) distinction between F-creativity and E-creativity. While more general definitions use adjectives like ‘original’ or ‘unconventional’ to capture creativity, Sampson brings the criteria of regularity into play: While *fixed creativity* (F-creativity) captures word-formation phenomena “that are based on existing word-formation patterns”, *extending creativity* (E-creativity) “denotes original formations that go beyond regular applications of productive patterns [...] and tend to be salient enough to be recognized as being motivated by the intention to be creative” (Schmid 2025: 98).²

Turning back to the phenomenon in focus, I would argue that in one sense we are dealing with a case of E-creativity. This applies if we relate phrasal compounding to the more productive pattern of ‘normal’ compounding (e.g. *Baumhaus*). Under this perspective, phrasal compounds “go beyond regular applications of productive patterns” (Schmid 2025: 98). This is due to the fact that the non-head position of phrasal compounds is filled by a syntactic unit, while the non-head position in a general schema for compounding normally is reserved for lexical categories. On the other hand, one could also argue that phrasal compounding itself is a productive pattern of German word-formation and thus can be seen as a manifestation of F-creativity.

The observation that phrasal compounds are salient in the sense that the intention to be creative is apparent is in favor of their treatment as a case of E-creativity (cf. Schmid 2025: 98). This at least applies for one of the two sub-patterns which have been introduced above, more precisely for the more expressive variant of phrasal compounds (e.g. *Guck-ich-hab-es-wieder-aus-dem-Bett-geschafft-Rap*). For sure, the intention of being creative is less apparent in formations like *1000-Stunden-Jahr*.

To conclude this excursion to creativity, it has to be highlighted that phrasal compounds – regardless of whether you label them E-creative or F-creative, display a “balance of originality and effectiveness”, i.e. their creativity does not jeopardize “effortless mutual understanding” between speaker and hearer (Schmid 2025: 98). Among others, their use in advertising underlines that phrasal compounds are a very effective way of communicating.

3. Phrasal compounds and the lexicon

What do phrasal compounds tell us about the lexicon? And what implications do they bring for more general questions like “What is a lexeme?” or “How lexical is morphology?” (Gaeta 2016)?

² For a critical discussion of the dichotomy between E-creativity and F-creativity and extensions of Sampson’s (2016) distinction cf. Stumpf (under review).

The theoretical ‘attractiveness’ which I have sketched out in section 2.1 will be used below to discuss how an adequate conception of the lexicon has to look like from the perspective of phrasal compounding. For that purpose, I will build on the approach developed by Gaeta (2016). This approach has emerged in a broader context, with the aim to clarify “how the lexicon has to be concretely conceived” (Gaeta 2016: 110). Gaeta states that the relation between morphology and lexicon has received relatively little attention to date. In what follows, he argues for a “consistently lexical morphology” (Gaeta 2016: 110). In this context, among other phenomena, he also explicitly refers to phrasal compounding.

3.1 Gaeta’s “Lexicality of the Input” principle

Crucial for Gaeta’s (2016) approach is that it combines two traditionally competing views by assuming an interplay of two different conceptions of the lexicon. This idea is captured in the so-called „Lexicality of the Input Principle” (LIP): On the one hand, the term lexicon has to be understood in a Bloomfieldian sense. Bloomfield defines the lexicon (= Lex₁) as a collection of any sort of entrenched or idiomatic expressions. Consequently, the lexicon in a Bloomfieldian sense „consists of a list of form/meaning arbitrary pairs coming from different procedures and sources as it contains all expressions presenting idiomatic traits, including those larger than one word“ (Gaeta 2016: 116). On the other hand, the term lexicon denotes an infinite set of potential lexemes for any given language (= Lex₂). In this understanding, the lexicon consists of the (infinite) set of all potential lexemes in a given language and therefore results exclusively from lexeme-formation procedures“ (Gaeta 2016: 116).

The ‘double nature’ of the lexicon sketched above is comparable to the idea of the so-called ‘construction’ in Construction Grammar (e.g. Goldberg 2006), where the strict line between lexicon and syntax is blurred and Lex₁ lexemes and Lex₂ lexemes (in the sense of Gaeta 2016) are equally part of the Constructicon, the inventory of constructions.

A consequence of combining two senses of the term ‘lexicon’ is that Gaeta’s approach “accounts for complex words formed on the basis of possible but unattested words (i.e. pure Lex₂-Lexemes)” and at the same time “is not biased against complex units being stabilized in the Lex₁“ (Gaeta 2016: 131–132).

Gaeta (2016: 134) lists different phenomena which can be explained by the interaction between Lex₁ and Lex₂ mentioned above: „Even if lexicalization might at first sight appear as an inertial force [...] we will see [...] that its interaction with the Lex₂ provides us the key to understanding a number of phenomena which have been traditionally considered marginal [...]”. One of these phenomena is phrasal compounding.

In this paper, I will argue that Gaeta’s observation clearly holds for phrasal compounds whose non-head can be analyzed as a fixed multi-word unit, but is less clear for non-heads which seem to be formed ‘on the fly’, i.e. which are not part of the Lex₁.³ In what follows, this hypothesis is discussed on the basis of corpus data from written and spoken German.

3.2 The ‘lexical status’ of phrasal modifiers from a corpus-based perspective

As several corpus-based investigations clearly indicate (Hein & Antonioli 2025; Hein 2023, 2015), the modifier of a phrasal compound is not necessarily a fixed multi-word unit. Rather, the modifiers vary in terms of their fixedness:

³ To my mind, modifiers formed ‘on the fly’ are not part of the Lex₂ either because they are syntactical units.

- (1) Fixed multi-word expression
Weniger-ist-mehr-Motto
‘Less-is-more credo’
- (2) Multi-word expression formed ‘on the fly’
Okay-wir-sind-nur-ne-Woche-da-Abschnitt
‘Okay-we’re-only-there-for-a-week section’
- (3) Completely lexicalized phrasal compounds (with own dictionary entry)
Armeleuteessen
‘Poor-people meal’

On the one hand, the modifier can be formed by a fixed multi-word expression (i), often in the form of a proverb or a slogan (cf. Steyer & Hein 2018), e.g. *Weniger-ist-mehr-Motto* ‘less-is-more credo’. While cases like that are clearly covered by Gaeta’s LIP (cf. section 3.1), this does not apply to cases like (ii) *Okay-wir-sind-nur-ne-Woche-da-Abschnitt* ‘Okay-we’re-only-there-for-a-week section’, an example from spoken German (cf. Hein & Antonioli 2025) where the modifier quite obviously is formed ‘on the fly’. The third case (iii) that has to be mentioned in this context is illustrated by phrasal compounds such as *Armeleuteessen*. The latter is lexicalized from a synchronic point of view and has a separate lexical entry (<https://www.duden.de/rechtschreibung/Armeleuteessen>). Consequently, formations of that type are not meaningful for the question of what phrasal compounds tell us about the lexicon. From the perspective of the lexicon, they have the same status as completely lexicalized N+N compounds such as *Haustür*, even if internally they display a marked structure.

Considering cases like *Okay-wir-sind-nur-ne-Woche-da-Abschnitt*, my hypothesis is that the inventory of phrasal compounds can only be captured in part through the conception of the lexicon outlined by Gaeta (cf. section 3.1).

The systematic use of corpus data allows us to have a look at the quantitative dimension of the lexical status of phrasal non-heads (cf. Table 1) – instead of claiming without any empirical evidence that phrasal compounds only occur with lexicalized non-heads (cf. Hein 2011: 332–334) or categorically assigning the theoretical status of a quotation to phrasal non-heads (cf. Wiese 1996; and more recently Gutzmann & Turgay 2024).⁴ See also Goldberg & Shirtz, who emphasize that the non-heads “very often have grammatical features associated with quotes”, while not all PALs are even hypothetical quotes” (Goldberg & Shirtz 2025: 298–299). A look at corpus data from written German (DeReKo) (cf. Hein 2015) and spoken German (FOLK corpus) (cf. Hein & Antonioli 2025)⁵ reveals that formations with lexicalized modifier are the most frequent subtype of phrasal compounding in both modalities. Lexicalized non-heads occur with a proportion of ~ 51 % respectively 64 %, e.g. *Leck-mich-am-Arsch-Tag* ‘Kiss-my-ass day’.

⁴ In my opinion, it is insufficient and unsatisfactory to regard the syntactic units in phrasal compounds categorically as quotations and not to examine their degree of fixedness in greater empirical depth. Nonetheless, it is true that quotes play a role in phrasal compounding, for example for the realisation of stereotypes described in section 3.3).

⁵ We considered 1.576 types from written German and 269 types from spoken German. Even if those inventories differ clearly in size (as well as the underlying corpora DeReKo and FOLK, cf. Hein & Antonioli 2025 : 71), the results obtained from the qualitative analysis of the two datasets can be compared and related to one another.

Table 1. Lexical status of phrasal modifiers: DeReKo vs. FOLK (cf. Hein & Antonioli 2025: 73)

Lexical status of the modifier	Examples	Proportion in FOLK (%)	Proportion in DeReKo (%)
Fixed	<i>Leck-mich-am-Arsch-Tag</i>	51,3	63,83
Built ‘on the fly’	<i>Wir-spielen-Spiele-Tag</i>	18,6	35,91
Lexicalized phrasal compound	<i>Acht-Stunden-Tag</i>	29,73	0,25

However, it has to be highlighted that non-heads formed ‘on the fly’ like in *Wir-spielen-Spiele-Tag* ‘We-play-games day’ occur with a non-marginal proportion of ~ 19 % (spoken German) respectively of ~ 36 % (written German). The fact that such non-heads are more frequent in written than in spoken German might be due to the fact that their production as well as their understanding involves more effort, which probably makes them more difficult to use in direct interaction.

Lastly, also the aforementioned type of lexicalized phrasal compound is attested in the data, e.g. *Acht-Stunden-Tag*. Interestingly, the proportion of such formations is particularly high in spoken German (~ 30 %), while it is particularly low (~ 0,3 %) in written German. This finding is intuitively expectable: It can be assumed that phrasal compounds which have a separate lexical entry can be produced (and understood) more easily and faster compared to their non-lexicalized counterparts. This for sure favours the use of completely lexicalized phrasal compounds in spoken language (cf. Hein & Antonioli 2025: 73–74).

If one takes empirical data seriously, one must conclude that phrasal compounds with a non-lexicalized modifier are not marginal at all. Even if the operationalisation of a concept such as lexicalization is not trivial⁶, determining the lexical status of the modifier is an important part of the empirical description of phrasal compounding. To my mind, a proportion of ~19 % respectively ~36 % of modifiers built on the fly cannot be ignored in the debate about the theoretical implications of phrasal compounding. Thus, the question arises how we can account for this phenomenon from the point of view of the lexicon.

The hypothesis that I will push forward is that the LIP (Gaeta 2016) might be too strict to explain the entire range of phrasal compounds. Moreover, the situation is even more complex, as there are two kinds of ‘special cases’ that also have to be discussed in the context of lexicalization, namely the role of abstract patterns and stereotypes for the coinage of phrasal non-heads.

3.3 Abstract patterns and stereotypes as blueprints for phrasal modifiers

From the perspective of multi-word units, some modifiers of phrasal compounds can also be classified as instantiations of more abstract underlying linguistic patterns. An example of such pattern is [X *oder* ‘or’ Y] (cf. Steyer 2013):

⁶ The following criteria were used to classify the lexical status of the modifier : (1) ‘fixed’: the modifier is listed in the dictionary (www.duden.de) or occurs recurrently in DeReKo ; (2) ‘built on the fly’: the modifier is not listed in the dictionary and does not occur recurrently in DeReKo ; (3) ‘lexicalized phrasal compound’: the entire compound is listed in the dictionary with an own entry (cf. Hein 2015: 205-209).

[X oder ‘or’ Y]*Zu-mir-oder-zu-dir-Gequatsche* ‘my-place-or-yours chatter’*Alles oder nichts-Devise* ‘all-or-nothing motto’*Friß-oder-stirb-Rede* ‘eat-or-die speech’

The idea illustrated here is that phrasal non-heads like *Zu-mir-oder-zu-dir*, *Alles oder nichts* or *Friß-oder-stirb* can be seen as pair formulas leading back to the more abstract pattern [X oder Y] (cf. Steyer & Hein 2018). At the same time, they are fixed multi-word units in their own right, and thus are supposed to be captured by the Lex₁ component in Gaeta's model (cf. section 3.1). The status of [X oder Y] as pattern, in turn, is externally motivated, i.e. it does not depend on the frequency of corresponding non-heads in phrasal compounds.

The specific quality of modifiers such as *Zu-mir-oder-zu-dir* (and many others) results from their double nature: On the one hand, they are lexicalized multi-word expressions, while on the other hand they are concrete instantiations of more abstract underlying linguistic patterns. Again, the idea of a Constructicon which allows for the representation of both (semi)abstract patterns and concrete lexical structures and also captures the relation between different kinds of linguistic information seems to be most suitable to account for this property of phrasal non-heads.

As a second type of ‘special case’, I would like to outline the role of stereotypes for phrasal compounding. There are many non-heads which are characterized by the use of stereotypes, e.g. *“Ich-kann-Golf-Ski-und-Wandern-und-bin-schöner-als-die-andern”-Franz* ‘I-can-golf-ski-and-hike-and-I-am-more-beautiful-than-the-others Franz’. By using a fictional quotation⁷, i.e. a potential utterance, this phrasal compound makes use of the cliché of Franz Beckenbauer being arrogant.⁸

In general, stereotypes play an important role in phrasal compounding. They are used to characterize a person, an attitude, etc. They can be easily realized by using „original or fictional quotation“ (Finkbeiner & Meibauer 2016: 39; cf. also Pascual, Królak & Janssen 2013) and are very effective:

Quotes can be powerful. They provide an effective way of identifying or characterizing the content of political movements (e.g. *fight racism*), familiar situations (e.g. *it's not you, it's me*), attitudes (e.g. *life is good*), or types of people (e.g. *don't mess with me*). The fact that a well-chosen quote can evoke so much is why quotes are used as bumper stickers, worn on T-shirts, included in email. (Goldberg & Shirtz 2025: 299)

But quotations are not the only way to express stereotypes. This is illustrated in *Dumme-Mädchen-Grill* ‘stupid-girls grill’, an example from spoken German (cf. Hein & Antonioli 2025: 75-76). Here, the stereotype of girls being stupid is played with without the use of a direct quote.

Complex words like *“Ich-kann-Golf-Ski-und-Wandern-und-bin-schöner-als-die-andern”-Franz* are not understandable without the help of world knowledge. Those who do not know Franz Beckenbauer, the famous football player and coach, will not understand the meaning of this complex word. Such extra-linguistic knowledge is not necessarily part of the lexicon. In

⁷ The fact that stereotypes in phrasal compounds are sometimes realized through quotations does not mean, in my view, that the modifier always has the status of quotations (cf. section 3.2).

⁸ Interestingly, in the data of Hein (2015) there are several phrasal compounds with the head *Franz*, all referring to Franz Beckenbauer (<http://uwv.ids-mannheim.de/phraskomp>).

this sense, Goldberg & Shiritz (2025: 292) argue that “speakers who produce a PAL construction are perceived to share more common knowledge with their intended audience than speakers who produce a close paraphrase without a PAL”.

Turning back to our initial question about the nature of the lexicon, I would like to emphasize that *Ich-kann-Golf-Ski-und-Wandern-und-bin-schöner-als-die-andern* is ‘fixed’ with regard to the underlying conceptual pattern, but it is not a typical multi-word unit. Thus, it is not the kind of lexicalized structure which is supposed to be part of the Lex₁ in Gaeta’s (2016) model of the lexicon.

All in all, this section has illustrated that not only lexicalization, but also (semi)abstract patterns and non-linguistic concepts such as stereotypes play a role for phrasal compounding. When addressing the question of what phrasal compounds tell us about the lexicon, such blueprints must be taken into account.

4. Conclusion

In this article, I tried to work out the consequences that phrasal compounds stipulate for an adequate conception of the lexicon.

Phrasal compounding is a productive process of German word-formation and by no means irregular or marginal. Aside from the fact that the empirical investigation of the phenomenon in various text types – newspaper texts (Hein 2015), spoken German (Hein & Antonioli 2025) and song lyrics (Hein 2023) – is very fruitful and brings to light many highly creative formations, phrasal compounding is also particularly suitable for theoretical questions due to its status as an interface phenomenon between morphology and syntax (cf. section 2). This theoretical potential is explored in this article.

In order to show which conception of the lexicon is needed to deal with phrasal compounds, I took Gaeta’s approach (2016) as a starting point (cf. section 3.1). In a next step, I tried to demonstrate on the basis of several corpus-based studies that the LIP is too strict to explain the whole range of phrasal compounds. The reason is that non-lexicalized modifiers are neither part of the Lex₁ nor part of the Lex₂ in the sense of Gaeta (2016) (cf. section 3.1)

Given the remarkable number of modifiers which are built ‘on the fly’, phrasal compounding cannot be fully explained by assuming an interplay of two different conceptions of the lexicon (cf. sections 3.2 and 3.3).⁹ Rather, the formation of phrasal compounds requires, at least in certain cases, an interaction between morphology and syntax. To my mind, phrasal compounds with modifiers built ‘on the fly’ are the result of integrating a syntactically formed unit into a product of word-formation. Consequently, the phenomenon does not provide clear evidence of a “consistently lexical morphology” (Gaeta 2016: 110) (cf. section 3).

Then the question arises how this interaction between morphology and syntax can be embedded in a theoretical framework (cf. Hein 2011). A solution might be a construction grammar approach in which the strict line between morphology and syntax is blurred and where all kind of grammatical knowledge is stored in a so-called Constructicon – whether they are (semi)abstract patterns or concrete linguistic structures at word-level or above it. I already demonstrated that phrasal compounding can be fruitfully modelled within the framework of Construction Grammar, considering a large empirically gained inventory of 1.576 phrasal compounds from the German Reference Corpus (DeReKo) (cf. Hein 2015). Also Gaeta (2016:

⁹ This is unless one assumes that ‘intended creativity’ does not fall within the scope of word-formation patterns and consequently does not need to be covered by them either (cf. Schultink 1961) – a point of view that I do not find satisfactory, given the high proportion of freely formed modifiers (cf. section 3.2).

109) claims that a phenomenon like phrasal compounding can only be explained “if the multifaceted perspective of the Constructicon is adopted as the interface of different modules of grammar”.

All in all, like many other phenomena, phrasal compounds underline the need for a dynamic, “richly textured” concept of the lexicon which is “not just an unstructured list of exceptions” (Jackendoff & Audring 2020: 83), but that contains abstract patterns or schemata in addition to concrete lexical realizations (cf. also Bauer 2017). Moreover, the crucial role of patterns is confirmed: „Patterns are the key to why speakers understand each other in everyday communication in spite of ever-changing lexical material, syntactic variance and strong idiosyncrasies“ (Steyer 2020: 275; cf. Schmid 2014).

With regard to the general role of phrases in word-formation, further investigation is definitely needed. For example, neither adjectival phrasal compounds nor other types of word-formation which involve syntactical units (cf. Lawrenz 2006) have been researched with empirical depth. Since such phenomena provide important insights into the nature of the lexicon and its relation to morphology and syntax, their further investigation is an important issue for future work.

References

- Aronoff, Mark. 1994. *Morphology by itself*. Cambridge: MIT Press.
- Bauer, Laurie. 2017. *Compounds and compounding*. Cambridge: Cambridge University Press.
- Booij, Geert E. 2010. *Construction morphology*. Oxford: Oxford University Press.
- Eitelmann, Matthias & Dagmar Haumann. (eds.) 2022a. *Extravagant morphology*. Amsterdam/Philadelphia: John Benjamins.
- Eitelmann, Matthias & Dagmar Haumann. 2022b. Extravagance in morphology. Introduction. In Matthias Eitelmann & Dagmar Haumann (eds.), *Extravagant morphology. Studies in rule-bending, pattern-extending and theory-challenging morphology*, 1-18. Amsterdam/Philadelphia: John Benjamins.
- Finkbeiner, Rita & Jörg Meibauer. 2016. Boris “Ich bin drin“ Becker’ (‘Boris “I am in” Becker’). Syntax, semantics and pragmatics of a special naming construction. *Lingua* 181. 36-57.
- Gaeta, Livio. 2016. How lexical is morphology? The constructicon and the quadripartite architecture of grammar. In Pavel Štekauer & Livia Körtevelyessy (eds.), *Word-Formation across Languages*, 109–146. Cambridge Scholars Publishing.
- Goldberg, Adele E. 2006. *Constructions at work. The nature of generalization in language*. Oxford: Oxford University Press.
- Goldberg, Adele E. & Shahr Shirts. 2025. The English Phrase-As-Lemma Construction: When a phrase masquerades as a word, people play along. *Language* 101. 291-320.
- Günther, Christine, Sven Kotowski & Ingo Plag. 2020. Phrasal compounds can have adjectival heads. Evidence from English. *English Language & Linguistics* 24. 75-95.
- Gutzmann, Daniel & Katharina Turgay. 2024. Phrasal compounds are quotational compounds. In Anke Himmelreich, Daniel Hole & Johannes Mursell (eds.), *To the left, to the right, and much in between. A Festschrift for Katharina Hartmann*, 319–331. Frankfurt: Goethe University Frankfurt.
- Haspelmath, Martin. 1999. Why is grammaticalization irreversible? *Linguistics* 37(6). 1043-1068.
- Hein, Katrin. 2011. Phrasenkomposita - ein wortbildungs-fremdes Randphänomen zwischen Morphologie und Syntax? *Deutsche Sprache. Zeitschrift für Theorie, Praxis, Dokumentation* 39(4). 331-361.
- Hein, Katrin. 2015. *Phrasenkomposita im Deutschen. Empirische Untersuchung und konstruktionsgrammatische Modellierung*. Tübingen: Narr.
- Hein, Katrin. 2017. Modeling the properties of German phrasal compounds within a usage-based constructional approach. In Carola Trips und Jaklin Kornfilt (eds.), *Further investigations into the nature of phrasal compounding*, 119-148. Berlin: Language Science Press.
- Hein, Katrin. 2023. Beinahe-ums-Leben-kommen-in-Regenpfützen und Chauvi-Macho-Macker-Stuss – kreative Wortbildungen in Songtexten. *Journal for Language Technology and Computational*

- Linguistics* 36(1). Special Issue: *Challenges in Computational Linguistics, Empiric Research & Multidisciplinary Potential of German Song Lyrics*. 73-92.
- Hein, Katrin & Giorgio Antonioli. 2025. Phrasenkomposita im gesprochenen Deutsch an der Schnittstelle von Wortbildungs- und Gesprächsforschung. In Adriano Murelli (ed.), *Gesprochenes Deutsch heute. Variation, Wandel, Sprachkontakt.*, 65–90. Berlin & Boston: De Gruyter.
- Jackendoff, Ray & Jenny Audring. 2020. *The texture of the lexicon. Relational Morphology and the Parallel Architecture*. Oxford: Oxford University Press.
- Lawrenz, Birgit. 2006. *Moderne deutsche Wortbildung. Phrasale Wortbildung im Deutschen: linguistische Untersuchung und sprachdidaktische Behandlung*. Hamburg: Dr. Kovač.
- Lieber, Rochelle & Sergio Scalise. 2006. The Lexical Integrity Hypothesis in a new theoretical universe. *Lingue e linguaggio* 1. 7–32.
- Meibauer, Jörg. 2003. Phrasenkomposita zwischen Wortsyntax und Lexikon. *Zeitschrift für Sprachwissenschaft* 22, 153–188.
- Meibauer, Jörg. 2007. How marginal are phrasal compounds? Generalized insertion, expressivity, and I/Q-interaction. *Morphology* 17(2). 233–259. doi:10.1007/s11525-008-9118-1.
- Meibauer, Jörg. 2015. On “R” in phrasal compounds – a contextualist approach. *Language typology and universals* 68(3). Special issue: *Phrasal compounds from a typological and theoretical perspective*. 241–261.
- Pascual, Esther, Emilia Królak & Theo Janssen. Direct speech compounds. Evoking socio-cultural scenarios through fictive interaction. *Cognitive Linguistics* 24(2). 345-366.
- Radimský, Jan. 2015. *Noun+Noun Compounds in Italian*. České Budějovice: University of South Bohemia in České Budějovice.
- Sampson, Geoffrey. 2016. Two ideas about creativity. Martin Hinton (ed.), *Evidence. Experiment and argument in linguistics and philosophy of language*, 15–26. Bern: Peter Lang.
- Schlückner, Barbara. 2012. Die deutsche Kompositionsfreudigkeit. Übersicht und Einführung. In Livio Gaeta & Barbara Schlückner (eds.), *Das Deutsche als kompositionsfreudige Sprache. Strukturelle Eigenschaften und systembezogene Aspekte*, 1–25. Berlin: de Gruyter.
- Schmid, Hans-Jörg. 2025. What is creative to whom and why? Creativity in word-formation and phraseology against the backdrop of shared Conventions and individual routines. In Sabine Arndt-Lappe & Natalia Filatkina (eds.): *Dynamics at the lexicon-syntax interface. Creativity and routine in word-formation and multi-word expressions*, 97–130. Berlin & Boston: de Gruyter.
- Schultink, Henk. 1961. Produktiviteit als morfologisch fenomeen. *Forum der Letteren* 2. 100–125.
- Spencer, Andrew (1991): *Morphological theory. An introduction to word structure in generative grammar*. Oxford: Blackwell.
- Steyer, Kathrin. 2013. *Usuelle Wortverbindungen. Zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht*. Tübingen: Narr.
- Steyer, Kathrin & Katrin Hein. 2018. Satzwertige usuelle Wortverbindungen und gebrauchsbasierte Muster. In Stefan Engelberg, Henning Lobin, Kathrin Steyer & Sascha Wolfer: *Wortschätze. Dynamik, Muster, Komplexität*, 107-130. Berlin: de Gruyter.
- Stumpf, Sören. under review. Analogy, creativity, and Diachronic Construction Morphology: Tracing the emergence of a semi-schematic compound in German. In Carmen Mellado Blanco, Pedro Ivorra Ordines & Maricel Esteban Fonollosa (eds.), *Creativity in the Intersection between Phraseology and Construction Grammar*. Amsterdam & Philadelphia: Benjamins.
- Trips, Carola & Jaklin Kornfilt. (eds.) 2017. *Further investigations into the nature of phrasal compounding*. Berlin: Language Science Press.
- Wiese, Richard. 1996. Phrasal compounds and the theory of word syntax. *Linguistic Inquiry* 27 (1). 183–193.

Refining quantitative methods for Diachronic Construction Morphology: a study of three compound families in French

Jan Radimský
University of South Bohemia in České Budějovice
radimsky@ff.jcu.cz

Florence Villoing
Paris Nanterre University & MoDyCo, UMR 7023
villoing@parisnanterre.fr

1. Introduction

French nominal compounds consisting of two nouns (henceforth, “N+N compounds”) (see (1)) have been extensively investigated from a synchronic point of view, but much less attention has been paid to the diachronic perspective.

- (1) a. BOUCHERIE-CHARCUTERIE (lit. butcher shop.delicatessen) ‘butcher shop’
b. ENFANT-ROI (lit. child.king) ‘child as a king’
c. CONTRÔLEUR-QUALITÉ (lit. controller.quality) ‘quality controller’

They seem to represent a relatively recent innovation in Romance languages. According to Rainer (2021), the pattern does not display any continuity from Latin compounding and probably stems from a variety of different syntactic constructions whose number is extremely limited in French, at least until the end of the 19th century.

Recently, a large-scale study was conducted on the Attributive-Appositive Noun + Noun compounds (henceforth, “N+N ATAP compounds”, see § 2. for the classification) of Italian (Radimský 2024 ; Radimský & Micheli 2025). It investigated the origin and diachronic development of these compounds, with a particular focus on a detailed analysis of the history of three semi-schematic families: *N-modello*, *N-base*, and *N-limite* (see (2) for some examples). These three families have been chosen because of the crucial role they have played in shaping the broader N+N ATAP pattern in Italian.

- (2) a. *podere modello* (lit. farm.model) ‘model farm’, *lettore modello* (lit. reader.model) ‘model reader’
b. *prezzo base* (lit. price.base) ‘base price’, *campo base* (lit. camp.base) ‘base camp’
c. *caso limite* (lit. case.borderline) ‘borderline case’, *età limite* (lit. age.borderline) ‘threshold age’

Since two of the three compound families under investigation – namely *N-limite* and *N-modèle* – also developed in French, the present study replicates Radimský’s (2024) research on French data, pursuing several objectives. The principal aim is to further refine a methodology designed to trace the diachronic development and constructionalization of semi-schematic compound families by exploiting large-scale data from Google N-grams. In addition to the *N-limite* and *N-modèle* families, which have evolved in parallel in both Italian and French, we include a third French family, *N-surprise*, which exhibits a comparable overall type frequency but is unattested

in Italian. In a broader perspective, this research contributes to the question of how the Attributive-Appositive (ATAP) N+N compound pattern emerged and evolved in French. We proceed from the hypothesis that it remains an open question whether the ATAP pattern as such constitutes a productive higher-order construction in present-day French, or whether its apparent rise in type frequency is primarily driven by a limited number of lower-level, semi-schematic constructions (Bauer 2017:74; Rainer 2016:2714).

The analysis will be conducted within the theoretical frameworks of *Construction Morphology* (cf. § 2). It is based on an original methodology drawing on very large corpus data (§ 3.) and will be both quantitative (§ 4) and qualitative in nature (§ 5).¹

2. Noun+Noun compounds in French

The identification of sequences of two nouns as belonging to a morphological composition rather than a syntactic sequence has been the subject of much debate (see, for French, Corbin 1997:84; Fradin 2009: 192-206; Villoing 2012: 35-36; Van Goethem & Amiot 2019: § 3; Amiot 2020: 1920-1921, among others). The difficulty of defining this type of compounds in French partly explains why few studies have been devoted to it. However, adopting the theoretical perspective of *Construction Morphology* (Booij 2010, 2016) renders obsolete the problem of modular boundaries in grammar between lexicalized constructions of syntactic or morphological origin. Thus, we will accept as a binomial compound noun any form-meaning pair with a conventionalized denotating function, notably because these pairs are constructions in the sense of *Construction Grammar*.

Nominal compounds of two nouns in French share some key properties:

- (i) it is a productive pattern that forms complex nominal units ;
- (ii) it involves two bare common nouns without determiner ;
- (iii) there is an implicit relationship between the two nouns, without a preposition ;
- (iv) compounding is endocentric, and the compound is left-headed.

Despite these general properties, N+N compounds do not all fall under the same semantic type. As for the typology of compounds, we followed the proposal by Scalise and Bisetto (2009), adapted to Romance languages by Radimský (2015). In this typology, compounds of two nouns appear in three subtypes : Coordinative compounds / Attributive-Appositive compounds / Subordinative compounds. Our research will focus on ATAP compounds.

¹ This article builds on the study by Radimský & Villoing (2025) written in French, which is based on the same dataset. While Radimský & Villoing (2025) focus in detail on the qualitative analysis of individual compound families – an aspect of particular interest to a Francophone readership and only briefly outlined here in Section 5 – the present study offers a more in-depth discussion of the methodological issues related to the quantitative description of these families, presented in Sections 3 and 4. The two studies are thus complementary.

Table 1. Main subtypes of NN compounds in French

Coordinative compounds (COORD)	Attributive-Appositive compounds (ATAP)		Subordinative compounds (SUB)	
	Attributive	Appositive	Verbal-nexus	Ground
BOUCHERIE-CHARCUTERIE 'butcher shop .delicatessen'	ARTISTE-AMATEUR 'artist.amateur'	ENFANT-ROI 'child.king'	CONTRÔLEUR-QUALITÉ 'controller.quality'	APPAREIL-PHOTO 'camera.photo'
MONTRE-BRACELET 'watch.bracelet'	TRAIN-MINIATURE 'train.miniature'	TOMATE-CERISE 'tomato.cherry'	PRÉVISION-MÉTÉO 'forecast.weather'	ZONE-EURO 'zone.euro'

These N+N ATAP compounds have two interpretations:

- (i) The **Attributive** interpretation which is characterized by the fact that the second noun behaves like a modifier of the first, expressing a property or a quality of the head: ARTISTE AMATEUR is interpreted as an *artist* which is an *amateur*.
- (ii) The **Appositive** interpretation which is close to the attributive, except that the attributive relationship is metaphoric: ENFANT-ROI denotes a child (ENFANT) which is like a king (ROI).

The interpretation of these compounds is triggered by the modifier (i.e. the rightmost element, the second noun). They tend to form strong modifier-based families, which is why selected modifiers with highest type frequencies have sometimes also been analysed as nouns with adjectival value (Noailly 1990).

3. Data

3.1 Underlying corpus

Empirically, this study relies on large-scale diachronic data extracted from the Google Books corpus, specifically in the form of raw frequency lists published by Google as the third version of the French and Italian Google N-grams datasets.² The total size of the underlying Google Books corpus for the period 1800–2019 amounts to 318,631,007,871 tokens for French and 120,410,089,963 tokens for Italian. However, corpus size varies considerably over time. For French, this diachronic variation is visualized in Figure 1. The smallest corpus size is recorded for the year 1944, with 296,475,658 tokens, while the largest sample corresponds to the year 2016, with 4,594,932,651 tokens.

