Verbal-nexus and attributive-appositive N+N compounds in Italian

A diachronic study

Jan Radimský* & M. Silvia Micheli
*University of South Bohemia in České Budějovice (CZ)

Outline

- N-N compounds in Italian
  - Key properties
  - Main subtypes
- State-of-the-art
  - N-N compounds
  - Focusing on VNX and ATAP compounds
- Research questions
- Methodology
  - Data gathering
  - Quantitative analysis
  - Theoretical framework
- Results
  - VNX compounds
  - ATAP compounds
- Conclusions

N-N compounds in Italian

- Some **examples**
  - *parola chiave* (keyword), *agenzia viaggi* (travel agency), *noleggio auto* (car rental), *studente lavoratore* (student worker)

- Key **properties**
  - Productive pattern (construction) that forms complex naming units
  - Involves 2 bare common nouns (no determiner)
  - Implicit relationship between nouns (no preposition)
  - Order of constituents: mostly endocentric, left-headed
    - *trattamento rifiuti (treatment_vaste$_{PL}$)* – waste treatment
    - *trattamento rifiuti È UN trattamento* (*waste treatment* IS A /kind of/ treatment)

Three main subtypes (Scalise & Bisetto 2009; Radimský 2015)

1. **Subordinative compounds**

- Verbal-nexus (VNX): ***trasporto merci*** (transport.goods) “freight transport”
  - Deverbal head + Non-head element (interpreted as its argument)
  - Verb-complement or Verb-adjunct relationship
  - Interpretation triggered by the deverbal head
  - These compounds are expected to form head-based ‘families’ or ‘semi-schematic constructions’ (such as *trasporto*-N – N-transport) **(?)**
  - 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) **(?)**

- Grounding: ***sala stampa*** (room.press) “press room”
  Other kind of subordinate relationship (R-relation)

Three subtypes (Scalise & Bisetto 2009; Radimský 2015)

2. **ATAP compounds**
   - Attributive: ***luogo simbolo*** (place.symbol) “symbolic place”
     Literal attributive relationship: N1 is (a) N2

   - Appositive**:** ***parola chiave*** (word.key) “keyword”
     Metaphoric attributive relationship: N1 is a kind of / is like a N2

     - Interpretation triggered by the modifier (i.e., the rightmost element)
     - They tend to form strong modifier-based families, which is why selected modifiers with highest type frequencies have sometimes also been 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 **(?)**

3. **Coordinative compounds**
   - ***lavoratore studente*** (worker.student) **“student worker“**
     Attributive relationship: N1 is (a) N2

State-of-the-art

- Italian N-N compounds have been extensively investigated from a **synchronic** point of view (cf. Radimský 2015 for an overview)
  - Studies focused on specific patterns (cf. Grandi 2009; Grandi, Nissim & Tamburini 2011 and Radimský 2016 on the attributive-appositive compounds, or Baroni, Guevara & Zamparelli 2009 and Lami & van den Weijer 2022 on verbal-nexus compounds)

- On the other hand, much less attention has been paid to the **diachrony** of NN compounds
  - They seem to represent a relatively **recent innovation** in Romance
  - According to Rainer (2021), the SUB pattern does not display any continuity with Latin compounding
  - The SUB pattern seems to stem from a variety of heterogeneous syntactic constructions whose number seems extremely limited in Italian, at least until the end of the 19th century

State-of-the-art

- More specifically, as far as **subordinative compounds** are concerned:
  - The existing literature does not report cases of subordinate N-N compounds attested before 1950 (Tollemache, 1945; Micheli, 2020a, 2020b)
  - 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),

- As for **ATAP compounds**:
  - 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)
  - She assumes that the pattern has reached real **productivity and dissemination only since the 21st century** (Micheli 2020b, 120)

- It can be therefore assumed that substantial **turning points** in the evolution of Italian NN compounds **occurred in the past two centuries**

Our study

- This study aims to investigate the **diachronic profile** of VNX and ATAP compounds, taking into consideration a period ranging from 1850 to the present

- The analysis will be both **quantitative and qualitative** in nature and allow us to answer the following questions:

1. How does the history of VNX and ATAP compounds begin and develop in Italian?
2. Do VNX/ATAP compounds represent a productive higher-order construction?
3. Do VNXs only form head-based families? Do ATAPs only form modifier-based families?

