



ATAP N-N compounds in French and Italian

A paradigmatic approach

Jan Radimský

University of South Bohemia (Czech republic)

Plan

- ▶ Compounds such as:
 - ▶ IT *parola chiave*, FR *mot-clé* (keyword)
 - ▶ IT *città simbolo*, FR *ville symbole* (symbolic city)
- ▶ Basic properties and delimitation of ATAP compounds
- ▶ A paradigmatic approach to ATAP compounds:
 - ▶ Do ATAP compounds form paradigms? At what level of abstraction? How do these paradigms interact?
 - ▶ Can we describe the inflection of ATAP compounds as a paradigm?



N-N attributive-appositive (ATAP) compounds

- ▶ Part of N-N structures
 - ▶ On the edge between morphology and syntax – a long debate
 - ▶ Bauer (1998): “there is no strong evidence for a distinction between two [i.e. syntactic and morphological] fundamental types of noun+noun construction”
 - ▶ Unitary treatment: consider almost all N-N structures in Romance languages as “compounds” (Gaeta-Ricca, 2009)
 - ▶ In the Construction grammar, the morphology vs. syntax distinction is not clear-cut
- ▶ N-N ATAP compounds – non-coordinate, N2 is a qualifier (=modifier, attribute) of N1
 - ▶ Scalise-Bisetto classification (2009), three types of relations between N1 and N2 in non-coordinate compounds:
 - ▶ R relationship → Subordinate “grounding” compounds
 - ▶ Unpredictable
 - ▶ Predicate-argument relationship → Subordinate “verbal nexus” compounds
 - ▶ N1 (head) is the predicate, N2 is its argument. E.g.: *trasporto rifiuti* (“waste transport”)
 - ▶ Attributive relationship → ATAP compounds
 - ▶ N2 is the modifier of the head noun N1
- ▶ How to recognize ATAP compounds on formal grounds?



Recognizing ATAP compounds

- ▶ The modifier (N2) may become a predicate
 - ▶ N1 is (a) N2 (Noailly, 1990)
 - ▶ FR *un taux record* – *ce taux est record*
“a record rate” – “this rate is a record”
 - ▶ FR *une ville symbole* – *cette ville est (un) symbole*
“a symbolic town” – “this town is a symbol”
 - ▶ Abstract nouns: FR *record, symbole, limite, victime..*
 - ▶ Concrete nouns with a lexicalized metaphor: FR *clé* (“key”), *phare* (“lighthouse”, i.e. “leading”), *béton* (“concrete”, i.e. “solid”)...
 - ▶ Concrete nouns with a very transparent metaphor (FR *jardin* - “garden” in *ville-jardin* – “garden city”)...
 - ▶ N1-N2 is (a) N2 from the point of view of X (Fradin, 2009)
 - ▶ FR *un requin marteau est un marteau du point de vue de la forme*
“a hammerhead is a hammer from the point of view of its shape”
 - ▶ Concrete nouns with a non-lexicalized metaphor. Ex: *pomme-allumette* - “match (=very thin) fries”
- ▶ ATAP compounds are close to other similar types (the same formal tests are applicable)
 - ▶ Hyponym-superordinate compounds (Bauer)
 - ▶ IT *carcere-lager* (“prison + [concentration camp]”), *medico-dentista* (“doctor-dentist”)
 - ▶ ATAP, N2 is a hyponym of N1
 - ▶ Symmetrical (“reversible”) ATAP compounds – coordinate or coordinate-like
 - ▶ IT *studente-lavoratore* (“student worker”)
 - ▶ N1 and N2 denote closely related concepts situated on the same hierarchy level
 - ▶ Paraphrase: N1 is (a) N2 OR N1 and N2 ?



Classification of N-N ATAP compounds

ATAP					
Attributive (literal interpretation of the modifier)				Appositive (metaphorical interpretation of the modifier)	
Endocentric			Exocentric	Endocentric	Exocentric
Reversible (coordinate-like, symmetric)	Irreversible		-		
Closely related concepts	Hyponym-superordinate	Non-related concepts			
<i>studente-lavoratore</i> <i>lavoratore-studente</i> “student-worker” (and the opposite)	<i>medico-dentista</i> “doctor-dentist”	<i>luogo simbolo</i> lit. “place-symbol” “symbolic place”	?	<i>parola chiave</i> “key word”	<i>madrelingua</i> lit. “mother-tongue” “native speaker”

(see Radimský, 2015:114, 158)

