EURALEX XIX

Congress of the European Association for Lexicography

Lexicography for inclusion

7-11 September 2021

Ramada Plaza Thraki

Alexandroupolis, Greece

www.euralex2020.gr

Proceedings Book
Volume 1

Edited by Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

EURALEX Proceedings

ISSN 2521-7100

ISBN 978-618-85138-1-5

Edited by: Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

English Language Proofreading: Lydia Mitits and Spyridon Kiosses

Technical Editor: Kyriakos Zagliveris

2020 Edition

Frame Semantics in the Specialized Domain of Finance: Building a Termbase to Aid Translation

Pilitsidou V.[1], Giouli V.[2]

[1] *National and Kapodistrian University of Athens, Greece*
[2] *Institute for Language and Speech Processing, ATHENA Research Centre, Athens*

Abstract

Frame semantics (Fillmore 1977, 1982, 1985) is one of the most important developments for lexicography in the 20th century. The semantic frames approach to lexicon building and semantic representation of meaning at word and phrase level – or even beyond – has been the focus of research in computational linguistics and in Natural Language Processing. The present paper is aimed at describing completed work for the creation of a domain-specific frame-semantic lexicon in Greek (EL) and its alignment to the English (EN) FrameNet. Building on Fillmore's Frame Semantics (Fillmore 1977, 1982, 1985) and on the example set by the FrameNet project (Baker et al. 1998), we developed a bilingual EL-EN lexical resource in the financial domain based on corpus evidence. Our motivation was two-fold: (a) to better account for the semantics of the specialized lexicon – especially the verbs and predicative nouns of the financial domain, and (b) to make cross-lingual alignments at the word level in a way that is meaningful for the translation process.

Keywords: frame semantics; FrameNet; frame; financial domain; translation; terminology; terminological resource

1 Introduction

The development of FrameNet (Baker et al. 1998) and FrameNet-compatible language resources as both human- and machine-readable lexica is based on annotating examples of how words are used in actual texts. Frame Semantics has many possible applications, terminography and domain-specific translation being one of them. In this paper we describe completed work for the creation of a terminological resource for the specialised domain of finance. The resource will be based on the principles of Frame Semantics (Fillmore 1977, 1982, 1985) and the example of FrameNet (Baker et al. 1998). The paper is organised as follows. In section 2 we briefly present the theoretical framework and the previous work on related projects, especially regarding the Greek language. The research scope and goals are set out in section 3. Section 4 and its subsections outline the methodology adopted towards developing the bilingual resource, whereas section 5 provides a detailed description of the resource and its components. In section 6 we discuss different aspects of the procedures and the results of our work and, finally, in section 7 we provide our conclusions and prospects for future research.

2 Theoretical Framework and Related Work

The theory of Frame Semantics by Charles J. Fillmore (Fillmore 1977, 1982, 1985) focuses on the continuity that exists between language and experience (Petruck 1996). According to this theory, words gain their meaning in a semantic frame which can be an event or a relation. In this context, the term "semantic frame" or "frame" refers to any system of meanings which is connected in a way that to understand any one of these meanings we must be able to understand the whole structure to which it belongs; when one of the elements of such a structure is used in a text or a discussion, then all the other elements automatically become available (Fillmore 1982: 111). Fillmore calls these elements "Frame Elements" (FEs). The words that evoke the semantic frames are called "Lexical Units" of the frame (LUs) and they are predicates which are mainly verbs, other parts of speech (names, adjectives, adverbs) as well as multi-word expressions (Tantos et al. 2015: 167).

Frame Semantics is the theoretical framework on which FrameNet (Baker et al. 1998), a lexical resource for the English language, is based. This semantic representation includes Frames and their LUs and allows the connection of all the grammatical categories (noun, adjective, verb, adverb) with a Frame.

Consequentially, the theory of the semantic Frames has been further utilised for the formulation of the Frame-based Terminology (FBT) theory and for the concomitant creation of terminological bases. According to this approach, the way that the senses which belong to a thematic field are realised and connected with each other depends on the events of the field (task-oriented). FBT is, according to Faber (2011, 2012, in Faber 2014:14), a cognitive approach to terminology which directly connects the specialised knowledge with Cognitive Linguistics and Semantics. It uses a modified version of Fillmore's Frames (Fillmore 1982, 1985) along with the premises of Cognitive Linguistics (Faber 2011).

In this context, Faber (2011) describes the specialised language as dynamic and supports that its representation should be dynamic as well, although this approach is not being used adequately in the terminological resources. She further supports that the way that items are represented in our brain means that current methods and ways of creating representations of specialised knowledge should be modified in order to take this information into account. Specialized language concepts cannot be activated in isolation unless they are part of a larger structure or event. Our knowledge about a concept initially gives us the context or the event in which the concept has a meaning for us. Consequently, concept representations should, instead of being presented out of context and being static, be presented inside their context and be dynamic (Faber 2011).