Data gathering

- The study is based on extensive diachronic data drawn from the **Google Books corpus** (size: 120,410,089,963 tokens) available in the form of raw frequency lists
  - Data for the extraction of N+N compounds come from pre-treated bigrams and trigrams to capture compounds with **space-separated** and **hyphen-separated** constituents, respectively (cf. Radimský 2022)
- We extracted a sample of roughly 2.645 ATAP and 1.772 VNX compounds
  - **Manual filtering**: based on previous research (Radimský 2015), N1 and N2 families, N2 modifiers listed by the Zingarelli dictionary
  - **Manual verification** in Google Books in order to achieve a higher accuracy (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
  - This allows us to analyse diachronically:
    - **the relative token frequencies of single compounds**
    - **the** relative type frequencies **of semi-schematic constructions** (e.g., N-chiave – “key-N”) as well as of the **fully schematic constructions**
    - their interaction

Quantitative analysis

- To identify **diachronic trends and draw regression lines**, we used:
  - the **Theil-Sen estimator** 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 (Herman & Kovář 2013)

- the ***Variability-based neighbour clustering method*** (Hilpert & Gries 2009) in order to identify potential turning points in the evolution of patterns

Theoretical framework

- In this study, we adopt the assumptions of two usage-based models, i.e., **Construction Morphology** (Booij, 2010) and **Relational Morphology** (Jackendoff & Audring 2020)

  - Schemas capture generalizations over a critical mass of already attested words (i.e., "**constructionalization**" based on previous individual "**innovation**", in the sense of Traugott & Trousdale, 2013)
    - **How can we identify** these innovations (or *leader words*)?

  - **Structural intersection** (Jackendoff & Audring, 2020:223-225)
    - establishes relational links between words based on their similarities

