# Description of predicative nouns in a Modern Greek financial corpus

**Rania Voskaki, Eleni Tziafa & Kyriaki Ioannidou**

*Aristotle University of Thessaloniki*
*rvoskaki@hotmail.com, etziafa@auth.gr, kiroanni@auth.gr*

**Abstract**
This paper reports on a corpus-based description of predicative nouns in a register-diversified financial corpus. Structural linguistics (Chomsky 1981) and register analysis (Biber & Conrad 2009) are the theoretical backgrounds of this research. As predicative noun, we define a noun derived from a verb, an adjective or a noun that occurs in support verb constructions (Gross 1981).

In order to identify the predicative nouns occurring in a Modern Greek financial corpus we applied a. five Lexicon-Grammar tables containing predicative nouns, along with their distributional and transformational properties (Tziafa 2012); b. 122 finite state automata (Ioannidou 2013), representing noun phrases.

**Keywords:** Modern Greek, predicative nouns, noun phrases, financial corpus, financial terminology, Natural Language Processing

## 1. Introduction

The current research deals with the description of predicative nouns in a Modern Greek financial corpus[1]. Initially, the objectives of the current research will be presented, along with the theoretical and methodological framework adopted, the linguistic resources created, which are the Modern Greek financial corpus, the Lexicon-Grammar tables and the finite state automata. Furthermore, we will present the results obtained and we will discuss on the conclusions we came up with as well as on the perspectives that we will endeavor to fulfill through future research.

The main objective of our research was the corpus-based description of predicative nouns in a register-diversified financial corpus. For that reason we divided the financial corpus in four sub-corpora according to their register, in order to study the

[1] It pertains to the field of Natural Language Processing and it is part of the research activities of the Laboratory of Translation and Language Processing, at the School of French Language and Literature, in the Aristotle University of Thessaloniki (http://www.frl.auth.gr/index.php/gr/structure-gr/laboratories-gr/laboratory-translation-speech-gr). The Laboratory of Translation and Language Processing is a member of RELEX, a network of laboratories in the domain of computational linguistics.

frequency of use of predicative nouns in each sub-corpus. Therefore, the results of our research could be useful in enhancing learning of Modern Greek as a foreign/second language in special interested groups. Also, we studied the context of predicative nouns using electronic linguistic resources.

In order to formalise the linguistic data[2], we were based on the syntactic framework of the Transformational Grammar as defined by Chomsky (1981). For the description of predicative nouns we have adopted the methodology framework developed by Gross (1981) and Giry-Schneider (1978). For the analysis of the corpus in registers, we were based on Biber & Conrad (2009). As it concerns the representation of our description we have used the finite state automata as they are proposed by Silberztein (1993) and Roche (1992). The creation of the finite state automata, the corpus processing as well as the concordances, were produced by means of Unitex, a corpus processing system based on automata oriented technology, developed by Paumier (2003).

It has to be mentioned that there is no measurement test on the implementation part of our research. Nevertheless, we intend to use in forthcoming research two basic evaluation measures: precision and recall.

## 2. Predicative nouns

As regards predicative nouns, we relied on the work conducted at the Documentary and Linguistic Automation Laboratory in Paris-Est Marne-la-Vallée[3], for French language and also the research conducted for Greek by Fotopoulou (1989), Moustaki (1995), Sklavounou (1994), Lamprou (1997), Kyriacopoulou & Sfetsiou (2002), Sfetsiou (2007), Fista & Kyriacopoulou (2009). In support verbs constructions, predicative nouns are considered as the core part of the phrase, selecting their own complements. Their function is that of a predicate. Their combination with the verb indicates an action (e.g. κάνω επίθεση, make an assault), semantically defined by the predicative noun (e.g. επίθεση, assault) and is placed in a timeline with the help of the verbal element (e.g. κάνω, make, έκανα, made, etc.). The notion of support verb initially appeared in Jespersen (1965), and in the publications of the Documentary and

[2] Part of this research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" – "Investing in knowledge society" of the National Strategic Reference Framework (NSRF) 2007-2013 (Research Funding Program: Heraclitus II. Investing in knowledge society through the European Social Fund).
[3] http://infolingu.univ-mlv.fr/

Linguistic Automation Laboratory, by Daladier (1978) and Giry-Schneider (1978). It was further developed by Gross (1981).

The undermentioned types of structures, illustrated in square brackets, occur, in large-scale frequency, among predicative nouns:

*(i)* $N_{pred}$ *[N* $N_{gen}$*]*

(1) [Αύξηση μετοχικού κεφαλαίου] του Ιασώ General
([Increase of capital] of Iaso General) (Literal translation)

*(ii)* $N_{pred}$ *[N Prep* $N_{acc}$*]*

(2) [Αύξηση στο μετοχικό κεφάλαιο] του Ιασώ General
([Increase in the capital] of Iaso General) (Literal translation)

*(iii)* $N_{pred}$ *[Abbr]*

(3) [ΑΜΚ] του Ιασώ General
(Capital increase of Iaso General)

In the first example, there is a predicative noun in the form of a compound word 'αύξηση μετοχικού κεφαλαίου' (increase of capital). In the second example there is a prepositional variation of the same predicative noun 'αύξηση στο μετοχικό κεφάλαιο' (increase in the capital). In the third one, the predicative noun is used as an abbreviation (ΑΜΚ, capital increase).

### *2.1 Linguistic resources and frequency study*

In the framework of our research we have created a Modern Greek financial corpus consisting of almost 19,000,000 running words (Tziafa 2012), which we compared to a reference corpus of general discourse comprising 4,500,000 running words. Also, we used a morphological dictionary of 2,000,000 flectional forms of general discourse, elaborated by the Laboratory of Translation and Language Processing and we created a morphological dictionary of 72,017 flectional forms of financial terms.

