



Tracing back the history of Italian Attributive-Appositive Noun+Noun compounds: First outcomes

Jan Radimský (University of South Bohemia, České Budějovice)

ABSTRACT

The paper presents the initial findings of a research project that investigates the origins of Italian ATAP NNs, such as *parola_N chiave_N* (keyword) or *luogo_N simbolo_N* (place.symbol — symbolic place). This study is based on a meticulously chosen sample of 1,924 NN types that include 47 different modifiers as their rightmost constituent, sourced from Google n-grams frequency lists (2020), which represent the most extensive diachronic linguistic data currently available. While recent literature suggests that Italian ATAP NNs emerged in the latter half of the 20th century and gained significant productivity only after 2000, our data indicate that this process occurred approximately 50 years earlier, particularly for a notable subset of N₂-based semi-schematic constructions. With respect to the theoretical frameworks of Construction Morphology, Relational Morphology, and Diachronic Construction Grammar, the analysis of the data provides an insight into the interaction between the fully schematic ATAP NN construction and various semi-schematic constructions (N₂-based families). Moreover, the study attempts to identify key diachronic stages in the development of the pattern.

KEYWORDS

ATAP, attributive-appositive compounds, compounding, diachrony, Italian

DOI

<https://doi.org/10.14712/18059635.2023.2.3>

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 raising productivity and the wide variety of patterns attested in contemporary Italian. In particular, Attributive-Appositive NNs (ATAP, henceforth), such as *parola_N chiave_N* (keyword) or *luogo_N simbolo_N* (place.symbol — symbolic place), which have been emerging at a fast pace especially in the past two decades, raised a lively debate concerning their theoretical status and classification (see references in Section 2.1.), as well as different morphosyntactic properties, including, namely, inflection or lexical insertion (Radimský 2018, Grandi, Nissim and Tamburini 2011). However, much less attention has been paid to the diachrony of NN compounds, which seem to represent a relatively recent innovation in Romance. In order to fill this gap, this article presents first outcomes of a research that aims to trace back the origin and development of Italian ATAP NNs based on Google n-grams frequency lists (2020), which are the most extensive diachronic linguistic data currently available. The observations presented below draw on a carefully

selected sample of 1,924 ATAP NN compounds that contain 47 different modifiers as their rightmost constituent, with frequency data covering the period from 1800 to 2019.

The paper is structured as follows: Section 2 outlines basic properties of Italian ATAP NNs and summarises existing knowledge about their diachronic development. Section 3 briefly introduces the underlying theoretical background that stems from Construction Morphology (Booij 2010, 2016), Relational Morphology (Jackendoff and Audring 2020), and Diachronic Construction Grammar (Hilpert 2015, Traugott and Trousdale 2013, Goldberg 2019, Hilpert 2021, among others). Section 4 describes the data-gathering process, while Section 5 presents the diachronic profile of the complete ATAP sample and the diachronic profile of selected N₂-based families or “semi-schematic constructions”. Finally, Section 6 summarises the results.

2 ATTRIBUTIVE-APPOSITIVE NN COMPOUNDS

2.1 DELIMITATION OF ATAP COMPOUNDS

The group of ATAP compounds, conceived in the influential classification of compounds by Bisetto and Scalise (2005; 2009) as one of the three major compounding patterns, along with coordinate and subordinate compounds, have given rise to stimulating discussions (cf. Baroni, Guevara and Pirrelli 2009; Grandi 2009; Grandi, Nissim and Tamburini 2011; Arcodia, Grandi and Montermini 2009) that point to the fact that this class of compounds is particularly hard to define on formal grounds. In short, Scalise and Bisetto (2009) argue that ATAP compounds feature an attributive relationship between the head and its modifier, the latter expressing a “property” or “quality” of the head, but they do not provide clear criteria on how to test it in the case of NN structures, besides the fact that the modifier “is often to be interpreted metaphorically” (Scalise and Bisetto 2009: 51–52). Without entering into details of this debate (see Radimský 2015: 92–112 for a detailed overview), we will delimit here the group of Italian ATAP NNs on formal grounds using the transformational test introduced in Radimský (2015: 92–102, see also Arnaud 2003: 12–13 or Fradin 2009 for French), as outlined in the following paragraphs.

The test is based on a widely shared assumption (cf. Salvi and Vanelli 2011) that the attributive relationship typically exists between a qualifying adjective and the corresponding head noun, and that this relationship can be consistently paraphrased using a predicative expression with a copular verb. Since nouns may also feature as subject complements¹ of copular verbs, the same test can be used to detect an attributive relationship in NN structures. Therefore, ATAP NNs with an abstract modifier may be transformed in a predicative expression directly (1), while ATAP NNs with a concrete noun used as a metaphoric modifier may be transformed at least indirectly, using an operator such as Italian *come* (“like a”) (2).

1 In Italian: *complemento predicativo del soggetto*.



- (1) *luogo simbolo* → *questo luogo è un simbolo*
 [place]_N [symbol]_N
 symbolic place → this place is a symbol
- (2) *pesce palla* → *questo pesce è come una palla*
 [fish]_N [globe]_N
 globefish → this fish is like a globe

Notice that the test yields ungrammatical or inappropriate outputs for subordinate NN compounds (3) as well as for NA phrases with a relational adjective (4), which is precisely what makes it suitable to set apart ATAP NN structures (1, 2) from subordinate NNs (3).²

- (3) *protezione persone* → **questa protezione è persone*
 [protection]_N [persons]_N
 protection of persons → *this protection is persons
- (4) *sistema nervoso* → **questo sistema è nervoso*
 [system]_N [nervous]_A
 nervous system → *this system is nervous

A more fine-grained classification of Italian ATAP NNs, introduced in Radimský (2015: 158–159), further sets apart attributive NNs with an abstract modifier (1) from appositive NNs with a concrete modifier (2), since only in the latter group modifiers have a metaphorical interpretation. However, as clear as this difference may seem at first sight, its empirical application comes up against the fact that — especially in the case of nominal modifiers with higher type frequencies in ATAP compounds — the metaphorical interpretation of the noun is often well “lexicalised” even without being limited to the NN pattern. For instance, the monolingual Zingarelli (2011) dictionary lists among the basic meanings of the words *chiave* (key) and *lampo* (lightning, flash) their metaphorical interpretations “important” (5) and “short / quick” (6), respectively, describing these interpretations as common even outside of the NN pattern, as the examples within (5–6) show.

- (5) *chiave 2* — *Elemento, dato o persona di importanza vitale per comprendere, interpretare, risolvere qlco. [...] Quel personaggio è la chiave di tutta la vicenda.*
chiave 2 — Element, fact or person of vital importance for understanding, interpreting, resolving sth. [...] [Example:] That character is the key to the whole affair.

2 For the sake of completeness, it should be added that this transformational test also yields positive results for endocentric coordinate NN compounds, such as *lavoratore-studente* (student worker), in which both components denote similar concepts situated on the same hierarchy level (see Radimský 2015: 102–112 for a detailed discussion). These NNs are not included within the ATAP class of NNs here.

