

Number inflection in AN and NA Italian compounds

M. Silvia Micheli
University of Pavia – University of Bergamo
silvia.micheli@unibg.it

1. Introduction

Although number inflection of Italian compound words is generally considered to be an irregular phenomenon, it has often been neglected in the literature devoted to morphological variation. In this paper, our aim is to provide a deeper analysis of the number inflection of two types of Italian compounds, i.e. Noun-Adjective (e.g. *rocca_Nforte_A* ‘stronghold’) and Adjective-Noun (e.g. *mezza_Aluna_N* ‘half-moon’) compounds. These kinds of forms are generally considered to be compounds by grammars and dictionaries, even though they are often the result of the lexicalization of originally syntactic expressions (e.g. *camposanto* ‘cemetery’, lit. ‘holy field’). As in many other cases, the boundary between syntax and compounding is not so clear. With regard to number inflection, these forms are very interesting since they display both cases of double inflection (e.g. *casse_{PL}forti_{PL}* ‘safes’, *doppi_{PL}vetri_{PL}* ‘double glasses’), in which there are two plural markers, and cases of external inflection (e.g. *rocca_{SG}forti_{PL}* ‘strongholds’, *doppio_{SG}petti_{PL}* ‘double-breasted’), in which the plural marker of the word is in the canonical position, namely on the right. Furthermore, in some cases the same compound shows both kinds of inflection: e.g. *caposaldo* ‘cornerstone’ (pl. *capo_{SG}saldi_{PL}* – *capi_{PL}saldi_{PL}*) or *terracotta* ‘terracotta’ (pl. *terra_{SG}cotte_{PL}* – *terre_{PL}cotte_{PL}*). These cases can be considered as instances of ‘overabundance’, i.e. “the situation in which there are two or more forms available to realize the same cell in an inflectional paradigm” (Thornton 2012: 183).

This work aims at describing this twofold variation through the analysis of both quantitative and qualitative data from a corpus of Contemporary Italian, i.e. itWaC (Baroni *et al.* 2009). By investigating the number inflection of AN and NA Italian compounds, we hope to enrich the description of these forms which seem to be halfway between syntax and morphology.

The paper is organized as follows. In sections 2 and 3, we provide a general introduction on Italian compounding and the plural inflection of compound words. Section 4 is devoted to the data retrieval and the parameters of the analysis. In section 5, we illustrate the results of the analysis of AN and NA compounds, focusing on the most relevant factors which can impact their number inflection. Finally, in section 6 we sum up our findings and provide some theoretical remarks on compounding in Italian.

2. AN and NA Italian compounds

Compounding in Italian (as in other Romance languages) is generally considered not to be very productive,¹ especially compared to derivation, which represents the most productive word formation mechanism throughout the history of the Italian language.

¹ For an overview of Italian compounding see Masini and Scalise (2012).

As in other Romance languages, the most productive compound type is the Verb Noun pattern (e.g. *portachiavi* ‘keyring’, *giradischi* ‘record player’), which represents also the most accurately described pattern from both a synchronic and a diachronic point of view (Ricca 2010, 2005; Štichauer 2015a-b). Except for this pattern, which is arguably the result of a purely morphological mechanism, for the other main compound types (i.e. AN, NA, NN, AA) the boundary between word formation and syntax is problematic.² As far as compounds involving adjectives are concerned, i.e. AN and NA, it is often hard to distinguish between compounds (1) and phrasal lexemes (2).³

- | | | |
|-----|---|--------------------|
| (1) | acquasanta
water:F.SG-holy:F.SG
‘holy water’ | [NA] _N |
| (2) | tessera sanitaria
card:F.SG health.F.SG
‘health card’ | [NA] _{NP} |

Although these two forms seem very similar from a formal point of view, the former is generally considered by Italian dictionaries to be a compound, whereas the latter is seen as a phrasal lexeme. Unlike languages such as Russian (Benigni and Masini 2010), in which compounds and phrasal lexemes can be distinguished from a formal point of view, since the former display a bound form as first constituent whereas the latter display an autonomous word, in Italian both compounds and phrasal lexemes are made up of two juxtaposed independent words.