The Modern Greek financial corpus used is specialised, monolingual and synchronic; it consists of four sub-corpora. The first sub-corpus comprises posts in public discussions with Stock Market as subject (web business discourse). The second sub-corpus concerns journalistic articles in printed or electronic form (journalistic discourse). The third sub-corpus consists of press releases, annual reports of the Athens Stock Exchange and official documents (official business discourse) and the

fourth one comprises academic texts (academic discourse). The overall corpus covers the time period from 1999 to 2010. The corpus structure and registers are illustrated in the following table:

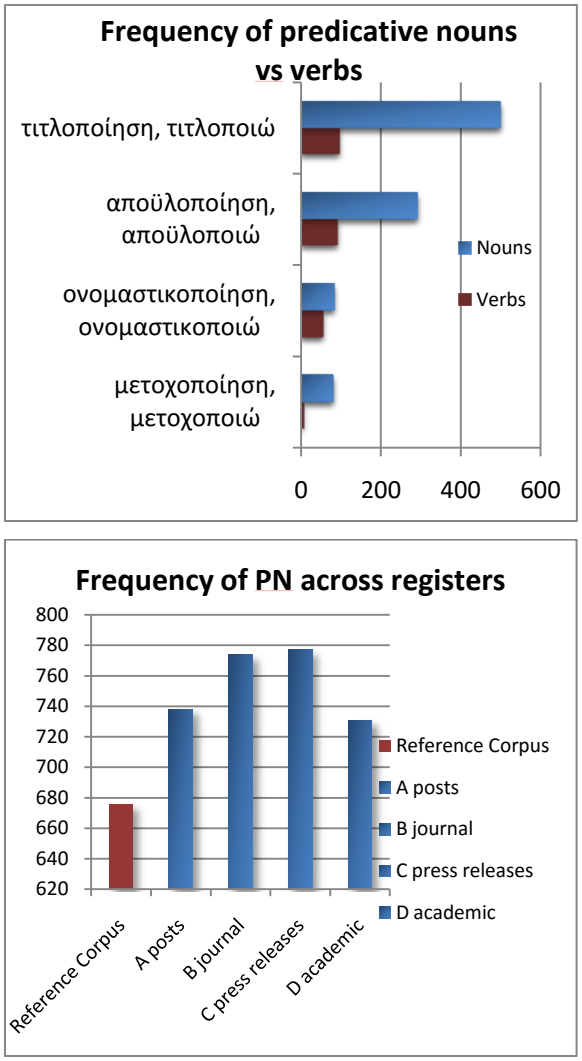
| Sub-corpora | Source | Time Period | Number of Texts | Words |
|---|---|---|---|---|
| A. Posts in public discussions with Stock Market as subject | Internet fora (http://www. neoforum.gr http://www. capital.gr) | 2009-2010 2009-2010 | 46,912 posts | 4,459,377 |
| B. Journalistic articles, in printed or electronic form | Financial newspapers Stock Market web-pages | 1999–2000 2000-2010 | 12,407 articles | 5,114,460 |
| C. Press Releases, Annual Reports of the Athens Stock Exchange, official documents | Web-page of the Athens Stock Exchange (http://www.ase.gr) | 2000-2010 | 18,199 press releases, 21 annual reports, 47 documents | 5,663,526 |
| D. Academic texts | Lecture notes, dissertations and articles | 2002-2010 | 350 texts | 3,567,942 |
| | | | **Total** | **18,805,305** |

*Table 1. Corpus structure and registers*

We studied the importance of predicative nouns in the Modern Greek financial corpus and we found out that there is a clear preference to predicative nouns versus verbs of the same root and meaning in all sub-corpora. For example, the predicative nouns ‘securitisation’ and ‘dematerialisation’ are used more frequently than the verbs ‘securitise’ and ‘dematerialise’. It is worth mentioning that the Greek terms ‘τιτλοποίηση’ and ‘μετοχοποίηση’ are synonymous in English (meaning ‘securitisation’).

Furthermore, we studied the importance of predicative nouns in Modern Greek financial corpus compared to a reference corpus of general discourse. The predicative nouns are used in a significantly higher frequency in the financial corpus. In parallel, we studied the frequency of predicative nouns across registers. The results show that predicative nouns prevail in press releases, followed by journalistic texts, computer mediated communication and academic texts.

In the following diagrams it is possible to make observations about the frequency and importance of the predicative nouns in the financial corpus used in the framework of this research.



