

NOUN-CLAD ADJECTIVES ON THE ADJECTIVAL STATUS OF NON-HEAD CONSTITUENTS OF ITALIAN ATTRIBUTIVE COMPOUNDS

NICOLA GRANDI

MALVINA NISSIM

FABIO TAMBURINI

ABSTRACT: In Italian appositive compounds like *parola chiave* ‘keyword’, the non-head constituent (N2) often undergoes a metaphorical interpretation and behaves like an adjective, emphasising a property of the head (N1). The main research question of the work described in this paper is: Are these really noun-clad adjectives? Specifically: (i) Do N2s “formally” show a morpho-syntactic behaviour typical of adjectives? (ii) Do N2s “semantically” show a selection of features that are normally associated with real adjectives modifying N1 (e.g. *riunione lunga* lit. meeting long ‘extremely long meeting’ and *riunione fiume* lit. meeting river ‘a never-ending meeting’)? We have designed some corpus-based studies to address these questions, and we observed that (i) “formally”, N2s do share some of the typical properties of adjectives, but not all, thus implying gradience in the definition of part-of-speech classes; (ii) “semantically”, N2s do behave like adjectives and the implied sense of N2 can be obtained via objective criteria, exploiting corpus data.

KEYWORDS: compounding, parts of speech, adjective, corpus evidence.

1. INTRODUCTION

Compounds have been classified according to different parameters: the syntactic relation between the members, the presence/absence of a head, the nature of linguistic items involved in compounding processes, etc. (see Bauer, 2001 and Montermini, 2010 for an exhaustive survey). Interestingly, all existing classifications of compounds (and irrespective of parameters adopted for classification) feature a class characterised by a modification relation between the constituents of a compound. In recent literature on compounding, this class has been named “attributive/appositive” (ATAP henceforth) (Scalise & Bisetto, 2009). In ATAP compounds one member acts as modifier of the other (usually a noun), thus performing the function usually ascribed to property words:

- (1) English *redskin*
Italian *pesce palla* lit. fish ball ‘globefish’

According to Bauer (2001: 697), such class of compounds covers the vast majority of compound words, and is a “quasi universal”: “this is the majority pattern for compounds in the languages of the world, and there are very few languages which do not have compounds of this type”.

The distinctive feature of this kind of compounds is usually identified in the metaphorical (re)interpretation of the modifying constituent. Bisetto & Scalise, for instance, describe ATAP compounds as formations “where the non-head very often is used somehow metaphorically, expressing an attribute of the head” (2005: 327); the same authors also states that “[i]n appositives [...] the noun plays an attributive role and is often to be interpreted metaphorically” (Scalise & Bisetto, 2009: 52).

However, through a deeper survey of the data, we can observe that the metaphorical (re)interpretation of the non-head constituent is not a homogeneous phenomenon. As already pointed out by Bisetto & Scalise, it is not observed at all in compounds in which the modifying constituent is a typical property word, i.e. an adjective (we will come back later to the issue of defining the core value of parts of speech). In Italian ATAP compounds such as:

- (2) *pellerossa* ‘redskin’
bassorilievo ‘bas-relief’
gentildonna ‘gentlewoman’

the modifying members (the adjectives *rossa* ‘red’, *basso* ‘low/bass’ and *gentil(e)* ‘gentle’) do not undergo any metaphorical reinterpretation. So, we might assume that metaphorical reinterpretation is a specific feature of a subclass of ATAP compounds, namely those whose second member is not inherently a property word. Moreover, even when a metaphorical reinterpretation exists, different degrees can be observed. If we focus on NN endocentric ATAP compounds, such as *pesce palla* ‘globefish’ or *parola chiave* ‘keyword’, we see that the meaning of the non-head constituent usually undergoes a process of semantic bleaching, through an isolating abstraction, which separates out one particular property or feature of that concept (Heine, Claudi & Hünemeyer, 1991: 43). For example, the Italian *pesce palla* ‘globefish’ is a compound where the modifier *palla* ‘ball’ provides the shape (the isolated feature of the concept *palla*, in Heine, Claudi & Hünemeyer’s terms) for the referent of the head noun (a fish). Similarly, in *riunione fiume* (lit. river meeting) ‘extremely long meeting’, the contextual property of the head noun (meeting) is obtained through the isolation of one specific feature of the modifier (river), namely the very long duration, probably deriving from the constant streaming of water in a river. Nevertheless, in compounds such as *pesce palla* ‘globefish’ or *pesce spada* ‘swordfish’, usually mentioned as prototypical instances of appositive compounds in the literature, the metaphorical (re)interpretation of the non-head is very weak, since a *pesce*

spada is a fish characterized by a long, flat bill, that looks like a sword. And a *pesce palla* has unequivocally the shape of a ball. In these cases, the feature of the meaning of *palla* (that is, being round) that has been singled out (“isolated” in the terms of Heine, Claudi & Hünemeyer 1991) is the core or nucleus characteristic of that concept, and, as a consequence, it characterises all the items that instantiate it.