- ▶ We are primarily concerned with “asymmetric” ATAP compounds (in blue)
- ▶ ATAP compounds are difficult to delimit
 - ▶ There is no clear-cut borderline with other types of compounds
 - ▶ Abstract (non-metaphorical) and concrete (metaphorical) modifiers might not necessarily behave in the same way; but the same holds true for lexicalized and non-lexicalized metaphorical modifiers
 - ▶ The formal tests do not always give decisive results

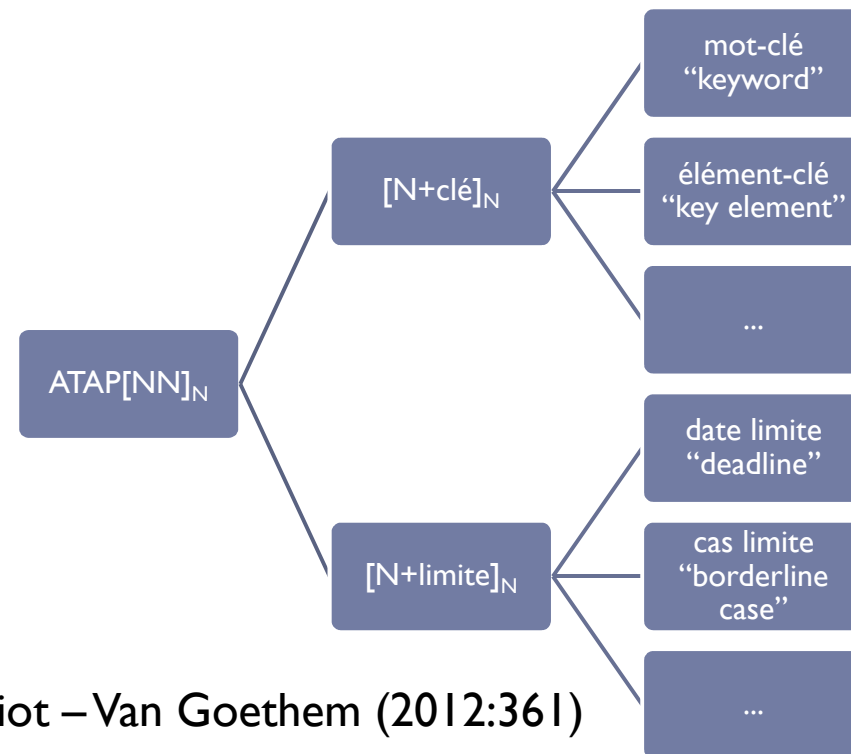


Do ATAP compounds form paradigms?

- ▶ Interpretation of ATAP compounds is triggered by the modifier (N2)
 - ▶ Each modifier might form a “paradigm” – a semi-schematic construction such as [N+clé]_N
- ▶ Do modifiers tend to have an “adjectival distribution” (Baroni-Guevara-Pirreli, 2009)?
 - ▶ High type frequency – modifiers combine with many different head nouns
 - ▶ High relative frequency of a noun in the N2 position
 - ▶ “*pilota* è in posizione N2 in 1/4 delle sue occorrenze nel corpus” – “*pilota* is in the position of nominal modifier in 1/4 of its occurrences in the corpus”

- ▶ Do the following schematic constructions correspond to a “paradigm of relationships” (Bauer, 2017)?
 - ▶ ATAP[N₁N₂]_N (= [N₁ is (a) N₂])

- ▶ Can we describe the inflection of ATAP compounds as a paradigm?
 - ▶ IT: [parola]_{sg} [chiave]_{sg} – [parole]_{pl} [chiave]_{sg}
 - ▶ FR: [mot]_{sg} [clé]_{sg} – [mots]_{pl} [clés]_{pl}



ATAP compounds in Wacky corpora

- ▶ Corpora: ItWac and FrWac
 - ▶ large web corpora, size: $1,9 \times 10^9$ (ItWac) and $1,6 \times 10^9$ (FrWac) positions
- ▶ Databases of binominals based on non-lemmatized frequency lists: ItWac binominals and FrWac binominals
 - ▶ Art/Prep – N1 – N2 (ex.: *la parola chiave, des mots clés*)
 - ▶ N1-N2 hyphenated (ex: *mot-clé*)
 - ▶ Art/Prep in the non-hyphenated version helps eliminating false results (lists of nouns)
 - ▶ Frequency filter: token fq > 3 per type
 - ▶ Automatic filtering and lemmatization
 - FR: GLAFF (Hathout – Sajous – Calderone, 2014)
 - IT: Morph-It (Zanchetta – Baroni, 2005)
 - ▶ In total approx. 700.000 N1-N2 combinations (types) for each language
- ▶ Subsequent manual filtering