The criteria which have been proposed in order to distinguish between compounds and phrasal lexemes vary from language to language.⁴ As shown in Booij (2009), the agreement inflection on the adjective represents a criterion for distinguishing AN compounds and AN phrasal lexemes in Dutch from a formal point of view. As already pointed out by Masini and Scalise (2012: 74), the agreement criterion in Italian is not as efficient as in Dutch, since agreement (at least gender agreement) appears in both compounds and in phrasal lexemes.

3. Number inflection in Italian compound words

Unlike other languages such as English, the plural marker in Italian compounds appears not only in its canonical position (3), namely on the right (as in simple words: sg. *cane* ‘dog’ – pl. *cani* ‘dogs’), but also inside the compound word (4) or in both constituents (5):

² They represent problematic phenomena if we assume that morphology and syntax are two distinct modules of grammar. Because of their in-between status, this kind of forms (for Germanic languages see Schlücker and Hüning 2009; Hüning 2010) has been effectively analyzed within a framework that treats lexicon and grammar (as well as word formation and syntax) as a continuum, i.e. Construction Grammar (Goldberg 2003). In this approach, grammar is seen as an inventory of constructions, i.e. conventionalized pairings of form and meaning components. Thus, both phrases and compounds can be considered as constructions.

³ See Booij (2009) for a cross-linguistic overview on phrasal lexemes. Masini (2012, 2009) for Italian.

⁴ The criteria which are generally adopted for distinguishing between compounds and phrasal lexemes in Italian are essentially three: 1) internal agreement; 2) the presence of a conjunction or a preposition, which are typically used in syntax, between the two constituents; 3) the presence of elements from other minor lexical categories. See Masini and Scalise (2012: 73) for a more detailed discussion of these parameters.

- | | | |
|-----|---|--|
| (3) | banconota
bank:SG-note:SG
'banknote' | banconote
bank:SG-note:PL
'banknotes' |
| (4) | capostazione
chief:SG-station:SG
'station master' | capistazione
chief:PL-station:SG
'station masters' |
| (5) | cassaforte
box:SG-strong:SG
'safe' | casseforti
box:PL-strong:PL
'safes' |

Thus, Italian compound words display three kinds of inflection, i.e. internal, external and double. The first case is due to the fact that inflection generally operates on the head of the compound, which is usually the left constituent in Italian. As already pointed out by Gaeta (2011: 79-80), the internal inflection represents an anomaly within the Italian inflectional system, as in simple and derived words the plural marker is always external. Diachronically, such an anomaly can be solved by means of the so-called “externalization of inflection” (Haspelmath 1994), through which the internal plural marker is moved to the right constituent, as shown in Figure 1, which is produced by Google books Ngram Viewer by querying for *pomodoro* ‘tomatoes’ (apple:PL-golden:SG) and *pomodori* (apple:SG-golden:PL)

Figure 1: Number inflection of *pomodoro* ‘tomato’ on Google Ngram Viewers corpus



Externalization of the inflection goes together with univerbation and loss of transparency of the internal structure of the compound word, which is interpreted (and inflected) as a simple word. This process affects especially very frequent and no longer productive compounds.

As shown in Micheli (2016a), both formal (i.e. syntactic category and gender of the constituents, head position) and semantic (i.e. semantic transparency of the constituents, semantic relation between them) factors are involved in the inflection of compound words.

Although these factors cannot be easily organized in a hierarchy, semantic factors seem to be more relevant. However, while formal factors can be analysed from both a qualitative and a quantitative point of view, semantic factors (especially semantic transparency) cannot be measured and are crucially dependent on the speakers' interpretation.

