Verbal-nexus and attributive-appositive N+N compounds in Italian
A diachronic study

Jan Radimský
University of South Bohemia

M. Silvia Micheli
University of Milano - Bicocca

1 Introduction

Italian Noun+Noun compounds (NN compounds, henceforth) have been extensively investigated from a synchronic point of view (see, for an overview, Radimský 2015) due to their significant productivity and the wide variety of patterns attested in Contemporary Italian. Over the last two decades, several studies have focused on the classification of this type of compounds (see, among others, Baroni, Guevara & Pirrelli 2009), which includes a rather heterogeneous set of words, as well as on properties of specific subtypes (see, e.g., Grandi 2009; Grandi, Nissim & Tamburini 2011 and Radimský 2016 on the attributive-appositive compounds, or Baroni, Guevara & Zamparelli and Lami & van den Weijer 2022 on verbal-nexus compounds, according to the classification proposed by Scalise & Bisetto 2009).

On the other hand, much less attention has been paid to the diachrony of NN compounds, that seem to represent a relatively recent innovation in Romance. According to Rainer (2021), the pattern does not display any continuity from Latin compounding and rather stems from a variety of heterogeneous syntactic constructions whose number seems extremely limited in Italian, at least until the end of the 19th Century.

The aim of this contribution is to examine thoroughly the diachronic profile of two prominent Italian N+N compounding patterns, namely verbal-nexus NNs (such as *ritiro*$_{N}$ *bagagli*$_{N}$ – "baggage claim") and attributive-appositive NNs (such as *parola*$_{N}$ *chiave*$_{N}$ – "keyword"), based on a large sample of more than 4.000 manually filtered compounds (types) and their diachronic frequency profiles drawn from the Google n-gram data. With reference to the theoretical frameworks of Construction Morphology (Booij 2010, 2016), Relational Morphology (Jackendoff & Audring 2020) and Diachronic Construction Grammar (Hilpert & Gries 2009, Traugott & Trousdale 2013, Goldberg 2019, Hilpert 2021, among others) we will analyse the progressive coinage of constructions at different levels of abstraction (i.e. substantial, semi-schematic and schematic), the relationship between them and the factors that trigger the "productivity upgrade" of the respective schemas.

Indeed, it is not very often that a new compounding pattern appears and develops in a modern language, in a diachronic period that is quite richly documented by written sources. Therefore, the analysis of this process will not only make it possible to show the specific situation of Italian NN compounds, but also to discuss general theoretical questions concerning the emergence of compounding patterns within the selected framework, such as *coverage* (Goldberg 2019) or *structural intersection* (Jackendoff & Audring 2020), as well as methodological tools designed for analysis of diachronic corpus data, such as *variability-based neighbour clustering* (Hilpert & Gries 2009).

2 Verbal-nexus and attributive-appositive NN compounds

2.1 Key properties

Verbal-nexus NNs (henceforth VNX NNs, also referred to as Argumental NNs) and attributive-appositive NNs (henceforth ATAP NNs) represent two prominent patterns of left-headed present-day Italian NNs (see Radimský 2015, among others).

VNX NNs (such as *trasporto*$_{N}$ *merci*$_{N}$ – transport of goods) are a subtype of left-headed endocentric subordinate compounds consisting of a deverbal head and a non-head element

which is interpreted as its argument. The interpretation of a VNX NN is triggered by the deverbal head (i.e. the leftmost element), so that these compounds are expected to form head-based 'families' or 'semi-schematic constructions' (such as *trasporto-N* – N-transport); however both synchronic and diachronic data surprisingly suggest that they also form argument-based families (such as N-*merci* – N-goods) (cf. Radimský 2020, 2023 in press). According to various scholars, Italian VNX NNs represent the most – if not the only really – productive higher-order subordinate NN construction in Romance (Rainer 2016, Baroni, Guevara & Zamparelli 2009, Radimský 2018).

ATAP NNs (such as *parola*$_{N}$ *chiave*$_{N}$ – "keyword") feature a head-modifier attributive relationship that may be paraphrased as 'N1 is a (kind of) N2'. The modifier may have either a metaphoric (in 'appositive NNs') or a literal (in 'attributive NNs') interpretation, such as in *parola*$_{N}$ *chiave*$_{N}$ ("keyword" – the word is 'key', important) and *luogo*$_{N}$ *simbolo*$_{N}$ ("symbolic place" – the place is a symbol), respectively. In both cases, the interpretation of ATAP compounds is triggered by the modifier (i.e. the rightmost element) and they tend to form strong modifier-based families, which is why selected modifiers with highest type frequencies have sometimes also be analysed as 'noun-clad adjectives' (Grandi, Nissim & Tamburini 2011). It is still debatable whether the ATAP pattern as such represents a productive higher-order construction in contemporary Italian or whether its type frequency growth is rather carried out by a small subset of lower-order semi-schematic constructions. Our current data show that the latter solution is probably more in line with reality.