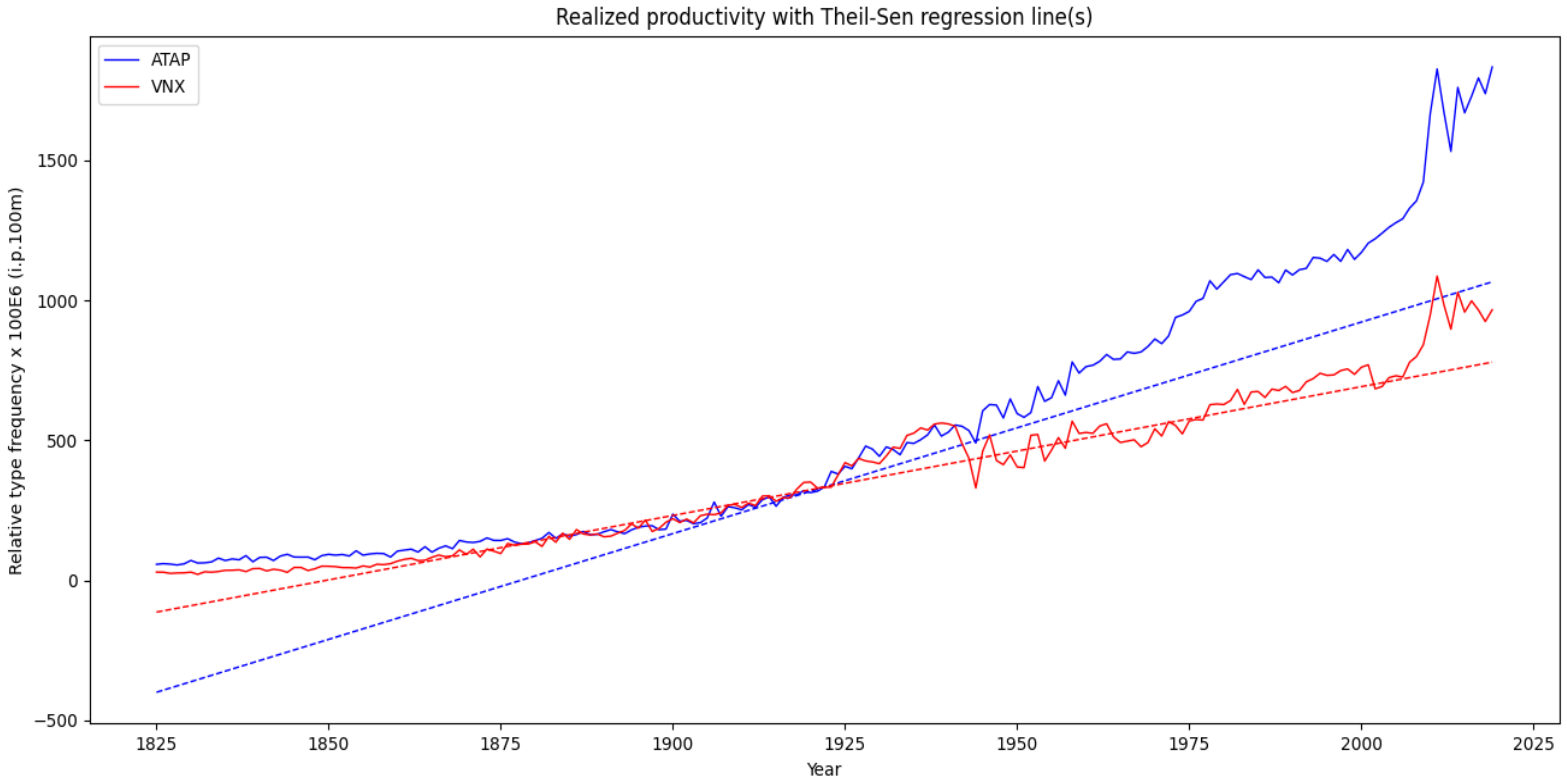
  - **Productivity**
    - according to RM, productivity is an "upgrade" (schemas are firstly declarative)
    - how to determine whether patterns are productive?
      - *uneven and even coverage* (Goldberg, 2019)
    - what is the role of **semi-specified constructions** (families) in diachrony?
      - It is not the N+N pattern of compounding which is productive, but patterns with individual lexemes within that (Bauer, 2017:74; cf. Rainer, 2016: 2714 for Romance NNs – e.g., *parola chiave* – "keyword")

Pattern overview (1): **Realized productivity (Baayen, 2009) of compounds**

- **Relative type frequency**: $F_{rel.} = V/Nx10^8$
  - number of types / corpus size in the respective year x constant (the result intuitively approaches the order of magnitude of the original type frequency data)
  - Realized Productivity restricted to “past achievement” – this drawback is irrelevant with diachronic data

- **Very similar curves**
  - 1st examples 19th century
  - Steady increase 1900-2000
  - Exponential increase 2000+

- **Is the coverage by families even?**
  - N1-based families for VNX
  - N2-based families for ATAP