Over the years, a number of frame-based language resources have been developed for other languages for general purposes (FrameNet Brazil (Salomão 2009), Spanish FrameNet (Subirats 2009) and Japanese FrameNet (Ohara 2009), and the Swedish FrameNet++ (Ahlberg et al. 2014), inter alia).
As far as the Greek language is concerned, there has been previous work in language for general purposes, but no work has been reported for language for specific purposes. In fact, an initial attempt to build a frame semantics lexical resource for Greek is reported in Gotsoulia et al. (2007); however, this work was conceived of as the preliminary phase of a pilot project for the development of the basic infrastructure and design of the actual resource. Later, a frame-driven approach was followed by Dalpanagioti (2012) to the bilingual lexicographic process for creating a bilingual lexical database of motion verbs for EL and EN. Yet, these studies are fragmented and represent only a part of the general language. Finally, Giouli et al. (2020) report on work towards the development of the Greek (EL) counterpart of the Global FrameNet in the context of the Shared Annotation Task, where the annotation methodology employed, the current status and progress made so far, as well as the problems raised during annotation of the EL corpus are put into focus. In this paper, we have tried to compile a frame-based lexical resource for the domain of finance.

3 Research Scope and Goals

An important – yet still under investigation – characteristic of Frame Semantics and FrameNet is the universal nature of frames. In theory, frames are universal, as are the different concepts across languages. In practice, however, this has yet to be proven. One of the purposes of this work was to examine whether a new approach to terminology, the FBT, can have this characteristic. We tried to see if the frames that include terminology of a specific domain can also be used in another language. To this end, a parallel knowledge base of finance terms has been created for Greek and English.
The present work has been conducted from a translation perspective, as well. One of the aims is to use the resulting resource for facilitating the translation procedure. This is feasible, since FrameNet is structured in a way which makes it readable by people as well as computers. Therefore, it would be possible to utilise the term base in tools that assist translation (machine translation tools).

4 Methodology

For the creation of the term base, the methodology was viewed as a four-step approach: as a first step, the corpus was compiled, consisting of two sub-corpora – one for each language. The second step was the extraction, selection and grouping of terms; in particular, the terms were classified into groups according to their underlying meaning in order to be assigned a frame. The third step was the frame-creation step and other procedures entailed by it; after the creation of the frames that was led by our terms, definitions of FEs were provided and the terms were assigned a frame in both languages. At fourth step, the frames' sentences-examples were annotated in various layers. These steps are described in detail in the following subsections.

4.1 Corpus Creation and Term Selection

For the purposes of this work, a special-purpose comparable corpus was created comprised by two sub-corpora – the EL sub-corpus and the EN sub-corpus. The EL sub-corpus is comprised by 73,069 tokens και 8,146 types, while the EN sub-corpus is comprised by 92,105 tokens and 7,129 types, size which was considered adequate for this work. As Pearson (1998: 56) mentions, when the corpus is designed for special purposes, then a smaller corpus which is derived from a given thematic field is more appropriate than a wider corpus.
The texts selected for this corpus belong to the same genre or thematic field and they were chosen according to their terminology content and coherence.[1] The sources are divided into two groups: (a) journalistic/news texts with finance content and (b) banks' financial results reports. The first group includes Greek and English articles from online newspapers with financial content as well as some with more general content; the second group, on the other hand, contains only financial texts or documents, which are very rich in terminology. This combination allowed us to find very specialised and less specialised terms. Both groups' articles cover topics such as economy, business, markets, stock market, bonds and banks - in other words, the corpus is compiled with texts from specialised fields, with high term ratio, which are suitable for terminological research.
The table below (Table 1) provides details on quantitative data of the two sub-corpora.

	Group (a) - EL	Group (b) - EL	Total EL	Group (a) - EN	Group (b) - EN	Total EN
Texts	121	10	131	92	5	97
Tokens	48,344	24,725	73,069	54,670	37,435	92,105
Types	7,333	2,118	9,451	6,579	1,845	8,424

Table 1: Quantitative data of the two sub-corpora.

[1] For the EL sub-corpus, the sources are: Capital.gr (https://www.capital.gr/), naftemporiki (https://www.naftemporiki.gr/), ΚΕΡΔΟΣ (http://www.kerdos.gr/), Η ΚΑΘΗΜΕΡΙΝΗ (https://www.kathimerini.gr/), ΕΘΝΙΚΗ ΤΡΑΠΕΖΑ (https://www.nbg.gr/) and ΤΡΑΠΕΖΑ ΠΕΙΡΑΙΩΣ (https://www.piraeusbank.gr/el/idiwtes).
For the EN sub-corpus, the sources are: CITY A.M. (https://www.cityam.com/), Bloomberg (https://www.bloomberg.com/europe), REUTERS (https://www.reuters.com/), The Guardian (https://www.theguardian.com/international), Barklays (https://www.barclays.co.uk/) and LLOYDS BANKING GROUP (https://www.lloydsbankinggroup.com/).

