An instance of productive overabundance: the plural of some Italian VN compounds

(Article begins on next page)

23 April 2024

An instance of productive overabundance:
The plural of some Italian VN compounds

Matteo Pellegrini (Bergamo) and Davide Ricca (Torino)

Abstract

This paper investigates a case of overabundance in the plural cell of an open subclass of Italian VN compounds. The empirical basis includes: (i) a 163-item list of relevant compounds, for which the relative frequency of cell mates has been estimated by means of Web data; (ii) a naming questionnaire based on visual input, with 30 images submitted to about 200 informants, including those of several objects whose names are scarcely established in the lexicon; (iii) a further questionnaire, adapted to each informant, asking for acceptability judgements to detect overabundance at the single speaker's level. Results show that the given subclass of VN compounds provides an instance of systematic and productive overabundance in the Italian morphological system, differently from the examples usually discussed for this language.

Keywords: Italian, Verb-Noun compounds, overabundance, productivity, Web linguistics

1. Introduction

The phenomenon of overabundance (i.e., multiple filling of the same cell in a paradigm) has been the object of much recent work, especially by Thornton (e.g., 2011, 2012, in press). In particular, proposals have been made to locate overabundance within the framework of Canonical Typology. In this perspective, canonical instances of overabundance should be the most idiosyncratic, unpredictable, isolated ones.

In this contribution[1], a different perspective will be taken. In §2 we argue for the theoretical relevance of also detecting and taking into account instances of systematic and productive overabundance, and in the following sections we investigate such a case in detail for Italian, namely the plural cell of an open subclass of VN compounds (of the type *copriletto* 'bed-cover', whose plural can be both *copriletto* and *coprileti*), which is definable by semantic criteria. A semantic-driven classification of all Italian VN compounds, according to the different options in the number marking of N displayed in both their singular and plural cells, is outlined and discussed in §3. In §4, after some methodological discussion of the reliability of Web-driven investigations, data from an extensive Web-based compound list are presented, which show that for the abovementioned class of VN compounds overabundance is highly attested (for about 50% of the types, with over 16% of them displaying a very balanced ratio between the cell mates). Moreover, overabundance appears to be equally widespread among both well-entrenched and low frequency items, suggesting that it is productive.

The issue of productivity is further explored in §5, in which we also investigate very rare items as approximating an online process of word formation, by means of a naming questionnaire based on visual input, and we partially tackle the difficult question of single speakers' competence, by

---

[1] The whole paper is the result of the close collaboration of both authors. However, for academic purposes, D.R. is responsible for §§1,2,3,6, and M.P. for §§4,5. The main contents of the work were presented at the 11th Mediterranean Morphology Meeting, held in Nicosia, Cyprus, on June 22-25, 2017.

submitting a second questionnaire asking for acceptability judgements and adapted to each informant depending on his/her answers to the first one.

2. Which canonicity for a non-canonical phenomenon?

The term "overabundance" may have only recently caught on in the general morphological literature (cf. Thornton 2011), but the phenomenon is surely quite widespread, as can easily be seen in the wealth of instances given in Thornton (in press), which refer to a mere handful of European languages.

As is now well known, we can define overabundance as "the situation in which two or more inflectional forms are available to realize the same cell in an inflectional paradigm" (Thornton in press). Within the framework of Canonical Typology (Corbett 2005; Brown, Chumakina & Corbett 2013 among others), overabundance, as a property of a given paradigm, is inherently non-canonical, because it violates the principle of "uniqueness of realization" or "univocality" (Thornton 2011: 361, Thornton in press): "every cell in a canonical paradigm contains only one form".

Thornton (in press), elaborating on Thornton (2011), is the most extensive attempt to characterize overabundance itself according to a scale of canonicity, following the approach applied by Corbett (2007: 10–11) to suppletion – another non-canonical phenomenon within inflectional paradigms. In this perspective, the "best", i.e. canonical, instances of an inherently non-canonical phenomenon should be the most irregular (formally idiosyncratic, unpredictable, isolated) ones. Consequently, Thornton proposes at least four independent criteria to locate instances of overabundance along a scale of decreasing canonicity:

> a. number of cells involved (1 > many > all);
> b. number of lexemes involved (1 > many > all);
> c. frequency ratio between the cell mates (best case: 1:1 ratio);
> d. relevance of conditioning factors (unconditioned > conditioned: conditions can be of diatopic, sociolinguistic, pragmatic, grammatical or contextual character)[2].

Criteria a.-d. are obviously very different in nature. We have no quarrel at all with c. and d. It makes perfect sense that a prototypical instance of overabundance should display a maximum of interchangeability between the cell mates, which means a minimum of internal and external linguistic factors conditioning the choice (criterion d.), the absence of which should reflect itself in the highest possible balancing between the options (criterion c.). The greater the interchangeability, the greater the unpredictability of the choice for the speakers.

In our view, criteria a. and b. are more problematic. They undoubtedly reflect the fact that the maximum of irregularity for the paradigm of a *single* lexeme is reached when a form is maximally unpredictable. And only a fully isolated instance of overabundance – both within the paradigm (criterion a.) and at the lexeme level (criterion b.) – rules out completely the possibility of detecting some kind of sub-regularity: any amount of systematicity reduces unpredictability.

However, from a different point of view, we can argue for exactly the opposite: namely, that systematic overabundance is much more "damaging" for the *global* regularity/canonicity of the inflectional system of a given language or word class. Recently studied and quoted examples of systematically overabundant cells include: GEN.SG of about 100 and LOC.SG of about 400 Czech nouns belonging to the same inflectional class (Bermel & Knittl 2012); INSTR.SG in some (phonetically and/or morphologically identifiable) subclasses of Croatian first declension nouns

---

[2] The precise labels for the different types of conditions given in Thornton (in press) are different, but they cover more or less the same – very extensive – spectrum mentioned here. For the purposes of this paper, it does not seem useful to go into more detail.

(Lečić 2015); one dual and three singular cells in the paradigm of neuter *i*-stem adjectives in Sanskrit (Stump 2016: 149).

At least in the two Slavic cases given above, it is clear that the overabundant lexemes still represent a minority within the inflectional class involved: this seems to add further complexity to the speakers' task, especially if one takes into account that this minority does not constitute a closed class (some borrowings and foreign names are included).

The latter point inserts a further relevant factor into the picture, namely the productivity of the overabundance-generating process. This seems a very interesting factor to evaluate, since from a diachronic perspective, overabundance should be commonplace – probably even unavoidable at least at the community level – as a transitional phenomenon: every instance of analogical change should entail a phase of coexistence/competition of the older and the innovative form in the same cell, at least within the speakers' community and/or in speakers' passive competence. But generally these transitional instances of overabundance are not expected to display productivity, even when they have systematic character: so one is not allowed *a priori* to equate systematicity with productivity. In Czech, the *-u* ending of the genitive comes from the inflectional class of the (historical) *-u-* stems and has extended to most, but not (yet?) all, members of the *-o-* stem class, previously marked by *-a* in the genitive. Thus the *-a/-u* doublets could be seen, in principle, as a nucleus of partial resistance to the change, together with the real stronghold of the lexemes keeping the older ending *-a* exclusively. However, this cannot be the whole truth, because among the *-a/-u* items attested in corpora Bermel & Knittl (2012) find clearly recent entries, like *blackjack, cadillac, trabant* etc. This points to a productive component in Czech *-a/-u* overabundance, and suggests the phenomenon holds a greater relevance for the morphological system as a whole.

An extreme case is probably given by the Spanish Past Subjunctive (1SG *hubiera/hubiese* etc, cf. Stump 2016: 151), whose overabundance is maximally systematic (displayed by all verbs) and fully productive, being automatically extended to all new entries in the mental lexicon. In Thornton's approach, this kind of overabundance would be ranked very low according to a canonicity scale, but on the other hand its significance for the non-canonicity of the Spanish morphological system looks much greater compared to that of, e.g., the isolated Italian lexeme (in just two cells) *devo/debbo* '(I) must', *devono/debbono* '(they) must'. It is also true, however, that the Spanish case is *so* systematic that it reduces unpredictability in a way that does not apply to Czech.

Finally, systematic/productive overabundance probably has a greater chance of being uniformly transmitted between generations, and thus of being part of (most) speakers' active competence, rather than simply being present in the linguistic community. This distinction is very important in characterizing the phenomenon, as Thornton (in press) also admits, although it is surely very difficult to investigate.

In the following sections we will try to focus especially on the facets of systematicity and productivity, by investigating in depth an instance of overabundance in Italian which has remained rather unnoticed till now (see, however, Micheli 2016: 252), namely the plural cell of a semantically definable subclass of Verb-Noun compounds.

3. The plural of Italian VN compounds: a semantic approach

There are arguably few topics in Romance word formation which have been investigated more thoroughly than Verb-Noun compounds (hereafter also VNCs), given the many peculiar features of this procedure. Salient issues include (at least): exocentricity, wide polysemy, double nominal/adjectival function, diachrony, and morphology-syntax interface. We will not even try to give an account of the existing bibliography. For single languages, we may just mention Rainer (1993: 265–278), Val Álvaro (1999: 4788–4799) and Moyna (2011) for Spanish, Villoing (2009) and Rosenberg (2011) for French, Bisetto (1999), Ferrari-Bridgers (2005), Ricca (2010) and von Heusinger & Schwarze (2013) for Italian. For more references and a general perspective at the

Romance level, the reader is referred to Gather (2001), and for a recent brief synthesis to Ricca (2015).

However, some issues are still not entirely settled, especially in areas which display cross-linguistic variation, despite the widely parallel behaviour of such compounds in all Western Romance languages. One such territory of variation is surely the distribution of singular vs plural marking in the noun (both for singular and plural compounds), and, consequently, the options available for the plural marking of the whole compound.

3.1 The number marking of N in singular VNCs in Italian

The starting point has to be the number marking of N which occurs when the whole compound has a singular meaning.

In the following we will focus on Italian only, but a brief comparison may be made with Spanish, as described by Rainer (1993), Val Álvaro (1999) and others. In Spanish, singular VNCs display the *-s* plural marker on N in the great majority of cases. The *-s* has extended even to most instances of unique referents and mass nouns, as in (1a-b):

(1) a. *un trotamundos*   'a:M.SG trot + world:**PL**, globetrotter'
    b. *un quitanieves*   'a:M.SG sweep + snow:**PL**, snowplough'

Even in Spanish, the generalization of the *-s* marker is certainly not complete (a list of exceptions can be found in Val Álvaro 1999: 4798), and probably has different relevance across the Spanish-speaking world (cf. Rainer 1993: 272). Nevertheless, this process implies a substantial change in the very description of the *-s* marker. As already noted by Rainer & Varela (1992: 130), the *-s* in Spanish VNCs has mainly taken the role of marking the composition process itself, with no consistent semantic motivation: i.e., it has come a long way towards becoming a morpheme in the sense of Aronoff (1994).

In Italian, on the contrary, the number marking of N in the singular form of the VN compound can be dealt essentially in semantic terms. For a similar approach, cf. von Heusinger & Schwarze (2013: 332–336) and – very briefly – Ricca (2015: 701). We propose here a six-grade scale, given in Table 1. Proceeding from grade 1 to grade 6, the semantic plausibility of plural N-marking increases, and the behaviour of Italian compounds overwhelmingly reflects this state of affairs.