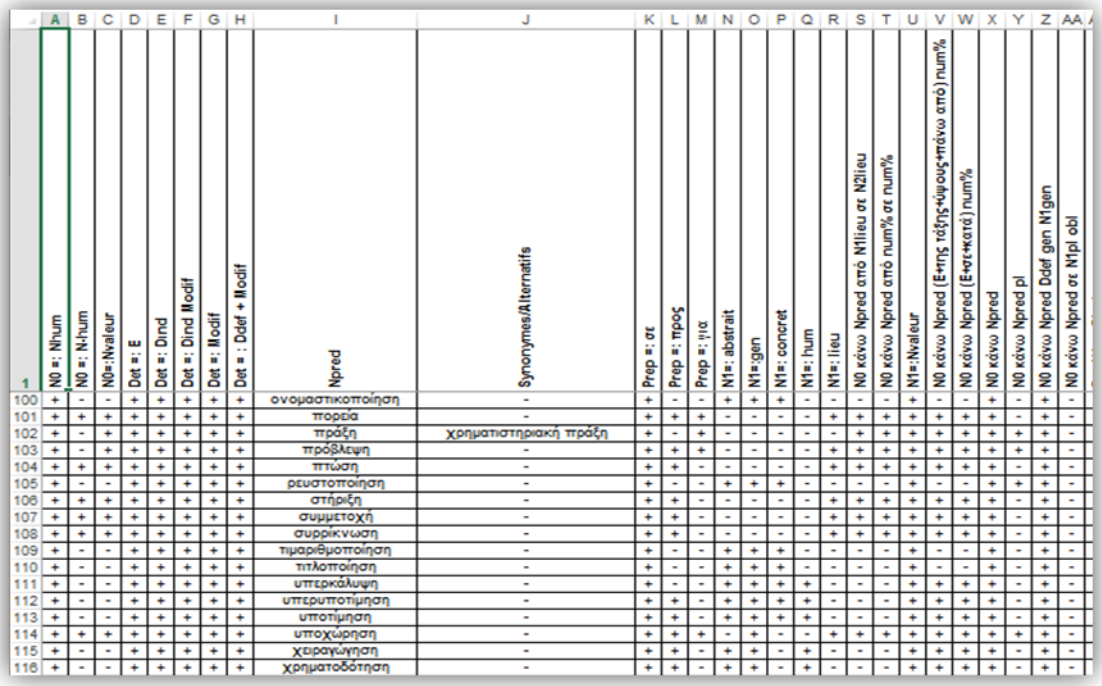

*Diagrams 1 & 2. Frequencies of predicative nouns in the Modern Greek financial corpus*

As it concerns the frequency of structures, the results show that the structure consisting of a predicative noun and a noun in genitive prevail in official documents, in journalistic and academic texts. On the contrary, the most frequent structure in computer mediated communication is the predicative noun followed by a determiner and a noun in genitive.

### *2.2 Lexicon-Grammar tables for predicative nouns*

As for the description of predicative nouns, we were based on the methodology framework proposed by Gross (1981). More precisely, Gross (1975) proposed the creation of syntactic and semantic lexicons of predicates that have the form of matrices and are called Lexicon-Grammar tables. These tables constitute an extremely rich and exploitable linguistic resource, as they contain a set of syntactic and semantic properties, which are studied individually for each predicate.

In figure 1, follows an extract of a Lexicon-Grammar table of predicative nouns:



| 1 | N0 =: Nhum | N0 =: N-hum | N0=:Nvaleur | Det =: E | Det =: Dind | Det =: Dind Modif | Det =: Modif | Det = : Ddef + Modif | Npred | Synonymes/Alternatifs | Prep =: σε | Prep =: προς | Prep =: για | N1=: abstrait | N1=:gen | N1=: concret | N1=: hum | N1=: lieu | N0 κάνω Npred από N1lieu σε N2lieu | N0 κάνω Npred από num% σε num% | N1=:Nvaleur | N0 κάνω Npred (Ε+της τάξης*ύψους*πάνω από) num% | N0 κάνω Npred (Ε+σε+κατά) num% | N0 κάνω Npred | N0 κάνω Npred pl | N0 κάνω Npred Ddef gen N1gen | N0 κάνω Npred σε N1pl obl |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 100 | + | - | - | + | + | + | + | + | ονομαστικοποίηση | - | + | - | - | + | + | + | - | - | - | - | + | - | - | + | - | + | - |
| 101 | + | + | + | + | + | + | + | + | πορεία | - | + | + | + | - | - | - | - | + | + | + | + | + | + | - | + | - |
| 102 | + | - | + | + | + | + | + | + | πράξη | χρηματιστηριακή πράξη | + | - | + | - | - | - | - | - | + | + | + | + | + | + | + | - |
| 103 | + | - | + | + | + | + | + | + | πρόβλεψη | - | + | + | + | - | - | - | - | + | + | + | + | + | + | + | + | - |
| 104 | + | + | + | + | + | + | + | + | πτώση | - | + | + | - | - | - | - | - | + | + | + | + | + | + | - | + | - |
| 105 | + | - | - | + | + | + | + | + | ρευστοποίηση | - | + | - | - | + | + | + | - | - | - | - | + | - | - | + | + | + | - |
| 106 | + | + | + | + | + | + | + | + | στήριξη | - | + | + | - | - | - | - | - | + | + | + | + | + | + | + | - | + | - |
| 107 | + | + | + | + | + | + | + | + | συμμετοχή | - | + | + | - | - | - | - | - | + | + | + | + | + | + | + | - | + | - |
| 108 | + | + | + | + | + | + | + | + | συρρίκνωση | - | + | + | - | - | - | - | - | + | + | + | + | + | + | + | - | + | - |
| 109 | + | - | - | + | + | + | + | + | τιμαριθμοποίηση | - | + | - | - | + | + | + | - | - | - | - | + | - | - | + | - | + | - |
| 110 | + | - | - | + | + | + | + | + | τιτλοποίηση | - | + | - | - | + | + | + | - | - | - | - | + | - | - | + | - | + | - |
| 111 | + | - | - | + | + | + | + | + | υπερκάλυψη | - | + | - | - | + | + | + | + | - | - | - | + | + | + | + | - | + | - |
| 112 | + | - | - | + | + | + | + | + | υπερυποτίμηση | - | + | + | - | + | + | + | + | - | - | - | + | + | + | + | - | + | - |
| 113 | + | - | - | + | + | + | + | + | υποτίμηση | - | + | + | - | + | + | + | + | - | - | - | + | + | + | + | - | + | - |
| 114 | + | + | + | + | + | + | + | + | υποχώρηση | - | + | + | + | - | + | - | - | + | + | + | + | + | + | + | + | + | - |
| 115 | + | - | - | + | + | + | + | + | χειραγώγηση | - | + | + | - | + | + | - | + | - | - | - | + | + | + | + | - | + | - |
| 116 | + | - | - | + | + | + | + | + | χρηματοδότηση | - | + | + | - | + | + | - | + | - | - | - | + | + | + | + | - | + | - |

*Figure 1. Extract of the Lexicon-Grammar table for $N_{pred}$ + Prep + $N_1$ (VSNAN) in Modern Greek*

As we can observe, in the column I, titled '$N_{pred}$', we inventoried predicative nouns registered in the $N_{pred}$ + *Prep* + $N_1$ table. Columns A to C describe the semantic features that concern the subject ($N_0$). For instance, the property '$N_0$ =: $N_{hum}$' refers to the semantic feature 'human' characterising the subject. Columns D to H refer to the presence and the structure of elements determining the subject (Det). Columns S to AA describe support verb constructions in which predicative nouns can be found.