- (6) *lampo 3* — *Ciò che ha brevissima durata: la giovinezza è un lampo.*
lampo 3 — What is very short-lived: [Example:] youth is a flash.

- lampo 4* — *Persona, animale o veicolo molto veloce: quel cane è un lampo.*
lampo 4 — Very fast person, animal or vehicle: [Example:] that dog is a flash.

In these cases, it would be a purely academic question to ask whether NNs such as *parola chiave* (keyword) or *treno lampo* (fast train, lit. “flash train”) are rather appositive (since the interpretation of the modifier is metaphorical with respect to the primary concrete meaning of that noun) or attributive (since the interpretation of the modifier is based on one of its meanings which happens to be metaphorical). To put it differently, the metaphorical interpretation of modifiers, such as (5–6), is not bound to the NN pattern. That is why no further distinction will be made between attributive and appositive NN compounds in this article.

Finally, a word is in order about the fact that some authors suggest that selected modifiers in ATAP compounds (i.e., such as *chiave* or *lampo* in the examples mentioned above) should rather be analysed as “noun-clad” invariable adjectives (Grandi, Nissim and Tamburini 2011, Grandi 2009, but see Thornton 2004: 528–530 or Radimský 2015: 22–29). Since these elements are primarily nouns, with selected syntactic (but not morphological) features typical for qualificative adjectives, the debate seems inconclusive on formal grounds. We thus consider that the choice of classifying these modifiers combined with head nouns as “ATAP NN compounds” is one of the two possible methodological options, whose particular advantage is that it makes it possible to describe ATAP NNs as an integral part of Italian NNs that have entered a phase of dynamic evolution in the past century.³

2.2 HISTORY OF ATAP COMPOUNDS

As Rainer (2021) points out, the current knowledge about the evolution and modern proliferation of Romance NN compounds is still extremely limited. The pattern does not display any continuity from Latin compounding, but rather stems from a variety of heterogeneous syntactic constructions whose number seems very low in Italian, at least until the end of the 19th century.

In an extensive diachronic study of Italian compounds based on the Codit corpus of literary Italian, Micheli (2020a: 91–93) found only 3 ATAP NNs in Old Italian (7a–c) and 15 ATAP NNs in Middle Italian (Micheli 2020a: 145, 152–155); the latter refer to animals — especially fishes — (8a), plants, vegetables, or other concrete substances (8b).

³ On the other hand, let us add that some specific regular NN structures were deliberately kept apart from NN compounds. These involve especially cases where the modifier is a proper noun, i.e. a referential expression, such as *sorella Maria* (sister Mary), that we consider as instances of regular syntax. Others involve self-referential modifiers, such as *l'aggettivo “buono”* (the adjective “buono”) or *il numero quattro* (the number four). For a more detailed discussion see Radimský (2015:40–44).



- (7a) *pescespada* — swordfish (fish+sword)
 (7b) *pesceporco* — grey triggerfish (fish+pig)
 (7c) *arcamensa* — large cupboard (ark+table)

- (8a) *pesceca* (dogfish, shark, fish+dog), *pescecavallo* (horseface loach, fish+horse), *pescesega* (sawfish, fish+saw), *pesceluna* (ocean sunfish, fish+moon), *formicaleone* (ant-lion, ant+lion), *grillotalpa* (mole cricket, cricket+mole)
 (8b) *melarancia* (sweet orange, apple+orange), *cavolfiore* (cauliflower, cabbage+flower), *cavolrapa* (kohlrabi, cabbage+turnip), *casciofiore* (type of cheese similar to pecorino, cheese+flower), *cartastraccio* (looking like wastepaper, paper+shredded), *erbaspada* (agave, grass+sword), *ceralacca* (sealing wax, wax+type of resin), *gommalacca* (shellac, gum+type of resin)

Unfortunately, the data extraction in the study by Micheli (2020a) considers only compounds written in the “tight” or “hyphenated” form. If compounds in the “loose” form (i.e., written as two separate words) were also extracted, a few other examples might be found in the Codit corpus to complete the sample (for instance, *edizione principe* — editio princeps). Nevertheless, this still indicates an extremely limited presence of ATAP NNs in the literary language before the mid-20th century.

Though the data from the Codit corpus (Micheli 2022) do not go further than the year 1947, based on secondary sources — or, rather, the lack of reliable information in secondary sources — Micheli (2020b: 120) assumes that the ATAP pattern has reached real productivity and dissemination only since the 21st century.

The data analysed in this article aim to shed more light on the evolution of the Italian ATAP NNs in the past two centuries.

3 THEORETICAL FRAMEWORK

As Construction Morphology (Booij 2010; 2016) and Relational Morphology (Jackendoff and Audring 2020) are usage-based models, schemas available in the Constructicon capture generalizations over a critical mass of already attested words. “Constructionalization” must therefore be based on previous individual “Innovation” (Traugott and Trousdale 2013). The process of constructionalization may produce a complex hierarchical network of schemas that capture different levels of generalization in the Constructicon. In the case of ATAP NN compounds, such a network may be schematised in (9a-d), showing in turn individual instances of ATAP NNs (9d), semi-schematic constructions based on a repeated component (9c), the fully schematic ATAP NN schema (9b) and the most general left-headed NN pattern (9a).

- (9a) Left-headed NN construction

$$[N_i N_j]_{Nk} \leftrightarrow [N_{i\text{-head}} N_{j\text{-non-head}}]_k$$



(9b) ATAP NN construction

$$\begin{aligned} [N_i N_j]_{Nk} &\leftrightarrow [N_{i\text{-head}} \text{È (COME) UN(A) } N_{j\text{-non-head}}]_k \\ [N_i N_j]_{Nk} &\leftrightarrow [N_{i\text{-head}} \text{IS (LIKE) A } N_{j\text{-non-head}}]_k \end{aligned}$$

(9c) ATAP NN semi-schematic constructions

c.1 SUB NN semi-schematic constructions based on the same modifier (N2)

$$\begin{aligned} [N_i \text{chiave}]_{Nk} &\leftrightarrow [N_{i\text{-head}} \text{È chiave/importante}]_k \\ [N_i \text{chiave}]_{Nk} &\leftrightarrow [N_{i\text{-head}} \text{IS key/important}]_k \end{aligned}$$

c.2 SUB NN semi-schematic constructions based on the same head (N1)

$$\begin{aligned} [\text{pesce } N_j]_{Nk} &\leftrightarrow [\text{pesce che È COME UN(A) } N_{j\text{-non-head}}]_k \\ [\text{pesce } N_j]_{Nk} &\leftrightarrow [\text{fish that IS LIKE A } N_{j\text{-non-head}}]_k \end{aligned}$$

(9d) Individual instances of NNs

- d.1 *parola chiave* (keyword)
- d.2 *pesce spada* (swordfish)
- d.3 *ruolo chiave* (key role)
- d.4 *pesce palla* (globefish)

A study that aims at describing and accounting for the emergence of a compounding word-formation pattern within this theoretical perspective has to ask the following questions, namely:

First, it is necessary to identify in the diachronic data primary lexical innovations (“leading words” or “attractors”, see e.g., Hartmann 2019: 316) that may be subject of subsequent etymological analyses. Such leading words appear in the early stages of the development of patterns, where they are expected to have higher token frequencies, though later, their token frequency may decrease significantly. Even though other interesting “attractors” with high type frequencies may also emerge in subsequent stages, their origin is potentially different: early attractors cannot result from the application of productive schemas, so that their etymology must be studied and accounted for individually.