Let us now consider a compound as *riunione fiume*. We all know what is a *fiume* ‘river’; and we all know what is a *riunione* ‘meeting’. What is really puzzling is the meaning of the whole compound: it designates a very long, never-ending, and often boring meeting. The semantic feature ‘long’ (with its consequent nuances of ‘never-ending’ and ‘boring’) is not the “nucleus characteristic” of the concept ‘river’, since not all rivers are long by default, though they all stream endlessly. So, while *palla* in the above mentioned compound conveys more or less the same meaning it has when used as an independent word, *fiume* in a compound pattern displays a meaning which is different from the one it has as an autonomous word. Thus, whereas the link between *palla* and *pesce palla* is fully transparent, the link between *fiume* and *riunione fiume* is instead less evident.

So, in some endocentric ATAP compounds made up of two nouns, the non-head constituent retains most of its primary connotation: in *pesce spada* ‘swordfish’ the second noun designates an entity which indeed resembles a sword. On the contrary, in other cases, when a metaphorical reinterpretation of the non-head member takes place, the second noun loses all its referential power, becoming a mere property word, and performing a plain modification function.

Moreover, what distinguishes compounds such as *riunione fiume* from other appositive compounds as *pesce palla* is that *riunione fiume* belongs to a family of compound words, in which the modifier is always the same, it always undergoes the same metaphorical shift:

- (3) $[[X]_N \textit{fiume}]_N$
fiume ‘river’ > ‘long and boring’
 X = *riunione* ‘meeting’, *interrogatorio* ‘examination’, *processo* ‘trial, lawsuit’, *discorso* ‘speech’, *discussione* ‘discussion’, etc.

Such a situation is all but rare in appositive compounds whose non-head constituent has undergone a strong metaphorical reinterpretation, while it is quite unusual in other compounding patterns (including ATAP compounds without metaphorical reinterpretation of the non-head constituent: in Italian no compound other than *pesce spada* is attested with *spada* ‘sword’). Other examples follow below:

- (4) a. $[[X]_N \textit{bomba}]_N$
bomba ‘bomb’ > ‘sensational’
 X = *notizia* ‘piece of news’, *intervista* ‘interview’, *rivelazione* ‘revelation’, etc.

- b. $[[X]_N \textit{lampo} \textit{'lightning'}}]_N$
lampo 'lightning' > 'quick, istantaneous'
 X = *processo* 'trial, lawsuit', *guerra* 'war', *viaggio* 'journey', *operazione* 'operation, action', etc.
- c. $[[X]_N \textit{chiave}]_N$
chiave 'key' > 'crucial'
 X = *parola* 'word', *ruolo* 'role', *concetto* 'concept', *elemento* 'element', *posizione* 'position', etc.
- d. $[[X]_N \textit{ombra}]_N$
ombra 'shadow' > 'parallel, alternative'
 X = *governo* 'government', *ministro* 'minister', *senatore* 'senator', *presidente* 'president', *Cancelliere* 'Chancellor', etc.

Furthermore, the occurrence of such words is strictly restricted to the non-head position of endocentric ATAP compounds. But it is well known that in Romance languages there are no constraints on the position of a word within a compound (cf. e.g. *ingegnere capo* 'head engineer' vs. *capoufficio* 'head clerk') and, in a wide typological perspective, a rigidly fixed position within a linguistic complex form is typically an affixal property.

Although the phenomenon is well-known, widely discussed in the literature and at first sight appears to be quite trivial, nearly all studies of this compounding pattern are based essentially on subjective judgments and are rarely supported by corpus-based evidence. Thus, some issues still require further, more objective investigation. More specifically, we believe the following questions need to be addressed:

- if the feature of the non-head which is projected to the head is not the "nucleus characteristic" of the concept, are there objective parameters which justify that choice and which make it foreseeable?
- is there a link between these parameters and the meaning of the nouns involved when used as autonomous words, i.e. outside a compound pattern?

We aim at answering these questions by focusing on Italian endocentric NN appositive compounds and exploiting corpus evidence to obtain the implied sense of non-head constituents while at the same time constraining our subjective judgement in its determination.

2. TYPOLOGICAL FRAMEWORK

In the typological approach to the often debated issue of delimiting word classes, "noun, verb and adjective are not categories of particular languages" (Croft, 2000: 65), but are "language universals – that is, there are typological prototypes [...] which should be called noun, verb and adjective" (Croft, 2000: 65). In this picture, actual constructions are the primitive elements of

syntactic representation and categories are derived from them: it is a sample of different constructions with a common function that defines the boundaries of a category or of a part of speech. Functions are cross-linguistically universal or, at least, recurrent. Single constructions are language-specific. So, typological comparison will sketch a pattern of variation and every single language will fit somewhere in this pattern of variation. In this framework word classes are described as language universals or, better, as typological prototypes determined by the simultaneous occurrence of some semantic classes and some pragmatic functions:

| | | PRAGMATIC FUNCTIONS | | |
|------------------|------------|--|--|-------------------------------|
| | | REFERENCE | MODIFICATION | PREDICATION |
| SEMANTIC CLASSES | OBJECTS | unmarked nouns | genitive, PPs on nouns, adjectivalisations | predicate nominals, copulas |
| | PROPERTIES | deadjectival nouns | unmarked adjectives | Predicate adjectives, copulas |
| | ACTIONS | Action nominals, complements, infinitives, gerunds | participles, relative clauses | unmarked verbs |

TABLE 1. PROTOTYPICAL NOUNS, ADJECTIVES, AND VERBS (CROFT, 1991: 67).