² <https://storage.googleapis.com/books/ngrams/books/datasetv3.html>

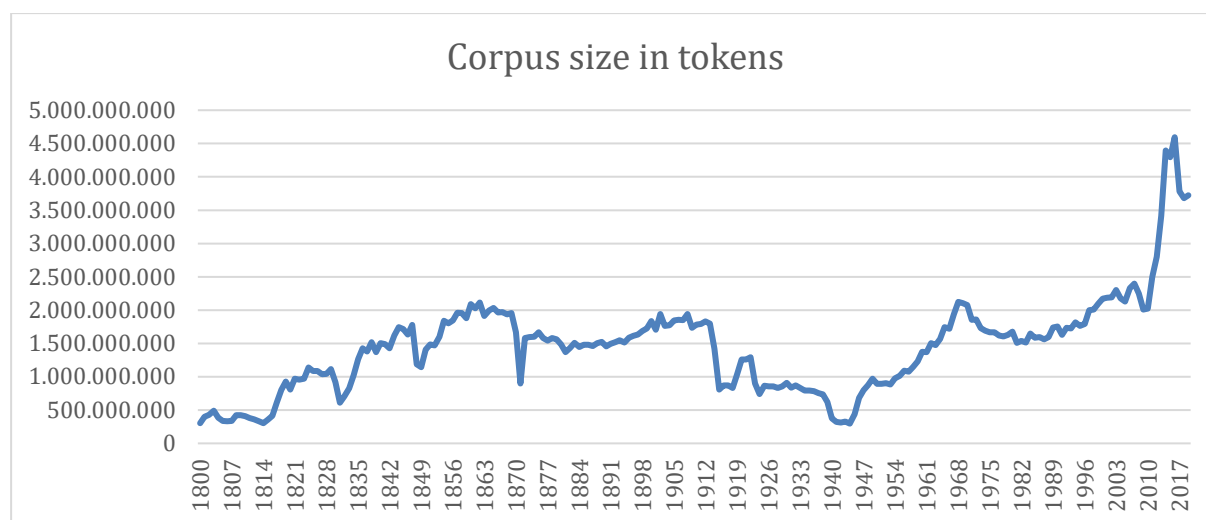


Figure 1. Size of the underlying Google Books corpus for the French Google N-grams frequency lists

These figures demonstrate that we are dealing with data whose magnitude and diachronic granularity are unparalleled in comparison with those typically employed in diachronic linguistic studies to date. A recurring challenge in previous research lies in the limited size of available corpora – often only a few million tokens per diachronic period – and the small number of time slices analyzed (frequently around five). As Hartmann (2018) has shown, this often necessitates the use of complex extrapolation techniques. As noted by Barðdal et al. (forthcoming), “[t]ypically, historical corpora offer a much greater abundance of texts for more recent periods, while the earlier periods are less documented, thus often yielding data scarcity. For instance, [...] in Frantext, the subcorpus for the 13th century is 74 times smaller than the one for the 20th century.”

In the dataset used here, diachronic granularity corresponds to the year of publication as recorded in Google Books, which yields 219 discrete time slices (i.e., individual years) for the period 1800–2019.³ In terms of corpus size per slice, the lowest-resourced year – 1944 – still comprises 296,475,658 tokens, a figure approximately three times larger than that of a so-called “third-generation” synchronic corpus such as the British National Corpus, which has served as a standard benchmark since the 1990s. In this light, while the overall size of the underlying corpus may seem unnecessarily excessive – indeed, it “represents many times the linguistic input that a single person would receive in their lifetime,” potentially diluting stylistically marked phenomena (Stefanowitsch 2020: 37) – the sizes of individual subcorpora, especially for the less represented periods, are sufficiently robust to obviate the need for extrapolation.⁴

Nevertheless, the corpus remains quantitatively imbalanced across time slices: the most extensive subcorpus (2016) is 15.5 times larger than the smallest one (1944). It is therefore essential to normalize type frequency values relative to corpus size, as will be discussed below.

This imbalance is not limited to volume but also extends to corpus composition, particularly in terms of genre representation – another longstanding challenge in diachronic research (Barðdal et al. forthcoming). In principle, this issue is intractable for historical corpora, given

³ Naturally, the assigned year corresponds to the publication date of the document rather than to the date of the text’s composition, which means that the re-edition of older sources may affect the accuracy of dating to some extent.

⁴ Stefanowitsch (2020: 38) estimates that a corpus generally needs to contain between one million and five hundred million tokens in order to allow for the study of any type of linguistic phenomenon.

that “the texts transmitted to the present represent a random subsample of the whole population, due to largely extralinguistic accidents. Thus, historical corpora can never even remotely capture the full variety of language” (Claridge 2009: 247). Genre balancing is problematic even in synchronic corpora. For instance, the latest conception of the Czech National Corpus (CNC) – one of the largest of its kind – does not define representativeness in terms of proportional genre distribution based on actual usage, but rather as the inclusion of maximal genre diversity, with relative weightings assigned arbitrarily (Cvrček, Čermáková & Křen 2016).

In this regard, the Google Books corpus offers a potentially more representative dataset than traditional historical corpora, due to its greater diversity of text types and genres. Despite its name, it is not restricted to books or literary texts; it includes various types of printed materials archived in libraries, including periodicals, and encompasses not only fiction but also administrative and technical language. This is of particular importance for the present study: Italian data show that N+N compounds began to emerge in the mid-19th century, primarily in the bureaucratic and economic language of the newly unified Kingdom of Italy. This stylistic anchoring explains their absence before the 1950s in diachronic corpora such as CODIT (Micheli 2022), which consist exclusively of literary texts (Radimský 2023).

In conclusion, the uneven distribution of genres across time periods does not necessarily constitute a methodological flaw. Over time, certain genres may naturally emerge, disappear, or change in frequency within the textual population. Thus, an unequal genre representation in large-scale datasets may not indicate sampling bias but may, in fact, reflect structural changes in the textual ecosystem over time.

3.2 Data extraction

The data used for the extraction of N+N compounds were drawn from the bigram and trigram datasets of the third version of the Google N-grams corpus, pre-processed in accordance with the procedure detailed in Radimský (2022). This step was necessary due to the frequent coexistence of two competing orthographic variants of such compounds: the hyphenated form and the spaced form (e.g., *mot-clé* vs. *mot clé*). In the raw data, the hyphenated variant appears as a trigram, while the spaced variant is listed as a bigram, given that the hyphen is treated as a standalone unigram. During pre-processing, these orthographic variants were consolidated into a unified bigram file to allow for consistent treatment. No lemmatization was applied, however, so as to enable future analyses to distinguish between different inflected forms. In technical terms, then, *mot-clé* and *mot clé* (= the singular form in two orthographic variants) on the one hand, and *mots-clés* and *mots clés* (= the plural form in two orthographic variants) on the other, are treated as two distinct compounds in the present analysis.

The pre-processed bigram file comprised a total of 29,697,536 candidate forms, from which actual N+N compounds had to be identified. This identification followed a multi-stage process. First, an automatic filtering was applied to retain only those bigrams whose constituents – word1 and word2 – could theoretically correspond to French nouns. This filtering excluded, in particular, cases where the second constituent (word2) was in fact a primarily adjectival form with only marginal nominal usage. To this end, the *Lexique 3.83* dictionary⁵ and part-of-speech frequency information extracted from the *Araneum Francogallicum Minus* corpus⁶ were employed to compile a list of nominal forms in French (a total of 48,231 forms), which served as the basis for automatic filtering of both constituents in the pre-processed bigrams. The filtering criteria for word2 were more stringent than for word1: adjectives with only marginal

⁵ Available at <http://www.lexique.org/> (cf. New et al. 2004).

⁶ See http://ucts.uniba.sk/aranea_about/index.html

nominal use (e.g., *cardiaque*), lexicalized adjectival nouns (e.g., *civil* in *un civil américain*, or *exclu* in *les exclus de la société*), as well as demonyms and toponyms (e.g., *espagnol*) were manually removed. This process reduced the candidate list to 3,324,646 bigram types (non-lemmatized).

Subsequently, the identification of genuine N+N compounds – representing only a small fraction of these candidates – was carried out manually. For the present analysis, we selected three N2-based families for detailed study: N- *limite*, N- *modèle*, and N- *surprise*, within which we identified 143, 144, and 145 compound types, respectively. To ensure maximal accuracy in data extraction, each compound type was manually verified by a native speaker against a sample of contextualized occurrences (tokens) in Google Books. Compounds that exhibited a high rate of false positives at the token level were excluded from the final dataset.

4. Diachronic profile of the families

4.1 Methodological preliminaries

The traditional methodological framework for analysing morphological productivity relies on Baayen's (2009) measures of Realized Productivity, Expanding Productivity, and Potential Productivity – two of which depend on hapax legomena. It is important to recall that these metrics were originally developed to infer productivity patterns even in relatively small or strictly synchronic corpora. As Berg (2020: 1118) observes, hapax-based measures offer only indirect approximations of productivity, and their interpretive value diminishes considerably when applied to very large corpora. For these reasons, our quantitative analyses will prioritize direct approaches wherever feasible. However, as Hüning (2019: 485) emphasizes, genuinely diachronic methods for tracing productivity change are still in their infancy and face substantial methodological challenges.

We assume that the diachronic development of constructions can be analysed directly by observing two complementary dimensions: (i) the gradual emergence of new types, observable through their first attestations (a direct approach advocated by Berg 2020), and (ii) changes in type frequency across successive time periods. Although these approaches constitute promising alternatives to hapax-based methods, their practical implementation is far from straightforward, particularly when working with large-scale datasets such as Google N-grams.

Given the nature and limitations of the available data, we therefore propose a set of quantitative procedures focusing on: (i) estimating the period during which the vast majority of types become attested for the first time (the “sample saturation period”, opposed to the “onset period”), (ii) examining diachronic changes in realized productivity, including the detection of potential breakpoints, and (iii) identifying salient leader words – i.e., particularly frequent items in specific periods that may have influenced the evolution of the respective constructional patterns. The former two methodological components will be introduced and applied in the following subsections, whereas the identification of leader words will be explored qualitatively in Section 5.

4.2 First attestations: Identifying the onset, expansion and saturation periods of the sample

In theory, it should be possible to directly retrieve the year of first attestation for each compound in our sample from the Google N-grams data. In practice, however, this is not feasible. Despite careful manual selection procedures, each compound still contains a proportion of false positives, and due to OCR-related errors, the likelihood of such false positives increases in earlier periods. Consequently, the directly observed year of first attestation cannot be considered fully reliable. Nevertheless, the diachronic evolution of absolute type frequency

allows us to approximate the initial emergence of each family and to identify key moments in their constructionalization. A suitable procedure is proposed below.

Let us first examine the raw type frequency for the three families – i.e., the absolute number of N+N compounds belonging to each family across time, as visualized in Figure 2.

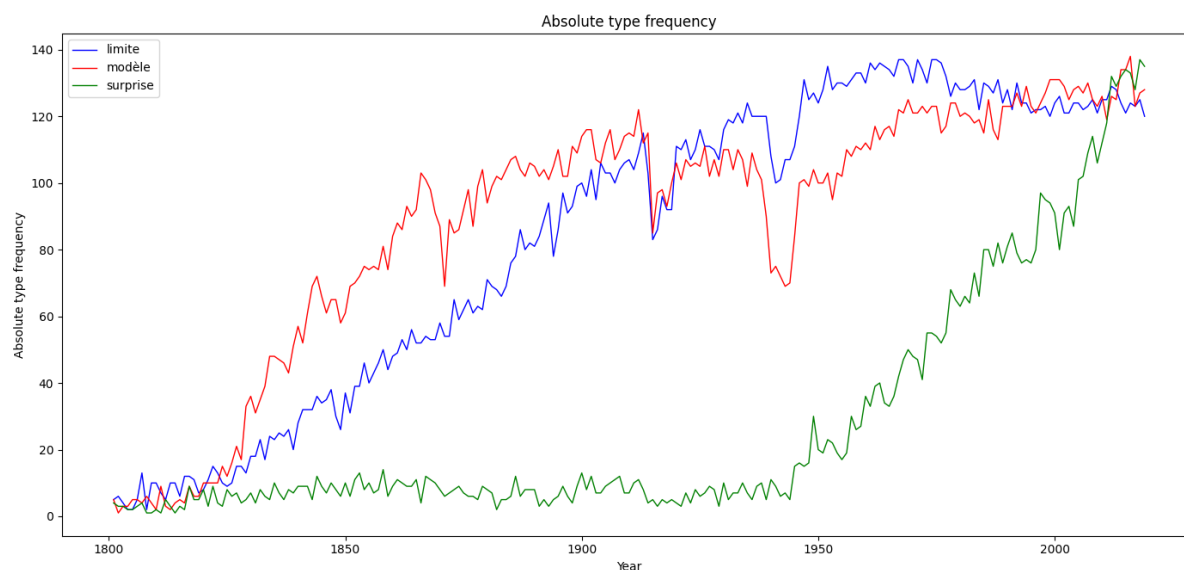


Figure 2. Absolute type frequency of the sample corresponding to the N-*limite*, N-*modèle*, and N-*surprise* families

The specific shape of these curves is not of primary relevance, since the underlying corpus size varies considerably across years (see Figure 1). Moreover, with Google N-grams, it is not possible to control for corpus size per time slice – as recommended by Berg (2020) – because the full underlying text data are not publicly accessible. The presence of false positives also implies that the real number of types is slightly lower than indicated in Figure 2. Nonetheless, the data visualized in Figure 2 offers a reasonably reliable indication of the period in which new compounds within each family first emerged. This emergence period – corresponding to the constructionalization of the respective families – can be divided into three distinct diachronic stages: (i) the *onset period*, (ii) the *expansion period*, and (iii) the *saturation period*.

Let us begin by defining and determining the sample saturation period. To put it simply, sample saturation is reached when nearly all compounds of a given family have already been attested for the first time, so that no new items (types) appear after this date. Given the methodological limitations outlined above, and as a precaution, we define and operationalize the beginning of the saturation period as the time span in which at least 95% of the NNs belonging to each family have appeared in the data in a more systematic manner – that is, over at least three consecutive years. If we take the first year of such a triennium as the milestone, saturation is reached in 1957 for N-*limite*, and in 2014 for both N-*modèle* and N-*surprise*. In other words, after these time points, the respective families ceased to produce new compounds.

The onset period can be viewed as the exact opposite of the saturation period, i.e. as the time span in which the attested compounds are so few that they are not likely to be formed according to the respective family schema which, at this time, does not exist yet. To operationalize the end of the onset period, two factors are to be considered: (i) the minimum number of types that can make up a higher-level construction is to be determined rather in absolute than in relative

figures – the lowest possible figure being 2 or 3 compounds⁷; and (ii) in older time periods the probability of false positives increases, so by precaution the required minimum type frequency should be higher. Let us therefore define the end of the onset period as the first year of the triennium in which at least 10 different compounds are attested in three consecutive years. According to this criterion, the onset period ends in 1816 for *N-limite*, in 1824 for *N-modèle*, and only in 1945 for *N-surprise*.

As a result, the expansion period covers the time span between the end of the onset period and the beginning of the saturation period, corresponding to the time in which the newly established construction (family) produces new compounds. The three periods for the respective families are summarized in Table 1.

Table 2. Onset, expansion and saturation periods for the *N- limite*, *N- modèle*, and *N-surprise* families

	Onset	Expansion	Saturation
N-limite	< 1816	1816-1957	> 1957
N-modèle	< 1824	1824-2014	> 2014
N-surprise	< 1945	1945-2014	> 2014

The longest expansion period is attested for the *N-modèle* family (1824–2014). Conversely, the expansion period of *N-limite* (1816–1957) and that of *N-surprise* (1945–2014) hardly overlap, which suggests that these constructions developed diachronically in two different historical periods.

4.3 Realized productivity

The data presented in Figure 2 reflect the number of N+N compounds for each of the studied families, which broadly corresponds to the notion of *rentabilité* ‘profitability’ as defined by Corbin (1987: 177), or to Baayen’s (2009) concept of *realized productivity*. However, before interpreting the trajectories of these values, it is necessary to normalize them relative to the size of the underlying corpus for each year.

A straightforward normalization approach would involve dividing the absolute type frequency (V) by the number of tokens in the corresponding subcorpus (N). However, as illustrated by Baayen’s (2009: 903) *Vocabulary Growth Curve*, the number of new types (V) does not increase proportionally with corpus size (N); rather, the rate of type emergence decreases as corpus size expands. This poses a particular challenge for large-scale corpora such as the one used in the present study. While the largest diachronic subcorpus (for the year 2016) in the French Google Books dataset is 15.5 times larger than the smallest one (1944), the actual difference in the number of new types is expected to be considerably smaller. Consequently, it is more appropriate to normalize type frequency relative to the number of unique word forms within each subcorpus.

For this purpose, we used the number of alphabetic unigram types (i.e., unigrams consisting exclusively of alphabetic characters) per diachronic slice as an estimate of the number of unique word forms. The range between the minimum (925,378 unique forms in 1814) and the maximum (2,849,533 in 1970) corresponds to a factor of approximately three (more precisely, 3.08). A diachronic overview of the number of unique word forms is presented in Figure 3.

OBJ

⁷ See Jackendoff & Audring (2020: 224-225).

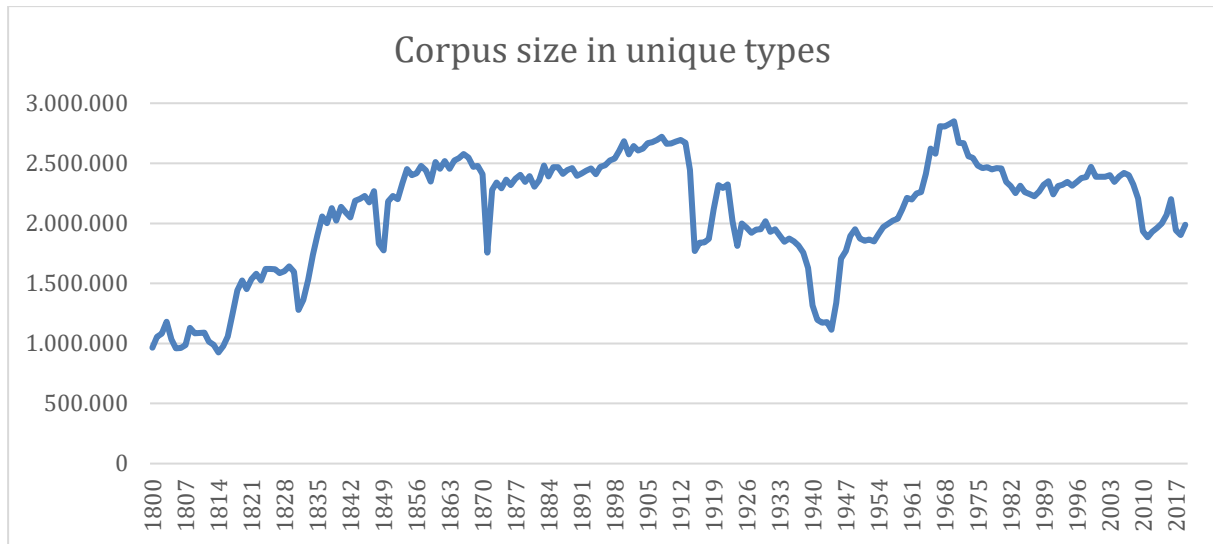


Figure 3. Number of unique word forms in the alphabetic unigrams of the French Google N-grams dataset

It is worth noting that the overall shape of the curve in Figure 3 closely mirrors that of Figure 1, though with considerably less pronounced fluctuations. A particularly interesting divergence appears in the most recent time periods, from 2010 onward: while the token-based curve (Figure 1) shows a steep upward trend reflecting a dramatic growth in corpus size, the curve for unique word forms (Figure 3) remains within the bounds of earlier trends. This suggests that the corpus may have reached a threshold at which the addition of large volumes of texts from similar registers and domains no longer significantly increases lexical diversity.

The normalized type frequencies, calculated relative to the number of unique word forms, are shown in Figure 4.

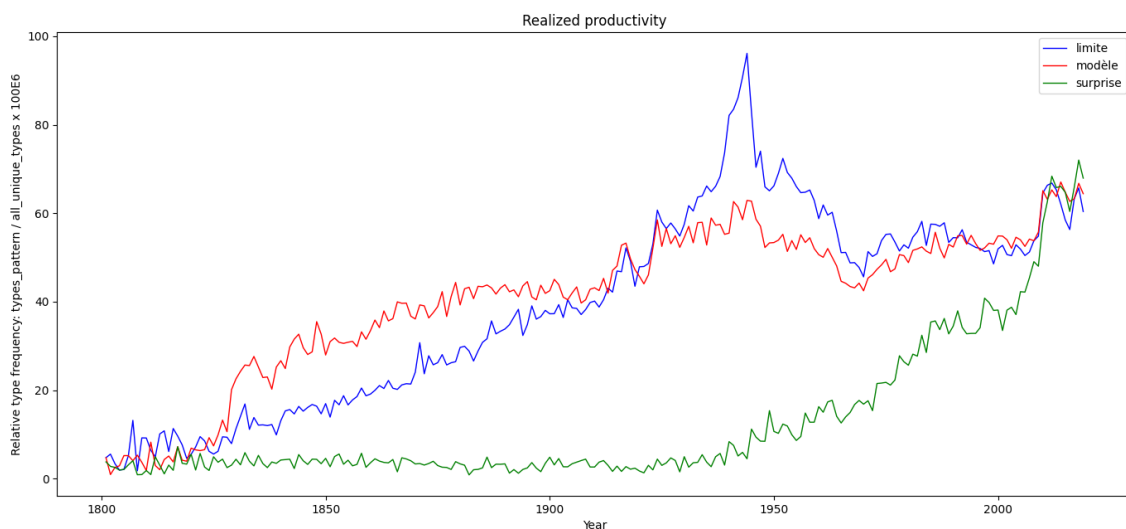


Figure 4. Relative type frequency (realized productivity), normalized by the number of unique word forms

According to Figure 4, the *N-limite* and *N-modèle* families exhibit a continuous increase in type frequency from the early 19th century until 1944, at which point a discontinuity occurs, followed by a decline and then a phase of stabilization or mild increase. After normalization, the apparent dips observed in the absolute values (Figure 2) during the 1920s and 1940s no longer appear, indicating that both families maintained steady growth in realized productivity until the mid-20th century. By contrast, the *N-surprise* family shows a similar trajectory in both

the absolute (Figure 2) and normalized (Figure 4) data, indicating a sustained influx of new forms and a continued growth in productivity since the 1940s.

In the next section, we will show how the shape of these curves can be divided into distinct phases, and how these phases relate to the onset, expansion, and saturation periods identified above.

4.4 Phases in the realized productivity curve

The diachronic trends in constructional patterns, as visualized through realized productivity in Section 4.3, can be further analyzed to identify potential change points marking shifts in their evolution. An initial attempt to apply the variability-based neighbour clustering method proposed by Gries and Hilpert (2008) did not yield satisfactory or intuitively interpretable results for the trajectories observed in our dataset. We therefore adopted an alternative method, tested in previous studies (Radimský 2024, 2025; Radimský & Micheli 2025), to analyze the *N-limite* and *N-modèle* families, which display a pronounced degree of variability.

To visualize diachronic trends via regression lines, we employed the Theil-Sen estimator (Sen 1968), supplemented by the Mann-Kendall trend test (Kendall 1975) to assess the statistical significance of the identified trends (Python implementation based on Hussain & Mahmud 2019). These non-parametric, rank-based methods are suitable for detecting various types of dependencies (not limited to linear relationships), do not require normally distributed residuals, and are robust to outliers. They are therefore well-suited to trend analysis in lexical usage based on large-scale diachronic corpora (see Kovář & Herman 2013). Key change points can be estimated – within a reasonable margin of approximation – through visual inspection of the realized productivity curve. Their validity is then confirmed (or rejected) if the Mann-Kendall test detects a statistically significant change in trend between the segments preceding and following the proposed point.

The realized productivity curve for the *N-limite* family can thus be divided into four phases, delineated by three change points, as shown in Figure 5.

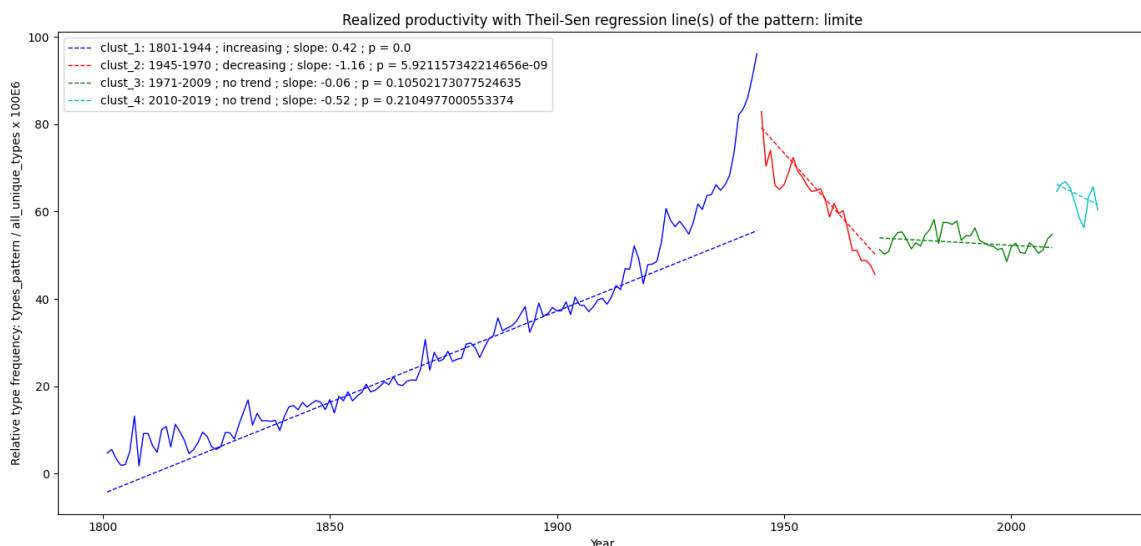


Figure 5. Relative type frequency (realized productivity) of the *N-limite* family, with diachronic segmentation

The first change point, in 1944, corresponds to the curve's absolute maximum, while the second, in 1970, marks its lowest value since the mid-20th century. In other words, after an extended initial period of growth (1800-1944), the curve undergoes a sharp trend reversal. The

second phase, which begins only 13 years before the start of the saturation period identified in the previous section (1957), is characterized by a declining trajectory.

From 1971 onward, two interpretive possibilities emerge. The first is to treat the entire 1971–2019 period as a single phase; in this case, the Mann–Kendall test reveals a mildly increasing trend, although the relatively high p -value ($p = 0.048$) suggests that this trend is only marginally significant. The second possibility, illustrated in Figure 5, isolates a third phase between 1971 and 2009, representing nearly four decades of stability, during which the slope of the regression line is close to zero. The final portion of the curve – post-2009 – is difficult to interpret. This segment is relatively short and marked by considerable variability, such that small adjustments to the boundaries of the cluster yield different trend outcomes in the Mann–Kendall test. A longer diachronic perspective will therefore be required before this final phase can be meaningfully analysed.

The diachronic segmentation visualized in Figure 5 aligns quite closely with the key phases identified in Section 4.2. Although no significant trend change appears at the boundary between the onset and expansion phases (which is not to be expected), the trend reversal in 1944 lies close to the end of the expansion phase (1957), suggesting that saturation effectively takes place during the decade 1945–1955. After this point, no significant increase in type frequency is detected.

The realized productivity curve for the *N-modèle* family can be divided into five phases, delineated by four change points, as shown in Figure 6.

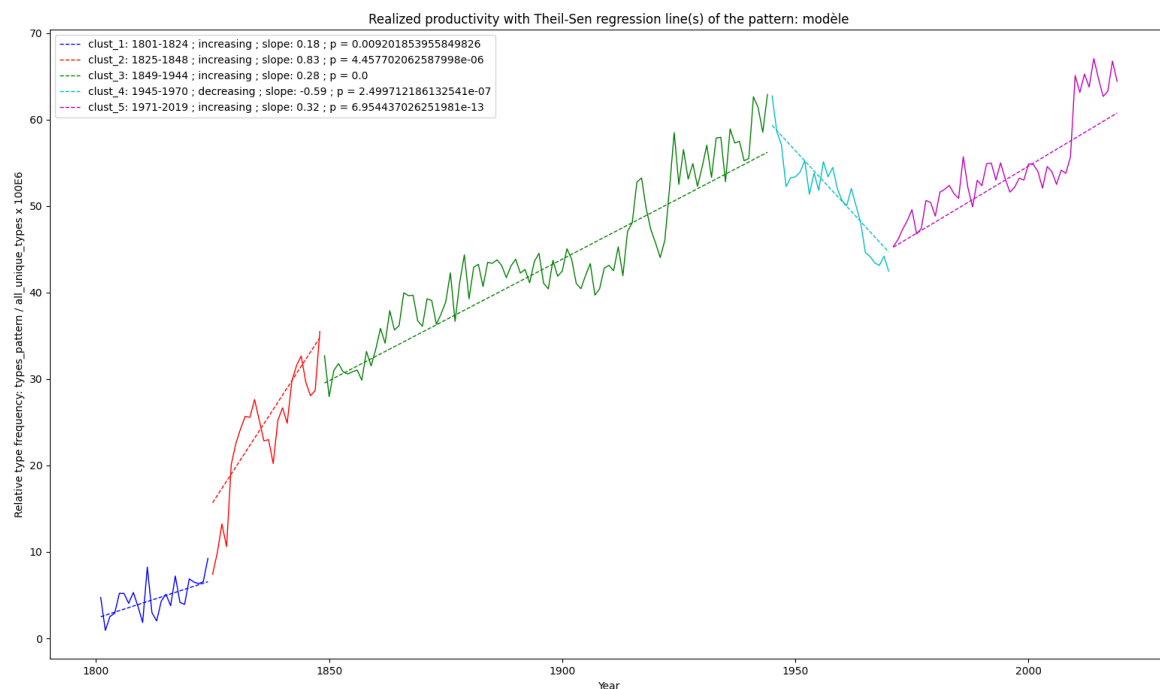


Figure 6. Relative type frequency (realized productivity) of the *N-modèle* family, with diachronic segmentation

Similarly to the *N-limite* family, the *N-modèle* family displays a mild increase during the onset period (cluster 1, from 1800 to 1824) and a decreasing tendency between 1945 and 1970 (cluster 4). However, the expansion period of *N-modèle* is longer and more articulated than in the case of *N-limite*. While the latter exhibits a continuous increase throughout its entire expansion period until the 1940s–1950s, the development of the former is structured into a sharp increase (cluster 2), a slower increase (cluster 3), a decline (cluster 4), and a renewed increase (cluster 5). This confirms the observation made in Section 4.2 that the expansion period of the *N-modèle* family extends into the 2010s.

5. Qualitative analysis: example of the emergence of a productive family, the N-MODÈLE family

The quantitative analyses presented above can be complemented by a detailed qualitative investigation into the history of each compound family. In this case, our methodology involves both an analysis of available dictionaries and an examination of so-called *leader words* (see Burdy 2019), that is, compounds with the highest relative token frequency in different diachronic stages, which most likely played a crucial role in the emergence and development of the respective families. As an illustration, we will briefly discuss the leader words of the N-*modèle* family as identified in the Google Books corpus (for a thorough analysis of the three families under investigation, see Radimský & Villoing 2025).

The first leader word before the 1930s and up until the 2000s was *ferme-modèle* (lit. farm.model) ‘model farm’, conveying the meaning of the family “organization to imitate” (as in Italian). This compound emerged at the time of the development of modern agriculture in France and Europe (1930-45), then in the colonies (Africa, Asia), and the Near and Middle East. It is found in literary works and in history, geography, and economics.

The second leader word in this family is *école-modèle* (lit. school.model) ‘model school’. It emerged at the time of the rise of public and secular education, to designate schools training teachers; then the term spread to describe other types of schools (agriculture, commerce, architecture, military) during the years 1930-45-60, and in other countries (China, the USSR, Israel; African and Asian countries). It remains in literature.

The third leader word is *lecteur-modèle* (lit. reader.model) ‘model reader’, which emerged in the 1980s. It belongs to the entire semantic subfamily of compounds meaning ‘perfect in their kind’ that have been present since before the 1930s, with *époux-modèle* (lit. husband.model) ‘ideal husband’, *femme-modèle* (lit. woman.model) ‘model woman’, *fille-modèle* (lit. girl.model) ‘model girl’ and others. But it is the compound *lecteur-modèle* that marks the turning point for the entire semantic subfamily, rising to the status of leader word from 2000. It immediately leads to about ten compounds of the same type. It was completely absent from the top twenty leading words during the four previous periods, but it emerged from the publication of Umberto Eco’s novel in which he proposes the notion that becomes extremely popular in the intellectual world. The notion is taken up in works of literary, linguistic, cinematographic, psychoanalytic, and sociological studies. It therefore seems that this compound has caused a real revival of the family of N-*modèle* compounds, in particular those whose N1 refers to a person.

6. Conclusion and future work

This study has explored the diachronic evolution of three semi-schematic compound families in French (N-*modèle*, N-*limite*, and N-*surprise*) using large-scale data extracted from Google N-grams. It complements previous research on Italian (Radimský 2024), where parallel families (N-*modello*, N-*limite*, and N-*base*) developed during the same historical period (1800-2000).