Second, progressively emerging higher-order constructions have to be detected and described. Recent studies on compounding emphasise the prominent role of lower-order schemas that require an in-depth investigation (see Hilpert 2015; Hartmann 2019). Indeed, the same general properties of higher-order constructions may correspond to very different configurations of lower-level constructions: a high type frequency of a general pattern, such as (9a) or (9b), may be due either to an increasing number of lower-level constructions (9c) or rather to a high type frequency of just a few lower-level constructions (9c). In the latter case, the higher-order constructions do not have sufficient coverage (Goldberg 2019: 51–73), which entails that the increasing type frequency of lower-order constructions does not strengthen the mental representation of higher-order constructions (Hilpert 2015; 2019).

This article will focus on the latter question from a quantitative point of view. On a sample of 1,924 ATAP NNs with 47 different N2s (modifiers), we will analyse



progressively emerging higher-order constructions (9b) and (9c), as well as their interaction in time.

In order to identify diachronic trends and draw regression lines, Theil-Sen estimator will be used and supplemented with the Mann-Kendall test for significance testing (Python implementation by Hussain and 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 and Herman 2013).

4 DATA

The research is based on extensive diachronic data drawn from the Google books corpus in the form of raw frequency lists that have been published as the 3rd version of Italian Google n-grams.⁴ The size of the underlying Google books corpus is 120,410,089,963 tokens. Data for the extraction of NN compounds come from pre-treated bigrams and trigrams to capture compounds with space-separated and hyphen-separated components, respectively.⁵ The n-grams were filtered and merged together into a single dataset labeled “it2020_bi”⁶ using the procedure described in detail by Radimský (2022); the whole it2020_bi dataset (from which function words have been filtered out) comprises 19,319,372 non-lemmatised types. From this list, only forms that potentially match the forms of Italian nouns according to the MorphIt dictionary (Zanchetta and Baroni, 2005) were filtered out, which yielded a list of 2,723,566 NN candidates. As even this reduced list still contains an extremely high rate of false positives, further filtering of ATAP NNs had to be done manually.

ATAP compounds are known to cluster around modifier-based families: some authors even assume that nominal modifiers in ATAP compounds constitute a relatively limited set of nouns (Baroni, Guevara, and Pirrelli 2009). That is why their manual extraction consisted in identifying potential modifiers on the rightmost position of NNs followed by systematic manual checking of all corresponding heads on the leftmost position.

Potential nominal modifiers (henceforth N₂s, by virtue of their rightmost position in the N₁N₂ structure under investigation) were identified in two ways. First, N₂s that appeared in the sample of 1,800 present-day Italian ATAP NNs from Radimský (2015) were put on the list. Second, the list was completed by further N₂s extracted manually from the Zingarelli (2011) dictionary that marks such nouns either by a specific secondary part-of-speech category labeled “invariable adjective

4 <https://storage.googleapis.com/books/ngrams/books/datasetsv3.html>

5 Compounds written as one graphical word have not been considered as this orthography is typical only to neoclassical right-headed compounds in modern Italian (see Radimský 2015: 49–50).

6 The it2020-bi dataset is available for download at: <https://osf.io/46qcd/>

function” (*in funzione di agg. inv.*), or by a note explaining that they may be used as non-head nouns in NN structures (*posto a un sost. / posto al sost.*). The dictionary data was added to ensure that some prominent older modifiers that are no longer in use were not missing from the sample. The current list of N2 candidates contains 319 non-lemmatised word forms (though further filtering may show that many of them are false positives).

In the subsequent manual identification of head nouns (N1s), we focused prominently (though not exclusively) on N2s with higher type frequencies. Up to this point, it was possible to check and filter whole families of head nouns (N1s) that correspond to 47 different N2s, their complete list including respective type frequencies is given in Appendix. In order to achieve the highest accuracy during data extraction, the following measures were taken.

- a) The majority of compounds were manually verified through Google Books. A preliminary review of a sample consisting of 10–20 contextualised instances (tokens) often allows for distinguishing between authentic ATAP NN compounds and false positives. Factors contributing to the presence of false positives on the candidate list include errors arising from optical character recognition (OCR) and complexities in morphological or syntactic ambiguity. This verification process successfully helped eliminate numerous false positives. However, the frequency counts for the identified true positive types may still contain some noise, which is unavoidable due to the restricted access to original Google Books data and the sheer volume of data involved.
- b) Word forms were used as basic units instead of lemmas, as it has been observed that due to morphological ambiguity, many inflected forms often contribute to false positives in actual texts. As most modifiers in Italian ATAP NNs are invariable (Radimský 2015: 49–50; 2018), the word/lemma distinction has a relatively limited relevance in the case of N2s.

The current sample comprises 1,924 ATAP NNs, understood as non-lemmatised types, featuring 47 different N2s (modifiers) and 1,148 different N1s (head nouns). For each compound (type), dated numbers of occurrences in Google books are available from 1800 to the present with a year-by-year precision.

5 RESULTS

5.1 DIACHRONIC PROFILE OF THE SCHEMATIC ATAP PATTERN

The diachronic evolution of ATAP compounds can be analyzed on various levels of abstraction, namely, substantial constructions (individual types), semi-schematic constructions (N1- or N2-based families), and the fully schematic ATAP construction. As outlined in Section 3, we will focus on the latter two cases. Let us first adopt the bird’s eye view by making a global overview of the whole sample of 1,924 ATAP NNs that corresponds to the schematic construction (10).



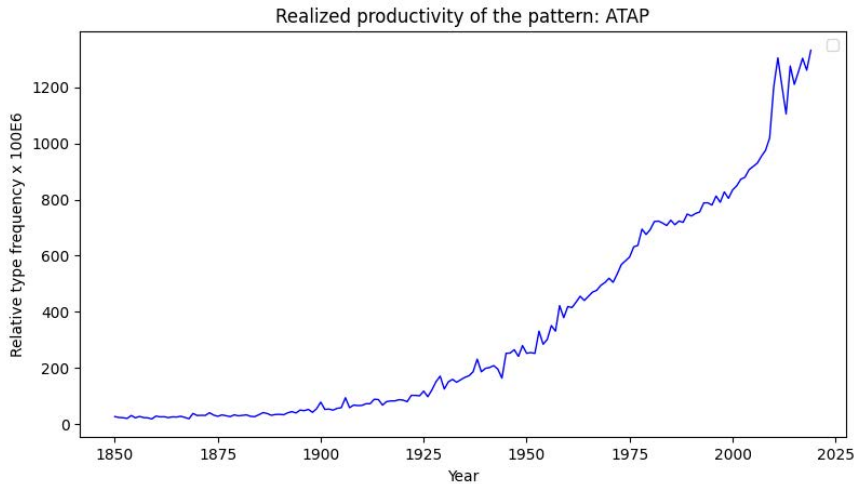


FIGURE 1: Realised productivity of ATAP NNs in diachrony
($V/N \times 10^8$, where V is the number of types, and N is the corpus size in the respective year)