In Micheli (2016b), the analysis of the inflection of compounds with *capo-* 'chief' as first constituent shows that they have different inflections depending on the position of the head and the different meanings conveyed by *capo-*.⁵ When *capo-* means 'chief' and represents the head of a subordinative compound (e.g. *capostazione* 'stationmaster'), the inflection is almost always internal, even in very frequent compounds. When *capo-* means 'main', as in *capolavoro* 'masterpiece' or *capoluogo* 'principal town', the inflection is always external, since in these cases the head is on the right. In coordinative compounds meaning 'x is the chief among $x_1, x_2, x_3 \dots$ ' (e.g. *capocuoco* 'head chef'), the inflection can be both double and external. In some cases, the same compound has two plural forms. For instance, overabundance is shown by *caporedattore* 'editor-in-chief' which has both double inflection (*capiredattori*, 72 occurrences on itWaC corpus) and external inflection (*caporedattori*, 64 occurrences on itWaC corpus).

There are many other exceptions which are pointed out by grammarians but not sufficiently well examined. In this paper, we will investigate two case studies, in order to enrich the description and the explanation of this case of morphological variation. The first case concerns AN compounds, e.g. compounds made up of *alto-* 'high' and *basso-* 'low' as first constituents, which present both double and external inflection. Such a scenario is also shown by NA compounds, e.g. compounds made up of *-forte* 'strong' as second constituent.

4. Methodology

The data for this study have been manually retrieved from a Contemporary Italian dictionary, i.e. *Devoto Oli* (2015), and from the wordlist of a reference corpus, i.e. the *New Basic Vocabulary* corpus (Chiari and De Mauro 2014; henceforth NVDB).⁶ The sample was made up of 83 NA compounds and 150 AN compounds. However, compounds in which the first constituent lacks the inflectional marker because of elision or apocope (e.g. *buonsenso* 'common sense'), have not been taken into account in the analysis. Thus, the final sample was made up of 50 NA compounds and 90 AN compounds. Since compound words in Italian are not very frequent, we have retrieved quantitative data from a large web corpus of Contemporary Italian, i.e. itWaC. In order to detect all kinds of inflection, we have checked each compound in all possible plural forms, e.g. in the case of *altopiano* 'upland', itWaC has been queried for *altopiani* and *altipiani*. Furthermore, since compound words in Italian can occur in three orthographic variants (i.e. juxtaposed, unverbated and hyphenated), each plural form has been checked in three different spellings, e.g. *alto piani*, *altopiani* and *alto-piani*.

We have grouped compounds into three groups, depending on whether they show double inflection, external inflection or both. Each compound has been analyzed from both a quantitative and a qualitative point of view. The parameters which have been taken into consideration are the following. From a quantitative standpoint, we have focused on whole-compound frequency (in both singular and plural forms). From a qualitative point of view, we have considered whether each compound shows a compositional or a non-compositional meaning. From a diachronic point of view, we have investigated whether compounds are

⁵ Here we have adopted the semantic classification of *capo-* compounds provided by Serianni (1989: 153).

⁶ The corpus is not yet freely available. I would like to thank Isabella Chiari for allowing me to use it for this study.

calques from other languages and when they have been attested for the first time. Furthermore, we have considered whether each compound belongs to a series with a lexically specified constituent.

5. Results

5.1 NA compounds

NA compounds are not very productive in Contemporary Italian. Since they follow the constituent order which is canonical for syntax, it is often hard to distinguish them from phrasal lexemes or phrases. However, it is possible to identify some groups of compounds (i.e. series) which share one constituent, even though they are often not still productive in Contemporary Italian. As an instance, we can think of compounds made up of *-forte* as second constituent (e.g. *roccaforte* ‘stronghold’ or *cassaforte* ‘safe’) or *acqua-* as first constituent (e.g. *acquasanta* ‘holy water’ or *acquacotta* ‘soup’, lit. ‘water+cooked’). In the former case, compounds with *-forte* can be considered as ‘constructional idioms’ (in the sense of Booij 2009) in which one position is lexically specified and shows a specific meaning; i.e. in these compounds *-forte* means ‘impregnable’. Thus, they seem to be closer to morphology than to syntax. However, except for these and very few other cases, NA compounds are isolated forms which do not belong to any series: they originate in syntax and, in some cases, undergo a lexicalization process, with a meaning which is not completely compositional anymore. For instance, *bancarotta* ‘bankrupt’ originates at the syntactic level (originally, it referred to a failed bank whose stall was broken) and shows a lexicalized meaning, but it does not form any series.