On step two, the terms of both sub-corpora were extracted semi-automatically using the software AntConc (Laurence 2016), and in particular its word list, keyword list, concordance and clusters functions for frequency analyses. To be more precise, initially the keywords were extracted and then they were examined in order to locate candidate terms, which later were individually processed through concordance and clusters in order to find candidate multi-word terms. Said procedures were followed in the same order for both sub-corpora.
At this initial stage, a total of 561 candidate terms were identified and selected that pertain to the Noun, Adjective and Verb grammatical categories; being a terminological work, we could locate a substantial number of multi-word nouns (also referred to as "multi-word terms") as well. The distribution of the so-identified lexical items per grammatical category or part-of-speech (POS) is depicted in Table 2.

POS	EL	EN
Nouns (single)	100	100
Nouns (multi-word)	110	184
Adjectives	24	13
Verbs	14	16
Total	248	313

Table 2: POS of candidate terms.

4.2 Creation of Frames and Annotation

The third step, creation of the semantic frames, was the most important and the most laborious one. Our initial attempt was to use frames which have already been created for the lexical resource FrameNet (Ruppenhofer et al. 2016). To this end, the EL terms were grouped according to their meaning and the scene or frame that they could evoke, and then effort was made to assign FrameNet's frames to them. For some of the terms this procedure was quite straightforward, such as those which evoke the frames of lending or borrowing, or those which belong to the scene of commerce. However, for terms of the language for special purposes – which formed the majority of our terms – the creation of new frames was necessary. This was accomplished with deep search in the FrameNet resource and extensive lexicographical research.
As a result of this process, the EL terms were grouped and divided into 9 scenes and 39 frames. At the next stage, we examined whether the frames which had been created according to the EL data could be used for the EN data as well, in order to find out if the already created frames can be used in a language other than the one they have been created. The decision to start with the creation of the EL frames was based on two considerations: firstly, in this way we could ensure that the terminology has been fully understood and organised into frames correctly, and secondly, that the conceptual structures created (the frames) adequately represent the respective terminology.
Since the lexical resource to be created was a terminological resource, we had to annotate (at this stage, manual annotation was performed to assign Core and Non-core FEs to the frames) beginning from nouns instead of verbs, so we deviated from FrameNet's method, as the majority of terms are nouns (single- and multi-word) and only a small proportion of them are verbs. Therefore, the frames were defined according to the terms and the states or events that they represent. The method for the EL frames creation was corpus-driven, whereas that for the EN frames creation was corpus-based, because the already created frames had to be used. This is why the web as a secondary source was used for the EN frames in addition to the primary source for collecting examples in English in cases where the EN sub-corpus was not enough (the frames for which such examples were used are: Withdrawal, Bank_account_management, Stock_exchange_transactions, Owing and Social_contributions).
Regarding the LUs that result from the frames, there are some which have not been derived from the corpus but needed to be added in order to complete the frame. These LUs were found either through the respective frame of the other language or from our general knowledge on a topic. For example, only few of the LUs of the frame Commerce existed in the corpus; however, having the LUs *αγοράζω.v* (to buy) and *αγοραστής.n* (buyer) without their opposites *πουλάω.v* (to sell) and *πωλητής.n* (seller) would make no sense. Additionally, some LUs have been assigned more than one frames indicating cases of polysemy or not very specialised terms – for example, the Greek noun *επενδυτής.v* (investor) could be used in the frames Obtaining_a_loan, Bond_issuing and Commerce.
Another important element that had to be added were the definitions of the FEs in the form of glosses; they were created according to each frame's needs in a way that they represent the concepts involved in a frame and, consequently, the terminology itself as concretely as possible. Moreover, an attempt was made to create these definitions as language independent as possible. The example of FN and Frame Semantics was once again followed for this process: FEs' definitions connect the concepts of each frame. To better account for the creation of the appropriate glosses (definitions), we consulted various lexica and reference works, such as dictionaries, term bases and language portals, namely: the Greek database of financial terms[2], various dictionaries that are available on the Greek Language Portal[3], EcoLexicon (Faber

[2] Available online at https://www.euretirio.com/
[3] Available online at http://www.greek-language.gr/greekLang/index.html

2011)[4], the English FrameNet database[5], Glosbe[6] parallel dictionary, the EU's IATE (Interactive Terminology for Europe) terminology database[7], Investopedia[8], the EL and EN branches of Wikipedia, Linguee EL-EN database[9], Merriam-Webster online dictionary[10], Oxford Dictionaries[11], and WordReference[12] online dictionary. These were used not only for assisting us to better understand the terminology, but also to properly create the FEs' definitions. An example of a frame (namely, the Obtaining_a_loan one) and the encoding performed in terms of annotated FEs and the definitions provided is depicted in Table 4.

The fourth and last step was the annotation of a second corpus including the sentences which were used as examples of the frames (also referred to as "sentences-examples"). This corpus is comprised of 255 sentences (EL: 130, EN: 125), or 6,923 words (EL: 3,832 words, EN: 3,091 words). In order to make the most out of this corpus, automatic pre-processing was in order. For this reason, the web tool UDPipe (Straka & Starková 2017) was used for processing the corpus comprising the sentences-examples of both languages, and tokenization, POS tagging and dependency parsing were performed. This pre-processing is essential because it provides a set of characteristics which are necessary for the following semantic analysis. Finally, annotation on lexical level was performed manually using the web annotation tool WebAnno (Yimam et al. 2013), which allows users to annotate texts in any level of lexical analysis by defining their own annotation scheme. Our annotation scheme were the FEs of our semantic frames, which appeared in the tool as tags. The number of tags we defined are 253, 147 of which are Core FEs and 106 Non-Core FEs. The tags are of course common for both languages, as the EL semantic frames were also used for EN.

