

FTA Corpus: a parallel corpus of English and Spanish Free Trade Agreements for the study of specialized collocations

Pedro Patiño¹

¹ NHH Norwegian School of Economics, Norway / Universidad de Antioquia, Colombia

Abstract

This paper describes the Corpus of Free Trade Agreements (henceforth FTA), a specialized parallel corpus in English and Spanish from Europe and America and a smaller subcorpus in English-Norwegian and Spanish-Norwegian that was prepared and then aligned with Translation Corpus Aligner 2 (Hofland & Johansson, 1998). The data was taken from Free Trade Agreements. These agreements are specialized texts officially signed and ratified by several countries and blocks of countries in the last twenty years. Thus, FTAs are a rich repository for terminology and phraseology that is used in different fields of business activity throughout the world. The corpus contains around 1.37 million words in the English section and 1.48 million words in its Spanish counterpart, plus 60,000 words each in the Spanish-Norwegian and English-Norwegian subcorpus. The corpus is being used primarily to study the terms and specialized collocations that include these terms in this kind of specialized texts.

Keywords: specialized collocation, specialized parallel corpus, corpus linguistics, Free Trade Agreement

* Principal contact:

Pedro Patiño, PhD student

Department of Professional and Intercultural Communication, NHH, Helleveien 30, 5045 Bergen, Norway

Tel.: +47 55 95 98 12

E-mail: pedro.patino@nhh.no / ppatino@idiomas.udea.edu.co

1. Introduction

This paper presents the FTA corpus, a specialized parallel corpus using Free Trade Agreements (FTA) in English, Spanish and a Norwegian subcorpus. The data that was used to create the corpus are texts taken from FTAs.

The remainder of the paper is organized as follows. First, the introduction attempts to define the basic notions of collocation, specialized collocation and parallel corpus. The second section of the paper describes the corpus itself and the data that justifies its compilation. Last, the conclusion offers some remarks regarding the corpus and its applications.

In the present work, the definition of collocation proposed by Bartsch (2004: 76) is adopted: “lexically and/or pragmatically constrained recurrent co-occurrences of at least two lexical items which are in a direct syntactic relation with each other”. The phenomenon of collocation in general texts has been studied extensively, especially after Firth's famous statement: “You shall know a word by the company it keeps” (1957:11). The corpus was compiled to perform research that is focused on the collocations that include terms and that appear in FTAs. Other authors have previously described some features of this type of lexical units in specialized texts and some have named the phenomenon LSP phraseology (Picht 1990; Kjær 1990; Budin 1990; Galinski 1990; Pavel 1993; Cabré 1998; Lorente 2002a, 2002b; L'Homme 1998, 2001, 2003, 2006, 2009; L'Homme and Bertrand 2000; Gozdz-Roszkowski 2011). However, to date, these units remain underdescribed in the literature and underrepresented in dictionaries, in part, because LSP phraseology is not the direct object of study neither of phraseology nor terminology (Kjær 2007). Some verbal collocations appearing in the FTA corpus are given as examples: In English (*prescribe a conformity assessment; cover a procurement; apply a safeguard measure*) and other Spanish verbal collocations (*suministrar un servicio; extender un arancel aduanero; autorizar la entrada temporal*). To refer to these lexical units, the term “specialized collocation” is used in this paper. A specialized collocation is defined in this way in my PhD research: A specialized collocation is a multiword expression composed of at least a term that serves as the node and its collocates, where the constituents (nouns, verbs, adjectives and adverbs) are in direct syntactic relation to each other and make a lexical combination that can be unpredictable and semi-compositional and have an internal and statistical tendency of preference.

