International Journal of Language, Translation and Intercultural Communication
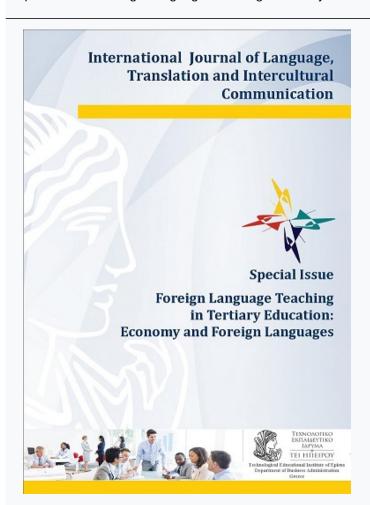
Special Issue: Foreign Language Teaching in Tertiary Education: Economy and Foreign Languages

International Journal of Language, Translation and Intercultural Communication

Special Issue

Foreign Language Teaching in Tertiary Education: Economy and Foreign Languages

ΤΕΧΝΟΛΟΓΙΚΟ
ΕΚΠΑΙΔΕΥΤΙΚΟ
ΙΔΡΥΜΑ
+
ΤΕΙ ΗΠΕΙΡΟΥ
Technological Educational Institute of Epirus
Department of Business Administration
Greece

Language resources to be used in teaching of LSP and specialised translation: Annual Reports in the banking sector as ad-hoc parallel and comparable corpora, in English, French and Greek

ELENI TZIAFA

doi: 10.12681/ijltic.10350

To cite this article:

Language resources to be used in teaching of LSP and specialised translation: Annual Reports in the banking sector as ad-hoc parallel and comparable corpora, in English, French and Greek

Eleni Tziafa
University of Cyprus
tziafa.eleni@ucy.ac.cy

Abstract

This study investigates the use of corpora in teaching of LSP and specialised translation, summarising the findings of the analysis of Annual Reports in the banking sector in English, French and Greek. The objectives of this study are: i. to connect texts and translations from these three languages, whereas English, serving as an intermediate language will help us make the proper choices of terms in Greek and French financial choices, based on informed choices of professional translators, ii. to prepare teaching materials tailored to the communication needs of students of economics or participants of business English courses, as a result of a contrastive linguistic research on the basis of parallel and comparable corpora. For this purpose, we apply easy to use and appealing concordance programs, so that teachers and learners are not put off from working with corpora because the software is too complex or not user-friendly enough. Both teachers and learners would profit a lot more from getting their hands on corpora themselves.

Keywords*: LSP, annual reports, banking sector, parallel corpora, comparable corpora*

1 Introduction

"The mobility of Europe's boundaries and people within them, together with growing opportunities for cross-border trade, adds to the complexity of language use across Europe and encourages the development of an international language or languages" (Rogerson-Revell, 2007: 105).

While there is an obvious necessity, there are also certain stumbling blocks, as regards specialised translation: Medium density languages, such as Greek, lag behind these high density languages as regards resources, and highly market-demanded specialised translation poses serious problems which go well beyond finding well established terminology equivalents (Pastor, 2007), while Henry & Roseberry (2001) even suggest that to compile genre-specific compendia or glossaries based on a specialised corpus would bring the learner more "success in job hunting". Thus, the modern way of teaching LSP and specialized translation appears to be a one way road: corpora, monolingual, parallel and comparable. However, such corpora, the lifeblood of modern language technology efforts, may be either non-existent or difficult to obtain, or require the use of more and more sophisticated tools of modern computational linguistics.

Nevertheless, if we cannot provide students with ready-made corpora, we can still teach them how to construct a specialized corpus, tailored to their needs. The objective of this study is to prepare teaching materials tailored to the communication needs of students of economics or participants of business English courses, as a result of a contrastive linguistic research on the basis of parallel and comparable corpora. For this purpose, we apply easy to use and appealing concordance programs, so that teachers and learners are not put off from working with corpora

because the software is too complex or not user-friendly enough. Both teachers and learners would profit a lot more from getting their hands on corpora themselves.