From a coarse-grained perspective, the diachronic trajectories of these three French and Italian families closely mirror each other, both in terms of timeline and domain-specific diffusion. Upon closer inspection, however, important divergences arise, confirming Franz Rainer’s intuition that each family has probably followed its own specific historical trajectory. First, not all three families are parallel to the same extent: while Italian exhibits a robust development of the N-*base* family, its French equivalent failed to achieve comparable productivity; conversely, the N-*surprise* family, absent in Italian, demonstrates a sustained

trajectory in French, with increasing productivity well into the contemporary period. Second, the data reveal important differences in diachronic trajectories within each language. The expansion of the French families *N-limite* and *N-surprise* was continuous, but took place in two distinct, non-overlapping periods. By contrast, the expansion of the *N-modèle* family extended over a longer time span and unfolded in two phases, separated by a period of decline between the mid-1940s and the mid-1970s. Google N-gram data also make it possible to identify and qualitatively analyse the different *leader words* that may have played an important role either in the initial formation of the respective families or in their subsequent development. This was only briefly illustrated here by a few examples from the *N-modèle* family.

Beyond its empirical contributions, this study also proposes several methodological innovations in the quantitative analysis of diachronic constructional change. It introduces a set of procedures to identify onset, expansion, and saturation periods for compound families, based on first attestations and type-frequency trajectories derived from Google N-gram data. Furthermore, it normalizes realized productivity using the number of unique word forms – rather than token counts – thereby addressing the limitations of applying traditional frequency metrics to large, genre-heterogeneous corpora such as Google N-grams. The use of Theil-Sen regression and the Mann-Kendall trend test provides a robust statistical framework for detecting change points in constructional development, complementing and refining existing models of productivity based on hapax legomena or vocabulary growth curves.

Future research should extend this approach to a broader set of N+N ATAP families in French and in other Romance languages, as such a systematic cross-linguistic comparison will be essential to fully reconstruct the mechanisms of interlingual influence. Another important issue left for future research, touched upon in Radimský & Micheli (2025), concerns a better understanding of how the expansion of individual families contributes to the emergence of more abstract higher-order constructions, such as the construction corresponding to N+N ATAP compounds.

Acknowledgements

This research was supported by the research project No. 25-15350S, entitled *How do word-formation patterns emerge? An empirical diachronic analysis of Italian N+N compounds*, funded by the Grant Agency of the Czech Republic (GAČR) and conducted at the Faculty of Arts, University of South Bohemia in České Budějovice.

References

- Amiot, Dany. 2020. Procédés morphologiques de création lexicale. *Grande Grammaire Historique du Français (GGHF)*. 1894-1927.
- Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Handbooks of Linguistics and Communication Science*, 899–919. Berlin, New York: Mouton de Gruyter. <https://doi.org/10.1515/9783110213881.2.899>.
- Barðdal, Jóhanna, Renata Enghels, Quentin Feltgen, Sven Van Hulle & Peter Lauwers. Forthcoming. Productivity in Diachrony. In *Wiley Blackwell Companion to Diachronic and Historical Linguistics*. Hoboken: Wiley-Blackwell.
- Bauer, Laurie. 2017. *Compounds and Compounding*. Cambridge University Press.
- Berg, Kristian. 2020. Changes in the productivity of word-formation patterns: Some methodological remarks. *Linguistics* 58(4). 1117–1150. <https://doi.org/10.1515/ling-2020-0148>.
- Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.
- Booij, Geert. 2016. Construction Morphology. In Andrew Hippisley & Gregory Stump (Eds.), *The Cambridge Handbook of Morphology*, 424–448. Cambridge: Cambridge University Press.

- <https://doi.org/10.1017/9781139814720.016>.
- Burdy, Philipp. 2019. On the importance of leader words in word formation: The popular transmission of the Latin abstract-forming suffix *-io* in French. *Word Structure*. 12(1). 42–59. <https://doi.org/10.3366/word.2019.0138>.
- Claridge, Claudia. 2009. Historical corpora. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: an international handbook* (Handbücher Zur Sprach- und Kommunikationswissenschaft = Handbooks of Linguistics and Communication Science Bd. 29.2), 242–259. Berlin, New York: Walter de Gruyter.
- Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique* (Linguistische Arbeiten 193). Tübingen [Lille]: M. Niemeyer [diff. Presses universitaires de Lille].
- Corbin, Danielle. 1997. Locutions, composés, unités polylexématiques: lexicalisation et mode de construction. In Martins-Balter (éd.), *La locution entre langue et usage*, 53–101. Lyon: ENS Éditions. <https://doi.org/10.4000/books.enseditions.18713>.
- Cvrček, Václav, Anna Čermáková & Michal Křen. 2016. Nová koncepce synchronních korpusů psané češtiny. *Slovo a slovesnost* 77(2). 83–101.
- Fradin, Bernard. 2009. Compounding in French. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford Handbook of Compounding*, 417–435. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199695720.013.0022>.
- Gries, Stefan & Martin Hilpert. 2008. The identification of stages in diachronic data: variability-based neighbour clustering. *Corpora* 3(1). 59–81. <https://doi.org/10.3366/E1749503208000075>.
- Hartmann, Stefan. 2018. Derivational morphology in flux: a case study of word-formation change in German. *Cognitive Linguistics* 29(1). 77–119. <https://doi.org/10.1515/cog-2016-0146>.
- Hüning, Matthias. 2019. Morphological Theory and Diachronic Change. In Jenny Audring & Francesca Masini (eds.), *The Oxford Handbook of Morphological Theory*, 475–492. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199668984.013.28>.
- Hussain, Md. & Ishtiaq Mahmud. 2019. pyMannKendall: a python package for non parametric Mann Kendall family of trend tests. *Journal of Open Source Software* 4(39). 1556. <https://doi.org/10.21105/joss.01556>.
- Jackendoff, Ray & Jenny Audring. 2020. *The texture of the lexicon: Relational Morphology and the Parallel Architecture*. Oxford, New York: Oxford University Press.
- Kendall, M. G. 1975. *Rank correlation methods* (4th ed.). London: Charles Griffin.
- Kovář, Vojtěch & Ondřej Herman. 2013. Methods for Detection of Word Usage over Time. In Aleš Horák & Pavel Rychlý (eds.), *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2013*, 79–85. Tribun EU. <https://www.muni.cz/en/research/publications/1131909>. (4 March, 2022).
- Micheli, M. Silvia. 2022. CODIT. A new resource for the study of Italian from a diachronic perspective: Design and applications in the morphological field. *Corpus* (23). <https://doi.org/10.4000/corpus.7306>.
- New, Boris, Christophe Pallier, Marc Brysbaert & Ludovic Ferrand. 2004. Lexique 2 : A new French lexical database. *Behavior Research Methods, Instruments, & Computers* 36(3). 516–524. <https://doi.org/10.3758/BF03195598>.
- Noailly, Michèle. 1990. *Le substantif épithète*. Paris: PUF.
- Radimský, Jan. 2015. *Noun+Noun compounds in Italian: a corpus-based study* (Episteme. Theoria). České Budějovice: Jihočeská univerzita v Českých Budějovicích.
- Radimský, Jan. 2022. Towards a diachronic analysis of Romance morphology through Google n-grams. *Corpus* (23). <https://doi.org/10.4000/corpus.7205>.
- Radimský, Jan. 2023. Where did the Italian Verbal-Nexus N+N compounds come from? *Proceedings of the Mediterranean Morphology Meetings* 13. 71–82. <https://doi.org/10.26220/mmm.4407>.
- Radimský, Jan. 2024. I composti N+N attributivi in diacronia: le famiglie N-modello, N-base e N-limite. *Études romanes de Brno* (3). 10–34. <https://doi.org/10.5817/ERB2024-3-2>.
- Radimský, Jan. 2025. I composti NN romanzi: una storia di famiglie. Il caso di N-chiave e N-fantasma. In Sara Matrisciano, Johannes Schnitzer, Elisabeth Peters & Franz Rainer (eds.), *Patterns, variants and change: through the prism of morphology: studies in honor of Franz Rainer* (Travaux de Linguistique Romane Morphologie, Syntaxe, Grammaticographie 5), 351–370. Strasbourg: ELiPHi - Éditions de Linguistique et de Philologie.
- Radimský, Jan. & M. Sylvia Michelli. 2025. Expanding Horizons: A comprehensive insight into the evolution of Italian Attributive-Appositive Noun+Noun Compounds using Google Books Data. *Lexique, Numéro spécial*. <https://doi.org/10.54563/lexique.1948>.
- Radimský, Jan. & Florence Villoing. 2025. Le rôle de trois familles semi-spécifiées dans le développement du modèle de composition N-N de type Attributif. *Langue française*, 228. 9–21.
- Rainer, Franz. 2016. Italian. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *An International Handbook of the Languages of Europe*, Volume 4 Word-Formation, 2712–2731. Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110379082-017>.
- Rainer, Franz. 2021. Compounding: From Latin to Romance. In Aronoff (Ed.), *Oxford Research Encyclopedia of Linguistics*, 1–31. Oxford: Oxford University Press. <https://doi.org/10.1093/acrefore/9780199384655.013.691>

- Scalise, Sergio & Antonietta Bisetto. 2009. The Classification of Compounds. In Rochelle Lieber, & Pavol Štekauer (eds.), *The Oxford Handbook of Compounding*, 34–53. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199695720.013.0003>
- Sen, P. K. 1968. Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63(324), 1379–1389.
- Stefanowitsch, Anatol. 2020. *Corpus linguistics: A guide to the methodology*. Zenodo. <https://doi.org/10.5281/ZENODO.3735822>.
- Trésor de la langue Française Informatisé*. <http://www.atilf.fr/tlfi>
- Van Goethem, Kristel & Dany Amiot. 2019. Compounds and multi-word expressions in French. In Barbara Schlücker (ed.), *Complex lexical units*, 127–152. Berlin: De Gruyter. <https://doi.org/10.1515/9783110632446-005>.
- Villoing, Florence. 2012. French compounds. *Probus*, 24(1). 29-60.

Superlative events. Intensifying verbs through the elative suffix *-issimo*

Luisa Troncone
University of Salerno &
University of Lille (STL UMR 8163)
ltroncone@unisa.it

Flavio Pisciotta
University of Salerno
fpisciotta@unisa.it

1. Introduction

In Italian, the most prototypical affix for the expression of intensification is the elative suffix *-issimo* (Cacchiani 2011; Napoli & Ravetto 2017; for the widespread perception of *-issimo* as the prototypical Italian intensification suffix, see also Giacinti 2024; Zilg 2011; Montermini 2024). Such strategy is most typically used with adjectives (1a) (Merlini Barbaresi 2004: 448-449) but is also exploited with nouns (1b) (Grandi 2002: 177), where it conveys a positive or prototypical evaluation of the referent, and to adverbs (1c) (Gaeta 2011), where it expresses the maximal degree of a property.

- (1) a. *bell-issim-o* beautiful-ELAT-M.SG ‘very beautiful’
b. *partit-issim-a* match.ELAT-F.SG ‘big match’
c. *subit-issimo* immediately-ELAT ‘literally immediately’

Regarding verbs, research concentrated on past participles in adjectival function (Merlini Barbaresi 2004: 449; Napoli 2013). However, usage data show that the suffix also occurs in other verbal forms (such as the present progressive, infinitive, and present simple; see 2a-c), especially in playful and/or informal contexts such as social media posts.¹ Notably, *-issimo* had already been used as a testing ground for word-formation rules on the Internet (Rainer 2003; Berezowska 2017). In fact, the Internet constitutes an ideal environment for observing creative and non-standard morphological uses. It has been observed that such creativity does not primarily aim at expanding morphological paradigms, but rather at increasing the expressive force (Santiloni 2026: 114).

- (2)
- | | | | | |
|----|---------------|-----------------|---------------------|--------------|
| a. | <i>Non</i> | <i>ho</i> | <i>visto</i> | <i>mezzo</i> |
| | NEG | AUX.PRS.IND.1SG | watch.PTCP.PST.M.SG | half.M.SG |
| | <i>minuto</i> | <i>dell’</i> | <i>Atalanta</i> | <i>ma</i> |
| | minute.SG | of.DET.M.SG | Atalanta | but |

¹ All the examples shown throughout the paper are social media posts, collected from X (formerly known as Twitter).

sto ***godend-issimo***
 AUX.PRS.IND.1SG **enjoy.PROG-ELAT**

‘I didn’t even watch half a minute of Atalanta BC’s match but **I’m enjoying [the result] a lot**’

b.	<i>Non</i>	<i>saprei</i>	<i>dire</i>	<i>con</i>
	NEG	can.PRS.COND.1SG	say.INF	with
	<i>certezza,</i>	<i>ma</i>	<i>mi</i>	<i>ha</i>
	certainty.SG	but	1SG.DAT	AUX.PRS.IND.3SG
	<i>fatto</i>	<i>comunque</i>	<i>rider-issimo</i>	
	make.PTCP	anyway	laugh.INF-ELAT	

‘I couldn’t say for sure, but it made me **laugh a lot** anyway’

c.	<i>ma</i>	<i>io</i>	<i>non</i>	<i>ho</i>
	but	1SG.NOM	NEG	AUX.PRS.IND.1SG
	<i>capito</i>	<i>la</i>	<i>differenza</i>	<i>ve</i>
	understand.PTCP	DET.F.SG	difference.SG	2SG.DAT
	<i>lo</i>	<i>giur-issimo</i>		
	3SG.M.ACC	swear.PRS.IND.1SG-ELAT		

‘but I didn’t understand the difference **I sincerely swear** to you’

These uses of *-issimo* are particularly interesting since evaluative suffixation in the verbal domain in Italian is rather unusual, and other means, e.g., prefixation or analytic strategies (3), are much more frequent (Napoli & Ravetto 2017: 339).

- (3) *Non **ultradrammatizziamo** queste circostanze*
 ‘Let’s not over-dramatise these circumstances’ (Napoli & Ravetto 2017: 340)

Moreover, evaluative suffixation in verbs has been found to exclusively express meanings related to reduction (4a), attenuation (4b), pejoration (4c), speed (4d), and iteration (4e) (Grandi 2008: 28; Grandi & Iacobini 2008: 486).

- (4) a. *canti-cchiare*
 sing-DIM
 ‘to sing softly’
 b. *mangiu-cchiare*
 eat-DIM
 ‘to eat a little/to snack’
 c. *studia-cchiare*
 study-DIM
 ‘to half-study’
 d. *becche-ttare*
 peak-DIM
 ‘to peck repeatedly’

- e. *salte-llare*
jump-DIM
'to hop'

Occurrences like (2a-c), instead, suggest that suffixation is indeed employed with verbs, at least in some communicative contexts, as a means to convey a proper “intensifying” semantics.

However, unlike its degree-modifying use with adjectival bases, the scope of the intensification in verbs is less clear: while in (2a) it possibly marks a stronger intensity, in (2b) the higher degree could also entail a higher duration; finally, in (2c) what is modulated is the illocutionary force of the utterance.

The goal of our study is thus twofold: on the one hand, we aim to offer a description of the use and distribution of *-issimo* with verbal bases in Italian, employing occurrences found on X posts; on the other hand, we attempt to understand *what* is intensified in such verbs, and if the intensifying sub-meaning depends on (or is influenced by) some semantic features proper of the verbal bases.

2. Data and Methods

Our study involved three separate steps: verb choice, verb annotation, and annotation of the type of intensification.

Regarding the selection of the verbs, to create the dataset, we extracted all the verbs from the fundamental vocabulary of Italian (De Mauro 2003) and ranked them by their frequency in the VoLIP corpus (Voghera *et al.* 2014), a corpus of spoken Italian. Considering the first 300 verbs of the so-obtained frequency ranking, we checked for the use of *-issimo* with each of these verbs, searching the Italian section of X for specific verbal forms, following Grandi's (2008) approach. The forms searched included the first- and third-person singular of the present simple (5a), the first- and third-person singular of the imperfective (5b), the first- and third-person singular of the present perfect (5c), as well as the infinitive (5d) and the gerund (5e). The past participle form was intentionally excluded, as its frequent adjectival uses could have obscured the distinction between verbal and adjectival intensification, thereby introducing noise into the analysis. It should be noted that the *-issimo* forms were constructed by the authors on the basis of their native-speakers' intuition and subsequently used as search strings for data retrieval.

(5)

- a. *CIAO MARIA IO ME NE VAD-ISSIMO*
bye MARIA 1SG SI from.here **leave.1SG.PRS.IND-ELAT**
'Bye Mary I'm immediately leaving'
- b. *Verso sera, ieri, l'affluenza mi aveva fatto ben sperare, per un momento. Ci vedevo la possibilità che indicasse una reazione al timore per un esito dato per scontato.*
Mi sbagliav-issimo e per
SI be.wrong.IPFV.IND.1SG-ELAT and for
questo, ho ancora gli
this.M.SG AUX.PRS.IND.1SG still DET.M.PL
occhi sbarrati
eyes.PL wide.open.PL

'Towards evening yesterday, the turnout had given me some hope, for a moment. I thought I could see the possibility that it reflected a reaction to the fear of a taken-

for-granted outcome. **I was so very wrong**, and partly for that reason I'm still in shock.'

c.	<i>io</i>	<i>non</i>	<i>ho</i>	<i>capit-issimo</i>
	1SG	NEG	AUX.PRS.IND.1SG	understand.PTCP-ELAT
	<i>il</i>	<i>monologo</i>	<i>della</i>	<i>Leotta,</i>
	DET.M.SG	monologue	of.F.SG	Leotta
	<i>proprio</i>	<i>come</i>	<i>logica</i>	
	quite	as	logic	

'I **didn't completely understand** the logic behind Leotta's monologue'

d.	<i>Perché!?</i>	<i>Io</i>	<i>volevo</i>	<i>dormir-issimo.</i>
	why	1SG	want.1SG.IPFV.IND	sleep.INF.ELAT

'Why!? I wanted to **sleep in**'

e.	[...]	<i>ti</i>	<i>sto</i>	<i>legend-issimo</i>
	[...]	2SG.ACC	AUX.1SG.PRS.IND	read.GER-ELAT

'[...] I'm **reading** your work **so often**'

We expect *-issimo* to occur with non-finite forms (past participles, gerunds, and infinitives), due to the marked and creative nature of such suffixation, also since these forms can display adjectival properties and thus provide a natural locus for the extension of an adjectival superlative suffix.

2.1 Annotation workflow

Given the technical limitations of X, which prevent systematic scraping and frequency extraction, our workflow involved the manual check of the availability of *-issimo* forms across different verb forms. Each form was treated as a distinct type, and only the existence (as opposed to the frequency) of each form was recorded. The resulting counts, therefore, reflect the presence/absence of patterns rather than token counts.

A second step of our analysis consisted of annotating the verbs obtained according to (a) Aktionsart, and (b) semantic class, information for which was retrieved from Open Multilingual Italian WordNet's SemFields (Pianta *et al.* 2002). Manual labeling was necessary because of the possibility, for each verb, to display more than one sense, leading it to belong to multiple SemFields.

Annotation of Aktionsart was manually carried out by the two authors, each on one half of the dataset. The annotation of both the semantic class and the actional class of the verbs allowed us to distinguish multiple senses for each verb, resulting in a final dataset of 350 verbs.²

The diagnostic tests used to determine the Aktionsart of the verbs are reported in Table 1.

² For the complete list of verbs, and their annotation, see below Section 5.

Table 1. Tests employed for the Aktionsart of the verbs

Test	Value	Implication
<i>is X-ing</i>	1 = yes, 0 = no	Compatibility with progressive/gerundive form; indicates dynamic/durative nature
<i>for T</i>	1 = yes, 0 = no, 2 = if iterated, semelfactive interpretation is triggered	Compatibility with “for + duration”
<i>in T</i>	1 = yes, 0 = no, 2 = culminative if telic	Compatibility with “in + duration”

Table 1 summarizes the main verbal properties annotated in our dataset and their operationalization. Aktionsart categorizes each verb according to the type of event it denotes, distinguishing between State, Punctual (Telic and Atelic), Activity, and Resultative verbs. The *is X-ing* test indicates whether the verb can appear in a progressive or gerundive form, serving as a measure of dynamicity and durativity. The columns *for T* (= time period) and *in T* assess temporal compatibility. Notably, the value 2 carries a specific semantic interpretation: in *for T*, it signals that iterativity points to a semelfactive interpretation, whereas in *in T*, it signals that a telic verb leads to a culminative interpretation when combined with *in T* structures. Values of 1 indicate compatibility without these additional properties, and 0 denotes incompatibility.

Table 2 summarizes the main patterns in the interaction between Aktionsart and the temporal/dynamic properties of verbs, as captured by the tests presented above.

Table 2. Implications of the tests on the Aktionsart of the verbs

Aktionsart	<i>is X-ing</i>	<i>for T</i>	<i>in T</i>	Example
Stative	0 / 1	1	0 / 2	<i>ricordare</i> ‘to remember’, <i>dimostrare</i> ‘to demonstrate’
Atelic punctual	0 / 1	0 / 2	0	<i>citare</i> ‘to cite’, <i>colpire</i> ‘to hit’
Telic punctual (Achievement)	0 / 1	0 / 2	2	<i>chiamare</i> ‘to call’, <i>creare</i> ‘to create’
Activity	1	1	0	<i>dormire</i> ‘to sleep’, <i>ridere</i> ‘to laugh’
Resultative	1	0 / 1	1	<i>sostituire</i> ‘to substitute’, <i>consegnare</i> ‘to deliver’

The last step consisted of annotating (c) the semantics of the intensified feature, that is, the aspect of the event intensified by the suffix, based on the annotation schema illustrated below. For this purpose, the authors were assigned half of the dataset that they had not annotated before, in order to prevent the influence of prior analyses of verbal properties on the interpretation of intensification.

2.2 Intensified feature(s) annotation

For the purpose of identifying the intensified aspects, we employed an original annotation schema (partly based on previous works about evaluation in the verbal domain: see Portero Muñoz 1997; Fleischhauer 2016; Napoli & Ravetto 2017; Audring *et al.* 2021), including both values intensifying the extension and the intensity of an event (cf. Bolinger 1972). The potentially intensified categories are INTENSITY, DURATION, COMPLETION, ARGUMENT QUANTITY, REITERATION, MANNER, and ILLOCUTIONARY FORCE, as described in Table 3. Based on this schema, we annotated the verbs to be found on X in the *-issimo* form, specifically obtaining sometimes more than one intensified aspect per verb. Operationalization relied on paraphrasing strategies, compatibility with temporal expressions or adverbs, and contextual inference (see Table 3).

Table 3. Tagset and tests for annotation of intensified aspect

Category	Explanation	Test / Operationalization
INTENSITY	Increase in the degree or force of the event compared to a neutral use of the verb	Paraphrase with <i>molto/tantissimo</i> ‘a lot’; substitutability with degree adverbs (e.g. <i>fortemente</i> ‘strongly’, <i>intensamente</i> ‘intensely’)
DURATION	Temporal extension of the event without necessarily implying a higher degree	Compatibility with <i>a lungo/per molto tempo</i> ‘for a long time’; inference of prolonged duration
COMPLETION	Emphasis on achieving the final point or outcome of the event	Compatibility with <i>completamente/del tutto</i> ‘completely’; inference that the result is fully achieved
ARGUMENT QUANTITY	Increase in the number or quantity of participants or objects involved in the event	Paraphrase with <i>molti/tanto</i> referring to objects or participants (e.g. <i>mangiare tantissimo</i> ‘eat a lot’)
REITERATION	Multiplication of occurrences of the event over time	Paraphrase with <i>ripetutamente/più e più volte</i> ‘again and again’
MANNER	Emphasis on the way the action is carried out (energy, accuracy, emphasis)	Paraphrase with manner adverbs (e.g., <i>con forza, con enfasi</i> ‘with strength, with emphasis’)
ILLOCUTIONARY FORCE	Emphasis on the speaker’s commitment or the pragmatic force of the speech act	Paraphrase with <i>davvero</i> ‘really’/ <i>te lo giuro</i> ‘I swear’/ <i>assolutamente</i> ‘absolutely’
AMELIORATION	Action is performed in a better way.	Paraphrase with <i>meglio</i> ‘better’

Using this schema, we annotated the types attested with *-issimo*, occasionally identifying more than one intensified aspect for a single verb. Each category captures a distinct dimension of intensification.

3. Results and discussion

In the following subsections, we briefly present the results of the annotation. First, we describe the occurrence of *-issimo* in our dataset, in relation to verb forms and semantics (Section 3.1). Next, we show how the annotated formal and functional categories intersect with the different types of intensified meaning (Section 3.2).

3.1 Distribution of *-issimo* among verbs

The query carried out on X shows that *-issimo* does not occur with all verbs in our dataset. Specifically, out of 350 verb lemmas (including relevant sense distinctions), only 201 (57.4%) are attested with *-issimo*. Interestingly, intensified forms with *-issimo* appear to occur more consistently with verbs characterized by a higher frequency (as recorded in the VoLIP corpus): testing the presence of *-issimo* against verb frequency rank using a Wilcoxon rank-sum test indicates a tendency for more frequent verbs to undergo intensification.

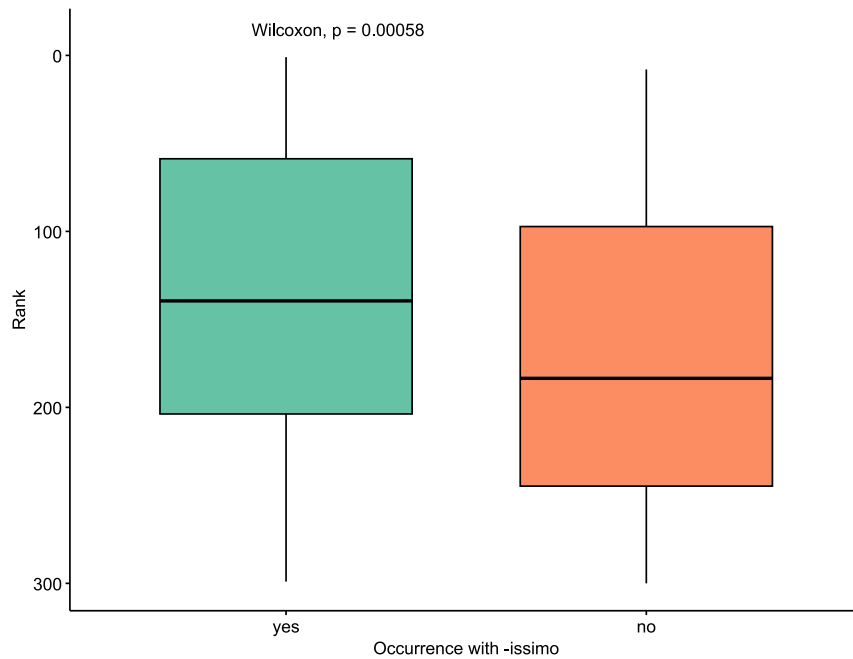


Figure 1. Frequency rank of the verb versus its occurrence with *-issimo*

In the following subsections, we address whether (and how) formal and semantic factors shape the distribution of *-issimo* in the dataset.

3.1.1 Verb forms

Not all verbal forms were found consistently in the dataset (Table 4). In fact, the most frequent forms are the Present Simple and the Present Perfect, attested for 60% and 62% of the verbs, respectively. By contrast, the Gerundive and the Imperfective appear to be the forms in which *-issimo* occurs least frequently, displayed only by 16% and 6% of the verbs.

Table 4. Number of verbs in which each form was found

Form	n. of verbs
Present Simple	125
Present Perfect	121
Infinite	67
Gerundive	33
Imperfective	14

As already mentioned (Section 1), we expected Present Perfect to be the most available verb form for suffixation in *-issimo*, due to its formal features. In fact, Present Perfect is an analytic tense made up of the auxiliary and the past participle. Thus, since *-issimo* is already employed productively with past participles functioning as adjectives, Present Perfect might have constituted a trigger for the analogic extension of this suffix from adjectival bases to verbal ones.

By contrast, the frequent use of *-issimo* with the Present Simple is unexpected, also because of its idiosyncratic behavior with respect to verbal inflectional morphemes. Generally, in non-finite forms, *-issimo* follows the last consonant of the verb form, whether the truncated vowel is morphemic or not (6), repeated from (2) above.

(6)

<i>Non</i>	<i>ho</i>	<i>visto</i>	<i>mezzo</i>
NEG	AUX.PRS.COND.1SG	see.INF	half.M.SG
<i>minuto,</i>	<i>dell'</i>	<i>Atalanta</i>	<i>ma</i>
minute.SG	of.DET.M.SG	Atalanta	but
<i>sto</i>	<i>godend-issimo</i>		
AUX.PRS.IND.1SG	pleasure.PROG-ELAT		

‘I haven’t watched half a minute of Atalanta’s match, but **I am strongly pleased** by the result.’

However, it may be the case that *-issimo* is inserted between the verb stem and the inflectional morpheme(s). This happens only in the case of participles, as shown by unaccusative verbs, where it allows the expression of gender and number (7).

(7)

<i>La</i>	<i>ship</i>	<i>ormai</i>	<i>è</i>	<i>partit-issim-a</i>
DET	(relation)ship	already	AUX.PRS.IND-3SG	start.PTCP-ELAT-F
<i>Date-mi</i>		<i>gioie</i>		
give.IMP.2PL-1SG.DAT		joy.PL		

‘The (relation)ship **has already powerfully started**. Give me joy’

In finite forms, overt person marking is often missing and is to be inferred from the context. However, sometimes it can be marked either before (8a) or after (8b) *-issimo*, but this last strategy is quite rare.

(8)

- a. *Lo* ***s-o-issimo*** *che tu lo sai*
 PRON know.PRS.IND-1SG-ELAT COMP 2SG.NOM PRON know.2SG
 ‘**I know very well** that you know’
- b. *La lontananza aiuta.* ***Aiut-issim-a***
 DET distance help.PRS.IND.3SG help.PRS.IND-ELAT-3SG
 ‘Distance helps. **It helps a lot**’

It is evident, thus, that the fact that the use of *-issimo* often obscures verbal inflection does not constitute a relevant constraint.

3.1.2 Semantic factors

Beyond formal factors, it is interesting to take into consideration the distribution of *-issimo* across verbs in relation to their semantic properties. First, we can consider the Aktionsart of the verb. As shown in Table 5, our dataset contains verbs from all Aktionsart classes, and all Aktionsart classes are modified by *-issimo*.

Table 5. Occurrence of *-issimo* by Aktionsart of the verbs

Class	Found	Not Found	Total
Achievement	61	39	100
Activity	46	28	74
Atelic	27	26	53
Punctual			
Resultative	31	34	65
State	36	22	58
<i>Total</i>	<i>201</i>	<i>149</i>	<i>350</i>

Chi-square test does not reveal any statistically significant differences between the Aktionsart classes in the presence of intensification with *-issimo* ($p = 0.27$). Analysis of the standardized residuals shows that classes characterized by the [-durative] feature (i.e., Resultative and Atelic Punctual) tend to occur less frequently. This is coherent with the observation made in the literature that evaluation is generally less felicitous with telic verbs or verbs used in their telic sense, i.e., resultatives and punctuals (Iacobini 2004; Grandi & Iacobini 2008; Napoli & Ravetto 2017). However, the differences are not pronounced enough to yield significant results.

Similarly, verbal semantics does not appear to constrain the presence of *-issimo* (Table 6). In this case as well, the chi-square test did not reveal a significant relationship between semantic classes and the possibility of intensification with *-issimo* ($p = 0.08$).

Table 6. Occurrence of *-issimo* by semantic class of the verbs³

Class	Found	Not Found	Total
Change	14	13	27
Communication	27	25	52
Mental (cognition, perception, emotion)	46	22	68
Motion	30	16	46
Object Related (consumption, creation)	16	20	36
Physical (body, contact)	13	5	18
Possession	10	12	22
Social (social, competition)	27	14	41
Stative	18	22	40
<i>Total</i>	<i>201</i>	<i>149</i>	<i>350</i>

It thus appears that the semantic distinctions considered do not constitute a real constraint on the distribution of *-issimo*, which is fairly generalized across quite diverse types of verbs.

3.2 Type of Intensification

In our dataset, all the Types of Intensification included in the classification (see Table 3) are attested. Table 7 shows how many types (i.e., individual verb-form combinations) are associated with each intensification subclass.

³ Semantic classes were lumped together in order to avoid sparseness due to low-frequency classes. We retained the original WordNet class labels in cases where no lumping was necessary, while for composite classes, we indicate in parentheses the WordNet semantic fields that compose them.

Table 7. Type frequency of the intensification subclass in the dataset

Intensification	n. types
ILLOCUTIONARY FORCE	121
INTENSITY/DEGREE	70
MANNER	62
COMPLETION	36
QUANTITY	29
REITERATION	19
DURATION	12
AMELIORATION	4

As can be seen, the most frequent class is by far ILLOCUTIONARY FORCE, suggesting that the primary function of *-issimo* with verbs may be pragmatic rather than semantic. Nevertheless, semantic values such as INTENSITY/DEGREE and MANNER are also well represented, whereas AMELIORATION, DURATION, and REITERATION occur only marginally.

To assess whether the type of intensification is influenced by any formal or semantic properties of the annotated verb forms, we cross-tabulated the intensification types in Table 7 with the variables annotated (Verb Form, Aktionsart, and Semantic Class). We then applied chi-square tests of independence, using simulated p-values in R, to assess significant associations. Since multiple tests were performed, we applied a Bonferroni correction to control for family-wise error rate, adjusting the significance threshold to $\alpha = 0.0167$ ($0.05/3$). In addition, we examined standardized residuals in order to identify which categories contribute most strongly to any observed associations. Finally, we computed Cramér's *V* to assess the strength of association between intensification type and each of the variables considered.