What fills the cells in Table 1 represents a generalisation on how single languages encode the corresponding pragmatic function and semantic class. So, in a wide cross-linguistic perspective, every linguistic item that performs a modification by property (that is, a descriptive function) can be labelled as adjective, independently of its formal behaviour, since each single language will devote different formal means to the expression of these prototypical functions. In such a picture, both cases of multiple class membership and occasional occurrence of linguistic items or constructions in atypical roles cannot be excluded. The crucial assumption of this framework, known as “typological theory of markedness”, is that linguistic items or constructions in atypical roles never have a higher number of morphemes, of inflectional distinctions, and of contexts of occurrence than typical members of the same class.

In the data we are discussing, we observe the occurrence of a noun in an atypical role. Nouns typically have a referential function (see Table 1). However, in the Italian compounds we are discussing here, such as *riunione fiume*, *parola chiave* or *viaggio lampo* (lit. journey lightning) ‘a (very) short journey’, the second constituent (N2) emphasises a property of the head noun (N1), thus playing an adjectival role: in *riunione fiume*, N2 does not designate a river, but it performs a modification function, indicating a feature that helps to establish the identity of the referent of N1.

In the framework sketched above, when a linguistic item moves towards a typological prototype, it tends to acquire the formal (i.e. structural, inflectional,

and distributional) properties associated with this prototype in each single language. Therefore, a noun encoding the pragmatic function of modification by property will naturally shift towards the class of unmarked adjectives, thereby tending to acquire their formal possibilities. As said, the formal behaviour of parts of speech depends on the typological shape of each single language. Thornton (2004: 528) lists six typical properties of Italian adjectives:

- (5) Number agreement with the head noun
- Gender agreement with the head noun
- Comparative form
- Superlative form
- Adverb formation with *-mente*
- Occurrence in prenominal position

So, we can add a research question to the two listed at the end of section 1: is the nominal nature of non-head members of ATAP compounds still prevalent even if they perform an adjectival function, or do they acquire typical adjectival properties, as a consequence of “playing” the adjective role? We will try to answer this question, as well as the previous ones, exploiting corpus evidence.

3. CORPUS-BASED EXPERIMENTS

The corpus-based analyses that we carried out were aimed at a more objective and throughout characterisation of the non-head members of ATAP compounds with respect to two aspects, directly related to the questions we posed at the end of the introduction and at the end of the previous section. The former has to do with which (possibly metaphoric) semantic feature is singled out for the contextual description of the head noun; the latter regards the formal aspect concerning their acquired degree of adjectivehood.

For our study, we selected a few nouns that typically occur as non-head members in the compounds described above, namely *cardine* ‘hinge’, *chiave* ‘key’, *fiume* ‘river’, *lampo* ‘lightning’, etc., and observed their behaviour when featuring in relevant constructions in the two corpora of Italian which we have selected as our sources of empirical data: a balanced one (CORIS/CODIS), especially used for the investigation of semantic features, and a larger but less controlled one (ItWac), which we exploited for finding possibly rarer but attested interesting expressions from a structural, inflectional, and distributional perspective.

CORIS/CODIS (Rossini Favretti, Tamburini & De Santis, 2002) – henceforth CORIS – is an electronically-based, balanced, general reference corpus of contemporary written Italian (1980-2007), freely available for research purposes, containing authentic texts in their full size. It contains 120 million words and is updated every three years by means of a monitor

corpus. It consists of a collection of authentic and commonly occurring texts in electronic format chosen by virtue of their representativeness of contemporary Italian. The corpus is PoS-tagged and lemmatised (Tamburini, 2007). The other corpus used for our experiments, ItWaK (Baroni *et al.*, 2009), is a huge, web-crawled PoS-tagged and lemmatised corpus composed of 1.5 billion tokens.

3.1 Semantic features

Although the interpretation of *cardine*, *chiave*, *fiume*, and *lampo*, etc. as non-head constituent in the ATAP compounds we are discussing here is rather clear for a native speaker of Italian, we exploited corpus evidence to obtain the implied (shifted) sense, thereby allowing for more than one interpretation to emerge and, at the same time, constraining our subjective judgement in its determination. One crucial assumption in our procedure is that we are dealing with a “family” of compounds, rather than isolated or idiosyncratic cases and we can therefore exploit generalisations (see (3) and (4)).

The experiment we describe involves 3 steps in data extraction and analysis:

1. For each noun under investigation, frequently used in Italian as non-head of appositive compounds, we extracted from CORIS all “N1 N2” compounds where this noun (for instance *cardine*) features as N2, and collected a list of all N1s (for *cardine*, for instance, we get *punto* ‘point’, *principio* ‘principle’, *elemento* ‘element’, ...). Table 2 reports some examples together with frequencies of extracted compounds.