Manual filtering

- ▶ Identify the modifiers with a high type frequency
 - ▶ But: important Noun/Adjective homonymy
 - ▶ ATAP N-N compounds do not display gender agreement
 - ▶ Good N2 candidates: high type frequency with “gender mismatch” between N1 and N2
 - ▶ *ruolo, punto, elemento, fattore, concetto* (M) - *chiave* (F)
 - ▶ Italian: 32.000 N2 candidates reduced to 4.700 which were filtered by hand
 - ▶ 147 N2s – 1.800 ATAP compounds
 - ▶ Good results for Italian, but not for French!
- ▶ Identify modifiers through an electronic dictionary
 - ▶ Only a few additional items for Italian (Zingarelli)
 - ▶ French: lower type frequency of N2s, majority of N2 retrieved using a dictionary (Robert)



Type frequency of N2s

▶ ItWac binominals

- ▶ 147 modifiers – 1,800 ATAP compounds
- ▶ On average **12.2** compounds per each modifier

▶ FrWac binominals

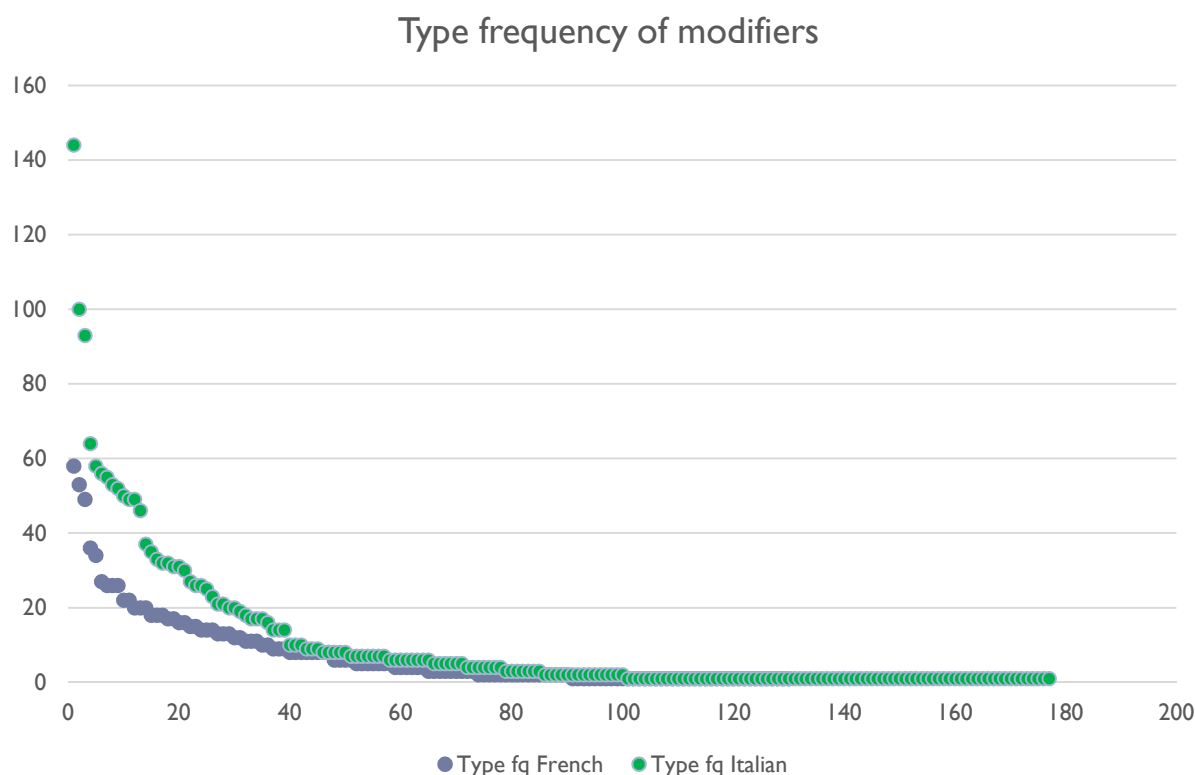
- ▶ 130 modifiers – 1,022 ATAP compounds
- ▶ On average **7.8** compounds per each modifier

▶ Results:

- ▶ The type frequency of Italian modifiers is higher
- ▶ In both languages, 1/3 of modifiers do not have any paradigm (only one head noun)
- ▶ The type frequency of modifiers in ATAP compounds does not substantially differ from the type frequency of non-head nouns in Grounding compounds

▶ Conclusion

- ▶ Only some semi-schematic ATAP constructions form paradigms



Paradigm of number inflection

- ▶ Can we describe the inflection of ATAP compounds as a paradigm?