This study investigates the use of corpora in teaching of LSP and specialised translation, summarising the findings of the analysis of Annual Reports in the banking sector in English, French and Greek. Through parallel and comparable corpora not only the right target language terms or multiword lexical units can be identified, but we can also select the right collocational profile. A positive result is that as well as getting instant access to real usage, we also get translation equivalents unavailable in existing lexicographic resources. Our objective is to connect texts and translations from three languages. English, serving as an intermediate language will help us make the proper choices of terms in Greek and French financial choices, based on informed choices of professional translators.

2 Objectives & Methodology

Greek is a low resourced language, or a medium density language, as regards availability of digitally stored material. However, while English may well be the most widely used business language in Europe, in a survey of language use in European businesses, Hagen (2010) claims that, in order to do cross-border business successfully, companies need to be able to communicate in all three of Europe's 'Big Languages', namely English, French and German, among other, emerging languages. These are the stars that we have to hitch our wagon to, in order to move on to the digital age.

We are missing large open source dictionaries, specialized dictionaries and corpora of all kinds, available as a whole. A solution to this problem is to make use of tailored-made or ad hoc corpora.

Our objective in this study is to show that a fairly simple text processing and only minimally language-specific data cleaning steps can produce useful basic language resources, in particular corpora and frequency dictionaries that are considerably larger than those currently available.

Furthermore, as Kornai (2013) suggests:

> For vital languages this is not a hard barrier to cross. We can generally put together a gigaword corpus just by crawling the web, and the standardly translated texts form a solid basis for putting together a parallel corpus. But for borderline languages this is a real problem, because online material is so thinly spread over the web that we need techniques specifically designed to find it, and even these techniques yield only a drop in the bucket. To make matters worse, the results of this crawl are not available to the public for fear of copyright infringement, yet in the digital age what cannot be downloaded does not exist. Remarkably, the fruits of such state- and EU-funded efforts tend to remain behind query interface walls: individual results pertaining to individual words may be accessible, but the data as a whole is not available for download (Kornai, 2013).

Thus, there are many stumbling blocks as regards specialised translation that we have to overcome.

Language teachers and learners can, as Barlow (1996) points out, treat parallel corpora like an online contextualized dictionary. Thus, teachers and learners can use parallel corpora to show or study how a particular word or phrase is translated and used in another language. Parallel corpora are especially useful when the word or phrase that the teacher or student wishes to examine does not appear in a dictionary, as so often happens with new terminology (Bowker & Pearson, 2002). As Barlow (1996) notes, "[t]he results of a corpus-based investigation can serve

as a firm basis for both linguistic description and, on the applied side, as input for language learning."

On the other hand, according to Gavioli & Zanettin (1997), comparable corpora provide "a repertoire of naturally occurring contexts in the target language onto which hypothesised translations can be mapped." They hence "problematize the choices of the translator" (ibid.) or trainee translator, and help him or her find the most adequate and acceptable translation.

The steps we took in this case study involved the collection of data, searching for parallel and comparable data, the use of every available software/tool for terminology extraction, the creation of a glossary and of a web page, in order to share our glossary and help other translators. These findings shed light on some of the terminology issues that may be present in such contexts and the possible communication difficulties that may result.

3   Case study: Annual reports of British, French and Greek banks

3.1 Annual reports' text and structure

For all their perceived dullness and lack of readability, annual reports are controversial documents, generating disagreement regarding audience, objectives, and credibility. Research suggests that most individual investors consider the annual report a fair to poor source of information, glossing over the bad news to make management look good (Hill & Knowlton, 1984). They are like gigantic brochures, nevertheless, offering valuable information. Communicating an objective message effectively and presenting oneself in a favourable light at the same time is of utmost importance in business. This is the aim of company annual reports, to reinforce positive image of the company in the eyes and minds of different sorts of its stakeholders (shareholders, regulators, the financial media, employees, creditors, customers, suppliers, but also academic researchers, etc.). Company law requires European firms to publish an annual report containing audited financial statements together with substantial narrative disclosures on a range of business and governance related topics. Banks in France and Greece publish their annual reports in English as well, and these are usually more strictly parallel than the rest of their web pages. However, as Varga et al. (2007) suggest, "there is no denying that the identification of such resources, negotiating for their release, downloading, format conversion, and character-set normalization remain labor-intensive steps, with good opportunities for automation only at the final stages."