5 Lexical Resource Description

The result of the previously described methodology is a bilingual terminological resource consisting of three components: (a) the EL and EN frames including examples, (b) the lexicon comprised of the LUs of the frames, and (c) the annotated corpus with the sentences-examples of the frames. All three components of this work are available and are described in detail below.

5.1 Frames and Scenes

The core of this work are the semantic frames. Together with the examples and the LUs that have been assigned to them, they are a way of presenting and understanding the terminology in both languages. Nine scenes have been created, which are divided into 39 frames. The scenes are the context in which a frame belongs. The following table (Table 3) illustrates the scenes and the frames.

[4] Available online at http://ecolexicon.ugr.es/en/index.htm
[5] Available online at https://framenet.icsi.berkeley.edu/fndrupal/
[6] Available online at https://el.glosbe.com/
[7] Available online at https://iate.europa.eu/home
[8] Available online at https://www.investopedia.com/
[9] Available online at https://www.linguee.com/english-greek
[10] Available online at https://www.merriam-webster.com/
[11] Available online at https://www.oxforddictionaries.com/
[12] Available online at http://www.wordreference.com/

SCENES	FRAMES
Scene 1: Lending	1. Obtaining a loan 2. Loan reimbursement 3. Increase/reduction of interest rate 4. Interest
Scene 2: Bank transactions	5. Deposit 6. Withdrawal 7. Bank account management 8. Way if transaction
Scene 3: Bonds	9. Bond issuing 10. Bond yield
Scene 4: Stock market	11. Stock 12. Stock market results 13. Stock indices 14. Stock exchange transactions 15. Distribution of earnings to shareholders
Scene 5: Financial results	16. Earnings 17. Profit 18. Expenditures 19. Ratio 20. Assets 21. Balance sheet data 22. Final results 23. Risks
Scene 6: Domestic economy	24. Lending a state 25. National budget 26. Facilitation for indebted state 27. Tax payment 28. Return to the financial markets 29. Owing 30. Economic performance index 31. Social contributions
Scene 7: Consumption	32. Commerce 33. Consumer spending 34. Price trend
Scene 8: Economy	35. Financing 36. Positive/negative economic activity 37. Change in price level 38. Financial crisis
Scene 9: Change in economy	39. Change position in a scale

Table 3: The scenes and the frames for the field of finance.

The frames are common for both languages and they are constructed as follows: the name of the frame is followed by its FEs. After that, the LUs are noted, which are the words or terms that evoke said frame, and at the end the frame's examples are listed, which are sentences from the corpus (or the web) that comprise the terminology of the frame. All frames include Core FEs and most of them Non-core FEs as well; the former are necessary elements for the conceptual structure that they describe, and the latter are non-obligatory elements.

As an example, the frame Obtaining_a_loan is shown in Table 4, along with the LUs and some of the examples in both languages. As one can see, the ID number of the frame (in our example: 1.), the name of the frame as well as all the Core and Non-core FEs with their definitions in the form of glosses are common in both languages and they are both written in English for reasons of consistency and better visualisation. The effort to create language independent FEs and definitions seems to have been successful, as for all our frames the same FEs were used for both languages, just like the example of the Obtaining_a_loan frame. This means that the general concepts exist in both EL and EN, given that the scene is common and the concepts, the states and the events that it expresses do not differ between the languages. The differentiation begins at the level of the LUs. In other words, the difference between the two languages lies on the lexicalisation of the concepts, rather than the concepts themselves. More details about the lexical units are provided at the next section.