Italian	French
[[N1] _{sg} [N2] _{sg}] _{sg} [[N1] _{pl} [N2] _{sg}] _{pl} (inflection mark on the head only)	[[N1] _{sg} [N2] _{sg}] _{sg} [[N1] _{pl} [N2] _{pl}] _{pl} (inflection mark on both the head and the modifier)
[[<i>parola</i>] _{sg} [<i>chiave</i>] _{sg}] _{sg} [[<i>parole</i>] _{pl} [<i>chiave</i>] _{sg}] _{pl}	[[<i>mot</i>] _{sg} [<i>clé</i>] _{sg}] _{sg} [[<i>mots</i>] _{pl} [<i>clés</i>] _{pl}] _{pl}

- ▶ Such a description would make sense if the inflection were regular.
- ▶ Unfortunately, this is not the case in either language



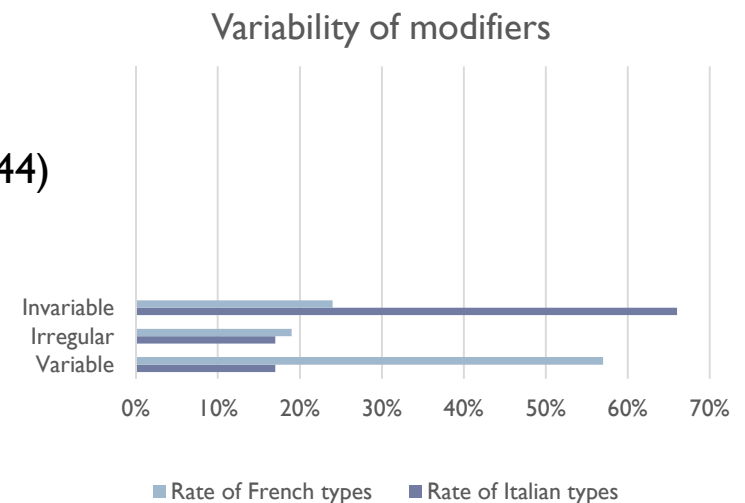
Number inflection of modifiers

▶ Italian: most modifiers are invariable

- ▶ Previous research gave divergent results:
 - ▶ Grandi-Nisim-Tamburini (2011:173): regular inflection of N2s
 - ▶ Baroni-Guevara-Pirreli (2009): rare inflection – only 5,83% tokens
- ▶ Radimský (2015:177): variability of modifiers with plural heads
 - ▶ 77 N2s almost always invariable (=66%)
 - ▶ 20 N2 almost always variable (=17%)
 - ▶ 20 N2s irregular (=17%)

▶ French: most modifiers are variable (Noailly, 1990:44)

- ▶ FrWac binominals database:
 - ▶ 60 N2s almost always variable (=57%)
 - ▶ 25 N2s almost always invariable (=24%)
 - ▶ 21 N2s irregular (=19%)



▶ Note

- ▶ “Almost always (in)variable” means that more than 95% tokens are (in)variable
- ▶ All counts are based on token frequency of compounds with plural heads

Inflection

- ▶ Should a N, as a modifier, be variable? Intuitive answers...
 - ▶ YES: analogy with N-A constructions; moreover, when transformed to predicates, both N and A are variable
 - ▶ *les villes-symboles* – ces villes sont [*des symboles*]_{pl} “these towns are symbols”
 - ▶ NO: assymmetric relationship between N1 and N2; when transformed to a predicate Ns may be invariable
 - ▶ *le città-simbolo* – queste città sono [*un simbolo*]_{sg} “these towns are a symbol”
- ▶ Why are some modifiers variable in Italian?
 - ▶ Only very symmetric modifiers are variable – the modifier functions like a second head
 - ▶ *paesi membri* (“member states”): *i paesi membri dell’UE* = *i membri dell’UE* (“member states of EU = members of EU”)
 - *questi paesi sono membri* / * *un membro* “these countries are members / *a member”
 - ▶ *persone vittime* (“person ‘victim’”): *le persone vittime* = *le vittime*
 - *queste persone sono vittime* / * *una vittima* “these persons are victims / *a victim”
- ▶ Why are some modifiers invariable in French?
 - ▶ Only very asymmetric modifiers are invariable, mass nouns, constructions on the edge of the ATAP class
 - ▶ *dossier béton* (“concrete dossier”, i.e. very solid) – *béton* is a mass noun
 - ▶ *spécialités maison, desserts maison* “‘house’ (=home-maid) specialities / desserts”
 - *ces spécialités sont ?maison* / * *maisons* “these specialities are ‘house’ (home-maid) / *a house”