It is pertinent to study the specialized collocations that appear in FTAs because current lexicographic and terminographic resources do not customarily include the collocations but only the terms, which are normally nouns. This collocational information is a key component in any language and therefore should be included in dictionaries, not only in general dictionaries but also in specialized dictionaries as some authors have pointed out (Benson 1985; Cop 1991; L'Homme 2006; Orliac 2004; Moon 2008). Besides, each subject field exhibits a peculiar phraseology and thus, since collocations can be unpredictable, even native speakers do not know how to combine terms in a particular subject field (L'Homme 2006; Bartsch 2004; Gozdz-Roszkowski 2011). Therefore, language professionals such as translators, terminologists, lexicographers, LSP learners and instructors, and technical writers need this lexical information. The need for this information is increased when working with a foreign language where a perfect command of phraseology is not an easy task to attain.

A combination of corpus-based and corpus-driven techniques are used to extract this information from the corpus. It is corpus-based because first the terms that constitute the node of the collocation are previously selected. The corpus-driven techniques are used to find which collocates are statistically relevant and are retrieved using software tools.

2. Trade Agreements

Trade is a very important human activity that emerged since the beginning of civilization. It has promoted not only economic exchange, but also cultural and political integration among nations. Nowadays, in our globalized world, trade continues to be an important activity and several

supranational organizations have developed a framework to regulate international trade in the form of FTAs. In 2010, according to WTO data, world merchandise exports reached a total of USD 14,851 billion and the European Continent alone has a share of 37.9².

Free trade agreements are specialized official documents that set the norms for the trade of goods among two or more parties and thus can be a rich repository for terminology and phraseology that is used in different fields of business activity throughout the world. Thus, whether it is Norwegian salmon, Colombian coffee, Peruvian avocado or any other product, the trade of these goods among nations is regulated by some trade agreement. Free trade refers to the “absence of government policies designed to regulate international trade, especially import limitations such as tariff or quotas” (Moon 2000a:574). Thus, a trade agreement is any initiative that involves the cooperation of two or more governments to facilitate or regulate trade between their economies. Trade agreements can be either bilateral, minilateral (or regional) and multilateral (or global) (Moon 2000b). A free trade area is defined as “a regional bloc made up of two or more countries which agree to liberalize their bilateral trade, while maintaining their restrictions on trade within third countries” (Nicolaidis 2000a:575).

Most of the agreements included in the FTA corpus have been officially signed and ratified by several national and supranational countries and entities in the last two decades. Specifically, these entities are the World Trade Organization, the European Free Trade Association, the European Union, the Free Trade Area of the Americas, Canada, the United States of America, Mexico, the Caribbean Community, the Dominican Republic, Colombia, Perú and Chile. Since the FTA have in principle equal legal status, they are not translations of each other. However they were written in different forms: some of the agreements were written in English and then translated into Spanish or vice versa. In other cases, as the negotiation advances using normally English as a lingua franca, the teams of free trade experts write a bilingual draft (Šarčević 2000), with each team writing in its mother tongue. Nevertheless, a parallel corpus is understood here in the same sense as Aijmer (2008), as a collection of source texts and their translations, aligned at the sentence level. The corpus is specialized because it gathers texts from a specific subject field, in this case, texts taken from FTAs. This allows to study the terms and collocates of this type of texts in their context (Koester 2010).

FTAs are specialized texts aimed at expert-expert communication (Spang-Hanssen 1983; Pearson 1998). The nature of these texts allows for the presence of terminology from several subject fields, like the fields of law (arbitral tribunal), economics (issuance of bills of lading) and from the disciplines involved in the goods subject to trade (*multimodal transport operator*).

3. Corpus description

The FTA corpus consists of 233 XML source files in each language. The corpus contains approximately 1370000 words in the English section and 1483000 words in its Spanish counterpart. A Norwegian subcorpus, comprising three FTAs and 60,000 words was also aligned for the language pairs English-Norwegian and Spanish-Norwegian. Today’s huge “mega-corpora” comprising 500 million words are being built, and therefore by today's standards a corpus with 1.5 million words is rather small. However, there is a general agreement that a small corpus contains up to 250,000 words (Koester 2010).