Moreover, two more factors restrict the ability of researchers to analyse these important disclosures on a large sample basis. According to Athanasakou et al. (2014), "first, contrary to the structured reporting system under which US registrants operate for mandatory annual filings such as 10-Ks, European firms have considerable discretion regarding both the content of their annual report narratives and the format in which these narratives are presented. A Form 10-K is an annual report required by the U.S. Securities and Exchange Commission, which gives a comprehensive summary of a company's financial performance".

In practice, European annual reports therefore display considerable variation both across firms and over time for a given firm. Second, electronic versions of European annual reports are normally provided in PDF format, which severely restricts researchers ability to directly access and extract specific categories of narrative disclosure.

In contrast to the USA, stock exchange-listed firms in Europe do not present their financial information and accompanying narratives in a standardised format when creating annual reports. Instead, European firms have much more discretion regarding the structure and content of the annual report. Finally, whereas financial filings made by firms in the USA are presented in plain text format, UK firms' annual reports are published as PDFs. The combination of these factors makes identifying document structure much more problematic for European firms compared to their counterparts in the USA.

A typical structure of an annual report includes, more or less, the following parts:

- Highlights
- Chairman's Statement
- CEO's Review
- Finance Director's Review
- Other Reviews
- Strategy and Business Model
- Directors Report
- Risk Report
- Corporate Social Responsibility Report
- Governance and Remuneration
- Auditors Report
- Accounting Policies
- Financial Statements
- Notes to the Accounts
- Other Statutory Information

3.2 The corpus under study

The corpus data are interweaved, in the way that French and Greek companies, along with companies all over the world, traditionally translate their annual reports and financial documents/statements in English. Therefore, English serves as the connecting link between two other languages.

The year selected was 2014, published on March, 2015, that is we selected the latest reports available. Each year, the annual report is published for the previous financial year. Since annual reports have a standard structure, it wouldn't be very useful to collect data from previous years, as the narrative is usually presented in much the same template as in the year before.

Therefore, the corpus in study is a specialized corpus of financial texts of the same year, within the same settings, belonging to the same genre or register (formal) and the same field (business).

The parallel data include the European Central Bank (ECB) annual reports of the year 2014, published on March 2015, which are available in 22 languages. For each language, English, French and Greek, a corpus of about 65-80,000 words is formed.

Regarding the comparable data, the top five (largest) banks[1] from the respective country were selected, which are listed on each country's Stock Exchange, forming a corpus of 481,279 words for Greek, 1,131,823 words for UK English and 1,781,664 words for French. The selection of the top five banks is justified, according to Subramanian, Insley, & Blackwell (1993), who reviewed the relationship between the performance of companies and the readability of their annual reports, finding that the annual reports of companies that performed well were easier to read than those of companies that did not perform well.

The annual reports for the UK English corpus are derived from the following banks:

- HSBC Holdings
- Barclays PLC
- Royal Bank of Scotland Group
- Lloyds Banking Group
- Standard Chartered PLC

These are not necessarily the best UK banks (based on best current accounts with the highest AER percentage). This ranking is on the basis of their size and not customer satisfaction level. For example, Barclays is one of the largest banks in the UK but is not considered a best UK bank due to its many customer service issues and complaints[1]. Although headquartered in the UK, each of these five largest British banks conducts business around the world and has a major global footprint. HSBC, for example, is this year's top largest UK bank, based on total assets, and it does business across 80 countries. Barclays is the 2nd biggest British bank (headquartered in London), but it provides banking services across Europe, the Americas, Africa, and Asia.

The annual reports for the French corpus come from:

- BNP Paribas SA (4th in World Ranking)
- Crédit Agricole SA
- Société Générale
- Groupe BPCE
- Natixis

For the Greek corpus, we actually included the only 5 Greek banks listed on the Athens Stock Exchange:

- Alpha Bank
- Eurobank Ergasias
- National Bank of Greece
- Piraeus Bank

---

[1] Based on bank rankings compiled from balance sheet information included on Bankers Almanac available at 24 June 2015, http://www.accuity.com/useful-links/bank-rankings.