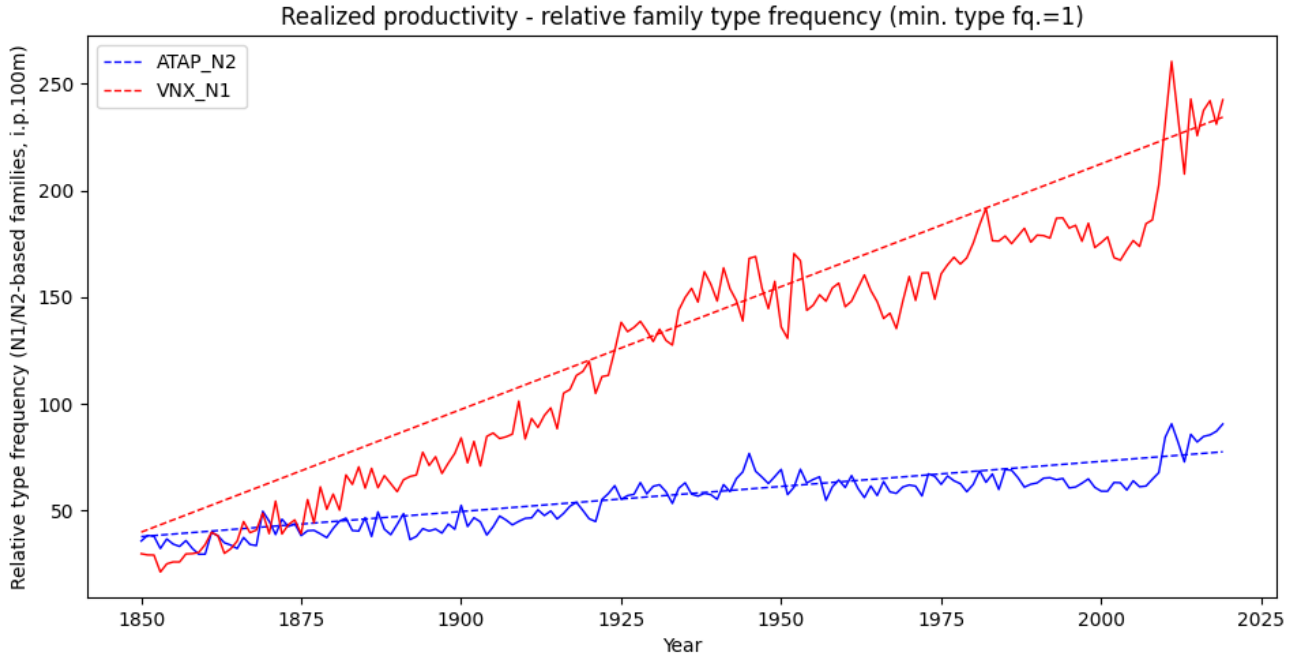
Pattern overview (2): **Realized productivity of N1 / N2 based families**

- Relative "**Family type frequency" (FTF)**

  ATAP: *N-chiave* (*key-N*) VNX: *trasporto-N* (*N-transport*)

  - ***Coverage*** (Goldberg, 2019): increase of one limited sub-pattern does not contribute to the productivity increase of the whole pattern. Only an increasing number of families entails a more even coverage of the VNX/ATAP construction and strenghten its mental representation
  - How many N1 / N2 based families ("triggers") are in the data?
  - Also expresses *realized productivity*, but all the types with the same N1/N2 are counted as just one
  - How many members make a "family"? (Here: 1 member – "trigger")