As regards the number inflection, NA compounds show overabundance in 22 cases, as shown in Table 1.

Table 1: Distribution of NA compounds with regard to plural type

Double inflection	External inflection	Overabundance	Total
22	6	22	50
44%	12%	44%	100%

Most of the NA compounds that show two kinds of plural are feminine (i.e. 72%) and belong to the inflectional class *-a/-e* (50%). They are native compounds, except for two forms which are calques from French (i.e. *cassaforte* ‘safe’ from *coffre-fort* and *manomorta* ‘mortmain’ from *mainmorte*). Table 2 illustrates the frequency of the different cell-mates of some NA compounds which show overabundance. Following Thornton (2012) and Fehringer (2004), we have established the strength of overabundance taking as a measure the ratio between tokens in two cell-mates. Following Thornton (2012: 189), we have also assumed that in a given synchronic corpus ratios in the range of units indicate existence of overabundance, ratios in the range of tens indicate that overabundance is on the verge of extinction or has recently ceased to exist, and higher ratios indicate complete extinction of overabundance in the synchronic stage represented by the corpus.

Table 2: NA compounds which show overabundance

Compound	External inflection	Double inflection	Ratio external infl. : double infl.	Freq. of the citation form	First occurrence	Calque
<i>acquamarina</i> water+marine 'aquamarine'	8	23	1 : 2,8	249	1571	no
<i>acquatinta</i> water+dyed 'engraving technique'	7	56	1 : 8	204	1875	no
<i>camposanto</i> field+holy 'cemetery'	44	9	4,9 : 1	722	XIV cent.	no
<i>cartapesta</i> paper+crushed 'papier-mâché'	10	2	5 : 1	2078	XVII cent.	no
<i>cassaforte</i> box+strong 'safe'	72	547	1 : 7,6	4920	1866	yes
<i>manomorta</i> hand+dead 'mortmain'	17	35	1 : 2	244	1676	yes
<i>piazzaforte</i> square+strong 'fortified town'	13	122	1 : 9,4	503	1551	no
<i>roccaforte</i> hold+strong 'stronghold'	693	81	8,5 : 1	3049	XIII cent.	no
<i>terracotta</i> earth+fired 'terracotta'	143	639	1 : 4,5	4705	1342	no
<i>terraferma</i> land+still 'dryland'	1	10	1 : 10	3962	XIII cent.	no

Table 2 shows NA compounds in which there is overabundance in Contemporary Italian. The highest strength of overabundance (ratio < 5) is shown by *acquamarina* 'aquamarine', *manomorta* 'mortmain', *camposanto* 'cemetery' and *terracotta* 'terracotta'. Thornton (2012: 191-197) discusses the following factors which help to preserve overabundance: 1) the fact that overabundant cells belong to an independently established morphomic pattern; 2) the existence of particular constructions that require a given form; 3) the fact the forms in overabundant cells are 'learned late' by speakers; 4) frequency. The first two cannot explain the maintenance of overabundance in these Italian compounds. On the other hand, the factor of late learning can be invoked for these NA compounds (except for *camposanto* and *cassaforte*), as they do not belong to the New Basic Vocabulary and often are used in specialized varieties. If we analyze the Italian data for the last parameter, i.e. frequency, we can observe that it can explain the maintenance of overabundance of *acquamarina*, *camposanto*, *manomorta*, as these compounds have a low overall lexeme frequency. However, it cannot explain the overabundance of *terracotta*, which has the second highest frequency but does not have the highest ratio between the two cell-mates in its overabundant cells. From a diachronic point of view, it is interesting to note that this case of overabundance

is stable over time: as shown in Figure 2, which is produced by Google books Ngram Viewer by querying for *terrecotte* and *terracotte* in the Italian corpus, the two plural forms coexisted permanently during the last two centuries of the history of written Italian. Also, it is worth noting that compounds which belong to a series, i.e. compounds that share *-forte* as second constituent, show overabundance. It appears from the data examined until now that none of these factors can explain the maintenance of overabundance on its own; instead, each case seems to need an *ad hoc* explanation.