The corpus includes texts from different language variants as it comprises texts written in English from the United States and Canada, European Union texts, as well as Spanish from many Latin American countries and blocks of countries. It also includes the texts of the Free Trade Area of the Americas (FTAA) and the World Trade Organization. The FTAA agreement has not been ratified so the version offered is a draft version that includes several terminology options that the teams of negotiators have suggested and that could be interesting to study

²http://www.wto.org/english/res_e/statis_e/its2011_e/its11_world_trade_dev_e.htm

terminological variation, for example: “Each Party shall maintain [or establish] judicial, [quasi-judicial,] [arbitral] or administrative tribunals”.

The original files were downloaded as PDF, HTM and RTF files that were converted to a XML code that is readable by the Translation Corpus Aligner 2 software. The data was downloaded from publicly available web pages from the Foreign Trade Information System of the Organization of American States³ and the European Union EUR-LEX⁴. Regarding copyright issues, the website of the former institution includes this statement:

The General Secretariat of the Organization of American States (GS/OAS) holds copyright on the information available on this website, unless otherwise stated. Copyright in any third-party materials found on this website must also be respected. Anyone may use or reproduce any information presented on this website for educational and other non-commercial purposes, provided that the use of such information is accompanied by an acknowledgement of the GS/OAS as the source. The GS/OAS reserves the right to withhold permission to reproduce this material.⁵

Thus, since the objective of the corpus is educational and non-commercial, to the best knowledge of the author, the corpus does not infringe the copyright.

Some of the PDF files were not directly convertible to plain text files and had to be first processed with Abbyy Fine Reader 9.0⁶, an application for Optical Character Recognition.

Before alignment of the data, the longer files were split into smaller ones. Also, since each FTA includes a glossary section with definitions of commonly used terms in the field of international trade, these sections had to be reorganized manually in Spanish to match the alphabetical order in English. Then, the data was aligned using the 2010 version of the software Translation Corpus Aligner 2 (Hofland and Johansson 1998), which allows exporting XML files compliant with the Text Encoding Initiative.

The FTA texts included in the corpus comprise the Spanish and English versions of the agreements signed by several countries or blocks of countries, as shown in Table 1. They differ in length: two of the FTAs, the FTAA draft and NAFTA, account roughly for 30% of the data whereas the shortest text, signed by CARICOM and the Dominican Republic, accounts for less than 1% of the data.

Table 1

Free Trade Agreements included in the corpus.

Countries	English words	Percentage	Year
Canada - Peru	69930	5,10	2008
CARICOM – Dominican Rep.	9458	0,69	1998
CARIFORUM – EU	51483	3,76	2008

³ http://www.sice.oas.org/agreements_e.asp

⁴ <http://eur-lex.europa.eu/JOHtml.do?uri=OJ:C:2010:083:SOM:es:HTML>

⁵ <http://www.sice.org/disclaim.asp>

⁶ <http://www.abbyy.com>

Countries	English words	Percentage	Year
Chile - EU	34381	2,51	2002
Chile – Australia	64841	4,73	2008
Chile – EFTA	16671	1,22	2003
Chile – USA	86112	6,28	2003
Colombia - Perú - EU	121003	8,83	2011
Colombia - USA	160091	11,68	2006
Colombia – EFTA	69569	5,08	2008
EFTA - Perú	24201	1,77	2010
EU	196494	14,34	1992 / 2007
FTAA (draft)	179747	13,12	2003
Mexico – EFTA	14862	1,08	2000
NAFTA	182990	13,35	1992
World Trade Organization	88548	6,46	1994
Total	1370381		

The data was annotated with Part of Speech tags in English and Spanish using Treetagger (Schmid 1994) and then encoded using IMS Corpus Work Bench (Christ 1994). This way, morphosyntactic patterns can be used to perform queries aimed at retrieving specialized collocations. Figure 1 presents a screen capture of the corpus query interface which allows searching for up to three consecutive tokens in English or Spanish and optionally including stop words. It will allow to perform queries using Part-of-Speech tags as well. Figure 2 shows a screen capture of the results of a query, with parallel source and target segments visible in their context.