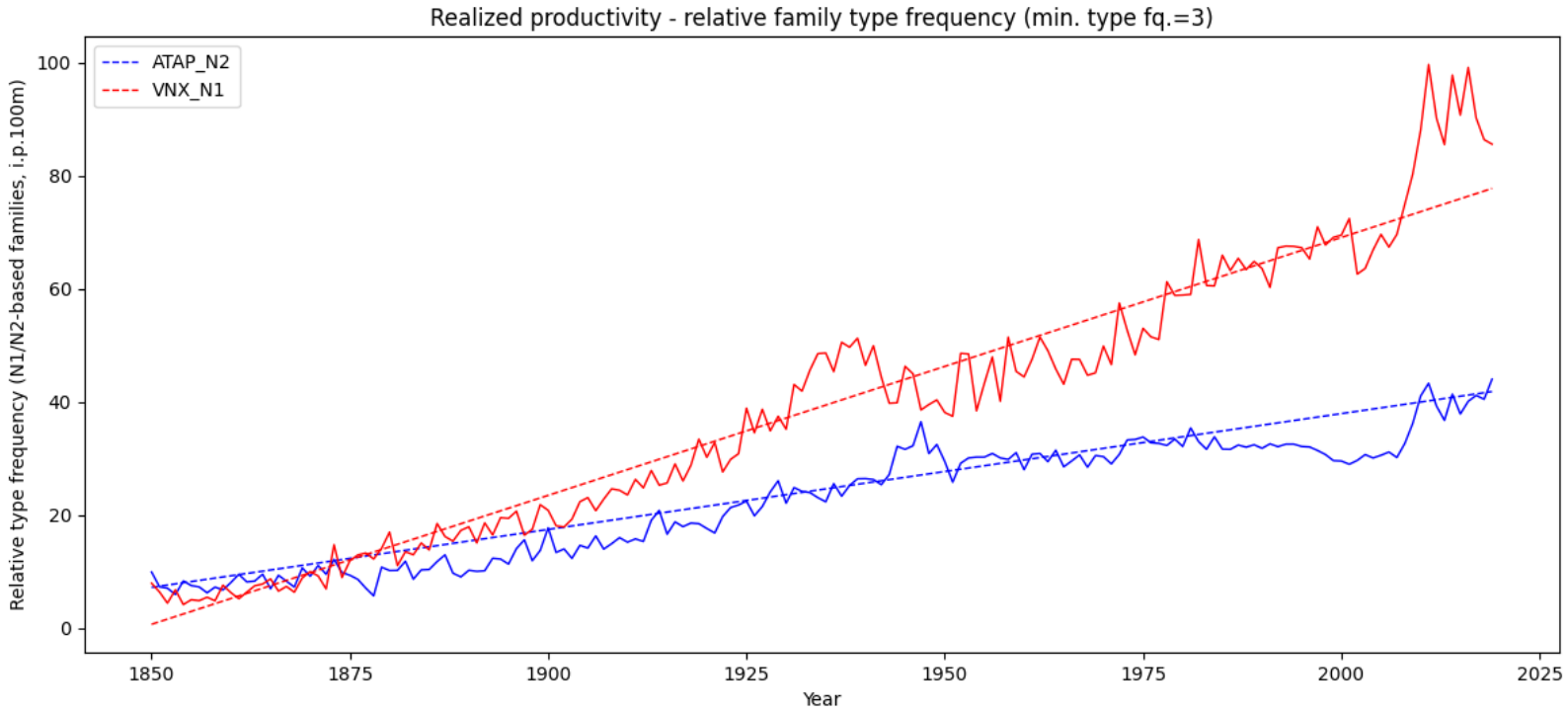
- Very different curves
- ATAP
  - Total: 123 N2s
  - No trend between: 1945-2008 (p=0.43, slope = -0.021)
  - Low coverage by families
- VNX
  - Total: 356 N1s
  - Stronger increase, despite shorter periods of "no trend": 1983-2006 (p=0.17, slope=-0.20)



Pattern overview (3): **Relative Family type frequency**

- An N1/N2-based “family” has at least 3 members (compounds) – *tres faciunt colegium*
- The only relevant measure for SUB_GROUND compounds, where both N1 and N2 may be ”triggers”
  - A new member of N1-based family may automatically yield a new N2 and vice-versa