1. Obtaining a loan - EL	**1. Obtaining a loan - EL**
Core FEs: Borrower: The person or institution who receives the Theme from the Lender for a Duration. Lender: The person or institution who gives the Theme to the Borrower for a Duration. Theme: The object that is transferred from the Lender to the Borrower for a Duration. Non-Core FEs: Duration: The amount of time in which the Borrower has possession of the Theme. Manner: The way in which the Lender lends the Theme. Place: The location in which the Lender lends the Theme to the Borrower. Purpose: The aim of the Lender which they believe will be accomplished by lending the Theme to the Borrower. Time: The time when the lending event occurs. Amount: The amount of money of the Theme.	Core FEs: Borrower: The person or institution who receives the Theme from the Lender for a Duration. Lender: The person or institution who gives the Theme to the Borrower for a Duration. Theme: The object that is transferred from the Lender to the Borrower for a Duration. Non-Core FEs: Duration: The amount of time in which the Borrower has possession of the Theme. Manner: The way in which the Lender lends the Theme. Place: The location in which the Lender lends the Theme to the Borrower. Purpose: The aim of the Lender which they believe will be accomplished by lending the Theme to the Borrower. Time: The time when the lending event occurs. Amount: The amount of money of the Theme.
LUs	**LUs**
δανείζω.v, δανείζομαι.v, δάνειο.n, παίρνω δάνειο.v, λαμβάνω δάνειο.v, δίνω δάνειο.v, δανεισμός.n, τραπεζικός δανεισμός.n, δανειστής.n, δανειστικός.a, δανειακός.a, δανειοδοτώ.v, δανειοδότηση.n, κόκκινο δάνειο.n, τράπεζα.n, εμπορική τράπεζα.n, κεντρική τράπεζα.n, συστημική τράπεζα.n, αγροτική τράπεζα.n, επενδυτής.n, κεφάλαια ρευστότητας.n, χρήματα.n, χρηματοδότηση.n, χρηματοδοτικός.a, πιστωτής.n, πίστωση.n, πιστωτικός.a, πιστώνω.v, επαγγελματικό δάνειο.n, επιχειρηματικό δάνειο.n, στεγαστικό δάνειο.n, ομολογιακό δάνειο.n, εποχικό δάνειο.n, άτοκο δάνειο.n	borrow.v, lend.v, take out a loan.v, borrowing.n, lending.n, lender.n, creditor.n, issue.v, loan.n, home loan.n, mortgage.n, unsecured loan.n, consumer credit.n, defaulted loan.n, bond issue.n, bank.n, Bank of England.n, central bank.n, investment bank.n, investor.n, funding.n, secured lending.n, corporate lending.n, bank lending.n, corporate debt.n
Examples	**Examples**
Οριστική λύση για τη διαχείριση [theme κόκκινων δανείων] συνολικού ύψους [amount 1,75 δισ. Ευρώ] που ΕΙΧΑΝ ΛΑΒΕΙ [borrower 82.000 αγρότες και κτηνοτρόφοι] από την πρώην [lender Αγροτική Τράπεζα] δίνει το υπουργείο Αγροτικής Ανάπτυξης. Η [borrower κεντρική τράπεζα] μείωσε επίσης το αντίστροφο επιτόκιο επαναγοράς –το επιτόκιο με το οποίο ΔΑΝΕΙΖΕΤΑΙ [theme χρήματα] από τις [lender εμπορικές τράπεζες] - κατά 0,25% στο 5,75%.	The Commission hopes that by improving the securitisation process, where assets such as mortgages and consumer credit are bundled together and sold on to investors as bonds, it will unlock [amount up to €150bn] [theme of funding] for [lender banks] to LEND to [borrower consumers] and [borrower growing businesses]. Today we get a new healthcheck on Britain's economy, with new figures showing how much new credit was lapped up by consumer last month, and how many [borrower people] TOOK OUT [theme mortgages].

Table 4: The frame Obtaining_a_loan (EL, EN).

A fundamental principle for the creation of the frames was to use FrameNet's frames to the greatest extent possible. However, the FrameNet's frames that could be adopted without any modification were not a lot due to the different nature of the two resources. Particularly, this work's frames Stock (FN: `Capital_stock`) and Change_position_on_a_scale (FrameNet: `Change_position_on_a_scale`) are the only ones which are exactly the same as the corresponding ones in FrameNet. To the frames Obtaining_a_loan (FrameNet: `Lending`) and Lending_a_state (FN: `Lending`) the FE *amount* was added, as our data required. Similarly, the frame Commerce (FN: `Commerce_Sell`) is the same as its corresponding one in FrameNet, with the addition of the FE *payment*. For the frame Loan_reimbursement some FEs have been used (BORROWER, THEME, LENDER, TIME, AMOUNT, from `Lending`), while the rest were newly created. In the frame Earnings one FE from FrameNet's frame `Earnings_and_losses` was used (TIME, which was renamed into *time period*) as well as the names of two FEs (EARNINGS and EARNER), but not their definitions because they did not completely fit to the specific terminology. Moreover, for the definition of *profit.n* in the frame Profit we used FrameNet's *profit.n* as lexical entry. Similarly, for the definition of *asset.n* in the frame Assets FrameNet's definition of *asset.n* as lexical entry was used. Finally, in a number of frames (for example, Change_in_price_level, Stock_market_results, Stock_exchange_transactions, Earnings, Profit, Expenditures, Ratio) the FEs FINAL VALUE and FINAL STATE of FrameNet's frame `Change_position_on_a_scale` have been used.

It should be mentioned that some FEs can be found in more than one frames; an example are the FEs Borrower and Lender of the frames Obtaining_a_loan, Loan_reimbursement, Increase/reduction_of_interest_rate and Interest of the Lending

scene as well as the frame Lending_a_state of the Domestic_economy scene. All the frames except for one belong to the same scene, the scene that is about lending, which shows that scenes are the general context into which the different frames belong (in this case, the context of lending or borrowing).