Table 1: Semantic subclasses of Italian VN compounds: singular cells

VNC-type		Example	Gloss	Translation	N-marking
1. Mass-N		*spazzaneve*	sweep + snow:**SG**	'snowplough'	always SG
2. Unique referent-N		*prendisole*	take + sun:**SG**	'sundress'	always SG
3. Stable-N		*copriletto*	cover + bed:**SG**	'bedcover'	mostly SG[a]
4. Variable-N	4a.	*cacciavite*	thrust + screw:**SG**	'screwdriver'	all options: SG, PL or both alternatives in the same compound
	4b.	*fermacravatta/-e*	hold + tie:**SG/PL**	'tiepin'	
	4c.	*schiaccianoci*	crush + nut:**PL**	'nutcracker'	
5. Multiple-N		*contapassi*	count + step:**PL**	'pedometer'	always PL
6. *Plurale tantum*-N		*portaocchiali*	carry + glasses:**PL**	'eyeglasses case'	always PL

[a] For an explanation of some exceptions, see Footnote 4 below.

The semantic motivation for the extreme grades of the scale, i.e. 1-2 vs 6, is quite obvious, as it is directly related to the inherent properties of N alone. There is simply no singular form available

for *pluralia tantum* nouns[3], and, on the other hand, no plural form available for mass nouns and unique referents, at least with the intended meaning. Notice, however, that in Spanish even these inherent properties are usually overcome by the plural-N generalization rule, as shown in (1).

As for the intermediate grades 3-5, the semantic motivation has to involve the interplay of V and N.

In Italian, N is invariably marked as plural in those compounds in which each single state of affairs associated with the compound (typically) involves a plurality of N referents. This is class 5, the type *contapassi* 'pedometer', *lavapiatti* 'dishwasher' (compare the unmarked singular in the English equivalent of the latter) which we will label "multiple-N compounds".

The opposite situation occurs in cases when the object denoted by the VNC typically acts on one and the same N referent throughout its existence, like *copriletto* 'bedcover' (an object usually associated to the same bed in a given room). This is class 3, labelled "stable-N compounds": since a single N referent is involved, the singular marking of N is to be expected semantically, and indeed normally occurs[4].

A more complex situation occurs in class 4, where two motivations compete, both of which are semantically plausible, with the consequence of licensing both options. Clear instances of this can be seen with items like 'nutcracker' or 'screwdriver'. In these cases we are dealing with objects which typically act on a single N referent each time they are used, but on different referents on different occasions. Therefore, the plural marking of N reflects a sort of external quantification (over all occasions), and the singular marking an internal one (a distributive reading). The semantic unpredictability of the choice is well reflected in the two items mentioned above: the relationship between N and V for the activities associated with a nutcracker and a screwdriver is essentially the same, but the Italian VNC equivalents − in the singular cell − only display a plural N for the former (*lo schiaccianoci*) and a singular N for the latter (*il cacciavite*). In many instances the lexicalization process has ended up by fixing (arbitrarily) only one variant as available; however, it is to be expected that other items − even some well-established in the lexicon − may display similar frequencies for both alternatives: e.g., the ratio of *il fermacravatta* vs *il fermacravatte* 'the tiepin' is currently around 6:1 on the Web.

Therefore, the items in class 4, which we have labelled "variable-N compounds", are the only ones among which we may expect relevant instances of overabundance *in the singular cell* of the paradigm. These are identified as the subclass 4b in Table 1. However, they will not be our main concern in the following, as will be made clear in §3.2.

For the sake of comparability, examples in Table 1 are all instances of [-animate] Instrument VNCs, but the same pattern is found with [+animate] Agent VNCs. For instance, a *portalettere*

---

[3] Following the remark of a reviewer, we may note that the singular noun *occhiale* indeed exists in Italian (usually marked as uncommon in dictionaries), either with the meaning of 'each lens found in glasses or similar optical instruments', or simply as a rarer alternative for *occhiali* 'glasses'. A similar state of affairs is not unusual in Italian, cf. *pantaloni/pantalone* 'trousers', *forbici/forbice* 'scissors'. In all such cases, we think that the label of *plurale tantum* remains basically adequate.

[4] An interesting kind of − partial − exception is given by an item like *copricerchio/copricerchi* 'cover + circle:SG/PL, wheel cover'. Given that a single wheel cover is usually firmly associated with one and the same wheel of a given car, we should assign this compound to the stable-N class, and therefore expect a singular N-marking (*un copricerchio*). This is indeed usually the case, but the alternative plural N-marking *un copricerchi* is equally found in a relevant minority of occurrences (the ratio on the Web, 8/6/2018, is about 5:1). However, this exception is easily explained, given that such objects are usually dealt with in sets of four. Therefore, the plural *i copricerchi* is more frequent than the singular (to give just a rough estimate, the current ratio on the Web is 3:1). But when the compound occurs in the plural, the plural N-marking is semantically justified, see §3.2, and overwhelmingly prevails (current Web ratio 36:1). The singular form with plural N-marking, *un copricerchi*, can be then plausibly seen as a back-formation. Similar − and stronger − considerations hold for an item like *un parastinco/parastinchi* 'a:M:SG protect + shin:SG/PL, shin guard', whose variants with singular/plural N-marking display a ratio around 1:1 on the Web: the back-formation interpretation for the singular *parastinchi* is supported by a ratio of about 14:1 of plural vs singular occurrences for the whole compound (shin guards obviously are met and used in pairs in the real world, and in this case there is no competition at all in the plural).

'carry + letter:PL, postman' definitely belongs to the multiple-N class, while a *prestanome* 'lend + name:SG, front-man' is a stable-N compound (an individual only "lends" his/her own name for illegal purposes). Finally, cases like *spazzacamino* 'sweep + chimney:SG, chimney sweep' vs *cantastorie* 'sing + story:PL, storyteller' illustrate the two options available for the number marking of N in the case of variable-N Agent VNCs.

Needless to say, when it comes to assigning every single VNC to the different subclasses, the semantic classification in Table 1 is not always as clear-cut as it would appear at first sight. Difficulties may arise essentially for two reasons:

(i) on the one hand, many compounds are polysemic, and the different meanings can (typically) belong to different classes: this point is illustrated in §4.1.1 below;

(ii) perhaps more problematically, the distinction between multiple-N and variable-N compounds depends crucially on how strictly the concept of "single occasion" is taken.

The very example of *cantastorie* 'storyteller' given above may illustrate the latter point. It has been taken as an instance of variable-N compound under a rather strict interpretation of "single occasion": of course a storyteller cannot tell two or several stories simultaneously. But (s)he may certainly tell several stories in a row, in the street, in a single recital and the like (indeed is typically expected to do so). If we take the whole recital as a single event/occasion, the compound in question shifts to the multiple-N class: thus, perhaps, given the borderline status of *cantastorie*, its exclusive plural N-marking is not as unpredictable as the exclusive singular N-marking of *spazzacamino*.

Probably a strict notion of "single occasion", limiting the multiple-N compounds to instances in which several N referents are involved more or less simultaneously, is easier to implement consistently than a loose one. Nevertheless, fuzzy boundaries are unavoidable.

The same holds for the border between variable-N and stable-N compounds. For instance, a neck warmer (It. *scaldacollo,* warm + neck:SG) is typically a personal piece of clothing (hence a stable-N), and a headrest (It. *poggiatesta*, lean + head:SG) is expected to meet quite a few heads in the course of time (hence a variable-N). But neck warmers can be borrowed and perhaps even rented, and if the headrest is part of the boss' favourite armchair, its variable-N status may be much reduced.

Fortunately, these unavoidable border problems do not have much influence on the methodology followed in this investigation, as will be discussed in §4.

The semantic classification given in Table 1 is very similar to the description found in von Heusinger & Schwarze (2013: 334–335). Their "condition 1" covers mass Ns (our class 1), and their "condition 2" comprises both our classes 2 and (at least partly) 3. Furthermore, their "condition 3" predicts plural N-marking for the same items which we define as multiple-N. The only discrepancy concerns our variable-N items, which apparently fall under their "elsewhere condition" and so are predicted to display singular N-marking only. However, as shown above, this holds only for a subset of them (of the type *cacciavite* 'screwdriver', class 4a), but not for items like *schiaccianoci* 'nutcracker', which definitely does not fall under von Heusinger & Schwarze's condition 3, since the typical event associated with it does not "involve more than one object (*at a time*)" (von Heusinger & Schwarze 2013: 335; emphasis added), and nevertheless obligatorily displays plural N-marking.

3.2 Overabundance in the plural cells of Italian VN compounds

The classification of Italian VN compounds described in §3.1 clearly indicates the cases in which we may expect the emergence of a kind of overabundance with features of systematicity and productivity, as founded on general semantic grounds and not due to local idiosyncrasies of single items. Variable-N compounds have been shown to be the locus of such overabundance *for the singular cells*: the overabundant items have been identified as the subclass 4b (items like *cacciavite* and *schiaccianoci*, both belonging to class 4 and displaying opposite marking of N, are not

overabundant at all in the singular). However, overabundance can be expected to have a greater relevance in the *plural* cells. This is shown in Table 2.

Table 2: Overabundance in singular and plural cells of Italian VN compounds

VNC-type		Singular cell	Plural cell	
1. Mass-N		*spazzaneve*	*spazzaneve*	No overabundance – N:SG only
2. Unique referent-N		*prendisole*	*prendisole*	(with few exceptions)
3. Stable-N		*copriletto*	***copriletto/-i***	**One-cell overabundance**
4. Variable-N	4a.	*cacciavite*	***cacciavite/-i***	**(in plural only)**
	4b.	***fermacravatta/-e***	***fermacravatta/-e***	**Two-cell overabundance**
	4c.	*schiaccianoci*	*schiaccianoci*	No overabundance – N:PL only
5. Multiple-N		*contapassi*	*contapassi*	
6. *Plurale tantum*-N		*portaocchiali*	*portaocchiali*	

Clearly, VN compounds with N:PL in the singular (i.e., classes 6, 5 and 4c in Table 2) are not expected to display any overabundance at all in the plural cell, because they are necessarily invariable. There is no further slot in Romance morphology to mark the external plural (i.e., the plural of the VNC as a whole) if N is already marked as plural in the singular cell.[5]

At the other end of the scale, the VNCs where N is a non-numerable noun or unique referent (classes 1 and 2) could formally show a singular-plural contrast, because they display N:SG in the singular. This, however, would be semantically unjustified. Indeed, most of them are invariable, and also keep N:SG in the plural:

(2)   a. *lo/gli spazzaneve* 'the snowplough(s)', *lo/gli scolapasta* 'the drain + pasta:SG, colander(s)'
       b. *il/i prendisole* 'the sundress(es)', *il/i giramondo* 'the turn + world:SG, globetrotter(s)'

The scattered exceptions of N:PL marking in the plural – as well as the instances of overabundance – are mostly interpretable as cases of strong opacification of the compound (e.g., *i girasoli* 'the:M.PL turn + sun:PL, sunflowers'; other examples in Micheli 2016: 248–249). In some cases of class 1 items, the N:PL marking may perhaps be favoured by the existence of plural forms for the mass noun involved, although semantically idiosyncratic. Cf. *fanghi* 'mud:PL, mud baths', *ceneri* 'ash:PL, ashes (of the dead)', and the plural VNCs *parafango/-ghi* 'protect + mud:SG/PL, mudguards', *portacenere/-i* 'carry + ash:SG/PL, ashtrays'.