Search with [IMS CWB](#)

30.8.09

Corpus **Sub-part**

Texts ▾ All ▾

	Word Original (English)	Word Translation (Spanish)		
Search Ignore case <input checked="" type="checkbox"/> RESET FORM	Whole word ▾	customs	Whole word ▾	
	POS (not yet)		POS (not yet)	
	Stop words		Stop words	
	2nd word			
	Whole word ▾	duty	Whole word ▾	
	POS (not yet)		POS (not yet)	
	3rd word			
	Whole word ▾		Whole word ▾	
	POS (not yet)		POS (not yet)	

Figure 1. IMS CWB Interface to query the FTA Corpus.

[word="customs"%c] [word="duty"%c]:PEDROT [word="#"%c];

Total: 168

Collocates

Except as otherwise provided in this Agreement, no Party may increase any existing customs duty , or adopt any new customs duty, on an originating good. (CAN_PER-1e.s102)	Salvo que se disponga algo distinto en este Tratado, ninguna Parte podrá incrementar ningún arancel aduanero existente, ni adoptar ningún arancel aduanero nuevo, sobre una mercancía originaria. (CAN_PER-1s.s103)
Except as otherwise provided in this Agreement, no Party may increase any existing customs duty, or adopt any new customs duty , on an originating good. (CAN_PER-1e.s102)	Salvo que se disponga algo distinto en este Tratado, ninguna Parte podrá incrementar ningún arancel aduanero existente, ni adoptar ningún arancel aduanero nuevo, sobre una mercancía originaria. (CAN_PER-1s.s103)
Notwithstanding Article 2001 (Administration of this Agreement - The Joint Commission), a joint decision by the Parties to accelerate the elimination of a customs duty on a good shall supersede any duty rate or staging category determined pursuant to Annex 203.2 for that good when approved by each Party in accordance with its applicable legal procedures. (CAN_PER-1e.s106)	No obstante el Artículo 2001 (Administración del Tratado - La Comisión Conjunta), una decisión conjunta de las Partes para acelerar la eliminación de un arancel aduanero de una mercancía, prevalecerá sobre cualquier arancel aduanero o periodo de desgravación definido en el Anexo 203.2 para tal mercancía, cuando sea aprobado por cada una de las Partes de conformidad con sus procedimientos legales aplicables. (CAN_PER-1s.s107)
increase a customs duty to the level established in Annex 203.2 following a unilateral reduction ; and (CAN_PER-1e.s108)	tras una reducción unilateral, incrementar un arancel aduanero al nivel establecido en el Anexo 203.2 ; o (CAN_PER-1s.s109)
maintain or increase a customs duty as authorized by this Agreement, the Dispute Settlement Body of the WTO or any agreement under the WTO Agreement. (CAN_PER-1e.s109)	mantener o aumentar un arancel aduanero cuando sea autorizado por este Tratado, por el Órgano de Solución de Controversias de la OMC o por cualquier acuerdo bajo el Acuerdo de la OMC. (CAN_PER-1s.s110)
the customs duty and any other charge that would be owed on entry or final importation of the good ; and (CAN_PER-1e.s133)	el arancel aduanero y cualquier otro cargo que se adeudaría normalmente por la entrada o importación final de la mercancía ; y (CAN_PER-1s.s134)

Figure 2. Query to the FTA corpus of the term customs duty.

4. Current and future work

A list of 448 terms was obtained automatically by extracting the pattern “X means Y” with regular expressions. These terms appear in the data in the “Definitions” section of each FTA, typically in the first part of each one of the agreements. Therefore, they are internally validated as terms. They were also validated externally as terms by consulting specialist dictionaries of the field of economics, finance and international trade. Later, an automatic extraction of a list of candidate terms for the different subsets of the corpus in English and Spanish was carried out using Termostat (Drouin 2003). This list is being manually checked to discard non-terms. Later, the candidate specialized collocations that include the terms from this list will be extracted semi-automatically.