| N2 | <i>cardine</i> ‘hinge’ | <i>chiave</i> ‘key’ | <i>fiume</i> ‘river’ |
|----|---|---|--|
| N1 | <i>punto</i> 47 (<i>punto cardine</i>) ‘point’ ‘crucial aspect’ | <i>parola</i> 548 (<i>parola chiave</i>) ‘word’ ‘keyword’ | <i>riunione</i> 8 (<i>riunione fiume</i>) ‘meeting’ ‘very long meeting’ |
| | <i>principio</i> 20 (<i>principio cardine</i>) ‘principle’ ‘key principle’ | <i>punto</i> 130 (<i>punto chiave</i>) ‘point’ ‘key point’ | <i>omonimo</i> 8 (<i>omonimo fiume</i>) ‘homonym’ ‘homonymic river’ |
| | <i>elemento</i> 14 (<i>elemento cardine</i>) ‘element’ ‘key element’ | <i>ruolo</i> 105 (<i>ruolo chiave</i>) ‘role’ ‘key role’ | <i>via</i> 5 (<i>via fiume</i>) ‘way/via’ ‘through the river’ |
| | ... | ... | ... |

TABLE 2. EXAMPLE OF N1 N2 APPOSITIVE COMPOUNDS EXTRACTED FROM CORIS AND THEIR FREQUENCY

2. For each N2 (e.g. *cardine*), for all the N1s collected in step 1 (for *cardine*: *punto*, *principio*, etc.) we searched CORIS for adjectives (ADJ) associated with them, allowing for the small variety of syntactic constructions outlined by the following patterns (CONJ_C stands for coordinating conjunction, ADV for adverb and the square brackets mark optional elements):

- N1 [ADV] ADJ
- N1 [ADV] ADJ CONJ_C ADJ
- ADJ N1
- ADJ CONJ_C ADJ N1

This search provided us with a list of adjectives used in the contexts of the N1s for each of the N2s we considered. In the case of *cardine*, for instance, Table 2 shows that frequent N1s are *punto*, *principio*, *elemento*. For each of these, the adjectival patterns sketched above returned a set of adjectives. The most frequent ones are listed in Table 3.

| N2 | <i>cardine</i> | | | | | |
|-----|--|--|--|--|--|-----|
| N1 | <i>punto</i> | | <i>principio</i> | | <i>elemento</i> | ... |
| ADJ | <i>percentuale</i> 551 'percentage' | | <i>generale</i> 225 'general' | | <i>essenziale</i> 377 'essential' | |
| | <i>debole</i> 404 'weak' | | <i>attivo</i> 151 'active' | | <i>fondamentale</i> 321 'fundamental' | |
| | <i>fermo</i> 342 'steady' | | <i>fondamentale</i> 148 'fundamental' | | <i>nuovo</i> 310 'new' | |
| | <i>decisivo</i> 254 'decisive' | | <i>costituzionale</i> 93 'constitutional' | | <i>costitutivo</i> 233 'constituting' | |

TABLE 3. EXAMPLE OF ADJECTIVES (ADJ) USED IN THE CONTEXTS OF N1S CONNECTED WITH THE N2 *CARDINE*.

The rationale behind this step is that the set $A = \{ADJ_1, \dots, ADJ_n\}$ formed by those adjectives occurring “most prominently with most N1s” would be indicative of the implied, used sense of N2. Determining in a precise way the essence of “most prominent” adjectives co-occurring with the various N1s is a matter of designing a proper set of measures to rank all the retrieved adjectives.

3. We developed a set of statistical/stochastic measures to obtain a scoring model able to capture regularities and predominant behaviours between the actors we considered. In this work we ranked the obtained adjectives considering the contribution of three kinds of indicators:

a) Frequency: the first measure aims at capturing the co-occurrence rate between the three actors, namely N2, the N1s and the ADJs, into a stochastic model using the information connected with the co-occurrence frequency between N1-N2 and N1-ADJ. If we consider the two events

$$\begin{aligned}
 X1 &= \text{co-occurrence between N1 and N2} \\
 X2 &= \text{co-occurrence between N1 and ADJ}
 \end{aligned}$$

the probability of these two events can be written, following also Figure 1, as:

$$p(X) = p(X1) * p(X2 | X1) = p(N1 | N2) * p(ADJ | N1, N2)$$

and summing over all possible N1s between N2 and ADJ we obtain the definition of the first index we used, a stochastic measure of the connection between a single ADJ and the N2 through all the N1s co-occurring with them,

$$\text{freqm}_{N2}(\text{ADJ}) = \frac{\sum_j p(N1_j | N2)}{\sum_j f(N1_j, N2) / f(N2)} * \frac{p(\text{ADJ} | N1_j, N2)}{f(\text{ADJ}, N1_j) / f(N1)}$$

where f stands for the frequency of occurrence of its arguments.

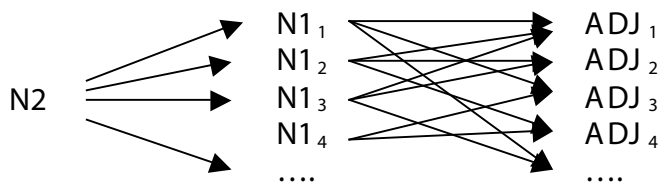


FIGURE 1. VISUALISATION OF THE RELATIONS BETWEEN THE ENTITIES INVOLVED

b) Word Spaces: the second indicator concerns the information connected with the distributional similarities between the N2 and the associated ADJs using a *Word Space Model* (Lenci, 2008). We used a well-known off-the-shelf customisable package, namely *Infomap-NLP* (Widdows, 2004), able to build a Word Space Model (WSM) based on a word co-occurrence matrix inside a fixed context (± 15 words). After a dimensionality reduction (through standard Singular Value Decomposition) each row of this matrix represents a vector in the WSM, and words that exhibit similar distributional behaviour have higher (cosine) similarity in the model.