Therefore, the types of VNCs which are expected to display systematic number variability are those in class 3 (stable-N compounds) and in the subclass 4a of variable-N compounds[6]. It must be stressed that for these items, semantic considerations just point to the *possibility* of marking number

---

[5] Occasionally, a VNC of this sort may exhibit a distinctive plural form. For instance, It. *paracadute* 'stop + fall(F):PL, parachute' belongs to class 4c, with no overabundance in the singular. However, besides the standard invariable plural *i paracadute* 'the:M:PL parachutes', a plural form *i paracaduti* – although not found in dictionaries – is rather frequent on the Web (approximate ratio 1:4). This deviant form necessarily implies the reanalysis of the (F):PL *-e* ending as the homophonous (M):SG ending of the *-e/-i* inflectional class (cf. *can-e* 'dog (M)'), which in turn means that some speakers do not perceive the compound as an analysable unit any more (Ricca 2015: 702).

[6] Notice that von Heusinger & Schwarze (2013: 332–333) seem not to recognize number variability for any subclass of Italian VNCs, as they treat all of them as basically invariable, apart from exceptional cases like *i portaceneri* 'the ashtrays', which are analysed as instances of external plural marking, along the lines of our Footnote 5 and Scalise's general approach (see Footnote 7). However, this does not find confirmation in the empirical data, as will be clearly seen in the following sections (see especially §4.2) and can also be inferred from data in Micheli (2016: 248–253), who, however, does not look for any motivation through further internal classifications within VNCs.

distinctively, not the *necessity* of doing so. Let's take *copriletto* 'bedcover', a stable-N compound in class 3, as an illustration. Given that one bedcover (usually) covers the same bed throughout, we expect just the singular form *il copriletto*, with N:SG. But in the plural we expect both N:PL *i copriletti* (different bedcovers normally cover different *beds*) and N:SG *i copriletto* (different bedcovers cover *one bed* each).[7]

Variable-N compounds are not unitary, precisely because they already behave differently – and unpredictably – in the singular cell, as seen in §3.1. Those compounds patterning like multiple-N VNCs (class 4c, like *schiaccianoci*) obviously will remain invariable. On the contrary, those patterning like stable-N VNCs in the singular (class 4a, like *cacciavite*) are expected to follow the behaviour of the latter in the plural as well, allowing for overabundance in the plural. Finally, those which already display overabundance in the singular (class 4b items like *fermacravatta/-e*), may well double it in the plural, giving rise to a different overabundance pattern, with both cells involved, which will not be further discussed here.

Therefore, the items in classes 3 and 4a are the topic of the following sections. Of course, what has been said up to this point does not mean that every item in these classes will display overabundance in the plural cells. The semantic analysis discussed above only implies that classes 3 and 4a are the ones in which overabundance is expected to occur, possibly with a systematic character. Its extent and systematicity are an empirical issue, and have been tested with Web and informant data, as will be discussed in the following sections.

4. Web data

4.1 Corpus, data and methodological issues

In this subsection, we describe the procedure followed to gather our corpus of VN compounds (§4.1.1), and then (§4.1.2) we justify the choice of using Web counts in our analysis, briefly discussing the much debated topic of the so-called "Web as corpus" (Kilgarriff & Grefenstette 2003).

4.1.1 A corpus of stable-N and variable-N compounds

In GRADIT (*Grande dizionario italiano dell'uso*, De Mauro 1999), an extensive dictionary of modern Italian usage, any verb base which is used in three or more VN compounds is classified as a *confisso* (literally 'confix', a term used in the Italian tradition to designate the bound formatives involved in the so-called (neo)classical compounds, such as *biography*, *monarchy*) and, consequently, lemmatized. This choice, although questionable from a theoretical point of view (cf. Ricca 2005: 469), is very useful in practice, since it enables one to find with ease all these elements, which are listed in Ricca (2005: 469–470). Our first step was then to extract from the abridged version, De Mauro (2000) – readily accessible and presumably more focused on current use – all the nominal[8] VN compounds whose first constituent belongs to this list.

First, we searched for lemmas displaying the same initial sequence of characters as the verb base, for example *salva-*, from the verb *salvare* 'to save'. Then, we manually discarded words which were not VN compounds, such as the agent noun *salvatore* 'saviour'. De Mauro (2000) also provides information on the frequency of usage of each word: we decided to exclude VN compounds marked as "low usage", "obsolete", "literary", "regional", "dialectal" or "exotic",

---

[7] For a completely different analysis, which interprets any occurrence of plural N-marking in all classes – when absent in the singular – just as an "external" marker of the *-o/-i* inflectional class, see e.g. Scalise (1994: 139) and the discussion in §4.2.

[8] Compounds whose output was exclusively adjectival have been discarded.

because of their marginality. As an example, *salvaroba* (instead of *guardaroba*, 'wardrobe'), is marked as "obsolete" and it has therefore been discarded. Compounds appearing only in some particular collocations, such as *battibaleno* (found only in the idiom *in un battibaleno* 'in a flash'), have been excluded as well.

The second step in our procedure consisted in tagging each compound following the classification proposed in §3.1. Even during this step some exclusions were made, when the classification was particularly problematic. A reason could be that the compound was so opaque as to make it impossible (at least synchronically) to apply the distinctions made in §3.1. Another difficulty could come from polysemic items, in cases where the different meanings would be assigned to different classes. A good example is the word *reggicanne* ('fishing rod holder/carrier'), for which one finds the following definitions (based on De Mauro 2000, with simplifications):

1. 'fishing rod holder made of a sharp tube to be driven into the ground'
2. 'fishing rod carrier placed on boats or car roofs'
3. 'fisherman's belt provided with a fishing rod holder'.

The same compound is used to denote three different objects. While the second definition would clearly be appropriate for an item in the multiple-N class, the other two tools, on the contrary, seem to require an assignment to the variable-N class, or maybe – despite the plural marking of N – even to the stable-N class (if we think of a fisherman with only one fishing rod). Anyway, it is impossible to assign the compound as a whole to a single class.

The result of these first steps was a corpus of 736 VN compounds, whose distribution among the various classes is summed up in Table 3.

Table 3: **Distribution of VN compounds among the different semantic classes**

Class	Number	Percentage
1. Mass-N	130	17.6%
2. Unique referent-N	8	1.1%
3. Stable-N	139	18.9%
4. Variable-N	292	39.7%
5. Multiple-N	154	20.9%
6. *Plurale tantum*-N	13	1.8%
TOTAL	736	

As the data in the table show, the two classes in which overabundance is expected to be present in a massive – and perhaps productive – way (i.e. stable-N and variable-N compounds), constitute a very significant subset of VN compounds: variable-N compounds alone amount to almost 40% of VN compounds, therefore representing the largest class; as for stable-N compounds, though fewer in number, they nevertheless constitute a non-negligible part, accounting for about a fifth of the total.

As said in §3.1, in this study we decided to focus on compounds that are expected in principle to show overabundance in their plural cell only: therefore, we have only considered stable-N compounds and variable-N compounds in which the ratio between the two cell mates was beyond 10:1 in the singular cell[9].

---

[9] Although, in principle, we were interested in studying only the items assignable to classes 3 and 4a in Tables 1-2, we included items with a slight percentage of overabundance also in the singular cell, given that the border between subclasses 4a and 4b in Tables 1-2 cannot be rigid by definition (the semantics associated to all variable-N compounds is the same), and that even some stable-N compounds could display, especially at the community level, some limited instances of overabundance in the singular, particularly reflecting back-formation processes of the kind discussed in Footnote 4.

In our case, it is not possible to obtain quantitative data on the usage of the different cell mates from traditional corpora, because such compounds are often very rare, up to the point that even in a large corpus such as *itWaC* (1.585.620.279 tokens, cf. Baroni et al. 2009: 212) many searches returned very few results. As a consequence, in order to have an idea of the quantitative dimension of overabundance in this class of compounds, Web frequencies have been retrieved using a commercial search engine, namely Google[10]. The searches were not performed on isolated word-forms, but instead on phrases, making it possible to distinguish the number marking on N from the number value referred to the compound as a whole. As an example, in the case of *reggirullante* ('snare drum holder') we looked for exact matches of the following sequences, where each of the possible forms (*reggirullante* and *reggirullanti*) were preceded by the definite article (*il* in the singular, *i* in the plural), as summarized in Table 4 below.

Table 4: **A sample of Web searches**

Search	N marking	VNC number
"il reggirullante"	singular	singular
"il reggirullanti"	plural	singular
"i reggirullante"	singular	plural
"i reggirullanti"	plural	plural

In some cases, the DefArt-N phrase alone was not sufficient to obtain a significant amount of data: in order to get a higher number of results, a first possibility was to add the graphical variant in which the two components are separated by a hyphen or a blank (e.g. *reggirullante ~ reggi-rullante ~ reggi rullante*, 'snare drum holder'); another option was to search for other phrases too, for example "dei ('of the/some of the') reggirullante/-i", "due ('two') reggirullante/-i", "molti ('many') reggirullante/-i". Different phrases were also used when the results of the search relative to the sequence comprising the definite article seemed to be biased, for example because it came out to be a title of any sort (books, films, websites, blogs etc.), or the name of a firm (like "Il paralume" for a company selling lampshades)[11]. If, even with the help of the expedients described above, it proved impossible to obtain a minimum of 50 results in each cell of the paradigm, the compound was excluded from the corpus.

To the 109 compounds thus extracted from De Mauro (2000), we added another 54 compounds, which seemed to be interesting to analyse even if they are not present in the dictionary. This choice is motivated by the fact that we are interested also in observing the behaviour of very rare compounds, ones that we do not expect to find in dictionaries, since it is exactly this kind of compounds that can tell us something on the productivity of overabundance. These additional compounds were found by taking the attested lexical bases as a starting point, and checking to see whether the results of the combination of such bases with other nominal constituents yielded a compound that was solidly present on the Web, albeit not found in dictionaries.

The procedure described above resulted in a corpus of 163 stable-N and variable-N compounds, upon which the quantitative observations given below are based. All 163 items are listed for reference in the Appendix.

Since using Web data in linguistics, especially for a quantitative analysis, is controversial to a degree, that issue will be considered in more detail in the following subsection.

4.1.2 On the use of the "Web as corpus"

---

[10] The Web searches were performed, for the most part, between January and September 2016.

[11] For simplicity, in (3)-(8) the compounds are always quoted using the DefArt-N phrase, even if the ratios may be based on different searches.

Since the beginning of the new millennium (cf. the seminal paper by Kilgarriff 2001), the World Wide Web has been used as a sort of corpus surrogate. The Web is, indeed, a very attractive source of language examples, and it seems to be very useful, especially when corpus data are not sufficient. However, it was soon noted that using Web data poses some serious challenges that need to be faced, an issue that we will briefly discuss in this section.

