

## Automatic Processing of technical texts in Modern Greek: Term Annotation

Tita Kyriacopoulou  
Aristotle University of Thessaloniki

Claude Martineau  
Université Paris-Est Marne-la-Vallée

Olympia Tsaknaki  
Aristotle University of Thessaloniki

Eleni Tziafa  
Aristotle University of Thessaloniki

Rania Voskaki  
Université Paris-Est Marne-la-Vallée

### 1 Introduction

The paper focuses on the achievement of the targets as follows: automatic term extraction stem from technical corpora (stock exchange, telecommunications), term annotation derived from stock exchange corpora and text alignment. We also calculate the success rate of an electronic dictionary of stock exchange terms and of a FSA which permits the automatic extraction of acronyms and initialisms in corpora. The above-mentioned matters will be successively presented.

The current research has evolved around the Modern Greek language processing and it is integrated into a program for the development and the gradual completion of the Unitex Greek version. Unitex<sup>1</sup> is a collection of programs developed for the analysis of texts in natural languages by using linguistic resources and tools. This work pertains to the field of NLP and is a product of collaboration between the Laboratoire d'Informatique Gaspard-Monge (Université Paris-Est Marne-la-Vallée) and the Laboratory of Translation and Language Processing, part of the School of French Language and Literature (Aristotle University of Thessaloniki).

### 2 Term annotation

The annotation procedure can identify and delimit words or sequences of words which constitute terms. The terms can be:

- a) single words (πληθωρισμός / inflation, μετοχή / stock),
- b) compound words (ελεύθερο εμπόριο / free trade),
- c) compound words composed of simple words or initialisms and compound words that are already terms (ζώνη ελεύθερου εμπορίου / free trade area, οικονομικές καταστάσεις βάσει ΔΠΧΠ<sup>2</sup> / financial statements under IFRS<sup>3</sup>);

---

<sup>1</sup> <http://www-igm.univ-mlv.fr/~unitex/>

<sup>2</sup> Διεθνή Πρότυπα Χρηματοοικονομικής Πληροφόρησης

<sup>3</sup> International Financial Reporting Standards

components can sometimes be two compound words (εντολή πώλησης με καθορισμένη ελάχιστη τιμή / sell limit order).

We use a linguistic approach and we surround the terms with tags: <TAG> (opening) <\TAG> (closing).

In the environment of the automatic analysis system Unitex, the recognition of word sequences can be achieved by local grammars (they are represented in FSA<sup>4</sup> and/or FST<sup>5</sup> graphs comprising eventually one or more sub-graph levels) and/or electronic dictionaries (lists of inflected forms provided by grammatical, inflectional and syntactico-semantic information). The strict approach by local grammar is more appropriate in the recognition of highly productive sequences based on a limited vocabulary and characterized by a significant distributional and combinatorial aspect, e.g. the dates.

We decided to list and represent the terms of areas of interest using a significant and varied vocabulary in the form of a dictionary. The specialized dictionary we used, created by Eleni Tziafa (2007), turned out to be an essential linguistic tool. It includes 9,634 terminological lexical units, single words (e.g. μετοχοδάνειο / share investment loan) and compound words (e.g. ομόλογουλούδι / flower bond). Sequences of letters not bounded by a dash, an apostrophe or a space are considered as single words (Anastassiadis-Symeonidis 1986, Kyriacopoulou *et al.* 2002, Silberztein 1990). Untranslated Greek, English and French terms are contained in texts concerning the stock exchange market. These terms are written either in Greek or in Latin characters (e.g. σουάπ / swap). For that reason, both forms were included in the dictionary.

The dictionary's typology consists of the following:

- Adjectives: ενδοσυνεδριακός / intraday, οφ-σορ / off-shore
- Nouns: απούλοποίηση / dematerialization, αγορά ευρωμετοχών / eurostock market
- Adverbs: εξωσυνεδριακά / pre-market, κάτω από τον πάγκο / under the counter
- Symbols: T, AAA, AA-, A+, BB+
- Verbs: απούλοποιώ / dematerialize
- Acronyms<sup>6</sup>: ΕΛ.ΠΙΣ. (Ελληνικά Πιστοποιητικά / Hellenic Certificates)
- Initialisms: Ο.Α.Σ.Η.Σ. (Ολοκληρωμένο Αυτόματο Σύστημα Ηλεκτρονικών Συναλλαγών) / Integrated Automated Electronic Exchange System

Information has also been provided, concerning the semantic characteristics of entries. In the example which is given here

απούλοποίη"ση.N265,N+Abst+[Action]+[Eco]

---

<sup>4</sup> FSA: Finite State Automaton.