Figure 2: Number inflection of *terracotta* 'terracotta' on Google Ngram Viewers corpus



As regards the NA compounds that always show external inflection (Table 3), we can observe that they are isolated and lexicalized. In two cases, i.e. *bancarotta* ‘bankrupt’ and *terrapieno* ‘embarkment’, they also show high frequency. In the latter case, the external inflection can be due to two facts, i.e. it is a calque from the right-headed Medieval Latin compound *terrapienum* and the two constituents have different genders.

Table 3: NA compounds which show external inflection

Compounds	Double infl.	External infl.	Frequency of the citation form	First occ.	Calque
bancarotta bank+broken ‘bankrupt’	0	149	4744	XVI cent.	no
terrapieno earth+full ‘embarkment’	0	334	1102	XVI cent.	yes
gattabuia tunnel+dark ‘clink’	0	3	206	1850	no
ceramolle wax+soft ‘wax’	0	1	20	XIX cent.	no
beccofrusone beak+cone-shaped ‘waxwing’	0	11	21	XIX cent.	no

The remaining NA compounds (around 44%) show double inflection. This makes it very hard to distinguish them from phrasal lexemes or noun phrases. Around one half (i.e. 52%) is made up of isolated forms, while the other half is made up of compounds which belong to a series that is no longer productive, i.e. compounds which have *acqua-* (e.g. *acquamarina* ‘aquamarine’) or *carta-* (e.g. *cartavelina* ‘tissue paper’) as first constituent. The data seem to suggest that when NA compounds are part of a series, even though it is no longer productive, they tend to form the plural by inflecting both constituents. However, this represents merely a tendency and not a systematic behavior, as shown by the compound series with *-forte* as second constituent, whose members present overabundance. Further, it is worth noting that the time of first occurrence of these compounds does not seem to affect their inflection, since they are all rather ancient words but show different behaviors.

5.2 AN compounds

AN compounds can be considered as closer to morphological structures than NA compounds, as they show a marked constituent order and, in most cases (around 83%), belong to series. Table 3 illustrates the size of these series,⁷ i.e. the number of compounds which share a common first constituent. It is worth noting that, whereas series of NA compounds may share both first and second constituent, the shared constituent in AN compounds series is always the adjective.

Table 4: AN compound series

First constituents	Num. of compounds which share the first constituent	Examples
alto- ‘high’	6	altopiano high+plain ‘plateau’
basso- ‘low’	5	bassorilievo low+relief ‘bas-relief’
buono- ‘good’	6	buonafede good+faith ‘good faith’
doppio- ‘double’	7	doppiopetto double+chest ‘double-breasted jacket’
falso- ‘false’	4	falsopiano false+plain ‘slight slope’
madre- ‘mother’	3	madreperla mother+pearl ‘mother-of-pearl’
malo- ‘bad’	21	malafede bad+faith ‘bad faith’
mezzo- ‘half’	23	mezzobusto half+bust ‘bust’

⁷ We have considered as ‘series’ the compound groups which contain at least 3 compounds sharing a common constituent.

Except for *malo-* ‘bad’, each first constituent reported in Table 4 is attested as a free lexeme and belongs to the New Basic Vocabulary of Italian. In Contemporary Italian, *malo-* (especially in its feminine form *mala*) occurs only as first constituent of compounds;⁸ as shown in Micheli (in prep.), it can be considered as an affixoid, in the sense of Booij and Hüning (2014).