First of all, a series of general problems arises whenever the Web is used to retrieve linguistic data (cf. Kilgarriff & Grefenstette 2003, Lüdeling et al. 2007, Hathout et al. 2008, Lew 2009, Fletcher 2012, Gatto 2014): the text is neither lemmatized nor tagged for Part-of-Speech, and a similar lack of detail emerges when moving from linguistic annotation to metadata; another issue is the complete absence of balancing of the Web, which, therefore, cannot be considered representative of anything other than itself; lastly, a very serious limitation is the fact that the size of the Web is not definite, with new pages constantly being added and old ones being removed. Then there are specific problems related to the so-called "Web frequencies": firstly, Web frequency figures refer not to distinct occurrences of the word forms searched for, but instead to the pages that contain those forms; secondly, there are duplicates – i.e. different Web pages containing the exact same text – and near-duplicates; lastly, there is the problem of the instability of Web counts, mainly because of its indefinite size, but also due to other reasons, such as the approximation of the number of results and the not entirely reliable precision and recall of search engines.

However, answers may be provided to at least some of the problems alluded to in these observations. As far as representativeness is concerned, it can be argued that, although it is certainly lacking in the Web, the situation is not that different from at least some traditional corpora. Even if it is certainly true that the Web is not representative of anything other than itself, the same holds true not only for the Web-crawled *itWaC* corpus, but also for reference corpora such as the British National Corpus: the detailed discussion of Kilgarriff & Grefenstette (2003) shows that representativeness can be considered a very challenging theoretical issue for corpus linguistics in general, rather than being just a problem of the Web. What is peculiar about the World Wide Web, as opposed to a corpus, is rather its chaotic composition, leading to a considerable amount of noise and to very dirty data. However, it has been shown that the presence of incorrect forms is not as significant as one might expect it to be (cf. Kilgarriff & Grefenstette 2003: 342). Regarding the lack of linguistic annotation and the limited search syntax, these are certainly serious limitations for linguistic research in general, but they are not particularly significant for our own research, where the simple searches described in §4.1.1 proved to be an efficient way to retrieve the relevant information. As for metadata, having them in a more systematic fashion would have been preferable, but again also by using a corpus such as *itWaC* we would not have solved the problem, since this kind of information is lacking even there. Moving on to the problems specifically concerning Web counts, the goal of the research must be kept in mind: in our case, the objective is simply to try to quantify, at least roughly, the strength of overabundance in each paradigm cell. From this standpoint, the aforementioned inconsistencies of Web frequencies do not seem to be too relevant, since they should all involve both cell mates equally.

The fact that Web counts have been used effectively to perform a series of NLP tasks (cf. the works cited in Nakov & Hearst 2005: 347) strongly suggests that these data are not useless, despite their drawbacks. Furthermore, an interesting study by Keller & Lapata (2003) has shown that Web frequencies correlate in a significant way with data obtained from traditional corpora, and, even more interestingly, with acceptability judgements from native speakers. Therefore, it seems that quantitative data extracted from the World Wide Web can be used, although with care, and that the enormous quantity of material manages to counterbalance the problems that have been mentioned.

The most serious limitation of the Web is its instability and, as a consequence, the impossibility of replicating results, which is crucial in a quantitative study. This seems to be a genuine problem, although some tentative solutions have been proposed (cf. Gatto 2014: 69). However, it can be argued that this fact alone is not sufficient to warrant a complete refusal to use such a large amount of material as the one contained in the World Wide Web, and that the Web can, nevertheless, be

considered a legitimate object of inquiry, in the same way as it is possible to make experiments on river water even if its chemical composition will inevitably be slightly different at each experiment (this convincing analogy was originally proposed by Kilgarriff 2001: 342). This is true especially in the field of morphology, where Web data can be particularly useful (cf. Hathout et al. 2008, Montermini 2015), since they allow the frequent problem of scarcity of data to be overcome. Generally speaking, a non-extremist attitude (such as the one advocated by Montermini 2015) towards the use of the Web seems to be advisable, avoiding both uncritical enthusiasm and total rejection.

4.2 Results

It is now possible to move on to an analysis of the results of the Web-based investigation. The data are summed up in Table 5. In each row, a different group of compounds is considered, allowing us to isolate three possible conditioning factors, namely: (i) the distinction between stable-N and variable-N compounds (rows 2-3), (ii) the degree of lexical entrenchment of the compound (rows 4-5), and (iii) the presence or absence of a gender mismatch between the internal noun and the whole compound (rows 6-7). In what follows, we will discuss each one of these factors in greater detail, in order to evaluate their relevance in influencing the presence and strength of overabundance in the plural cell.

Table 5: Overabundance in the plural cell: evaluation of different factors

row	group of compounds	balanced overab. ($< 2:1$)	strong overab. ($< 10:1$)	weak/no overab. ($> 10:1$)	total n.
1	all compounds	16.6%	32.5%	50.9%	163
2	stable-N	15.4%	35.4%	49.2%	130
3	variable-N	21.2%	21.2%	57.6%	33
4	more entrenched	10.8%	39.2%	50%	74
5	less entrenched	21.3%	27%	51.7%	89
6	no gender mismatch	23.5%	43.8%	32.6%	98
7	gender mismatch	6.2%	15.4%	78.5%	65

Overall, the data in the first row of the table show a remarkable presence of overabundance, empirically confirming our initial expectations, discussed in § 3.2. In almost half of the cases (49.1%), overabundance can be considered quite strong, the ratio between the occurrences of the cell mates being lower than 10:1 (the same threshold used in Thornton 2012: 189). Furthermore, in a smaller – but still significant – portion of compounds (16.6%), the ratio is lower than 2:1: since in this case there is not a clear preference for one of the cell mates, we refer to this situation using the term "balanced overabundance". Differently from Thornton (2012) – as we are mainly interested in productive overabundance – , we did not consider it useful for our purposes to make further distinctions above the 10:1 ratio.

Rows 2 and 3 of Table 5 show that the distinction between stable-N and variable-N compounds does not influence the results substantially: qualitatively, it is possible to find cases of very strong – and even balanced – overabundance in both kinds of compounds. For instance, in (3a-b) it is shown that in the indisputably stable-N compound *apricancello*, the two cell mates *(gli) apricancello/apricancelli* are used with a very similar frequency (1.2:1 ratio) [12], and the same holds for *mettifoglio*, which is certainly a variable-N compound (1:1.4 ratio between *(i) mettifoglio/-gli*).

---

[12] In (3) and the following examples, the ratios are always given by putting the values for the invariable form (singular N-marking) on the left.

(3) a. *gli   apricancello/-i*   (stable-N)   1.2:1 ratio
          the:M.PL   open + gate:SG/PL
          'the gate openers'

      b. *i   mettifoglio/-gli*   (variable-N)   1:1.4 ratio
          the:M.PL   put + sheet:SG/PL
          'the sheet placers'

Data in rows 4-5 of Table 5 have a significant import on the theoretical analysis of the plural marker *-i* in instances like (3a-b). As already mentioned in Footnote 7, a different way to account for the presence of a plural marker on N (when it does not occur in the singular form of the compound, of course) could be to consider it as marking the external plural of the whole compound, which then would be treated simply as a member of the *-o/-i* inflectional class: this is suggested, for instance, by Scalise (1994: 139). Following this approach, the plural marking of, say, *apricancell-i* in (3) would not differ from the one of a simple noun like *tavol-i* from *tavolo* 'table', apart from the fact that in the former case it competes with invariability. However, if this analysis were correct, one might expect to find equally frequent instances of plural N-marking in classes 1 and 2 of Table 2 (mass nouns and unique referents), where they are rather exceptional. Moreover, concerning the classes 3 and 4a we are discussing here, the expectation would be to find more instances of plural N-marking – and thus a stronger overabundance – in compounds that are firmly entrenched in the lexicon, and therefore could more plausibly be treated by speakers as an unanalysable whole. To see if this is indeed the case, in rows 4-5 we divided the compounds in two groups, according to an estimate of their degree of lexical entrenchment. To be classified as "more entrenched", a compound had to satisfy two conditions, namely: (i) being attested in De Mauro (2000), and (ii) occurring at least 10 times in the *itWaC* corpus. Otherwise, the compound was considered as "less entrenched". The figures of the table do not show a clear difference between these two groups of compounds, seriously weakening the viability of analyses like Scalise's. To provide a concrete example, while the remarkable presence of plural marking on the internal noun in a compound like *copricapo* – which is quite frequent and is plausibly stored in the lexicon as such, although a segmentation into constituents is of course not impossible – could be considered as a consequence of its high degree of entrenchment, such an analysis cannot easily account for the comparable situation found in a very rare word like *coprimanubrio*:

(4) a. *i   copricapo/-i*   (more entrenched)   1:1.4 ratio
          the:M.PL   cover + head:SG/PL
          'the hats'

     b. *i   coprimanubrio/-bri*   (less entrenched)   1:1.3 ratio
          the:M.PL   cover + handlebar:SG/PL
          'the handlebar covers'

In rows 6-7 of Table 5, another factor is considered, namely the presence or absence of a mismatch between the gender of the internal noun and the gender of the whole compound. This distinction seems to have a non-negligible effect on the strength of overabundance in our data. To illustrate this aspect, it is useful to start from some examples. In (5a-b), the agentive compound *portavoce*, literally 'voice carrier', can refer to a male person – meaning 'spokesman' – or to a female person – meaning 'spokeswoman'. In the first case, the compound is masculine, and there is a gender mismatch, since the internal noun *voce* is feminine: therefore, pluralization of N is disfavoured, and *i portavoce* is much more frequent than *i portavoci*. On the other hand, when also the compound is feminine, and so there is no gender mismatch, both forms – *le portavoce* and *le portavoci* – appear with comparable frequency.

(5) a. *i*   [[*porta*][*voce/-i*$_\mathbf{F}$]]$_\mathbf{M}$ (gender mismatch, *-e/-i* N)   24:1 ratio
        the:M.PL   carry + voice:SG/PL
        'the spokesmen'

     b. *le*   [[*porta*][*voce/-i*$_\mathbf{F}$]]$_\mathbf{F}$ (no gender mismatch)   3:1 ratio
        the:F.PL   carry + voice:SG/PL
        'the spokeswomen'

The qualitative picture described above is also quantitatively confirmed by a chi-square test: the distinction between compounds with and without a gender mismatch is indeed statistically significant ($p < 0.00001$), while both the stable-N vs variable-N distinction and the degree of entrenchment of the compound are not significant ($p > 0.1$ in both cases).

The effect of gender mismatch appears to be even stronger when the compound is masculine and N belongs to the *-a/-e* inflectional class, as is clearly shown by the very different overabundance ratios displayed by the fully synonymous compounds *coprisella* and *coprisellino* in (6a-b)[13]. The contrast is often even categorical in the converse case, when the compound is feminine and N belongs to the *-o/-i* inflection class: see, for instance, the agentive compound *ficcanaso* in (7a-b), which is always invariable (no overabundance at all!) when it denotes a woman, and is therefore feminine.