<sup>5</sup> FST: Finite State Transducer.

<sup>6</sup> As acronym, we define an abbreviation consisting of the first syllabic fragment of the constituent words. As initialism, we consider a type of abbreviation that normally consists of the initial letters of the constituent words (Anastassiadis-Symeonidis 1986).

the entry *αποϋλοποίηση* (dematerialization) is qualified as an abstract noun (Abst), it expresses action ( [Action] ) and it belongs to the economic domain terms ( [Eco] ). Similarly, as illustrated in the example which follows, the entry *ευρωγέν* (euroyen) is a concrete noun (Conc), it constitutes a monetary unit ( [Monn] ) and it also forms part of the economic field ( [Eco] ):

ευρωγέν.N305,N+Conc+[Monn]+[Eco]

To describe semantically the aforementioned terms, we consulted the same resources as for their collection and their morphological description, which is the following:

- specialized dictionaries. It should be noted that we have already lemmatized terms residing in these dictionaries;
- the terminology database IATE of the European Union;
- several termbases, available on Internet;
- finance and economic magazines and newspapers.

Regarding the local grammar, the only FST we used (Figure 1), allowed after the application of our dictionary, the identification of the terms recognized in the surrounding tag. The meta-symbol <DIC> of this grammar, matches with any word present in our dictionary in the processed text.

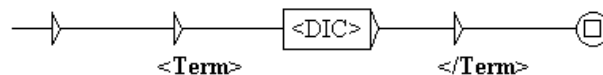


Figure 1: FST for term annotation

The Annual Report and Review of the 6th Fiscal Year 2006 of the company *Hellenic Exchanges S.A.*<sup>7</sup>, written in Greek, was the corpus we used for the term annotation, in order to extract and identify the initial performance. The overall size of the corpus was around 714 KB (51,185 words, 158 pages). To evaluate the FST (see chapter 3) we did not apply it on the whole corpus. Due to its length, we decided to apply it on a sample corpus of 231 KB (16,820 words, 60 pages).

The stock exchange vocabulary is specific; it is used by particular professional groups and covers a whole range of their issues of interest and activities. The 1993-2000 period was determinant for the formulation of the specific Greek stock exchange vocabulary. In these seven years, there was experienced a remarkable increase in the operations and the interest shown by population. This growth created new needs in the market and pointed out the imperative necessity of using new words, which entered the domain of stock exchange as neologisms or neonyms. During this period, new terms were introduced constantly and continually into the stock exchange vocabulary. Nevertheless, after the market decline in 2000 until today, due to the decrease of interest for this specific

<sup>7</sup> <http://www.ase.gr>

domain, this vocabulary became more restricted. Therefore, its extensive description is permitted to a certain degree.

As pointed out by Bowker and Pearson (2002), there is no standard size for a corpus to be representative. Corpora of millions or hundreds of millions of words can be proved useful for the research of special languages. The electronic corpora of special use were authentic, i.e. without having induced any modifications. They include texts originated from a specific thematic domain and the high frequency occurrence of terms related to it. Hence, stock exchange and telecommunications terminology is expected to occur to a great extent in these corpora.

The first step was to preprocess the corpus, which was incorporated in Unitex. The corpus preprocess included the undermentioned operations: normalization of separators, identification of tokens, normalization of non-ambiguous forms, splitting into sentences and application of dictionaries. We then applied the FST.

The corpus processing systems, search in texts concrete morphosyntactic structures, which are treated and presented under the forms of concordances. It is worth mentioning, to be precise, that concordancers generate a list, including all the words of a given text, chosen on the basis of the dictionary, retrieved alphabetically, and presenting each word within the bilateral context. Concordances are of great utility, because they allow the possibility to study the use of a word in question in texts. Moreover, they provide information about the sort of context, narrow or broad, where the word occurs.