In the intersection of a property (column) and an entry of predicative noun (line), the symbol '+' means that this property concerns this predicative noun, whereas the symbol '–' means that the predicative noun is not characterised by this property. Let

be the two first cells (first line) of the Lexicon-Grammar table in figure one. According to the description provided, the predicative noun ‘ονομαστικοποίηση’ (share registration) accepts a ‘human subject’ ($N_0$ =: $N_{hum}$) and not a ‘non-human subject’ ($N_0$ =: $N\text{-}_{hum}$). For example:

(4) [Οι μέτοχοι] [$N_0$ =: $N_{hum}$] έκαναν την ονομαστικοποίηση των μετοχών τους
[The shareholders] [$N_0$ =: $N_{hum}$] made the share registration
(Literal translation)
[The shareholders] [$N_0$ =: $N_{hum}$] registered the company's shares

(5) * [Το συμβόλαιο] [$N_0$ =: $N\text{-}_{hum}$] έκανε την ονομαστικοποίηση των μετοχών τους
[The contract] [$N_0$ =: $N\text{-}_{hum}$] registered the company's shares

In order to study the predicative nouns, we constructed 5 Lexicon-Grammar tables –in accordance with the theoretical and methodological approach we adopted– with 244 predicative nouns that combine with 311 verbs besides ‘κάνω’ (do/make) and ‘έχω’ (have). The tables include information about the subject ($N_0$) and object complements ($N_1$ and/or $N_2$) of the predicative nouns constructions, their alternative forms, their combination with numbers, their transformations, and their combination with verbs; also, examples from the corpus are included. More precisely, we created the following Lexicon-Grammar tables:

*(i) Table VSN: $N_0 + V_{sup} + N_{pred}$*

(5) Η μετοχή έκανε limit up
(The share hit limit up)

The table VSN contains 31 predicative nouns. The predicative nouns in this table are not combined with any complements ($N_1$ and/or $N_2$). In most cases, they are not combined with any determiner.

*(ii) Table VSNAN: $N_0 + V_{sup}\ N_{pred}$ σε (=at, to) $N_1$ or $N_0 + V_{sup}\ N_{pred}\ N_{1gen}$*

(6) Η τράπεζα έκανε μείωση στα επιτόκια
(The bank decreased the interest rates)

The *VSNAN* table consists of 115 predicative nouns. In this table, the object complements that follows the predicative nouns are introduced by a preposition (Prep $N_1$), or they appear in genitive case ($N_{1gen}$). The preposition is usually ‘σε’ (at), but also ‘προς’ (to) and ‘για’ (for). It has to be mentioned that $N_1$ is usually an abstract

noun in genitive case. As for prepositional object complements (Prep $N_1$), they are usually combined with numbers or percentages (e.g. η τράπεζα έκανε μείωση στα επιτόκια της τάξης του 1%, the bank decreased the interest rates by 1%). The complement is never a subordinate clause.

*(iii) Table VSNDNAN:* $N_0 + V_{sup}\ N_{pred}\ N_{1gen}$ *σε (to)* $N_2$

(7) Το χρηματιστήριο προχώρησε στη μετατροπή της εταιρίας σε αμοιβαίο κεφάλαιο
(The stock market went on to transform the company into a mutual fun)

69 predicative nouns have been registered in the *VSNDNAN* table. In this table, besides $N_1$, there is also a second complement ($N_2$), introduced with the preposition 'σε' (to). $N_1$ is mostly in genitive case and semantically it is something measurable ($N_{valeur}$), while the semantic characteristic of $N_2$ is that of 'non-human' ($N_{-hum}$) or 'location' ($N_{lieu}$).

*(iv) Table VSNNDEN:* $N_0 + V_{sup}\ N_{pred}\ N_1$ *από (from)* $N_2$

(8) Ο Όμιλος Χ προέβη στη διαγραφή των μετοχών της από το χρηματιστήριο
(Group X had its shares delisted from the stock market)

11 predicative nouns are noted down in the *VSNNDEN* table. The complements of the predicative nouns in this table are also two, as in the previous table, but in this case the second complement is introduced with a different preposition, 'από' (from), while $N_1$ is in genitive case or introduced with another preposition. A new table was created, because the different preposition used in these structures leads to completely different transformations.

*(v) Table VSNPN:* $N_0 + V_{sup} + N_{pred}$ *με (with)* $N_1$ — $N_0 + N_1 + V_{sup} + N_{pred}$, *(symmetrical structures)*

(9) Η Εταιρεία προχώρησε σε συμφωνία με την Cogetech
(The Company conducted a deal with Cogetech)

18 predicative nouns are registered in the table *VSNPN*. The syntactic structures described in this table are defined as symmetrical (Pantazara 2003), because the kind of action described is impossible without the cooperation of a second agent (Giry-Schneider 1987). The symmetric structure expresses a binary relationship, but "its two arguments have necessarily identical participation in any event described by the predicate" (Dimitriadis 2008: 329).

### *2.3 Parameterised finite state automata*

In order to exploit all information registered in a Lexicon-Grammar table, we created a parameterised finite state automaton (Paumier 2003) for each table. By the term finite state automata, or graphs, we mean electronic grammars that are used to represent linguistic phenomena (Gross 1993; Sklavounou 1995; Roche & Schabès 1997). By the term parameterised finite state automata we mean meta-graphs that allow us to convert tables into graphs. That is, parameterised graphs generate automatically a group of graphs, according to the information contained in Lexicon-Grammar tables.