(10) ATAP NN construction

$$\begin{aligned} [N_i N_j] N_k &\leftrightarrow [N_i\text{-head } \dot{E} \text{ (COME) UN(A) } N_j\text{-non-head}]_k \\ [N_i N_j] N_k &\leftrightarrow [N_i\text{-head IS (LIKE) A } N_j\text{-non-head}]_k \end{aligned}$$

Figure 1 gives a diachronic overview of the number of types (type frequency) of the ATAP pattern from the 1850s to 2019. Since the size of the underlying corpus varies in diachrony, type frequency is expressed in relative terms as the number of types divided by the corpus size in the respective year, which corresponds to the classical productivity measure referred to as Realised Productivity by Baayen (2009). The result was multiplied by the constant 10^8 so that it intuitively approaches the order of magnitude of the original type frequency data.⁷ The fact that Realised Productivity is restricted to “past achievement”, usually quoted as a drawback of this measure when applied to synchronic corpora, does not represent an issue when diachronic data are available.

The curve in Figure 1 shows a relatively steady increase in the type frequency from the beginning of the 20th century on, with two periods of steeper slope in 1950–1980 and 2000–2019, respectively.

Nevertheless, such an overall picture does not show how the lower-level families contribute to the increase, i.e., whether the general increase in type frequency is due to an increasing number of many different lower-level constructions or rather to a high type frequency of just a few lower-level constructions. Indeed, the N_2 -based family with N_2 *chiave* (key), which is the one that has by far the highest type frequency in the sample, also has an increasing realised productivity since the 1930s, as

⁷ For instance, in the year 2019, the absolute type frequency 1,603 corresponds to the relative type frequency 1,331.16 displayed in Figure 1.

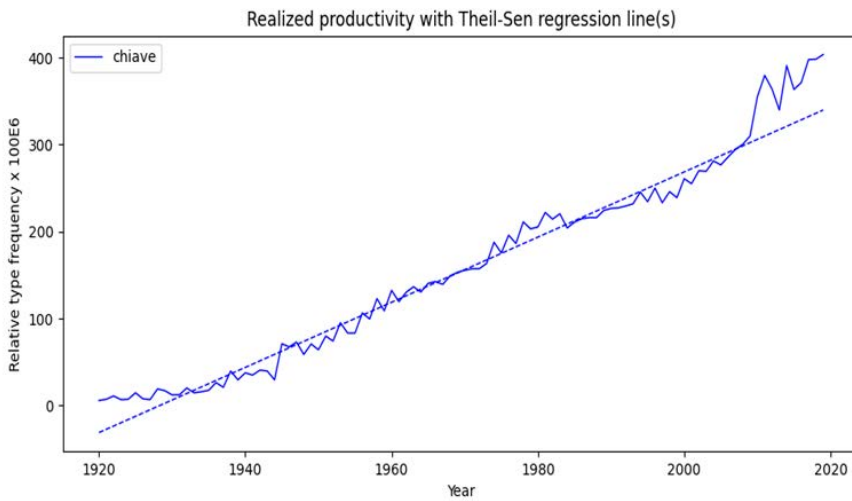


FIGURE 2: Realised productivity of the N-2 based family N-*chiave* (N-key) ($V/N \times 10^8$, where V is the number of types, and N is the corpus size in the respective year)

shown in Figure 2, which entails that it might contribute substantially to the increasing realised productivity observed in Figure 1.

Since ATAP compounds cluster around N₂-based families, we will hypothesise that a more accurate picture of the realised productivity of the whole ATAP pattern can be obtained by asking how many different N₂-based families are present at a given point of time. Indeed, only an increasing number of N₂-based families would entail a more even coverage of the ATAP construction (10), so that it may strengthen its mental representation. This measure, referred to as Family type frequency in Figure 3, yields a different view on Realised productivity of ATAP NNs observed in Figure 1. Though, technically speaking, it also expresses Realised productivity of ATAP NNs, all the types with the same N₂ are counted as just one.

Compared to Figure 1, Figure 3 indicates three different stages in the evolution of ATAP NNs. Within the first period going from the 1850s to 1950s, the slowly increasing type frequency of ATAP NNs observed in Figure 1 is due to an increasing number of N₂-based families observed in Figure 3. In other terms, this is the period when almost all N₂-based families under investigation emerge. In the second period between 1950 and 2000, the rapid type frequency increase of ATAP NNs (Figure 1) is imputable to a relatively limited number of families because the family type frequency in Figure 3 displays no trend ($p=0.591$ according to the Mann-Kendall test) and the global number of N₂-based families is lower than in the 1940s. Thus, the ATAP pattern as such does not evolve, only some N₂-based families do. The next turnpoint took place after 2000, when the family type frequency in Figure 3 began to grow again.

When comparing Type frequency and Family type frequency in Figures 1 and 3, the most striking difference emerges in the first part of the second period, going

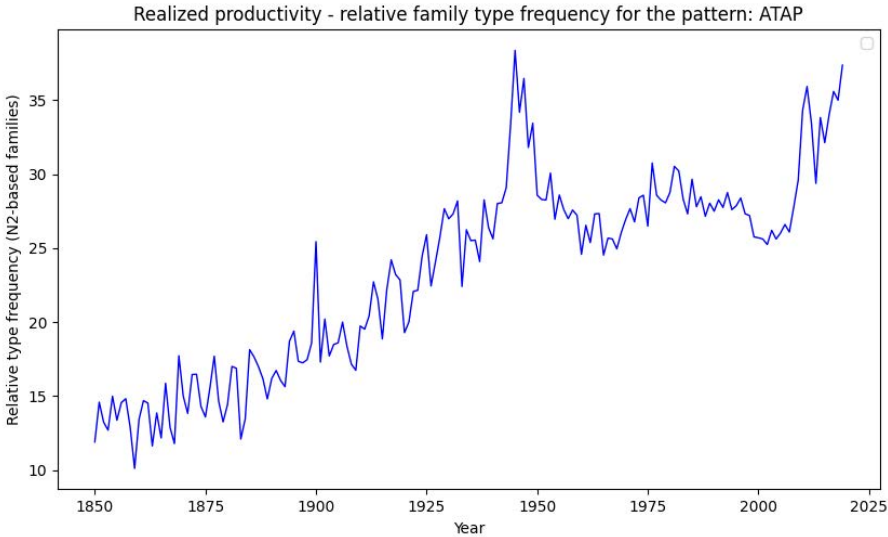


FIGURE 3: Family type frequency of N2-based families

from 1950 to 1980, where the Type frequency displays a strong increasing trend ($p=1.62 \times 10^{-13}$ according to the Mann-Kendall test), while the Family type frequency has no trend ($p=0.918$).