Economic terms are listed in this extract of concordances:

καινίασε μια νέα πολιτική προβολής της <Term>ελληνικής αγοράς</Term>. {S} Τον Ιούνιο, το ΧΑ διο  
 θεμελιώδη μακροοικονομικά στοιχεία της <Term>ελληνικής οικονομίας</Term>, καθώς επίσης και από  
 one) παρουσιάσεις σε περίπου 80 διεθνή <Term>επενδυτικά κεφάλαια</Term>. {S} Προβάλλοντας και α  
 χώνευση με απορρόφηση από την ΕΧΑΕ των <Term>θυγατρικών εταιρειών</Term> «ΚΕ-ΝΤΡΙΚΟ ΑΠΟΘΕΤΗΡΙΟ  
 μός Έκτης Εταιρικής Χρήσης 2006 012 Τα <Term>καθαρά κέρδη</Term> του Ομίλου το 2006, μετά την  
 ιέβλεπον στη θωράκιση του θεσμι-κού και <Term>κανονιστικού πλαισίου</Term> που ρυθμίζει τη λειτ  
 ν 30.10.2006 ξεκίνησε η λειτουργία της <Term>κοινής πλατφόρμας</Term> υποστήριξης της λειτουργ  
 μένων των εσόδων από τη λειτουργία της <Term>κοινής πλατφόρμας</Term> ΧΑ-ΧΑΚ), κα-θώς και από  
 ορούς Ενεργητικό 45,117 10,241 341% Μη <Term>κυκλοφορούν Ενεργητικό</Term> 43,618 45,046 -3% Σ  
 ι) Διαθέσιμα 120,103 179,674 -33% Λοιπό <Term>κυκλοφορούν Ενεργητικό</Term> 45,117 10,241 341%  
 προσπάθεια της διοίκησης για μείωση του <Term>λειτουργικού κόστους</Term> συνεχί-ζεται και αποδ  
 υ θα βοηθήσει περαιτέρω στη μείωση του <Term>λειτουργικού κόστους</Term> του Ομίλου. {S}1. {S}  
 ι φορολογικού ελέγχου 37 2. {S}9 Εξέλιξη <Term>μετοχικού κεφαλαίου</Term> 37 2. {S}10 Μετοχική σύ  
 ρς ΕΧΑΕ να ανέλθει σε 70.271.463 και το <Term>μετοχικό κεφάλαιο</Term> να αυξηθεί κατά €71.750  
 λελών, έσοδα από συνδρομές και αυξήσεις <Term>μετοχικών κεφαλαίων</Term> εισηγμένων εταιριών, έ  
 άρωση» των κριτηρίων εισαγωγής για τις <Term>ναυτιλιακές εταιρίες</Term>. {S} Το 2006 ήταν διεθ  
 ικητικό Συμβούλιο της Εταιρείας. {S} Οι <Term>οικονομικές καταστάσεις</Term> του Ομίλου ΕΧΑΕ έχ  
 αλούμε όπως εγκρίνετε τις ενοποιημένες <Term>οικονομικές καταστάσεις</Term> της Εταιρεί-ας για  
 : ΕΤΕΣΕΠ από τη ΕΧΑΕ) είναι εμφανή στις <Term>οικονομικές καταστάσεις</Term> του 2006, με τον π  
 αναλυτικά στις ση-μειώσεις των ετησίων <Term>οικονομικών καταστάσεων</Term>, οι οποίες έχουν σ  
 λίξεις της αγοράς. {S} Τα απο-τελέσματα <Term>περιστολής των δαπανών</Term> είναι εμφανή στον ι  
 η κερδοφορία του Ομίλου. {S} Σημαντικές <Term>ππνές εσόδων</Term> για τον Όμιλο, εκτός από τις  
 αλλαγής για τον τελικό επενδυτή. {S} Στο <Term>πλαίσιο ενίσχυσης</Term> της ανταγωνιστικότητας τ  
 νομή μερίσματος €0,5 ανά μετοχή. {S} Το <Term>συνολικό ποσό</Term> που θα διανεμηθεί ανέρχεται

Table 1: This is an example of concordances of Greek economic terms.

