

A taxonomic study of economics and business genre colonies. A corpus-based analysis of metadiscourse markers in English, French and Spanish

Daniel Gallego-Hernández & Patricia Rodríguez-Inés

Universidad de Alicante (Spain) & Universitat Autònoma de Barcelona (Spain)

daniel.gallego@ua.es, patricia.rodriguez@uab.es

Abstract

The COMENEGO (Corpus Multilingüe de Economía y Negocios [Multilingual Economics and Business Corpus]) project aims to compile a corpus that will be freely available online and useful for LSP trainers, translators, trainee translators and translator trainers. The project's stages include designing a virtual platform for access to the corpus's texts, and compiling and analysing a pilot corpus to evaluate the proposed text categories, which will be used as a means of filtering searches in the corpus once it is online. Many text classifications are based on prior experience or other classifications and are never empirically validated. While the classification proposed here draws on previous proposals and observation of the professional and academic arenas, we have also looked to validate it empirically, and it is that validation with which this paper is concerned. We have used corpus linguistics tools to analyse metadiscourse in three subcorpora (English, French and Spanish), taking Hyland's (2005) work as a basis in the case of English, to determine if our proposed categories have internal linguistic characteristics that support or confirm their taxonomic validity.

Keywords: metadiscourse markers, taxonomic validity, economics, contrastive analysis, corpus.

Resumen

Estudio taxonómico de colonias de géneros de economía y negocios. Análisis basado en corpus de marcadores metadiscursivos en inglés, francés y español

El proyecto COMENEGO (Corpus Multilingüe de Economía y Negocios) trata de compilar un corpus especializado en los ámbitos mencionados, de libre acceso a través de Internet y de utilidad para, entre otros, profesores de lenguas para fines específicos, traductores en formación, formadores y traductores. Tal proyecto pasa por varias etapas entre las cuales se encuentra el diseño de una plataforma virtual que dé acceso a los textos a través de Internet, y la compilación y análisis de un corpus piloto que permita valorar la clasificación de los textos en distintas categorías. En muchos casos las clasificaciones de textos no se validan empíricamente y nacen de experiencia previa u otras clasificaciones. En este caso la clasificación propuesta también se basa en propuestas anteriores y la observación del mundo profesional y académico, pero se ha querido aportar una validación empírica que la sustente. El objetivo del presente trabajo tiene que ver con el análisis de las categorías textuales del corpus, cuya clasificación servirá, por ejemplo, para ayudar a filtrar búsquedas, una vez el corpus se encuentre disponible en Internet. En concreto, analizamos tres subcorpus (inglés, francés y español) desde el punto de vista del metadiscurso tomando como referencia el trabajo de Hyland (2005) para el inglés, con el propósito de responder a la pregunta de si estas categorías tienen rasgos lingüísticos internos que apoyan o confirman su validez taxonómica. Para ello, utilizamos herramientas de la lingüística de corpus.

Palabras clave: marcadores metadiscursivos, validez taxonómica, economía, análisis contrastivo, corpus.

1. Introduction

Translators and LSP writers of economics, financial and business texts can use the Internet to access a wealth of information to solve writing and translation problems and difficulties. Thanks to technology, there are various strategies for making the most of sources of such information. In addition to terminographic resources, such as dictionaries, glossaries and terminological databases, the Internet offers translators and LSP writers parallel texts, i.e. texts comparable, in terms of function, subject matter and/or communicative situation, to that which they are translating or writing. Here we will focus on the development of a resource that can meet translators' terminographic needs.

Translators can work with parallel texts in different ways. They can use their intuition or experience to locate texts online and read the segments they deem useful; use the Internet as if it were a corpus (web-as-corpus approach); compile texts on their computer and study them by means of

corpus analysis applications (web-for-corpus approach); or consult pre-compiled online corpora (stable corpora).

The web-as-corpus approach involves using search engines (e.g. Google) as if they were concordancers. Specific research in which this approach is applied to the practice of economic translation is scarce (Gallego-Hernández, 2012a). The methodology requires the establishment of a series of parameters related to original and parallel texts, as well as the strategic use of search engine functions and operators. Rather than downloading texts, translators consult them online on the basis of the descriptions and results search engines provide.

There are various studies that discuss the web-for-corpus approach application to translation (Bernardini & Zanettin, 2000; Zanettin, Bernardini & Stewart, 2003; Sánchez Gijón, 2004; Beeby, Rodríguez-Inés & Sánchez Gijón, 2009; Rodríguez-Inés, 2009), and even some looking at its application to economic translation (Gallego-Hernández, 2012a; Krüger, 2012; Barceló & Delgado, 2014). Designing and compiling a corpus in this way essentially entails swiftly gathering together a number of texts whose terminological, phraseological and conceptual information allows a translator to solve a particular translation's problems and difficulties. Generally speaking, compilation involves locating and downloading texts, making them compatible with concordancing software and, if necessary, formatting them.

Stable corpora are corpora of a relatively fixed size and design which, in general, have been pre-compiled and prepared for online exploitation, and most thus have an interface for retrieving information. Some specialised economics corpora exist, such as Pompeu Fabra University's IULA Technical Corpus, which is multilingual (Spanish, English and Catalan) and freely available (Cabré & Martorell, 2004: 174). Vigo University's CLUVI Corpus includes EGAL, an economics subcorpus containing 0.4 million words in Spanish, and CONSUMER, a subcorpus on consumption containing 1.8 million words in Spanish. The MLCC Multilingual and Parallel Corpora, created in 2005 by the European Languages Resources Association (ELRA), feature a generic subcorpus of financial articles from newspapers in various languages. Other corpora containing economics or business texts include the Cobuild Business Corpus (initially created for lexicographical purposes), the Wolverhampton Business English Corpus (with texts from 1999 to 2000), Mike Nelson's Business English Lexis Site, the Business

Letter Corpus, and the Cambridge and Nottingham Spoken Business English Corpus (CANBEC).

Working with parallel texts has advantages for economic translation and its teaching, but also certain drawbacks. The web-as-corpus methodology entails using a tool for a purpose for which it is not intended. Consequently, the content obtained (which might not correspond to the original text), users (who might be incapable of making the most of the Internet) and search engines (whose functions translators might find limited, depending on their needs) can all be obstacles to benefiting fully from online texts. The web-for-corpus approach, meanwhile, requires various extra skills and an investment of time that could put translators off. In this case too, the content obtained depends on the scope conventional search engines offer for retrieving texts. Lastly, existing stable corpora do not seem to be adapted to translators' needs. Some are obsolete and tend to be predominantly in English (see all the aforementioned corpora); their range of text genres¹ is rather limited (e.g. CONSUMER, MLCC), bearing in mind the variety of texts economic translation encompasses; and their interfaces are not designed with translators' needs in mind (e.g. CANBEC, Wolverhampton Business English Corpus).

Against that backdrop, we are working on the design of a virtual platform geared specifically to the exploitation of multilingual corpora by translators of economics texts. Filtering searches is one of the functions the platform must have, for which purpose text categories are necessary. Several authors have written about taxonomies of economics texts (e.g. Fernández Antolín & López Arroyo, 2008; Pizarro Sánchez, 2009; Socorro-Trujillo, 2010; Herrero Rodes & Román Mínguez, 2015; Álvarez García, 2017). It can be said, in general, that business text genres are numerous and are affected by terminological variation, that researchers have failed to agree on their classification, and that many of their texts are of a hybrid nature and can thus also be associated with other areas, such as law or advertising.

With regard to the platform's filter system, if the diversity of business-related text genres means presenting users with long lists of names of genres or text types (some of which they may know by other names), filtering by genre or text type could prove largely ineffective, hindering corpus exploitation. A solution could be to group genres² or text types together in categories, although, here too, there is a lack of consensus on how to do so.