Table 5 : Distribution of AN compounds with respect to plural type

Double inflection	External inflection	Overabundance	Total
45	14	31	90
50%	15,5%	34,4%	100%

As shown in Table 5, also in number inflection of AN compounds there are many cases of overabundance, though within this compound category the double inflection represents the most frequent plural type. The fourth column of Table 6 illustrates to what extent overabundance is spread among the series examined in Table 4.

Table 6: Number inflection in AN series

Series	Double inflection	External inflection	Overabundance	Total
alto-	1 16,6%	1 16,6%	4 66,7%	6
basso-	2 40%	0	3 60%	5
buono-	4 66,7%	0	2 33,3%	6
doppio-	4 57%	0	3 43%	7
falso-	2 50%	0	2 50%	4
madre-	0	1 33,3%	2 66,7%	3
malo-	13 62%	1 5%	7 33%	21
mezzo-	15 65%	3 13%	5 22%	23

It appears from the data in Table 6 that overabundance is frequent in AN compounds with *alto-* and *basso-* as first constituent. These compounds show irregular behavior also in the singular form: in analogy to AN Latin compounds,⁹ their first constituent sometimes ends in the vowel *-i* (i.e. *altipiano* or *bassipiano*), which has the same spelling as the plural marker for masculine nouns.

⁸ The construction [*malo-* ‘bad’ + Noun] can be considered quite productive in Contemporary Italian, as shown by many neologisms such as *mala-sanità* (lit. ‘bad health service’), *mala-finanza* (lit. ‘bad finance’), etc.

⁹ In Latin compounding, the vowel *-i* functions as the most common linking element between the two constituents. Following Ralli (2006), it can be considered as a compound marker that identifies the compounding process. According to Oniga (1992), it is the result of the phonetic change of the thematic vowel of the first constituent.

On the other hand, it is worth noting that the two largest series, i.e. compounds with *malo-* and *mezzo-*, show a strong tendency to maintain double inflection. Furthermore, we can observe that in AN compounds that form a series external inflection is very exceptional.

Instead, if we analyze AN compounds which do not belong to a series, we notice that they frequently show external inflection.

Table 7: Number inflection of isolated AN compounds

Double inflection	External inflection	Overabundance	Total
4 28,6%	7 50%	3 21,4%	14
albogatto (1) biancospino (617) drittofilo (4) primadonna (694)	bellosguardo (1) cortometraggio (9.878) mediocredito (34) mediometraggio (277) piattabanda (94) piattaforma (54.441) vanagloria (612)	biancofiore (53) cortocircuito (3.016) primogenitore (9)	

Table 7 illustrates the distribution of isolated AN compounds with regard to the plural type. For each plural type, compounds with their frequency in brackets (i.e. frequency of the citation form on itWaC corpus) are reported. It should be noted that *piattaforma* ‘platform’ and *piattabanda* ‘lintel’ come from French (i.e. *plateforme* and *plate-bande*) and *vanagloria* ‘boastfulness’ originates from Latin (i.e. *vanagloria*): they are probably quite opaque to speakers and thus inflected as simple words. For *cortometraggio* ‘short film’ and *piattaforma*, frequency seems to be the most important factor in explaining the external inflection. As far as compounds with an adjective belonging to the *-o* inflectional class are concerned (i.e. singular masculine; e.g. *medio* ‘middle’ or *corto* ‘short’), it can be noticed that this *-o* is homophonous to the linking element in AA compounds (e.g. *italo-americano* ‘Italian-American’), as shown in Grossmann and Rainer (2009) and D’Achille and Grossmann (2009). Since the number inflection in these compounds is always external, we can assume that *cortometraggio*, *mediometraggio* ‘medium-length film’ and *mediocredito* ‘medium-term credit’ show external inflection in analogy with AA constructions.

6. Discussion

Based on descriptions provided in the previous sections, in what follows we will discuss our findings on number inflection in NA and AN compounds and some theoretical issues with regard to their status.