More precisely, parameterised graphs are constructed manually and describe, by means of paths, all possible constructions of one table. Each column of the table (that is, each distributional and transformational property) is referred to as variable (parameter). For each line of the table (predicate), the parameterised graph creates one graph, the paths of which depend on the presence of the symbol '+', or the symbol '–' of each property. The symbol '+' allows the creation of the correspondent path, whereas the symbol '–' eliminates the path.

The following image corresponds to a parameterised automaton created for the table VSNAN.

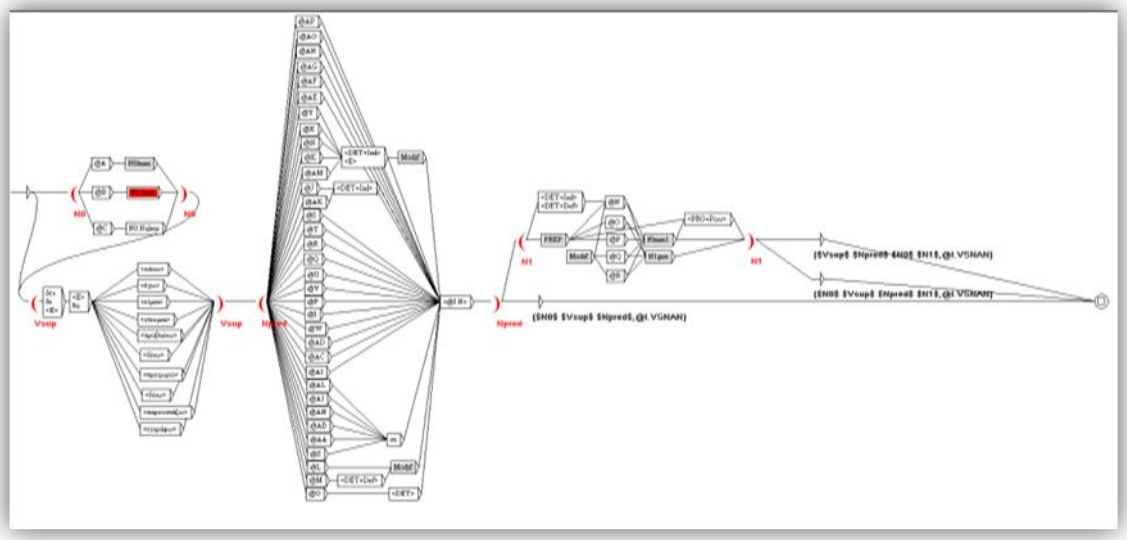


*Figure 2. Extract of the parameterised automaton for the table VSNAN in Modern Greek*

### *2.4 Finite state automata*

As described above, a parameterised finite state automaton allows the application of all syntactic and semantic information contained in the Lexicon-Grammar tables to corpora. According to the adopted methodological framework (cf. 1. Introduction), this information concerns all transformational and distributional properties of the registered predicates; that is, in our case predicative nouns. For instance, by applying a parameterised automaton, we recognise the following structures:

(10) Οι επενδυτές προχώρησαν χθες σε αναδιάρθρωση των χαρτοφυλακίων τους
The investors proceeded yesterday to the portfolios' restructuring

(11) Έγινε χθες αναδιάρθρωση των χαρτοφυλακίων από τους επενδυτές
The portfolios' restructuring was completed by the investors yesterday

Nevertheless, a parameterised automaton fails to recognise a great number of sentences, due to the complex structure of their arguments (Voskaki 2011). For example the argument 'του χρέους της εταιρείας' (the company's debt) is composed of two nouns in genitive case:

(12) (Έγινε) αναδιάρθωση του χρέους της εταιρείας
Restructuring of the company's debt (was done)

Hence, we enriched this parameterised automaton with a set of electronic grammars (finite state automata) that recognise complex structures of noun phrases (Ioannidou 2013).

The difference between the parameterised finite state automaton and the set of smaller finite state automata is that the former is performing syntactic analysis, whereas the latter performs chunking[4]. In other words, the parameterised finite state automaton focuses on the description of the structural rules that concern the composition of phrases (in our case predicative nouns and their arguments), whereas the set of finite state automata is applied locally and recognises complex structures of phrases, regardless of their syntactic position.

The set of finite state automata consists of 122 electronic local grammars, every one of which corresponds to a different structure. The great number of grammars

[4] By *chunking* or 'shallow/light parsing' we mean the segmentation of a sentence into groups of words (noun phrases, verb phrases etc.), without recognising their syntactic role in the main sentence (Abney 1991; Ramshaw & Marcus 1995; Stamatatos, Fakotakis & Kokkinakis 2000; Bai et al. 2006; Blanc, Constant & Watrin 2007a, b; Antoine, Mokrane & Friburger 2008; Mokrane, Antoine & Friburger 2008.

recognising noun phrases can be explained by the existence of several morphological features characterising a noun phrase (in terms of case, gender and number). These features concern most of the constituents of a noun phrase (determiner, adjective, noun etc.) and, as a result, during automatic processing, the words of the noun phrase are tagged with all different combinations of features. In Natural Language Processing, when at least two morphological features of the same kind (e.g. nominative and accusative case) are attributed to the same word, then this word is called morphologically ambiguous (Courtois 1996; Gross 2001; Kyriacopoulou 2005). An example follows of a predicative noun and its object, with all morphological ambiguities:

(13) Εισαγωγή [N:Nfs/N:Afs] πολλών [A:Gmp/A:Gfp/A:Gnp] τίτλων [N:Gmp][5]
Listing of many shares

In order to overcome this difficulty, we decided to describe in detail all different combinations of features of the constituents of noun phrases. Hence, we assured agreement in gender, case and number, eliminating as many morphological ambiguities as possible, based on the context, as proposed by Koskenniemi (1990), Roche (1992) and Laporte (2001). In the following image, we can see a finite state automaton describing a masculine noun phrase in accusative case of singular. This combination of morphological features is indicated by the letters 'Ams' ('A' for accusative, 'm' for masculine and 's' for singular) in every node.

