Electronic Database of New Words in Greek

Anna Anastassiadis-Syméonidis
Aristotle University of Thessaloniki
A.U.TH. Department of Philology, GR-54124, Thessaloniki, Greece
ansym@lit.auth.gr

Georgia Nikolaou
Aristotle University of Thessaloniki
A.U.TH. Department of French Language and Literature, GR-54124, Thessaloniki, Greece
ngeorgia@lit.auth.gr, georgetten48@hotmail.com

Abstract
This paper deals with the development of an electronic database of new words in Greek, which constitutes a part of the research project "Pythagoras". Our aim is (a) to create an electronic inventory of all the new Greek words that are used in non scientific Greek language magazines, (b) to offer precise information about the semantic frame, the morphological features and the special register of all the lexical units which will be included in the database, (c) to make useful correlations between those units, as far as their form and meaning are concerned, and (d) to elaborate a lexicographic product that will enable us to use it for future linguistic descriptions and/or analyses of the Greek language. In the present paper our corpora and scientific methods will be presented, as well as a small sample of the lexical units which are going to be included in our database.

1 Description of the lexicographic product

The lexicographic product that we are going to present constitutes an electronic means of recording, sorting and classifying the lexical units of Modern Greek language which are not currently included in any general dictionary of Greek language, special terminology vocabulary or electronic database of Greek terms. More specifically, the main topic of this paper is the creation of an electronic database of new Greek words, which have been found and selected during the last eight years (1998-2006) under the supervision of the professor of the Aristotle University of Thessaloniki Ms. Anna Anastassiadis-Syméonidis.

2 Development of the scientific application

2.1 Framework

This scientific application constitutes a part of a national research project called *Pythagoras*, which is sponsored by the Greek Ministry of National Education and Religious Affairs. The duration of the project is set for twenty-one months, starting in April 2005 and finishing in December 2006. The first part of the project, in which we are at the moment, consists of: (a) the selection of literature, references and methodological tools that are going to be used

for the elaboration of the electronic database, and (b) the laying down of the list of head-words that are going to be included in the final lexicographic product. During the second phase of our work, we plan to proceed to the technical development of our application: the input of the new Greek words and the linguistic information that accompany them, as well as the presentation of a part of the electronic database on a special web page.

2.2 Description of our corpus

Our corpus includes approximately 10 000 lexical units, which all belong to the *Archive of Neologisms*, created and run by Ms Anna Anastassiadis-Syméonidis. This Archive has been enriched with new lexical items every year since 1979. It includes new words of the general vocabulary and does not include literary, scientific or technical terms, since the aim of its creation has been to record new words that are used in everyday life by the average Greek speaker; see Anastassiadis-Syméonidis et al. (1997: 362). Only a part of this Archive is going to be used for the elaboration of an up-to-date electronic database of new words in Greek. The units which are going to form the list of headwords are going to be selected by three criteria:

(a) The apparition of lexical units in written texts published during the last eight years (1998-2006). We have chosen to include the new words of the last eight years, since 1998 was the year when the two main general dictionaries of Modern Greek language were published.[1] These dictionaries are extensively used by the majority of Greek population and have not been considerably renewed since their first edition. Some largely used lexical items which first appeared in texts published before 1998 and are not registered in the two dictionaries mentioned above (e.g. *σπόιλερ*)[2] can also be included in our final inventory, since they need to be explained and codified for the average users of the Greek language.

(b) The list of headwords in the two main general dictionaries of Modern Greek language. As mentioned above, our final list will not include words that are found in those two dictionaries. Some of the lexical units which will be included in our database are registered in the *Reverse Dictionary of Modern Greek Language* by Anna Anastassiadis-Syméonidis (2002) (see *References*), which consists of 180 000 words and constitutes the largest and most updated dictionary of Modern Greek language at the moment. Nevertheless, we have chosen to record those units as well, since the *Reverse Dictionary of Modern Greek Language* does not offer any kind of semantic or etymological information that could help the average users of Modern Greek.

(c) The high frequency of the lexical units found in published texts which address the average Modern Greek speaker. Our aim is to give priority to the units that are most largely used by the native Greek speakers, since our inventory will contain a finite number of lexical items (see also below). For instance, if our Archive of Neologisms includes 4 000 units that

---

[1] These two dictionaries are: *The Dictionary of Standard Modern Greek Language (Λεξικό της Κοινής Νέας Ελληνικής)* of the Institution Manolis Triantaphyllidis and the *Dictionary of Modern Greek Language (Λεξικό της Κοινής Νέας Ελληνικής Γλώσσας)* by G. Babiniotis.

[2] Its meaning is *spoiler* (= a kind of wind that is mounted on the rear of the car).

have appeared during the last eight years and have not been registered in any other lexico-graphic source, our final selection will be based upon the number of times that every unit appeared in the sources, so that the units that are left out will be the less popular ones. Therefore, lexical items which have only appeared once in the sources will most likely be excluded from our final list of headwords.

The number of lexical units that will finally be selected for our list of headwords is 3 000. The lexical items included will belong to all parts of speech (apparently there will not be any grammatical words, since they belong to the part of vocabulary that is not prone to renewal). We will also include items which consist of more than one words (e.g. οθόνη οπίσθιου φωτισμού)[3] as well as items which are written in Latin characters, since they appear in Greek texts in a large scale. In cases where the lexical items appear in more than one forms (e.g. κάζουαλ and *casual*), all the forms of the lexical units will be recorded.