5.2 Lexical Units

The resulting terminological resource consists of 374 LUs for EL and 368 LUs for EN. The LUs of the frames do not only include the terms that have been extracted from the corpus, but other LUs that evoke the frames as well. Therefore, we can say that the plurality of the LUs are the terms of this field, even though some of them are also used in general language. For example, the LUs *borrow.v*, *lend.v*, *take out a loan.v* of the frame Obtaining_a_loan are also used in language for general purposes.

Following again the example of FrameNet, each frame is accompanied by the LUs that evoke it. A bilingual lexicon has thus been created including the information shown on Table 5 which comprises some of the EN LUs of the frames Deposit and Withdrawal.

Entry ID	Lexical unit	Part of speech	Frame	Full form	Abbreviation	Alternative form	Synonym	Antonym	Hypernym	Example
EN_049	saving	noun	Deposit							yes
EN_050	transaction	noun	Deposit							no
EN_051	withdrawal	verb	Withdrawal					deposit		yes
EN_052	withdrawal	noun	Withdrawal					deposit	transaction	no
EN_053	money	noun	Withdrawal							yes
EN_054	bank	noun	Withdrawal						credit institution	yes

Table 5: Example of LUs of the lexicon.

This type of representation allows us to explore distinct meanings of the terms, particularly through the lexical relations (synonymy, antonymy, hypernymy/hyponymy) and the definitions of the FEs, and it provides a helpful tool for cases of polysemy. The lexical relations do not only connect the terms, but some frames as well, in a way that they make it easier for someone to understand the concepts that they represent. Here, it should be mentioned that the LUs are listed per frame and there are cases where an LU appears in more than one frames; this is one more way to study instances of polysemy.

It must be clear that there is no full form, abbreviation, alternative form, synonym, antonym and hypernym/hyponym for every LU. Also, even though we tried to provide examples for the majority of the terms, there are some terms or other LUs that evoke certain frames that are not in the sentences-examples of the frames. These are evoked through the lexical relations and we have listed them in the LUs' tables. With a possible future expansion of our resource, more terms and examples can be added.

Additionally, effort was made to align the EL and the EN terms. The unique EL terms are 190 and the unique EN terms 172, out of which there are 137 aligned sets of terms. This number shows us that starting with parallel corpora and following the above-mentioned procedures for two languages, we can end up with a bilingual term base which is useful in many ways, and especially for the translation process. With a future extension of the corpus and subsequent extraction of more terms, the aligned terms will increase, providing us with a term base with valuable semantic information.

5.3 Annotated Sentences-Examples

The annotated sentences are the third component of our resource. In total, the annotated corpus consists of 255 sentences (130 EL and 125 EN), which correspond to 6,923 words (3,832 EL and 3,091 EN). These sentences, which are the frames' examples bare two layers of linguistic annotation via UDPipe and WebAnno tools, and are available for use and further extension.

An example of annotated sentences with WebAnno is shown in Figure 1, where EN sentences of the frame Expenditures are annotated. Above each annotated word or phrase the frame to which it has been assigned and the FE appear. For example, the LU "operating expenses" of sentence 16 evokes the frame Expanditures and has been annotated with the FE "expenditure". The tags in colour make it easier to see the different FEs and may also be useful for discovering repeated patterns in a particular frame. Here, for example, in the frame Expenditures there seems to be a tendency of mentioning first the theme (expenditure), then the predicate and after the FEs *final state* and *final value*. The FE *cause* seems to be used at the end of the sentence.

Essentially, we have provided a semantically annotated corpus which is a useful resource for seeing the relations between the elements that comprise each sentence.

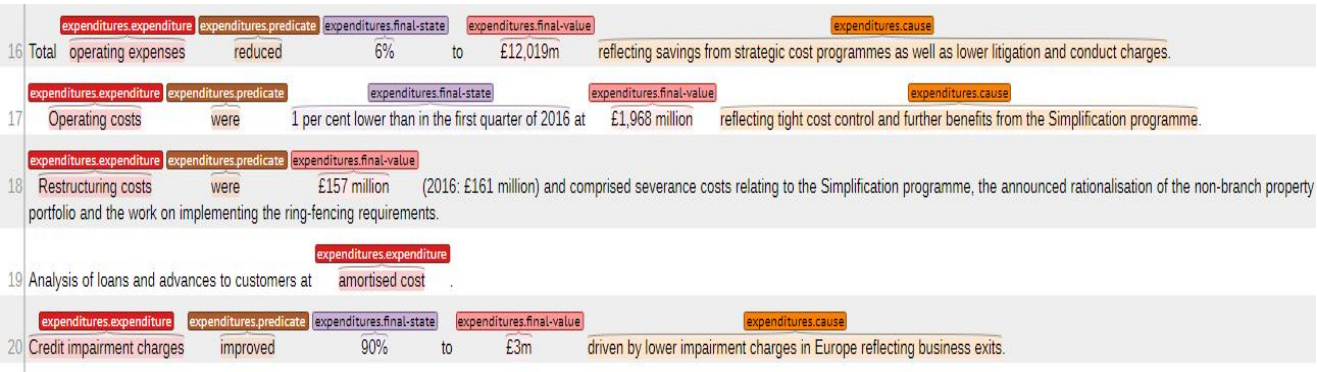

Figure 1: The frame Expenditures (EN) in WebAnno (Yimam et al. 2013).