3.2.1 Type of intensification and verb tenses

The first variable considered is Verb Form. Table 8 shows the cross-tabulation between Verb Form and Type of Intensification. The chi-square test showed a weakly significant association ($p = 0.01$). Figure 2 shows the standardized residuals of the test:⁴ cells with values greater than 2 are likely to contribute to the significance of the test, indicating a positive association, whereas those with values lower than -2 indicate that they are likely to contribute as a negative association.

Table 8. Intensification Type by Verb Form

	Gerund	Infinitive	Imperfective	Present Perfect	Present
AMELIORATION	0	2	0	1	1
COMPLETION	3	2	1	25	5
DURATION	4	2	0	4	2
ILLOCUTIONARY FORCE	6	16	5	29	65
INTENSITY/ DEGREE	12	8	6	25	19
MANNER	15	5	2	15	25
QUANTITY	3	7	0	16	3
REITERATION	1	4	0	6	5

⁴ Cross-tables for all the investigated variables can be found at the online repository.



Figure 2. Standardized residuals for the chi-square test between Verb Form and Type of Intensification

Some relevant associations can be observed, in particular: on the one hand, the Present Simple (9) is strongly associated with ILLOCUTIONARY FORCE, while the Present Perfect (10) is associated with intensification understood as COMPLETION, consistently with the boundedness of the event expressed by this tense. Nonetheless, Cramér's V indicates a rather low correlation between Verb Form and Types of Intensification ($V = 0.14$).

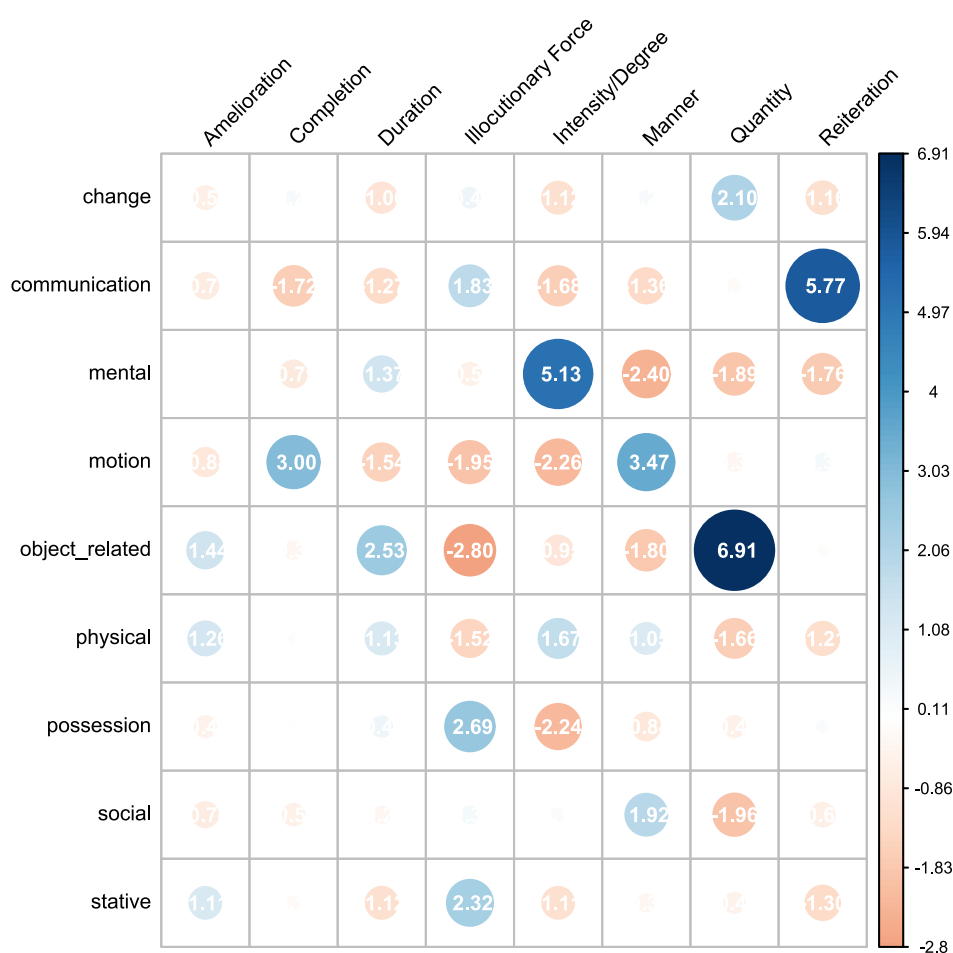
- (9) *e quando troverò una donna ossessionata da questa canzone come lo sono io avrò fatto scacco matto ve lo **giurissimo***
 ‘and when I’ll find a woman obsessed with this song as much as I am I’ll have checkmated I **sincerely swear**’
- (10) [...] *Solo altri quattro giorni di scuola E HO **FINITISSIMO***
 ‘[...] just four more school days and **I AM DONE FOREVER**’

3.2.2 Type of intensification and semantic classes

The second variable tested is the semantic class of the base. The contingency table crossing Semantic Class and Types of Intensification is shown in Table 9. In this case, the chi-square test yields a significant result ($p = 0.001$), and Cramér's V points to a higher association than with Verb Form (0.27). Figure 3 shows the standardized residuals of the test.

Table 9. Intensification by Semantic Class

	change	communication	mental	motion	object related	physical	possession	social	stative
AMELIORATION	0	0	1	0	1	1	0	0	1
COMPLETION	3	1	7	12	2	3	2	3	3
DURATION	0	0	5	0	3	2	1	1	0
ILLOCUTIONARY FORCE	10	19	28	13	2	6	12	14	17
INTENSITY/DEGREE	3	4	34	5	3	9	0	8	4
MANNER	5	4	8	19	1	7	2	11	5
QUANTITY	5	3	3	4	11	0	1	0	2
REITERATION	0	9	1	3	1	0	1	1	0

**Figure 3.** Standardized residuals for the chi-square test between SemField and Intensification

Indeed, not all classes display particularly strong associations. There are, however, some exceptions: verbs belonging to the object-related class (which was in turn based on the

SemFields of consumption and creation) are associated with intensification intended as ARGUMENT QUANTITY augmentation. For instance, in (11), what is intensified (i.e., augmented) is the amount, or the length, of the hair chopped off.

- (11) *Ho tagliatissimo i capelli [...]*
‘I cut a lot of hair [...]

Mental verbs (including the emotion, cognitive, and perception classes) show an association with INTENSITY/DEGREE (12), while communication verbs appear to be associated with a REITERATION interpretation (Y).

- (12) *vi amissimo*
‘I love you so much’

- (13) *Magari lo avete già dettissimo, ma [...]*
‘Maybe you already **talked about it over and over**, but [...]

There are also weaker but still meaningful associations, such as those between motion verbs and the two classes of MANNER and COMPLETION.

3.2.3 Type of intensification and actional classes

Table 10 shows how Types of Intensification intersect with Actional Classes in our dataset. The chi-square test yields a significant association ($p = 0.001$). Figure 4 shows the presence of many significant associations between Actional Classes and Types of Intensification, as reflected in the correlation coefficient ($V = 0.34$), which proves to be the highest among the variables investigated in our dataset.

Table 10. Intensification by Aktionsart of types

	Achievement	Activity	Punctual_atelic	Resultative	State
AMELIORATION	1	1	1	0	1
COMPLETION	20	3	3	6	4
DURATION	0	12	0	0	0
ILLOCUTIONARY FORCE	47	10	24	10	30
INTENSITY/DEGREE	14	27	2	4	23
MANNER	15	20	8	15	4
QUANTITY	3	8	1	12	5
REITERATION	4	5	6	0	1

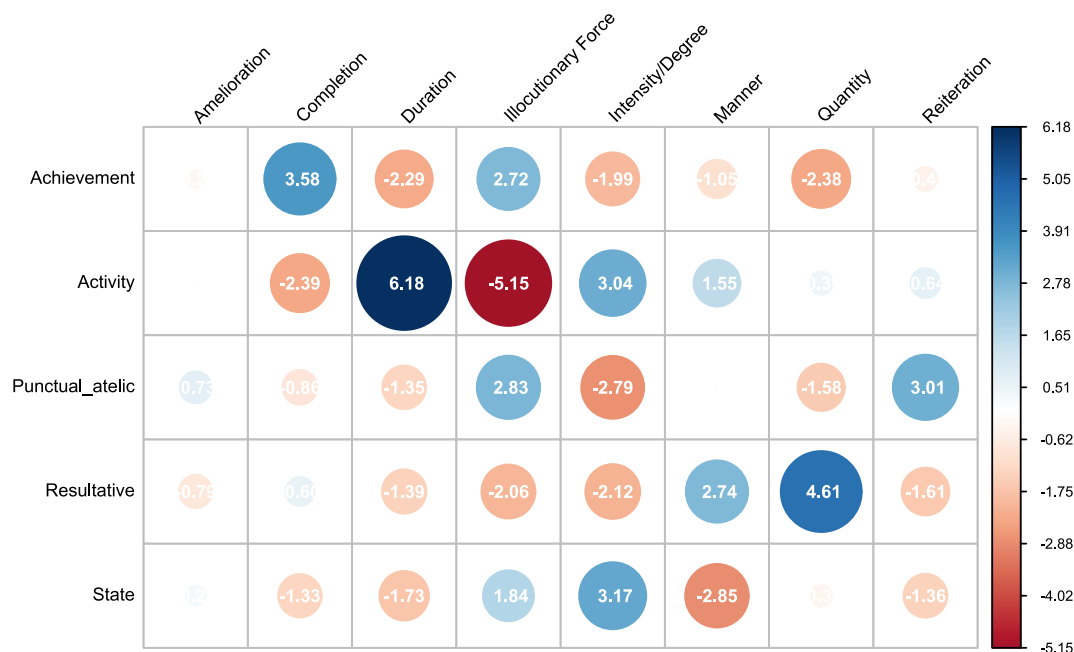


Figure 4. Standardized residuals for the chi-square test between Aktionsart and Intensification

As already observed, ILLOCUTIONARY FORCE is the most frequent type of intensification and is expressed across all actional classes. In (14), *-issimo* modifies a Resultative (6a) verb, a State verb (6b), and an Achievement verb (6c).

- (6) a. *Vabbè mi lancissimo [...]*
 ‘Anyway now **I go and throw myself out**’
 b. *[...] Ci sta! Ci staissimo!*
 ‘[...] I agree! I **totally agree!**’
 c. *Non approvavo Holland in generale, ma ho cambiatissimo idea*
 ‘I didn’t approve of Holland in general, but **I totally changed** my mind’

In these occurrences, *-issimo* functions as a way of strengthening the speaker’s statements, either signaling a higher commitment to the assertion made or by emphasizing its emotional content. Thus, because of the inherent underspecification of this category, it may need a finer-grained subdivision of the pragmatic functions performed. Nonetheless, the widespread presence of this class in the data is hardly surprising, considering that *-issimo* has been described as a marker that strengthens illocutionary force (Dressler & Merlini Barbaresi 1994: 509) or, more generally, as carrying expressive and emotional functions (Wierzbicka 1986). In the literature regarding adjectival evaluation, such pragmatic values arise secondarily from semantic intensification. In the case of verbal bases, it is quite true that pragmatic values are also present in occurrences in which *-issimo* performs other, more clearly semantic intensifying functions, but, in occurrences in which such semantic shades are not clear, the resorting to its pragmatic functions remains available.

Activity verbs constitute the only class showing some resistance to purely pragmatic uses, as shown by the negative association with ILLOCUTIONARY FORCE, and, on the contrary, by the strongly positive association with DURATION. Nonetheless, this association should not lead us to the assumption that most types of activity verbs express this semantics: in fact, these are only 12, but they run counter to the general trend of the other actional classes, which cannot intensify DURATION. Likewise, the negative association with ILLOCUTIONARY FORCE indicates that Activity verbs express this meaning less frequently than the others when evaluated with *-issimo*.

How can this distribution be explained? In the case of Activity verbs, intensification (or, more broadly, the evaluation) is more natural (i.e., frequent) than that of other actional types, as observed in previous studies (Grandi & Iacobini 2008; Napoli & Ravetto 2017). Consequently, the interpretation may be more straightforward, also because of the actional traits of this class of verbs. In particular, DURATIVITY emerges as a salient trait in these cases and is readily conceptualized, providing a clear basis for the intensifying effect of *-issimo* (15).

- (15) [...] *questa è dura, ci devo **pensarissimo***
 ‘[...] that’s hard, I need to **think about it a lot**’

However, in some cases, even when such an interpretation would be possible, the highlighted aspect is rather the greater INTENSITY with which the action is performed (16).

- (16) *Qui ho **risissimo***
 ‘Here I **laughed a lot**’

The fact that other actional classes lack more prototypically gradable features does not necessarily imply that the relevant occurrences are automatically interpreted in terms of ILLOCUTIONARY FORCE. In fact, all classes display significant associations with other types of intensification, which, although not the most frequent, are the most distinctive for each class.

For instance, State verbs display a high association with INTENSITY/DEGREE. This pattern is particularly pronounced when the semantic class of a certain verb is Mental, to which many State verbs belong, as we see in (11). In such cases, *-issimo* seems to denote a higher intensity of the psychological state experienced.

- (17) *Se succedesse, e io **sperissimo** di no, ci sarà da piangere e da mandare giù parecchi rospi*
 ‘If that happens, and I **wholeheartedly hope** it doesn’t, we’ll have to cry and swallow the bitter pill’

An interpretation overtly connected to actional traits is often triggered by Atelic Punctual verbs, which, when combined with the elative, are characterized, on the one hand, by their resistance to intensification in terms of INTENSITY/DEGREE, and on the other, by a tendency to convey the REITERATION of the event (18).

- (18) *mizzica lo **usissimo** anche quando parlo [...]*
 ‘I **use mizzica very often** also when I speak [...]

Some correspondence with the most salient trait is also found in the case of Achievement verbs, which seem to get intensified for COMPLETION: in (19), for example, the spike of “being done” corresponds to “being done *forever*”, to not have to step into a school anymore in the future. The highlight is, then, the definitiveness of the fact.

- (19) [...] *Solo altri quattro giorni di scuola E HO **FINITISSIMO***
 ‘[...] just four more school days AND I AM **DONE FOREVER**’

Lastly, unlike these cases, Resultative verbs do not display an immediate connection to the aspect they get intensified by, but instead, their compatibility with the ARGUMENT QUANTITY class is better understood in terms of semantic class (see Section 3.2.2).

To sum up, *-issimo* displays the ability to express a wide range of intensification meanings. While it is true that its use with verbs that are not typically gradable tends to trigger primarily pragmatic interpretations, Aktionsart and certain semantic classes nonetheless appear to

contribute significantly to determining (or further specifying) which aspect of the event is being intensified. Table 11 summarizes some observations regarding the connections between Aktionsart, semantic class and intensified features.

Table 11. Schematization of intensification type based on Aktionsart and Semantic Class

Activity verbs		→		activity is PROLONGED
State verbs	→	<i>mental</i>	→	state is INTENSIFIED
Atelic punctual verbs		→		action is REPEATED
Achievement verbs		→		action is COMPLETED
Resultative verbs	→	<i>object-related</i>	→	result is INCREASED

4. Limitations and future outlook

In the previous sections, we investigated a creative verbal evaluation strategy, primarily employed in online and playful contexts, and proposed a preliminary schema of the possible aspects that may constitute the scope of intensification for verbal bases. Our results confirm the relation between some actional classes and some intensification values already noticed in the literature (Napoli & Ravetto 2017). While, as we saw, some recurrent tendencies emerged, our analysis also shows that intensified features are largely unpredictable when relying exclusively on verbal classes. Consequently, the meaning of the suffix *-issimo* remains underspecified when occurrences are considered out of context.

More generally, classifying evaluation in the verbal domain proves to be a non-trivial task, as the traditional oppositions BIG-SMALL and GOOD-BAD do not map transparently onto verbal semantics. This difficulty becomes particularly evident when considering categories such as MANNER and AMELIORATION. With respect to the latter, the (few) attestations (20) raise the question of whether a corresponding PEJORATION class of verbal evaluation may exist, or whether both phenomena should instead be subsumed under a broader QUALITATIVE evaluation type. Notably, PEJORATION was not attested in our dataset, likely due to the semantics of the suffix under investigation, despite being among the evaluative values previously identified for verbal morphology (see above, Section 1).

- (20) [...] sanno *suonarissimo*.
 ‘[...] they can **play very well**’

By contrast, the distribution of MANNER intensification appears less problematic (21): its strong association with motion verbs, and its absence with stative predicates can plausibly be accounted for in terms of the semantic compatibility between manner-related scaling and eventive structure.

- (21) *Scappissimo dall'afa romana ciao*
 ‘I’m immediately running away from Roman heat byeeee’

Finally, it should be acknowledged that the informal and playful nature of the environments from which our data were drawn may have influenced both the frequency and the functions of the evaluative strategies observed. For this reason, the proposed classification should be regarded as preliminary and open to further validation. Future research could test its robustness by extending the analysis to other verbal intensification strategies in Italian, such as prefixal formations, and by considering a wider range of registers and communicative contexts.

Data availability statement

Data, including the annotation, verb list, and complete outcomes of the investigation are available through the Github repository: <https://github.com/fla-pi/verb-issimo>.

Abbreviations

DIM = diminutive

ELAT = elative

References

- Audring, Jenny, Sterre Leufkens & Eva van Lier. 2021. Small events: Verbal diminutives in the languages of the world. *Linguistic Typology at the Crossroads* 1(1). 223-256. <https://doi.org/10.6092/issn.2785-0943/13427>.
- Berezowska, Małgorzata. 2017. Intensificazione del significato tramite l'uso dei prefissi e dei suffissi superlativi nel linguaggio giovanile e nello stile femminile [Intensification of meaning through elative prefixes and suffixes in youth language and feminine style]. *Studia de Cultura* 9(1). 19-28.
- Bolinger, Dwight. 1972. *Degree words*. Boston: DeGruyter. <https://doi.org/10.24917/20837275.9.1.2>.
- Cacchiani, Silvia. 2011. Intensifying affixes across Italian and English. *Poznań Studies in Contemporary Linguistics* 47(4). 758-794. <https://doi.org/10.2478/psicl-2011-0038>.
- De Mauro, Tullio. 2003. *Guida all'uso delle parole: Parlare e scrivere semplice e preciso per capire e farsi capire* [Guide to use of words. To talk and write easy and precise to understand and be understood]. Roma: Editori Laterza.
- Dressler, Wolfgang U. & Lavinia Merlini Barbaresi. 1994. *Morphopragmatics. Diminutives and Intensifiers in Italian, German, and Other Languages*. Boston: DeGruyter.
- Fleischhauer, Joachim. 2016. *Degree gradation of verbs*. Düsseldorf: Düsseldorf University Press.
- Gaeta, Livio. 2011. Superlativo [Superlative]. In Raffaele Simone (ed.), *Enciclopedia dell'italiano* [Encyclopedia of Italian]. Roma: Istituto della Enciclopedia Italiana Giovanni Treccani. [https://www.treccani.it/enciclopedia/superlativo_\(Enciclopedia-dell'Italiano\)/](https://www.treccani.it/enciclopedia/superlativo_(Enciclopedia-dell'Italiano)/).
- Giacinti, Fernando. 2024. Il suffisso valutativo -issimo nell'italiano nel mondo [The -issimo evaluative suffix in world Italian]. In Raffaella Bombi, Francesco Costantini, Fernando Giacinti & Diego Sidraschi (eds.), *Italiano nel mondo e lingue minoritarie dentro e fuori i confini italiani: valori identitari e imprenditorialità* [Italian in the world and minority languages within and outside Italian borders: identity values and entrepreneurship], 39-57. Udine: Forum Editrice Universitaria Udinese.
- Grandi, Nicola. 2002. *Morfologie in contatto: Le costruzioni valutative nelle lingue del Mediterraneo* [Morphologies in contact. Evaluative constructions in the languages of the Mediterranean area]. Milano: FrancoAngeli.
- Grandi, Nicola. 2008. *I verbi deverbali suffissati in italiano: Dai dizionari al web* [Deverbal suffixed verbs in Italian. From dictionaries to the web]. Alessandria: Caissa Editore.
- Grandi, Nicola & Claudio Iacobini. 2008. L'affissazione valutativa nei verbi dell'italiano [Evaluative affixation in Italian verbs]. In Emanuela Cresti (ed.), *Nuove prospettive nello studio del lessico. Atti del IX Congresso Internazionale SILFI (Firenze, 15-17 giugno 2006)* [New perspectives on the study of the lexicon. Proceedings of the IX International Conference SILFI (Florence, 15-17 June 2006)], 483-490. Firenze: Firenze University Press.
- Iacobini, Claudio. 2004. Prefissazione [Prefixation]. In Maria Grossmann & Franz Rainer (eds.), *La formazione delle parole in italiano* [Word formation in Italian], 97-164. Tübingen: DeGruyter Mouton.
- Merlini Barbaresi, Lavinia. 2004. Aggettivi deaggettivali [Deadjectival adjectives]. In Maria Grossmann & Franz Rainer (eds.), *La formazione delle parole in italiano* [Word formation in Italian], 444-449. Tübingen: Niemeyer.
- Montermini, Fabio. 2024. Same source, different outcomes -issime vs. -(i)ssimo in contemporary French. Presented at *Unlocking Evaluative Morphology Berlin, 14-15 March 2024*.
- Napoli, Maria. 2013. On Italian past participles with -issimo: The superlative of events between intensification and pluractionality. *Linguistica e Filologia* 33. 85-126.

- Napoli, Maria & Miriam Ravetto. 2017. Ways to intensify: Types of intensified meanings in Italian and German. In Maria Napoli & Miriam Ravetto (eds.), *Exploring intensification: Synchronic, diachronic and cross-linguistic perspectives*, 327–352. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/slcs.189.17nap>.
- Pianta, Emanuele, Luisa Bentivogli & Christian Girardi. 2002. MultiWordNet: developing an aligned multilingual database. In *Proceedings of the First International Conference on Global WordNet*, Mysore, India, 293–302. Global WordNet Association.
- Portero Muñoz, María Carmen. 1997. Syntactically relevant semantic features. *Alfinge. Revista de Filología* 9. 333–354. <https://doi.org/10.21071/arf.v9i.7173>.
- Rainer, Franz. 2003. Studying restrictions on patterns of word-formation by means of the Internet. *Italian Journal of Linguistics* 15.131–140.
- Santiloni, Viola Amelia. 2026. “Ma che stai dicendo?” A Linguistic Overview of Teen Representation in Italian Television Series. In: Luca Barra, Danielle Hipkins, Catherine O'Rawe & Dana Renga (eds.), *Contemporary Italian Youth Television*, 109–119. Cham: Palgrave Macmillan. https://doi.org/10.1007/978-3-031-98064-0_8.
- Voghera, Miriam, Claudio Iacobini, Renata Savy, Francesco Cutugno, Aurelio De Rosa & Iolanda Alfano. 2014. VoLIP: a searchable Italian spoken corpus. In Ludmila Veselovská & Merkéta Janebová (eds.), *Complex visibles out there. Proceedings of the Olomouc Linguistics Colloquium: Language Use and Linguistic Structure*, 628–640. Olomouc: Palacký University.
- Wierzbicka, Anna. 1986. Italian reduplication: cross-cultural pragmatics and illocutionary semantics. *Linguistics* 24. 287–315.
- Zilg, Antje. 2011. THAT'S AMORE: brand names in the Italian food market. *International Journal of Applied Linguistics* 21. 1–25. <https://doi.org/10.1111/j.1473-4192.2009.00252.x>.

A manner of evaluating: Italian deverbal evaluative motion verbs

Luisa Corona
University of L'Aquila
luisa.corona@univaq.it

Luisa Brucale
University of Palermo
luisa.brucale@unipa.it

Egle Mocciaro
Masaryk University
egle.mocciaro@mail.muni.cz

1. Introduction

This paper examines the interaction between the semantic-functional domain of evaluation and that of manner as encoded in Italian deverbal evaluative verbs (hereinafter DEVs), that is, verbs derived from other verbs by means of evaluative suffixes, focusing in particular on motion verbs (e.g. *saltellare* ‘to hop’ < *saltare* ‘to jump’).

Against the backdrop of the relevant literature on evaluation (Dressler & Merlini Barbaresi 1994; Grandi 2002, 2017; Grandi & Körtvélyessy 2015) and manner (Slobin 2004, 2005, 2006; Stosic 2020), we propose that the overlap between the two domains produces a semantic area in which manner corresponds to a specific evaluative nuance through which the speaker refers to a non-uniform movement that an agent performs in a disengaged way and/or in a scattered space. In other words, the speaker represents the motion event as being executed in a non-standard or non-prototypical manner.

This view is consistent with the idea, which has gained ground in recent decades, that evaluation plays a central role in categorisation. According to Grandi (2002: 29), an evaluative construction encodes a subjective judgement about the conformity or deviation of an entity or event with respect to a culturally shared standard. The morphological encoding of evaluation entrusts the expression of the standard to the base (e.g. *salt-*) and that of deviation to the evaluative affix (e.g. *-ellare*). In this way, the speaker signals that the meaning of the base is realised in an atypical or *inexact* way (Grandi 2017: 147-148). We will describe this inaccuracy in terms of *approximation* (see Masini et al. 2023 and, in particular, Brucale & Mocciaro 2023) meaning the degree of distance that separates a certain entity from the prototypical realisation of the category to which it belongs (see also Stosic & Amiot 2011: 425).

As regards verbs, the deviation expressed by evaluative affixation has been described in works focusing on different languages, such as French (Amiot & Stosic 2011; Moline & Stosic 2016; Stosic & Amiot 2011, 2019), Italian (Bertinetto 2004; Grandi 2008; Tovenà 2011, 2015), Italian and Spanish (Artusi 2019), and Serbian (Stosic 2013). These studies highlighted the role played by the actional properties of bases (mainly, atelic verbs expressing durative activities such as ‘to read’) in the construction of evaluative meaning (Section 2), which, in the case of motion verbs, translates into a preference for bases containing a manner component rather than a path component (e.g. ‘to run’ vs ‘to leave’; see Section 3).

However, evaluation in the verb is still a largely unexplored area, and many questions remain unanswered. The work presented here aims to help bridge this gap by addressing the following research questions:

1. How and through which properties is the notion of manner specified in the case of motion DEVs, and how does it interact with evaluation?
2. What role does the context in which the DEV occurs play in determining the specific evaluative nuance?

To answer these questions, we extracted from the ItWac wordlist the verbs derived with the most productive evaluative suffixes in Italian (Grandi 2008), that is, *-Vcchiare* (93 verbs), *-ellare* (7 verbs), *-er-ellare* (7 verbs), and *-ottare* (7 verbs). We then classified both the DEVs and their bases based on (an adaptation of) the lexical categories of the WordNet ontology, as reported in Table 1.

Table 1. WordNet-based lexical categories used in the classification

WordNet lexical category	Examples
Bodily functions and care	'sweat', 'wash'
Change	<i>-ize</i> and <i>-ify</i> verbs
Communication	'lisp'
Consumption	'drink'
Contact	'grasp', 'squeeze'
Cognition	'infer'
Creation	'sew'
Motion	'travel', 'swim'
Emotion and psychological states	'fear'
Stative	'surround'
Perception	'see'
Possession	'have'
Social interaction and competition	'play'
Weather	'rain'

Manner is a key component in the WordNet ontology and works as a hyponymic semantic relation, called “troponymy”, crossing all the lexical classes and based on which “To V_1 is to V_2 in some particular manner” (Fellbaum 1990: 285), e.g. ‘limping’ is a manner of ‘walking’, which is a manner of ‘moving’. This approach proved extremely useful for our analysis, because it revealed which lexical classes are more sensitive to a manner reading (see Brucalè, Corona & Mocciaro forthcoming), as we will show in Section 4. However, its rigorously non-decompositional nature proves inefficient when it comes to understanding what the DEVs most prone to manner interpretation have in common. As the data analysis suggests, in fact, the manner interpretation of DEVs does not depend on the overall semantics of verbs but on certain more specific semantic properties, such as agentivity and telicity. In many cases, these properties seem to be metonymically activated (or not) by the context in which DEVs occur. Verifying this hypothesis requires narrowing the focus to a smaller group of DEVs from the 114 extracted in order to give more space to the analysis of their context of use. For this reason, we have decided to limit the analysis, at this stage, to motion verbs only, which constitute the most frequent lexical type (see Section 4.1). Based on this qualitative analysis, we will draw some conclusions and propose hypotheses for further investigation (Section 5).

2. Deverbal evaluative verbs: What has been said

To date, still a few studies are dedicated to evaluative verbs in Romance languages. In general terms, evaluative suffixation in verbs does not produce a transcategorisation (Bertinetto 2004: 465) but modifies the semantics of the event, blending in its intensity, uniformity or manner of execution. For instance, Tovenà (2011: 43) describes the action of *mangiucchiare* ‘to nibble’ “as being performed with a kind of reduced engagement by the agent, that shows in smaller bites and lack of continuity”. Bertinetto (2004: 470) talked about a constellation of frequently overlapping meanings, including events occurring: a) without real commitment and with an attenuated result; b) with exaggerated insistence and/or with a worse-than-normal result; c) with great intensity and/or not precisely spatially directed; d) repeatedly or iteratively. Grandi (2008, 2009) identified four semantic nuclei: superficiality (e.g. *studiacchiare* ‘to study halfheartedly’), attenuation (e.g. *vivacchiare* ‘to get by, to subsist’), (re)iteration (e.g. *saltellare* ‘to hop’), and rapidity (e.g. *becchettare* ‘to peck’). Reiteration can be further specified as event-internal pluractionality (Amiot & Stosic 2011; Tovenà 2011, 2015; see also Cusic 1981; Mattioli 2019), whereby a single event is fragmented in a series of smaller and repeated subevents. In terms of *Aktionsart*, the evaluative process favours activity (‘to read’) or active accomplishment (‘to read a book’) bases, which are durative, dynamic and atelic (or non-inherently telic), as are their derivatives; accomplishments are less represented; statives and achievements are insignificant (Bertinetto 2004; Grandi 2008; Tovenà 2015). In Italian, most activities and active accomplishments exhibit unergative morphosyntax involving the perfective auxiliary ‘to have’ (vs ‘to be’ for unaccusatives) (Grandi 2015: 1476).

3. Manner and motion

In the panorama of approaches to the notion of manner, we take as our reference point the studies by Slobin (2004, 2006) within the framework of motion events lexical typology (Talmy 2000). Unlike other authors who consider it a semantic primitive (Jackendoff 1983: 48-56; Levin & Rappaport Hovav 2013 inter al.), Slobin (2004: 255) defines Manner as “an ill-defined set of dimensions that modulate motion”. This set of dimensions remains, however, largely undefined (Corona 2025: 195), and several studies have recently highlighted the need to focus on the complex set of linguistic means and strategies used to encode this notion (Corona & Pietrandrea 2021; Minoccheri & Stosic 2022). In fact, manner can be encoded at different levels. The most studied level is syntax, in particular the use of various circumstantial modifiers (adverbs, phrases, subordinate clauses, gerunds). By contrast, morphological and suprasegmental encodings remain still less explored (see Amiot & Stosic 2011; Brucale, Corona & Mocciaro forthcoming; Stosic 2013). The relationship between different levels of encoding, for example the intertwining of morphological and syntactic markers, also remains unexplored. On the other hand, on a semantic level, the intricate network of connections with a series of related notions, including attenuation, approximation and pluractionality, remains to be thoroughly investigated.

4. Data analysis

4.1 General overview

As mentioned in Section 1, we extracted from ItWac all the DEVs in *-Vcchiare*, *-ellare*, *-er-ellare*, and *-ottare*, i.e., the most productive suffixes, which form first conjugation verbs in *-are* (Bertinetto 2004: 467; Grandi 2008). Of the 114 DEVs obtained (from 103 verb bases), 81.58% are in *-Vcchiare* (93), 6.14% in *-er-ellare*, in *-ottare* and in *-ellare* (7 verbs each).

We classified both the bases and DEVs based on WordNet lexical classes. As Graphs 1 and 2 show, the most frequent lexical class in terms of type is, in both cases, motion (see Stosic 2019), followed by communication and social interaction (respectively, 16, 14, and 12 bases; 21, 18 and 13 DEVs); the less frequent are creation, weather and emotion/psychological states, while some fluctuation is observable in the middle of both hierarchies.