With the above in mind, we have compiled a pilot corpus in three languages and categorised its texts according to criteria described further below. The objective of this article is to study whether our proposed categories can be considered valid insofar that there are intra-linguistic differences between them. To that end, we performed discourse analysis in the hope of identifying internal linguistic features supporting or confirming our categories' taxonomic validity, with the ultimate aim of determining whether, from a discursive perspective, our current category structure is justified or ought to be changed in any way.

We will now describe our theoretical framework, which revolves around the concept of “metadiscourse”. We believe studying metadiscourse on the basis of corpora to be a suitable way to undertake an initial text analysis because, as will be shown later, doing so helps establish the texts' metadiscourse profile and, thus, the differences and similarities between text genres and between languages. We will then explain our corpus analysis methodology (pilot corpus description and exploitation), which is fundamentally based on corpus linguistics tools, before presenting our results, in the form of metadiscourse profiles, and discussing the most significant differences and similarities between categories and between languages.

2. Theoretical framework

The concept of “metadiscourse” was introduced by Vande Kopple (1985) and Crismore, Markkanen & Steffensen (1993). Hyland (2005: 37) defines it as “the cover term for the self-reflective expressions used to negotiate interactional meanings in a text, assisting the writer (or speaker) to express a viewpoint and engage with readers as members of a particular community”.

Functionally, metadiscourse helps a writer establish links with readers (through entertainment, persuasion, dissuasion, etc.). Socially, metadiscourse may vary according to the purposes different communities have when communicating.

Hyland (2005: 48-54) identifies two main categories of metadiscourse, namely interactive resources, which are used “to organize propositional information in ways that a projected target audience is likely to find coherent and convincing”; and interactional resources, which “involve readers and open opportunities for them to contribute to the discourse by alerting them to the author's perspective towards both propositional information and

readers themselves”. He distinguishes five categories of interactive resources (see appendix for examples):

- i. Transition markers, which are primarily logical connectors, such as conjunctions and adverbial phrases, for helping readers interpret pragmatic connections between propositions. They indicate additive, contrastive and consequential steps in discourse.
- ii. Frame markers, which signal text boundaries and elements of schematic text structure.
- iii. Endophoric markers, which refer to other parts of the same text.
- iv. Evidentials, which refer to ideas from external sources.
- v. Code glosses, which help rephrase, explain or elaborate on what has been said, to ensure readers understand.

Likewise, he distinguishes five categories of interactional resources:

- i. Hedges, which help emphasise a position’s subjective nature by allowing information to be presented as opinion rather than fact.
- ii. Boosters, which, in contrast to hedges, help close down alternatives or head off conflicting views.
- iii. Attitude markers, which convey the writer’s affective attitude to propositions, such as surprise, agreement, obligation or frustration.
- iv. Self mentions, which are quantifiable on the basis of the frequency of first-person pronouns, possessive adjectives and/or verbs.
- v. Engagement markers, which help the writer address readers directly to focus their attention or include them as participants in the discourse.

Hyland (2005) studies metadiscourse and its markers from an academic viewpoint, which has proved insufficient for some corpora and for areas of specialisation such as tourism, business and journalism. He actually highlights (2005: 87) that one of the concept’s main features is its dependence on context, and it is thus closely related to the standards and expectations of certain text genres. That explains why more study has

apparently been conducted on the interactional dimension, an essential part of Halliday's (1978) "register", than on the interactive dimension, particularly in non-academic texts. Non-academic texts' metadiscourse (especially interactional) has seemingly been studied more than that of academic texts, with examples including journalistic genres of opinion (Dafouz-Milne, 2008), web pages promoting tourism (Suau-Jiménez, 2006) and letters from CEOs (Gallego-Hernández, 2012b). The studies in question show just how important context is for establishing markers that effectively describe a specialised discourse or text.

There have also been studies of the metadiscourse of some genres of economics texts. Valero-Garcés (1996), for example, compares two texts in English written by non-native speakers with two written by native speakers, focusing on four subtypes of metatext (connectors; reviews or earlier markers; previews or later markers; and action markers). Her results show that Spanish-speaking writers use less metatext and favour a more impersonal style than Anglo-American writers (1996), suggesting that choice of rhetorical strategies depends not only on the individual author but also on their culture and the context of the target culture. Moreno-Fernández (1998) analyses hedges and boosters in a bilingual corpus of economics and business research articles in Spanish and English written by native speakers of the respective languages. According to her results, Spanish texts contain fewer hedges than English texts. In another study of the use of boosters, in texts written in English by non-native speakers in this case, Carrió-Pastor and Calderón (2015) analyse a corpus of 100 emails composed by Spanish and Chinese employees of an export company. Their results reveal that the two groups are similar in terms of their frequent use of certain markers (*know*, *confirm*), but differ in that the Spaniards are more assertive when communicating in English (use of *must*, for instance). Pizarro-Sánchez and Bravo-Gozalo (2006) study the translation and integration of tables in a parallel corpus (English-Spanish) and a comparable corpus (English) of corporate annual reports. The main difference observed in the way tables are integrated into the texts is that all the translations and the Spanish originals analysed use verbal integrating elements (e.g. *a continuación, en el siguiente, following, as follows*), while less than half of the English originals analysed do, leading the authors to state that by presenting the information more explicitly, "translators do not respect the conventional use of this rhetoric element in the TL polysystem" (2006: 170). Mur-Dueñas (2010) studies metadiscourse in a comparable corpus of research articles on business

management written in US English and Peninsular Spanish. She focuses on attitude markers (in the form of verbs, adjectives, nouns, adverbs and phrases) and compares aspects such as frequency of use, tendencies and the values expressed. Her results indicate that attitude markers are used very frequently in both subcorpora (attributable to the interest of the authors of such articles in presenting and promoting their findings) and that there are certain similar tendencies in both languages, suggesting that the authors, whose sociocultural contexts are different, share disciplinary values.

Looking beyond the results of any particular study, this brief literature review indicates that the presence of a given metadiscourse resource can help establish what we refer to in this article as *metadiscourse profiles*. The metadiscourse profiles of text genres contextualised in didactic communication situations, for example, could be expected to reflect a high presence of interactive markers, especially code glosses, used by senders to teach or transmit specialised knowledge to an audience possibly unfamiliar with the subject being dealt with. It is our understanding that metadiscourse analysis is not limited to a single text genre but can extend to groups of genres, i.e., in our case, the different categories into which our pilot corpus's text types/genres are distributed. We believe that studying metadiscourse profiles to identify differences in those of each of COMENEGO's text categories could contribute to determining the categories' taxonomic validity..

3. Methodology

3.1. Pilot corpus

The pilot corpus's texts were chosen primarily on the basis of their potential usefulness for economic translation training and practice (Gallego-Hernández & Krishnamurthy, 2013; Rodríguez-Inés, 2014). Table 1 gives examples of the genres and genre colonies involved and shows how they are categorised at present.

GENRES and GENRE COLONIES	CATEGORIES
descriptions of bank products, financial products and insurance; corporate web pages (commercial websites)	COMMERCIAL (COM)
online courses; guides for consumers, investors and bank clients (web pages of teachers, universities, institutions and companies)	DIDACTIC (DID)
laws, codes, decrees (websites of ministries and agencies)	LEGAL (LEG)
articles of association, regulations, minutes of annual meetings, rules (corporate and informative websites)	ORGANISATIONAL (ORG)
press releases, news, newsletters (corporate websites and newspapers)	PRESS (PRS)
academic papers (general-interest and specialised websites, specialised journals)	SCIENTIFIC (SCI)
financial prospectuses, annual accounts, annual reports, financial results, corporate responsibility reports, management reports, analyses, country- and sector-specific reports, marketing plans, quarterly results (corporate and informative websites)	TECHNICAL (TEC)

Table 1. Examples of texts as distributed in COMENEGO.