### 3 Evaluation of dictionary effectiveness



The results we obtained after the dictionary application are presented below:

1,857 matches
2,925 recognized units
6.365% of the text is covered

In order to assess the quality of the dictionary, we used two basic evaluation measures: precision and recall. In the context of the present study, precision can be defined as the ratio of relevant terms over the total number of terms retrieved by the corpus processing system. Recall is the ratio of relevant terms retrieved by the corpus processing system over the total number of relevant terms in the corpus. For the precision percentage, it is essential to calculate the terms retrieved from the text, while for the recall percentage it is required to locate the exact number of correct terms included in the text. That calculation has to be made manually.

More precisely, out of the 1,857 matches retrieved from the 231 KB text, the relevant terms were 1,786, that is a percentage of 96.1%. As for the recall ratio, the corpus processing system retrieved 1,786 terms, while it should have returned 2,100 terms. Thus, it was covered at a percentage of 85.04%. Moreover, there was a percentage of 3.82% corresponding to words that should not have been retrieved. The number of irrelevant terms constitutes the noise. Noise is due to ambiguous words which may come either from the special vocabulary or from the general. By applying the specialized dictionary the system recognized words that were not only terms but also they were registered in the general dictionary. For instance the words *αποτέλεσμα*, *εκτέλεση*, *πράξη* are term entries occurring in both the specialized and general dictionaries.

The undermentioned examples are provided to illustrate the method of lemmatization in the specialized dictionary, the recognized term as well as the common word:

- *αποτέλεσμα*,.N+[Eco]:Nns:Ans:Vns / effect, result  
*αποτέλεσμα φορολογικού ελέγχου* / audit results  
*αποτέλεσμα με μια ματιά* / result in a glance
- *εκτέλεση*,.N+Abst+[Eco]:Nfs:Afs:Vfs / execution  
*εκτέλεση εντολής* / order execution  
*εκτέλεση των καθηκόντων* / duty execution
- *πράξεις*,*πράξη*.N+[Eco]:Nfp:Afp:Vfp / trading  
*πράξεις από τον δικαιούχο των αξιών* / shareholder trading  
*πράξεις ατιμωτικές* / dishonest acts

#### 4 Text Alignment

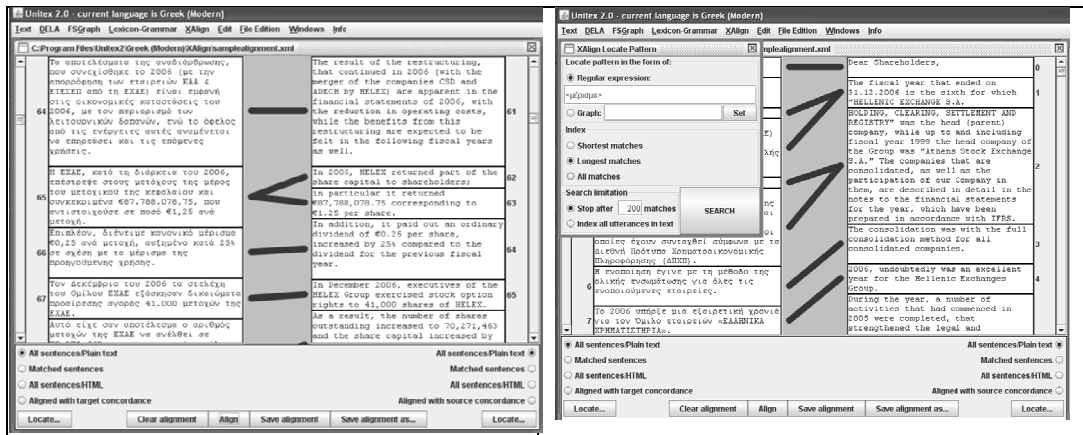
For the term extraction of bilingual texts we proceeded to the text alignment. This process is based on the principle that parallel texts, which usually are a source text and the target text(s) produced, can be aligned at a sentential level. More

concretely, it is possible to search a word or a phrase in both texts and find their translation in sentences aligned with those containing the word or phrase in question. However, it should be stressed that word alignment is still not possible. Patrick Watrin integrated into Unitex the Open Source text alignment tool XAlign, developed at the LORIA (Laboratoire Lorrain de Recherche en Informatique et ses Applications).

The corpora we had at our disposal are of two types:

- Annual Report and Review of the 6th Fiscal Year 2006 of Hellenic Exchanges S.A. The size of the Greek and the English corpora amount to a total of 714 KB and 672 KB respectively.
- Annual Report 2008 of the Hellenic Telecommunications and Post Commission (EET)<sup>8</sup>. The length of the Greek corpus corresponds to 708 KB, while the English corpus is totaling 673 KB.