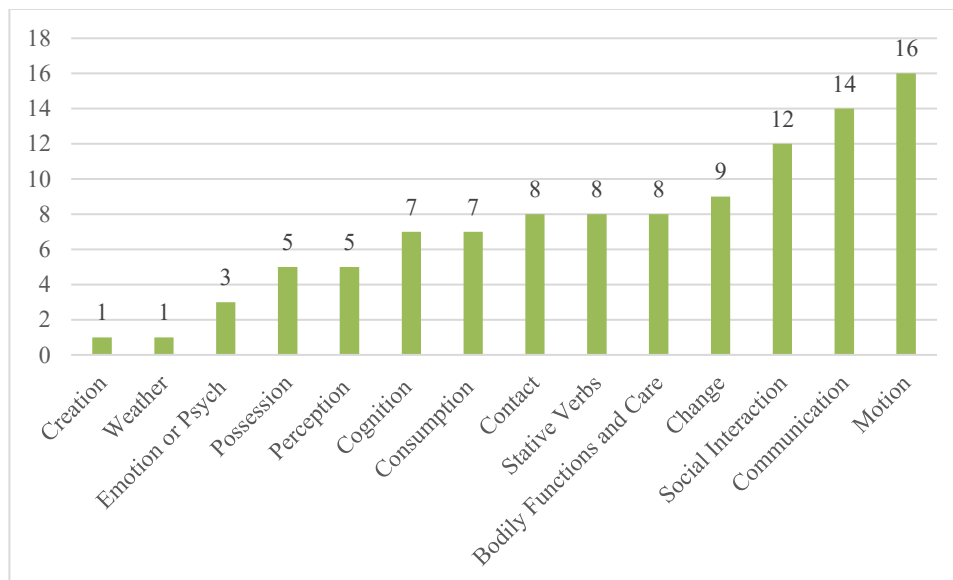


Figure 1. Types of verbal bases in ItWaC according to the WordNet Ontology

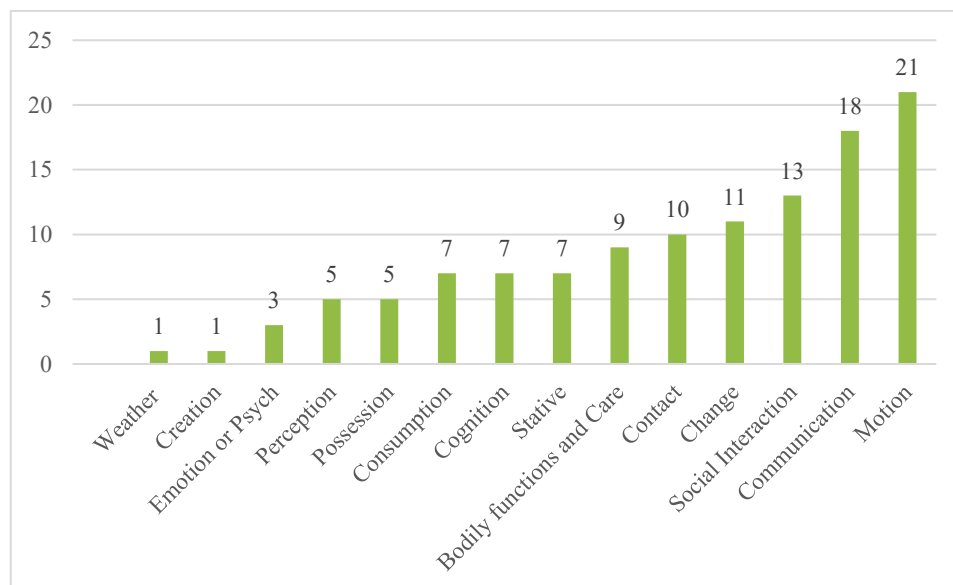


Figure 2. Types of DEVs in ItWaC according to the WordNet Ontology

To obtain more detailed information, we also extracted DEV token frequency. Graph 3 shows that motion drops to second place after communication, while maintaining a still significant frequency (3375).

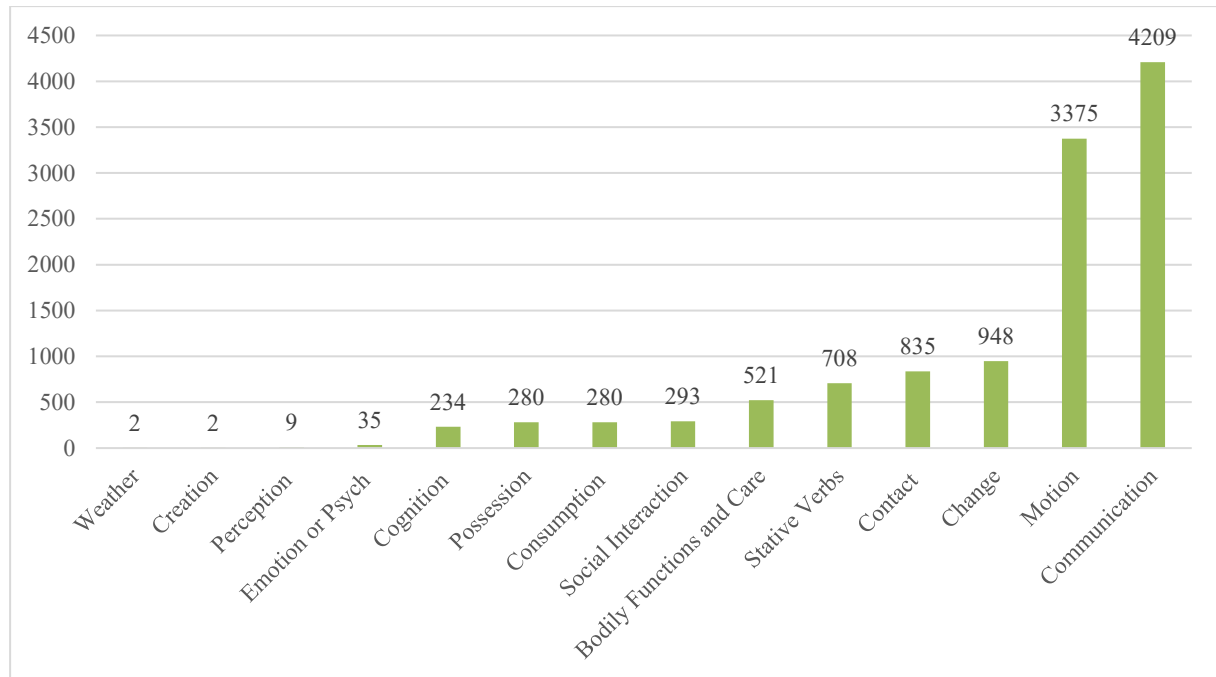


Figure 3. Tokens in ItWaC according to the WordNet Ontology

In what follows, we will propose a quantitative and qualitative description of motion DEVs and the discourse context in which they occur.

4.2. Motion Deverbal Evaluative verbs

4.2.1 Motion verbs as bases

Graph 4 shows a marked imbalance between a small number of high-frequency motion verbs and a long tail of low-frequency verbs. Three verbs cover 94.74% of occurrences, i.e., *saltellare* (2645 occurrences, i.e., 78.37%), *trotterellare* (389 occurrences, i.e., 11.52%) and *girellare* (164 occurrences, i.e., 4.85%). The remaining verbs have values close to zero.

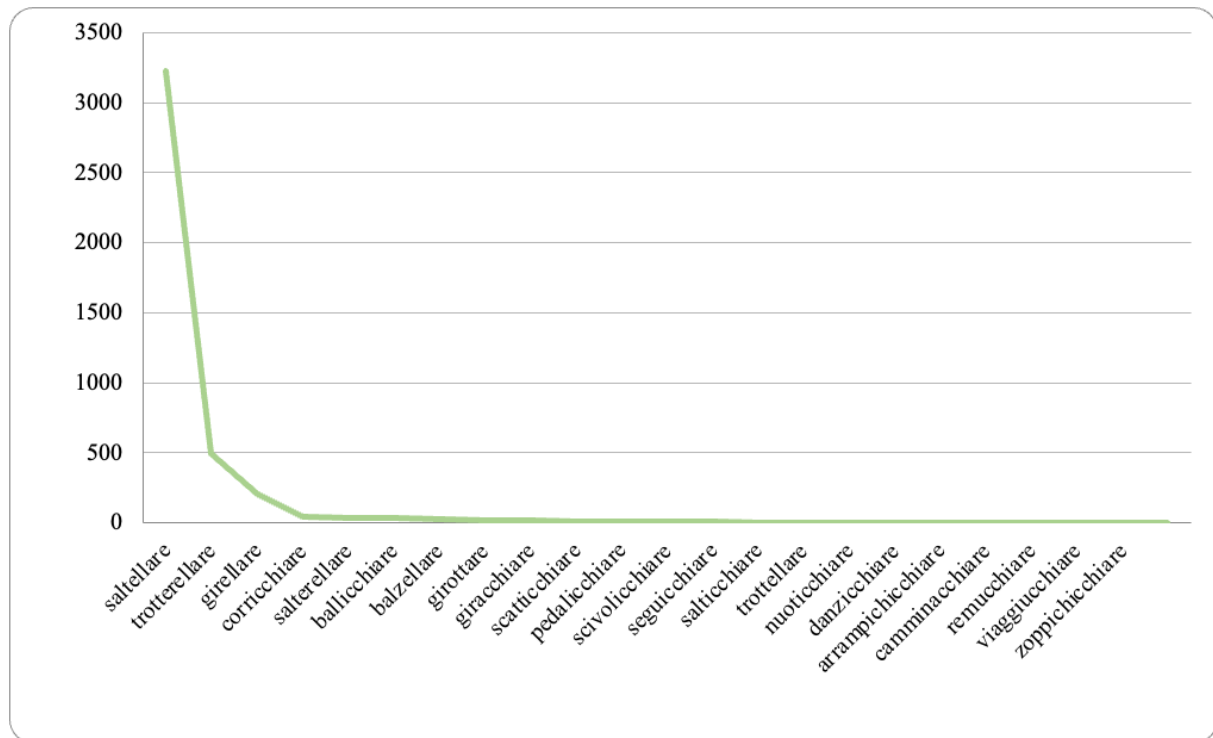


Figure 4. Motion verbs

As Table 2 shows, 12 out of 16 bases are manner of motion verbs, two encode the path (*girare* ‘to turn’, *seguire* ‘to follow’), one both path and manner (*arrampicare* ‘to climb’; cf. Levin & Rappaport Hovav 2013), while the last is neutral (*viaggiare*), i.e., it does not encode deixis, manner or path (cf. Verkerk 2014: 47, 52).

Table 2. The bases of evaluative motion verbs.

MANNER	PATH	PATH+MANNER	NEUTRAL
<i>ballare</i> ‘to dance’	<i>girare</i> ‘to go around’	<i>arrampicare</i> ‘to climb’	<i>viaggiare</i> ‘to travel’
<i>balzare</i> ‘to leap’	<i>seguire</i> ‘to follow’		
<i>camminare</i> ‘to walk’			
<i>correre</i> ‘to run’			
<i>danzare</i> ‘to dance’			
<i>nuotare</i> ‘to swim’			
<i>pedalare</i> ‘to pedal’			
<i>saltare</i> ‘to jump’			
<i>scattare</i> ‘to burst forth’			
<i>scivolare</i> ‘to slide’			
<i>trottare</i> ‘to trot’			
<i>zoppicare</i> ‘to limp’			

The distribution in Table 2 is consistent with the previously observed preference of DEVs for atelic, dynamic and durative bases (Bertinetto 2004; Grandi 2008; Stosic 2019; Tovenà 2011). The qualitative analysis of the context provides further information on (i) what semantic contribution suffixes make to bases that already denote manner; (ii) whether and how a manner reading can emerge from DEVs derived from bases expressing path; (iii) the possible use of DEVs in telicity contexts; iv) agentivity as a central property for manner interpretation.

4.2.2 Manner of motion bases

Two distinct classes of manner evaluative verbs can be identified.

The first class includes:

- (i) low-manner bases, i.e., “everyday verbs like run and walk” (Slobin 2004: 74), e.g., *corricchiare*, *camminicchiare*;
- (ii) highly lexicalised verbs, with high frequency and early attestations (between 15th and 19th century), e.g. *saltellare*, *salterellare*, *balzellare*, and *trotterellare*.

These DEVs may co-occur with contextual elements reinforcing the specific manner of carrying out the activity, e.g. *a testa bassa* in (1) and *come cuccioli allegri e irrefrenabili* in (2):

- (1) *I giallorossi hanno corricchiato a testa bassa, scambiandosi poche parole*
‘The giallorossi barely ran, heads down, exchanging a few words.’
- (2) *Senza paura di farsi vedere sui prati o sulle strade asfaltate mentre saltellano come cuccioli allegri e irrefrenabili*
‘Without fear of being seen on the fields or the paved streets while they hop around like happy and uncontrollable puppies.’

Contexts are most typically atelic, including for DEVs deriving from bases allowing for an “aspectual reclassification” (Cennamo & Lenci 2019: 381), i.e., showing both atelic (*ho corso nel parco* ‘I ran in the park’; *ha saltato su e giù per due ore* ‘he jumped up and down for two hours’) and telic patterns (*sono corsa in casa* ‘I ran home’; *è saltato dal quarto piano* ‘he jumped off the fourth floor’). This alternation is reflected in a different perfective auxiliary selection (*avere* in the first case, *essere* ‘to be’ in the second). The characteristics of atelicity or telicity can be reinforced or even entirely entrusted to the context. *Corricchiare* (39 tokens) has only atelic usages, possibly locally reinforced by prepositional phrases, such as the extended ground *per il campo* ‘throughout the field’ in (3), or adverbials expressing spatial distribution/pluractionality, e.g. *di qua e di là* ‘here and there’ in (4):

- (3) *I daltonici corricchiavano per il campo cercando d’evitare il pallone*
‘The color-blind players were running around the field, trying to avoid the ball.’
- (4) *io non vado a corricchiare di qua e di là per tentare di conservare il mio posto*
‘I’m not running here and there just to try and hold onto my place.’

On the other hand, DEVs based on *saltare* (*saltellare*, 2,644 tokens; *salterellare*, 30 tokens; *salticchiare*, 3 tokens) may sometimes occur in telic contexts (218 out of 2,677 examples, i.e., less than 8% of the tokens). Telicity is entirely given by locative particles, as in *saltellare giù* in (5) (cf. Iacobini & Masini 2006), or prepositional phrases encoding the goal of motion, as *fino alla sua sedia* in (6):

- (5) *Saltelliamo giù, senza mai guardare in basso*
‘We jump down, never looking down.’
- (6) *Il ranocchio entrò e, sempre dietro a lei, saltellò fino alla sua sedia*
‘The frog came in and, still following her, hopped all the way to her chair.’

The second class includes verbs based on high manner of motion verbs, i.e., “more expressive verbs such as crawl, stomp, roll” (Slobin 2006: 74), which are less frequent in the corpus and not attested in the main Italian lexicographic resources, e.g. *ballicchiare* < *ballare* ‘to dance’ (7), *pedalicchiare* < *pedalare* ‘to pedal’, *scivolicchiare* < *scivolare* ‘to slip’, *nuoticchiare* < *nuotare* ‘to swim’, *danzicchiare* < *danzare* ‘to dance’, and *zoppichicchiare* < *zoppicare* ‘to limp’. In these cases, the manner component is only encoded in the base, while the suffix specifies that the meaning of the base is only attenuated; this may be contextually reinforced, e.g. through the paucal quantifier *un po’* ‘a little’ in (8):

- (7) *Debutta il nuovo spettacolo di Dado dal titolo " Suono, canto, recito... e ballicchio!"*
 ‘The new show by Dado makes its debut, titled ‘Sound, Song, Acting... and Little Dancing!’
- (8) *Adesso si vedono anche le prime biciclette elettriche: si pedalicchia un po' per aiutare il motorino, vanno piano e hanno poca carica, ma sono antisudore.*
 ‘Now we’re also seeing the first electric bikes: you pedal a bit to help the motor, they go slow, and the battery doesn’t last long, but they’re sweat-proof.’

4.2.3 Low-agentive manner of motion bases

Among the high-manner verbs discussed in 4.2.2, we also find a few low-agentive DEVs. *Scivolicchiare* < *scivolare* ‘to slip’ has 2 occurrences in ItWaC, while *zoppichicchiare* < *zoppicare* ‘to limp’ is attested only once. However, this hapax, reported in (9), is used (just like its co-occurring base) metaphorically to encode a purely “fictive motion” (Talmy 2000) that actually describes the *instability*, that is, unreliability of examples of certain types of vandalism. To compensate for this lack of information, we conducted a targeted Google search for contexts in which the verb encodes physical movement involving animate agents; these are reported in (10)-(11).

- (9) *gli esempi zoppicano sempre e non portano a nulla se vogliamo fare paragoni prendiamo il vandalo della Pietà e quelli del Biancone, chi allaga un liceo o chi da fuoco a un teatro (anche questo zoppichicchia un po')*
 ‘The examples always limp and lead nowhere; if we want to make comparisons, we should look at the vandal of the *Pietà* and those of the *Biancone*, or at those who flood a high school or set a theatre on fire (this example, too, limps a bit)’
- (10) *Ieri sera Viola (4 anni meno 3 mesi...) e' caduta dall'altalena... (...) stamattina quando lo poggia dice che le fa "Solo un po' male, ma non tanto..." e "zoppichicchia". E' andata all'asilo, mi ha detto "sto bene e zoppico solo un po'"...*
 ‘Last night Viola (four years old minus three months...) fell off the swing. (...) This morning, when she puts her foot down, she says it “only hurts a little, not much,” and that she’s “limping a bit.” She went to kindergarten and told me, “I’m fine, I just limp a little.’
- (11) *zoppica, o meglio, zoppichicchia quando stanca, dalla zampa diametralmente opposta a quella della ipotetica (...) displasia.*
 ‘It limped, or rather, it still limps when tired, from the paw directly opposite to the one affected by the hypothetical dysplasia.’

While *zoppicare* and *zoppicchiare* involve low control by an animate subject, the roll-verb *scivolicchiare* may refer, like its base, to both animate, as in (12), and inanimate entities with no control over the movement involving them (Levin 1993: 265), as in (13):

- (12) *quando cammini per casa a piedi scalzi scivolicchi sulle pozze di sudore da te in continuazione secreto*
 ‘When you walk around the house barefoot, you slip on the puddles of sweat you constantly secrete.’
- (13) *In discoteca non sarebbe più appropriata una scarpa normale che possa scivolicchiare ma che sia dal look più elegante?*
 ‘Wouldn’t a regular shoe that can slip a bit but has a more elegant look be more appropriate for the club?’

In all cases, low or no agentivity inhibits manner interpretation, while DEVs are used to approximate the meaning of the base by mitigating the prototypical values of the process. This can be reinforced contextually by the use of *un po’* ‘a little’ already in the fictive motion example in (9) as well as in (10), where the paucal quantifier + verb base replicates the DEV used by the speaker immediately before. In (11), a similar effect is achieved by using the base verb *zoppicare* ‘to limp’ as contrast device (see Corona & Russo 2023: 108).

4.2.4 Path and path/manner bases

DEVs derived from path bases are *girellare* (164 tokens), *giracchiare* (13 tokens) and *girottare* (9 tokens) < *girare* ‘to go around’, and *seguicchiare* (3 tokens) < *seguire* ‘to follow’; *arrampicchiare* (1 token) derive from *arrimpicare* ‘to climb’, classified as path/manner verb since it describes an upward movement carried out with the assistance of hands and feet on a vertical or inclined surface (see Levin & Rappaport Hovav 2013).

The base of the most frequent DEVs of this group, *girare* (< Late Latin *gyrare* < *gyrus* ‘circle’), originally encodes a subcomponent of path defined “vector” by Talmy (2000) (see also Grinevald’s 2011 notion of “spatial orientation”): it indicates the action of making a circular movement or, by extension, a movement along a multilinear trajectory with no endpoint, with an implication of aimlessness and lack of commitment (‘I roamed the streets for hours’). (*Girare* also means ‘to turn’, as in *ho girato pagina* ‘I turned the page’, but this telic meaning does not produce any DEV.) DEVs from *girare* preserve the atelic sense of the base.¹ In 44 out of 186 occurrences, they indicate repetitiveness and disengagement, as well as pluractionality (iteration/spatial distribution) (14):

- (14) *ci consente di girellare, guardare, entrare nei negozi, fotografare, senza alcun problema*
 ‘We can stroll around, browse, pop into stores, take photos, no problem at all.’

¹ As pointed out by major Italian lexicographic sources, these DEVs are described as conveying nuances of manner (e.g. aimlessness or idleness) and pluractional meanings, both spatially distributive and iterative, which are often encoded by the same devices cross-linguistically (cf. Mattioli 2019: 52). For instance, De Mauro glosses the verb as ‘andare qua e là senza meta’ (‘to go here and there aimlessly’); Treccani as ‘andare qua e là per passatempo o per ozio’ (‘to go here and there for leisure or out of idleness’); and Tommaseo-Bellini as ‘andare qua e là oziosamente, non per grandi spazi’ (‘to wander here and there idly, within a limited area’).

In a further 44 tokens, DEVs' sense is reinforced by contextual elements, such as the spatial adverbials in (15) and reduplication in (16), which evoke an event distributed through space, or the similitive construction (cf. Haspelmath & Buchholz 1998; Treis & Vanhove 2017; for Italian, Corona 2025) in (17), which specifies the specific manner of *girellare*:

- (15) *girellando qua e là, capitò per caso in una vigna, piena di grappoli bruni e dorati*
'Wandering here and there, they happened to stumble upon a vineyard full of golden and brown bunches of grapes.'
- (16) *girellando girellando, trovò parole più adatte delle proprie*
'Wandering and wandering, they eventually found words better than their own.'
- (17) *Una gazzella dei Carabinieri girellava come uno scarafaggio insonnolito*
'A police car was slowly circling like a drowsy cockroach.'

The sense of disengagement may also result from a strategy of attenuation operating in the context, such as the selection of the vague paucal quantifier *un po'* 'a little' (cf. examples 9-10), which undergoes a quantity > quality > manner reinterpretation (see Russo 2023; Voghera 2017: 392-393). A clear case is in (18), where *un po'* contributes to describe the activity of *girellare* as carried out in a relaxed and undemanding manner:

- (18) *mi volevo regalare il viaggio dei miei sogni... così ho girellato un po' su internet e nelle agenzie viaggi e vi volgio [sic!] rendere partecipi dei risultati della mia ricerca.*
'I wanted to treat myself to the trip of my dreams... so I poked around a bit online and in some travel agencies, and now I want to share what I found with you all.'

In the remaining 98 occurrences, the space through which the entity moves is specified in context by means of locative adverbials denoting an extended but bounded area, characterised by non-linearity or discontinuity, as *tra le strade* in (19):

- (19) *è bello girellare tra le strade e vedere quanto i bielorussi siano diversi da noi*
'It's nice just going around the streets and seeing how different Belarusians are from us.'

Arrampicchiare suggests similar observations. In (20), it occurs in the gerund (which is, more in general, typically used to express manner) and is accompanied by the spatial distributive *qua e là* 'here and there', encoding non-uniformity in space. This non-standard way to perform the activity is reinforced by other contextual evaluative expressions such as *orrendo stile* 'dreadful style' and the similitive *con la solita agilità da felini pensionati* 'with the usual agility of retired felines':

- (20) *Arrampicchiando qua e là, con orrido stile e con la solita agilità da felini pensionati, arriviamo al meandrino da vedere indicatoci da Poppi*
'Clambering here and there, with dreadful style but the usual agility of retired felines, we reach the small meander to see pointed out by Poppi.'

In contrast with the DEVs based on *girare*, *seguicchiare* is never used in a literal spatial sense, but only in its figurative sense of 'to keep up with someone regularly; to follow; to support'. The suffix adds a sense of intermittence and inconsistency, which is ultimately a non-prototypical manner, as in (21):

- (21) *la marvel odierna (...) la segugiochio (fra un periodo di astinenza ed un altro) per affezione verso i personaggi ma è raro che condivida le scelte editoriali*
 ‘I still kind of follow modern Marvel (on and off) out of affection for the characters, but it’s rare that I actually agree with their editorial choices.’

A targeted Google search revealed that, even in web occurrences outside the ItWaC corpus, *segugiochiare* is used exclusively in a figurative sense, denoting an inconsistent and/or intermittent way of being a fan, a supporter or a follower (22), of following someone on social media (23), or of engaging in activities that would require regular commitment (e.g. *segugiochiare* a diet, a school, see 24-25).

- (22) *è arrivato un periodo in cui ho cominciato a segugiochiare anche qualche altro sport. Tipo il ciclismo*
 ‘A period came along when I started kind of following a few other sports as well, like cycling’
- (23) *vedo che anche se c’è qualcuno che mi segugiochia il profilo nessuno recensisce...*
 ‘I see that even though some people sort of follow me, nobody leaves any reviews...’
- (24) *sta settimana sto segugiochiando la dieta senza contare le calorie*
 ‘This week I’m kind of half-heartedly following the diet without counting calories’
- (25) *Al liceo prendevo voti decenti limitandomi a segugiochiare le lezioni*
 ‘In high school I got decent grades just by kind of half-paying attention in class’

5. Conclusions

This work focused on the interaction between evaluation and manner expressed in evaluative verbs derived from motion bases. In particular, we sought to answer the following research questions:

1. How and through which properties is the notion of manner specified in the case of motion DEVs, and how does it interact with evaluation?
2. What role does the context in which the DEV occurs play in determining the specific evaluative nuance?

The data were extracted from the ItWac corpus and then analysed on the basis of WordNet lexical classes. The quantitative analysis showed a predominance of manner of motion bases, which is fully consistent with what has been already observed in the literature on the topic, that is, DEVs strongly prefer activity verbs, which encode atelic, dynamic, and durative events and exhibit unergative morphosyntax (Bertinetto 2004; Grandi 2008; Stosic 2019; Tovenà 2011, 2015).

The qualitative analysis of the contexts in which DEVs occur (reinforced by some targeted search on the Web) has made it possible to refine the overall picture. In particular, it was found that:

- (i) DEVs derived from low manner bases express, like their bases, atelic and agentive events; the suffix helps to specify the manner as non-prototypical motion realised with little agentive effort and/or in a non-homogeneous space (*corricchiare di qua e*

- di là* ‘to run around’). These meanings are typically supported by contextual elements (adverbs, prepositional phrases, etc.).
- (ii) Only in cases where the manner is highly specific, as is the case with high manner verbs (*si pedalicchia un po’* ‘one pedals a bit’), do the suffix and context suggest an attenuative interpretation of the properties of the base.
 - (iii) However, the suffix does not merely reiterate the manner value of the base, but itself conveys a meaning of manner, as demonstrated by the fact that even bases expressing path give rise to manner DEVs (*girellare qua e là* ‘wandering here and there’); in this case too, we are dealing with agentive and typically atelic situations, in which a certain event is realised with little agentive effort and/or is connoted in a pluractional sense.
 - (iv) Atelicity represents a strong restriction both in the selection of bases and in the output of the evaluative process, as widely observed in the literature; however, it is not a total constraint: motion DEVs in telic constructions are rare and marginal, but not entirely absent, although telicity is entirely entrusted to other context elements (*saltelliamo giù* ‘we jump down’).
 - (v) The only real restriction in manner reading is agentivity, which is therefore a defining characteristic of the concept. In fact, where there is a decrease in agentivity, as is the case with some high manner verbs (e.g. *zoppicchiare* ‘to limp a little’), a manner reading seems to be inhibited.

The alternation of manner and attenuative meanings in motion DEVs occurs entirely within the subjective domain of evaluation. In other words, we are dealing with a manner evaluation that operates on the semantic characteristics of the base and manifests itself in exclusively agentive contexts, indicating low commitment and/or non-uniform movement on the part of the agent. Manner evaluation and attenuation are specifications of approximation operations, through which the speaker grades the distance of the referent from the centre of the category in which it is classified.

References

- Amiot, Dany & Dejan Stosic. 2011. Sautiller, voleter, dansoter: évaluation, pluriactionnalité, aspect. In Eugénia Arjoca-Ieremia, Cécile Avezard-Roger, Jan Goes, Estelle Moline et Adina Tihu (Eds.), *Temps, aspect et classes de mots: études théoriques et didactiques*, 277–297. Arras: Artois Presses Université.
- Artusi, Andrea. 2019. Appunti contrastivi sui verbi valutativi suffissati italiani e spagnoli: Analisi di due corpora sincronici. *Cuadernos de Filología Italiana* 26. 11–28.
- Bertinetto, Pier Marco. 2004. Verbi deverbali. In M. Grossmann & Franz Rainer (Eds.), *La formazione delle parole in italiano*, 465–472. Tübingen: Niemeyer.
- Brucalé, Luisa & Egle Mocciaro. 2023. Approximation through suffixation: -ddu/-a in Sicilian. *Journal of Word Formation* 7(1). 27–51.
- Brucalé, Luisa, Luisa Corona & Egle Mocciaro. Forthcoming. Where evaluation meets manner: The case of Italian deverbal evaluative verbs. *Lingue e Linguaggio* 25(1).
- Cennamo, Michela & Alessandro Lenci. 2019. Gradience in Subcategorization? Locative Phrases with Italian Verbs of Motion. *Studia Linguistica* 73(2). 369–397.
- Corona, Luisa. 2025. Confini sfumati tra nozioni contigue: Le costruzioni similative e l’espressione della maniera in italiano. In Massimo Cerruti, Cristina Onesti, Riccardo Regis & Davide Ricca (Eds.), *Continuo e discreto nelle scienze del linguaggio*, 193–218. Milano: Officinaventuno.
- Corona, Luisa & Paola Pietrandrea. 2021. In a manner of speaking: The expression of manner in spoken Italian dialogues. In Caterina Mauri, Eugenio Gorla & Ilaria Fiorentini (Eds.), *Building categories in interaction: Linguistic resources at work*, 415–438. Amsterdam: Benjamins.

- Corona, Luisa & Gina Russo, 2023. Italian deverbal verbs in *-Vcchiare*: A manner of approximating. *Lingue e Linguaggio* 22(1). 89-116.
- Cusic, David D. 1981. *Verbal plurality and aspect*. Stanford: Stanford University dissertation.
- De Mauro, Tullio (Ed.). s.d. *Dizionario della lingua italiana*. Nuovo De Mauro. <https://dizionario.internazionale.it>.
- Dressler, Wolfgang U. & Lavinia Merlini Barbaresi. 1994. *Morphopragmatics: Diminutives and intensifiers in Italian, German, and other languages*. Berlin: De Gruyter.
- Fellbaum, Christiane. 1990. English verbs as a semantic net. *International Journal of Lexicography* 3(4). 278–301.
- Grandi, Nicola. 2002. *Morfologie in contatto: le costruzioni valutative nelle lingue del Mediterraneo*. Milano: FrancoAngeli.
- Grandi, Nicola. 2008. *I verbi deverbali suffissati in italiano*. Cesena & Roma: Caissa.
- Grandi, Nicola. 2009. Restrictions on Italian verbal evaluative suffixes: The role of aspect and actionality. *York Papers in Linguistics Series* 2 10. 46–66.
- Grandi, Nicola. 2015. Word-formation and lexical aspect: Deverbal verbs in Italian. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (Eds.), *Word-formation: An international handbook of the languages of Europe*, vol. 2, 1467–1482. Berlin: De Gruyter.
- Grandi, Nicola. 2017. I diminutivi come marche di attenuazione e indeterminatezza. In Oana-Dana Balaş, Adriana Ciama, Mihai Enăchescu, Anamaria Gebăilă, Roxana Voicu (Eds.), *L'expression de l'imprécision dans les langues romanes*, 139–152. Bucureşti: Ars Docendi.
- Grandi, Nicola & Livia Körtvélyessy (Eds.). 2015. *Edinburgh handbook of evaluative morphology*. Edinburgh: Edinburgh University Press.
- Grinevald, Colette. 2011. On constructing a working typology of the expression of path. *Faits de Langues. Les Cahiers* 3. 43-20.
- Haspelmath, Martin & Oda Buchholz. 1998. Equative and similative constructions in the languages of Europe. In Johan van der Auwera (Ed.), *Adverbial constructions in the languages of Europe*, 277-334. Berlin: de Gruyter.
- Iacobini, Claudio & Francesca Masini. 2006. The emergence of verb-particle constructions in Italian: locative and actional meanings. *Morphology* 16.155–188.
- Jackendoff, Ray. 1983. *Semantics and cognition*. Cambridge, MA: MIT Press.
- Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.
- Levin, Beth & Malka Rappaport Hovav. 2013. Lexicalized meaning and manner/result complementarity. In Boban Arsenijević, Berit Gehrke & Rafael Marín (Eds.), *Studies in the Composition and Decomposition of Event Predicates*. 49–70. Dordrecht: Springer.
- Masini, Francesca, Muriel Norde & Kristel Van Goethem (Eds.). 2023. Approximation in morphology. *Journal of Word Formation* 7(1).
- Mattiola, Simone. 2019. *Typology of pluractional constructions in the languages of the world*. Amsterdam: Benjamins.
- Minoccheri, Chiara & Dejan Stosic. 2022. La manière dans tous ses états: Une étude exploratoire sur corpus. *8 Congrès Mondial de Linguistique Française (CMLF 2022)*. SHS Web of Conferences 138, 11008.
- Moline, Estelle & Dejan Stosic. 2016. *L'expression de la manière en français*. Paris: Éditions Ophrys.
- Russo, Gina. 2023. A refinement of the definition of Manner and its expression through Italian indefinites. *Studi e Saggi Linguistici* 61(2). 145-191.
- Slobin, Dan I. 2004. The many ways to search for a frog: Linguistic typology and the expression of motion events. In Sven Strömquist & Ludo Verhoeven (Eds.), *Relating events in narrative*, vol. 2, 219-257. Mahwah, NJ: Lawrence Erlbaum Associates.
- Slobin, Dan I. 2005. Relating narrative events in translation. In Dorit Ravid & Hava Bat-Zeev Shyldkrot (Eds.), *Perspectives on Language and Language Development: Essays in Honor of Ruth A. Berman*, 115-129. Dordrecht: Kluwer.
- Slobin, Dan I. 2006. What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In Maya Hickmann & Stéphane Robert (Eds.), *Space in languages: Linguistic systems and cognitive categories*, 59-81. Amsterdam: Benjamins.