(6) a. *i*   [[*copri*][*sella/-e*$_\mathbf{F}$]]$_\mathbf{M}$ (gender mismatch, *-a/-e* N)   292:1 ratio
        the:M.PL   cover + saddle:SG/PL
        'the saddle covers'

     b. *i*   [[*copri*][*sellino/-i*$_\mathbf{M}$]]$_\mathbf{M}$ (no gender mismatch)   1:4 ratio
        the:M.PL   cover + saddle:DIM:SG/PL
        'the saddle covers'

(7) a. *i*   [[*ficca*][*naso/-i*$_\mathbf{M}$]]$_\mathbf{M}$ (no gender mismatch)   7:1 ratio
        the:M.PL   stick + nose:SG/PL
        'the male meddlers'

     b. *le*   [[*ficca*][*naso*$_\mathbf{M}$]]$_\mathbf{F}$ (gender mismatch, *-o/-i* N)   ∞ ratio
        the:F.PL   stick + nose:SG   (N:SG only)
        'the female meddlers'

The gender mismatch effect could be interpreted, at first sight, as supporting Scalise's claim that the presence of plural marking in such compounds is simply due to external plural assignment, predictable from the final vowel of N: masculine compounds ending in *-o* and, similarly, feminine compounds ending in *-a* could easily be assigned to the productive *-o/-i* and *-a/-e* inflection classes, while, on the contrary, a gender mismatch compound like *coprisella*, if treated as a single whole, would belong to the masculine *-a/-i* class which is basically unproductive in Italian (and invariability would follow). However, it is very important to observe that these are just tendencies: there are compounds where overabundance is very strong despite the presence of gender mismatch, even in the stronger case of masculine compounds containing feminine nouns of the *-a/-e* inflection class. A particularly clear example is *reggimensola* in (8), where the ratio between the frequencies of usage of the two cell mates is almost completely balanced.

---

[13] This kind of data is probably the basis for the generalization often found in normative descriptions of Italian – from Goidanich (1918) up to, e.g., Serianni & Castelvecchi (1988: 134), and several websites depending on it –, which suggest that VN compounds are variable if and only if there is no gender mismatch between N and the whole compound. However, this generalization clearly does not hold, as noted also by von Heusinger & Schwarze (2013: 333, Footnote 9): it would predict systematic pluralization of masculine class 1-2 items when N ends in *-o* (but cf. $^?$*i perditempi* 'the:M.PL lose+time(M):PL, lazy persons', $^?$*i giramondi* etc.), and no pluralization at all for masculine compounds with feminine N, contrary to instances like Example (8).

(8)   *i*   [[*reggi*][*mensola/-e*$_\mathbf{F}$]]$_\mathbf{M}$ (gender mismatch, *-a/-e* N)   1:1 ratio
      the:M.PL   hold + shelf:SG/PL
      'the shelf brackets'

In such cases, the plural marker *-e* cannot in any way be interpreted as referring to the number opposition of the whole compound, since in Italian masculine nouns cannot be assigned to the *-a/-e* inflection class. As said above, masculine compounds containing *-a/-e* Ns, if they were treated like simple words, should remain invariable, or perhaps – much less probably – be assigned to the *-a/-i* inflection class of, e.g., *il sistema / i sistemi* 'the system(s)$_M$'. Therefore, the fact that also masculine compounds containing Ns belonging to the *-a/-e* inflection class do display overabundance provides strong evidence that Scalise's explanation in terms of external plural assignment cannot fully account for the presence of plural N-marking in such compounds.

5. Questionnaire data

The Web data presented in the previous section are very useful in order to obtain a basic quantitative assessment of the phenomenon. However, by means of such data it is of course possible to investigate the presence of overabundance only at the level of the global speaker community, making it very difficult to evaluate the relevance of sociolinguistic conditioning factors. To reduce the impact of this problem, by taking a small and well identifiable speaker community into account, and to investigate overabundance also in the competence of individual speakers, two different questionnaires have been employed. The procedure and the data are described in detail in the following subsections.

5.1 Questionnaire 1

5.1.1 Description

The first questionnaire consisted in a naming task based on visual input. The participants were 207 students of the University of Turin, attending an introductory linguistics course. They were asked to give a name to the objects depicted in some images[14], in written contexts in which the empty slot for the noun was inserted in an unambiguously plural NP (or DP). The questionnaire consisted of 30 images preceded by an introductory example, to which 4 distractors were added – namely, objects named by VN compounds belonging to a different semantic subclass, like *scolapasta* 'colander', containing a mass noun, or *schiaccianoci* 'nutcracker', a variable-N compound belonging to subclass 4c, with plural N-marking also in the singular cell. Several images intentionally displayed very unfamiliar objects, for which we could assume that many informants had never before been faced with the need to name them.
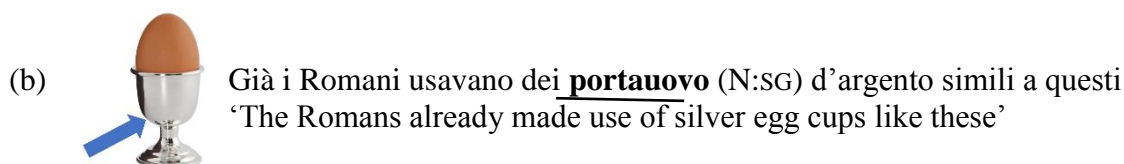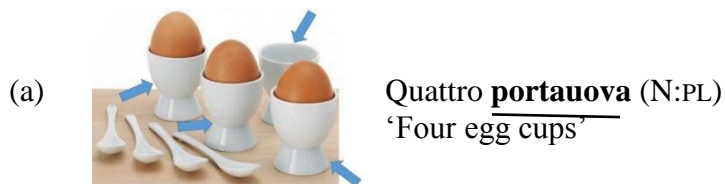 As illustrated in the two examples in (9), two slightly different versions of this questionnaire were submitted. In the first one, (9a) – which was filled out by 92 students – there were several objects in the image, and only the numeral in the linguistic context; in the second one, (9b) – filled out by 115 students – there was only one object in the image and a more complex linguistic context, involving various agreement targets, such as definite or indefinite articles, demonstratives and/or adjectives. While numerals other than 'one' are of course semantically plural, but do not carry

---

[14] Besides being on each participant's questionnaire, next to each sentence, the images were also shown one by one on a screen, in bigger size and in colour, so as to make them more easily recognizable. Each image was exposed for a period of 18 seconds, which allowed us to keep the answering time under control, avoiding too much afterthought by the informants.

explicit formal markers of number/gender, determiners and adjectives do formally show agreement in number, and often also in gender (see for instance the contrasting forms of the adjective 'beautiful': *bell-o* M.SG / *bell-a* F.SG / *bell-i* M.PL / *bell-e* F.PL).

(9)

(a) Quattro **portauova** (N:PL)
'Four egg cups'

(b) Già i Romani usavano dei **portauovo** (N:SG) d'argento simili a questi
'The Romans already made use of silver egg cups like these'

This difference could in principle have some effect on the choice of the form of the compound. We could expect the second version to favour N:PL answers in compounds with no gender mismatch – since in a sequence like *questi segnalibri* 'this:M.PL mark + book:(M).PL' the gender and number values (and even forms!) of the determiner coincide with the corresponding values of N in the compound. On the contrary, N:SG answers could be favoured in compounds with gender mismatch: for them, a sequence like *questi poggiateste* 'this:M.PL lean + head:(F).PL' makes the clash between markers more evident, whereas a sequence like *questi poggiatesta* 'this:M.PL lean + head:(F)SG' coincides formally with what we could have if the compound were treated as a simple noun (recall that masculine borrowings ending in *-a* are invariable in Italian).

On the other hand, considering the different kind of visual input, the first version of the questionnaire might possibly favour answers with plural N-marking (because several objects were shown), while the second one could be expected to favour forms with singular N-marking.

To evaluate the effect of the different design of the two questionnaires, a chi-square test was performed, considering only compounds with gender mismatch, where there is alignment of the two factors mentioned above: only for such compounds N:PL answers should be consistently disfavoured in the second version. The chi-square test showed that the difference was indeed significant ($p < 0.001$).

However, in what follows we will discuss only aggregate data, since the difference between the two versions – though significant – is not so great as to alter the overall behaviour of the compounds. In other words, the generalizations that we draw would nonetheless hold for both versions.

The data obtained with this first questionnaire are similar to Web data in that they allow us to investigate only overabundance within a community, not intra-speaker overabundance. However, they differ in two important respects: first, we are now dealing with a well identifiable and much more homogeneous community of speakers; moreover, we can say more about productivity, since we are directly observing the naming process of very rare items, which is a good approximation of an online process of word formation.

5.1.2 Naming options

It is important to stress that the participants were not explicitly required to answer with a VN compound: this was merely indirectly suggested by one example given at the beginning of the questionnaire – the compound *portaocchiali* 'eyeglasses case' – and by the insertion, as distractors,

16

of some very common VN compounds belonging to other subclasses (as said above in §5.1.1). The main reason for this choice was trying to minimize the informants' metalinguistic awareness while performing the task. We hoped that, submitted to an essentially visual input, the informants would focus on the naming task, and not on the morphological structure, which was of course the true aim of the investigation.

Naturally, due to the rarity of some of the objects represented in the images, and to the fact that no explicit metalinguistic instruction to produce a VN compound was given, different naming options were expected, and did in fact occur. In what follows, we briefly describe the different possibilities that we had to deal with, explaining and justifying our choices in treating the data.

Only in the best-case-scenario did we receive (almost) a single lexical answer, the compound that we were expecting. In other cases, we got several different compounds with the same N, but different verbal bases. For instance, when confronted with an image of a 'tongue cleaner', our informants consistently used the noun *-lingua*, but they came up with 17 different verbal bases, some of which were simply synonyms of the expected *pulisci-* (e.g. *netta-* 'clean', *lava-* 'wash'), but others had a different meaning, due to a misunderstanding of the purpose of the object (e.g. *abbassa-* 'lower', *ferma-* 'block'). In such cases, it seemed reasonable to group the compounds with the same N together in the evaluation of the results, provided that – of course – the meaning of the verbal base did not alter the semantic classification of the compound.

The opposite possibility also occurred: sometimes we elicited compounds with the same verbal base, but with different synonymous nouns. For instance, to name a 'steering-wheel lock', the verb *blocca-* 'block' was consistently used – although not exclusively, see Table 6 below – but we registered a competition between two different nouns, *-volante* and *-sterzo*, both meaning 'steering-wheel'. In such cases, it did not seem advisable to group the answers, because of the possible effect of differences in gender and inflection class. Normally we simply discarded the minority answer(s); however, in cases like *bloccavolante* vs the expected *bloccasterzo*, where both alternatives were frequent, it was possible to keep the former as a new item to be treated independently.

On the other hand, we had to discard answers giving a structurally different compound, for instance *portadocumenti* 'document holders', which is arguably a plural-N compound in Italian, differently from the expected *portapatente/-i* 'driving licence holders', clearly a stable-N compound. Of course, we also excluded answers which were not a VN compound at all, like the generic *tazze* 'cups' or *bicchieri* 'glasses' instead of the expected *portauovo/-a* 'egg cups' for the object shown in (9).

In a few cases, the discarded answers were so numerous that we had to eliminate the item from our data set. We established a lower limit of 40 usable answers, and 3 items fell below this threshold, namely the bottom three in Table 6, which gives a quantitative evaluation of the possible naming options. Overall, answers with VN compounds were largely dominant, and for almost two thirds of the items the grouped set of the eligible VN compounds totalled more than half of the answers (as shown by the light grey cells in Table 6). This seems to confirm the general opinion that VN compounding is a very productive word formation strategy in Italian. The cases of strong competition between different verbal bases can be interpreted as another piece of evidence supporting this claim. In Table 6, the cells showing a non-negligible difference between the third column – which refers only to the main verbal base – and the fourth column – which sums the percentages of all the compounds with same N but different V – are highlighted in dark grey. A particularly clear example is provided by 'helmet holder': for this item we elicited two different verbal bases – *appendi-* 'hang' and *porta-* 'carry' – with comparable frequency, together with other, more marginal, options.