For designing our second indicator, we built a WSM using the lemmatised and PoS-tagged version of CORIS (see Table 4 for an example) and used the cosine similarity to evaluate a possible meaning overlap between N2 and the connected ADJs, defining the second index as

$$WSM_{N_2}(ADJ) = cosine_similarity(N2, ADJ)$$

| | |
|--|---|
| fiume_NN | <i>rive_NN</i> :0.466725 ‘banks’ |
| <i>lago_NN</i> :0.640202 ‘lake’ | <i>sabbia_NN</i> :0.458300 ‘sand’ |
| <i>riva_NN</i> :0.623108 ‘bank’ | <i>roccia_NN</i> :0.450610 ‘rock’ |
| <i>valle_NN</i> :0.528168 ‘valley’ | <i>acque_NN</i> :0.431660 ‘waters’ |
| <i>fiumi_NN</i> :0.523994 ‘rivers’ | <i>ruscello_NN</i> :0.431298 ‘stream’ |
| <i>ponte_NN</i> :0.518435 ‘bridge’ | <i>diga_NN</i> :0.430327 ‘dam’ |
| <i>torrente_NN</i> :0.505725 ‘torrent’ | <i>collina_NN</i> :0.430273 ‘hill’ |
| <i>mare_NN</i> :0.504719 ‘sea’ | <i>scorreva_V_GVRB</i> :0.428204 ‘streamed’ |
| <i>montagne_NN</i> :0.494485 ‘mountains’ | <i>spiaggia_NN</i> :0.425290 ‘beach’ |

TABLE 4. THE MOST SIMILAR WORDS OF THE NOUN *FIUME* IN THE WORD SPACE MODEL.

c) Spread: the third index is a measure indicating how frequently a single ADJ co-occurs with the various N1s associated with a N2 (*spread*), defined as

$$spread_{N_2}(ADJ) = \sum_{i(N1, ADJ) \neq 0} 1 / \#N1$$

where $\#N1$ stands for the total number of N1s associated with N2. This measure is used both as a general index capable of outlining prominent combinations as well as a threshold to eliminate less relevant combinations. The rationale behind this is that if an adjective tends to occur with many of the extracted N1s, it is quite likely that this adjective is indicative towards the determination of the semantics of N2.

| <i>cardine</i> | | <i>chiave</i> | | <i>fiume</i> | |
|--------------------------------------|----------|--------------------------------------|-----------|--|-----------|
| N1=14 (f>1) ADJs=8791 | | N1=112 (f>1) ADJs=12665 | | N1=7 (f>1) ADJs=5672 | |
| ADJ | score | ADJ | score | ADJ | score |
| <i>fondamentale</i> 'fundamental' | 3.77e-02 | <i>solo</i> 'one and only' | 1.62e+00 | <i>lungo</i> 'long' | 9.24e-06 |
| <i>nuovo</i> 'new' | 1.74e-02 | <i>ultimo</i> 'last' | 1.29e+00 | <i>ultimo</i> 'last' | 2.86e-06 |
| <i>essenziale</i> 'essential' | 1.36e-02 | <i>seguinte</i> 'following' | 2.61e-01 | <i>solo</i> 'one and only' | 2.16e-06 |
| <i>unico</i> 'unique' | 1.15e-02 | <i>unico</i> 'inique' | 2.59e-01 | <i>numeroso</i> 'numerous' | 1.65e-06 |
| <i>importante</i> 'important' | 8.39e-03 | <i>politico</i> 'politic' | 1.41e-01 | <i>privato</i> 'private' | 1.36e-06 |
| <i>preciso</i> 'precise' | 4.34e-03 | <i>singolo</i> 'single' | 9.53e-02 | <i>grande</i> 'big' | 1.29e-06 |
| <i>politico</i> 'politic' | 2.81e-03 | <i>semplice</i> 'simple' | 7.91e-02 | <i>successivo</i> 'following' | 1.03e-06 |
| <i>tradizionale</i> 'traditional' | 2.02e-03 | <i>principale</i> 'principal' | 7.32e-02 | <i>continuo</i> 'continuous' | 8.81e-07 |
| <i>solo</i> 'one and only' | 1.97e-03 | <i>importante</i> 'important' | 6.86e-02 | <i>interminabile</i> 'never-ending' | 6.76e-07 |
| <i>vecchio</i> 'old' | 1.66e-03 | <i>fondamentale</i> 'fundamental' | 2.37e-02 | <i>famoso</i> 'famous' | 6.50e-07 |
| <i>ultimo</i> 'last' | 1.62e-03 | <i>nuovo</i> 'new' | 1.21e-02 | <i>intero</i> 'whole' | 6.23e-07 |
| <i>singolo</i> 'single' | 1.05e-03 | <i>storico</i> 'historical' | 8.30e-03 | <i>breve</i> 'brief' | -2.03e-07 |
| <i>comune</i> 'common' | 1.01e-03 | <i>vero</i> 'true' | 6.28e-03 | <i>vero</i> 'true' | -2.60e-07 |
| <i>storico</i> 'historical' | 1.01e-03 | <i>grande</i> 'big' | 6.11e-03 | <i>importante</i> 'important' | -9.56e-07 |
| <i>sociale</i> 'social' | 7.36e-04 | <i>numeroso</i> 'numerous' | -4.46e-03 | <i>tradizionale</i> 'traditional' | -1.45e-06 |
| <i>determinato</i> 'determined' | 2.09e-04 | <i>particolare</i> 'particular' | -5.40e-03 | <i>unico</i> 'unique' | -1.01e-05 |
| <i>semplice</i> 'simple' | 1.64e-04 | <i>determinato</i> 'determined' | -1.55e-02 | <i>nuovo</i> 'new' | -1.03e-05 |
| <i>particolare</i> 'particular' | 1.31e-04 | <i>specifico</i> 'specific' | -2.45e-02 | | |
| <i>interessante</i> 'interesting' | 6.51e-05 | <i>relativo</i> 'relative' | -6.65e-02 | | |