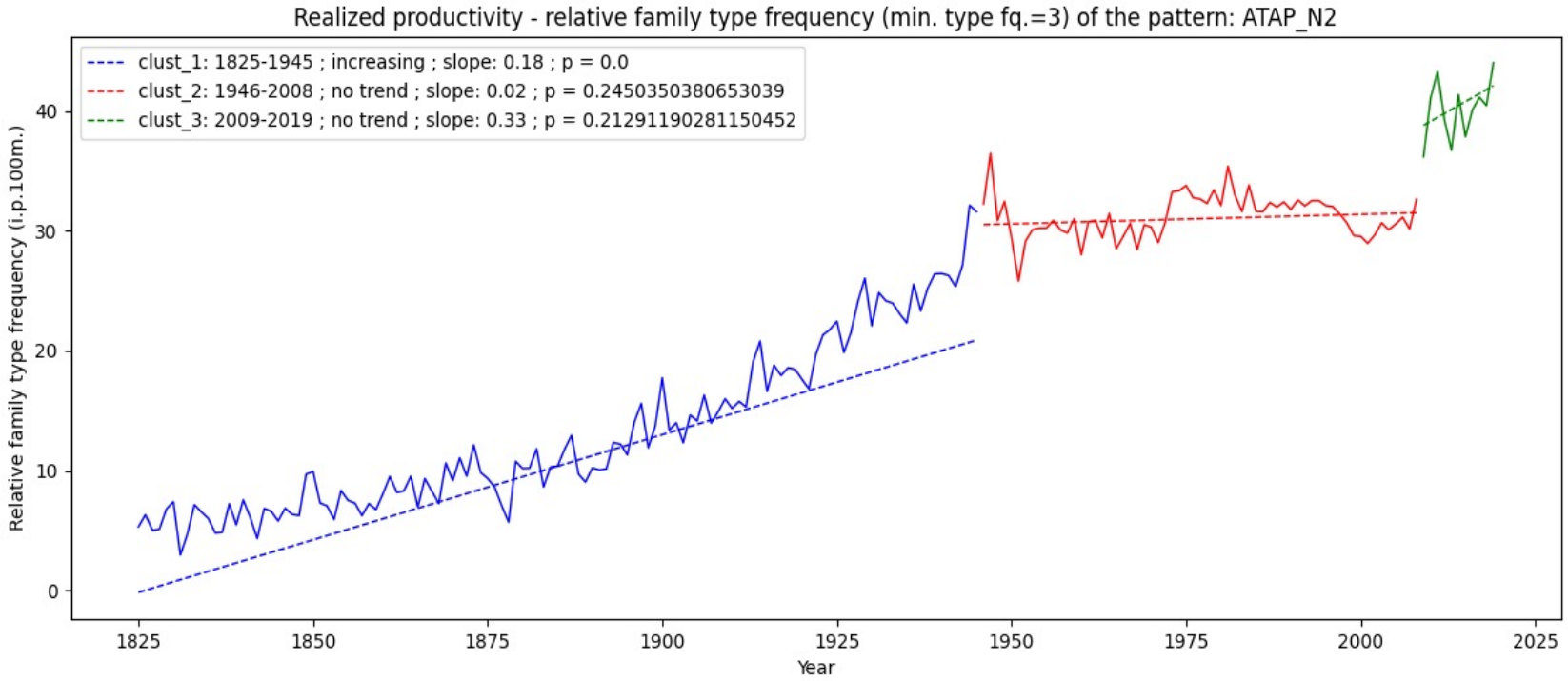
- Still very different curves
- ATAP
  - Total: 56 (of 123) N2s
  - Still no trend between: 1945-2008 (p=0.078, slope=0.026)
- VNX
  - Total: 156 (of 356) N1s
  - Stronger increase - no important periods of “no trend”
- Low coverage of ATAP construction by families
  - ATAP itself is not a vital pattern yet



Focus on the curve of the ATAP_N2 pattern

- Clusters are identified manually, and trends are verified using the Mann-Kendall test
  - R-implementation of the *Variability-based neighbour clustering* (Hilpert & Gries 2009) algorithm does not yield results that seem intuitively meaningful

- ATAP
  - No trend since 1945!



Structural intersection: Do non-triggers make up families?

- **Structural intersection in compounds**
  - Establishes relational links between words based on their similarities (Jackendoff & Audring, 2020:223-225)
  - Based on N1/N2 families in compounds – the only intersection in form (no affix)
  - Triggers are relevant (N2s for ATAP, N1s for VNX)
  - What about non-triggers?
    - In synchronic data, the family-size effect is prominent with both a specified N1 and N2 (Radimský, 2020)

SUB NN semi-schematic construction

$[trasporto\ N_j]_{Nk}$ ⇔

$[TRASPORTO_{i\text{-}head}\ REL\ N_{j\text{-}non\text{-}head}]_k$

$[N_i\ merci_j]_{Nk}$ ⇔

$[N_{i\text{-}head}\ REL\ MERCI\ _{j\text{-}non\text{-}head}]_k$

Individual instances of NNs

noleggio auto

trasporto merci

trattamento rifiuti

scarico merci

trasporto persone

Structural intersection: Do non-triggers make up families?