Table 6: Q1 data: different naming options

main VNC	gloss	main VNC only	grouped VNC	discarded VNC	not a VNC / no answer
segnalibro/i	mark+book	99%	99.5%	0%	0.5%
cacciavite/i	thrust+screw	98.6%	99.5%	0%	0.5%
copridivano/i	cover+couch	97.6%	97.6%	0.5%	1.9%
poggiatesta/e	lean+head	94.2%	95.7%	0%	4.3%
portauovo/a	carry+egg	89.4%	89.9%	0.5%	9.7%
segnaposto/i	mark+place	86.5%	87%	5.3%	7.7%
portaspazzolino/i	carry+toothbrush	74.9%	84.5%	0%	15.5%
copriob(b)iettivo/i	cover+lens	79.2%	79.7%	1.4%	18.8%
tappanaso/i	plug+nose	76.8%	79.2%	2.4%	18.4%
copripoltrona/e	cover+armchair	78.7%	78.7%	9.2%	12.1%
portapenna/e	carry+pen	59.4%	78.3%	7.7%	14%
puliscilingua/e	clean+tongue	37.0%	69.6%	2.4%	28%
coprisellino/i	cover+saddle	63.3%	63.3%	28%[a]	8.7%
appendiborsa/e	hang+bag	22.0%	63.3%	3.9%	32.9%
appendicasco/-chi	hang+helmet	20.8%	60.4%	7.7%	31.9%
copricostume/i	cover+swimsuit	59.4%	59.4%	9.2%	31.4%
copricatena/e	cover+chain	47.1%	59.4%	6.3%	34.3%
scaldacollo/i	warm+neck	52.2%	57%	0.5%	42.5%
portaspazzolone/i	carry+ toilet brush	50.7%	51.2%	28.5%	20.3%
copriletto/i	cover+bed	46.4%	46.4%	6.8%	46.9%
poggiachitarra/e	lean+guitar	25.8%	44.9%	4.3%	50.7%
copriteiera/e	cover+teapot	43.5%	44%	4.3%	51.7%
bloccasterzo/i	block+ steering wheel	39.6%	43%	30.9%[a]	26.1%
portagiornale/i	carry+newspaper	24.4%	43%	15%	42%
reggimensola/e	hold+shelf	18.8%	31.4%	9.7%	58.9%
salvaschermo/i	save+screen	10.1%	29.5%	3.9%	66.7%
copridito/a[b]	cover+finger	6.3%	21.7%	3.4%	74.9%
tagliamelone/i	cut+melon	11.8%	15.5%	68.1%	16.4%
poggiatamburo/i	lean+drum	4.3%	12.6%	6.3%	81.2%
portapatente/i	carry+ driving licence	8.2%	8.7%	81.6%	9.7%

light grey: > 50% of usable VNC answers
dark grey: cases of significant competition between V bases

---

[a]These two very high percentages of "discarded VNC" are mainly due to the two synonymous items *coprisella* and *bloccavolante*, which were frequent enough to be treated autonomously, as said in §5.1.2 (see Tables 7 and 9).
[b]*Copridito* is surely not a familiar word for most (perhaps all) Italian speakers, as remarked by a reviewer. The expected word was the slightly more familiar *salvadito* 'save + finger, finger guard (e.g. for archers)'; presumably the image was not explicit enough and most speakers opted for the semantically more general 'cover' as a suitable verb base.

The most stable base was *copri-* 'cover', with practically no competitors whenever it was the expected choice. This base was dominant also in some compounds where it was not expected: see for instance the case of 'bicycle chain guard', where *copricatena* 'cover + chain' was preferred over the expected (because more current in lexicons) *paracatena* 'protect + chain'. On the contrary, compounds with *reggi-* 'hold' (e.g. *reggiborsa* 'hold + bag, bag holder') seem to be rather unstable, since the verbal base was often replaced by alternatives like *porta-* 'carry' or *appendi-* 'hang', or even by non-VNC options.

5.1.3 Overabundance in data from questionnaire 1: overall results

Coming now to the analysis of the results, Table 7 lists all eligible VNCs elicited in the questionnaire, ranked according to their "overabundance value". The value is simply the highest number occurring in the ratio between singular and plural occurrences, when the lowest one is normalized to 1: i.e., both a ratio of 2.5:1 and 1:2.5 would correspond to an overabundance value of 2.5. Obviously, the lower this value, the more balanced the overabundance. As said above, in most cases we had to group different verbal bases occurring with the same N: the ranking is made considering the total of the grouped VN compounds. However, even if only the main verbal bases were to be considered, the ranking would not be much different[15].

In Table 8, a comparison is drawn between the data from the Web and from this first questionnaire, both of which deal with community overabundance, although two very different kinds of community are considered. Due to the substantial differences in the two VNC samples, an overall comparison between Web data and Q1 data (given in the first and second row of Table 8) may be of limited value. However, it suggests a higher rate of both balanced and strong overabundance in the informants' answers. When the comparison is restricted to the 19 items shared between the two corpora, the same tendency is confirmed, but only when medium- vs low-ranked items are compared (see the third and fourth row of the table).

Regarding the three possible conditioning factors considered for Web data in §4.2, Table 5, the picture that emerges from Q1 data is given in the third, fourth and fifth column of Table 7. Neither the different semantic classification (stable-N vs variable-N) nor the different degree of lexical entrenchment of the various items seem to influence much the presence and strength of overabundance, and this observation is comparable to what could be deduced from Web data. A difference, however, occurs with the gender mismatch effect described in examples (5)-(8), which does not emerge in questionnaire data: indeed, even in the stronger case of masculine compounds where N is feminine and belongs to the *-a/-e* inflection class, there is a fair number of cases where overabundance is strong (like, for instance, in *copriteiera* 'teapot cover') or even balanced (in two items, *appendiborsa* 'bag holder' and *portapenna* 'pen holder'). This is another piece of evidence which contradicts Scalise's (1994) hypothesis of plural N-marking as "external marking" only: in a naming task in which many informants were presumably confronted with unfamiliar items that they had not previously stored in their mental lexicon, the N-pluralizing strategy is nevertheless quite frequent, even in cases of gender mismatch when it gives an impossible output for Scalise's theory (which could at most predict **i copriteieri*, but never *i copriteiere* as attested).

On the other hand, another factor, which we did not discuss before, emerges very clearly in the questionnaire data: if the internal N refers to a unique body part, like for instance *collo* in *scaldacollo* 'neck warmer', the singular form largely prevails, leaving hardly any room for overabundance. This fact is not unexpected, given that in Italian there is a strong preference for distributive constructions also in syntax when body part nouns are involved, as the contrasts in examples (10a-b) clearly show.

---

[15] Notice that, for brevity and readability, the first column of Table 7 reports only the "main VNC" – i.e. the one with the expected verb base, or the most frequent in few cases where it did not coincide with the former.

(10) a. *alzate la mano destra /* [??]*le mani destre* 'raise your right hand(s)'

    b. *non soffiatevi il naso/*i nasi in pubblico* 'do not blow your nose(s) in public'.

Table 7: Q1 data: items ranked by decreasing overabundance

main VNC	gloss	type	entren-ched	gender mismatch	N. of answers	overabun-dance value
cacciavite/i	thrust+screw	var	yes	yes	206	**1**
appendiborsa/e	hang+bag	var	no	yes	131	**1.2**
portauovo/a	carry+egg	var	yes	yes[a]	186	*1.3*
portapenna/e	carry+pen	stab	yes	yes	162	*1.3*
copridito/a	cover+finger	stab	no	yes[a]	45	*1.4*
portaspazzolino/i	carry+toothbrush	stab	no	no	175	**1.4**
portagiornale/i	carry+newspaper	var	no	no	89	*1.5*
copriob(b)iettivo/i	cover+lens	stab	no	no	165	**1.5**
coprisellino/i	cover+saddle	stab	no	no	131	**1.7**
reggimensola/e	hold+shelf	stab	no	yes	65	**2.1**
segnaposto/i	mark+place	stab	yes	no	180	**2.2**
copriteiera/e	cover+teapot	stab	no	yes	91	**2.8**
copriletto/i	cover+bed	stab	yes	no	96	**3.2**
portaspazzolone/i	carry+toilet brush	stab	no	no	106	**3.2**
copridivano/i	cover+couch	stab	yes	no	202	**3.3**
appendicasco/-chi	hang+helmet	stab	no	no	125	**3.5**
segnalibro/i	mark+book	var	yes	no	206	*3.6*
bloccavolante/i	block+steering wheel	stab	no	no	54	**3.9**
copricatena/e	cover+chain	stab	yes	yes	123	**3.9**
bloccasterzo/i	block+steering wheel	stab	yes	no	89	**6.4**
copricostume/i	cover+swimsuit	var	yes	no	123	**6.7**
copripoltrona/e	cover+armchair	stab	no	yes	163	**7**
tappanaso/i	plug+nose	stab	no	no	164	**11.1**
poggiachitarra/e	lean+guitar	stab	no	yes	93	**14.5**
salvaschermo/i	save+screen	stab	yes	no	61	**19.3**
scaldacollo/i	warm+neck	stab	yes	no	118	**25.2**
puliscilingua/e	clean+tongue	stab	no	yes	144	**27.8**
poggiatesta/e	lean+head	var	yes	yes	198	**32**
coprisella/e	cover+saddle	stab	yes	yes	44	**43**

dark grey: balanced overabundance (N:SG / N:PL around 1:1 ratio, between 2:1 and 1:2)
light grey: strong overabundance (N:SG / N:PL between 10:1 and 1:10)
white: (very) weak overabundance (N:SG / N:PL higher than 10:1)
bold: prevalence of N:SG; *italics:* prevalence of N:PL

---

[a] In these cases, the gender mismatch is detectable only in the plural, because *uovo* and *dito* belong to the residual class of nouns with alternating gender.

Table 8: Overabundance in Web vs Q1

	balanced overabundance (< 2:1)	strong overabundance (2:1 – 10:1)	weak/no overabundance (> 10:1)	total n.
Q1 data	31%	44.8%	24.1%	29
Web data	16.6%	32.5%	50.9%	163
comparable VNC only: Q1	15.8%	52.6%	31.6%	19
comparable VNC only: Web	15.8%	36.8%	47.4%	19

To sum up, the data elicited in the first questionnaire confirm a solid presence of overabundance in the plural cell of the compounds of the two subclasses considered, maybe one even more solid than in Web data, at least as far as community overabundance is involved. The fact that this conclusion can be drawn even considering a much smaller and more homogeneous community of speakers might be taken as evidence that the presence and strength of overabundance is not much influenced by sociolinguistic factors, such as geographical and social background, or by stylistic differences.

5.2 Questionnaire 2

5.2.1 Description

The second questionnaire was submitted to the same group of students (at least in principle, see below) one week later. For each Q1 informant a customized Q2 was prepared, where they were required to give acceptability judgements on the same compounds provided in Q1. However, in this second questionnaire the N number marking was the opposite of the one given by the informants in their answers to Q1. Therefore, the input of Q2 was different for each informant. If, for example, an informant had answered *quattro portauova*, 'four carry + egg:PL' in Q1a (see 9a), (s)he received as Q2 input a sentence containing the form *portauovo*, 'carry + egg:SG'. Like in Q1b, the context required the plural interpretation of the VNC and consisted in full sentences, different from the ones used in Q1b and identical for all informants.