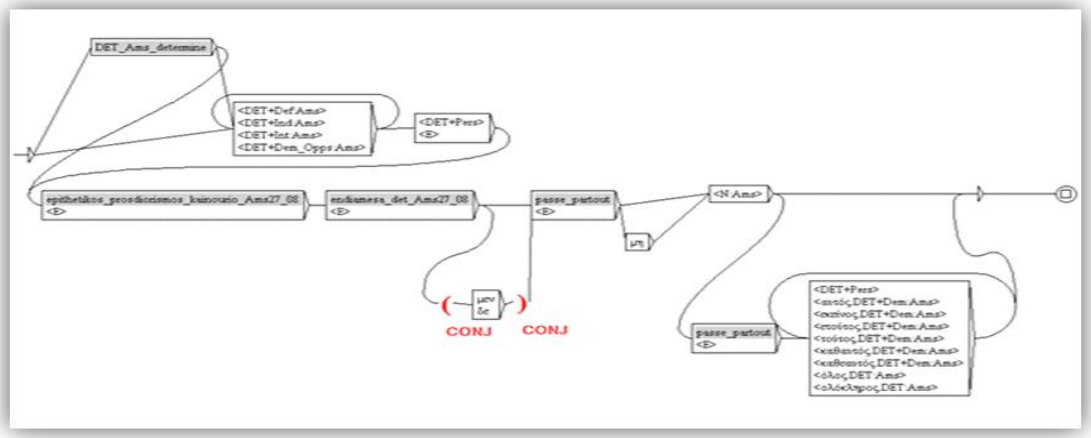



*Figure 3. Finite state automaton recognising masculine noun phrases in accusative case*

[5] 'N' stands for 'noun'; 'Nfs' for 'Nominative, feminine singular'; 'Afs' for 'Accusative, feminine, singular; 'A' for 'adjective'; 'Gmp' for 'Genitive, masculine, plural'; 'Gfp' for 'Genitive, masculine, plural'; 'Gnp' for 'Genitive, neuter, plural'.

By applying it on the corpus, we obtain the following result, where some of the morphological ambiguities presented in the example (13) are eliminated. More precisely, 'πολλών' is correctly recognised only as 'masculine':

(14) Εισαγωγή [N:Nfs/N:Afs] πολλών [A:Gmp] τίτλων [N:Gmp]
Listing of many shares

Moreover, apart from detailed information as morphology is concerned, we also described complex structures of noun phrases. More precisely, we recognised maximal-length noun phrases, that is, noun phrases modified by other smaller noun phrases, called 'base noun phrases' (Ramshaw & Marcus 1995; Tjong Kim Sang 2000; Bai et al. 2006). The following maximal-length noun phrase consists of four 'base noun phrases':

(15) [[Οι μετοχές] [της εταιρείας] [κατασκευών] [του ομίλου X]]
[[The shares] [of the construction] [company] [part of the group X]]

The automatic recognition of maximal-length noun phrases has been achieved by implementing a part of the finite state automata created by Ioannidou (2013), as shown in the following chapter.

## 3. Implementation results

In this chapter we present the results obtained after the application of the created finite state automata. At first, we will show an extract of the concordances retrieved after the application of the parameterised finite state automaton (as described in chapter 2.3). We will then compare them with the concordances obtained after the application of the enriched parameterised finite state automaton containing the set of electronic grammars of noun phrases (described in chapter 2.4).

As mentioned above, parameterised graphs focus on the syntactic bonds between predicates and arguments. More precisely, in our case, they aim at recognising predicative nouns and their arguments, according to their distributional and transformational properties. In the following image, we can observe an extract of the concordances obtained after the application of the parameterised finite state automata corresponding to the constructed tables.

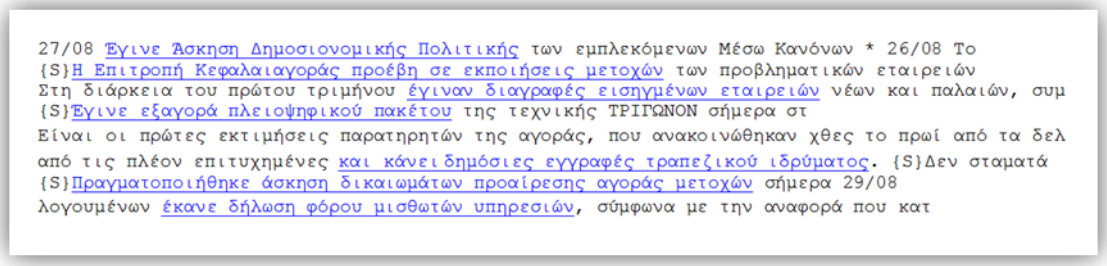



*Figure 4. Extract of concordances after the application of the parameterised finite state automata*

For instance, in the second row of the concordances we can observe the following example:

(16) Η Επιτροπή Κεφαλαιαγοράς προέβη σε εκποιήσεις μετοχών
The Capital Market Commission proceeded to divestiture of shares

The full structure of the $N_{pred}$ was recognised, along with the subject (Επιτροπή Κεφαλαιαγοράς, Capital Market Commission), the support verb (προέβη σε, proceeded), the predicative noun (εκποίηση, divestiture) and the object in genitive (μετοχών, of shares). Nevertheless, certain complements of the main arguments were not recognised.

During corpus automated processing, the presence of complex arguments hinders syntactic analysis. Therefore, they are not recognised as an entire noun phrase; that is, as an entire argument.