6 Discussion

The outcome of this work is a bilingual term base which can be used either in its present form or for the creation of an electronic data base. The term base can also be utilised for the assistance of the translation procedure, for example in software designed for this purpose. In fact, the possibility to use an annotated corpus like the one described in this paper breaks new ground for terminography for assisting translation.

One of the biggest challenges that we had to face during the alignment of the EL and EN frames was that of equivalence. The major problem of language and the primary concern of linguistics is "equivalence in difference" (Jacobson 1959, in Munday 2002: 71), meaning that, regardless of differences someone may deal with while translating, the total equivalence between source and target language must be ensured; and since translation from one language into another includes two equivalent messages in two different codes, equivalence in difference is the only acceptable form of equivalence (Kentrotis 1996: 283). An approach to terminography like the one of FBT aims at exactly this kind of equivalence.

During this effort, we needed to make alignments at two levels: first, at the frame level, and then at the LU level. Frame alignment was not problematic; however, there were some problems when we tried to align the EL and EN terms. Firstly, there is a number of pragmatic elements that is hard to translate, even if the concept exists in both languages. One example is the frame Social_contributions, as it contains terms from two different financial systems. Additionally, the expressive meaning (Baker 1992: 23) of the terms might differ, like the EL term *κεφαλαιακοί περιορισμοί.n* (*capital controls.n*), which is more emotionally loaded than the English equivalent due to the Greek financial crisis.

If we look again at the LUs of the frame Obtaining_a_loan that is available in Table 4, we can see that most terms can be aligned, which proves the universal nature of specialised terminology and of the semantic frames. There are also cases where a translational equivalent exists, but it has not been found in our corpus – probably in a future expansion of the resource these terms could also be added.

A notable example of difference in the lexicalisation of concepts is the concept of *δανείζω.v* (give a loan) and *δανείζομαι.v* (take a loan). In Greek, the same verb is used in active and passive voice for expressing two different equivalent verbs in English: *lend.v* and *borrow.v*, respectively. In FrameNet these two concepts form two distinct frames which are called Lending and Borrowing. In our resource, however, we decided to combine them under one common frame (Obtaining_a_loan), because our aim was to gather all the terms that pertain to bank lending in one frame. This difference in lexicalisation can also lead to other issues, for example difficulty in annotation due to discrepancies among languages in the lexicalisaton of concepts. The majority of the discrepancies, however, tend to be with verbs denoting financial-related events, rather than nouns, the latter being more technical and specific.

The present resource is of course smaller in size than the one of FrameNet, and as result the frames are not so extended, in a sense that they include only the FEs that are essential for covering the terms. In a possible future expansion of the resource, more frames and more FEs to the existing frames can be added, so that the resource can cover the biggest possible part of financial terminology.

It should be taken into account that the development methodology of our resource differs in a few fundamental aspects than the one of FrameNet. The most important one is that the frames were created based on the terminology that had to be accounted for; to put it another way, the frames, the FEs and their definitions were all created in order to express the terms in the most precise way possible. They were also viewed as a way to examine whether frames can be used in both languages. Also, it is worth mentioning that the aim of this work was not to compile a complete terminological resource, but rather to explore the methodology and the process for developing a bilingual frame-based resource of a specialized domain.

7 Conclusion

We have presented work which includes the development of a bilingual resource for the domain of finance which is based on Frame Semantics (Fillmore 1977, 1982, 1985). The outcome of our work is: (a) a bilingual lexical resource in electronic format which contains LUs in Greek and English (c. 560 terms, 740 LUs in total). The LUs are described both in terms of their semantic frames and through the listing of the lexical relations with which they are linked to each other; (b) a number of scenes and semantic frames for the semantic field of finance; in particular, the specialised vocabulary is organised around

9 scenes and 39 frames which are common for both languages; and (c) a fully annotated corpus composed of the sentences-examples in which the LUs are attested.

Future work has already been planned towards enriching the Greek component of FrameNet, as well as making comparisons between the Greek and English language. In particular, we participate in the Global FrameNet project, which is a joint effort to bring together FrameNets in different languages. Another possible future prospect is the expansion of our resource, which can be done in a number of ways. More data can be added to the corpus in order to extract more terms and find more examples for our frames. In this way, more FEs can also be added to the frames. From another perspective, the resource can be extended to other domains of language for specific purposes, as a way to examine whether the same principles would apply when following the above-described methods. The resource has been made freely available for research purposes via CLARIN-EL repository.

8 References

Ahlberg, M., Borin, L., Dannélls, D., Forsberg, M, Toporowska Gronostaj, M., Friberg Heppin, K., Johansson, R., Kokkinakis, D., Olsson, L.J., Uppström, J. (2014). Swedish FrameNet++ The Beginning of the End and the End of the Beginning. In *Proceedings of the Fifth Swedish Language Technology Conference*, Uppsala, 13-14 November 2014.

Baker, M. (1992). *In other words. A coursebook on translation*. New York: Routledge.