The evaluation scale was articulated in three levels, from full acceptability (A) to unacceptability (C), with an intermediate possibility (B) if the form was judged as acceptable, but another form was preferred. In case of answers B or C, informants were required to write the alternative, preferred form, while in case of judgements of full acceptability (A) an alternative could, but did not have to be provided. Such alternative forms were useful in order to ensure that what the informants evaluated was really the number marking of the compound: if, for instance, the form *copriobbiettivi* 'lens covers' was judged as unacceptable, but the suggested alternative form was *tappi* 'corks, covers', it is clear that the informant did not reject the use of plural N-marking, but the use of a compounding strategy altogether. Therefore, answers like these have been discarded.

Unfortunately, among the 207 students that had answered Q1 in the first lesson, only 121 were also present when Q2 was performed: this is why there are less data for Q2 than for Q1 (see the second column of Table 9 compared with the sixth of Table 7).

Table 9 contains 28 compounds, and not 30 like Table 6. The reason is that four Q1 items were discarded because too few relevant answers were provided. Conversely, each of the two variants of the Q1 items already split in Table 7 (*coprisella/coprisellino* 'saddle cover' and *bloccasterzo/bloccavolante* 'steering wheel lock') received enough answers to be treated independently in Q2 as well.

Unlike Web data and Q1 data, this second questionnaire allowed us to explore the possibility of overabundance as reflected in the competence of each single speaker – although, of course, only passive competence can be investigated in an acceptability judgement task like Q2.

5.2.2 Overabundance in data from questionnaire 2: overall results

The overall data are given in Table 9, in which items are ranked according to decreasing percentages of full intra-speaker overabundance – namely, A answers ("OK") in Q2. The considerable variation in the number of relevant answers, as displayed in the second column, is of course linked to the percentage of VN answers in Q1: if an informant did not use a VN compound in the first place (i.e., in Q1), it was obviously impossible to invert the number value of N in Q2, so that nothing could be said about intra-speaker overabundance.

Generally speaking, the informants showed a very high rate of positive acceptability judgements for the choice opposite of their own: full acceptance (i.e., A answers) rates above 60% for 17/28 items (cells in light grey in Table 9, 60.7%), and at least conditioned acceptance (A + B answers) remains above 75% for all the 17 items above; moreover, only 3 items (10,7%) are under the 60% threshold in this second column. Of course, acceptability judgements still concern passive competence only. Nothing certain can be said about the extent of overabundance as part of the active competence of each speaker, although answer A may suggest that the informant would also produce both alternatives (at least according to his/her own evaluation).

Not unexpectedly, there appears to be also a good correlation between the ranking of intra-speaker (at least passive) overabundance and the strength of overabundance at the community level. All but one instance of (very) weak community overabundance as resulting from Table 7 are placed in the lowest quarter of the ranked list in Table 9 (the dark grey cells). The only exception is *poggiachitarra* 'guitar stand', which is strongly oriented towards N:SG at the community level (*i poggiachitarra* 'the:M.PL lean + guitar:SG), but still displays a high acceptance of its alternative *i poggiachitarre* 'the: M.PL lean + guitar:PL'. At any rate, even for the lowest ranked items in the table the informants show a relatively high tolerance towards the minority alternative. For an item like *puliscilingua* 'tongue-cleaner', just 6 speakers chose the plural N-marking (*i puliscilingue*) in Q1; nevertheless, still about a third of the Q2 informants, when this alternative was presented to them, accepted it without problems, and another third judged it to be worse, but nonetheless acceptable.

Finally, an interesting observation can be made about the correlation of overabundance with one of the factors considered above for both Web data and Q1 data, namely the degree of entrenchment of the compound (cf. Tables 5 and 7): the acceptability of the alternative options tends to rank higher for items which can be evaluated as being less entrenched in the lexicon (those highlighted in bold in Table 9). Perhaps intra-speaker overabundance is more frequent in less entrenched compounds, because they are more easily created from scratch as opposed to simply being retrieved from the mental lexicon. The very same effect, although concerning only overabundance *in the singular* and evaluated across all semantic classes of VN compounds, has been observed by von Heusinger & Schwarze (2013: 333–334): overabundance was very limited for their sample of 1350 "lexicalized" items (only 3%), but much more relevant for their "non-lexicalized" sample of 100 items (26%): significantly, they concluded that "original variation may be blocked by lexicalization and by subsequent pattern formation". These results suggest very neatly that overabundance is a very productive phenomenon in Italian VN compounds, since productivity is best tested on new formations, which give a picture of a word forming process "in real time".

Table 9: Intra-speaker overabundance: acceptability judgements compared with community data

main VNC compound	n. of relevant answers	full intra-speaker overabundance ("OK" only)	partial intra-speaker overabundance ("OK"+"acceptable")	community overabundance
appendiborsa/e	75	78.7%	92.0%	balanced
portaspazzolone/i	56	78.6%	92.9%	strong
reggimensola/e	40	77.5%	90.0%	s
portaspazzolino/i	86	76.7%	90.7%	b
copriob(b)iettivo/i	90	76.7%	87.8%	b
copriteiera/e	48	75.0%	93.8%	s
copricatena/e	62	74.2%	85.5%	s
appendicasco/-chi	74	73.0%	93.2%	s
copripoltrona/e	95	70.5%	85.3%	s
poggiachitarra/e	59	69.5%	88.1%	weak
bloccavolante/i	28	67.9%	82.1%	s
segnaposto/i	101	64.4%	80.2%	s
portagiornale/i	41	63.4%	87.8%	b
coprisellino/i	128	62.5%	76.6%	b
copridivano/i	111	61.3%	75.7%	s
portauovo/a	102	60.8%	78.4%	b
cacciavite/i	119	60.5%	76.5%	b
segnalibro/i	117	53.8%	73.5%	s
coprileto/i	52	51.9%	61.5%	s
portapenna/e	90	51.1%	68.9%	b
copricostume/i	71	49.3%	71.8%	s
salvaschermo/i	33	48.5%	63.6%	w
tappanaso/i	103	47.6%	71.8%	w
bloccasterzo/i	47	46.8%	68.1%	s
poggiatesta/e	111	44.1%	58.6%	w
coprisella/e	24	41.7%	58.3%	w
puliscilingua/e	53	35.8%	64.2%	w
scaldacollo/i	57	24.6%	45.6%	w

bold italics: less entrenched items

6. Conclusions

In this paper we have investigated a particular case of overabundance in Italian nominal inflection, namely the plural cell of a subclass of VN compounds (among those displaying singular N-marking in the singular), which is definable with semantic criteria: the union of what we called "stable-N" compounds and those "variable-N" compounds which have only N:SG in the singular. Our data show that this subclass displays systematic overabundance. This has been verified both at the level of the global speakers' community (at least if Web data can be considered an approximation of it), and at the level of a micro-community, where externally conditioned overabundance due to geo-

sociolinguistic factors should be much reduced: namely, a 200-person class of university students in Turin, who answered a naming questionnaire based on visual input.

Moreover, a further questionnaire submitted to the same informants has shown a fairly high amount of overabundance also at the level of individual speakers' – at least passive – competence, because most informants who chose a form for the plural of a stable-N compound also fully accepted the opposite alternative when confronted with it a week later.

Finally, this kind of overabundance in Italian morphology can plausibly be taken as productive, because in Italian the stable-N compounds with Instrument/Agent meaning are an open class, and overabundance occurs also with formations minimally entrenched in the lexicon. Therefore, it seems safe to hypothesize that for this subclass of items, a considerable number of speakers acquire both pluralizing strategies at the very moment of their acquisition in the mental lexicon.

From a more general, theoretical point of view, this case study might point out the importance of taking into account the instances of systematic and productive overabundance, which in recent approaches are perhaps downplayed as particularly non-canonical instances of the phenomenon.

Appendix: complete list of the Web VNC corpus discussed in §4.2

NB: The compounds highlighted in bold are those considered to be "less entrenched" in the counts reported in Table 5.

1. Stable-N compounds:

alzabimbo 'raise + child', ***alzatacco*** 'raise + heel', ***apricancello*** 'open + gate', ***aprifesta*** 'open + party', *apripista* 'open + track' (M/F), ***avvolgifiocco*** 'wind up + jib', ***avvolgilenza*** 'wind up + line', ***avvolgiranda*** 'wind up + mainsail', ***avvolgitubo*** 'wind up + tube', ***battiporta*** 'hit + door', *battiscopa* 'hit + broom', ***bloccadisco*** 'block + disc', *bloccaruota* 'block + wheel', *bloccasterzo* 'block + steering wheel', ***caricacellulare*** 'charge + mobile phone', ***caricatelefono*** 'charge + (mobile) phone', ***cavalcaferrovia*** 'overpass + railway', *cavalcavia* 'overpass + road', ***chiudibocca*** 'close + mouth', *chiudifila* 'close + line' (M), *chiudilettera* 'close + letter', ***chiudipacco*** 'close + package', *chiudiporta* 'close + door', *copriasse* 'cover + board', *copribara* 'cover + coffin', ***copricalorifero*** 'cover + radiator', *copricanna* 'cover + top tube', *copricapo* 'cover + head', ***copricasco*** 'cover + helmet', *copricatena* 'cover + chain', *copricuscino* 'cover + pillow', *copridivano* 'cover + couch', ***copriforno*** 'cover + oven', ***coprifronte*** 'cover + forehead', *coprigiunto* 'cover + expansion joint', *coprilavatrice* 'cover + washing machine', *coprileto* 'cover + bed', ***coprimanubrio*** 'cover + handlebar', *coprimaterasso* 'cover + mattress', *coprimozzo* 'cover + hub', ***coprimuro*** 'cover + wall', ***copriobiettivo*** 'cover + lens', ***copripasseggino*** 'cover + stroller', *copripiumino* 'cover + duvet:DIM', ***copripiumone*** 'cover + duvet:AUG', ***copripoltrona*** 'cover + armchair', ***copriradiatore*** 'cover + radiator', ***coprirete*** 'cover + bed base', *copriruota* 'cover + wheel', ***coprisedia*** 'cover + chair', *coprisella* 'cover + saddle', ***coprisellino*** 'cover + saddle:DIM', *copritastiera* 'cover + keyboard', ***copritavola*** 'cover + table', *copritavolo* 'cover + table', ***copriteiera*** 'cover + teapot', ***copritermosifone*** 'cover + radiator', ***coprivaso*** 'cover + vase', ***coprivolante*** 'cover + steering wheel', ***ferma(a)nello*** 'block + ring', ***fermadeviatoio*** 'block + (railroad)switch', ***fermaguaina*** 'block + sheath', ***fermascambio*** 'block + switch', *ficcanaso* 'stick + nose' (M/F), *guardaparco* 'watch + park' (M/F), *guardasala* 'watch + room' (M), ***marcadavanzale*** 'mark + windowsill', *marcapiano* 'mark + floor', ***paracamino*** 'protect + fireplace', *paraculo* 'protect + ass' (M), ***paraglomo*** 'protect + hoof', *paralume* 'protect + lamp', ***passanastro*** 'pass + ribbon', ***passaparete*** 'pass + wall', ***poggianaso*** 'lean + nose', *portabandiera* 'carry + flag' (M/F), *portacasco* 'carry + helmet', ***portacatino*** 'carry + basin', *portainnesto* 'carry + graft', ***portapatente*** 'carry + driving licence', ***portascopino*** 'carry + toilet brush', *portavoce* 'carry + voice' (M/F), *prestanome* 'lend + name' (M/F), ***prestavoce*** 'lend + voice' (M), ***proteggilama*** 'protect + blade',