Further evidence that a substantial breakpoint took place between the 1940s and 1950s is given by the diachronic evolution of Potential productivity of ATAP NNs visualised in Figure 4. Following Baayen (2009), potential productivity was calculated by dividing the number of hapax legomena of ATAP NNs by the total number of tokens of all ATAP NNs in the respective years. As it was not very clear what to count as “hapax legomenon” in the corpora that are both very large and variable in size (notice that, technically speaking, frequency data for yeach year come from a different subcorpus with a different size), the threshold was determined in the following way: based on the assumption that hapax legomena represent about 50% of a frequency list (commonly known as the “First Zipf’s law”), all ATAP NNs from all the subcorpora were put on a single frequency list ordered by the decreasing relative frequency; the dividing point situated roughly in the middle of the list (rank 44,055 out of 86,554, i.e. 50.9%) corresponds to the $F_{rel.}=2$. Therefore, any ATAP NN with $F_{rel.}<2$ was taken to be a hapax legomenon.

Baayen’s Potential productivity is supposed to capture the degree of saturation of a word formation pattern. In other terms, a decreasing potential productivity entails progressive saturation of the pattern. If the above discussed periodization is applied to the data in Figure 4, we can notice that potential productivity of the ATAP pattern has an increasing trend in the first period from 1850 to 1950 ($p=0.0111$) that contrasts with a clearly decreasing trend in the second period from 1950 to 2000 ($p=2.22 \times 10^{-16}$). This matches well the assumption that in the second period, a limited number of N2-based families are gradually saturating while new families do not emerge.

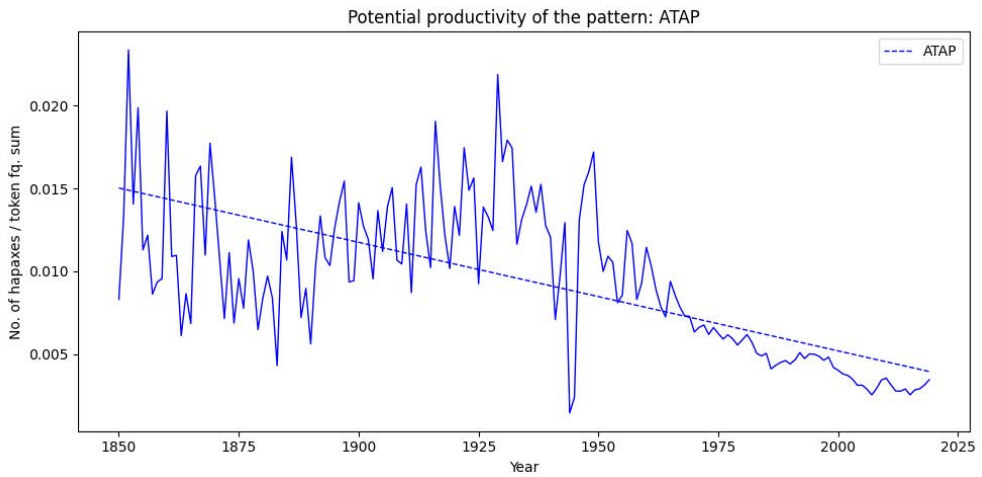


FIGURE 4: Potential productivity of ATAP NNs

The following section will provide a more fine-grained analysis of the phenomenon by focusing on the diachronic evolution of 16 most prominent N2-based families.

5.2 DIACHRONIC PROFILE OF SELECTED N2-BASED FAMILIES

The most prominent N2-based family *N-chiave* (11) has 559 attested types in the sample (in absolute figures). Though its type frequency slowly grows over the 19th century oscillating from 1 to 10 types, the steep and steady increase of the curve began only in the 1930s and lasted throughout the entire 20th century, as shown above in Figure 2. Indeed, from the 1930s to 2000s, the curve virtually follows the Theil-Senn regression line and then it begins to grow even faster.

(11) ATAP NN semi-schematic construction *N-chiave*

$$[N_i \textit{chiave}]_{Nk} \leftrightarrow [N_{i\text{-head}} \textit{È chiave/importante}]_k$$

$$[N_i \textit{chiave}]_{Nk} \leftrightarrow [N_{i\text{-head}} \textit{IS key/important}]_k$$

Graphs in Figure 5 show the diachronic curves of Realised productivity of the following 15 most important families with global type frequency higher than 30 in the sample. All of them appeared already in the 19th century (see Appendix) and in most cases, their Realised productivity curve has a very similar shape to the one of *N-chiave*, i.e., a steady increase that begins somewhere between the 1930s – 1960s (that is, 1–3 decades later than for *N-chiave*) and turns into an exponential increase after the 2000s.

In the second half of the 20th century, some families had their particular “period of glory” followed by a period of relative stability. For instance, the families *N-pilota* and *N-guida* increase quickly in the 1960s, the family *N-leader* in the 1980s.

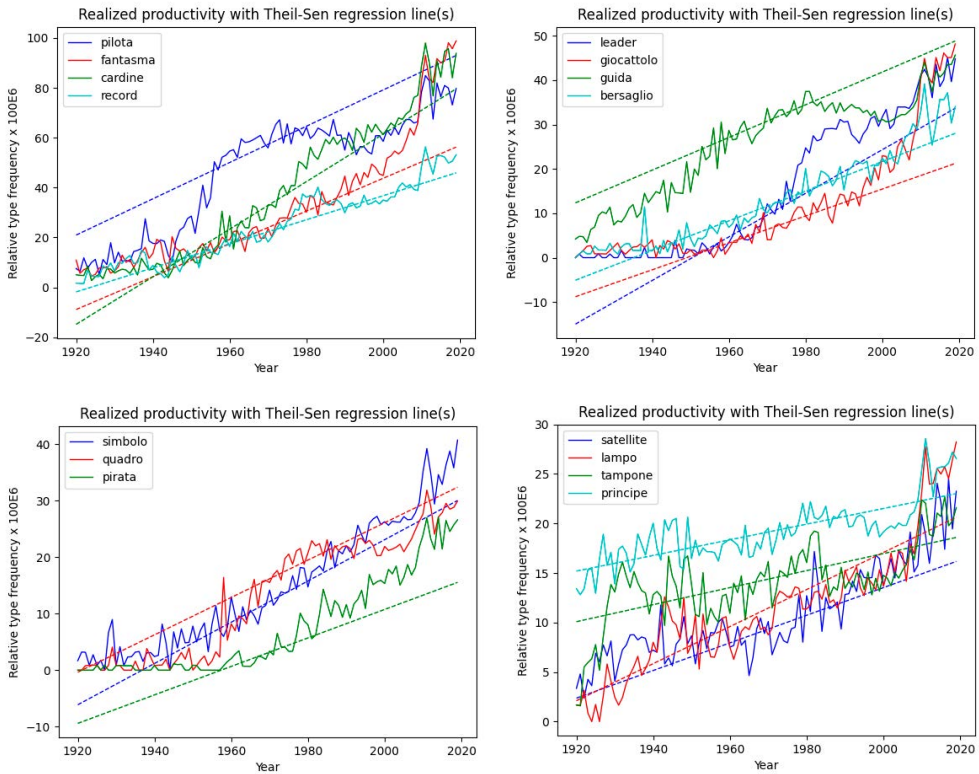


FIGURE 5: Realised productivity of selected N-2 based families *pilota* (pilot), *fantasma* (ghost), *cardine* (hinge), *record* (record), *leader* (leader), *giocattolo* (toy), *guida* (guide), *bersaglio* (target), *simbolo* (symbol), *quadro* (frame), *pirata* (pirate), *satellite* (satellite), *lampo* (lightning), *tampone* (buffer), *principe* (prince)⁸

Notice that the families N-*principe* and N-*tampone* differ from the others in that the first period of their rapid type frequency growth came even earlier than for N-*chiave*: N-*principe* displays a steep slope between the 1890s – 1940s, while the realised productivity of N-*tampone* grew very quickly in a short time span situated in the 1920s – 1930s, as shown in Figure 6.⁹

Interestingly, the realised productivity of all the 16 families (including N-*chiave*) start to rise again after 2000.