NA compounds do not show a clear tendency, as they present many cases of both double inflection (44%) and overabundance (44%). External inflection is displayed only by fully lexicalized and frequent compounds (e.g. *bancarotta*). Frequency seems to be crucial for both external inflection and overabundance, as very frequent words tend to be interpreted as simple words and therefore to be inflected as such.¹⁰ On the other hand, compounds which are part of a series tend to maintain double inflection. This unclear picture can be accounted for by the fact that this type of compounds is very heterogeneous, as it contains both lexicalized units (e.g. *bancarotta* or *cassaforte*) and combinations of a noun and an adjective which show a (at least partially) compositional meaning (e.g. *camposanto*, *terracotta*). The data seem to

¹⁰ The link between high frequency and loss of semantic transparency has been shown by several psycholinguistic experiments (specifically for Italian compounds, see, among others, Marelli and Luzzatti 2012).

suggest that the combination of a noun and an adjective in Italian is productively exploited only to create phrasal lexemes, which in some cases, due to high frequency, lose their semantic transparency, increase in cohesion from a formal point of view and show word-like behavior (e.g. they are graphically unverbated and display external inflection). Thus, it seems that a truly morphological (and productive) mechanism for creating NA compounds does not exist in Contemporary Italian.

AN compounds show a more regular behavior in number inflection. They often belong to a series and form the plural by inflecting both constituents. The more productive the series is, the stronger the tendency towards double inflection, as we saw for compounds with *malo-* and *mezzo-*. Thus, double inflection seems to be a feature that is codified within these constructions.

According to the data analyzed in this study, we claim that, as far as Italian is concerned, number inflection does not represent an effective criterion for distinguishing between AN and NA compounds and phrasal lexemes (or phrases). On the other hand, double inflection seems to be a characteristic of productive series of AN compounds. We therefore suggest that, although in most cases each compound seems to need an *ad hoc* explanation, frequency (especially for NA compounds) and belonging to a series (for AN compounds) are the most relevant factors for determining inflection.

7. Conclusions

In this paper, we have investigated the number inflection of AN and NA compounds, by basing our analysis on quantitative data from a corpus of Contemporary Italian, i.e. itWaC. We have shown that NA compounds are mostly ancient and infrequent forms which originate in syntax and, in some cases, undergo lexicalization. They are probably quite opaque to speakers and interpreted as simple words. Thus, this group of forms is made up of lexical units which are not the result of a productive morphological mechanism, but rather a repository of isolated lexicalized combinations of a noun and an adjective, which do not show consistent behavior in number inflection. On the other hand, AN compounds represent a more morphological pattern mostly consisting of (sometimes still productive) series which still display a transparent internal structure and a strong tendency for double inflection.

References

- Baroni, M., S. Bernardini, A. Ferraresi, & E. Zanchetta (2009) The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language resources and evaluation* 43(3): 209-226.
- Benigni, V. & F. Masini (2009) Compounds in Russian. *Lingue e linguaggio* 8(2): 171-194.
- Booij, G., & M. Hüning (2014) Affixoids and constructional idioms. In: R. Boogaart, T. Coleman & G. Rutten, (Eds.), *Extending the scope of Construction Grammar*. Berlin, Boston: De Gruyter Mouton, 77-105.
- Booij, G. (2009) Construction morphology and Compounding. In: R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of compounding*. Oxford: Oxford University Press, 201-216.
- Chiari, I. & T. De Mauro (2012) The new basic vocabulary of Italian: problems and methods. *Rivista di statistica applicata/Italian Journal of Applied Statistics* 22(1): 21-35.
- D'Achille, P. & M. Grossmann (2009) I composti aggettivo + aggettivo in italiano. In: M. Iliescu, H. Siller & P. Danler (eds.), *Actes du XXVe Congrès International de Linguistique et Philologie Romanes (3-8 sept. 2007, Innsbruck)*. Tübingen: Niemeyer, 405-413.
- Devoto G., G. C. Oli (2014) *Il Devoto Oli. Vocabolario della lingua italiana*, Luca Serianni e M. Trifone (Eds.), Firenze: Le Monnier.
- Fehrer, C. (2004) How Stable are Morphological Doublets? A Case Study of/[schwa]/~ Ø Variants in Dutch and German. *Journal of Germanic linguistics* 16(4): 285-329.

