

Variability in Greek multiword compound nouns in the form of Adjective (Adjective) Noun

In this presentation we are going to discuss morphosyntactic variability in Modern Greek (MG) multiword compound nouns (CN) in the form of Adjective (Adjective) Noun and their automatic processing. This type of CN is subject to shortenings due to adjective substantivation or omission of the adjective. Following Gaston Gross' criteria (1996), we define as CN as a multiword noun whose at least one of the semantic, syntactic or distributional properties cannot be deduced by the properties of each of its constituents.

Substantivation is the process in which adjectives (or participles) are converted into nouns and acquire the paradigm and syntactic functions of nouns. Based on the criteria of predictability and variation of gender and number, Anastasiadis-Simeonidis et al. (2012) differentiate Greek substantivised adjectives (SA) in two categories: SA whose formation is the result of ellipsis and whose gender is the same as the noun eg. **το** (neutral definite article in singular)/*the* εμφιαλωμένο νερό/*mineral water* → **το** εμφιαλωμένο/*the mineral* and SA which arise by conversion and whose gender is not always predictable eg. **η** (feminine definite article in singular)/*the* μαθηματική επιστήμη/*mathematical science* → **τα** (neutral definite article in plural)/*μαθηματικά/the mathematics*. In this example, we can observe that the feminine compound noun is converted into a neutral simple noun. However, the meaning is the same.

Another type of variation is the omission of the adjective of the CN. For instance, it is frequently used the simple noun πρόγραμμα/*show* instead of τηλεοπτικό πρόγραμμα/*TV show*. Moreover, some CN are subject to both phenomena (Kyriacopoulou 2005), eg. αστικό λεωφορείο/*urban bus* → αστικό/*urban* (SA) or λεωφορείο/*bus* (omission of the adjective).

These phenomena may also occur to CN in the form of Adjective + Adjective + Noun (AAN). In some cases, the shortened structure may appear in two different genders (Kyriacopoulou 2005). That happens mostly in CN which describe languages, eg. **η** μεσαιωνική λατινική γλώσσα/*the medieval latin language* (feminine AAN CN) → **η** μεσαιωνική λατινική (feminine AN CN) or **τα** μεσαιωνικά λατινικά/*the medieval latin* (neutral AN CN). In the following example, both an adjective and the noun of the compound structure could be omitted: το παιδικό τηλεοπτικό πρόγραμμα/*the kids TV show* → το παιδικό πρόγραμμα/*the kids show* or το παιδικό/*the kids*. In order to draw a list of CN and their possible variants, we studied two MG dictionaries: *Dictionary of the Modern Greek language* (G. Babiniotis) and *Dictionary of Common Modern Greek language* (Institute of Modern Greek Studies), and we also extracted data from a corpus of approximately 6,000,000 words of Journalistic and Educational Discourse.

Then we decided to extend the variation description to the automatic processing via Multiflex, a cross-language morphosyntactic generator of multiword units developed by A. Savary (2005a/b/c, 2007). This tool, which has been integrated in the Unitex processor (Paumier 2012), supports finite-state transducers that can model a number of combining conditions. Furthermore, Multiflex gives the possibility to generate the inflected forms of a compound structure as well as the inflected forms of its variants. Given the following inflection graph for αστικό λεωφορείο/*urban bus*:

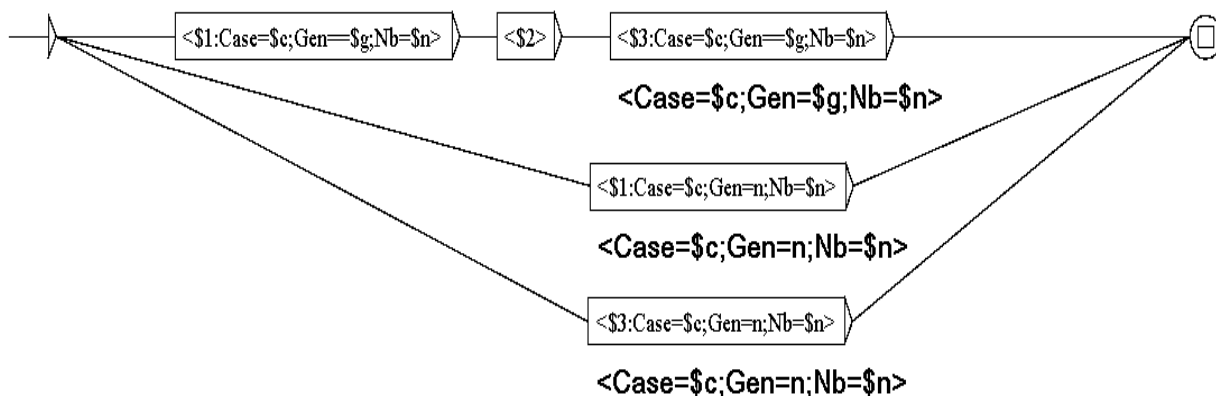


Figure 1: Inflection graph for αστικό λεωφορείο/*urban bus*

The upper path of the graph allows to generate all the inflected forms of the CN both in singular and in plural: three constituents are present in αστικό λεωφορείο/*urban bus*: αστικό/*urban* referred to as \$1, a space referred to as \$2, and λεωφορείο/*bus* referred to as \$3. The middle and the lower path represent the variants of the CN. Following the middle path, only the first constituent's inflected forms (\$1=αστικό/*urban*) will be generated and following the lower path only the third constituent's inflected forms will be generated (\$3=λεωφορείο/*bus*). As the graph on Figure 1 contains three different paths, the whole set of inflected forms generated for αστικό λεωφορείο/*urban bus* would be:

αστικό λεωφορείο, αστικό λεωφορείο.N:Nns	(urban bus)
αστικά λεωφορεία, αστικό λεωφορείο.N:Nnp	(urban buses)
αστικού λεωφορείου, αστικό λεωφορείο.N:Gns	(of the urban bus)
αστικών λεωφορείων, αστικό λεωφορείο.N:Gnp	(of the urban buses)
αστικό λεωφορείο, αστικό λεωφορείο.N:Ans	(urban bus)
αστικά λεωφορεία, αστικό λεωφορείο.N:Anp	(urban buses)
αστικό λεωφορείο, αστικό λεωφορείο.N:Vns	(urban bus)
αστικά λεωφορεία, αστικό λεωφορείο.N:Vnp	(urban buses)
αστικό, αστικό λεωφορείο.N:Nns	(urban)
αστικά, αστικό λεωφορείο.N:Nnp	(urbans)
αστικού, αστικό λεωφορείο.N:Gns	(of the urban)
αστικών, αστικό λεωφορείο.N:Gnp	(of the urbans)
αστικό, αστικό λεωφορείο.N:Ans	(urban)
αστικά, αστικό λεωφορείο.N:Anp	(urbans)
αστικό, αστικό λεωφορείο.N:Vns	(urban)
αστικά, αστικό λεωφορείο.N:Vnp	(urbans)
λεωφορείο, αστικό λεωφορείο.N:Nns	(bus)
λεωφορεία, αστικό λεωφορείο.N:Nnp	(buses)
λεωφορείου, αστικό λεωφορείο.N:Gns	(of the bus)
λεωφορείων, αστικό λεωφορείο.N:Gnp	(of the buses)
λεωφορείο, αστικό λεωφορείο.N:Ans	(bus)
λεωφορεία, αστικό λεωφορείο.N:Anp	(buses)
λεωφορείο, αστικό λεωφορείο.N:Vns	(bus)
λεωφορεία, αστικό λεωφορείο.N:Vnp	(buses)

The inflected forms of the CN αστικό λεωφορείο/*urban bus* in all four cases in singular and plural and of the simple words αστικό/*urban* and λεωφορείο/*bus* are followed by the morphological features N=noun, Nns=Nominative neutral singular, Nnp=Nominative neutral plural, Gns=Genitive neutral singular, Gnp=Genitive neutral plural, Ans=Accusative neutral singular, Anp=Accusative neutral plural, Vns=Vocative neutral singular, Vnp=Vocative neutral plural.

The graph presented above could be used for any compound noun which inflects in the same way as αστικό λεωφορείο/*urban bus* due to Multiflex's unification-based formalism. Actually, in order to predict all possible variants of Greek multiword compound nouns in the form of Adjective (Adjective) Noun, we have created thirty (30) inflection graphs.

Bibliography

- Anastassiadis-Symeonidis A. (1986). *Neology in Modern Greek* (in Greek), PhD Thesis, Thessaloniki: Aristotle University of Thessaloniki.
- Anastassiadis-Symeonidis A., Efthimiou A. & A. Fliatouras (2003). "Phenomena of substantivation in Modern Greek" (in Greek), In *Proceedings of the 6th International Conference "Intercultural Education-Greek as a second/foreign language"*, Patra: University of Patra, pp. 385-402.
- Babiniotis, G. (1998). *Dictionary of Modern Greek language* [in Greek]. Athens: Centre of Lexicology.
- Giannakidou, A. & M. Stavrou (1999). "Nominalization and Ellipsis in the Greek DP". *The Linguistic Review* 16: pp. 295-331.

- Gross G. (1996). *Les expressions figées en français : Noms composés et autres locutions*, Paris: Ophrys.
- Kyriacopoulou T. (2005). *Analyse automatique des textes écrits : le cas du grec moderne*. Thessaloniki, University Studio Press.
- Paumier S. (2012). *Unitex 3.0 – User Manual* [1st version: 2003], <http://www-igm.univ-mlv.fr/~unitex/UnitexManual3.0.pdf>.
- Savary A. (2005a). “Towards a Formalism for the Computational Morphology of Multi-Word Units”, In Vetulani (ed.) *Human Language Technologies as a Challenge for Computer Science and Linguistics, Proceedings of the 2nd Language & Technology Conference, Poznań, Poland, April 21-23*.
- Savary A. (2005b). “A Formalism for the Computational Morphology of Multi-Word Units”, *Archives of Control Sciences* 15: 3, pp. 437-449.
- Savary A. (2005c). *MULTIFLEX. User’s Manual and Technical Documentation. Version 1.0. Technical Report 285*, LI-François Rabelais University of Tours: France.
- Savary A., C. Krstev & D. Vitas (2007). “Inflectional non compositionality and variation of compounds in French, Polish and Serbian, and their automatic processing”, In Dziadkiewicz A. & I. Thomas (eds.) « Les langues slaves et le français : approches formelles dans les études contrastives », *Bulag* 32, Presses de universitaires de Franche-Comté, pp. 73-93.
- Triantafyllidis M. Foundation (1996). *Dictionary of Common Modern Greek language* [in Greek]. Institute of Modern Greek Studies: Thessaloniki.