2.2 History of Italian NNs

As Rainer (2021:17) puts it: „the evolution and modern proliferation of NN compounds in the Romance languages, unfortunately, has not yet been studied in detail from a diachronic point of view".

Single instances of Italian ATAP compounds are already attested in Old Italian. Based on the Codit corpus, Micheli (2020a:91-93) found 3 ATAP NNs in Old Italian (*pescespada* – swordfish, *pesceporco* – grey triggerfish, *arcamensa* – large cupboard) and 15 ATAP NNs in Middle Italian (Micheli 2020a:145, 152-155), but she assumes that the pattern has reached real productivity and dissemination only since the 21st century (Micheli 2020b, 120).

As for subordinate NNs, the existing studies based on literary Italian do not report cases of such compounds attested before 1950 (Tollemache, 1945; Micheli, 2020a, 2020b), but Rainer (2021:17) notes that they became more frequent in contexts related to commerce and industry already since the 19th century. In the journalistic style, first examples are assumed to appear around the 1970s (Dardano 2009:226-229), from where they gradually made their way beyond the narrow sphere of professional communication.

It can be therefore assumed that substantial turning points in the evolution of Italian NN compounds – still very little explored – occurred in the past two centuries.

3 Theoretical framework

Construction Morphology and Relational Morphology are usage-based models, which entails that schemas available in the Constructicon capture generalizations over a critical mass of already attested words. In a diachronic perspective thus, "constructionalization" must be based on previous individual "innovation" (in the sense of Traugott & Trousdale 2013). One of the targets of the research is therefore also to find a method for identification of such lexical innovations (or *leader words*) in the early stages of the development of patterns.

Once a critical mass of individual lexical innovations is in place, Constructionalization – within the Relational Morphology framework (Jackendoff & Audring 2020) – consists of two

steps. First, relational links between the existing words must be built through the process of "Structural Intersection", and then it is necessary to determine whether these new relational schemas are productive. In the case of Compounding, we assume that the Structural Intersection yields primarily semi-schematic constructions, in which chunks of forms (either the leftmost or the rightmost component) are shared. Such a view is consistent with the assumption of Laurie Bauer (2017: 74) that "it is not the N+N pattern of compounding which is productive, but patterns with individual lexemes within that", as well as with the observation of Franz Rainer (2016:2714) that within Italian N+N compounds, "neologisms tend to follow analogues or series of analogues with the same first or second constituent." We will show that, surprisingly, both VNX and ATAP NN compounds form N1- and N2-based families, though only some of them achieve higher type frequencies, including higher type frequencies of hapax forms, and can be therefore considered as productive.

If subsequent Constructionalization is to yield some higher-order constructions, these should correspond to areas in which examples encountered so far cluster (cf. the notion of *coverage* by Goldberg 2019: 51-73 and its application to compounds by Hilpert 2015). Our data suggest that these higher-order constructions may or may not correspond to "classes" or "types" of compounds.

4 Data & methodology

The research is based on extensive diachronic data drawn from the Google books corpus available in the form of raw frequency lists as the 3rd version of Italian Google n-grams,[1] the size of the underlying Google books corpus is 120,410,089,963 tokens. Data for the extraction of N+N compounds come from pre-treated bigrams and trigrams (to capture compounds with space-separated and hyphen-separated components, respectively) from which a sample of roughly 2.000 ATAP and 2.000 VNX compounds has been extracted. In order to achieve a higher accuracy, most compounds have been checked back manually in Google books and many false positives have been eliminated. For each compound, dated numbers of occurrences in Google books are available from 1850 to the present with a year-by-year precision, which makes it possible to analyse in diachrony not only relative token frequencies of single compounds, but also relative type frequencies of different higher-order constructions, such as head-based or modifier-based families (e.g. N-chiave – "key-N") and fully schematic constructions, and their interaction.

To identify diachronic trends and draw regression lines, Theil-Sen estimator was used and supplemented with the Mann-Kendall test for significance testing (Python implementation by Hussain & Mahmud 2019). These rank-based non-parametric methods are suitable to test any form of dependence (not only linear), they do not assume a normal distribution of errors and they are not sensible to outliers, which makes them particularly suitable for trend identification of word usage in diachronic corpora (Kovář & Herman 2013). Potential turning points in the evolution of patterns are detected using the Variability-based neighbour clustering method (Hilpert & Gries 2009).

5 References

Baroni, Marco, Guevara, Emiliano & Vito Pirrelli. 2007. NN compounds in Italian. Modelling category induction and analogical extension. *Lingue e linguaggio* 6(2). 263—90.