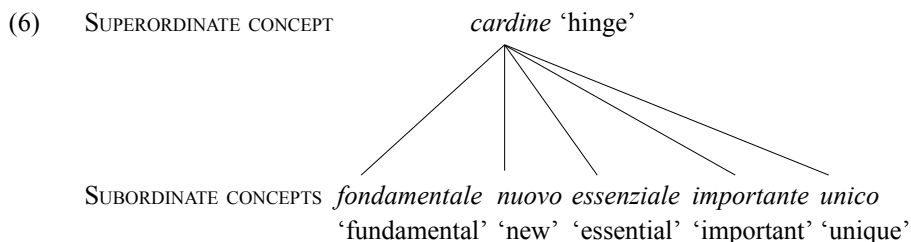
TABLE 5. THE RESULTS OBTAINED APPLYING THE GLOBAL SCORING MODEL FOR SOME EXAMPLE N2.

We can then define a “Global Scoring Model” by combining the different measures proposed before, able to rank every adjective associated with a specific non-head noun N2:

$$score_{N_2}(ADJ) = freqm_{N_2}(ADJ) * WSMm_{N_2}(ADJ) * spread_{N_2}(ADJ)$$

Table 5, in the previous page, shows the list of adjectives most prominently associated with the non-head nouns examined in this study. The top adjectives should provide the “isolated” (see above) semantic feature of N2 when combined in ATAP compounds.

The analysis carried out on CORIS shows a significant result: a noun filling the head slot in NN endocentric ATAP compounds, when used as an autonomous word, tends to co-occur with adjectives that sketch just the semantic field represented by the semantic features associated with the non-head noun. In other words, considering a family of appositive compounds sharing the same non-head constituent, our analysis shows that the set of adjectives occurring most frequently with most head nouns is indicative of the implied sense of the non-head. Consider, for example, the case of *cardine* ‘hinge’. According to subjective judgments of native speakers, in a compound pattern it is supposed to mean something like ‘crucial, central, the most important, etc’. The picture resulting from our corpus study can be represented as follows:



As for *cardine* such a situation is not surprising, since this noun very rarely occurs with its literal meaning even as an autonomous word; moreover, it often acquires a metaphorical reading which is similar to that observed in compounds:

- (7) STAMPAQuot: [...] *E rinunciare ad un **cardine** di tanti processi [...]*
 ‘[...] And renouncing to the pivot of many trials [...]’
 STAMPAQuot: [...] *Adesso , quel filmينو sarà il **cardine** dell ‘ accusa [...]*
 ‘[...] Now, that video will be the hinge of the prosecution [...]’
 STAMPAPeri: [...] *sulla banca che rappresenta il **cardine** della cosiddetta “ Galassia [...]*
 ‘[...] on the bank that represents the hinge of the so-called “Galaxy [...]’
 MISCVolumi: [...] *attraversa tutta la storia e ne è il **cardine** : La vita si è posata su tutti [...]*
 ‘[...] stands over all history and is its hinge: Life extends over all [...]’

However, the results of our corpus-based experiments are especially promising for nouns that, when used in isolation, usually retain their literal meaning, such as *fiume* or *chiave*. In these cases, data extracted from CORIS reveal that their meaning when featuring as N2s in NN appositive compounds is well expressed by adjectives that combine with N1s observed in such compounds, as for example *lungo* (ADJ) occurring with *riunione*

(N1), observed in *riunione fiume* (NN). This not only maps well to native speakers' judgments but also widens the lexico-semantic spectrum of the implied meaning. A few more examples are given in (8):

- (8) *chiave* 'key' *solo* 'one and only', *unico* 'unique', *singolo* 'single', *principale* 'main', *importante* 'important', *fondamentale* 'fundamental'
fiume 'river' *lungo* 'long', *numeroso* 'numerous', *grande* 'big', *continuo* 'continuous', *interminabile* 'never-ending'

Therefore, data suggest that the metaphorical (re)interpretation of the non-head constituent of an appositive NN endocentric compound is not arbitrary at all. The feature selected within the "body" of this noun (in Lieber, 2004's terms) is representative of a sample of meanings that display a preferential link with head nouns, when used as autonomous words:

- | | | |
|-----|----------------------|-------------------------------------|
| (9) | Autonomous word | Modifier in a compound |
| | <i>fiume</i> 'river' | <i>lungo</i> 'long' |
| | | <i>grande</i> 'big' |
| | | <i>numeroso</i> 'numerous' |
| | | <i>continuo</i> 'continuous' |
| | | <i>interminabile</i> 'never-ending' |

So the corpus based analysis allows us to make two kinds of predictions:

- a. extracting from the corpus the adjectives that more frequently combine with NN compounds' heads enables us to constrain our subjective judgments in the determination of the meaning of non-head members;
- b. extracting the sample of nuances that characterise the modifiers enables us to outline the number of nouns that are likely to occupy the head slot.