In the next three images is shown an example of the alignment process on the stock exchange corpora. The first image displays the sentential alignment whereas the second image represents the search of the term *μέρισμα* (dividend). Finally, in the third image, we obtained only the sentences which contain *μέρισμα* linked to those including *dividend*.



<sup>8</sup> <http://www.eett.gr>

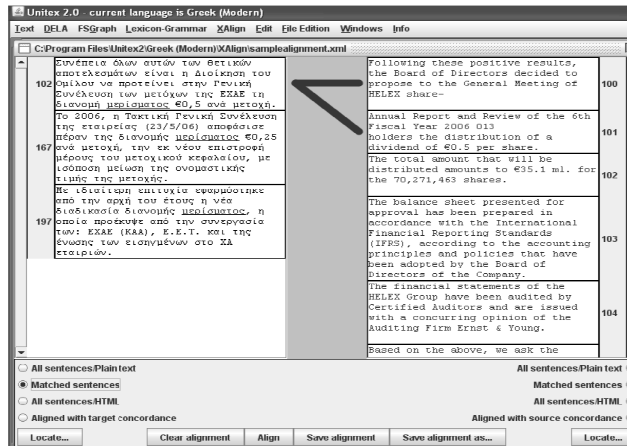


Table 2: These are examples of the alignment process on stock exchange corpora.

## 5 Term extraction using FSA

In the framework of automatic term extraction, we applied on the corpora the FSA shown below that allows the recognition of a noun followed by an acronym or initialism (Kyriacopoulou, Martineau and Tziafa: 2009).

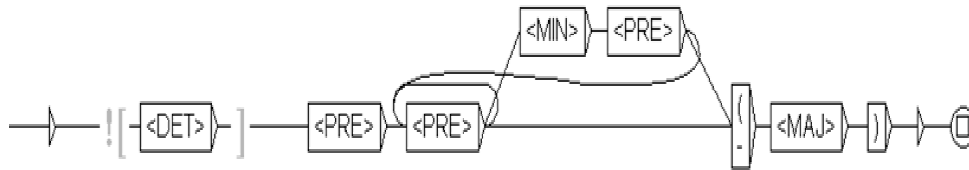


Figure 2: FSA for term extraction

We applied the FSA on the Greek corpus of stock exchange terms (714 KB: 51,185 words, 158 pages), the English corpus of stock exchange terms (672 KB: 53,152 words, 154 pages), the Greek telecommunications corpus (708 KB: 45,731 words, 100 pages) and, finally, the English telecommunications corpus (673 KB: 46,680 words, 98 pages). In the present article, we set forth the concordances we retrieved from the Greek corpora.

The occurrences took the following form of concordances:

Θάρτο Καθαρού Κέρδους 49,1% 36,7% 34% Απόδοση Ιδίων Κεφαλαίων (ROE) 37,6% 13,4% 180% Δείκτη 263.595 263. (S)595 διασύνδεση μέσω του Δικτύου Χρηματοπιστωτικών Συναλλαγών (ΔΧΣ) Συμβάσεις εσό ά πρότυπα και ειδικότερα τα πρότυπα του Ινστιτούτου Εσωτερικών Ελεγκτών (ΙΙΑ). (S)Κύρια καθήκον 41,590 102% Αποσβέσεις 2,128 2,951 -28% Λειτουργικό Αποτέλεσμα (EBIT) 81,732 38,639 112% Χρηματ 4 0725 0658 0677 0,413 0767 2,128 2,951 Λειτουργικό Αποτέλεσμα (EBIT) 20539 9,088 22,513 7,360 Λειτουργικές Δαπάνες 27,349 28,065 -3% Λειτουργικό Αποτέλεσμα προ Αποσβέσεων (EBITDA) 83,860 4 ,877 6,420 5,949 7738 7712 27349 28,065 Λειτουργικό Αποτέλεσμα προ Αποσβέσεων (EBITDA) 21,082 9

Table 3: This is an example of concordances taken from the corpus of Greek stock exchange terms.