Interaction of paradigms

- ▶ The paradigm of relationships makes it possible to explain the variation of inflection of ATAP compounds
 - ▶ $\text{ATAP}[N_1 N_2]_N = [N_1 \text{ is (a) } N_2]$
 - ▶ If a noun (N_2) is obligatorily variable in the predicate position, it will also vary in the corresponding modifier position (and vice-versa)
- ▶ What if the paradigm of relationships does not provide a clear answer?
 - ▶ Many nouns, such as *clé / chiave* (“key”) or *simbole / simbolo* (“symbol”), may be variables as well as invariables in the predicate position
 - ▶ However, their variability in the modifier position of ATAP compounds seems much more restricted
 - ▶ Then, the general schematic paradigm $\text{ATAP}[N_1 N_2]_N$ provides a “default” solution, which consists in having:
 - ▶ a variable modifier in French
 - ▶ an invariable modifier in Italian



Conclusions

In Romance ATAP N-N compounds it might be useful to observe at least three types of paradigms:

- ▶ Paradigms that correspond to semi-schematic constructions with selected modifiers such as $[N+cl\acute{e}]_N$
- ▶ A “paradigm of relationships” $ATAP[N_1N_2]_N$ where the relation between N_1 and N_2 may be made explicit by a syntactic paradigm $[N_1 \text{ is (a) } N_2]$
- ▶ An inflectional paradigm
 - ▶ IT: $[N_1]_{sg} [N_2]_{sg} - [N_1]_{pl} [N_2]_{sg}$
 - ▶ FR: $[N_1]_{sg} [N_2]_{sg} - [N_1]_{pl} [N_2]_{pl}$
- ▶ The assumption that the latter two types of paradigms interact makes it possible to explain the inflectional deviations observed in corpus data



References

- ▶ Amiot Dany, Van Goethem Kristel (2012), A constructional account of French -clé 'key' and Dutch sleutel-'key' as in mot-clé/sleutelwoord 'key word'. *Morphology*, 22, pp. 347–364.
 - ▶ Baroni Marco et al. (2006), The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. Online <http://wacky.sslmit.unibo.it/lib/exe/fetch.php?media=papers:wacky_2008.pdf>
 - ▶ Baroni Marco, Guevara Emiliano, Pirrelli Vito (2009), Sulla tipologia dei composti N+N in italiano: principi categoriali ed evidenza distribuzionale a confronto. In: Ruben Benatti, Giacomo Ferrari and Monica Mosca (eds.), *Linguistica e modelli tecnologici di ricerca* (Atti del 40esimo Congresso della Società di Linguistica Italiana). Roma, Bulzoni, pp. 73-95.
 - ▶ Bauer Laurie (1998), When is a sequence of two nouns a compound in English? *English language and linguistics*, Volume 2, Issue 1, Cambridge, 1998, pp. 65-86.
 - ▶ Bauer Laurie (2017), Notions of paradigm and their value in word-formation. ParadigMo Toulouse – Conference book of abstracts.
 - ▶ Fradin Bernard (2009), IE, Romance: French. In: Lieber R., Štekauer P., *The Oxford handbook of compounding*. Oxford University Press, pp. 417-435.
 - ▶ Gaeta Livio, Davide Ricca (2009), Composita solvantur: Compounds as lexical units or morphological objects? *Rivista di Linguistica*, 22/1, pp. 35-70.
 - ▶ Grandi Nicola, Nissim Malvina, Tamburini Fabio (2011), Noun-Clad Adjectives. On the adjectival status of non-head constituents of Italian attributive compounds. *Lingue e linguaggio*, X.1, pp. 161-176.
 - ▶ Hathout Nabil, Franck Sajous and Basilio Calderone (2014), GLÀFF, a Large Versatile French Lexicon. Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pp. 1007-1012, Reykjavik, Iceland.
 - ▶ Noailly Michèle (1990), *Le substantif épithète*. Paris, PUF.
 - ▶ Radimský Jan (2015), Noun+Noun Compounds in Italian. A corpus-based study. Jihočeská univerzita, edice Epistémé, České Budějovice,
 - ▶ Scalise Sergio, Bisetto Antonietta (2009), The classification of compounds, In: Lieber Rochelle, Štekauer Pavol (eds.), *The Oxford handbook of compounding*, Oxford, Oxford University Press, pp. 34-53.
 - ▶ Zanchetta Eros, Baroni Marco (2005): Morph-it! A free corpus-based morphological resource for the Italian language. *Proceedings of Corpus Linguistics 2005*, Birmingham: University of Birmingham.
-