3.2 Morphosyntactic features

We have observed that N2s in ATAP compounds lose their referential power (or their connotation) completely when they undergo a process of semantic bleaching and, consequently, of metaphorical reinterpretation. They acquire the status of property words and perform a mere descriptive and modifying function.

As seen in Table 1, the function labelled as "modification by property" is typically fulfilled by adjectives. However, this does not prevent other linguistic elements or constructions from acting as modifiers of a noun. As said above, the elements used in atypical roles are also expected to acquire the formal properties associated with the prototypes of such roles in each single language, although they might exhibit a "defective" behaviour with respect to the structural, inflectional, and distributional characteristics of unmarked members.

Data presented so far confirm that non-head constituents of appositive compounds, at least semantically, behave like adjectives. Assuming that

nothing prevents a linguistic item with a prototypical nominal function, i.e. designating an object, from being used with an adjectival function, thus modifying the meaning of another noun, we should be able to observe that N2s like *fiume* and *chiave* show in ATAP compounds a morpho-syntactic behaviour typically associated with adjectives, with possible gaps.

Out of the six features listed by Thornton that characterise adjectives (see Section 2), we selected those that could be checked in an automatic way, at least at a first stage, namely:

1. Agreement in number with N1 (*parole chiavi* lit. words keys, *riunioni fiumi* lit. meetings rivers)
2. Any sort of adverbial modification (*parola più chiave* lit. word more key, *ruoli molto chiave* lit. roles more key)
3. Superlative form (*recupero lamp-issimo* lit. recovery lightning-SUPERLATIVE, *notizia bomb-issima* lit. news bomb-SUPERLATIVE)

For testing the presence of these morphosyntactic features typical of adjectives in the non-head member of ATAP compounds, we used ItWac (Baroni *et al.*, 2009), a larger but less controlled corpus. We searched it by means of simple patterns that exploit the corpus' part-of-speech annotation. Results were subsequently checked by hand since the automatically obtained information was not at the desired level of detail (for instance, there is only one PoS tag for plural and singular nouns). Example patterns for each of the three criteria are given below:

1. Agreement in number with N1
Example pattern: [pos="NOUN"] "cardini";
Example expected match: "punti cardini".
2. Adverbial modification
Example pattern: "parola" [pos="ADV"] "chiave";
Example expected match: "parola più chiave".
3. Superlative form
Example pattern: "riunione" "fiumissim."
Example expected match: "riunione fiumissima".

Our findings indicate that, as expected, not all features are equally exhibited. More specifically, number agreement appears to be a common feature:

- (10) 29440497: *contenente*/VER:ppre <*parole*/NOUN *chiavi*/NOUN> e/CON [...]
 containing word.PL key.PL and
 16125469: *individuando*/VER:geru <*settori*/NOUN *chiavi*/NOUN> nei/ARTPRE [...]
 individuating field.PL key.PL in_the
 60670692: *dei*/ARTPRE <*naufraghi*/NOUN *fantasmi*/NOUN> del/ARTPRE 1996/NUM
 of_the castaway.PL phantom.PL of_the 1996
 74072749: *per*/PRE <*orchestre*/NOUN *fantasmi*/NOUN> [...]
 for orchestra.PL phantom.PL

52722172: *se/CON i/ART <punti/NOUN cardini/NOUN> della/ARTPRE [...]*
 if the point.PL hinge.PL of_the
 65238705: *in/PRE <settori/NOUN cardini/NOUN> quali/WH [...]*
 in field.PL hinge.PL like

Adverbial modification is present but less frequent:

- (11) 5765971: *il/ART <punto/NOUN già/ADV chiave/NOUN> della/ARTPRE [...]*
 the point already key of_the
 45825791: *un/ART <match/NOUN altrettanto/ADV chiave/NOUN> [...]*
 a match as_much key
 71345308: *in/PRE <posizioni/NOUN cosi/ADV chiave/NOUN> –/NOCAT [...]*
 in positions so key
 80624947: *una/ART <parola/NOUN oggi/ADV chiave/NOUN> :/PUN [...]*
 a word today key

Intensification absolutely never occurs in ItWac. Some sporadic occurrences can be found through a Google search:

- (12) *corradooooooooooooo se ci sei scrivi subito ho da darti una notizia bomb-issima sulla tua cantante preferitaaaaa* (www.tvblog.it)
 ‘Corrado, if you are there, write immediately, I have bomb-SUPERLATIVE news about your favourite singer’.

In compounds as *pesce palla*, with a low degree of metaphorical reinterpretation, both number agreement and adverbial modification are almost absent.

The discrepancy in the occurrence of these features is in line with the assumptions of the typological theory of markedness and might indicate that

- (i) (formal) nounhood is still perceived, and thus prevents some behaviour to show;
- (ii) some features are indeed more prototypically associated with adjectives (and thus farther away from nounhood) than others.