Our “technical” (TEC) and “scientific” (SCI) categories correspond to the functional types of communication in organisations Cassany (2004: 53-55) identified on the basis of pragmatic and discursive criteria (function, interlocutors, structure and style). The two categories’ discourse has a chiefly referential function, is intended to transmit information objectively and arises in an organisation’s technical activities (projects, auditing, research). Our “organisational” (ORG) category draws on Cassany’s conception of organisational discourse, which can have a conative, referential or metalingual function, is aimed at systematising and regulating an organisation’s activity, and tends to arise in specific areas of work (general management, human resource management, administration, appraisal, quality control). However, whereas Cassany’s organisational category includes what he calls *administrative and legal languages*, we have established a separate “legal” (LEG) category. While closely related to the organisational category, it comprises texts intended to regulate the activity not of a particular organisation but of all the organisations of one or more countries. We have also included a “commercial” (COM) category based on Cassany’s classification. Such discourse, the function of which lies between conative and referential, seeks to influence the reader’s opinion and behaviour, and generally occurs in specific activities (marketing, advertising, communication, sales).

We have established another two categories, “didactic” (DID) and “press” (PRS), mainly on the basis of pragmatic criteria. DID contains texts designed to educate, in either academic contexts (e.g. courses, notes) or professional contexts (e.g. consumer guides, explanations about the stock exchange for investors). PRS comprises organisations’ press releases, as well as informative texts on economic, financial or business matters published by the general or specialised press.

The pilot corpus has a French, a Spanish and an English subcorpus. The Spanish and French subcorpora were compiled on the basis of the

discourse-centred text classification described above, and their texts chosen using intuition, experience and previous works on economics, business and financial text taxonomies (AUTHOR 1). An effort was made to make these two corpora balanced in terms of number of tokens. The English subcorpus was compiled two years later and is smaller. It has the same categories as the other subcorpora, but was developed following more objective general and specific criteria (Rodríguez-Inés, 2014).

The primary general criterion applied when selecting texts in English was statistical. Tolosa-Igualada (2014) provides a range of data concerning economic translation from Spanish to English and vice versa, obtained in 2013 through a survey of professional translators. The English subcorpus includes texts from almost all the survey's main genres, such as sales contracts, licensing agreements, correspondence, letters, annual accounts and financial statements, deeds of sale, powers of attorney, memorandums and articles of association, shareholders' agreements, delivery notes, minutes of board meetings and general meetings of shareholders, audits of accounts, terms and conditions of contract, service contracts, employment contracts, advertising leaflets and brochures, reports, legal advice with economic and financial content, press articles and opinion pieces, company registration certification, certificates of employment, product descriptions and payment documents. The only document type we were unable to include was internal communication, due to the lack of availability of specimens. We also took into account the subjects survey respondents mentioned most frequently (Tolosa-Igualada, 2014: 16), notably including intrinsically economic activities, banking and insurance, all of which are well represented in COMENEGO's English subcorpus.

Our second general criterion was organisations' consent to add their texts to COMENEGO's virtual platform (all the consenting organisations are listed at <http://dti.ua.es/es/comenego/agradecimientos.html>).

Another general criterion was ease of access to texts, in terms of it being possible not only to locate them but also to download them and convert them to plain text. We looked for texts available via the public Internet (by searching for keywords in Google) which we could download directly (e.g. using HTTrack) as web pages or PDF files and then convert to plain text (using HTML2txt converter and Ultra Document Text Converter).

It is worth pointing out that all the English subcorpus's texts are complete texts rather than samples, although some are templates or models, especially

in the case of genres such as business letters, contracts, delivery notes, payment documents and application forms.

Lastly, we targeted diversity where sources are concerned. We collected texts from anglophone and non-anglophone countries and emerging and developed economies throughout the world, written by native and non-native speakers alike. They include both originals and translations, and correspond to businesses and bodies of different sizes and natures.

In addition to the general criteria described above, we made use of external quality criteria for certain categories. In the case of scientific texts, we decided to select articles on subjects prominent in the aforementioned survey from specialised journals included in either the Social Sciences Citation Index (e.g. *The Geneva Papers on Risk and Insurance; International Review of Finance; Journal of Corporate Finance; Emerging Markets Review; Accounting, Auditing & Accountability Journal*) or the Scopus database (e.g. *Journal of Internet Banking and Commerce; International Journal of Business Science and Applied Management*).

We also applied a specific criterion to select texts from the press. We took them either from well known or reputed regional, national or international dailies that have an online version and specialise in economics or similar matters, or from specialised sections in generalist dailies. The English subcorpus thus includes articles and opinion pieces from sources as varied as *The New York Times, Financial Times, The Economist, The Guardian, The Wall Street Journal, The Washington Post and The Telegraph*.

Table 2 shows the pilot corpus's content in terms of the finite number of files and tokens in each category and language.

LANG	ITEM	COM	DID	ORG	TEC	SCI	LEG	PRS	TOTAL
FR	FILES	3 909	1 121	634	133	203	21	2 859	8 880
	TOKENS	1 325 544	1 304 585	1 365 468	1 187 806	1 301 102	1 293 704	1 308 418	9 086 627
	MEAN	339	1 164	2 154	8 931	6 409	61 605	458	1 023
ES	FILES	5 255	1 491	429	351	99	211	2 214	10 050
	TOKENS	1 329 915	1 276 089	1 337 822	1 188 068	1 311 731	1 342 698	1 329 029	9 115 352
	MEAN	253	856	3 118	3 385	13 250	6 363	600	907
EN	FILES	247	253	172	308	67	12	535	1 594
	TOKENS	614 112	622 749	633 381	646 585	632 644	668 350	666 194	4 484 015
	MEAN	2 486	2 461	3 682	2 099	9 442	55 696	1 245	2 813

Table 2. Number of files and tokens in the COMENEGO pilot corpus by category and language.

3.2. Corpus queries

Despite showing the usefulness of interactional metadiscourse in economic contexts, the studies mentioned in the theoretical framework analyse specific economics and business genres in specific languages, and we were thus unable to use their results as a basis for validating our pilot corpus's categories. Nonetheless, their methodologies, particularly those of Gallego-Hernández (2013) and Suau-Jiménez (2014), who have already proposed a methodological approach to analysing the pilot corpus based on interactional markers, are used in this article.

Corpus linguistics has generated various analytical tools (e.g. WordSmith Tools and AntConc) capable of extracting data in the form of frequent word lists, concordances, collocates and n-grams. We used AntConc's concordance function to establish each subcorpus's metadiscourse profile, extracting concordances based on pre-established lists of keywords potentially representative of each category of interactive and interactional metadiscourse markers (see appendix).

Having found no list of French or Spanish keywords potentially representative of any particular category of metadiscourse in the aforementioned previous studies, we translated the list of words in English originally presented by Hyland (2005: 218-224). The words used to extract concordances from each subcorpus are included in the appendix to this article.

Through AntConc, we interrogated the corpus using regular expressions (all of which are set out in the appendix). We used the symbol | to search for all the words representative of a given category in a single query; the symbols * and + to allow for inflection, especially in Spanish and French³; and the symbol @ to allow for lexical variation (e.g. *dicho de otro modo* vs. *dicho en otras palabras*).

After extracting the concordances of each metadiscourse resource in each language, we read through every single concordance to verify that its keyword actually constituted metadiscourse and was not simply noise. Our aim was to count the number of times each resource appears in each category and subcorpus so as to establish COMENEGO's metadiscourse profile.