Finally, the question must be asked to what extent the steady increase in Realised productivity of the whole ATAP pattern in Figure 1 is imputable to the N-*chiave* construction (Figure 2), which has by far the highest type frequency in the sample. For

⁸ See Appendix for more detailed glosses and examples.

⁹ As one of the anonymous reviewers suggested, the curve of N-*principe* possibly finds its explanation in the rise of classical philology after 1850, which boosted the frequency of *editio princeps*.

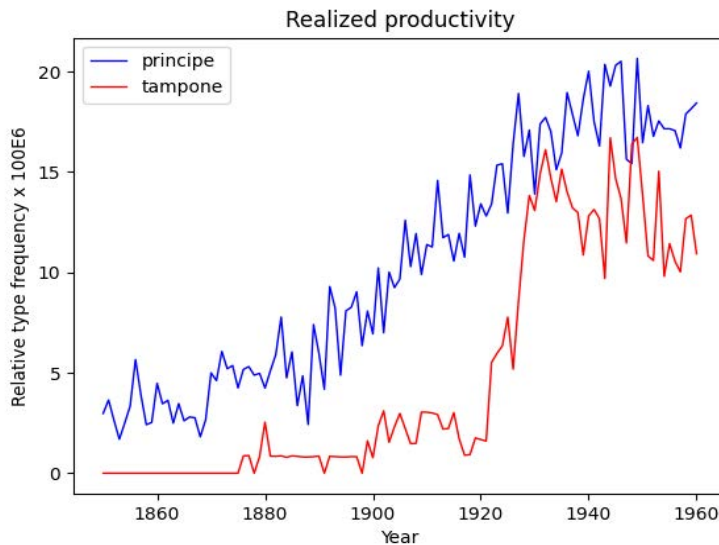


FIGURE 6: Realized productivity of selected N-2 based families *tampone* (buffer) and *principe* (prince) from 1850's to 1960's

this purpose, Figure 7 indicates the rate of N-*chiave* types in the whole sample of ATAP NNs in diachrony. Notice, however, that the data from the 18th century are based on very low absolute figures (with type frequency of N-*chiave* comprised between 0–10), which makes them quite volatile and less reliable.

The curve shows that the contribution of the N-*chiave* construction to the increase in Realised productivity of ATAP NNs was substantial especially between the 1930s and 1950s, where many new families emerge in data, but only a few (especially those in Figure 6) go productive. Since the 1950s, the rate of N-*chiave* types has remained high but stable: even though new N₂-based families do not emerge, the increasing Realised productivity of ATAP NNs is distributed among several N₂-based families. Also the faster curve growth after 2000 cannot be attributed to the growth of the N-*chiave* family only.

6 DISCUSSION AND CONCLUSIONS

The analysis of 1,924 Italian ATAP NNs with 47 different N₂s (modifiers) in Google n-gram data yields a detailed insight into the diachronic emergence of ATAP NNs with respect to the role of the different semi-schematic constructions (N₂-based families). From a global perspective, the data show a steady increase in the type frequency of ATAP NNs from the beginning of the 20th century on, i.e. roughly 50 years earlier than expected, as observed in Figure 1 above. However, this is but a small part of the story.

What is more, almost all N₂-based families appear for the first time as early as the 19th century. The subsequent development can be split into three parts, with the re-

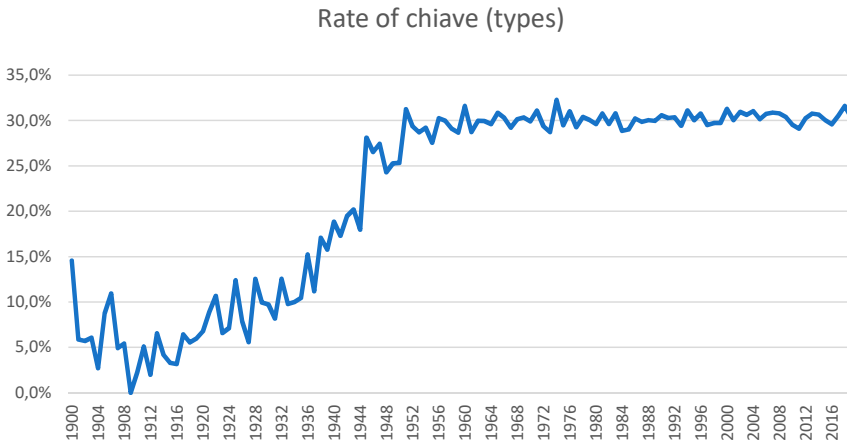


FIGURE 7: Rate of N-*chiave* (N-key) types in the sample of ATAP NNs

spective breakpoints situated around 1950 and 2000. During the first half of the 20th century, a high variability of different families emerge — the highest family type frequency is reached in the late 1940s — so that the ATAP pattern has a relatively even coverage by families. At the same time, only very few families go productive in this period: N-*principe* between the 1890s – 1940s, N-*tampone* in the 1920s – 1930s and N-*chiave* since the 1930s. Between the 1940s and 1950s, the situation changes. On the one hand, the number of families decrease, which entails a lower coverage of the ATAP pattern, but on the other hand, a substantial set of families go productive (though not following always the same curve). In other terms, the ATAP NN pattern itself does not seem to expand, its potential productivity even decreases, but a set of established families keep expanding. The following turnpoint takes place after 2000, where virtually all the families begin to expand faster and even their number starts to rise again.

These first outcomes will need to be refined and — if necessary — modified by further research. First of all, more N₂-based families with lower frequencies should be identified and added to the sample. A qualitative analysis of single families, their history, and their important early members or “primary lexical innovations” is also much needed. Then, the results should be also compared with the corresponding French pattern, on which the Italian pattern most probably depends in some way.¹⁰ Unfortunately, any analysis based on Google n-gram data will also necessarily suffer from some issues that cannot be overcome, however rich and interesting these data are. The underlying Google books corpora are not genre-balanced, complete access to original texts is not always possible, re-edition of older sources may perturb dating accuracy, and the n-gram data contain some part

¹⁰ French ATAP NNs have been usually described as NA_{inv.} phrases where the non-head corresponds to a converted denominal invariable adjective (see Noailly 1990: 35–64; similarly for Spanish see Rainer 2012).

of false positives which is impossible to eliminate — to mention at least the most important ones. Nevertheless, Google n-gram data still represent the richest and most accurate source for the diachronic analysis of written language of the past two centuries currently available.