By enriching the same parameterised automaton with electronic grammars describing noun phrases (cf. chapter 2.4), we obtain the following concordances:

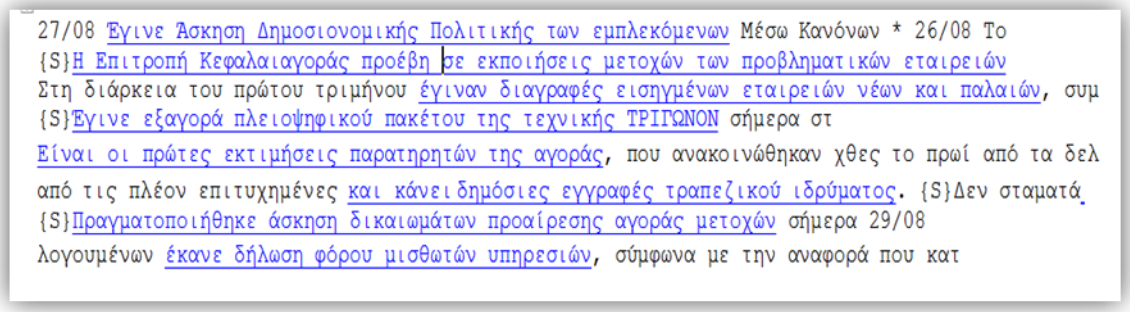



*Figure 5. Extract of concordances after the application of the enriched parameterised finite state automata*

As it concerns the example in the second row (referred in example (16)), a more complex syntactic structure is now recognised. It includes further arguments such as the noun phrase 'των προβληματικών εταιρειών' (companies in difficulties), complement of the noun 'μετοχών' (of shares).

## 4. Conclusions

In the framework of the current research, we have created 5 Lexicon-Grammar tables, 5 parameterised finite state automata, enriched by a set of 122 finite state automata describing complex noun phrases (Ioannidou 2013). Those data were applied on a Modern Greek financial corpus consisting of almost 19,000,000 running words (Tziafa 2012), in order to study the structure of predicative nouns. The Modern Greek financial corpus was divided in four sub-corpora, according to their register, so as to study the frequency of use of predicative nouns in each sub-corpus.

The frequency results showed that predicates as part of noun phrases prevail in press releases derived from the website of the Athens Stock Market, followed by academic texts, journalistic texts, and computer mediated communication. Frequency may not be 'automatically pedagogically useful' (Kaltenböck & Mehlmauer-Larcher 2005), as there is a set of decisions relating to teaching that must also be taken into account. However, it can at least help syllabus designers, materials writers and teachers alike to make better-informed and more carefully motivated decisions (Gavioli & Aston 2001). Therefore, the results of this work could be used for teaching Modern Greek as a second/foreign language in special interest groups, such as students interested in financial terminology.

In order to assess the quality of the created linguistic resources, we intend to use two basic evaluation measures: precision and recall. That calculation has to be made manually, by tagging part of the corpus in terms of syntax. That process, will also contribute to the enrichment of the existing Lexicon-Grammar tables.

## References

Abney, S. 1991. Parsing by Chunks. In R. Berwick, S. Abney and C. Tenny (eds), *Principle-based parsing*. Dordrecht: Kluwer, 257-278.

Antoine, J.-Y., A. Mokrane & N. Friburger. 2008. Automatic rich annotation of large corpus of conversational transcribed speech: The chunking task of the EPAC project. *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*, 678-683. Retrieved on 15 July 2016 from: http://www.lrec-conf.org/proceedings/lrec2008/

Bai, X.-M., L. Jin-Ji, K. Dong-Il & L. Jong-Hyeok. 2006. Identification of maximal-length noun phrases based on expanded chunks and classified punctuations in Chinese. *Computer Processing of Oriental Languages. Beyond the Orient: The Research Challenges Ahead Lecture Notes in Computer Science* vol. 4285: 268-276.

Biber, D. & S. Conrad. 2009. *Register, genre, and style.* Cambridge: Cambridge University Press.

Blanc, O., M. Constant & P. Watrin. 2007a. A finite-state super-chunker. *Implementation and Application of Automata Lecture Notes in Computer Science* vol. 4783: 306-308.

Blanc, O., M. Constant & P. Watrin. 2007b. Segmentation en super-chunks with finite-state approach. *Proceedings of TALN 2007*: 33-42. Retrieved on 15 July 2016 from: https://www.researchgate.net/publication/268338256_Segmentation_in_Super-Chunks_with_a_Finite-State_Approach

Chomsky, N. 1981. *Lectures on government and binding: the Pisa lectures*. Amsterdam: Foris.

Courtois, B. 1996. Formes ambiguës de la langue française. *Lingvisticæ Investigationes* 20(1): 167-202.

Daladier, A. 1978. *Quelques problèmes d'analyse d'un type de nominalisation et de certains groupes nominaux français,* PhD dissertation, University of Paris 7.

Dimitriadis, A. 2008. The event structure of irreducibly symmetric reciprocals. In J. Dölling & T. Heyde-Zybatow (eds), *Event structures in linguistic form and interpretation*. Berlin: Mouton de Gruyter, 327-354.

Fista, E. & T. Kyriacopoulou. 2009. Prédicats non verbaux préfixés en grec moderne : Le cas des préfixes ξε- et παρα-. *Proceedings of the International Conference « Supports et prédicats non verbaux dans les langues du monde »* : 34-39.

Fotopoulou, A. 1989. Étude comparative des extensions aspectuelles des verbes supports avoir, être Prép et faire en français et en grec moderne. *Séminaire de Linguistique Grecque, Mémoires du CERIL 4*.

Gavioli, L. & G. Aston. 2001. Enriching reality: Language corpora in language pedagogy. *ELT Journal* 55(3): 238-246.

Giry-Schneider, J. 1978. *Les Nominalisations en français: L'opérateur « faire » dans le lexique*. Genève: Droz.

Giry-Schneider, J. 1987. *Les prédicats nominaux en français.* Genève: Droz.

Gross, M. 1975. *Méthodes en syntaxe. Régime des constructions complétives.* Paris: Hermann.