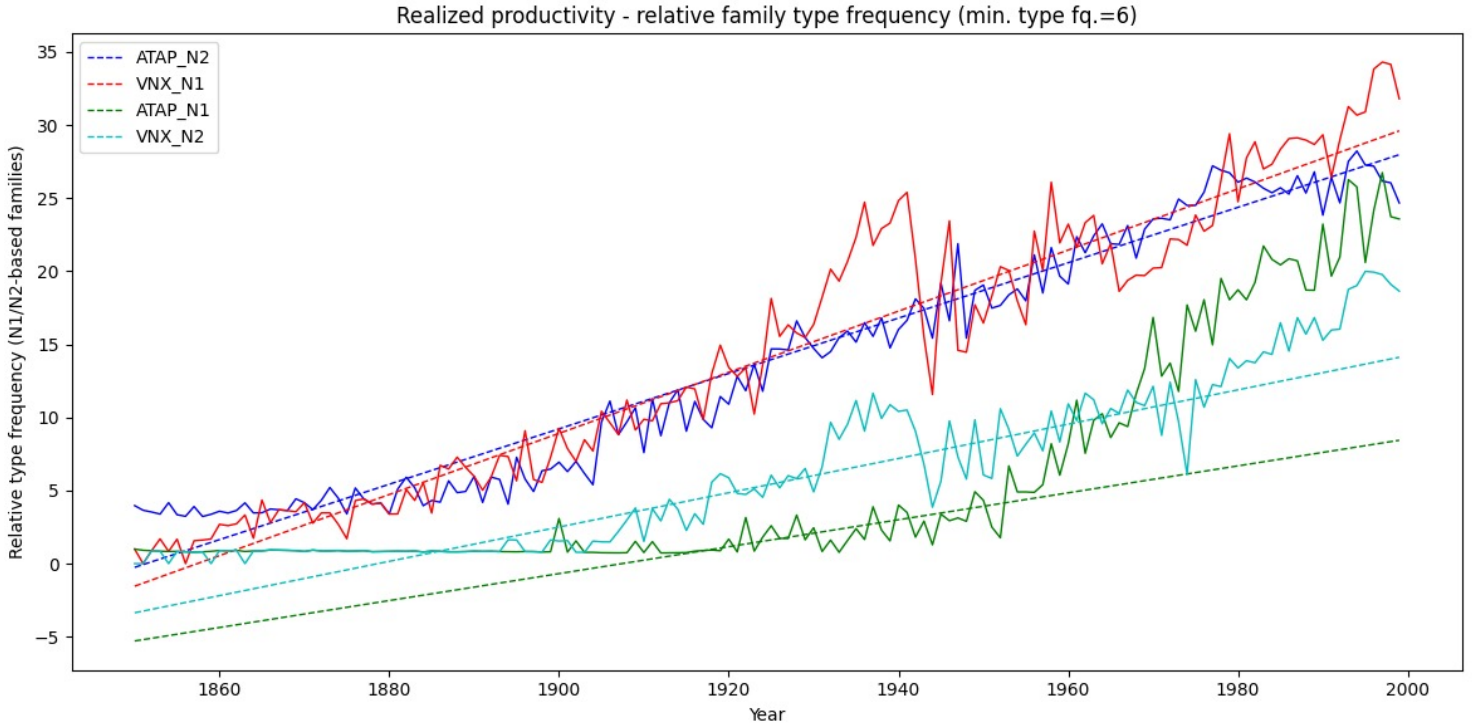
Relative Family type frequency of families with **at least 6 members**

- Families of “triggers” – ATAP_N2, VNX_N1
- Families of “non-triggers” – ATAP_N1, VNX_N2
  - Surprisingly similar values – esp. for ATAP N_2 & ATAP_N1 after 1990’s

- ATAP
  - No new N2 families since 1945
  - BUT: N1s present in a N2 family expand in other N2 families **within the ATAP pattern**

- ***città modello* (model city) →**
- *città* + *giardino* (garden), *bersaglio* (target), *simbolo* (symbol), *fantasma* (ghost), *matrigna* (stepmother), *satellite* (satellite), *dormitorio* (dormitory), *mito* (myth), *ghetto* (ghetto), *partner* (partner)...