- [Attica Bank](#)

All links to the annual reports consistuting the corpus under study are provided in the Annex.

3.3 The glossary produced

For term extraction there are both commercial tools, such as SynchroTerm, a Canadian-based statistical term extractor from Terminotix[2], or SDL MultiTerm Extract[3], and also open source tools, such as TaaS[4], TermoStat Web 3.0[5], and others, such as Lexterm, VocalGrabber, TermMine, Fivefilters, etc. For the purposes of this study, SDL MultiTerm Extract, a tool popular among translators, was used, leading to the formation of multilingual termbases.

In this procedure, there are certain weaknesses, due to noise, as many non-terms, not relevant expressions are extracted; there is also silence, as especially one-word terms cannot be extracted easily (unless creating even bigger noise). Moreover, not all official EU languages (in particular Greek) are supported.

The Greek alphabet has always been seen as an obstacle in the digital age (less programs, less suitable encoding). Nevertheless, in some cases, we can turn this difference into an advantage, in order to extract terms in texts. Moreover, a particularly interesting aspect of this corpus was the extraction and analysis of foreign words kept as they were or in parentheses in the documents that is the terms that the translators considered as too troublesome. These data provide insights into the cohesion of texts as well as into what kind of meaning is left unexpressed, demanding reader and translator interpretation. With simple regular expressions we extracted from the Greek corpora 350 terms in parentheses in Latin alphabet, as shown in the following example in Greek, translated also in English and French:

- *Επιπρόσθετα, πριν από την εκπνοή της οκταετούς μεταβατικής περιόδου θα αναπτυχθεί και ένας **κοινός μηχανισμός στήριξης (backstop)** με σκοπό τη διευκόλυνση του δανεισμού του ΕΤΕ. (GR)*

- *In addition, a common **backstop** to the SRF will be developed before the end of the eight-year transitional period of the SRF. This backstop will also facilitate borrowing by the SRF. (EN)*

- *En outre, un **filet de sécurité** commun au FRU sera mis en place avant la fin de la période de transition de huit ans. Il facilitera également les emprunts par le FRU. (FR)*

Additionally, the terms left untranslated are also especially interesting, as shown in the following example:

- *Προσφέρεται επίσης, ένα ευρύ φάσμα χρηματοοικονομικών προϊόντων και υπηρεσιών υψηλής ποιότητας, περιλαμβανόμενων της διαχειρίσεως κεφαλαίων και του private banking, της διαθέσεως ασφαλιστικών προϊόντων, του factoring και του*

---

[2] Available at http://www.terminotix.com.

[3] Available at http://www.sdl.com/cxc/language/terminology-management/multiterm/extract.html.

[4] Available at https://term.tilde.com.

[5] Available at http://termostat.ling.umontreal.ca. Greek is not included in languages supported.

leasing, της επενδυτικής τραπεζικής, των χρηματιστηριακών εργασιών και της διαχειρίσεως ακίνητης περιουσίας (Source: Alpha Bank AR 2014)[6].

Then, we looked for them in the other corpora, forming a core term base, a core for the glossary available online at http://annualreportsglossary.weebly.com as shown in Images 1 and 2. The terms found in Latin alphabet in the Greek corpus are marked with red color.
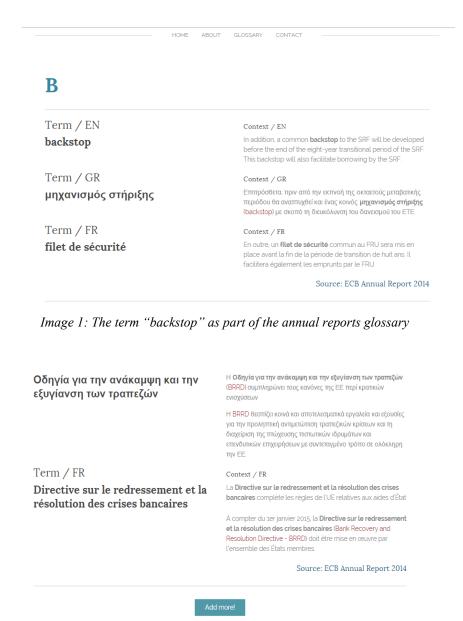

Image 1: The term "backstop" as part of the annual reports glossary

Image 2: The term "Bank Recovery and Resolution Directive (BRRD)" as part of the annual reports glossary

---

[6] Which is translated as: *"A broad range of financial products and services of high quality is also provided, including fund management and private banking, insurance products provision, factoring and leasing, investment banking, stock exchange and real estate management".*

In image 2, we can also see the invitation to users (or students) to add more terms to a site under development, where new terms are constantly added. In total, 1,183 terms were extracted from the corpora constructed, and they are gradually added to the online glossary, which includes the terms in three languages (English, French and Greek), their context and their source.