The occurrence of the candidate collocations that are being extracted from the Spanish and English data will be checked against reference material apart from the FTA corpus, using software tools and scripts whenever it is possible to automatically detect whether the specialized collocations occur in the reference material. This material is constituted by general

and specialized corpora and dictionaries, almost all available in electronic format. Seven specialist dictionaries of economics and five reference corpora (three of them annotated) will be used.

The following are the specialist dictionaries: First, the SICE-OAS online Dictionary of Trade Terms⁷, with 397 FTA terms. Second, the Dictionary of International Business Terms (Capela and Hartman 2000). Third, the Routledge Spanish Dictionary of Business, Commerce and Finance (Routledge 1998) which has 37,000 entries. Fourth, the Diccionario de comercio internacional: importación y exportación: inglés-español, Spanish-English (Alcaraz and Castro 2007). Last, the International Monetary Fund (IMF) Terminology⁸. In addition, two dictionaries will be used as monolingual resources: the Routledge Dictionary of Economics, Second edition (Rutherford 2002) and the Dictionary of International Business Terms (Shim, Siegel & Levine 1998).

Regarding the reference general corpora, in the first place the Corpus of Contemporary American English⁹ (Davies 2009) will be used, with 450 million words as of February 2013. To verify whether the Spanish candidate specialized collocations occur in external data, two corpora will be used: Corpus del Español with 100 million words¹⁰ (Davies 2002) and Corpus de Referencia del Español Actual (CREA-RAE)¹¹ with 200 million words. The CREA corpus is the biggest Spanish corpus but is not annotated with Part-of-Speech tags. Last, a reference specialized corpus will be used: the English and Spanish subcorpus of the Economics section of IULA's Technical Corpus. This corpus contains 1.6 million words and was developed at the Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, Barcelona, Spain (Bach, Saurí, Vivaldi & Cabré 1997)¹².

5. Conclusions

Building a small specialized parallel corpus is challenging and involves much work but is a rewarding task. Parallel corpora are valuable language resources to comparatively study different features of two languages in context. In the case of specialized texts, by using a corpus it is possible to derive terminology and its associated phraseology in a systematic way as compared to doing the same task manually.

The extraction and description of specialized collocations found in the FTA corpus can help to enrich term bases and specialized dictionaries from the field of international trade with terminological and phraseological information. It can also prove useful for the improvement of other language resources such as computational lexicons for machine translation and other tasks of natural language processing. These phraseology and terminology aware language resources are necessary tools for language professionals such as translators, interpreters, terminologists, technical writers and lexicographers.

Acknowledgements

This research has received support from the EU 7th Framework Program under a Marie Curie Initial Training Network, project CLARA, Common Language Resources and their Applications, under grant agreement no. 238405. I want to express my deep gratitude to Knut Hofland for his valuable assistance in the preparation of a parallel corpus. I also want to thank my PhD supervisors and other researchers for their valuable insights: Gisle Andersen and Marita Kristiansen from NHH and Koenraad De Smedt from the University of Bergen.