- Gaeta L. (2011) La composizione in italiano e oltre. In: Aa.Vv. (Eds.), *Scuola di formazione di italiano lingua seconda/straniera: competenze d'uso e integrazione*. Napoli: Edizioni Scientifiche Italiane, 73-92.
- Goldberg, A. (2003) Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences* 7 (5): 219-224.
- Grossmann, M., & F. Rainer (2009) Italian adjective-adjective compounds: Between morphology and syntax. *Italian Journal of Linguistics* 21(1): 71-96.
- Haspelmath, M. (1993) The diachronic externalization of inflection. *Linguistics* 31(2): 279-310.
- Hüning, M. (2010) Adjective+Noun constructions between syntax and word formation in Dutch and German. In: A. Onysko & S. Michel (Eds.), *Cognitive perspectives on word formation*. Berlin, Boston: De Gruyter Mouton, 195-215.
- Marelli, M., & C. Luzzatti (2012) Frequency effects in the processing of Italian nominal compounds: Modulation of headedness and semantic transparency. *Journal of Memory and Language* 66(4): 644-664.
- Masini, F. & S. Scalise (2012). Italian compounds. *Probus* 24: 61-91.
- Masini, F. (2012) *Parole sintagmatiche in italiano*. Bologna: Caissa Italia.
- Masini, F. (2009) Phrasal lexemes, compounds and phrases: a constructionist perspective. *Word Structure* 2(2): 254 -271.
- Micheli, M. S. (in prep.) Mal-constructions at work in Italian.
- Micheli, M. S. (2016a) Sul plurale delle parole composte nell'italiano contemporaneo. *Studi di lessicografia italiana* 33: 227-255.
- Micheli, M. S. (2016b) Limiti e potenzialità dell'uso di dati empirici in lessicografia. Il caso del plurale delle parole composte. *RiCOGNIZIONI. Rivista di Lingue e Letterature straniere e Culture moderne* 3(6): 15-33.
- Oniga, R. (1992) Compounding in Latin. *Rivista di linguistica* 4(1): 97-116.
- Ralli, A. (2006) Variation in word formation: the case of compound markers. In: P. Cuzzolin, M. Napoli (Eds.), *Fonologia e tipologia lessicale nella storia della lingua greca*, Milano: Franco Angeli, 238-264.
- Ricca, D. (2010) Corpus data and theoretical implications. With special reference to Italian VN compounds. In: S. Scalise & I. Vogel (Eds.), *Cross-disciplinary issues in compounding*. Amsterdam: John Benjamins, 311: 237-254.
- Ricca, D. (2005) *Al limite tra sintassi e morfologia: i composti aggettivali V-N nell'italiano contemporaneo*. In: M. Grossmann & A. M. Thornton (Eds.), *La formazione delle parole*. Roma: Bulzoni, 465-486.
- Serianni, L. (1991) *Grammatica italiana*. Torino: Utet.
- Schlücker, B. & M. Hüning (2009) Compounds and phrases. A functional comparison between German A+N compounds and corresponding phrases. *Italian Journal of Linguistics/Rivista di Linguistica* 21(1): 209-234.
- Štichauer, P. (2015a) *La formazione delle parole in diacronia: studi di morfologia derivazionale dell'italiano tra il Cinquecento e l'Ottocento*. Charles University in Prague: Karolinum Press.
- Štichauer, P. (2015b) The meaning of Italian VN compounds in a diachronic perspective: the development of a word-formation rule and its semantic instruction. In: J. Audring, N. Koutsoukos, F. Masini, I. Raffaelli (Eds.), *Morphology and Semantics. Proceedings of the 9th Mediterranean Morphology Meeting*, 138-142.
- Thornton, A. M. (2012) Reduction and maintenance of overabundance. A case study on Italian verb paradigms. *Word Structure* 5(2): 183-207.