Nevertheless, such discrepancy does not prevent us from arguing for an adjectival nature of non-head members of appositive compounds. In the theoretical picture we are adopting, according to Croft (1991; 2000), it is natural for marked members of a category to exhibit a “defective” behaviour with respect to the inflectional possibilities of unmarked members. And it is exactly what data presented so far reveal (Grandi, 2009).

4. CONCLUSIONS

The crucial question we have investigated in the work described here is: is a noun “used as an adjective” actually an adjective or is it merely “functioning

like” an adjective, retaining its nominal nature? The answer falls in the middle. Indeed, we are dealing with a typical case of gradience between two parts of speech. If we look beyond their formal nominal appearance, we have observed that nouns featuring as non-heads in ATAP compounds actually behave like adjectives, both from a functional point of view (loosing the referential capacity and performing a modification function), and from a formal point of view (agreeing in number with their head noun), at least as long as the language system allows it. However, whereas it is clear that such nouns fully function as adjectives, formally we hit another level of gradience, which is internal to the definition and classification of parts-of-speech. Not all properties that formally define an adjective are equally exhibited by all the members of the class “adjective”, since some are more prototypical than others. Therefore, a noun shifting towards an adjectival status might share some but not all of such properties. In Italian, number agreement is a core feature of adjectives. The nouns we have studied seem indeed to share that property. So, are these noun-clad adjectives? Yes, but to a point, since their nominal origin prevents some adjectival formal features to show.

Additionally, we face yet another kind of gradience, which affects the classification strategies we employ to define complex forms (such as the compounds in our study). Again, if we focus on their nominal appearance, we can keep on classifying these complex forms among compounds. But if we look beyond their formal structure and admit that the non-head member of the compound is a noun-clad adjective, then we are moving towards the boundary between the categories of compound and phrase. Indeed, we could postulate that we are dealing with phrases rather than compounds or, at least, with compound-like phrases.

REFERENCES

- Baroni, M., Bernardini, S., Ferraresi, A., Zanchetta E. (2009). The Wacky Wide Web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43 (3), 209-226.
- Bauer, L. (2001). Compounding. In M. Haspelmath, E. König, W. Oesterreicher & W. Raible (Eds.), *Language Typology and Language Universals* (pp. 695-707). Berlin: Walter de Gruyter.
- Bisetto, A. & Scalise, S. (2005). Classification of compounds. *Lingue e Linguaggio* 2, 319-332.
- Croft, W. (1991). *Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information*. Chicago: The University of Chicago Press.
- Croft, W. (2000). Parts of speech as language universals and as language-particular categories. In P. Vogel & B. Comrie (Eds.), *Approaches to the Typology of Word Classes* (pp. 65-102). Berlin: Mouton de Gruyter.
- Grandi, N. (2009). When morphology ‘feeds’ syntax. Remarks on noun > adjective

- conversion in Italian appositive compounds. In F. Montermini, G. Boyé & J. Tseng (Eds.), *Selected Proceedings of Décembrettes 6 - Morphology in Bordeaux* (pp. 111-124). Somerville, MA: Cascadilla Press.
- Heine, B., Claudi, U. & Hünemeyer, F. (1991). *Grammaticalization: A Conceptual Framework*, Chicago: The University of Chicago Press.
- Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics* 20 (1), 1-31.
- Lieber, R. (2004). *Morphology and Lexical Semantics*. Cambridge: Cambridge University Press.
- Montermini, F. (2010). *Units in compounding*. In S. Scalise & I. Vogel (Eds.), *Cross-Disciplinary Issues in Compounding* (pp. 77-92). Amsterdam: Benjamins.
- Rossini Favretti, R., Tamburini, F. & De Santis, C. (2002), CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model. In A. Wilson, P. Rayson & T. McEnery (Eds.), *A Rainbow of Corpora: Corpus Linguistics and the Languages of the World* (pp. 27-38). Munich: Lincom-Europa.
- Scalise, S. & Bisetto, A. (2009). Classification of compounds. In R. Lieber & P. Štekauer (Eds.), *The Oxford Handbook of Compounding* (pp. 49-82). Oxford: Oxford University Press.
- Tamburini, F. (2007). CORISTagger: a high-performance PoS tagger for Italian. *Intelligenza Artificiale IV* (2), 14-15.
- Thornton, A. M. (2004). Conversione in aggettivi. In M. Grossmann M. & F. Rainer (Eds.), *La formazione delle parole in italiano* (pp. 526-533). Tübingen: Niemeyer.
- Widdows, D. (2004). *Geometry and Meaning*. Stanford, CA: CSLI Publication.

Nicola Grandi

Dipartimento di Studi Linguistici e Orientali
 Università di Bologna
 via Zamboni 33, 40124 Bologna
 Italy
 e-mail: nicola.grandi@unibo.it

Malvina Nissim

Dipartimento di Studi Linguistici e Orientali
 Università di Bologna
 via Zamboni 33, 40124 Bologna
 Italy
 e-mail: malvina.nissim@unibo.it

Fabio Tamburini

Dipartimento di Studi Linguistici e Orientali
 Università di Bologna
 via Zamboni 33, 40124 Bologna
 Italy
 e-mail: fabio.tamburini@unibo.it