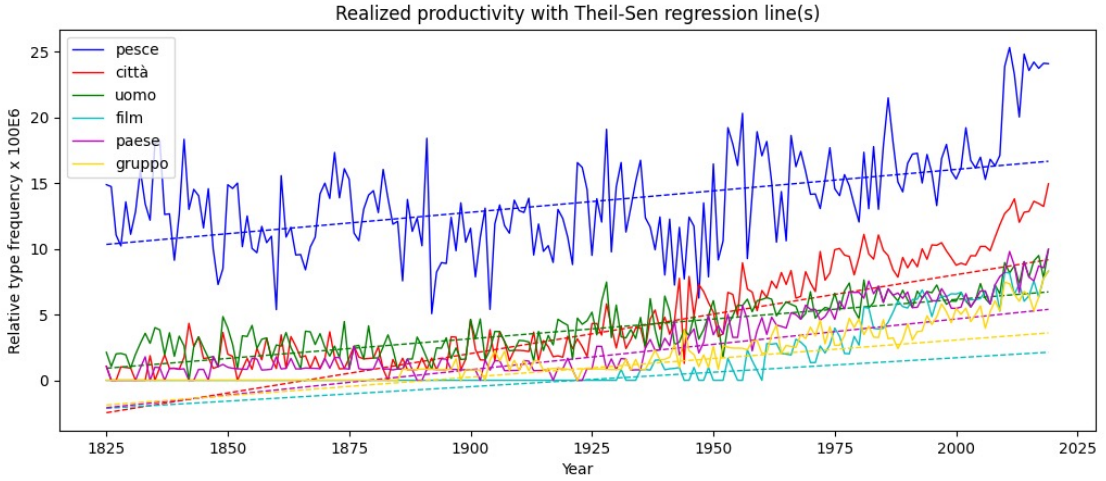
- ATAP pattern has some cognitive relevance?



Other relevant lower-order schemas?

- Form-based families
- *pesce-N* – ancient “island” of appositive compounds
  - *pesce spada* - swordfish
  - *pesce cane* – dogfish, shark
  - *pesce ago* – pipefish
  - *pesce porco* – grey triggerfish (fish.pig)
  - *pesce sega* – sawfish

Prominent N1-based ATAP families



- Further research: sense-based families within SUB_GROUND compounds?
  - Means of transport (car/wagon)
    - *carro + attrezzi / merci / bagagli / bestiame / cavalli...*
      - *towtruck, freight wagon, baggage wagon, cattle car, horse wagon*
    - *vagone + bar / ristorante / bestiame / fumatori / merci / salotto...*
      - *bar wagon, dining car, cattle car, smoking carriage, freight wagon, lounge car*

Conclusions & future work

- Tools for diachronic analysis of compounds
  - Realized productivity of patterns (relative type fq. of compounds)
  - Family type frequency (relative type fq. of form-based families with different size)
    - Analysis of *Coverage,* based on *Structural intersection*
  - Analysis of single families
    - Each family has its own history, identification of leader words (= source of *innovation*)
  - How to identify relevant clusters in the diachronic curves?

- CM/RM: Analysis of diachronic interaction of constructions at different levels of generalization
  - Which form-based families matter, how do they interact with higher-order constructions?
    - Non-trigger families matter
  - Which higher-order constructions are relevant?
    - Is “ATAP” a relevant category?
  - Are there some sense-based lower-order constructions?
  - Need to gather a complex sample with a variety of NNs

References

- Aronoff, Mark & Mark Lindsay. 2014. Productivity, blocking and lexicalization. In Rochelle Lieber & Pavol Štekauer (eds.), The Oxford handbook of derivational morphology, 67–83. Oxford: Oxford University Press.
- Baroni, Marco, Guevara, Emiliano & Roberto Zamparelli. 2009. The dual nature of Deverbal Nominal Constructions: Evidence from acceptability ratings and corpus analysis. Corpus Linguistics and Linguistic Theory 5(1). 27—60.
- Bauer, Laurie. 2017. *Compounds and Compounding*. Cambridge: Cambridge University Press.
- Booij, Geert. 2010. *Construction Morphology.* 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. 2011. Noun-clad adjectives. On the adjectival status of non-head constituents of Italian attributive compounds. *Lingue e linguaggio* 10(1). 161—76.
- 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.

References

- 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. 2022. Towards a diachronic analysis of Romance morphology through Google n-grams. *Corpus* 23. https://doi.org/10.4000/corpus.7205
- 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.