4   Conclusions

This study has shown that by collecting a minimal amount of corpus data we can form a glossary that can serve as a basis for translation and linguistic studies and also for translation practice. Although this study led to important findings, further analysis of annual reports is warranted. By using this study and the core termbase constructed as a benchmark, a stronger understanding of annual reports can be gleaned and also a translation aid, promoting digital literacy and contributing to the survival of less resourced languages into the digital age. It is hoped that this core termbase will continue to grow, both as regards its terms and its users and contributors.

References

Athanasakou, V., El-Haj, M., Rayson, P., Young, S., & Walker, M. (2014). Computer-based analysis of the strategic content of UK annual report narratives. American Accounting Association Annual Meeting, August 2-6, 2014, Atlanta, USA. Available at http://ucrel.lancs.ac.uk/cfie/AthanasakouEtAl_AAA_outline.pdf.

Barlow, M. (1996). Corpora for Theory and Practice. In: *International Journal of Corpus Linguistics* 1(1), 1-37.

Corpas Pastor, G. (2007). "Lost in specialised translation: the corpus as an inexpensive and under-exploited aid for language service providers". In *Translating and the Computer* 29, (pp. 1-18), Proceedings of the ASLIB Conference, London: Aslib.

Gavioli, L. & Zanettin, F. (1997). "Comparable corpora and translation: a pedagogic perspective". Available at http://www.sslmit.unibo.it/cultpaps/laura-fede.htm.

Hagen, S. (2010). Mapping successful language use in international business: How, when and where do European companies achieve success? In G. Stickel, *Language Use in Business and Commerce in Europe* (pp. 23-34). Frankfurt am Main: Peter Lang.

Henry, A., & Roseberry, R. (2001). "A narrow-angled corpus analysis of moves and strategies of the genre: 'Letter of application'." English for Specific Purposes, 20, (pp. 153-167).

Hill & Knowlton, (1984). The annual report: A question of credibility. New York: Hill & Knowlton.

Kornai, A. (2013). Digital Language Death. *PLoS ONE* 8(10): e77056. doi: 10.1371/journal.pone.0077056.

Rogerson-Revell, P. (2007). "Using English for International Business: A European case study". In *English for Specific Purposes* 26, (pp. 103-120).

Subramanian, R., Insley, R. G., & Blackwell, R. D. (1993). Performance and readability: A comparison of annual reports of profitable and unprofitable corporations. *The Journal of Business Communication* 30(1), (pp. 49-61).

Varga, D., Halacsy, P., Kornai, A., Nagy, V., Nemeth, L., et al. (2007). Parallel corpora for medium density languages. In: Nicolov, N. Bontcheva, K., Angelova, G., Mitkov, R. (eds) *Recent Advances in Natural Language Processing* IV. Selected papers from RANLP-05, Amsterdam: Benjamins (pp. 247-258).

Annex – Links of Annual reports

ECB 2014 Annual reports:

https://www.ecb.europa.eu/pub/pdf/annrep/ar2014en.pdf?4b7cd65cf2d8c49a280596c968bcd2b6 (EN)

https://www.ecb.europa.eu/pub/pdf/annrep/ar2014el.pdf?7e7e73704803e9ff5abe4fc573400fd1 (GR)

https://www.ecb.europa.eu/pub/pdf/annrep/ar2014fr.pdf?cb1f3ac57ad767128a7df54db7a6323e (FR)

UK banks Annual Reports:

HSBC Holdings:

http://www.hsbc.com/investor-relations/financial-and-regulatory-reports/annual-report-and-accounts-2014