η Γεωγραφικού Πληροφοριακού Συστήματος (Geographic Information System - GIS) στο δικτυακό τόπο ραφική θέση των κατόχων τους. (S)Βλ.(S) Geographic Information Systems (GIS). (S)Η υπηρεσία κατ

oup Η EETT, μέσω της συμμετοχής της στο Independent Regulators Group (IRG), συνέχισε τη συνεργασία με τις προβλεπόμενες διαδικασίες της International Telecommunication Union (ITU) και εναρμόσεων, το οποίο βασίζεται στην εφαρμογή «One Stop Notification (OSN)». (S) Το 2008, η EETT αναγνώδηγιών αναφορικά με Ταχυδρομικά θέματα (Postal Directive Committee - PDC), κατά τη διάρκεια της . (S)Ιδιωτικά Δίκτυα Κινητής Υπηρεσίας (Private Mobile Radio - PMR). (S) Ο όρος «Ειδικά Ραδιοδίκ

Table 4: This is an example of English term concordances taken from the corpus of Greek telecommunications terms.

ολικής Υπηρεσίας (ΦΠΚΥ), η πρόσβαση στα Δημόσια Ταχυδρομικά Δίκτυα (ΔΤΔ), η κοστολόγηση - τιμολογία υπηρεσιών ταχυμεταφοράς. (S) Η διάθεση του Δημόσιου Ταχυδρομικού Δικτύου (ΔΤΔ) σε εναλλακτικούς παρόχους, οι όροι και προϋποθέσεις διάθεσης του Δημόσιου Ταχυδρομικού Δικτύου (ΔΤΔ) σε εναλλακτικούς παρόχους θα χρησιμοποιηθούν από τη Διεθνή Ένωση Τηλεπικοινωνιών (ITU) για εμπλουτισμό των εργασιών του European Regulators Group - ERG, του Δικτύου Λογών Ανταγωνισμού (ECN), της Postal Directive που αρχίζει με «116» δεσμεύεται στα Εθνικά Σχέδια Αριθμοδότησης (ΕΣΑ) για εναρμονισμένους αριθμούς. (S)Ομάδα που απαρτίζουν οι 25 Εθνικές Ρυθμιστικές Αρχές (ΕΡΑ) της ΕΕ, καθώς και οι 8 Εθνικές Ρυθμιστικές Αρχές (ΕΡΑ), την Ευρωπαϊκή Επιτροπή, υποβλήθηκαν 3 αιτήματα από ευρωπαϊκές Εθνικές Ρυθμιστικές Αρχές (ΕΡΑ) προκειμένου να συντονιστούν οι εργασίες στο πλαίσιο διμερών επαφών με τις Εθνικές Ρυθμιστικές Αρχές (ΕΡΑ) άλλων κρατών. (S)Ειδική ομάδα εργασίας προς την Ευρωπαϊκή Επιτροπή και τις Εθνικές Ρυθμιστικές Αρχές (ΕΡΑ) το Σχέδιο Μέτρων αναφορικά με την αλλαγή, όπως αυτός ανακοινώνεται από την Εθνική Στατιστική Υπηρεσία της Ελλάδος (ΕΣΥΕ). (S) Η ΜΑΤ εγγραφή Μνημόνιο μεταξύ της EETT και της Εθνικής Ρυθμιστικής Αρχής της Ινδίας (TRAI), με στόχο τ

Table 5: This is an example of Greek term concordances taken from the corpus of Greek telecommunications terms.

The system failed to recognize some occurrences for several reasons (use of typographic symbols -punctuation marks, special purpose characters-, minuscule letters among majuscule letters, occurrence of the initialism before the complete form, etc.). E.g.:

- «ΧΡΗΜΑΤΙΣΤΗΡΙΟ ΑΞΙΩΝ ΑΘΗΝΩΝ Α.Ε.» (ΧΑ) / Athens Stock Exchange
- Διεθνή Πρότυπα Χρηματοοικονομικής Πληροφόρησης (ΔΠΧΠ-IFRS) / International Financial Reporting Standards
- Ανάπτυξη Συστημάτων και Υποστήριξης Κεφαλαιαγοράς Α.Ε. (Α.Σ.Υ.Κ. Α.Ε.) / Development of Capital Market Systems and Support S.A.
- Συνδέσμου Μελών των Χρηματιστηρίων Αθηνών (Σ.ΜΕ.Χ.Α.) / Members of Athens Stock Exchange Association
- Πολυμεσικών Μηνυμάτων (MMS) Προστιθέμενης αξίας (PMMS) / Premium Multimedia Messaging Service
- Νόμο Κοινές Υπουργικές Αποφάσεις (ΥΑ) / Common Minister's Decisions
- Voice over Internet Protocol (VoIP)
- Εθνικής Επιτροπής Τηλεπικοινωνιών & Ταχυδρομείων (ΕΕΤΤ) / Hellenic Communications & Post Commission
- Αστικά Κέντρα (Α/Κ) / Cities
- Χάρτη Υποχρεώσεων προς τον Καταναλωτή (ΧΥΚ) / Chart of Obligations to the Consumer

- γραμμή βοήθειας (HELP LINE)
- Σύντομων Γραπτών Μηνυμάτων προστιθέμενης αξίας (PSMS) / Premium Short Message Services
- SNG (Satellite News Gathering)

## 6 Evaluation of FSA effectiveness

To evaluate the FSA we did not apply it on the total Greek corpus of stock exchange terms, but on a representative sample of 231 KB (16,820 words, 60 pages). As regards the English corpus of stock exchange terms, we used 218 KB (17,216 words, 50 pages) out of 672 KB. Likewise, we made use of 396 KB (23,323 words, 50 pages) from the Greek telecommunications corpus, instead of its entire size and 383 KB (24,660 words, 50 pages) from the English telecommunications corpus instead of 673 KB. We set out the results taken out from the Greek corpora.

With respect to:

a) the corpus of stock exchange terms, the results we found are cited below:

14 matches
122 recognized units
0.265% of the text is covered

Evaluation measures

- Precision  $12/14 = 85.7\%$
- Recall  $12/19 = 63.1\%$

Noise = 14.2%

b) the corpus of telecommunications terms, the results we found are presented below:

71 matches
699 recognized units
0.847% of the text is covered

Evaluation measures

- Precision  $68/71 = 95.7\%$
- Recall  $68/107 = 63.55\%$

Noise = 4.22%

## 7 Conclusions and perspectives

In this paper, we have described the procedure of the automatic term extraction appearing in technical texts (stock exchange, telecommunications). By means of an electronic dictionary of stock exchange terms and a FST we have shown how to annotate terms in a stock exchange corpus. We have also focused on text alignment of parallel (Greek and English) technical corpora. A FSA has been

utilized to track acronyms and initialisms in these corpora. Finally, we proceeded to the evaluation of the electronic dictionary of stock exchange terms and of the FSA.

There are a number of potential areas for application of the topics discussed in this paper such as the automatic translation, the language teaching, the search engines, the spelling correctors as well as the syntactic correctors. The improvement of the performance of the stock exchange terms dictionary, the creation of a telecommunications terms electronic dictionary and a tool creation for the automatic term extraction, are among the variety of perspectives that we will endeavor to fulfill through future research. In addition, semantic term tagging would offer a subtle text analysis and a meticulous research by category.

### References

- Anastassiadis-Symeonidis, A. 1986. *Η Νεολογία στην Κοινή Νεοελληνική. Epistimoniki Epetirida Filosofikis Scholis*. Thessaloniki: Aristotle University of Thessaloniki.
- Bowker, L., and J. Pearson. 2002. *Working with specialized language: a practical guide to using corpora*. London: Routledge.
- Kyriacopoulou, T., C. Martineau, and E. Tziafa. 2009. Extraction Terminologique et Corpus Alignés Anglais Grec. *Arena Romanistica : The 28th Conference on Lexis and Grammar* 4. 214-223.
- Kyriacopoulou, T., S. Mrabti, and A. Yannacopoulou. 2002. Le dictionnaire électronique des noms composés en grec moderne. *Lingvisticae Investigationes* 25:1. 7-28.
- Silberztein, M. 1990. Le dictionnaire électronique des mots composés. *Langue Francaise* 87. 71-83.
- Tziafa, E. 2007. Κατασκευή ηλεκτρονικού λεξικού οικονομικής-χρηματιστηριακής ορολογίας – Γενικές παρατηρήσεις. In *Proceedings 6<sup>th</sup> Conference "Hellenic Language and Terminology"*. Athens. 289-298.