proteggischermo 'protect + screen', **puliscilingua** 'clean + tongue', *reggicoda* 'hold + tail', **reggifilo** 'hold + cord', **reggimensola** 'hold + shelf', *reggipetto* 'hold + breast', **reggirullante** 'hold + snare drum', *reggisella* 'hold + saddle', *reggiseno* 'hold + breast', **reggitenda** 'hold + curtain', **reggitubo** 'hold + tube', **rollafiocco** 'roll + jib', *rompigetto* 'break + jet', **rompitratta** 'interrupt + length', **salvamotore** 'save + engine', **salvapercussore** 'save + firing pin', *salvaschermo* 'save + screen', *scaldabagno* 'warm + bathroom', *scaldacollo* 'warm + neck', *scaldaletto* 'warm + bed', *scendibagno* 'get out + bath tub', *scendiletto* 'get out + bed', *segnagusto* 'mark + (ice cream) flavour', **segnalimite** 'mark + limit', *segnaposto* 'mark + place', *segnaprezzo* 'mark + price', **segnatavolo** 'mark + table', *segnavia* 'mark + way', **stringinaso** 'squeeze + nose', **tappanaso** 'plug + nose', *tendicatena* 'stretch + chain', **tendicinghia** 'stretch + belt', **tendicintura** 'stretch + seatbelt', *tendicollo* 'stretch + collar', **tendifilo** 'stretch + cord', **tendireggia** 'stretch + strap', *tergicristallo* 'wipe + windscreen', *tergilunotto* 'wipe + rear window', **tiracatena** 'pull + chain'.

2. Variable-N compounds with singular N-marking in the singular cell:

alzasedia 'raise + chair', *appoggiatesta* 'lean + head', **apribocca** 'open + mouth', **battilastra** 'hit + slab', *battipalo* 'hit + pole', *battipenna* 'hit + plectrum', *battipista* 'hit + track', *battistrada* 'hit + road', *cacciavite* 'thrust + screw', **cercafase** 'search + phase', *copricostume* 'cover + swimsuit', **copritovaglia** 'cover + tablecloth', **coprizaino** 'cover + backpack', *girarrosto* 'turn + roast', *giravite* 'turn + screw', *guidafilo* 'lead + cord', **infilaaghi** 'thread + needle', *lavatesta* 'wash + head', *leccaculo* 'lick + ass' (M), **mettifoglio** 'put + sheet', *parafulmine* 'protect + lightning', *parapetto* 'protect + chest', **pesafiltro** 'weight + filter', **poggiacapo** 'lean + head', **poggiaschiena** 'lean + back', *poggiatesta* 'lean + head', **portapranzo** 'carry + lunch', *reggispinta* 'hold + pressure', *salvavita* 'save + life', **scaldarancio** 'warm + meal', *segnalibro* 'mark + book', *spazzacamino* 'sweep + chimney' (M), **stringitubo** 'tighten + tube'.

References

Aronoff, Mark. 1994. *Morphology by itself: Stems and inflectional classes.* Cambridge, MA: MIT Press.

Baroni, Marco & Bernardini, Silvia & Ferraresi, Adriano & Zanchetta, Eros. 2009. The WaCky wide Web: A collection of very large linguistically processed Web-crawled corpora. *Language Resources and Evaluation* 43(3). 209–226.

Bermel, Neil & Knittl, Luděk. 2012. Morphosyntactic variation and syntactic constructions in Czech nominal declension: Corpus frequency and native-speaker judgments. *Russian Linguistics* 36. 91–119.

Bisetto, Antonietta. 1999. Note sui composti VN dell'italiano. In Benincà, Paola & Mioni, Alberto & Vanelli, Laura (eds.), *Fonologia e morfologia dell'italiano e dei dialetti d'Italia. Atti del XXXI Congresso internazionale di studi della Società di Linguistica Italiana*, 505–538. Roma: Bulzoni.

Brown, Dunstan & Chumakina, Marina & Corbett, Greville G. (eds.). 2013. *Canonical morphology and syntax.* Oxford: Oxford University Press.

Corbett, Greville G. 2005. The canonical approach in typology. In Frajzyngier, Zygmunt & Hodges, Adam & Rood, David S. (eds.), *Linguistic diversity and language theories* (Studies in language companion series 72), 25–49. Amsterdam: Benjamins.

Corbett, Greville G. 2007. Canonical typology, suppletion, and possible words. *Language* 83(1). 8–41.

De Mauro, Tullio (ed.). 1999. *Grande dizionario italiano dell'uso.* Torino: UTET.

De Mauro, Tullio (ed.). 2000. *Il dizionario della lingua italiana. Con cd-rom.* Torino: Paravia.

Ferrari-Bridgers, Franca. 2005. Italian [VN] compound nouns: A case for a syntactic approach to word formation. In Twan, Geerts & van Ginneken, Ivo & Jacobs, Haike (eds.), *Romance languages and linguistic theory 2003*, 63–80. Amsterdam: Benjamins.

Fletcher, William H. 2012. Corpus analysis of the World Wide Web. In Chapelle, Carol A. (ed.), *Encyclopedia of Applied Linguistics*, 1339–1347. Oxford: Wiley-Blackwell.

Gather, Andreas. 2001. *Romanische Verb–Nomen–Komposita: Wortbildung zwischen Lexikon, Morphologie und Syntax*. Tübingen: Narr.

Gatto, Maristella. 2014. *Web as corpus: Theory and practice*. London: Bloomsbury.

Goidanich, Pier Gabriele. 1918. *Grammatica italiana ad uso delle scuole. Con nozioni di metrica, esercizi e suggerimenti didattici.* Bologna: Zanichelli.

Hathout, Nabil & Montermini, Fabio & Tanguy, Ludovic. 2008. Extensive data for morphology: using the World Wide Web. *French Language Studies* 18. 67–85.

Keller, Frank & Lapata, Mirella. 2003. Using the Web to obtain frequencies for unseen bigrams. *Computational Linguistics* 29. 459–484.

Kilgarriff, Adam. 2001. Web as corpus. In Rayson, Paul & Wilson, Andrew & McEnery, Tony & Hardie, Andrew & Khoja, Shereen (eds.), *Proceedings of the Corpus Linguistics Conference (CL 2001)*, 342–344. (University Centre for Computer Research on Language Technical Papers, Vol. 13, Special Issue, Lancaster University (UK), 29 March-2 April 2001).

Kilgarriff, Adam & Grefenstette, Gregory. 2003. Introduction to the special issue on Web as corpus. *Computational Linguistics* 29(3). 333–347.

Lečić, Dario. 2015. Morphological doublets in Croatian: The case of the instrumental singular. *Russian Linguistics* 39. 375–393.

Lew, Robert. 2009. The Web as corpus versus traditional corpora: Their relative utility for linguists and language learners. In Baker, Paul (ed.), *Contemporary corpus linguistics*, 289–300. London: Continuum.

Lüdeling, Anke & Evert, Stefan & Baroni, Marco. 2007. Using Web data for linguistic purposes. In Hundt, Marianne & Nesselhauf, Nadja & Biewer, Carolin (eds.), *Corpus linguistics and the Web*, 7–23. Amsterdam: Rodopi.

Micheli, Maria Silvia. 2016. Sul plurale delle parole composte nell'italiano contemporaneo. *Studi di lessicografia italiana*, 33. 229–256.

Montermini, Fabio. 2015. Regole (e irregolarità) nella formazione delle parole. In Grandi, Nicola (ed.), *La grammatica e l'errore: Le lingue naturali tra regole, loro violazioni ed eccezioni*, 63–83. Bologna: Bononia University Press.

Moyna, Maria Irene. 2011. *Compound words in Spanish: Theory and history*. Amsterdam: Benjamins.

Nakov, Preslav & Hearst, Marti. 2005. A study of using search engine page hits as a proxy for n-gram frequencies. In Angelova, Galia & Bontcheva, Kalina & Mitkov, Ruslan & Nicolov, Nicolas & Nikolov, Nikolai (eds.), *International Conference Recent Advances in Natural Language Processing Proceedings*, 347–353. (Borovets, Bulgaria 21-23 September 2005.)

Rainer, Franz. 1993. *Spanische Wortbildungslehre*. Tübingen: Niemeyer.

Rainer, Franz & Varela, Soledad. 1992. Compounding in Spanish. *Rivista di Linguistica* 4(1). 117–142.

Ricca, Davide. 2005. Al limite tra sintassi e morfologia: I composti aggettivali V-N nell'italiano contemporaneo. In Grossmann, Maria & Thornton, Anna M. (eds.), *La formazione delle parole: Atti del XXVII congresso internazionale di studi della Società di Linguistica Italiana (SLI): L'Aquila, 25-27 settembre 2003*, 465–486. Roma: Bulzoni.

Ricca, Davide. 2010. Corpus data and theoretical implications: With special reference to Italian V-N compounds. In Scalise, Sergio & Vogel, Irene (eds.), *Cross–disciplinary issues in compounding*, 167−175. Amsterdam: Benjamins.

Ricca, Davide. 2015. Verb-Noun compounds in Romance. In Müller, Peter O. & Ohnheiser, Ingeborg & Olsen, Susan & Rainer, Franz (eds.), *Word-formation. An international handbook of the languages of Europe*, 688–707. Berlin: de Gruyter.

Rosenberg, Maria. 2011. Les composés français VN — aspects sémantiques. *Revue Romane* 46(1). 69 –88.

Scalise, Sergio. 1994. *Morfologia.* Bologna: Il Mulino.

Serianni, Luca & Alberto Castelvecchi. 1988. *Grammatica italiana: Italiano comune e lingua letteraria.* Torino: UTET.

Stump, Gregory. 2016. *Inflectional paradigms: Content and form at the syntax–morphology interface.* Cambridge: Cambridge University Press.

Thornton, Anna M. 2011. Overabundance (multiple forms realizing the same cell): A non–canonical phenomenon in Italian verb morphology. In Maiden, Martin & Smith, John Charles & Goldbach, Maria & Hinzelin, Marc–Olivier (eds.), *Morphological autonomy: Perspectives from Romance inflectional morphology,* 358–381. Oxford: Oxford University Press.

Thornton, Anna M. 2012. Reduction and maintenance of overabundance: A case study on Italian verb paradigms. *Word Structure* 5(2). 183–207.

Thornton, Anna M. In press. Overabundance: A canonical typology. In Rainer, Franz & Gardani, Francesco & Dressler, Wolfgang U. & Luschützky, Hans Christian (eds.), *Competition in inflection and word-formation.* Berlin: Springer.

Val Álvaro, José Francisco. 1999. La composición. In Bosque, Ignacio & Demonte, Violeta (eds.), *Gramática descriptiva de la lengua española*, vol. 3, 4757−4841. Madrid: Espasa.

Villoing, Florence. 2009. Les mots composés VN. In Fradin, Bernard & Kerleroux, Françoise & Plénat, Marc (eds.), *Aperçus de morphologie du français,* 175−198. Saint-Denis: Presses Universitaires de Vincennes.

von Heusinger, Klaus & Schwarze, Christoph. 2013. Italian V + N compounds, inflectional features and conceptual structure. *Morphology* 23(3). 325–350.