- Stosic, Dejan. 2013. Manner of motion, evaluative and pluractional morphology. *Oslo Studies in Language* 5(1). 61-89.
- Stosic, Dejan. 2019. Manner as a cluster concept: What does lexical coding of manner of motion tell us about manner? In Michel Aurnague & Dejan Stosic (Eds.), *The Semantics of Dynamic Space in French. Descriptive, Experimental and Formal Studies on Motion Expression*, 141-177. Amsterdam: John Benjamins.
- Stosic, Dejan. 2020. Defining the concept of manner: An attempt to order chaos. *Testi e linguaggi* 14. 127-150.
- Stosic, Dejan & Dany Amiot. 2011. Quand la morphologie fait des manières: les verbes évaluatifs et l'expression de la manière en français. In Dany Amiot, Walter de Mulder, Estelle Moline & Dejan Stosic (Eds.), *Ars grammatica. Hommage à Nelly Flaux*, 403-430. Bern: Peter Lang.
- Stosic, Dejan & Dany Amiot. 2019. Motion verbs and evaluative morphology. In Michel Aurnague & Dejan Stosic (Eds.), *The semantics of dynamic space in French: Descriptive, experimental and formal studies on motion expression*, 179-215. Amsterdam: Benjamins.
- Talmy, Leonard. 2000. *Toward a cognitive semantics*, vol. 2: *Typology and process in concept structuring*. Cambridge, MA: MIT Press.
- Tovena, Lucia M. 2011. When small is many in the event domain. *Lexis: E-journal in English Lexicology* 6. 41-58.
- Tovena, Lucia M. 2015. Evaluative morphology and aspect/actionality. In Nicola Grandi & Livia Körtevélyessy (Eds.), *Edinburgh handbook of evaluative morphology*, 108-120. Edinburgh: Edinburgh University Press.
- Tommaseo, Nicolò & Bernardo Bellini. 1865-1879. *Dizionario della lingua italiana nuovamente compilato*. Torino, Società Unione Tipografica-Editrice.
- Treccani = Istituto della Enciclopedia Italiana (s.d.). *Vocabolario Treccani online*. Roma: Istituto della Enciclopedia Italiana. <https://www.treccani.it/vocabolario/>
- Treis, Yvonne & Martine Vanhove (Eds.). 2017. *Similative and equative constructions: A crosslinguistic perspective*. Amsterdam: Benjamins
- Verkerk, Annemarie. 2014. *The evolutionary dynamics of motion event encoding*. PhD Thesis. Radboud University Nijmegen: Nijmegen.
- Voghera, Miriam. 2017. Quando vaghezza e focus entrano in contatto: Il caso di *un attimo*, anzi *un attimino*. In Roberta D'Alessandro, Gabriele Iannàccaro, Diana Passino & Anna M. Thornton (Eds.), *Di tutti i colori. Studi linguistici per Maria Grossmann*, 385-398. Utrecht: Utrecht University Repository.

Investigating the productivity of evaluative morphology: Diminutives of loanwords in Italian

M. Silvia Micheli
Università degli Studi di Milano Statale
maria.micheli@unimi.it

Matteo Pellegrini
University of Surrey
matteo.pellegrini@surrey.ac.uk

1. Introduction¹

The field of evaluative morphology has recently attracted renewed attention in studies on Italian morphology, as demonstrated by the growing body of research on approximation and intensification (see, e.g., Masini & Micheli 2020, Micheli 2023, Corona & Russo 2023, Masini & Di Donato 2023 on approximative strategies; Lacić 2025 and Micheli & Verdelli 2025 on intensifiers). Although current scholarly interest appears to have shifted toward what Grandi (2002, 2017) defines as *qualitative* evaluation, the *quantitative* evaluative domains examined in the seminal works by Dressler & Merlini Barbaresi (1992, 1994), namely diminution and augmentation, remain far from fully explored.

Focusing on diminution, Italian diminutive suffixes have traditionally drawn the interest of scholars because of their abundance and the wide range of their semantic and pragmatic functions (see, among others, Rainer 1990; Grandi 1998, 2017; Cinque 2015). On the other hand, research on the form of diminutives has been more limited, with a primary focus on the so-called *antesuffixal interfixes* (e.g., *-ic-* in *pianticina*, plant-INTERFIX-DIM; see Dressler & Merlini Barbaresi 1986, 1989; Merlini Barbaresi 2012). More recently, Micheli & Pellegrini (2025) – henceforth, M&P – have sought to address this gap by proposing a framework for the formal properties of Italian diminutive suffixes, drawing on the theoretical tools of Construction Morphology – henceforth, CxM (Booij 2010). In this paper, we aim to build on our previous work by examining diminutives of loanwords (e.g., It. *t-shirt* > *t-shirtina* ‘small t-shirt’). This perspective enables us to address the dimension of productivity, which had remained unexamined in our previous work.

The paper is structured as follows. After reviewing our previous work on the form of Italian diminutives (Section 2), we highlight that the issue of the productivity of different processes is left open in that work (Section 3). We then describe how we obtained our dataset, starting from collections of loanwords and searching for corpus attestations of their diminutive forms (Section 4), and analyse this dataset according to the same parameters used in our previous work (Section 5). Section 6 concludes and highlights possibilities for future work.

¹ The whole paper is the result of the close collaboration of both authors; however, for academic purposes, M. Silvia Micheli is responsible for Sections 1 and 3, Matteo Pellegrini for Sections 4 and 6. Both authors are responsible for Sections 2 and 5.

2. Alternation patterns and constructional schemas in Italian diminutives

Focusing on formal aspects, it can be observed that several factors are at play in Italian diminutive formation. There are several diminutive-forming suffixes; the most common are *-in-*, *-ett-* and *-ell-*, but many others are found in a non-negligible amount of cases. In most cases, these suffixes are transparent to the gender of the base they apply to. We thus have *abitinom*² from *abitom* ‘dress’ vs. *copertina_F* from *coperta_F* ‘blanket’. However, it is not uncommon to find masculine diminutives formed from feminine bases (e.g. *cordino_M* from *corda_M* ‘rope’), and – although more rarely – the reverse (e.g. *fiaschetta_F* from *fiasco_M* ‘flask’). There is a relationship between gender and the final segments of bases and diminutives. While there is an exceptionless implication for diminutives, which always end in *-o* if they are masculine and in *-a* if they are feminine, the situation is more complex for bases. This is due on the one hand to exceptional cases of feminine nouns ending in *-o* (e.g. *mano_F* ‘hand’, with diminutive *manina_F*) and masculine nouns ending in *-a* (e.g. *fantasma_M* ‘ghost’, with diminutive *fantasmino_M*); on the other hand, to nouns with other endings that can be either masculine or feminine (cf. nouns in *-e*, such as *dente_M* ‘tooth’ vs. *chiave_F* ‘key’). Furthermore, suffixes can be combined with one of a set of what Dressler & Merlini Barbaresi (1986) called “interfixes”, such as *-ic-* (e.g., *oss-ic-ino* from *osso* ‘bone’, *ort-ic-ello* from *orto* ‘garden’). Lastly, several diminutive suffixes are occasionally used cumulatively, as in *ros-ell-ina* from *rosa* ‘rose’. The examples just given show that all the factors at play – suffixes, interfixes, gender and final segment of the base and diminutive – are, at least partly, independent of each other: different combinations of suffixes and interfixes are possible, and both gender-preserving and gender-switching diminutives are found with all suffixes and interfixes, together with the related variation in the alternation of final segments.

To account for all these factors and the non-trivial relationship between them, M&P used tools originally conceived for the analysis of patterns of form alternation in inflection, in the tradition of word-based abstractive theories (cf. Ackerman et al. 2009; Bonami & Boyé 2014; Blevins 2016, and subsequent work), and provided a formalisation of these different patterns grounded in the framework of CxM (cf. Booij 2010 and many subsequent studies). CxM considers both individual words and the word-formation mechanisms instantiated through constructions (or schemas), enabling the identification of regularities and sub-regularities in morphologically complex words. This makes it particularly well-suited for capturing generalisations on the surface alternation patterns found between bases and diminutives. At the same time, applying CxM to Italian diminutives allows for the refinement of its analytical tools, which have traditionally focused on capturing semantic rather than formal regularities.

Based on a sample of about 1,300 base-diminutive pairs taken from the Zingarelli dictionary (2022, online version), M&P extracted surface alternation patterns using the Qumin toolkit (Beniamine 2018). These patterns are obtained by matching the form of the base and that of the diminutive, singling out the segments that are not the same in the two forms, as shown in the fourth column of Table 1. These alternation patterns serve as a proxy for paradigmatic relations

² Since the semantic effect of the diminutive formation process in Italian includes a broad range of values (e.g. small size, young age, low intensity/importance, affection, etc.), being also influenced by the context, we do not provide a gloss of the meaning of diminutives in this paper. In all cases, the gloss would simply be ‘<base meaning>.DIM’, e.g. in this case ‘dress.DIM’.

that hold between different constructional schemas, shown in the fifth column, in the notation of CxM.³

Table 1. Alternation patterns between some Italian base-diminutive pairs (from M&P)

Base meaning	Base	Diminutive	Alternation pattern	Paradigmatic relation (CxM notation)
‘movie’	/film/	/filmino/	$\Rightarrow$ ino	$\langle [x]_{\text{MORPH/SYN}_i} \leftrightarrow [\text{SEM}_i] \approx \langle [[x]_{\text{MORPH/SYN}_i}/\text{ino}/\text{MORPH/SYN}_j] \leftrightarrow [\text{DIM SEM}_i]_j \rangle$
‘van’	/furgone/	/furgon ^f ino/	$e \Rightarrow$ ^f ino	$\langle [x]/e/\text{MORPH/SYN}_i \leftrightarrow [\text{SEM}_i] \approx \langle [[x]_{\text{MORPH/SYN}_i}/\text{f}^f\text{ino}/\text{MORPH/SYN}_j] \leftrightarrow [\text{DIM SEM}_i]_j \rangle$
‘eye’	/okkjo/	/okkjolino/	$\Rightarrow$ lino	$\langle [x]_{\text{MORPH/SYN}_i} \leftrightarrow [\text{SEM}_i] \approx \langle [[x]_{\text{MORPH/SYN}_i}/\text{lino}/\text{MORPH/SYN}_j] \leftrightarrow [\text{DIM SEM}_i]_j \rangle$
‘cabbage’	/kavolo/	/kavolino/	$o \Rightarrow$ ino	$\langle [x]/o/\text{MORPH/SYN}_i \leftrightarrow [\text{SEM}_i] \approx \langle [[x]_{\text{MORPH/SYN}_i}/\text{ino}/\text{MORPH/SYN}_j] \leftrightarrow [\text{DIM SEM}_i]_j \rangle$

Such surface constructions were then organised into separate hierarchies, including sub-schemas at different levels of generality, to be able to express generalisations on each of the different, mutually independent factors described above. For instance, the hierarchy shown in Figure 1 accounts for the patterns of gender alternation found between bases and diminutives. In that hierarchy, we have the most specific constructions at the bottom, and progressively more abstract groupings going up the tree, e.g., b. for gender-preserving diminutives, and c. for gender-switching diminutives. A similar mechanism operates in the hierarchies drawn for other factors (final segment alternation, suffixes, and what we call “antesuffixal phonological material”, which covers both cumulated suffixes and interfixes). For reasons of space, we do not describe in detail these other hierarchies: the interested reader is referred to M&P.

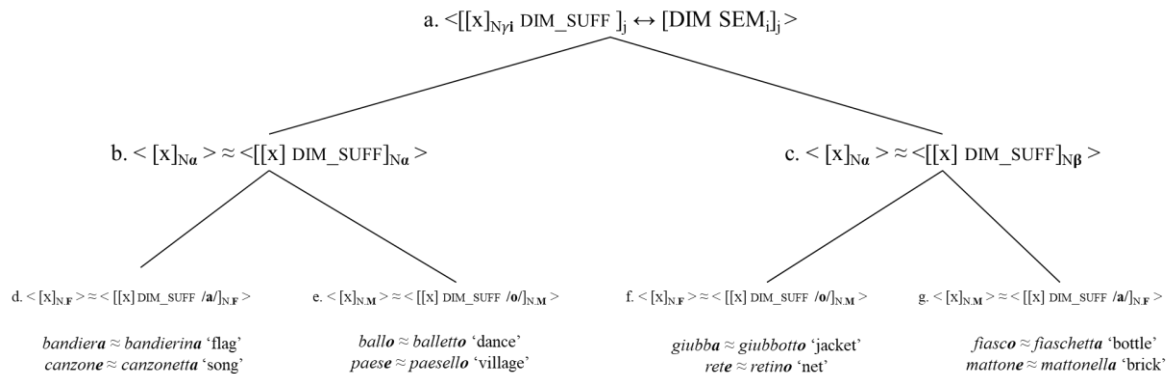


Figure 1. The gender hierarchy (from M&P)

The mechanism of multiple inheritance was then exploited to account for all the factors at play: surface alternation patterns inherit from the most specific schema available in each of the

³ For the sake of illustration, we adopt a simplified version of the notation commonly used in the CxM literature (Masini & Audring 2019: 368–369).

hierarchies. For instance, the pattern “ $e \rightleftharpoons ikolo$ ”, relating the base *rete* ‘net’ to the diminutive *reticolo*, inherits from sub-schema d. in the antesuffixal phonological material hierarchy, e. in the suffix hierarchy, f. in the final segment hierarchy, g. in the gender hierarchy, h. from the lexical category hierarchy, as shown in Figure 2.

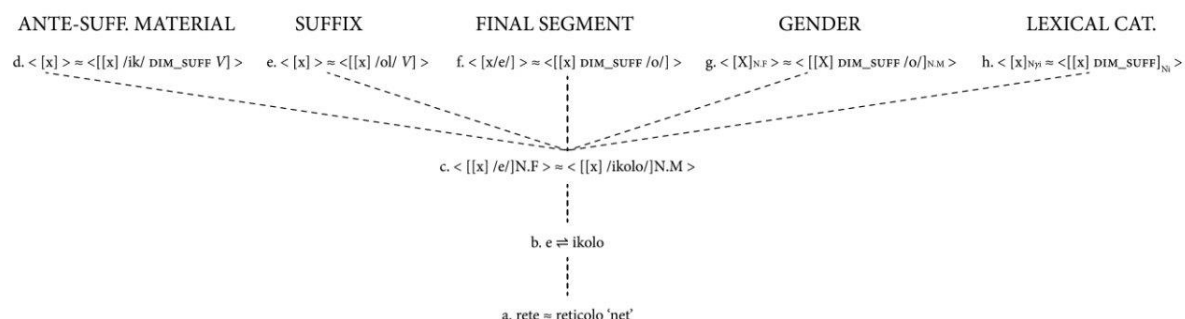


Figure 2. Exploiting the mechanism of multiple hierarchy to account for our data: the pair *rete-reticolo* (from M&P)

3. Open issues

As already noted, an advantage of integrating automatically extracted surface alternation patterns with CxM is that this approach offers both a theoretical representation and a quantitative account of the (sub-)schemas identified in the data. For example, the frequency of the sub-schemas corresponding to diminutive suffixes was calculated, thereby gaining an overview of which suffixes are most frequent in our dataset. The following table presents the most frequent suffixes in our database, specifically those with a type frequency > 10 .

Table 2. Sub-schemas corresponding to the most frequent suffixes within the database collected in M&P (p. 193)

Sub-schema	Example pair	Freq.	%
$\langle [x] \rangle \approx \langle [[x] /in/ V] \rangle$	<i>cappotto</i> > <i>cappottino</i> ‘coat’	573	43.54%
$\langle [x] \rangle \approx \langle [[x] /ett/ V] \rangle$	<i>bosco</i> > <i>boschetto</i> ‘wood’	417	31.69%
$\langle [x] \rangle \approx \langle [[x] /ell/ V] \rangle$	<i>campana</i> > <i>campanella</i> ‘bell’	194	14.74%
$\langle [x] \rangle \approx \langle [[x] /ol/ V] \rangle$	<i>goccia</i> > <i>gocciola</i> ‘drop’	55	4.18%
$\langle [x] \rangle \approx \langle [[x] /uttf/ V] \rangle$	<i>bocca</i> > <i>boccuccia</i> ‘mouth’	30	2.28%
$\langle [x] \rangle \approx \langle [[x] /ott/ V] \rangle$	<i>barile</i> > <i>barilotto</i> ‘barrel’	24	1.82%

However, because the study relied solely on type-frequency data for diminutives extracted from a dictionary (therefore encompassing a largely lexicalised set of forms), it offered only limited insight into the productivity of the processes under investigation. On the other hand, applying corpus-based quantitative measures (such as those proposed by Baayen 1992 and subsequent studies) to Italian diminutives presents several methodological challenges.

The most productive suffix, *-in-*, occurs not only in diminutives but also in the creation of a wide range of other derivatives. Firstly, it is used in the formation of (both denominal and deverbal) agent (see Rainer 2024) and instrument nouns. As a noun-forming suffix, *-in-* is also employed in the creation of feminine forms and, albeit rarely, in the creation of event nouns. As an adjectival suffix, *-in-* creates relational adjectives and ethnic adjectives, as well as adjectives indicating membership of a group. Moreover, it is also used with an attenuative

value, especially in combination with color terms. Finally, the segment *-in-* can also be found in simple nouns and inflected verb forms. Given that *-in-* is not the only diminutive suffix employed in non-evaluative derivational strategies (see, e.g., *-ett-*), it is evident that performing an analysis based on corpus data obtained through wildcard searches presents considerable difficulties. Indeed, evaluative affixation was not included among the processes examined in extensive corpus-based analyses of morphological productivity in Italian (e.g., Gaeta & Ricca 2003, 2006).

In this study, we therefore adopt a different approach, examining the productivity of diminutive affixes by looking at when they are applied to new entries in the lexicon, focusing on loanwords. Indeed, Dressler (2003: 37, 2007: 163–165; see also Gardani 2013: 34–38) proposes loanword integration as a diagnostic for determining the degree of productivity of a morphological mechanism. Following Wurzel (1984), Dressler (2003) argues that the integration of a loanword exhibiting unfitting properties is the strongest indicator of the highest productivity. As will be discussed below, the attachment of a diminutive suffix to an Italian base may fall within this category, since foreign bases typically exhibit unfitting properties (most notably their consonantal endings) and may therefore require phonological adjustments. Adopting this perspective, we compiled a dataset of English loanwords attested in Italian, and we searched diminutives formed from them in corpora, following the procedure discussed in the following Section.

4. Data

As a point of departure for our procedure to gather data for this study, we searched for potential bases of interest – i.e., loanwords – in two lexical sources. We extracted, on the one hand, 2,097 entries that are tagged as English loanwords in the Zingarelli dictionary (2024, online version); on the other hand, 3,998 neologisms were manually identified as non-Italian from the repertoire collected by the Treccani online portal.⁴ Together, these two data sources allow us to cover loanwords with different characteristics. The former is a reference dictionary, and thus mostly provides us with loanwords that are by now well established in the Italian lexicon (such as *film*⁵). The latter is a repertoire of neologisms, which consequently provides us with very recent and less entrenched loanwords (such as *foodblogger*).

We then automatically generated diminutives by applying to such bases the processes that were found to be most frequent in M&P, namely, the suffixes *-in-*, *-ett-*, *-ell-*, possibly preceded by the interfixes *-ic-*, *-ar/-er-*, or the cumulated suffix *-ol-*, both in masculine and feminine forms, to account for both gender preserving and gender switching diminutives. For instance, starting from *film*_{N,M}, we generated the diminutives *filmin*_M and *filmin*_F, as well as *filmetto/-a*, *filmello/-a*, *filmicino/-a*, *filmicetto/-a*, *filmicello/-a*, *filmarino/-a*, *filmaretto/-a*, *filmarello/-a*, *filmerino/-a*, *filmeretto/-a*, *filmerello/-a*, *filmolino/-a*, *filmoletto/-a*, *filmolello/-a*. We then checked whether such diminutives are found in a large web-crawled corpus of Italian, itTenTen20, by searching not only singular but also plural forms (e.g., *filmini*, *filmetti*, *filmelli*, etc.). After a first automatic check based on the generated strings, we manually excluded false positives. These are sometimes cases of forms that are actually created by one of the other

⁴ Specifically, the data come from the webpage *Neologismi della settimana* ('Loanwords of the week'), a repository compiled by the editors of the Treccani portal, also incorporating user submissions. The section is accessible at: https://www.treccani.it/magazine/lingua_italiana/neologismi/ (last accessed 19 December 2025)

⁵ The meaning of Italian loanwords is omitted because it is the same as in English. We only provide a gloss when there is a difference.

morphological processes listed in the previous section that also yield forms in *-ino* (e.g. *testino* is found in itTenTen20, but it is not the diminutive of the loanword *test*, but the third person plural subjunctive of the verb derived from it), or simple nouns or – more commonly – proper nouns that happen to display the same final segments as one of the diminutive suffixes or interfixes that we focus on (e.g. *Vannino*, an Italian name that happens to match the form of the diminutive of the loanword *van*).

An interesting case is that of names of objects, brands, websites, projects, etc., which are formed exploiting the morphological processes of diminutive formation. For instance, starting from the loanword *job*, by adding the suffix *-in-* preceded by the interfix *-er-*, *Jobberino* was formed as the name of an online platform for online work. Similarly, starting from the loanword *snack*, by applying the same processes, but with the *-a* final segment normally found with feminine diminutives, *Snackerina* was formed as the name of a brand of food-inspired stickers. While these examples offer further qualitative evidence of the productivity of diminutive formation processes, including interfixes, we decided not to include them in our dataset, due to their apparent status as conscious, metalinguistic coinages.

The outcome of our procedure is a dataset of 644 diminutives for 413 bases. Among those bases, 325 come from the Zingarelli dictionary, 37 from the Treccani neologisms, and 51 are attested in both sources. To be able to extract alternation patterns between the base and diminutive using Qumin, thus replicating the procedure followed in M&P, we obtained an IPA transcription of all the bases and diminutives in our sample, using the online tool IPAfonetica (Pedroncelli 2023), and manually correcting occasional mistakes. This is the dataset on which the results shown in the next section are based.

5. Results

5.1 Surface alternation patterns

The analysis of the surface alternation patterns generated using the Qumin toolkit enables us to determine whether loanword diminutives are formed through a wide range of formal strategies or rather through a more limited set of processes. Based on our dataset, 43 patterns were found. Notably, 12 of these had already been identified by M&P (p. 186), as presented in Table 3.

Table 3. Patterns attested in both our dataset and M&P

Pattern	Example	N. pairs
$\Rightarrow$ ino	<i>browser</i> > <i>browserino</i>	251
$\Rightarrow$:ino	<i>bug</i> > <i>bugghino</i>	84
o $\Rightarrow$ ino	<i>gazebo</i> > <i>gazebino</i>	5
o $\Rightarrow$ et:o	<i>pogo</i> > <i>poghetto</i>	3
e $\Rightarrow$ ino	<i>meme</i> > <i>memino</i>	2
$\Rightarrow$ na	<i>lobby</i> > <i>lobbina</i>	2
e $\Rightarrow$ iʃino	<i>monstre</i> > <i>monstricino</i>	1
a $\Rightarrow$ ino	<i>koala</i> > <i>koalino</i>	1
a $\Rightarrow$ ina	<i>koala</i> > <i>koalina</i>	1
a $\Rightarrow$ et:a	<i>koala</i> > <i>koaletta</i>	1
o $\Rightarrow$ el:o	<i>tornado</i> > <i>tornadello</i>	1
o $\Rightarrow$ olino	<i>shampoo</i> > <i>shampoolino</i>	1
Total: 12		353

The quantitative data indicate that the most frequent pattern involves the addition of the suffix *-ino* to a base ending in a consonant (e.g., *computer* > *computerino*), while patterns attested only once or twice involve bases ending in a vowel, which English in turn has borrowed from other languages (e.g., *tornado*). As will be discussed below, when the base ends in a consonant, it may also undergo consonant lengthening, as illustrated by the second most frequent pattern (e.g., *bug* > *bugghino*).

More interestingly, we identified as many as 31 patterns not attested in the previous study, thereby substantially enriching the picture outlined by M&P, as illustrated in Table 4.

Table 4. Patterns identified in our dataset (not attested in M&P)

Pattern	Example	n. pairs
⇒ ina	<i>box</i> > <i>boxina</i>	71
⇒ et:o	<i>blues</i> > <i>bluesetto</i>	64
⇒ :et:o	<i>flop</i> > <i>floppetto</i>	28
⇒ :ina	<i>chat</i> > <i>chattina</i>	26
⇒ el:o	<i>paper</i> > <i>paperello</i>	21
⇒ et:a	<i>cover</i> > <i>coveretta</i>	20
⇒ arel:o	<i>film</i> > <i>filmarello</i>	9
⇒ erel:o	<i>link</i> > <i>linkerello</i>	8
⇒ :et:a	<i>snob</i> > <i>snobbetta</i>	7
⇒ iŋino	<i>prompt</i> > <i>prompticino</i>	3
⇒ el:a	<i>sneaker</i> > <i>sneakerella</i>	3
⇒ :iŋino	<i>spot</i> > <i>spotticino</i>	3
...		2-1
Total: 31		290

As expected, all newly identified patterns involve bases ending in a consonant, a typical feature of borrowings from English. Interestingly, the most common pattern involves a base ending in a consonant and a diminutive formed with the suffix *-in-* followed by the final vowel *-a*, which typically marks feminine gender. One aspect that has already been noted concerns the lengthening of the final consonant in monosyllabic bases. This phenomenon is also observed in English loanwords when a verbal inflectional suffix is added to a base, such as in *chat* > *chattare* (as well as in the form attested in our dataset *chat* > *chattina*). M&P argued that the CxM mechanism of multiple inheritance can also be exploited to account for wider phonological generalizations potentially applicable also in contexts other than diminutives. Accordingly, both types of words can be considered to inherit from the same schema that accounts for this regularity, namely <[[x] VC#]> ≈ <[[x] VC: [suff]]>.

5.2 Final segments

Table 5 presents the patterns identified with respect to the final sound, which in the diminutive form corresponds to the inflectional ending. As shown by our data, the most frequent patterns involve bases ending in a consonant combined with diminutive forms ending in *-o* or *-a*, which typically encode masculine and feminine gender in Italian, respectively. In addition, a number of patterns with lower incidence were observed, some of which are unattested in the M&P sample. Among these, we found a pattern in which a base ending in *-i* is paired with a diminutive form ending in *-o* (e.g. *kiwi* > *kiwino*).

Table 5. Final segments of the diminutives in our sample and in M&P

Sub-schema	Example	Pairs	Pairs in M&P	Pairs in our sample
$\langle [xC] \rangle \approx \langle [[x]C \text{ DIM_SUFF } /o/] \rangle$	<i>bug</i> > <i>bugghino</i>	489 (25%)	8 (0.6%)	481 (74.8%)
$\langle [x/o] \rangle \approx \langle [[x] \text{ DIM_SUFF } /o/] \rangle$	<i>gazebo</i> > <i>gazebino</i>	481 (24.6%)	471 (35.8%)	10 (1.6%)
$\langle [x/a] \rangle \approx \langle [[x] \text{ DIM_SUFF } /a/] \rangle$	<i>koala</i> > <i>koalina</i>	439 (22.4%)	436 (33.2%)	2 (0.3%)
$\langle [x/a] \rangle \approx \langle [[x] \text{ DIM_SUFF } /o/] \rangle$	<i>koala</i> > <i>koalino</i>	165 (8.4%)	165 (12.5%)	1 (0.2%)
$\langle [x/e] \rangle \approx \langle [[x] \text{ DIM_SUFF } /o/] \rangle$	<i>meme</i> > <i>memino</i>	164 (8.4%)	161 (12.2%)	3 (0.5%)
$\langle [xC] \rangle \approx \langle [[x]C \text{ DIM_SUFF } /a/] \rangle$	<i>chat</i> > <i>chattina</i>	136 (6.9%)	-	136 (21.1%)
$\langle [x/e] \rangle \approx \langle [[x] \text{ DIM_SUFF } /a/] \rangle$	<i>legge</i> > <i>leggina</i> 'law'	52 (2.7%)	52 (4.0%)	-
$\langle [x/o] \rangle \approx \langle [[x] \text{ DIM_SUFF } /a/] \rangle$	<i>mano</i> > <i>manina</i> 'hand'	21 (1.1%)	21 (1.6%)	-
$\langle [x/i] \rangle \approx \langle [[x] \text{ DIM_SUFF } /o/] \rangle$	<i>kiwi</i> > <i>kiwino</i>	6 (0.3%)	-	6 (0.9%)
$\langle [x/i] \rangle \approx \langle [[x] \text{ DIM_SUFF } /a/] \rangle$	<i>wiki</i> > <i>wikina</i>	5 (0.3%)	1 (0.1%)	4 (0.6%)

5.3 Gender

The phenomenon of gender switch induced by diminutive suffixation is well-documented in the literature and has been further corroborated by M&P (pp. 189–190). Nevertheless, the productivity of this mechanism remains largely unexplored. While limited in scope, our data offer preliminary insights into this process. Table 6 illustrates that, alongside standard gender-preserving patterns, a pattern emerges where feminine bases are paired with masculine diminutives (e.g., *t-shirt_F* > *t-shirtino_M*). However, the reverse pattern (masculine to feminine) was not observed, further reflecting the lower frequency of this switch reported in the M&P sample. The data thus appear to indicate that gender switch is still productive, albeit confined to cases of feminine-to-masculine reassignment, with masculine commonly analysed as the unmarked gender in Italian. At the same time, it should be noted that our data are rather limited, and the gender attribution of loanwords in Italian is variable, as noted by Thornton (2009). For instance, we observe that the noun *slip* 'briefs underwear' (which is generally masculine in Italian) occurs in our sample in both the forms *slippino_M* and *slippina_F*, the latter possibly being based on a feminine variant (*la slip*, reflecting the influence of the Italian equivalent *mutanda_F* 'briefs underwear' – rather than being the outcome of a gender switch induced by the diminutive formation process. To err on the side of caution, in our annotation, we considered as cases of gender switch only the ones where that analysis appeared to be the only viable option. This motivates the much lower frequency of gender-switching diminutives in our sample compared to M&P, and makes it difficult to draw any more precise quantitative evaluation of the difference between the two samples.

Table 6. Gender of the diminutives in our sample and in M&P

Sub-schema	Example	Pairs	Pairs in M&P	Pairs in our sample
$\langle [x]_{N,M} \rangle \approx \langle [[x] \text{ DIM_SUFF } /o/]_{N,M} \rangle$	<i>bug</i> > <i>bugghino</i>	1,083 (55.3%)	588 (44.7%)	495 (77%)
$\langle [x]_{N,F} \rangle \approx \langle [[x] \text{ DIM_SUFF } /a/]_{N,F} \rangle$	<i>cover</i> > <i>coveretta</i>	620 (31.7%)	479 (36.4%)	141 (21.9%)
$\langle [x]_{N,F} \rangle \approx \langle [[x] \text{ DIM_SUFF } /o/]_{N,M} \rangle$	<i>chat</i> > <i>chattino</i>	176 (9%)	174 (13.2%)	2 (0.3%)
$\langle [x]_{N,M} \rangle \approx \langle [[x] \text{ DIM_SUFF } /a/]_{N,F} \rangle$	<i>mattone</i> > <i>mattonella</i> 'brick'	28 (1.4%)	28 (2.1%)	-

5.4 Suffixes

Table 7 shows the number of base-diminutive pairs in our dataset, displaying each of the three diminutive suffixes on which we focus, also providing the number of pairs recorded in M&P for those suffixes based on a sample of the full Italian lexicon.

Table 7. Base-diminutive pairs with suffixes *-in-*, *-ett-*, *-ell-* in our sample and in M&P

Sub-schema	Example	Pairs in M&P	Pairs in our sample
$\langle [x] \rangle \approx \langle [[x] /in/ V] \rangle$	<i>browser</i> > <i>browserino</i>	573 (43.6%)	465 (72.3%)
$\langle [x] \rangle \approx \langle [[x] /ett/ V] \rangle$	<i>cover</i> > <i>coveretta</i>	417 (31.7%)	128 (19.9%)
$\langle [x] \rangle \approx \langle [[x] /ell/ V] \rangle$	<i>paper</i> > <i>paperello</i>	194 (14.8%)	50 (7.8%)

As a general remark, we can observe that all three suffixes are attested with a remarkable number of non-native bases, and can thus be confirmed to be productive, as expected on the basis of qualitative observation (Merlini Barbaresi 2004: 265–266). However, a comparison of the distribution found in M&P with that of this work highlights remarkable differences that can help to gain a deeper insight into the relative productivity of suffixes. Specifically, the suffix *-in-* is much more frequent in the sample of loanwords than in the overall lexicon, while the suffixes *-ett-* and *-ell-* are less frequent in the sample of loanwords than in the overall lexicon. According to the principle described in Section 3, the preferential application to novel entries in the lexicon – such as loanwords – can be taken as an indication of productivity. Consequently, *-ett-* and *-ell-* are shown to be more weakly productive, and the difference with the most productive suffix, *-in-*, is even larger than it appears based on the distribution found in the overall lexicon. Indeed, while *-in-* was already found to be the most frequent suffix by M&P, in their data *-ell-* was also very common, being attested with a number of bases not so different from that of *-in-*. In the sample of loanwords, however, the distribution is much more unbalanced, with *-in-* covering most of the bases, and *-ett-* and *-ell-* only used in a minority of cases.