⁷ http://www.sice.oas.org/Dictionary/GT_e.asp

⁸ <http://www.imf.org/external/np/term/eng/>

⁹ <http://www.americancorpus.org/>

¹⁰ <http://www.corpusdelespanol.org/>

¹¹ <http://corpus.rae.es/>

¹² <http://bwananet.iula.upf.edu/>

References

- Aijmer, K. 2008. Parallel and comparable corpora. In A. Lüdeling, A. and M. Kytö (Eds.), *Corpus Linguistics. An International Handbook* (pp. 275–292). Berlin / New York: Walter de Gruyter.
- Bach, C., Saurí, R., Vivaldi, J. & Cabré, M. T. 1997. *El Corpus de l'IULA: descripció. Papers de l'IULA. Sèrie Informes, 17*. Barcelona: IULA, Universitat Pompeu Fabra.
- Bartsch, S. 2004. *Structural and functional properties of collocations in English: a corpus study of lexical and pragmatic constraints on lexical co-occurrence*. Tübingen: Gunter Narr Verlag.
- Benson, M. 1985. Collocations and idioms. In: R. Ilson, (ed.) *Dictionaries, lexicography and language learning* (pp. 61-68). Oxford: Pergamon Press.
- Budin, G. 1990. Terminological analysis of LSP phraseology. In Terminology Science and Research. Journal of the International Institute for Terminological Research (IITF), vol. 1 (1-2): 64-69.
- Cabré, M. T. 1998. *Terminology: Theory, methods and applications*. John Benjamins. Amsterdam / Philadelphia.
- Christ, O. 1994. A modular and flexible architecture for an integrated corpus query system. In Proceedings of COMPLEX'94. 3rd Conference on Computational Lexicography and Text Research. Budapest, Hungary, July 7-10, 1994, 23-32.
- Cop, M. 1991. Collocations in the bilingual dictionary. In Hausmann, F. J., Reichmann, O., Wiegand, H. E. & Zgusta, L., (Eds.), *Wörterbücher: ein internationales Handbuch zur Lexikographie*, volume 3, (pp. 2775-2778). Berlin / New York: Walter de Gruyter.
- Davies, M. 2002. Un corpus anotado de 100.000.000 palabras del español histórico y moderno. In *Actas del Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural*. Valladolid, Spain: SEPLN. 21-27.
- Davies, M. 2009. The 385+ Million Word Corpus of Contemporary American English (1990-2008+): Design, Architecture, and Linguistic Insights. In *International Journal of Corpus Linguistics* (14): 159-190.
- Drouin, P. 2003. Term extraction using non-technical corpora as a point of leverage. *Terminology*, 9(1): 99–115.
- Firth, J. R. 1957. *Papers in Linguistics 1934-1951*. London: Oxford University Press.
- Galinski, C. 1990. Terminology and phraseology. Terminology Science and Research. En Journal of the International Institute for Terminological Research (IITF), vol. 1 (1-2): 70-86.
- Gozdz-Roszkowski, S. 2011. *Patterns of Linguistic Variation in American Legal English. A Corpus-Based Study*. Lodz Studies in Language. Bern: Peter Lang.
- Hofland, K. & Johansson, S. 1998. The Translation Corpus Aligner: A program for automatic alignment of parallel texts. In S. Johansson and S. Oksefjell (Eds.) *Corpora and Cross-linguistic research. Theory, Method, and Case Studies* (pp. 87-100). Amsterdam/Atlanta: Rodopi.