REFERENCES

- Arcodia, G. F., N. Grandi and F. Montermini (2009) Hierarchical NN compounds in a cross-linguistic perspective. *Italian Journal of Linguistics* 21/1, 11–33.
- Arnaud, Pierre J. L. (2003) *Les composés timbre-poste* (Travaux du C.R.T.T.). Lyon: Presses Univ. de Lyon.
- Baayen, R. H. (2009) Corpus linguistics in morphology: Morphological productivity. In: Lüdeling, A. and M. Kytö (eds) *Handbooks of Linguistics and Communication Science*, 899–919. Berlin / New York: Mouton de Gruyter. doi:10.1515/9783110213881.2.899.
- Baroni, M., E. Guevara and V. Pirrelli (2009) Sulla tipologia dei composti N+N in italiano : principi categoriali ed evidenza distribuzionale a confronto. In: Ferrari, G., R. Benatti and M. Mosca (eds), *Linguistica e modelli tecnologici di ricerca : atti del XL Congresso internazionale di studi della Società di linguistica italiana (SLI) : Vercelli, 21–23 settembre 2006*, 73–96. Roma: Bulzoni Editore.
- Bisetto, A. and S. Scalise (2005) The Classification of Compounds. *Lingue e linguaggio* 2, 319. doi:10.1418/20728.
- Booij, G. (2010) *Construction morphology* (Oxford linguistics). Oxford / New York: Oxford University Press.
- Booij, G. (2016) Construction Morphology. In: Hippius, A. and G. Stump (eds) *The Cambridge Handbook of Morphology*, 424–448. 1. edn. Cambridge: Cambridge University Press. doi:10.1017/9781139814720.016.
- Fradin, B. (2009) IE, Romance: French. In: Lieber, R. and P. Štekauer (eds) *The Oxford Handbook of Compounding*, 417–435. Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199695720.013.0022.
- Goldberg, A. E. (2019) *Explain me this: creativity, competition, and the partial productivity of constructions*. Princeton, New Jersey: Princeton University Press.
- Grandi, N. (2009) When Morphology 'Feeds' Syntax: Remarks on Noun > Adjective Conversion in Italian Appositive Compounds. In: Montermini, F., G. Boyé and J. Tseng (eds) *Selected Proceedings of the 6th Décebrettes*, 112–124. Somerville (MA): Cascadilla Proceedings Project.
- Grandi, N., M. Nissim and F. Tamburini (2011) Noun-clad adjectives. On the adjectival status of non-head constituents of Italian attributive compounds. *Lingue e Linguaggio* 10/1, 161–176. doi:10.1418/34543.
- Hartmann, S. (2019) Compound worlds and metaphor landscapes: Affixoids, allostructions, and higher-order generalizations. *Word Structure* 12/3, 297–333. doi:10.3366/word.2019.0151.
- Hilpert, M. (2015) From hand-carved to computer-based: Noun-participle compounding and the upward strengthening hypothesis. *Cognitive Linguistics* 26/1, 113–147. doi:10.1515/cog-2014-0001.
- Hilpert, M. (2019) Higher-order schemas in morphology: What they are, how they work, and where to find them. *Word Structure* 12/3, 261–273. doi:10.3366/word.2019.0149.
- Hilpert, M. (2021) *Ten lectures on diachronic construction grammar* (Distinguished lectures in cognitive linguistics vol. 26). Leiden / Boston: Brill.
- Hussain, Md. and I. Mahmud (2019) pyMannKendall: a python package for non parametric Mann Kendall family of trend tests. *Journal of Open Source Software* 4/39. 1556. doi:10.21105/joss.01556.
- Jackendoff, R. and J. Audring (2020) *The texture of the lexicon: relational morphology and the parallel architecture*. First edition.



- Oxford / New York: Oxford University Press.
- Herman, O. and V. Kovář (2013) Methods for Detection of Word Usage over Time. In: Horák, A. and P. Rychlý (eds), *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2013*, 79–85. Brno: Tribun EU.
- Micheli, M. S. (2020a) *La formazione delle parole: italiano e altre lingue* (Studi superiori; Linguistica 1237). 1a edizione. Roma: Carocci editore.
- Micheli, M. S. (2020b) *Composizione italiana in diacronia: Le parole composte dell'italiano nel quadro della Morfologia delle Costruzioni*. Berlin / Boston: De Gruyter. doi:10.1515/9783110652161.
- Micheli, M. S. (2022) CODIT. A new resource for the study of Italian from a diachronic perspective: Design and applications in the morphological field. *Corpus* 23. doi:10.4000/corpus.7306.
- Noailly, M. (1990) *Le substantif épithète* (Linguistique nouvelle). Ire éd. Paris: Presses universitaires de France.
- Radimský, J. (2015) *Noun+Noun compounds in Italian: a corpus-based study* (Episteme. Theoria). Vydání 1. České Budějovice: Jihočeská univerzita v Českých Budějovicích.
- Radimský J. (2018) Inflection of binominal ATAP compounds in French and Italian: a paradigmatic account. *Lingue e linguaggio* 2, 261–272. doi:10.1418/91868.
- Radimský, J. (2022) Towards a diachronic analysis of Romance morphology through Google n-grams. *Corpus* 23. doi:10.4000/corpus.7205.
- Rainer, F. (2012) Escarceos sobre la conversión sustantivo → adjetivo en español. In: Fábregas, A., E. Feliú, J. Martín and J. Pazó (eds) *Los límites de la morfología: Estudios ofrecidos a Soledad Varela Ortega*, 369–382. Madrid: UAM Ediciones. doi:10.15366/1.morfologia2012.022.
- Rainer, F. (2021) Compounding: From Latin to Romance. *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. doi:10.1093/acrefore/9780199384655.013.691. [last accessed 01/08/2023].
- Salvi, G. and L. Vanelli (2011) *Nuova grammatica italiana* (Manuali Linguistica). Bologna: Il Mulino.
- Scalise, S. and A. Bisetto (2009) The Classification of Compounds. In: Lieber, R. and P. Štekauer (eds) *The Oxford Handbook of Compounding*, 34–53. Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780199695720.013.0003.
- Thornton, A. M. (2004) Conversione. In: Grossmann, M. and F. Rainer (eds) *La formazione delle parole in italiano*, 499–533. Tübingen: Max Niemeyer Verlag.
- Traugott, E. C. and G. Trousdale (2013) *Constructionalization and constructional changes* (Oxford Studies in Diachronic and Historical Linguistics). First edition. Oxford: Oxford University Press.
- Zanchetta, E. and M. Baroni (2005) Morph-it! A free corpus-based morphological resource for the Italian language. *Proceedings from the Corpus Linguistics Conference Series* 1(1). Available at <https://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-conf-e-journal.aspx> [last accessed 01/08/2023].
- Zingarelli (2011) *Zingarelli Nicola, Lo Zingarelli 2011, Vocabolario della lingua italiana*, CD-ROM. Bologna: Zanichelli.