1 s, téléchargez-le gratuitement en cliquant sur l'image suivante : Plan d'intervention **EXCEPTIONS** COM
 2 s, téléchargez-le gratuitement en cliquant sur l'image suivante : Plan d'intervention **EXCEPTIONS** T COM
 3 **SAINT** ne se préoccupe pas des incidences sur l'image de cet appareil et de ses utilisateurs. De **LONG** DID
 4 Profil **RAK** : Appel de marge Cliquez sur l'image pour agrandir Le Broker évalue le risque et **RE** DID
 5 il s'agirait d'un système d'écriture basée sur l'image. Toujours est-il que ces codex ont pu être **déjà** DID
 6 **ça** voient les européens, 2003, PUF. Livre sur l'image des produits européens en Chine (voir extrait en DID
 7 06 février 2009 Sur le thème " 2009, arrêté sur l'image ", l'image sous toutes ses formes est à l'**honneur** PRS
 8 Pour télécharger le pictogramme, cliquez sur l'image Contacts Pernod Ricard Francisco de la **VEGA** / PRS
 9 pouvant avoir une incidence significative sur l'image donnée par les comptes annuels, selon les **plans** TEC
 10 l'égard d'un confrère ou susceptible de ternir l'image de la profession. Ils s'efforcent de résoudre à **LEG**

Figure 1. Concordances of image.

For illustrative purposes, Figure 1 shows the concordances of the keyword image extracted from the French subcorpus to study endophoric markers. Concordances 1, 2, 4 and 8 all correspond to metadiscourse (e.g. *Cliquez sur l'image*, “click on the image”). However, *image* is not used in reference to another part of the same text in the remaining concordances, which we thus omitted from our quantitative analysis.

4. Results and discussion

Table 3 contains the results we obtained by applying the aforementioned methodology to each metadiscourse resource and each of the corpus's

LANG	CAT	INTERACTIVE						INTERACTIONAL				
		transition	frame	endophoric	evidential	gloss	booster	hedge	attitude	self mention	engagement	
EN	COM	Items	1888	735	1139	47	814	2239	5529	999	5526	7380
		%	7.2	2.8	4.3	0.2	3.1	8.5	21.1	3.5	21.1	28.2
	DID	Items	3396	1076	1701	73	1610	4532	6866	1806	2079	7345
		%	11.1	4.1	6.5	0.3	6.1	17.3	26.2	6.9	7.9	28.0
	ORG	Items	1356	15052	9837	52	317	3199	6475	816	63	1
		%	3.6	57.4	37.5	0.2	1.2	12.2	24.7	3.1	0.2	0.0
	TEC	Items	1500	2800	2360	68	366	2306	4961	801	2815	2547
		%	7.3	10.7	9.1	0.3	1.4	6.7	18.9	3.1	10.7	9.7
	SCI	Items	3677	446	374	288	770	2476	5623	1414	5097	1247
		%	17.2	1.7	1.4	1.1	2.9	9.4	21.5	5.4	19.4	4.8
	LEG	Items	4744	1038	2428	221	1745	3184	5780	1365	4525	156
		%	18.8	4.0	9.3	0.8	6.7	12.1	22.1	5.2	17.3	0.6
PRS	Items	1547	750	1693	100	567	1568	2954	946	5441	284	
	%	9.8	2.9	6.5	0.4	2.2	6.0	11.3	3.6	20.8	1.1	
FR	COM	Items	3231	328	378	15	776	1900	1010	5562	5922	44616
		%	5.1	1.3	0.7	0.1	3.0	7.3	3.9	21.2	22.6	170.3
	DID	Items	5550	777	1864	197	2766	2260	1918	5848	1750	1488
		%	22.7	3.0	7.1	0.8	10.6	8.6	7.3	22.3	6.7	5.7
	ORG	Items	4270	341	578	6	714	365	1369	3350	931	876
		%	32.8	1.3	2.2	0.0	2.7	2.2	5.2	12.8	3.6	3.3
	TEC	Items	3964	439	1099	223	898	1309	1954	4230	4068	613
		%	21.1	1.7	4.2	0.9	3.4	5.0	7.5	16.1	15.5	2.3
	SCI	Items	8301	1626	3156	5888	2174	2485	2997	5647	5769	391
		%	21.4	6.2	12.0	22.5	8.3	9.5	11.4	21.5	22.0	1.5
	LEG	Items	3053	12788	4389	0	318	269	392	1657	437	2433
		%	11.9	48.8	16.7	0.0	1.2	1.0	1.5	6.3	1.7	9.3
PRS	Items	4484	644	166	27	493	2291	1833	6765	3637	1912	
	%	20.2	2.5	0.6	0.1	1.9	8.7	7.0	25.8	13.9	7.3	
ES	COM	Items	3458	266	399	0	266	3857	4788	6916	8112	11570
		%	8.7	1.0	1.5	0.0	1.0	14.7	18.3	26.4	31.0	44.2
	DID	Items	5615	1148	1787	510	893	4211	6125	5742	8550	3318
		%	14.8	4.4	6.8	1.9	3.4	16.1	23.4	21.9	32.6	12.7
	ORG	Items	3077	401	2141	0	268	1472	2007	5485	1739	1070
		%	17.4	1.5	8.2	0.0	1.0	5.6	7.7	20.9	6.6	4.1
	TEC	Items	4752	1069	1544	119	475	4158	2257	5109	1069	475
		%	22.6	4.1	5.9	0.5	1.8	15.9	8.6	19.5	4.1	1.8
	SCI	Items	7608	1836	3017	2361	787	4985	6427	6952	2492	787
		%	20.4	7.0	11.5	9.0	3.0	19.0	24.5	26.5	9.5	3.0
	LEG	Items	3491	671	5774	0	269	1343	1074	3625	269	806
		%	20.2	2.6	22.0	0.0	1.0	5.1	4.1	13.8	1.0	3.1
PRS	Items	4652	665	133	133	399	4652	1994	6911	3323	930	
	%	19.6	2.5	0.5	0.5	1.5	17.8	7.6	26.4	12.7	3.5	

Table 3. COMENEGO's metadiscourse resources.

languages and text categories. As the three subcorpora are of different sizes, in order to facilitate comparability, we normalised the results by providing not only the number of markers observed for each category but also their percentage frequency distributions in each category.

To check whether the distribution of the values observed for each marker differs from category to category, we used the chi-square test of homogeneity for each language:

- English: $\chi^2 (54) = 86937.8, p < 0.001, V = 0.286$
- French: $\chi^2 (54) = 206981.1, p < 0.001, V = 0.410$
- Spanish: $\chi^2 (54) = 67063.5, p < 0.001, V = 0.240$

The results of the tests are statistically significant, as the p-value is lower than the preestablished significance level ($\alpha = 0.05$). It can therefore be concluded that the markers are distributed in a non-homogeneous manner in the three languages. That non-homogeneity is greatest in the case of French, given that it is the language with the highest Cramér's V.

The data contained in table 3 are graphically represented in figures 2-4, which show the metadiscourse profiles identified for English, French and Spanish respectively. COMENEGO's text categories are shown on each profile's vertical axis, while its horizontal axis shows the frequency of use of each metadiscourse resource, i.e. the number of markers observed for each category.

We will now describe and discuss the metadiscourse profiles identified for each language in each text category.

4.1. Metadiscourse profile in English

Figure 2 shows the metadiscourse profile of each text category in English. Logically, given that texts in the COM category seek to establish a relationship between sellers and customers and to influence the opinion and behaviour of the latter, it is notable for a high frequency of self mentions (especially the markers *we*, *our* and *us*) and, in particular, engagement markers (*you*). Examples include:

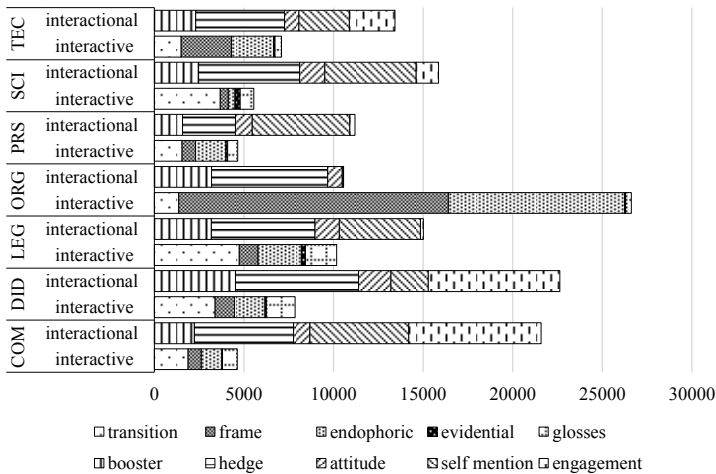


Figure 2. Metadiscourse profile of COMENEGO's English subcorpus.