- Kjær, A. 2007. Phrasemes in legal texts. In H. Burger, D. Dobrovolskij, P. Kühn & N. R. Norrick (Eds.) *Phraseology: An International Handbook of Contemporary Research, Volume 1* (pp. 506–516). Berlin / New York: Walter de Gruyter.
- Kjær, A. L. 1990. Phraseology research. State-of-the-art. In Terminology Science and Research. Journal of the International Institute for Terminological Research (IITF), vol. 1 (1-2): 3-20.
- Koester, A. 2010. Building small specialised corpora. In A. O'Keefe and M. McCarthy (Eds.) *The Routledge Handbook of Corpus Linguistics*. (pp. 66-79). London: Routledge.
- L'Homme, M. C. 2006. The processing of terms in dictionaries: New models and techniques. a state of the art. In Terminology, 12(2):181–188.
- L'Homme, M. C. 1998. Caractérisation des combinaisons lexicales spécialisées par rapport aux collocations de langue générale. In Proceedings of EURALEX, Liège, Belgium. 513-522.
- L'Homme, M. C. 2003. Les combinaisons lexicales spécialisées (CLS). Description lexicographique et intégration aux banques de terminologie. In F. Grossman & A. Tutin (Eds.) *Les collocations analyse et traitement*, (pp. 89-104). Amsterdam: De Werelt.
- L'Homme, M. C. 2009. A methodology for describing collocations in a specialised dictionary. In S. Nielsen and S. Tarp (Eds.) *Lexicography in the 21st century*. (pp. 237-256). Amsterdam/Philadelphia: John Benjamins.
- L'Homme, M.C. & Bertrand, C. 2000. Specialized Lexical Combinations: Should they be Described as Collocations or in Terms of Selectional Restrictions. In Proceedings of Ninth Euralex International Congress, Stuttgart, Germany, Stuttgart University. 497-506.
- L'Homme, M.C. (2001. Combinaisons lexicales spécialisées: Regroupement des mots clés par classes conceptuelles. In B. Daille, B. & G. Williams (Eds.) Journées d'étude de l'ATALA. La collocation. Rapport de recherche, Nantes: Institut de recherche en informatique de Nantes. 19-22.
- Lorente, M. 2002a. Altres elements lèxics. In J. Solà, *Gramàtica del català contemporani (Gcc)*. Volum I (pp. 831-888). Barcelona: Empúries.
- Lorente, M. 2002b. Terminología y fraseología especializada: del léxico a la sintaxis. In G. Guerrero, & L. F. Pérez Lagos, L. F. (Eds.), *Panorama actual de la terminología* (pp. 159-180). Granada: Comares, Colección Interlingua, 30.
- Moon, B. E. 2000a. Free trade area (FTA). In R. J. Barry Jones (Ed.), *Routledge Encyclopedia of International Political Economy* (pp. 574-575). London: Routledge.
- Moon, B. E. 2000b. Trade agreements. In R. J. Barry Jones (Ed.), *Routledge Encyclopedia of International Political Economy*. (pp. 1570-1571). London: Routledge.
- Moon, R. 2008. Dictionaries and collocation. In S. Granger and F. Meunier (Eds.) *Phraseology: An Interdisciplinary Perspective* (pp. 313-336). Amsterdam / Philadelphia: John Benjamins.
- Nicolaidis, P. 2000. Free trade area (FTA). In R. J. Barry Jones (Ed.), *Routledge Encyclopedia of International Political Econom.* (pp. 575-576). London: Routledge.
- Orliac, B. 2004. Automatisation du repérage et de l'encodage des collocations en langue de spécialité. PhD dissertation. Montreal: University of Montreal.

- Pavel, S. 1993. La phraséologie en langue de spécialité. Méthodologie de consignation dans les vocabulaires terminologiques. In *Terminologies Nouvelles*, 10. Brussels: RINT, (pp. 67-82).
- Pearson, J. 1998. *Terms in Context*. Amsterdam/Philadelphia: John Benjamins.
- Picht, H. 1990. LSP phraseology from the terminological point of view. In *Terminology science & research: Journal of International Institute for Terminology Research*, (1-2):33-48. Vienna: International Network for Terminology.
- Šarčević, S. 2000. Creativity in legal translation: how much is too much? In A. Chesterman; N. Gallardo & Y. Gambier, (Eds.) *Translation in context: selected contributions from the EST Congress, Granada, 1998* (pp. 281-292). Amsterdam / Philadelphia: John Benjamins.
- Schmid, H. 1994: Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of International Conference on New Methods in Language Processing*. Manchester, UK.
- Spang-Hanssen, H. 1983. Kommunikation og fagsprog [Communication and LSP]. In *Språk i Norden*. Oslo: Cappelen, 25-37.

Dictionaries

- Alcaraz, E. & Castro, J. 2007. *Diccionario de comercio internacional: importación y exportación: inglés-español, Spanish-English*. Barcelona: Ariel.
- Capela, J. & Hartman, S. 2000. *Dictionary of International Business Terms*. 2nd Ed. Hauppauge, United States: Barron's Educational Series.
- Routledge 1998. *Spanish Dictionary of Business, Commerce and Finance*. CD-ROM. London/New York: Routledge Software.
- Rutherford, D. 2002. *Routledge Dictionary of Economics*. London/New York: Routledge.
- Shim, J., Siegel, J. & Levine, M. 1998. *The Dictionary of International Business Terms*. Chicago: Glenlake Publishing Company.