Baker C. F., Fillmore C. J., Lowe J. B. (1998). The Berkeley FrameNet Project. In *Proceedings of the 17th International Conference on Computational Linguistics* – Volume 1 (COLING '98), Vol. 1. Association for Computational Linguistics, Stroudsburg, PA, USA: 86-90. Accessed at https://dl.acm.org/citation.cfm?doid=980451.980860 [29/05/2020]

Dalpanagioti, T. (2012). *A Frame-driven Approach to Bilingual Lexicography: Lexical Meaning and Usage Patterns in Greek and English Motion Verbs*. PhD Thesis. Εθνικό αρχείο διδακτορικών διατριβών. Accessed at https://www.didaktorika.gr/eadd/handle/10442/27162 [29/05/2020]

Faber, P. (2011). The dynamics of specialized knowledge representation: Simulational reconstruction or the perception–action interface. *Terminology* 17, no 1 (January 1): 9-29. Accessed at http://dx.doi.org/10.1075/term.17.1.02fab [29/05/2020]

Faber, P. (2014). Frames as a Framework for Terminology. In book: *Handbook of Terminology*, Vol. 1, edited by H. J. Kockaert and F. Steurs. Amsterdam/Philadelphia: John Benjamins: 14-33.

Fillmore, C. J. (1977). Scenes-and-frames Semantics. In *Linguistic Structures Processing. Fundamental Studies in Computer Science*, vol. 59. North Holland Publishing: 55-81.

Fillmore, C. J. (1982). Frame Semantics. Linguistics in the Morning Calm, ed. The Linguistic Society of Korea. Selected papers from SICOL-1981: 111-137.

Fillmore, C. J. (1885). Frames and the Semantics of Understanding. *Quaderni di semantica*. Vol. 6, no 2: 222-254.

Giouli, V., Pilitsidou, V., Christopoulos, H. (2020). Greek within the Global FrameNet Initiative: Challenges and Conclusions so far. In *Proceedings of the International FrameNet Workshop 2020, Towards a Global, Multilingual FrameNet. LREC 2020 Workshop, Language Resources and Evaluation Conference*, 11–16 May 2020, Marseille, France. 48-55. Accessed at https://lrec2020.lrec-conf.org/media/proceedings/Workshops/Books/framenet2020book.pdf [31/07/2020]

Gotsoulia, V, Desipri, E., Koutsombogera, M., Prokopidis, P., Papageorgiou, H., Markopoulos, G. (2007). Towards a Frame Semantics Lexical Resource for Greek. In *Proceedings of the Sixth International Workshop on Treebanks and Linguistic Theories*.

Kentrotis, G. (1996). *Theory and Practice of Translation* (in Greek). Athens: Diavlos.

Laurence, A. (2016). AntConc (Version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. Accessed at http://www.laurenceanthony.net/ [29/05/2020]

Munday, J. (2002). *Introducing Translation Studies: Theories and Applications*. (in Greek). Athens: Metaixmio.

Ohara, K. H. (2009). *Frame-based Contrastive Lexical Semantics in Japanese FrameNet: The Case of Risk and Kakeru*: 163–182. Berlin/New York: Mouton de Gruyter.

Pearson, J. (1998). *Terms in context*. Amsterdam: John Benjamins.

Petruck, M. R. L. (1996). *Handbook of Pragmatics.* Jef Verschueren, Jan-Ola Östman, Jan Blommaert, and Chris Bulcaen (eds.). Philadelphia: John Benjamins.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L, Johnson, C. R., Scheffczyk, J. (2006). *FrameNet II: Extended Theory and Practice*. Berkeley, California: International Computer Science Institute.

Salomão, M. M. M. (2009). Framenet brasil: um trabalho em progresso. Caleidoscópio 7(3): 171–182.

Straka, M., Starková, J. (2017). Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*. Association for Computational Linguistics: 8899. Vancouver, Canada. Accessed at http://www.aclweb.org/anthology/K/K17/K17-3009.pdf [29/05/2020]

Subirats, C. (2009). *Spanish FrameNet: A Frame-semantic Analysis of the Spanish Lexicon*. Berlin/New York: Mouton de Gruyter.

Tantos, A., Markantonatou, S., Anastasiadi-Symeonidi, A., Kyriakopoulou, T. (2015). *Computational Linguistics* (in Greek). [E-book] Athens: Hellenic Academic Libraries Link. Accessed at https://repository.kallipos.gr/bitstream/11419/2205/13/nlp2ndEdition.pdf [29/05/2020]

Yimam, S. M., Iryna, G., de Castilho, R. E., Biemann, C. (2013). *WebAnno: A Flexible, Web-based and Visually Supported System for Distributed Annotations*. ACL.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments to the manuscript that contributed to improving the final version of the paper. The research leading to the results presented here was conducted in the context of the Translation and Interpreting MA programme of the Faculty of Turkish and Modern Asian Studies of the National and Kapodistrian University of Athens. Also, they would like to specially thank Prof. E. Sella and Prof. I.E. Saridakis, for their support throughout this effort and their comments.