1. *For 50% of the capital provided, we will pay you 100% of the increase in the CPI*
2. *Doing up your home? Buying a car? We can offer you a Personal Loan tailored to meet your needs.*

Another notable aspect of this category is a high presence of hedges, especially *may*, *should* and *would*. Examples include:

3. *You should consult your tax adviser regarding specific questions*
4. *The tax position above may change at any time which may affect*
5. *You must advise us in writing if you would prefer us to pay you interest...*

The DID category stands out in two respects where interactive markers are concerned. The first is a high use of code glosses (especially *such as*, *for example* and *called*), confirming the explanatory nature of the texts in this category. Examples include:

6. *This can suggest a few things such as less repeat calls, better product stability...*
7. *A share Sicav can specialise, for example, by geographical region or sector*
8. *This is called a loss leader.*

The second is a high frequency (second only to SCI) of transition markers (*but, however, because*). Examples include:

9. *Most trusts will specialise in one of these, but some deliberately have a balanced strategy*
10. *With higher profits comes a higher tax bite, however, the higher income looks better to prospective investors and lenders*
11. *It is called preferred because the dividend must be paid before dividends are paid on the common stock.*

In relation to interactional markers, DID is the category that makes the greatest use of boosters (*important, even, usual*) and hedges (*may, should, would*). Examples include:

12. *The most important brokered securities markets are...*
13. *The usual formula for calculating the working capital...*
14. *As your company grows, you should consider contracting out as many tasks as possible*
15. *This type of borrowing may offer tax advantages.*

Like COM, DID features a considerable number of engagement markers, mainly the keyword *you*. Interaction between authors and readers is necessary in both categories, owing to the importance of the reader understanding the content in the case of DID texts, and to the importance of the customer being persuaded to do something in the case of COM texts.

The ORG category differs from COM and DID in that it has more frame markers (particularly in the form of numbered lists) and endophoric markers (chiefly the keywords *section, above* and *below*, with examples including:

16. *Paragraph (a) above applies to...*
17. *the following terms have the meanings specified below*
18. *procedures established under Section 5.4 and Section 5.5 hereof.*

ORG is the category with the third highest presence of engagement markers, and one of those with the fewest code glosses (along with LEG and TEC) and self mentions (along with DID and LEG). This may be due to the fact

that in ORG texts the author is irrelevant but the content is partly geared to regulating the addressee's behaviour.

While the TEC and SCI categories may seem similar *a priori*, they differ in some respects. For instance, SCI has more code glosses (*such as, for example, e.g.*), evidentials (*according to, cited*) and transition markers (*but, however, because*). Examples include:

19. *According to ISO 17021...*
20. *Recently, a number of studies, e.g. Bae and Goyal (2010) and Mitton and O'Connor (2012), have...*

This is logical, as scientific language requires connectors and explanations to articulate its discourse, and needs to refer to other works or texts. SCI has more interactional markers than TEC, except in the case of self mentions. Both categories contain very few engagement markers.

LEG is, by far, the category with the most frame markers (particularly in the form of numbered lists and keywords such as *section* and *chapter*) and endophoric markers. This is attributable to its texts (laws, regulations, etc.) being of a highly structured nature and, in many cases, long. LEG features virtually no self mentions, as its texts are impersonal.

The PRS category has certain distinguishing characteristics, including a low frequency of interactive markers other than transition markers. With regard to interactional markers, it does not seem to differ substantially from other categories. Engagement markers are its most common interactional resource, followed by hedges (*would, may, could, should*) and self mentions (*our, we*). Most of the keywords corresponding to self mentions appear in quotes, however, and were therefore disregarded in our analysis to avoid any misinterpretation of results.

4.2. Metadiscourse profile in French

Figure 3 shows the metadiscourse profile of each text category in French. The COM category stands out due to its large number of engagement markers (*vous* and *notre* in particular, and, to a lesser degree, interrogatives and directives), which, together with self mentions (*nous, nos, notre*) and attitude markers (especially exclamations, adjectives such as *grand, bon, simple, mieux* and *important*, and the expression *grâce à*), reinforce the idea of discourse in which senders (sellers) establish direct contact with receivers (customers) to

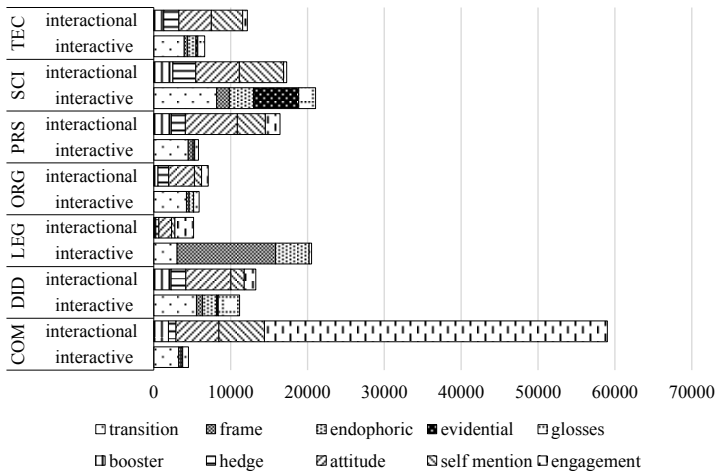


Figure 3. Metadiscourse profile of COMENEGO's French subcorpus.

influence their behaviour (persuade them to make a purchase). Examples include:

21. *Une grande souplesse de remboursement grâce au prêt Facilimmo*
22. *Devis d'assurance: assurez-vous à bon prix!*
23. *Les numéros importants pour votre tranquillité.*

The DID category is notable for a high presence of all types of transition markers (contrastive: *or, cependant, alors que*; additive: *de plus, également, par ailleurs*; and consequential: *afin*) and attitude markers (*grand, important, il faut, bon, simple*). As might be expected, it also has many code glosses (especially *par exemple, c'est-à-dire, appelé, considérer comme* and *autrement dit*). Examples include:

24. *Prenons par exemple l'hypothèse la plus simple*
25. *De plus, afin d'éviter le blocage de la succession, tout héritier ou créancier pourra...*

DID is actually the category with the most code glosses, ahead of even SCI. ORG does not differ greatly from the other categories in terms of a particular metadiscourse resource's presence. Along with LEG, it appears to

be the category with the fewest metadiscourse markers, other than transition markers (especially additive markers: *également, en outre, par ailleurs, de plus*; and consequential markers: *afin, en vue de, en conséquence*), which feature in every category, and attitude markers (particularly adjectives with nuances of positivity, simplicity and importance, such as *bon, grand, simple, important, approprié, mieux* and *significatif*). Even so, it has fewer attitude markers than any category other than LEG, confirming its regulatory nature.

The TEC and SCI categories' profiles are similar but differ in terms of the various markers' proportions. With the exception of engagement markers, which are present to practically the same degree in both categories, the frequency of every interactional and interactive resource is much higher in SCI. Significantly, SCI has more evidentials (mainly due to the use of bibliographic reference systems) and hedges (especially forms of the verb *pouvoir*; conditional forms of other verbs, such as *serai**, *devrai** and *aurai**; verbs such as *sembler* and *paraître*; and the adjective *possible*). The presence of signs of intertextuality, established through allusion to prior work and studies, and a tendency to qualify statements confirm the special nature of SCI, where academic language is prominent.