Barclays PLC:

https://www.home.barclays/annual-report-2014.html

Royal Bank of Scotland Group:

http://investors.rbs.com/~/media/Files/R/RBS-IR/2014-reports/annual-report-2014.pdf

Lloyds Banking Group:

http://www.lloydsbankinggroup.com/globalassets/documents/investors/2014/2014_lbg_annual_report.pdf

Standard Chartered PLC:

https://www.sc.com/annual-report/2014/

French banks Annual Reports:

BNP Paribas SA:

https://invest.bnpparibas.com/sites/default/files/documents/bnp_paribas_ra_2014_fr.pdf (FR)

http://www.bnpparibas.com.br/PDFs/Documentos_Geral/AnnualReport_2015.pdf (EN)

Crédit Agricole SA:

http://www.credit-agricole.com/Investisseur-et-actionnaire/Information-financiere/Rapports-annuels-et-resultats (FR)

http://www.credit-agricole.com/en/Investor-and-shareholder/Financial-reporting/Credit-Agricole-S.A.-financial-results (EN)

Société Générale:

http://www.societegenerale.com/sites/default/files/societegeneraleddr2014_0.pdf (FR)

http://www.societegenerale.com/sites/default/files/societegeneraleddr2014en_0.pdf (EN)

Groupe BPCE:

http://www.groupebpce.fr/Investisseur/Resultats/Documents-de-reference (FR)

http://www.groupebpce.fr/en/Investors/Results/Registration-documents (EN)

Natixis:

https://www.natixis.com/natixis/upload/docs/application/pdf/2015-03/natixis-docderef-2014.pdf (FR)

https://www.natixis.com/natixis/upload/docs/application/pdf/2015-03/natixis_-_2014_registration_document.pdf (EN)

Greek banks Annual Reports:

Alpha Bank:

https://www.alpha.gr/files/investorrelations/ANNUAL_REPORT_2014_EN.pdf (EN)

http://www.alpha.gr/files/investorrelations/ANNUAL_REPORT_2014_GR.pdf (GR)

Eurobank:

https://www.eurobank.gr/Uploads/pdf/Annual_Report_2014.pdf (EN)

https://www.eurobank.gr/Uploads/pdf/Etisios_Apologismos_2014.pdf (GR)

National Bank of Greece:

https://www.nbg.gr/english/the-group/investor-relations/annual-report-offerring-circular/Documents/ANNUAL%20REPORT%202014.pdf (EN)

https://www.nbg.gr/greek/the-group/investor-relations/annual-report-offerring-circular/Documents/Ετήσιες%20Εκθέσεις%20και%20Ενημερωτικά%20Δελτία/2014_ΕΤΗΣΙΑ_ΕΚΘΕΣΗ.pdf (GR)

Piraeus Bank:

http://www.piraeusbankgroup.com/~/media/Com/2015/Files/Investor-Relations/Financial-Data/Annual-Reports/Annual_Report2014_En.pdf (EN)

http://www.piraeusbankgroup.com/~/media/Com/Downloads/Investors/Financials/PB_EE_net.pdf (GR)

Attica Bank:

http://www.atticabank.gr/images%5Csiteimages%5C/annual_reports%5Cen%5CATTICA_BANK_ANNUAL_REPORT_2014_ENG_FINAL.pdf (EN)

http://www.atticabank.gr/images/siteimages/annual_reports/PR71%20ATTICA%20ANNUAL%20REPORT%202014%20GR%20210x297%20final%20web.pdf (GR)

About the Author

Eleni TZIAFA conducted this research as a Post-Doc Researcher at the Department of French Studies and Modern Languages at the University of Cyprus, where she worked on the project for the construction of a parallel Greek-French corpus of texts, the SOURCe project (http://sourcecorpus.eu/). In 2016 she was appointed as an Assistant Professor at the French Language and Philology Department of the National and Kapodistrian University of Athens. She holds a PhD in Translation and Corpus Linguistics from the Aristotle University of Thessaloniki. Her research interests focus on corpus linguistics and translation studies, with a special interest in the use of tools and methods of corpus linguistics to address problems, questions and issues in translation, discourse analysis, critical discourse analysis and second language acquisition. She has published articles on Greek corpus studies.