Jan Radimský

Department of Romance Studies

Faculty of Arts, University of South Bohemia in České Budějovice

Branišovská 31a, 37005 České Budějovice

ORCID ID: 0000-0002-9016-9293

radimsky@ff.jcu.cz

APPENDIX

List of N2s (modifiers) in the sample, ordered according to decreasing type frequency

N2	English gloss (N ₂)	Example N1	English gloss (N ₁ N ₂)	Type frequency	First appeared in	Last appeared in
<i>chiave</i>	key <i>important</i>	<i>parola</i>	word+key <i>keyword</i>	559	1806	2020
<i>pilota</i>	pilot <i>leading, tri- al</i>	<i>progetto</i>	project+pilot <i>pilot project</i>	153	1806	2020
<i>fantasma</i>	ghost <i>imaginary, false</i>	<i>città</i>	town+ghost <i>ghost town</i>	140	1804	2020
<i>cardine</i>	hinge <i>fundamen- tal, corner- stone</i>	<i>punto</i>	point+hinge <i>linchpin</i>	136	1830	2020
<i>record</i>	record	<i>livello</i>	level+record <i>record level</i>	83	1869	2020
<i>leader</i>	leader	<i>azienda</i>	company+leader <i>leading company</i>	64	1867	2020
<i>giocattolo</i>	toy <i>false</i>	<i>pistola</i>	gun+toy <i>toy gun</i>	62	1869	2020
<i>guida</i>	guide	<i>linee</i>	lines+guide <i>guidelines</i>	57	1802	2020
<i>bersaglio</i>	target	<i>organo</i>	organ+target <i>target organ</i>	52	1819	2020
<i>simbolo</i>	symbol	<i>luogo</i>	place+symbol <i>symbolic place</i>	49	1805	2020
<i>satellite</i>	satellite	<i>paese</i>	country+satellite <i>satellite country</i>	47	1825	2020
<i>quadro</i>	frame <i>general</i>	<i>accordo</i>	agreement+frame <i>framework agreement</i>	42	1804	2020
<i>pirata</i>	pirate <i>false, fraud- ulent, ag- gressive</i>	<i>copia</i>	copy+pirate <i>pirate copy</i>	38	1838	2020
<i>tampone</i>	buffer <i>provisional</i>	<i>zona</i>	zone+buffer <i>buffer zone</i>	38	1876	2020
<i>lampo</i>	lightning <i>quick, short</i>	<i>guerra</i>	war+lightning <i>blitzkrieg</i>	37	1813	2020
<i>principo</i>	prince <i>main, lead- ing</i>	<i>strumento</i>	instrument+prince <i>main instrument</i>	36	1802	2020



N2	English gloss (N ₂)	Example N1	English gloss (N ₁ N ₂)	Type frequency	First appeared in	Last appeared in
<i>fantasia</i>	fantasy <i>flashy</i>	<i>tessuto</i>	fabric+fantasy <i>patterned fabric</i>	27	1836	2020
<i>cult</i>	cult	<i>film</i>	movie+cult <i>cult movie</i>	27	1881	2020
<i>spazzatura</i>	trash	<i>televisione</i>	television+trash <i>trash television</i>	25	1900	2020
<i>partner</i>	partner	<i>paesi</i>	countries+partner <i>partner countries</i>	23	1882	2020
<i>spettacolo</i>	show enter- tainment	<i>politica</i>	politics+show <i>showbiz politics</i>	21	1809	2020
<i>ombra</i>	shadow	<i>governo</i>	government+shadow <i>shadow government</i>	19	1805	2020
<i>ponte</i>	bridge <i>temporary</i>	<i>prestito</i>	loan+bridge <i>bridge loan</i>	19	1824	2020
<i>civetta</i>	owl <i>coquette, de- coy</i>	<i>auto</i>	car+owl <i>decoy car</i>	17	1804	2020
<i>fiume</i>	river <i>excessively long</i>	<i>discorso</i>	speech+river <i>excessively long speech</i>	16	1861	2020
<i>stimolo</i>	trigger, in- centive	<i>frase</i>	sentence+trigger <i>trigger sentence</i>	15	1816	2020
<i>cuscinetto</i>	cushion <i>buffer</i>	<i>zona</i>	zone+buffer <i>buffer zone</i>	13	1902	2020
<i>matrigna</i>	stepmoth- er <i>unkind, wicked</i>	<i>natura</i>	nature+stepmother <i>wicked nature</i>	12	1802	2020
<i>feticcio</i>	fetish <i>idol</i>	<i>attore</i>	actor+fetish <i>fetish actor</i>	12	1828	2020
<i>dormitorio</i>	dormitory	<i>quartiere</i>	district + dormitory <i>dormitory district</i>	12	1895	2020
<i>fantoccio</i>	puppet	<i>regime</i>	regime+puppet <i>puppet regime</i>	11	1811	2020
<i>boia</i>	execution- er <i>extreme, in- human</i>	<i>freddo</i>	cold+executioner <i>freezing cold</i>	10	1868	2020
<i>ragno</i>	spider	<i>uomo</i>	man+spider <i>spiderman</i>	8	1802	2020
<i>cane</i>	dog <i>damn</i>	<i>freddo</i>	cold+dog <i>freezing cold</i>	6	1801	2020

N2	English gloss (N ₂)	Example N1	English gloss (N ₁ N ₂)	Type frequency	First appeared in	Last appeared in
<i>forza</i>	strength <i>basic, fundamental</i>	<i>idea</i>	idea+strength <i>basic idea</i>	5	1804	2020
<i>bomba</i>	bomb <i>sensational</i>	<i>notizia</i>	news+bomb <i>bombshell</i>	5	1885	2020
<i>lumaca</i>	snail <i>very slow</i>	<i>treno</i>	train+snail <i>very slow train</i>	4	1809	2020
<i>fotocopia</i>	photocopy <i>identical</i>	<i>giornali</i>	newspapers+ photocopy <i>identical newspapers (featuring the same content)</i>	4	1869	2020
<i>catenaccio</i>	bolt <i>blocking</i>	<i>legge</i>	law+bolt <i>decree-law (a decree entering in force immediately)</i>	4	1884	2020
<i>ghetto</i>	ghetto	<i>quartiere</i>	district+ghetto <i>ghetto district</i>	4	1938	2020
<i>trabocchetto</i>	trap	<i>domanda</i>	question+trap <i>trick question</i>	3	1929	2020
<i>ago</i>	needle <i>very thin</i>	<i>pesce</i>	fish+needle <i>needlefish</i>	2	1802	2020
<i>lager</i>	Nazi concentration camp	<i>manicomio</i>	madhouse+ concentration camp <i>asylum where ill-treatment is inflicted</i>	2	1966	2020
<i>avventura</i>	adventure	<i>viaggio</i>	trip+adventure <i>adventure trip</i>	2	1976	2020
<i>francobollo</i>	postage stamp <i>very small</i>	<i>formato</i>	format+postage stamp <i>postage stamp format</i>	1	1872	2020
<i>cavia</i>	guinea pig, lab animal	<i>città</i>	town + guinea pig <i>experimental town</i>	1	1956	2012
<i>panino</i>	sandwich	<i>giornali</i>	newspapers+ sandwich <i>newspaper that is sold together with another newspaper or magazine</i>	1	1976	2010