Gross, M. 1981. Les bases empiriques de la notion de prédicat sémantique. *Langages* 63: 7-52.

Gross, M. 1993. Local grammars and their representation by finite automata. In M. Hoey (ed.), *Data, Description, Discourse: Papers on the English Language in honor of John McH Sinclair*. London: Harper-Collins, 26-38.

Gross, M. 2001. Les ambiguïtés. *Lingvisticæ Investigationes* 24(1): 3-42.

Ioannidou, K. 2013. *Ονοματικά σύνολα της Νέας Ελληνικής: Αυτόματη αναγνώριση και εξάλειψη μορφολογικών αμφισημιών κατά την αυτόματη επεξεργασία κειμένων. Προτάσεις εφαρμογής στη μετάφραση.* PhD dissertation, Aristotle University of Thessaloniki. Retrieved on 15 July 2016 from: http://www.didaktorika.gr/eadd/handle/10442/29579

Jespersen, O. 1965. *The Philosophy of grammar.* New York: Norton.

Kaltenböck, G. & B. Mehlmauer-Larcher. 2005. Computer corpora and the language classroom: On the potential and limitations of computer corpora in language teaching. *ReCALL* 17 (1): 65-84.

Koskenniemi, K. 1990. Finite-state parsing and disambiguation. *COLING* 20: 229-232. Retrieved on 15 July 2016 from: http://dl.acm.org/citation.cfm?id=997979

Kyriacopoulou, T. 2005. *L'analyse automatique des textes écrits : Le cas du grec moderne.* Thessaloniki: University Studio Press.

Kyriacopoulou, T. & V. Sfetsiou. 2002. Les constructions nominales à verbe support en grec moderne. *Linguistic Insights* 5: 163-181.

Lamprou, E. 1997. *Les verbes supports en français et en grec.* MS dissertation, University of Paris 13.

Laporte, E. 2001. Reduction of lexical ambiguity. *Lingvisticæ Investigationes* 25(1): 67-103.

Mokrane, A., J.-Y. Antoine & N. Friburger. 2008. Cascades de transducteurs pour le chunking de la parole conversationnelle : l'utilisation de la plateforme CasSys dans le projet EPAC. *Procceedings of the 15ème Conférence sur le Traitement Automatique du Langage naturel (TALN 2008).* Retrieved on 15 July 2016 from: http://www.info.univ-tours.fr/~antoine/articles/2008_TALN.pdf

Moustaki, A. 1995. *Les expressions figées είμαι /être Prép C W en grec moderne*. PhD dissertation, University of Paris 8.

Pantazara, A.-V. 2003. *Syntaxe dérivationnelle du grec moderne: Les constructions verbales à un complément prépositionnel et les constructions nominales et adjectivales prédicatives associées*. PhD dissertation, University of Paris 8, Saint-Denis.

Paumier, S. 2003. *De la reconnaissance de formes linguistiques à l'analyse syntaxique*. PhD dissertation, University of Marne-la-Vallée.

Ramshaw, L.A. & M.P. Marcus. 1995. Text chunking using transformation-based learning. *Proceedings of the Third Workshop on Very Large Corpora (ACL)*, 82-94. Retrieved on 15 July 2016 from: http://www.aclweb.org/anthology/W95-0107

Roche, E. 1992. Text disambiguation by finite state automata, an algorithm and experiments on corpora. *Proceedings of the 15th International Conference on Computational Linguistics (COLING '92)*, 993-997. Retrieved on 15 July 2016 from: http://dl.acm.org/citation.cfm?id=993114

Roche, E. & Y. Schabès. 1997. *Finite-state language processing.* Cambridge, MA/London: MIT Press.

Sfetsiou, V. 2007. *Κατηγορηματικά ονόματα: Μέθοδος ανάλυσής τους για ηλεκτρονικές εφαρμογές.* PhD dissertation, Aristotle University of Thessaloniki.

Silberztein, M. 1993. *Dictionnaires électroniques et analyse automatique de textes, Le système INTEX*. Paris, Masson.

Sklavounou, E. 1994. Support nouns: Application to the special lexicon of tennis. In I. Philipakki-Warburton, K. Nicolaidis & M. Sifianou (eds), *Themes in Greek Linguistics. Papers from the 1st International Conference on Greek Linguistics. Current Issues in Linguistic Theories* 117, *Reading*. Amsterdam/Philadelphie: Benjamins, 515-520.

Sklavounou, E. 1995. Ένα παράδειγμα λεξιλογικής αναπαράστασης και επεξεργασίας με πεπερασμένα αυτόματα. *Studies in Greek Linguistics, Proceedings of the 15th Annual Meeting of the Department of Linguistics*, 810-819.

Stamatatos, E., N. Fakotakis & G. Kokkinakis. 2000. A practical chunker for unrestricted text. *Proceedings of the NLP 2000 Conference*. Heidelberg: Springer-Verlag, 139-150.

Tjong Kim Sang E.F. 2000. Noun phrase recognition by system combination. *Proceedings of the 1st North American chapter of the Association for Computational Linguistics Conference*. Stroudsburg, PA: Association for Computational Linguistics, 50-55. Retrieved on 15 July 2016 from: http://aclweb.org/anthology//A/A00/A00-2007.pdf

Tziafa, E. 2012. *Μελέτη της γλώσσας του χρηματιστηρίου με βάση σώματα κειμένων και στόχο την αυτόματη μετάφραση.* PhD dissertation, Aristotle University of Thessaloniki. Retrieved on 15 July 2016 from: http://invenio.lib.auth.gr/record/130913/

Voskaki, O. 2011. *Le lexique-grammaire des verbes du grec moderne. Constructions transitives non locatives à un complément d'objet direct.* PhD dissertation. University of Paris-Est. Retrieved on 15 July 2016 from: http://tel.archives-ouvertes.fr/tel-00776448/