The LEG category is notable, as it was in English, for its high number of endophoric markers, transition markers (especially additive markers, such as *également, en outre, de plus* and *d'autre part*; the contrastive marker *toutefois*; and consequential markers, such as *en vue de, afin* and *dès lors*) and, in particular, frame markers (chiefly the keywords *article, chapitre* and *section*, which appear in the titles of each text's *articles, chapters* and *sections*, as well as numbered lists). LEG is also characterised by a very low presence of other metadiscourse resources, some of which (e.g. evidentials) are almost totally absent.

Lastly, the PRS category is noteworthy for its high number of attitude markers (especially qualifiers with positive nuances, such as *grand, important, bon, meilleur, significatif* and *exceptionnel*, as well as the expression *grâce à*), transition markers (particularly the additive markers *également, de plus* and *par ailleurs*) and, in the case of press releases, self mentions (*nous, notre, nos*). It is second to SCI as the category with the most boosters (particularly adverbs such as *très, totalement, entièrement, parfaitement* and *pleinement*; the verb *démontrer*; and the adjective *véritable*). Examples include:

26. *Plusieurs mois un véritable engouement, jouant pleinement son rôle de valeur refuge...*

27. *Il démontre également la qualité du groupe, et de son équipe*
28. *La réaction très forte des investisseurs reflète clairement la confiance que les marchés accordent au Groupe.*

4.3. Metadiscourse profile in Spanish

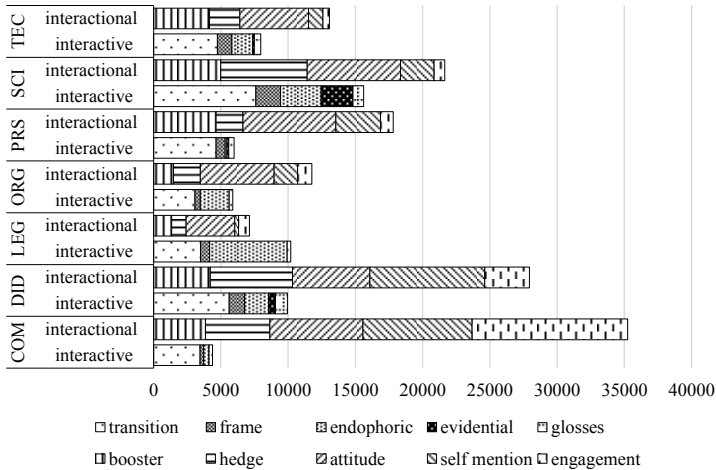


Figure 4. Metadiscourse profile of COMENEGO's Spanish subcorpus.

Figure 4 shows the metadiscourse profile of each text category in Spanish. In addition to being the category with the most engagement markers (the pronouns *tú* and *usted*, as well as interrogatives), COM is characterised by a high presence of self mentions (first-person plural verb forms and the pronouns and possessives *nosotros*, *nuestro* and *nos*) and attitude markers (mainly adjectives and adverbs with positive nuances, such as *mejor*, *principal*, *gran*, *especial*, *importante* and *buen*). This helps confirm that its texts involve senders (sellers) trying to persuade receivers (customers) to make a purchase, as suggested previously in relation to the other subcorpora. Examples include:

29. *Tenemos una amplia gama de hipotecas para usted. Encuentre la que mejor se adapta a sus necesidades*
30. *Gran profesionalidad: formarás parte de un equipo innovador, exigente y con talento.*

Like COM, the DID category has many engagement markers (especially interrogatives, as well as imperatives and pronouns such as *usted* and, to a lesser degree, *tú*), self mentions (mainly first-person plural verb forms) and attitude markers (adjectives and adverbs with positive nuances, such as *gran*, *mejor*, *principal*, *importante* and *buen*), possibly to aid learning by establishing direct contact between senders and receivers. DID differs from the other categories, including COM, in that, as it was in English and French, it is the category with the most code glosses (the participles *llamado* and *denominado*, expressions such as *es decir* and *por ejemplo*, and the verb *significar*). It also features many transition markers, especially contrastive markers (such as *aunque*, *sino*, *sin embargo*, *mientras que*, *no obstante*, *por el contrario* and *a pesar de*) and consequential markers (*por tanto*, *por lo que*, *por ello*, *por lo tanto*). Examples of such expressions include:

31. *Sin embargo, no significa decir que el método convencional tenga perdido...*
32. *La diferencia más importante está en que lo que habitualmente se ha denominado en Europa y, más concretamente, en Gran Bretaña.*

The ORG category's main trait is that attitude markers (expressions such as *de calidad*, *mejor* and *adecuado*) are its most common metadiscourse resource, followed by transition markers (*asimismo*, *además*, *no obstante*, *por tanto*, *igualmente*) and endophoric markers (keywords such as *apartado*, *capítulo* and *párrafo*). Evidentials, code glosses, self mentions and engagement and frame markers are scarce in this category.

The TEC and SCI categories are characterised by a high frequency of transition markers (the most common being *además*, *aunque*, *sin embargo* and *por tanto* in the case of SCI, and *además*, *aunque*, *sin embargo*, *por lo que* and *asimismo* in the case of TEC) and attitude markers (adjectives with positive nuances, the most common in both categories being *gran*, *mejor*, *importante*, *principal* and *significativo*). In the case of attitude markers, Dueñas (2010) found similar frequencies in American Business Management and local Spanish research articles (SCI). TEC and SCI are the categories with the most frame markers (in the case of SCI, expressions related to sequence, such as *finalmente*, *en primer lugar*, *por último* and *a continuación*; in the case of TEC, expressions related to sequence, such as *a continuación*, *por último* and *finalmente*, and others used to change topic, such as *en cuanto a*, *respecto de* and *con respecto de*), as might be expected given that their texts are very long and thus require such resources to structure their discourse. TEC and SCI differ

from one another in that the latter contains many more hedges (mainly conditional verb forms, the verbs *poder* and *parecer*, and the adjective *posible*) and code glosses (the verbs *denominar*, *significar* and *llamar*, and expressions such as *en particular*, *particularmente* and *de la siguiente forma*). This difference is logical, bearing in mind that the resources in question are widely used in academic language (corresponding to SCI).

As in English and French, the LEG category stands out due to its large number of frame markers (particularly the keyword *artículo*, used in the title of each of the articles into which texts are organised), endophoric markers (keywords such as *apartado*, *capítulo*, *artículo*, *párrafo* and *parte*) and transition markers (mostly additive markers, such as *asimismo*, *además* and *igualmente*, but also contrastive markers, such as *no obstante*, *aunque*, *sino* and *sin embargo*, and consequential markers, such as *por tanto*, *por lo que* and *en consecuencia*). It has few evidentials, code glosses or self mentions.

Finally, the PRS category is characterised by a high presence of attitude markers (again, adjectives with positive nuances, such as *mejor*, *gran*, *principal* and *importante*), transition markers (additive markers, such as *además*, *asimismo* and *igualmente*; and contrastive markers, such as *aunque*, *mientras que* and *sino*) and boosters (words or expressions such as *muy*, *demostrar*, *sobre todo*, *pleno* and *claramente*), as was the case in French.

5. Conclusion

This paper presents an initial analysis aimed at confirming or rejecting our pilot corpus's current text categories. Our object of study to that end has been metadiscourse markers.

The results of our analysis indicate differences and similarities between the corpus's categories. Each category seems to have at least one metadiscursive trait that distinguishes it from the rest. The COM category, for example, is mainly marked by interactional metadiscourse, particularly engagement markers and self mentions, owing to the direct contact established between senders (sellers) and receivers (customers). Attitude markers are also highly present in COM, reflecting its characteristic conative function.