Further evidence of a stronger productivity of *-in-* can be found looking at cases in which more than one diminutive is formed from the same base, using different suffixes (e.g., *filmino*, *filmetto* and *filmarello* from *film*), and comparing their frequency to that of cases where only one suffix can be applied. The relevant data are given in Table 8.

Table 8. Suffixes that can apply to the same base in our sample

Diminutive Suffixes	Example	N. cases in our sample
-in- only	<i>mouse</i> > <i>mousino</i> ‘computer mouse’	276
-in- and -ett-	<i>server</i> > <i>serverino/serveretto</i>	71
-ett- only	<i>zombie</i> > <i>zombietto</i>	26
-in- and -ell-	<i>computer</i> > <i>computerino/computerello</i>	18
all suffixes	<i>film</i> > <i>filmino/filmetto/filmarello</i>	10
-ell- only	<i>muffin</i> > <i>muffinello</i>	10
-ett- and -ell-	<i>fiction</i> > <i>fictionetta/fictionarella</i> ‘TV series’	2

What is interesting to observe is that the suffixes *-ett-* and *-ell-* are attested more frequently with bases to which also the suffix *-in-* can be attached. This suggests that some bases are particularly prone to act as bases for the formation of diminutives, and lend themselves easily to the application of several processes, including *-ett-* and *-ell-*. However, this also means that when only one process is applied, the default status of the suffix *-in-* is even stronger than it would appear, looking at all diminutives regardless of whether they are formed from the same base.

Another aspect that deserves attention is whether there are phonological contexts that possibly favor or disfavor the application of specific suffixes. To that end, in Table 9, we group the bases of our sample according to their final segment, and for each final segment, we show how many diminutives are formed with each of the suffixes under consideration, also providing the percentage with respect to the total number of diminutives formed with the corresponding suffix.

Table 9. Final segments of the bases from which diminutives of our sample are formed

Final segment of the base	Diminutives in:		
	<i>-in-</i>	<i>-ett-</i>	<i>-ell-</i>
/r/	132 (35.2%)	16 (14.7%)	13 (32.5%)
/t/	57 (15.2%)	-	7 (17.5%)
/k/	23 (6.1%)	14 (12.8%)	5 (12.5%)
/p/	25 (6.7%)	14 (12.8%)	1 (2.5%)
/l/	24 (6.4%)	13 (6.4%)	1 (2.5%)
/d/	20 (5.3%)	-	3 (7.5%)
/m/	11 (2.9%)	7 (6.4%)	2 (5%)
/s/	10 (2.7%)	7 (6.4%)	2 (5%)
/n/	11 (2.9%)	4 (3.7%)	3 (7.5%)
/g/	10 (2.7%)	6 (5.5%)	1 (2.5%)

We find evidence of some constraints that appear to be dissimilatory in nature. For instance, bases ending in /l/ rarely display the suffix *-ell-*, which also includes the same consonant. Interestingly, the only case where that happens (*singlerello* from *single* /'singol/ ‘unmarried’) is only an apparent counterexample, as the suffix *-ell-* is preceded by the interfix *-er-*, so that there is no repetition of the same segment in adjacent syllables. Similarly, bases ending in /t/

never display the suffix *-ett-*, which also includes the same consonant. In this case, the constraint appears to extend also to bases ending in the voiced counterpart /d/.

Furthermore, differences between suffixes regarding the distribution of final segments of their bases suggest that other kinds of phonological conditioning can play a role, although their nature is probabilistic. For instance, compared to the other suffixes, *-ett-* is used less frequently with bases ending in /r/, and more frequently with bases ending in /p/; *-in-* is used less frequently with bases ending in /k/. However, since our data are limited, these preferences should be confirmed by looking at a larger sample to be confident that they are indeed active in speakers' competence.

5.5 Antesuffixal phonological material

Table 10 shows the number of base-diminutive pairs that can be mapped to sub-schemas corresponding to the insertion of antesuffixal phonological material of different nature – namely, the interfix *-ic-*, the two different variants of the interfix *-er-/ar-*, and the cumulated suffix *-ol-*. Both the sample of this study and the one of M&P are taken into consideration.

Table 10. Base-diminutive pairs with antesuffixal *-ic-*, *-er-/ar-*, *-ol-* in our sample and in M&P

Sub-schema	Example	Pairs in M&P	Pairs in our sample
<[x]> ≈ <[[x] /iʃ/ DIM_SUFF V]>	<i>prompt</i> > <i>prompticino</i>	31 (2.4%)	40 (2%)
<[x]> ≈ <[[x] /er/ DIM_SUFF V]>	<i>link</i> > <i>linkerello</i>	9 (0.7%)	26 (1.3%)
<[x]> ≈ <[[x] /ar/ DIM_SUFF V]>	<i>film</i> > <i>filmarello</i>	10 (0.8%)	25 (1.3%)
<[x]> ≈ <[[x] /ol/ DIM_SUFF V]>	<i>shampoo</i> > <i>shampolino</i>	18 (1.4%)	20 (1%)

In general, those processes prove to be vital also in the sample of loanwords. This is an indication of their productivity that brings further support to previous qualitative observations by Dressler & Merlini Barbaresi (1989: 544, 1994: 533), by providing more examples and systematic quantitative data.

Again, differences in the distribution of these items in the two samples offer additional insight into their degrees of productivity. The interfixes *-er-/ar-* are more frequent in the sample of loanwords than in the overall lexicon, while the interfix *-ic-* and the cumulated suffix *-ol-* are less frequent in the former than in the latter. Consequently, the productivity profile that emerges from the analysis of new entries in the lexicon differs from what could be expected on the basis of data including more strongly lexicalised items. The interfix *-ic-* is the one that is most frequent in the overall lexicon, but it turns out to be less frequent than *-er-/ar-* when looking at loanwords only: the latter interfix is thus the one for which we find stronger evidence of productivity, despite the prevalence of the former in the overall lexicon.

Our quantitative data also provide a sample of novel entries that can be used to test claims that have been made about the preferential application of interfixes to bases with a specific phonotactic shape in terms of the number of syllables. Dressler & Merlini Barbaresi (1989: 245) claim that the shortness of the base does not favor the insertion of interfixes, citing examples of monosyllabic bases without interfixes such as *teino* from *tè* 'tea'. They also claim that trisyllabic words do not usually display interfixes, although they acknowledge the presence of occasional counterexamples. In contrast, in our data, we find that antesuffixal phonological material is mostly found with monosyllabic and disyllabic bases: the former possibility accounts for 23 diminutives of our sample, the latter for the other 10. The prevalence of monosyllabic bases is even more remarkable if compared with the distribution of the full sample of loanwords: monosyllables account for 35% (146/413) of the bases of our sample, but for

almost 70% of the ones used as bases of diminutives with some kind of antesuffixal phonological material. Furthermore, there is no example of diminutives with antesuffixal phonological material formed from a base with more than two syllables. This suggests that the length of the base in terms of number of syllables plays at least a probabilistic role in favoring the presence of antesuffixal phonological material.

6. Conclusions and future work

This study of the application of the most common diminutive formation processes to loanwords complements the one based on the overall lexicon provided by M&P in several ways. First of all, new patterns of form alternation emerge that were not attested in the sample of M&P. This is mostly due to the peculiar phonological characteristics of English loanwords. Unlike native Italian words, they often end in a consonant, thus giving rise to different overall patterns and alternations of final vowels. Gender-preserving diminutives are confirmed to be the most frequent scenario. We found a few clear examples of feminine-to-masculine switch, but no case of masculine-to-feminine switch, thus confirming the asymmetry between the two possibilities suggested by the quantitative data of M&P, and possibly suggesting an even more limited productivity of the latter – although the *caveats* on the difficulty in identifying cases of gender switch with these data should be borne in mind. Both the suffixes and the antesuffixal phonological material that M&P found to be most frequent are confirmed to be productive by their application to loanwords, even very recent ones, shown in the data presented in this paper. However, their frequency distributions are not the same in the two samples, shedding light on differences between what would be suggested by the frequency data of a study of the overall lexicon, as M&P, and claims that can be made on the actual productivity of morphological processes on the basis of a sample of loanwords, as in this study. The role of the suffix *-in-* as the default process seems to be even more pronounced than it would appear looking at the overall lexicon, while the productivity of *-ell-* and *-ett-* appears to be overestimated looking at frequency data due to their use in many well entrenched items; the observation that the latter suffixes are mostly applied to bases to which also the former can be applied brings further support to this claim. As for interfixes, *-er-/ar-* is the one that can be claimed to be most productive, as it is most frequently applied to loanwords, even if it is used less frequently than *-ic-* in the overall lexicon. Furthermore, these data provide a more systematic sample to test previous claims on phonological factors that motivate the preferential application of specific suffixes, or the insertion of antesuffixal phonological material. Summing up, focusing on loanwords provides new insight that is likely to more closely reflect productivity than looking at frequency distributions in the overall lexicon.

References

- Ackerman, Farrell, James P. Blevins & Robert Malouf. 2009. Parts and wholes: Patterns of relatedness in complex morphological systems and why they matter. In Juliette Blevins & James P. Blevins (eds.), *Analogy in grammar: Form and acquisition*, 54–82. Oxford: Oxford University Press.
- Baayen, Harald. 1992. On frequency, transparency and productivity. In *Yearbook of morphology 1992*, 181–208. Dordrecht: Springer.
- Beniamine, Sacha. 2018. *Classifications flexionnelles. Étude quantitative des structures de paradigmes* [Inflectional classifications. A quantitative study of paradigm structures]. Doctoral thesis. Université Sorbonne Paris Cité-Université Paris Diderot (Paris 7).
- Blevins, James P. 2016. *Word and paradigm morphology*. Oxford: Oxford University Press.
- Bonami, Olivier & Gilles Boyé. 2014. De formes en thèmes [From forms to themes]. In Florence Villoing, Sarah Leroy & Sophie David (eds.), *Foisonnements morphologiques. Études en hommage à Françoise Kerleroux*, 17–45. Paris: Presses Universitaires de Paris-Ouest.

- Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.
- Cinque, Guglielmo. 2015. Diminutive, endearing, augmentative and pejorative heads in the extended nominal projection. In Elisa Di Domenico, Cornelia Hamann & Simona Matteini (eds.), *Structures, Strategies and Beyond*, 67–81. Amsterdam: Benjamins.
- Corona, Luisa & Gina Russo. 2023. Italian deverbal verbs *in-Vcchiare*: a Manner of Approximating. *Lingue e linguaggio* 22(1). 89–116.
- Dressler, Wolfgang U. 2003. Degrees of grammatical productivity in inflectional morphology. *Italian Journal of Linguistics* 15(1). 31–62.
- Dressler, Wolfgang U. 2007. Productivity in word formation. In Gonia Jarema & Gary Libben (eds.), *The mental lexicon: Core perspectives*, 159–183. Amsterdam: Elsevier.
- Dressler, Wolfgang U., & Lavinia Merlini Barbaresi. 1986. How to fix interfixes? On the structure and pragmatics of Italian (and Spanish, Russian, Polish) antesuffixal interfixes and of English “intermorphic elements”. *Acta Linguistica Academiae Scientiarum Hungaricae* 36(1/4). 53–67.
- Dressler, Wolfgang U. & Lavinia Merlini Barbaresi. 1989. Interfissi e non-interfissi antesuffissali nell’italiano, spagnolo e inglese [Antesuffixal interfixes and non-interfixes in Italian, Spanish and English]. In Fabio Foresti, Elena Rizzi & Paola Benedini (eds.), *L’italiano tra le lingue romanze*, 243–252. Roma: Bulzoni.
- Dressler, Wolfgang U. & Lavinia Merlini Barbaresi. 1992. Italian diminutives as non-prototypical word formation. In Livia Tonelli & Wolfgang U. Dressler (eds.), *Natural morphology: Perspectives for the nineties*, 21–30. Padova: Unipress.
- Dressler, Wolfgang U. & Lavinia Merlini Barbaresi. 1994. *Morphopragmatics: Diminutives and intensifiers in Italian, German, and other languages*. Berlin-New York: Mouton de Gruyter.
- Gaeta, Livio & Davide Ricca. 2003. Frequency and productivity in Italian derivation: A comparison between corpus-based and lexicographical data. *Rivista di Linguistica* 15. 63–98.
- Gaeta, Livia & Davide Ricca. 2006. Productivity in Italian word formation: a variable-corpus approach. *Linguistics* 44(1). 57–89.
- Gardani, Francesco. 2013. *Dynamics of morphological productivity: The evolution of noun classes from Latin to Italian*. Leiden, The Netherlands: Brill.
- Grandi, Nicola. 1998. Sui suffissi diminutivi [On diminutive suffixes]. *Lingua e stile* 33(4). 627–654.
- Grandi, Nicola. 2002. *Morfologie in contatto. Le costruzioni valutative nelle lingue del Mediterraneo* [Morphologies in contact. Evaluative strategies in Mediterranean languages]. Milano: Franco Angeli.
- Grandi, Nicola. 2017. I diminutivi come marche di attenuazione e indeterminazione [Diminutives as markers of attenuation and indeterminacy]. In Oana-Dana Balaş, Adriana Ciama, Mihai Enăchescu, Anamaria Gebăilă & Roxana Voicu (eds.), *L’expression de l’imprécision dans les langues romanes*, 162–175. Bucharest: Ars docendi-Universitatea din Bucureşti.
- Lacić, I. 2025. InTens—A dataset of Italian intensified derivatives. Description and application in a productivity study. In Richard Huyghe, Megan Prudent & Justine Salvadori (eds.), *Proceedings of the Fifth International Workshop on Resources and Tools for Derivational Morphology*, 73–86. University of Fribourg.
- Masini, Francesca & Jenny Audring. 2019. Construction Morphology. In Jenny Audring & Francesca Masini (eds.), *The Oxford Handbook of Morphological Theory*, 365–389. Oxford: Oxford University Press.
- Masini, Francesca & M. Silvia Micheli. 2020. The morphological expression of approximation. The emerging *simil-* construction in Italian. *Word Structure* 13(3). 371–402.
- Masini, Francesca & Jacopo Di Donato. 2023. Non-prototypicality by (discontinuous) reduplication: the N-non-N construction in Italian. *Zeitschrift für Wortbildung / Journal of Word Formation* 7(1). 130–155.
- Merlini Barbaresi, Lavinia. 2004. *Paradigma dei suffissi alterativi* [Paradigm of alterative suffixes]. In Maria Grossmann and Franz Rainer, *La formazione delle parole in italiano*, 265–267. Tübingen: Niemeyer.
- Merlini Barbaresi, Lavinia. 2012. Combinatorial patterns among Italian evaluative affixes. *SKASE Journal of theoretical linguistics* 9(1). 2–15.
- Micheli, M. Silvia. 2023. The emergence of approximative values in the Italian prefixes *semi-* and *para-*. *Zeitschrift für Wortbildung / Journal of Word Formation* 7(1). 156–176.

- Micheli, M. Silvia & Matteo Pellegrini. 2025. Exploring the form of Italian diminutives: Alternation patterns and constructional schemas. *Constructions and Frames* 17(2). 175–210. <https://doi.org/10.1075/cf.23028.mic>
- Micheli, M. Silvia & Martina Verdelli. 2025. Old and New Intensifying Prefixes in Italian: The Case of *uber-*. *Zeitschrift für Wortbildung/Journal of Word Formation* 9(2). 28–68.
- Pedroncelli, Giovanni. 2023. *Sviluppo di un'applicazione per la trascrizione fonetica automatica dell'italiano standard* [Development of an application for the automatic phonetic transcription of standard Italian]. Master's thesis. Università degli studi Guglielmo Marconi, Roma.
- Rainer, Franz. 1990. Appunti sui diminutivi italiani in *-etto* e *-ino* [Notes on Italian diminutives ending in *-etto* and *-ino*]. In Monica Berretta, Piera Molinelli & Ada Valentini (eds.), *Parallela 4. Atti del V incontro di italo-austriaco della Società di Linguistica italiana (Bergamo, 2-4 ottobre 1989)*, 207–218. Tübingen: Narr.
- Rainer, Franz. 2024. Origin and development of agent nouns in *-ino*. *Études romanes de Brno* 45(3). 35–53.
- Thornton, Anna Maria. 2009. Constraining gender assignment rules. *Language Sciences* 31(1). 14–32.
- Wurzel, Wolfgang U. 1984. *Flexionsmorphologie und Natürlichkeit* [Inflectional morphology and naturalness]. Berlin: Akademie-Verlag.

A case-study on proper nouns formed by *-exit* in Contemporary Italian

Mauro Le Donne
University of South Bohemia
mauroledon@gmail.com

1. Introduction

This study investigates the micro-diachronic trend and the morphotactic features of the lexical innovations formed by *-(e)xit* (Mattiello 2018; Caffarelli 2018; Lalić-Krstin & Silaški 2019) in Contemporary Italian.¹ The data consists of English *complex loanwords* (1a-c) and Anglo-Italian *hybrid formations* (2a-c) (Seifart 2015) extracted from two different Italian web corpora:

- (1) a. *Irex* ‘hypothetical withdrawal of Ireland from the EU’ < *Ireland* / *Irish*
b. *Faragexit* ‘call for the resignation of Nigel Farage’ < (*Nigel*) *Farage*
c. *Nato-exit* ‘hypothetical withdrawal from NATO’ < *NATO*
- (2) a. *Romexit* ‘delocalisation of companies from Rome’ < *Roma* ‘Rome’
b. *Gentilexit* ‘withdrawal of support for Paolo Gentiloni’ < (*Paolo*) *Gentiloni*
c. *Tavexit* ‘suspension of the fundings for the TAV project’ < *TAV*²

Even if the source language cannot always be determined with certainty, it is argued that the items in (1a-c) were borrowed by remote-contact (mostly) with English, whereas the items in (2a-c) have been formed *extracting* and *extending* the pattern on native stems of the recipient language (i.e. Italian) (Seifart 2015; Gardani 2021). With regard to the morphological status of *-(e)xit* in this pattern, it could be argued that it is open to at least two different interpretations:

- (i) since it originally appeared as the right-hand constituent of a semi-complete blend (i.e. *Grex* < *Greek Euro Area exit*) (Ronneberger-Sibold 2006), *-(e)xit* could be considered an output of blending, that is, a splinter (Berman 1961; Adams 1973), or a “morphemised splinter”, occurring in other co-hyponymic blends (Lalić-Krstin & Silaški 2019: 228);³
- (ii) since it appears in different words in a paradigmatic series, *-(e)xit* could be considered an exponent of this series (Montermini 2018: 433-435).⁴

¹ The label Contemporary Italian indicates the variety of Italian spoken nowadays, whose specific features are described in D’Achille (2003).

² The extended form of the initialism *TAV* is It. *Treno ad Alta Velocità* ‘high-speed train’. When not otherwise specified, the second constituent (i.e. *exit*) is not reported to avoid redundancy.

³ Indeed, even if there is no formal reduction of *-(e)xit* in *Grex* or *Brex*, the reduced splinter *-xit* is also attested. However, not all scholars would agree on this hypothesis, according to the assumption that blends “cannot form series” (Fradin 2015: 390).

⁴ This exponent could be either classified as an affixoid or as a combining form. Indeed, it takes a more specific meaning than its verbal and nominal counterparts, i.e. *to exit* and *the exit*. For instance, it refers to a call for the resignation of a politician in (1b), or to the hypothetical withdrawal of a political entity in (1a) and (1c).

Leaving the treatment of *-(e)xit* as an exponent to future works, in this study, it will be considered as in (i). A unified treatment of the two lexical sets in (1a-c) and (2a-c) is justified on the basis of the same morphological (i) and semantic (ii) features, indeed:

- (i) both consist of blends formed by the combination of a proper noun and the sequence *-exit* which functions as the semantic head, yielding different degrees of reduction and possible overlap, e.g. *Bruxit* < *Bruxelles*, *Ponsexit* < *Ponsacco*; or even neither formal reduction nor overlap, e.g. *Greece-exit* < *Greece*, *Pragaexit* < *Praga* ‘Prague’;
- (ii) both exhibit many-to-many form-meaning relationships, that is, one form might correspond to multiple meanings, e.g. *Wexit* ‘call for the resignation of Arsène Wenger’ / ‘hypothetical secession of Canadian Western provinces’; and one meaning might correspond to multiple forms, e.g. *Itaexit* / *Italexit* / *Italiaexit* / *Italiexit* / *Itxit* ‘hypothetical withdrawal of Italy from EU’.

The research questions addressed in this study are the following:

- (i) How did this morphological process change through time in Contemporary Italian?
- (ii) Are there any differences in the morphological make-up of the two lexical sets as in (1a-c) and (2ac)?

The sociolinguistic and lexicographic background of this morphological process will be discussed in Section 2, the methodology by which the items in (1a-c) and (2a-c) were extracted from two different web corpora of Italian will be illustrated in Section 3, the micro-diachronic trend of the *-exit* words will be examined in Section 4, along with their morphotactic features (4.1). Finally, in section 5, our conclusions will be drawn.

2. The origin of *-exit* and its spread in Contemporary Italian

As is well known, the emergency of *-exit* in English begun with the appearance of the proper noun *Grexit* in 2012, originally, the blend of an (ethnic) adjective and a noun (i.e. *exit*), specifically coined to shorten the syntagm *Greek Euro Area Exit* (Thornton 2016; Caffarelli 2018: 447). Even if the model became successful, at least to the point that some lexical innovation based on *Grexit* followed its coinage; its fate cannot be compared to the global fame of its co-hyponym *Brexit* (< *British exit*), which appeared shortly thereafter and can be considered the *leader word* of many subsequent analogical coinages (Thornton 2016; Mattiello 2018: 5-6; Lalić-Krstin & Silaški 2019: 223).⁵ Indeed, due to its popularity, *Brexit* was listed into the dictionaries and the neological data banks of various European languages, Italian included (Caffarelli 2018).⁶

From a sociolinguistic point of view, the spread of *Brexit* in different natural languages can be interpreted as a symptom of globalization, due to the increasing role of social media in contemporary communication (Antonelli 2016). The language of social media is peculiar for

⁵ The term *leader word* refers to words that are particularly salient in a word formation pattern, and that may serve as analogical model to derive new words (Rainer 2013: 152; Lasserre & Montermini 2015: 169).

⁶ In 2016 *Brexit* was officially recorded in the *Oxford English Dictionary* (<https://www.oed.com/>). Beside English, *Brexit*’s entry has been listed in other lexicographic resources, for instance: in Italian (2016), see *Treccani*’s neological data bank (https://www.treccani.it/magazine/lingua_italiana/neologismi/); in Spanish (apparently in 2017), see the *Diccionario panhispánico del español jurídico*, co-authored by the *Real Academia Española* (<https://dpej.rae.es/>); in French (2019), see *Larousse*’s dictionary (<https://www.larousse.fr/>); in German (2017), see *Duden*’s dictionary (<https://www.duden.de/>).

both its diaphasic and diamesic features, as it is influenced by the social prestige of English as *lingua franca*, as well as other sectoral and professional communities (Antonelli 2016; Cortelazzo 1988). Furthermore, social media facilitate the spread of “innovations in geographically discontinuous settings” (Sayers 2014: 192), by connecting distant speech communities and engaging an ever-growing number of users with their informality and interactivity. In this way, lexical innovations may circulate without any direct contact among speakers (Buchstaller *et al.* 2010; Grieve *et al.* 2016; Brasolin *et al.* 2023). Another pertinent aspect concerns the informalisation of political jargon, a change that has been acknowledged as a significant development in social media discourse, rendering this variety more appealing and accessible to a broader audience, even by the coinage of occasional, catchy words (Lalić-Krstin & Silaški 2019: 224; Brasolin *et al.* 2023: 5, 10).

Zooming in on the Italian sociolinguistic situation, it is well known that the exclusive usage of dialects in Italy has gradually given way to situational and stylistic usages alongside Italian, or rather, alongside Italian regional varieties (Lepschy & Lepschy 2019: 51-52; De Mauro 2016: 111-120). This is confirmed by De Mauro (2016), who reports that “[w]ithin the same conversation or discourse, speakers [...] shift from the standard language to the dialect or vice versa (*code-switching*), or else they blend words from the two different idioms (*code-mixing*)” (2016: 119).⁷ In this respect, it has been argued that many bilingual speakers “do not consider Italian a valuable heritage”, and that issues related to contact-induced language change may come from both external and internal sources (Marello 2020: 168-169). In other words, the bilingual attitude of many Italian speakers and the sociopolitical prestige of English in the varieties under scrutiny (i.e. the journalese, the jargon of politics) provide a fertile ground that could have led to the borrowing of complex loanwords (henceforth, CLs) in (1a-c), and, thereafter, to the extraction and extension of the morphological pattern by which hybrid formations (henceforth, HFs) in (2a-c) were formed.

In the context of this study, HFs are defined as words formed by a combination of a native left-hand base and a borrowed right-hand base, whereas CLs consist of two borrowed bases, with *exit* functioning as the borrowed right-hand constituent for both types of word formation.⁸ The extension of this pattern in Italian journalese could be witnessed by the frequent combination with different toponyms of Southern, Central and Northern Italy:

- (3) a. regions: *Ligurexit* (< *Liguria*), *Marchexit* (< *Marche*), *Calabrexit* (< *Calabria*)
- b. cities: *Triexit* (< *Trieste*), *Pescarexit* (< *Pescara*), *Napolexit* (< *Napoli* ‘Naples’)
- c. towns: *Trexit* (< *Trebaseleghe*), *Prexit* (< *Preturo*), *PA-exit* (< *Piazza Armerina*)

The word-formations in (3a-c) consistently refer to an Italian toponym, although with different levels of granularity. In fact, the toponym may refer to regions (3a), cities (3b) or towns (3c).⁹ Moreover, the items in (3a-c) share (more or less) the same semantic content: ‘(hypothetical) withdrawal from a socio-politic entity X’. Computing the geocoordinates of all the toponyms

⁷ “Nell’ambito di una stessa conversazione o di uno stesso discorso i parlanti [...] passano dalla lingua al dialetto o viceversa (*code switching*), oppure mescolano parole dei due idiomi diversi (*code mixing*)” (2016: 119).

⁸ In Seifart’s paper (2015), HFs are defined as the “combinations of the borrowed affix with native stems”, while CLs are defined as “combinations of the borrowed affix with likewise borrowed stems” (2015: 517). Since the present study does not deal with affixed words properly, these definitions are adopted in a looser sense.

⁹ It could still be argued that the (original) ethnic adjective is involved, as in *Grexit* (i.e. *Greek*) and *Brexit* (i.e. *British*), e.g. *Calabrexit* < (?*Calabria* / ?*calabrese* ‘Calabrian’). However, for many -*exit* words this hypothesis seems unlikely, e.g. *PA-exit* in 3c, *Siexit* (< *Siena* / **senese* ‘Sienese’), *Terni-exit* (< *Terni* / **ternano*). Note also that in many dictionaries the etymology of *Brexit* is *Britain exit* (see also Thornton 2016).

involved in the HFs extracted in this study, in Figure 1, it is possible to grasp the areal extension in which such (sometimes alleged) claims of socio-political autonomy raised debates:¹⁰

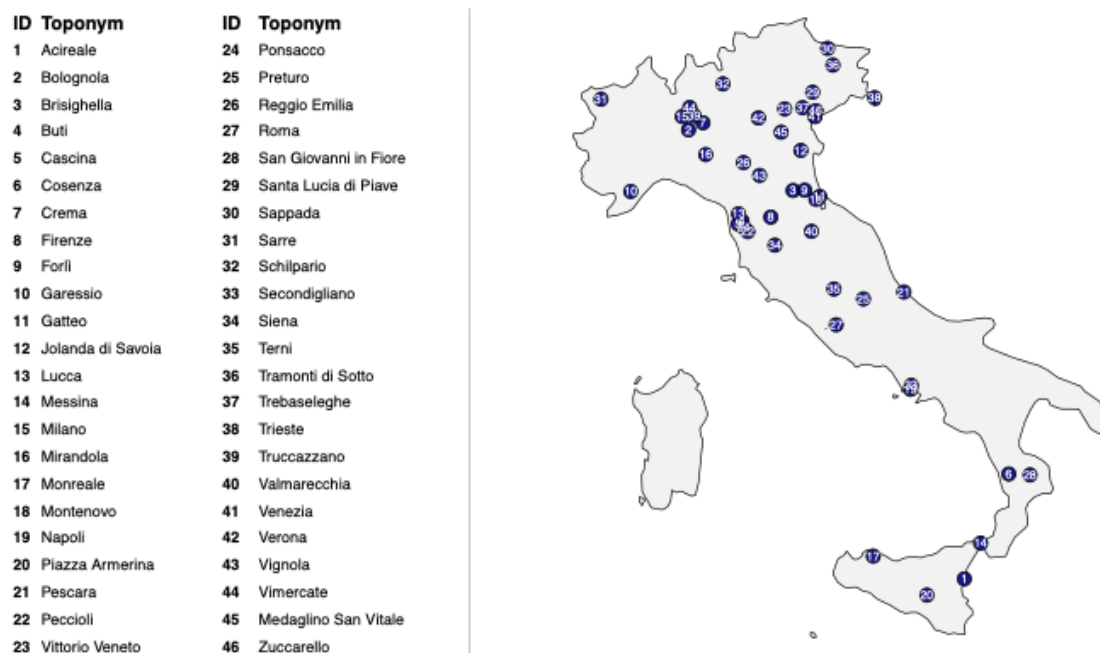


Figure 1. A map of Italian toponyms (cities and towns) involved in the -exit word formation process

3. Data gathering and annotation

The -exit words were extracted from two different web corpora of Italian: the static web corpus *Italian TenTen 2020* (henceforth ITT; Jakubíček *et al.* 2013); and a monitor, time-stamped corpus, *Italian Trends* (henceforth ITr; Herman *et al.* 2025), which is regularly updated with news articles provided by RSS feed since 2014 (Trampuš & Novak 2013). The two web corpora were selected for their suitability in relation to the specific phenomenon analysed in this study, a relatively recent word formation process, as the first -exit formation is attested only in 2012 (see Section 2). In this respect, both corpora are based on documents produced in the current and the preceding decade. Moreover, the large design of both makes it possible to retrieve many low-frequency items (e.g. *hapax* and *dis legomena*), thus ensuring a sufficiently rich sample.¹¹ In both cases, the process of semi-automatic extraction can be summarised in three steps:¹²

¹⁰ The map and the other plots in this paper were realised using the software R (R Core Team 2022) and the following libraries: *sf* (Pebesma 2018), *ggplot2* (Wickham 2016), *dplyr* (Wickham *et al.* 2023), *patchwork* (Pedersen 2024), *rnaturalearth* (Massicotte & South 2026).

¹¹ The ITT contains 14,514,566,714 tokens, 33,582,719 types, and 30,718,525 documents. At the time of writing, the ITr contains 12,333,693,644 tokens, 14,592,351 types, and 34,476,008 documents (accessed on 17/03/2026). As a monitor corpus, only .it documents of the decade 2014-2024 were considered in the extraction phase (see note 12). This subcorpus contains 8,182,953,879 tokens and accounts for 66.3% of the entire ITr.

¹² There is one specificity that must be mentioned. The time-stamped design of the ITr makes this corpus suitable for micro-diachronic analyses. Therefore, the extraction process was performed slightly differently. At first, the corpus was subdivided in 11 subcorpora, each one corresponding to a single year of the time span 2014-2024, i.e. 11 years in total. To ensure that the web articles were really produced in Italian, only the web pages with the .it domain were chosen for the extraction. The points listed in (i-iii) were thus repeated 11 times to obtain the ITr lexical dataset.

- (i) The CQL-tool available in *Sketch Engine* (Kilgariff *et al.* 2014) was used to retrieve all the tokens containing the string *-xit*, excluding very frequent false positives, e.g. the Latinism *resurrexit* or the loanword *exit*;
- (ii) The list of tokens (together with their co-textual occurrence) was downloaded in *.xml* format and manually-checked to obtain a list of types;
- (iii) The list of types was inserted into another *.xml* file and annotated according to different chronological, morphotactic, and morphosemantic variables.

This process was repeated for both corpora to obtain two separate datasets of types, whose raw and normalised values are reported in Table 1, notice that the numbers include doublets, that is, items that were found in both corpora, for example, *Brexit* and *Grexite*:

Table 1. Raw/normalised number of CLs and HFps formed by *-exit* in the ITT and ITr¹³

Corpus	CL/pmwp (%)	HF/pmwp (%)	Total/pmwp (%)
ITT	140/0.0096 (53.23%)	123/0.0085 (47.14%)	263/0.0181 (46.76%)
ITr	172/0.0210 (52.6%)	155/0.0189 (47.4%)	327/0.0400 (55.42%)
Total (%)	311 (52.71%)	279 (47.28%)	590 (100%)

As it could be noted, the *-exit* formations are more frequent in the ITr corpus than in the ITT. This difference may be explained by the more specific textual genre of the ITr corpus. Indeed, the latter consists predominantly of news articles, in which the use of *-exit* words is likely to be favoured. On the contrary, the ITT corpus is a static web corpus, thus it is more heterogeneous from this point of view.

4. Micro-diachronic trend of *-exit* in the ITr

After its first appearance in 2012, *Brexit* was included four years later in *Treccani*'s neological data bank (see note 6). Since then, the large number of CLs and HFps attested (see Table 1) makes a micro-diachronic examination worthwhile, in order to trace how this "process of language-internal analogical extension" developed over time (Seifart 2015: 511). To explore this issue, the design of the ITr was exploited to analyse the development of the *-exit* pattern during the years 2014-2024. Two variables were particularly suitable for this purpose, namely the number of new types recorded each year and their date of first attestation.