Regarding the fields where the new words belong, here are the main ones; also see Anas-tassiadis-Syméonidis et al. (1997: 363):
- fashion
- aesthetics, face and body care, daily hygiene
- computers and technical devices accessible to the average speaker
- cars and motorcycles
- popularized science (psychology, ecology, environmental issues, economy)
- recreation, amusement
- interpersonal relations, social life, lifestyle
- sports
- media

The sources of our corpus are articles from Greek magazines whose readers are mainly average speakers. The texts of our corpus do not address people with specialized interests.[4]

2.3 Linguistic information included in the electronic database

The linguistic information included in the electronic database after the headword is the following:
- the word class that the lexical unit belongs to, i.e. its part of speech (in case we have an item consisting of more than one words, its grammatical category will be defined by the word that constitutes the head of the lexical item, e.g. the item οθόνη οπίσθιου φωτισμού[5], which consists of two nouns and an adjective, will be classified as a noun)
- phonetic transcription of the lexical unit, with the use of the International Phonetic Alphabet
- the field (domain) to which the item belongs (e.g. fashion, sports)

---

[3] Its meaning is *back lighting screen*.
[4] The texts which constitute our sources are published mainly in the following editions: *Cosmopolitan, E (Epsilon), Elle, Κλικ, Madame Figaro, Marie Claire, PC Magazine, Vita*.
[5] *Back lighting screen*.

• the kind of text in which the lexical unit appears (e.g. article, advertisement, caption etc.)

• the meaning of the lexical unit

• the process of the morphological construction of the lexical unit (derivation, compounding, conversion) or the process of its borrowing from a language-source (direct loanword, calque etc.)

• the context in which the item is found in the texts

• a reference to lexical units that are related to the item, as far as their form or meaning are concerned. This kind of information largely depends upon the context of the item (e.g. $\kappa\lambda\iota\kappa$ "click", a term used in the computers jargon, and $\kappa\lambda\iota\kappa$ "switching on", a term used to describe interpersonal relations)

• the source of the new word, i.e. the title of the magazine, the number of the issue and the page where the new word was found

• an indicator of high frequency for the 500 most used new words which appear in our sources and are included in our electronic database (a number rating approximately at 15% of our list of new words).

2.4 Sample

The following table shows an example of how the new words are going to be treated in the electronic database.

ριαλιτικός-ή-ό Ε "adj."
rialitik'os
Διασκέδαση "amusement"
Άρθρο "article"
Ερμηνεία "meaning": που έχει σχέση με τα ριάλιτι, δηλαδή τα τηλεπαιχνίδια που έχουν ως θέμα τους την καθημερινή ζωή των ανθρώπων που συμμετέχουν σε αυτά "that has to do with reality shows"
Μονολεκτική κατασκευασμένη μονάδα που προήλθε από παραγωγή, με την εφαρμογή του επιθήματος -ικός στη δάνεια αγγλική βάση reality. "One-word constructed unit which was created by derivation, with the application of the affis -ikos to the english loan base form reality."
...Ορθά επιστρέφουν τα αρχαία στα σχολεία μας, τι θα γίνει όμως με τις ριαλιτικές διαλέκτους; "...It's good that the teaching of Ancient Greek is back in the school curricula, but what about reality shows dialects?"
Ε, τεύχος "issue" 709, 14 Νοεμβρίου 2004 "14 November 2004", σελ. "page" 10

Table 1. Sample of the electronic database

2.5 *Future Perspective*

To sum up, we would like to refer to the future perspectives and the advantages of our linguistic application:

1. quick and easy enrichment with new lexical units, since an electronically organised database already exists and is ready for further elaboration

2. development of the metalexicographic research, since the current linguistic application can serve as an example for the creation of other tools of linguistic description and study. It is also possible that with the help of statistical analysis the electronic database of new words can be used for the updating of the existing general dictionaries of Modern Greek language (the high frequency indicator could play an important role at this point)

3. contribution to the linguistic analysis of Greek language: with the help of the words included in the electronic database, a number of useful remarks can be made, concerning the productive patterns of Modern Greek language, the productivity of the linguistic items (prefixes, suffixes), the conditions of vocabulary renewal etc.

4. contribution to the teaching of Greek language as a foreign language, since the electronic database includes linguistic information that cannot be found in other books, dictionaries or databanks and is highly accessible to every user.

References
A. Dictionaries
Αναστασιαδη-Συμεωνίδη, Α. (2002), *Αντίστροφο Λεξικό της Νέας Ελληνικής*, Ινστιτούτο Νεοελληνικών Σπουδών ['Ιδρυμα Μανόλη Τριανταφφυλλίδη].

Ayto, J. (1990) *The Longman Register of New Words*, Vol. II. Longman.

Λεξικό της Κοινής Νεοελληνικής (1998), Ινστιτούτο Νεοελληνικών Σπουδών ['Ιδρυμα Μανόλη Τριανταφφυλλίδη] Θεσσαλονίκη: Α.Π.Θ.

Μπαμπινιώτης Γ. (2002²) [1998], *Λεξικό της Κοινής Νεοελληνικής Γλώσσας*. Αθηνα: Κέντρον Λεξικολογίας.

Tulloch, S. (1991), *The Oxford Dictionary of New Words*. Oxford University Press.

B. Other Literature
Αναστασιαδη-Συμεωνίδη, Α. (1986), *Η νεολογία στην κοινήν νεοελληνική*, Επιστημονική Επετηρίδα της Φιλοσοφικής Σχολής του Α.Π.Θ., Παράρτημα 65.

Αναστασιαδη-Συμεωνίδη, Α. & Βαβατζάνη, Π. & Γαλανή, Σ. & Σταυριανάκη Αικ. (1997), «Το λεξικό της νεοελληνικής». *Ελληνική Γλώσσα και Ορολογία-Ανακοινώσεις Iου Συνεδρίου*, Αθήνα, pp. 361-369.