DID and SCI differ from the other categories on the basis of interactive metadiscourse, specifically a high frequency of code glosses, which their texts use to fulfil their communicative goal (educating about aspects of

economics) through explanations, examples, etc. They also differ from each other on the basis of interactive resources, in this case evidentials, which are more common in SCI (which features academic language with many references to previous work). There are differences between the two in terms of interactional resources too, particularly self mentions and engagement markers, with SCI seemingly having a greater tendency towards impersonality.

ORG and LEG are the categories with the least marked interactional profiles in Spanish and French. In English, in contrast, both categories are marked by hedges and boosters (the main difference between the two lies in self mentions and engagement markers, which are practically absent in LEG). A tendency towards impersonality seems to be a distinguishing trait of LEG in particular (a category that includes essentially normative texts, with more interactive than interactional resources, potentially calling the current trend of basing analyses exclusively on interactional metadiscourse into question) in all three languages. In terms of interactive metadiscourse, ORG and LEG (particularly the latter) are chiefly characterised by the use of frame, transition and endophoric markers, which distinguishes them from the other categories and is attributable to their very long texts being highly structured (information organised into articles and titles; internal references; etc.).

The TEC category is very similar to SCI in all three languages as far as interactional resources are concerned. The two differ with regard to interactive metadiscourse, however, with TEC containing far fewer resources, especially evidentials, which it lacks almost entirely.

PRS, finally, is similar to other categories (e.g. COM) in terms of interactive resources, of which it has very few besides transition markers. In the case of interactional metadiscourse, however, its profiles appear to be unique, with little resemblance to those of other categories in any of the languages.

This paper presents merely an initial analysis of the COMENEGO pilot corpus, one that can be complemented with studies of other objects (terminology, phraseology, metaphor, etc.) for the purpose of validating or rejecting the corpus's current text categories.

With regard to the present configuration of our virtual platform for working with the corpus texts, the results of our analysis indicate that some of the categories, such as COM and PRS, have distinctive traits and may thus be retained. A decision is yet to be made as to whether having a single type of defining resource (e.g. code glosses in the case of DID and evidentials in that

of SCI) is sufficient to justify a category's inclusion among the platform's search options, or whether the most similar categories (e.g. TEC and SCI, or SCI and DID) should be merged. In the face of this dilemma, it seems appropriate to keep platform users informed of the text types or genres each category contains, and to enable them to filter searches not only by category but also by text genre. In any case, as stated previously, this analysis is by no means definitive. Further in-depth studies of the characteristics of the corpus's texts ought to be carried out, bearing in mind the hybrid, permeable nature they can have.

Article history:

Received 29 August 2018

Received in revised form 03 August 2020

Accepted 10 May 2021

References

- Álvarez García, C. (2017). Los textos en el ámbito del comercio exterior: una taxonomía para la formación de traductores. *Sendebarr*, 28, 113–133.
- Barceló, T., & Delgado, I. (2015). Las metáforas en los textos periodísticos franceses sobre la bolsa: estudio a partir de un corpus electrónico. In D. Gallego-Hernández (Ed.), *Current approaches to business and institutional translation* (pp. 105–118). Peter Lang.
- Bhatia, V. K. (2004). *Worlds of Written Discourse*. Continuum.
- Beeby, A., P. Rodríguez-Inés, & P. Sánchez Gijón (Eds.) (2009). *Corpus use and translating*. John Benjamins.
- Bernardini, S., & F. Zanettin (Eds.) (2000). *I corpora nella didattica della traduzione: Corpus use and learning to translate*. CLUEB.
- Cabré, M. T., & Martorell, C. (2004). El corpus tècnic del IULA: corpus textual especializado plurilingüe. *Panacea*, 16, 173–176.
- Carrió-Pastor, M. L., & Calderón, R. M. (2015). A contrastive analysis of metadiscourse features in business e-mails written by non-native speakers of English. *Procedia*, 173, 214–221.
- Cassany, D. (2004). Explorando los discursos de las organizaciones. In A. van Hooff Comajuncosas (Ed.), *Textos y discursos de especialidad* (pp. 49–70). Rodopi.
- Crismore, A., Markkanen, R., & Steffensen, M. (1993). Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students. *Written Communication*, 10(1), 39–71.
- Dafouz-Milne, E. (2008). The pragmatic role of textual and interpersonal metadiscourse markers in the construction and attainment of persuasion: A cross-linguistic study of newspaper discourse. *Journal of Pragmatics*, 40, 95–113.
- Dueñas, P. M. (2010). Attitude markers in Business Management research articles: A cross-cultural corpus-driven approach. *International Journal of Applied Linguistics*, 20, 50–72.
- Fernández Antolín, M., & López Arroyo, B. (2008). La indisolubilidad del lenguaje jurídico-económico. In M. Calvi, G. Mapelli, & J. Santos López (Eds.), *Lingue, Culture, Economia. Comunicazione e pratiche discorsive* (pp. 111–119). Franco Angeli.
- Gallego-Hernández, D. (2012a). *Traducción económica y corpus: del concepto a la concordancia. Aplicación al francés y al español*. Servicio de Publicaciones de la Universidad de Alicante.
- Gallego-Hernández, D. (2012b). Metadiscourse y traducción en el lenguaje de los negocios: estudio basado en corpus (francés-español). *RAEL: Revista Electrónica de Lingüística Aplicada*, 11, 13–24.
- Gallego-Hernández, D. (2013). COMENEGO (Corpus Multilingüe de Economía y Negocios): A metadiscursive analysis approach. *Procedia - Social and Behavioral Sciences* 95C: 146–153.
- Gallego-Hernández, D., & Krishnamurthy, R. (2013). COMENEGO (Corpus Multilingüe de

- Economía y Negocios): Design, creation and applications. *ELR Journal*, 8.
- García Izquierdo, I. (2011). New perspectives for the analysis and formalisation of specialised genres: the GENTT proposal. *Terminàlia*, 3, 13–21. <<http://revistes.iec.cat/index.php/Terminàlia/article/view/45249>>
- Halliday, M. A. K. (1978). *Language as social semiotics*. Edward Arnold.
- Herrero Rodes, L., & Román Mínguez, V. (2015). English to Spanish translation of the economics and finance genres. *INTRAlinea Special Issue: New Insights into Specialised Translation*.
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Continuum.
- Krüger, R. (2012). Working with corpora in the translation classroom. *Studies in Second Language Learning and Teaching*, 4, 505–525.
- Moreno Fernández, A.I. (1998). The explicit signalling of premise-conclusion sequences in research articles: A contrastive framework. *Text*, 18(4), 545–585.
- Mur-Deñás, P. (2010). Attitude markers in business management research articles: a cross-cultural corpus-driven approach. *International Journal of Applied Linguistics*, 20(1), 50–72.
- Pizarro Sánchez, I. (2009). La comunicación escrita en la empresa: criterios para una taxonomía. In C. Pérez-Llantada & M. Watson (Eds.), *Languages for business: A global approach* (pp. 149–160). Prensas Universitarias Universidad de Zaragoza.
- Pizarro Sánchez, I. & Bravo Gozalo, J. (2006): Contrastive analysis of non-verbal elements in the discourse of Economics in English and in Spanish. In J. Bravo Gozalo (Ed.), *Aspects of translation* (pp. 145-172). Secretariado de Publicaciones e Intercambio Editorial, Universidad de Valladolid.
- Rodríguez-Inés, P. (2009). Evaluating the process and not just the product when using corpora in translator education. In A. Beeby, P. Rodríguez-Inés & P. Sánchez-Gijón (Eds.), *Corpus use and translating: Corpus use for learning to translate and learning corpus use to translate* (pp. 129-149). John Benjamins.
- Rodríguez-Inés, P. (2014). COMENEGO: compilación del corpus piloto en inglés y primeros análisis. *Monográficos de la Revista Hermēneus*, 16, 187–199.
- Sánchez Gijón, P. (2004). *L'ús de corpus en la traducció especialitzada: compilació de corpus ad hoc i extracció de recursos terminològics*. UPF.
- Sinclair, J. (1996). Preliminary Recommendations on corpus typology. URL: <http://www.ilc.cnr.it/EAGLES/corpusyp/corpusyp.html> [05/08/18].
- Socorro Trujillo, K. (2010): A text typology for the international trade: A pedagogical tool in the translation classes for this specific purpose. In J. Cifuentes Honrubia, A. Gómez González-Jover, A. Lillo Buades, J. Mateo Martínez & F. Yus Ramos (Eds.), *Los caminos de la lengua. Estudios en homenaje a Enrique Alcaraz Varó* (pp. 1479–1490). Universidad de Alicante.
- Suau-Jiménez, F. (2006, July). El metadiscurso en el género servicios y productos turísticos en inglés y español: importancia de su traducción como recurso para la persuasión del cliente. Paper presented at the *I Congreso Internacional de Traducción Especializada*, Buenos Aires, Argentina.
- Suau-Jiménez, F. (2014). Preparación pre-traductológica de textos de economía a través del análisis metadiscursivo interpersonal: estudio piloto del corpus COMENEGO. *Monográficos de la Revista Hermēneus*, 16, 149–166.
- Tolosa-Igualada, M. (2014). Dime qué traduces y “les” diré quién eres. Estudio basado en encuestas acerca de los documentos traducidos por traductores económicos (inglés-español y español-inglés). *Monográficos de la Revista Hermēneus*, 16, 23–41.
- Valero-Garcés, C. (1996). Contrastive ESP rhetoric: Metatext in Spanish–English economics texts. *English for Specific Purposes*, 15(4), 279–294.
- Vande Kopple, W. J. (1985). Some exploratory discourse on metadiscourse. *College Composition and Communication*, 36, 82–93.
- Zanettin, F., S. Bernardini, & D. Stewart (Eds.) (2003). *Corpora in translator education*. St. Jerome.