In Figure 2, these variables were used to plot two complementary aspects of productivity: the degree of saturation of the word formation process over time and the distribution of new coinages per year.¹⁴ Both plots clearly point to a sharp increase of *-exit* formations in 2016. In

¹³ In a recent study, Wolfer & Koplenig (2025) show that normalised frequencies are generally reliable in comparisons involving large corpora. Nonetheless, "[f]or smaller corpora in particular, [...] the comparability for less frequent types can quickly decrease" (2025: 20), and even "larger corpora have the same problem in their own lower frequency ranges" (2025: 19). Following this insight from corpus linguistics, in Table 1 normalised frequencies are reported together with raw frequencies. The percentages refer to raw frequency values.

¹⁴ In the left plot, the values adjacent to the data points represent the cumulative number of types in percentage; in the right plot, they represent the number of new *-exit* formations per year. In the latter case, normalised values are used to account for the uneven distribution of documents across years in the ITr corpus, with 2014 and 2015

that year, the new coinages exceeded those of 2015 by more than 25%, reaching a normalised distribution of 0.19, the highest value in the time span considered. Of course, this peak is intimately tied to the related socio-political event of that year, namely the referendum in which 51.9% of British people voted in favour of leaving the European Union. Moreover, since the withdrawal was formally carried out only four years later, it seems not coincidental that the word formation process registers a doubling of new coinages in 2019 and 2020 (46 and 54 types, respectively) compared to 2018 (21 types).¹⁵ By contrast, in the final years of the timeline (2023-2024), there is only a small increase of -exit word formations. Indeed, from 2022 onwards, the cumulative sum of new types increases by only about 3%, while the normalised distribution of new types per year falls to 0.01, the lowest value registered in the time span considered.

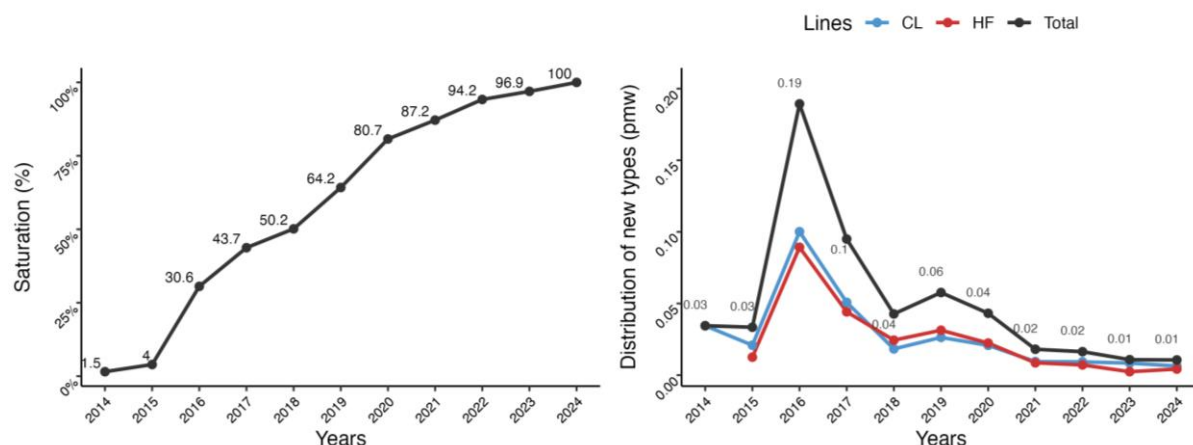


Figure 2. Cumulative distribution and diachronic trend of -exit words in the ITr corpus

As for the distribution of new HF and CLs, the right-hand plot in Figure 2 highlights similar micro-diachronic trends over the time span considered.¹⁶ The first HF is recorded in 2015, one year later than the first CLs, among which we find already in 2014 successful neologisms, such as *Brexit*, *Grex*, and *Nexit* (< *Netherland*). The first HF recorded is: *Crocexit* ‘call for the resignation of Rosario Crocetta’ < (*Rosario*) *Crocetta*; *Iridexit* ‘(hypothetical) decommission of the Iride beach club’ < (*lido*) *Iride* ‘name of a beach club in Sassari’; and *Telexit* ‘the crisis of TV programming’ < *televisione* ‘television’. Interestingly, HF even outnumber CLs in the period 2018-2020. Since the years between 2016 and 2020 represent the most productive phase of -exit in terms of new coinages, with 2016 as its peak, the predominance of HF in the final years of this period may reflect a shift towards an increased use of the -exit pattern in Italian journalese, where it is employed to refer to national and local socio-political events alongside international ones, by analogy with the most salient example of such events, namely *Brexit*.

containing fewer documents than 2020. For more information on the design of the ITr: see <https://www.sketchengine.eu/italian-trends-corpus/> (accessed on 19/03/2026).

¹⁵ Although the number of newly attested types increased in 2020, the normalised distribution of new coinages in the right plot decreases due to the substantial growth of the corpus size in that year, which attenuates the incidence of the value reported.

¹⁶ To test whether there is an association between the two variables (year of occurrence ~ origin), a chi-square test was performed on a contingency table containing the number of new CLs and HF across years during the time span 2014-2024. The results do not indicate a significant association between the two variables ($\chi^2 = 9.606$, $df = 10$, $p = 0.47$), suggesting that the number of new coinages across years is independent of their origin.

However, this does not necessarily entail a consolidation of the *-exit* pattern in Italian. On the contrary, apart from rare highly (token) frequent items such as *Grex* and *Brex*, subsequently followed by the successful HF *Italex* (2016), the category of *-exit* words consists almost exclusively of *hapax* and *dis legomena*, as shown in Table 2.

Table 2. *Hapax* and *dis legomena* among *-exit* words in the ITr

Type	Hapax (%)	Dis (%)
CLs	95 (51.63%)	22 (50%)
HFs	89 (48.36%)	22 (50%)
Total	184 (100%)	44 (100%)

More than 50% of the items are *hapaxes* (184/327; 56.26%). Furthermore, *hapax* and *dis legomena* together cover almost three quarters of the entire dataset (228/327; 69.72%). Such numbers clearly indicate that the *-exit* word formation pattern yields predominantly ephemeral items, since more than half of the items formed by *-exit* are occasional creations; in this respect, there is no significant difference between CLs and HFs. The ephemerality of many items is further confirmed by the median absolute frequency, which is always equal to 1 due to the substantial number of *hapaxes*. Conversely, the mean absolute frequency is influenced by a small number of highly frequent items, mostly CLs: $\bar{x}_{CL}$ 712.73; $\bar{x}_{HF}$ 60.63.

Another relevant aspect concerns the permanence of types during the time span 2014-2024. Figure 3 shows the number of years in which each *-exit* word is attested in the ITr. As already mentioned, nearly three quarters of the dataset (244/327 items; 74.6%) do not “remain” at all, disappearing after one year and revealing once again that this kind of word formation is mostly transient in nature. In this respect, there is no substantial difference between CLs (123/244) and HFs (121/244), e.g. *BRIexit* ‘withdrawal from the Belt and Road Initiative’ (< *Belt and Road Initiative*); *Insalex* ‘(hypothetical) dismissal of bagged salads’ (< *insalata* ‘salad’). In addition, 66/327 items (20.18%) are attested over a span of 2-4 years, and may be considered moderately ephemeral. Even in this case, we find balance between HFs (31/66; 46.96%) and CLs (35/66; 53.03%), e.g. *Monrealex* ‘secession of near-villages from the municipality of Monreale’ (< ?*Monreale* / ?*monrealese*); *coal-exit* ‘decommission of coal as a polluting source of energy’ (< *coal*). As mentioned before, the relevant difference between HFs and CLs concerns the most frequent items. Among the few forms attested for 5 or more years, 14/17 (82.35%) are CLs, for example, *Polexit* ‘(hypothetical) withdrawal of Poland from EU’, which is attested for 9 years. By contrast, only 3/17 (17.64%) are HFs: *Italex*, *Italiexit*, *Italiaexit* ‘hypothetical withdrawal of Italy from EU’, three forms with a slightly different degree of retention of the first constituent.¹⁷

¹⁷ *Polexit* and *Italex* became names of extreme right-wing political parties as well. In both cases, it is impossible (and maybe also irrelevant) to understand whether the first constituent is an ethnic or a toponym. Due to the occurrence of alternative forms, such as *Polaexit*, *Polandexit* and *Polonexit* (a HF), it seems plausible to assume that the toponym is more salient for speakers. However, in some cases, it is impossible to decide. Furthermore, sometimes it is hard to say with certainty in which language CLs were formed, as only HFs were meticulously cross-checked in parallel language corpora, such as *English Trends*. In the case of *Polexit*, it seems that the original word formation comes from English. However, this point does not seem to be crucial since, the pattern becoming clear enough in one language means that, whether *Pol-* comes from *Poland* (or *Polish*?) or from *Polska* (or *polski*?) could vary from speaker to speaker.

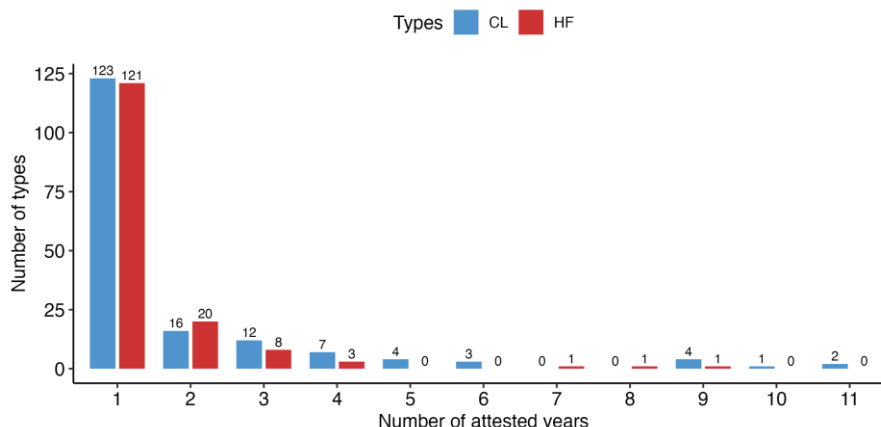


Figure 3. Distribution of -exit types according to occurrence(s) per year

4.1 The syllabic length and the shape of formative in -exit words

In this section, the morphotactic features of the -exit words attested in the ITT and the ITr are examined. It should be noted that, except for the most frequent items and a limited number of neologisms, the IPA transcripts of the phonetic realisation of -exit words are not available. For this reason, the following analysis is primarily based on orthographic evidence.

A particularly relevant aspect of the -exit word formation process concerns the shape of its constituents. As already mentioned, this morphological pattern emerged at first from the need to shorten a longer noun phrase, fulfilling an economic motivation or a “compacting function”, i.e. *Grexit* < *Greek Euro Area Exit* (Renner 2020: 9). In this respect, the types in (4a-e) preserve the original right-headed structure, while presenting prosodic and morphotactic differences in the shape of its formants:¹⁸

- (4) a. *campetto-exit* ‘dismantlement of the playground’ < *campetto* ‘playground’
- b. *Armanexit* ‘abandonment of Milan by Giorgio Armani’ < (*Giorgio*) *Armani*
- c. *Renxit* ‘withdrawal of support for Matteo Renzi’ < (*Matteo*) *Renzi*
- d. *Balo-exit* ‘sale of the football player Mario Balotelli’ < (*Mario*) *Balotelli*
- e. *PD-exit* ‘abandonment of the Democratic Party’ < *PD* ‘Democratic Party’

Notably, the prosodic output of the examples in (4a-e) does not always correspond to the original disyllabic model imposed by *Grexit* and *Brexit*. Instead, we find variation between three (4b), four (4b-d-e), and five syllables (4a). In terms of morphotactic features of the examples, while (4c) can still be analysed as in the domain of blending, in other cases -(e)xit is attached to different types of morphological units, exhibiting thus a more affix-like behaviour (Bauer *et al.* 2019: 61-63). In particular, the left-hand constituent may correspond to: a full word in (4a); a root in (4b);¹⁹ a splinter in (4c); a clipping in (4d);²⁰ an initialism in (4e). Even

¹⁸ The extended form of the initialism *PD* is *Partito Democratico*.

¹⁹ Even if the term “lexical root” applied to a proper noun may look theoretically problematic, the position of the split and its frequency in the sample (see Table 6) point to the same preference for roots (and stems) as bases in Italian word formation and inflection (Dressler & Thornton 1991). Moreover, as Cutler (1981) observes, “[t]he acceptability of neologisms [...] depends crucially on the ease with which their base word can be recognized within them” (Cutler 1981: 76-77).

²⁰ In this study, the distinction between clippings and splinters is primarily based on their different morphosyntactic status. Clippings are truncated words that typically retain the denotation of the full form, although their connotation

the shape of the “morphemised splinter” *-(e)xit* (Lalić-Krstin & Silaški 2019: 228) may occur either in its full form (4a-b-d-e) or in the reduced variant *-xit* (4c).

In Table 3, the number of syllables of the *-exit* words gathered from the ITT and the ITr is reported.²¹

Table 3. Syllabic distributions of *-exit* words in the ITr and the ITT

σ	Examples	ITr (%)	ITT (%)
2	<i>Bayxit</i>	77 (23.91%)	74 (28.46%)
3	<i>Vinexit</i>	105 (32.6%)	97 (37.3%)
4	<i>Corbynexit</i>	111 (34.47%)	70 (26.92%)
5	<i>Mezzogiornexit</i>	26 (8.07%)	14 (5.38%)
6	<i>Agroinvest-exit</i>	3 (0.93%)	3 (1.15%)
7	<i>San Giovanni in Fiorexit</i>	0 (0%)	2 (0.77%)
Total		322 (100%)	260 (100%)

Although the two samples were collected from corpora of different sizes, both distributions are comparable. In fact, both show clusters of values around 2-4 syllables, and both are centred around 3 syllables, with similar dispersion.²² The only differences in the ITT corpus involve fewer trisyllabic types and the attestation of items with 7 syllables., e.g. *cooperativExit* ‘end of participation in the cooperative compliance regime’ < *cooperative (compliance)*.²³ Table 4 shows how the CLs and HFes differ from each other in relation to the same variable (syllabic length).

Table 4. Syllabic distributions of CLs and HFes in the ITr and the ITT

σ	Examples	ITr		ITT	
		CL (%)	HF (%)	CL (%)	HF (%)
2	<i>Bayxit</i>	61 (36.3%)	16 (10.38%)	58 (41.42%)	16 (13.3%)
3	<i>Vinexit</i>	63 (37.5%)	42 (27.27%)	53 (37.85%)	44 (36.6%)
4	<i>Corbynexit</i>	38 (22.61%)	73 (47.4%)	24 (17.14%)	46 (38.3%)
5	<i>Mezzogiornexit</i>	6 (3.57%)	20 (13%)	5 (3.57%)	9 (7.5%)
6	<i>Agroinvest-exit</i>	0 (0%)	3 (1.94%)	0 (0%)	3 (2.5%)
7	<i>San Giovanni in Fiorexit</i>	0 (0%)	0 (0%)	0 (0%)	2 (1.6%)
Total		168 (100%)	154 (100%)	140 (100%)	120 (100%)

may shift, e.g. *Balo* < (*Mario*) *Balotelli* in (4d). By contrast, splinters are bound, usually submorphemic elements involved in blending, and generally lack word status. More specifically, they consist of phonetically reduced forms of one or more lexical units, e.g. *Ren-* < (*Matteo*) *Renzi* and *-xit* < *exit* in (4c) (Adams 1973: 149-150). Although *-exit* formally corresponds to a full word in many cases, it also occurs as a splinter, i.e. *-xit* (see note 3).

²¹ Note that in eight cases the number of syllables was not annotated for the dubious pronunciation of the item, e.g. *Fdl-exit* ‘opting out from regional elections in Calabria by Fratelli d’Italia’ < *Fdl* (< *Fratelli d’Italia* ‘name of a right-wing Italian political party’).

²² ITr: $\bar{x}_\sigma$ 3.3, M_σ 3, SD_σ 0.95. ITT: $\bar{x}_\sigma$ 3.15, M_σ 3, SD_σ 0.98.

²³ This item was considered a HF, although it falls out of the definition given in Section 2. However, it must be noted that *cooperativExit* was not found in any parallel English corpora. This item seems to be an occasionalism used in the juridical language variety of Italian: <https://loconteandpartners.it/wp-content/uploads/2019/02/II-Quotidiano-Ipsa-del-08.10.16.pdf> (accessed on 06/04/2026).

In Table 4, the difference in syllabic length between CLs and HFs emerges quite clearly from the reported percentages. In particular, CLs tend to cluster around 2 and 3 syllables, whereas HFs are more centred around 3 and 4 syllables, e.g. *Zuxit* ‘protest against the automated censorship operated by Facebook’ (< *Mark Zuckerberg*) (ITT) vs. *Melonexit* ‘reject of the European and constitutional principles by Giorgia Meloni’ (< *Giorgia Meloni*) (ITr), and this tendency is also corroborated by descriptive statistics²⁴. Moreover, compared to CLs, HFs seem to allow both shorter and longer prosodic structures. For instance, HFs reach an output of 6 syllables, or even 7 in the ITT sample, whereas CLs never adopt longer prosodic outputs, e.g. *Secondiglianexit* ‘(hypothetical) secession of Secondigliano from Naples’ municipality’ < *Secondigliano*, *San Giovanni in Fiorexit* ‘(hypothetical) secession of San Giovanni in Fiore from Cosenza’s municipality’ < *San Giovanni in Fiore*. Overall, the syllabic structure of CLs and HFs represents one relevant point of divergence between the two categories.

The final feature to be examined is the shape of the formatives (full words, morphs, splinters, clippings, initials) in the two samples of -exit words (for a similar analysis, see Bauer *et al.* 2019: 61-63). The global values are reported in Table 5.

Table 5. Distribution of formatives for the -exit types in the ITr and ITT

Constituent	Formatives	ITr (%)	ITT (%)	Example	Referent
left	full word	56 (17.1%)	40 (15.2%)	<i>Islamexit</i>	<i>Islam</i>
	root	90 (27.52)	70 (26.61%)	<i>Atalantexit</i>	<i>Atalanta</i>
	splinter	136 (41.6%)	114 (43.34%)	<i>Jolexit</i>	<i>Jolanda di Savoia</i>
	clipping	10 (3%)	5 (1.9%)	<i>Demexit</i>	<i>Democrats</i>
	initialism	21 (6.42%)	21 (8%)	<i>Ehxit</i>	<i>Euskal Herria</i>
	unclear	14 (4.28%)	13 (4.94%)	<i>Frontexit</i>	<i>Frontex</i>
	Total	327 (100%)	263 (100%)		
right	-exit	277 (84.7%)	207 (78.7%)	<i>Benettexit</i>	<i>Benetton</i>
	-xit	43 (13.15%)	48 (18.25%)	<i>Bielxit</i>	<i>Marcelo Bielsa</i>
	unclear	7 (2.15%)	8 (3.04%)	<i>Texit</i>	<i>Texas</i>
	Total	327 (100%)	263 (100%)		

Despite the different number of observations, there are no substantial differences. The two samples are coherent in terms of morphotactic segmentation, and even if the samples were collected from two different corpora, the difference between percentages remain relatively small, generally below 5 percentage points. By and large, the left-constituent is preferably retained as a splinter, although there is a certain degree of variation, since lexical roots and full words also occur with decreasing frequency. On the contrary, clippings appear to be less salient formatives. The right constituent tends to be -exit in most cases, whereas -xit is used less frequently. The variation in the shape of the left-hand constituent may reflect the fact that the -exit word formation process is still emerging in Contemporary Italian. On the other hand, even if analogy grasps much more than the mere shape of the items, it is evident that many do not follow the *Brexit*-model at all, being substantially longer and formed from meaningful morphological elements, such as words and lexical roots. At this point, it can be interesting to distinguish once again between CLs and HFs:

²⁴ ITr: $\bar{x}_\sigma$ 2.93, M_σ 3, SD_σ 0.85 (CLs) vs. $\bar{x}_\sigma$ 3.69, M_σ 4, SD_σ 0.89 (HF). ITT: $\bar{x}_\sigma$ 2.83, M_σ 3, SD_σ 0.83 (CLs) vs. $\bar{x}_\sigma$ 3.54, M_σ 3.5, SD_σ 1.01.

Table 6. Distribution of formatives for the CLs and HFs in the ITr and ITT

Constituent	Formatives	ITr		ITT	
		CL (%)	HF (%)	CL (%)	HF (%)
left	full word	33 (19.18%)	23 (14.83%)	22 (15.71%)	18 (14.63%)
	root	25 (14.53%)	65 (41.93%)	19 (13.57%)	51 (41.46%)
	splinter	86 (50%)	50 (32.25%)	75 (53.57%)	39 (31.7%)
	clipping	8 (4.65%)	2 (1.3%)	4 (2.85%)	1 (0.81%)
	initialism	11 (6.4%)	10 (6.45%)	10 (7.14%)	11 (8.94%)
	unclear	9 (5.23%)	5 (3.22%)	10 (7.14%)	3 (2.44%)
	Total	172 (100%)	155 (100%)	140 (100%)	123 (100%)
right	<i>-exit</i>	128 (74.41%)	149 (96.12%)	97 (69.28%)	110 (89.43%)
	<i>-xit</i>	38 (22.1%)	5 (3.22%)	37 (26.42%)	11 (8.94%)
	unclear	6 (3.48)	1 (0.65%)	6 (4.28%)	2 (1.62%)
	Total	172 (100%)	155 (100%)	140 (100%)	123 (100%)

In Table 6, the formatives involved in both categories (i.e. CLs and HFs) yield quite homogeneous values. More importantly, Table 6 highlights the important differences between the categories of CLs and HFs. In particular, the left-constituent of HFs is selected preferably among the lexical roots, whereas CLs show a preference for splinters. At the same time, CLs may present also a left-hand constituent corresponding either to a full word or to a lexical root; in turn, HFs may also involve splinters and full words. For both categories, the occurrence of clippings as formative elements is a marginal tendency. Turning to the right constituent, another relevant divergence lies in the selection of the splinter *-xit*, which seems to be a highly dispreferred variant for HFs. This fact could also depend on the shape of the preceding element or on other prosodic constraints, for instance, the tendency to assume the same syllabic length of the leader word *Brexit*, which, as we have already seen, is not common for HFs. In fact, among the 43 occurrences of *-xit* in the ITr, in 35 cases (81.4%) the *-exit* word is disyllabic. Similarly, in the ITT, among the 48 occurrences of *-xit*, 39 items (81.25%) are disyllabic. Since, as already observed before, CLs tend to prefer disyllabic outputs (see Table 4), it is not surprising that the splinter *-xit* appears almost exclusively in this category of *-exit* words.

5. Conclusions

In this study, the lexical innovations formed by the productive splinter *-(e)xit* were analysed, with particular reference to their micro-diachronic trend and to their peculiar morphotactic shape. The diachronic trend of *-exit* highlights its emerging nature, yielding many *hapax* and *dis legomena*, ups and downs that could be considered a symptom of its newness in the morphological system of Contemporary Italian. At the same time, the international outbreak of *Brexit* has had its impact, leading to the formation of new types and to an increasing combination of the borrowed formative *-exit* with native elements, mostly proper nouns. The micro-diachronic trend highlighted no notable difference between the distribution of CLs and HFs formed by *-exit* (see Figures 3-4) in the ITr, except for the higher number of frequent and resistant CLs recorded. Conversely, the Anglo-Italian HFs present few forms that are frequent enough in terms of tokens and recurrent enough to persist after one or two years. Taking into account the morphotactic shape of the *-exit* words, it has been established that, in spite of the model yielded by the leader word *Brexit* (i.e. reduction to a complex onset of the left-hand constituent, disyllabic length), both samples display notable variation: many types exhibit longer syllabic structures (3-4 syllable); the left-constituent is often reduced to a splinter, but it can be also preserved fully or retain the entire lexical stem. Again, this behaviour is in line with

the distinction between CL and HF. Indeed, HFs are on average 1 syllable longer than CLs, which in turn either prefer to adopt a disyllabic output or a trisyllabic one. Furthermore, whereas CLs select a splinter and the reduced splinter *-xit* more often, HFs tend to retain the lexical root of the left-constituent and to fully preserve the *-exit* component. What can be inferred from those observations? Our hypothesis is that after the success of *Brexit*, an *-exit* word formation pattern was indeed extracted and extended on native stems (Gardani 2021). However, since such pattern was firmly anchored in socio-political events that have by this point in time already passed, it is difficult to establish whether it will remain or disappear in the future. Nonetheless, its impact on the morphological system of Italian is interesting to analyse, since it sheds light on how the speakers tried to accommodate it, preferring longer forms over shorter ones, and lexical stems over splinters. So, in the end, we are left with a question of whether this morphotactic behaviour shown by HFs could have been caused by system-dependent morphological naturalness (Dressler *et al.* 1987).

Acknowledgements

The attendance at the 14th edition of the Mediterranean Morphology Meeting was made possible by funding provided through the Young Researcher and Innovator Conference Grant of the ENEOLI COST Action (CA22126 – European Network on Lexical Innovation). I would like to thank the organisers and participants of the conference. I would like to express my gratitude to all the editors of the volume, especially Frane Malenica (University of Zadar), whose careful reading and valuable feedback have helped to improve the quality of this work.

References

- Adams, Valerie. 1973. *An Introduction to Modern English Word-Formation*. London, New York: Longman.
- Antonelli, Giuseppe. 2016. *L'italiano nella società della comunicazione 2.0*. 2nd edn. Bologna: Il Mulino.
- Bauer, Laurie & Beliaeva, Natalia & Tarasova, Elizaveta. 2019. Recalibrating productivity: Factors involved. *Zeitschrift für Wortbildung / Journal of Word Formation* 3(1). 44–80.
- Berman, J. M. 1961. Contribution on Blending. *Zeitschrift für Anglistik und Amerikanistik* 9. 278–281.
- Brasolin, Paolo & Franzini, Greta E. & Spina, Stefania. 2023. “Ti blocco perché sei un trollazzo”. Lexical Innovation in Contemporary Italian in a Large Twitter Corpus. In Boschetti, Federico, Lebani, Gianluca E., Magnini, Bernardo & Novielli, Nicole (eds.), *Proceedings of the Ninth Italian Conference on Computational Linguistics (CLiC-it 2023)*, 100–112. Venice: CEUR Workshop Proceedings.
- Buchstaller, Isabelle, Rickford, John R., Traugott, Elizabeth Closs, Wasow, Thomas & Zwicky, Arnold. 2010. The sociolinguistics of a short-lived innovation: Tracing the development of quotative all across spoken and internet newsgroup data. *Language Variation and Change* 22. 191–219.
- Caffarelli, Enzo. 2018. Brexit, Grexit, Frexit, Prexit, Crexit, Catalexit et sim. *Rivista Italiana di Onomastica* 21(2). 447–448.
- Cortelazzo, Michele A. 1988. Italienisch: Fachsprachen: Lingue speciali. In Holtus, Günter, Metzeltin, Michael & Schmitt, Christian (eds.), *Lexikon der Romanistischen Linguistik*, vol. 4, 246–255. Tübingen: Niemeyer.
- Cutler, Anne. 1981. Degrees of transparency in word formation. *Canadian Journal of Linguistics / Revue canadienne de linguistique* 26(1). 73–77.
- D'Achille, Paolo. 2003. *L'italiano contemporaneo*. Bologna: Il Mulino.
- De Mauro, Tullio. 2016. *Storia linguistica dell'Italia repubblicana dal 1946 ai nostri giorni*. Roma, Bari: Laterza.
- Dressler, Wolfgang Ulrich, Mayerthaler, Willi, Panagl, Oswald & Wurzel, Wolfgang Ullrich. 1987. *Leitmotifs in Natural Morphology*. Amsterdam, Philadelphia: John Benjamins.

- Dressler, Wolfgang Ulrich & Thornton, Anna Maria 1991. Doppie basi e binarismo nella morfologia italiana. *Rivista di Linguistica* 3(1). 3–22.
- Fradin, Bernard. 2015. Blending. In Müller, Peter O., Ohnheiser, Ingeborg, Olsen, Susan & Rainer, Franz (eds.), *Word-Formation. An International Handbook of the language of Europe*, vol. 1, 386–413. Berlin, New York: De Gruyter Mouton.
- Gardani, Francesco. 2021. On how morphology spreads. *Word Structure* 14(2). 129–147.
- Grieve, Jack & Nini, Andrea & Guo, Diansheng. 2016. Analyzing lexical emergence in Modern American English online. *English Language & Linguistics* 21. 99–127.
- Herman, Ondřej, Jakubíček, Miloš, Kraus, Jan & Suchomel, Vít. 2025. From Word of the Year to Word of the Week: Daily-updated Monitor Corpora for 25 Languages. In Kosem, Iztok, Jakubíček, Miloš, Medved', Marek, Zgaga, Karolina, Arhar Holdt, Špela, Munda, Tina & Salgado, Ana (eds.), *Proceedings of the eLex 2025. Brno, Czech Republic: Lexical Computing*, 44–61. Brno: Lexical Computing.
- Jakubíček, Miloš, Kilgarrieff, Adam, Kovář, Vojtěch, Rychlý, Pavel & Suchomel, Vít. 2013. The TenTen corpus family. In Hardie, Andrew & Love, Robbie (eds.), *7th International Corpus Linguistics Conference CL*, 125–127. Lancaster: UCREL.
- Kilgarrieff, Adam, Baisa, Vít, Bušta, Jan, Jakubíček, Miloš, Vojtěch, Kovář, Michelfeit, Jan, Rychlý, Pavel & Suchomel, Vít. 2014. The Sketch Engine: ten years on. *Lexicography* 1. 7–36.
- Lalić-Krstin, Gordana & Silaški, Nadežda. 2019. 'Don't go brexin' my heart': The ludic aspects of Brexit-induced neologisms. In Koller, Veronika, Kopf, Susanne & Miglbauer, Marlene (eds.), *Discourses of Brexit*, 222–236. London: Routledge.
- Lasserre, Marine & Montermini, Fabio. 2014. How is the meaning of complex lexemes constructed? A study of neoclassical compounds in *-cratie / -crate* and *-logie / -logue*. *Rivista di Linguistica* 26(2). 157–181.
- Lepschy, Anna Laura & Lepschy, Giulio. 2019. *La lingua italiana. Storia. Varietà dell'uso. Grammatica*. 6th edn. Firenze, Milano: Bompiani.
- Marello, Carla. 2020. New words and new forms of linguistic purism in the 21st century: The Italian debate. *International Journal of Lexicography* 33(2). 168–186.
- Massicotte, Philippe & South, Andy. 2026. *rnaturalearth: World Map Data from Natural Earth*. R package version 1.2.0. URL: <https://CRAN.R-project.org/package=rnaturalearth>.
- Mattiello, Elisa. 2018. Paradigmatic Morphology. Splinters, Combining Forms, and Secreted Affixes. *SKASE Journal of Theoretical Linguistics* 15(1). 2–22.
- Montermini, Fabio. 2018. Les affixes dérivationnels ont-ils des allomorphes? Pour une modélisation de la variation des exposants dans une morphologie à contraintes. In Bonami, Olivier, Boyé, Gilles, Dal, Georgette, Giraudo, Hélène & Namer, Fiammetta (eds.), *The lexeme in descriptive and theoretical morphology*, 423–465. Berlin: Language Science Press.
- Pebesma, Edzer. 2018. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal* 10(1). 439–446. DOI: <https://doi.org/10.32614/RJ-2018-009>.
- Pedersen, Thomas Lin. 2024. *patchwork: The Composer of Plots*. R package version 1.3.0. URL: <https://CRAN.R-project.org/package=patchwork>.
- Rainer, Franz. 2013. Formación de palabras y analogía. Aspectos diacrónicos. In Pujol Payet, Isabel (ed.), *Formación de palabras y diacronía*, 141–172. A Coruña: Universidade da Coruña, Servizo de Publicacións.
- R Core Team. 2022. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. URL: <https://www.R-project.org/>.
- Renner, Vincent. 2020. An ecosystem view of English word-formation. *The Mental Lexicon* 15(1). 4–20.
- Ronneberger-Sibold, Elke. 2006. Lexical Blends: Functionally Tuning the Transparency of Complex Words. *Folia Linguistica* 40(1/2). 155–181.
- Sayers, David. 2014. The Mediated Innovation Model: A Framework for Researching Media Influence in Language Change. *Journal of Sociolinguistics* 18(2). 185–212.
- Seifart, Frank. 2015. Direct and indirect affix borrowing. *Language* 91(3). 511–532.
- Thornton, Anna Maria. 2016. *Il genere di Brexit*. (<https://accademiadellacrusca.it/it/consulenza/il-genere-di-brexit/1097>) (Accessed 2025-01-16).

- Trampuš, Mitja & Novak, Blaž. 2012. Internals of an aggregated web news feed. *Proceedings of the Fifteenth International Information Science Conference IS SiKDD 2012*. 431–434.
- Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
- Wickham, Hadley, François, Romain, Henry, Lionel, Müller, Kirill & Vaughan, Davis. 2023. *dplyr: A Grammar of Data Manipulation*. R package version 1.1.3. URL: <https://CRAN.R-project.org/package=dplyr>.
- Wolfer, Sascha & Koplenig, Alexander. 2025. Does corpus size influence normalised frequencies? *Corpus Linguistics and Linguistic Theory*. Ahead of Publication. DOI: <https://doi.org/10.1515/cllt-2024-0040>