Bauer, L. 2017. *Compounds and compounding*. Cambridge: Cambridge University Press.

Booij, Geert. 2010. *Construction Morphology.* Cambridge: Cambridge University Press.

[1] https://storage.googleapis.com/books/ngrams/books/datasetsv3.html

Booij, Geert. 2016. Costruction Morphology. In Andrew Hippisley & Gregory Stump (eds.), *The Cambridge Handbook of Morphology*, 424—448. Cambridge: Cambridge University Press.
Dardano, Maurizio. 2009. *Costruire parole*. Bologna: Il Mulino.
Goldberg, Adele E. 2019. *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton: Princeton University Press.
Grandi, Nicola. 2009. When Morphology 'Feeds' Syntax: Remarks on Noun > Adjective conversion in Italian Appositive Compounds. In *Selected Proceedings of the 6th Décembrettes*, 111—124. Somerville, MA: Cascadilla Proceedings Project.
Grandi, Nicola, Nissim, Malvina & Fabio Tamburini, Noun-clad adjectives. On the adjectival status of non-head constituents of Italian attributive compounds. *Lingue e linguaggio* 10(1). 161—76.
Hilpert, Martin 2015. From hand-carved to computer-based: Noun-participle compounding and the upward strengthening hypothesis. *Cognitive Linguistics* 26(1). 1—36.
Hilpert, Martin. 2021. *Ten lectures on diachronic construction grammar*. Brill. 10.1163/9789004446793
Hilpert, Martin & Gries, Stefan T. 2009. Assessing frequency changes in multistage diachronic corpora: Applications for historical corpus linguistics and the study of language acquisition. *Literary and Linguistic Computing* 24(4). 385—401.
Hussain, Manjurul & Ishtiak Mahmud. 2019. pyMannKendall: a python package for non parametric Mann Kendall family of trend tests. *Journal of Open Source Software* 4(39). 1556. https://doi.org/10.21105/joss.01556
Jackendoff, Ray & Jenny Audring. 2020. *The texture of the lexicon: relational morphology and the parallel architecture*. Oxford: Oxford University Press.
Herman, Ondrej & Vojtech Kovár. 2013. Methods for Detection of Word Usage over Time. In *RASLAN - Seventh Workshop on Recent Advances in Slavonic Natural Language Processing*, 79-85. ISBN 978-80-263-0520-0.
Lami, Irene & Joost van de Weijer. 2022. Compound-internal anaphora: evidence from acceptability judgements on Italian argumental compounds. *Morphology* 32. 1—30.
Micheli, M. Silvia. 2020a. *Composizione italiana in diacronia. Le parole composte dell'italiano nel quadro della Morfologia delle Costruzioni*. Berlin/New York: De Gruyter.
Micheli, M. Silvia. 2020b. *La formazione delle parole. Italiano e altre lingue*. Roma: Carocci.
Radimský, Jan. 2015. *Noun+Noun compounds in Italian: A corpus-based study*. České Budějovice: Jihočeská univerzita. Edice Epistémé.
Radimský, Jan. 2016. I composti NN attributivi nel corpus ItWac. In Annibale Elia, Claudio Iacobini & Miriam Voghera, (eds.), *Livelli di analisi e fenomeni di interfaccia. Atti del XLVII Congresso di studi della Società di Linguistica Italiana*, 189—204. Roma: Bulzoni.
Radimský, Jan 2018. Inflection of binominal ATAP compounds in French and Italian: a paradigmatic account. *Lingue e linguaggio* 17(2). 261—272.
Radimský, Jan 2020. A paradigmatic approach to compounding. In Jesús Fernández-Domínguez, Alexandra Bagasheva & Cristina Lara Clares (eds.), *Paradigmatic relations in word formation*, 164—185. Leiden: Brill.
Radimský, Jan. In press. Where did the Italian N+N compounds come from? In Jenny Audring et al. (eds.), *Mediterranean Morphology Meeting* 13. Patras, University of Patras. ISSN: 1826-7491.
Rainer, Franz. 2016. Italian. In Peter O. Müller, Ingeborg Ohnheiser & Susan Olsen (eds.), *Word-formation. An International Handbook of the Languages of Europe*, 2712—2731. Berlin/Boston: Walter de Gruyter.
Rainer, Franz. 2021. Compounding: from Latin to Romance. In *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.691
Scalise, Sergio & Antonietta Bisetto. 2009. The classification of compounds. In Rochelle Lieber & Pavol Štekauer (eds.), *The Handbook of Compounding*, 49—82. Oxford: Oxford University Press, Oxford.
Tollemache, Federico. 1945. *Le parole composte nella lingua italiana*. Roma: Edizioni Rores di Nicola Ruffolo.
Traugott, Elizabeth C. & Graeme Trousdale. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.