Daniel Gallego-Hernández is a Senior Lecturer in the Faculty of Arts of the University of Alicante in Spain. His research interests include corpus linguistics applied to translation, business translation and translation teaching. He teaches business translation between French and Spanish as

well as terminology at his faculty. His individual publications include a book entitled Business translation and corpora.

Patricia Rodríguez-Inés is a Senior Lecturer at the Department of Translation, Interpreting and East Asian Studies of the Universitat Autònoma de Barcelona in Spain. Her research interests include corpus linguistics applied to translation, translation competence and its acquisition, and translator education. She has authored and co-authored papers published in top ranking journals within Translation Studies.

NOTES

¹ “[genre] is a collective product that results from each particular circumstance of communication. Any form of conventionalised and culturally determined text, regardless of the field (specialised or not) in which the communication takes place, can therefore be considered a genre. Nevertheless, the notion is especially significant in the fields of specialised communication” (García Izquierdo, 2011: 14).

² “[Genre colony] A grouping of closely related genres ‘serving broadly similar communicative purposes, but not necessarily all the communicative purposes in cases where they serve more than one’” (Bhatia, 2004: 59).

³ For example, to locate engagement markers in Spanish, we used searches designed to cover future and imperative form endings: ?*ese|?*ate|no ?*e|no ?*es|*ás [futuro] |? [interrogación] |señoras|señores|te|tenga|permita|recuerde|deje|busque|compre|encuentre|haga|tú|tu|tus|usted++

Appendix

Hyland’s metadiscourse items in English (2005: 218-224). French and Spanish versions are own translations.

Endophoric markers:

article+|au+dess++s|cf.|chapitre+|diagramme+|encadré|exemple+|figure+|graphique+|image+|infra|lignes|p.|page+|paragraphe+|partic+|plus bas|plus haut|schéma+|section+|supra|table+|tableau+|v.|voir

Frame markers:

à ce point|à ce stade|à commencer|à l’égard|actuellement|alors|au bout du compte|autrement|avant tout|bref|brèvement|but de|ce qui m’intéresse|ce qui nous|ce qui nous intéresse|cela conduit à|cinquième lieu|cinquièmement|comme indiqué précédemment|comment+ons|concentr+ sur|conclusion|d’abord|d’ailleurs|dans ce travail|dans cette étude|dans cette leçon|dans le présent document|dans un sens général|d’autre part|dernier lieu|désir+|deuxième lieu|deuxièmement|d’un côté|d’une part|en ce qui concerne|en conclusion|en fin de compte|en résumé|en revanche|en somme|en troisième lieu|en un mot|enfin|ensuite|généralement|grossièrement|intention de|intention est|je voudrais|jusqu’alors|l’intention de|l’objet|maintenant|nous voudrions|objectif est|par ailleurs|par rapport|pour ce qui est|pour conclure|pour finir|pour le moment|pour terminer|précédemment|premier|premièrement|prétendons|principalement|puis|quant à|quatrième lieu|quatrièmement|récapitulatif|résumant|revenons|second lieu|septième lieu|septièmement|sixième lieu|sixièmement|souhait+|synthés+|tenons à|thème|troisièmement|un autre aspect|quà|voudr+

Transition markers:

à cause d+|à force d+|à tel point qu+|afin d+|afin qu+|ainsi|ajoutons qu+|alors|alors qu+|au contraire|au lieu d+|au lieu qu+|au point d+|au point qu+|aussi|aussi bien|bien qu+|ce n’est pas qu+|ce+ a pour effet|ce+ empêche|ce+ pousse à|ce+ provoque|ce+ s’oppose à|cependant|certes|c’est la raison pour laquelle|c’est pourquoi|conséquent|contrairement à|d’ailleurs|d’autre part|de crainte|de façon à|de façon qu+|de là|de la même manière|de manière à|de manière qu+|de peur d+|de peur qu+|de plus|de@sorte qu+|dès lors|done|d’où|du moins|du reste|d’un autre côté|également|en conséquence|en dépit d+|en outre|en plus|en revanche|en sorte d+|en sorte qu+|en sus d+|en vue d+|encore qu+|ensuite|faute d+|il est exact q+|il se peut qu+|inversement|loin d+|mais|malgré|même si|même s’il+|néanmoins|non seulement|or|outre|par ailleurs|par conséquent|par contre|par manque d+|par suite|par suite|pour|pour que|pour sa part|pourtant|puis|quant à|s’ajouter|sans compter q+|sans doute|seulement|si quand|sinon|sous prétexte d+|tandis qu+|tant de +nom que|tant et|tant que|tellement|toutefois

Evidentials:

(???)|, (???)|, (???)|, (???)|, (???)|comme en témoign+|comme l+ conseil+|comme l+ indiqu+|comme l+ not+|comme l+ suggèr+|comme le conseil+|comme le not+|comme le suggèr+|comme l’indiqu+|comme mentionné+

Code glosses:

à savoir | à titre d'exemple | appelé* | autrement dit | c.-à-d. | ce qui démontre qu* | ce qui explique qu* | ce qui illustre le fait qu* | ce qui montre qu* | ceci démontre qu* | ceci explique qu* | ceci illustre le fait qu* | ceci montre qu* | ceci prouve qu* | ceci prouve qu* | cela démontre qu* | cela explique qu* | cela illustre le fait qu* | cela montre qu* | cela prouve qu* | c'est à dire | c'est-à-dire | concrètement | connu* comme | connu* sous | consid* comme | de fait | défin* comme | désign* par | d'une manière concrète | effectivement | en d'autres termes | en effet | en fait | en réalité | en somme | en un mot | ex: | ne signifie pas qu* | notamment | on entend par | p. ex. | par exemple | par le fait | plus simplement | porter la mention | précisément | prenons le cas | qualifié* d* | que l'on appelle | qu'on appelle | si l'on prend le cas | signifie qu* | spécialement | spécifiquement | suffit de rappeler qu* | tel est le cas | tel* qu* | un autre exemple | veut dire qu* | voilà en outre

Figure 5. Interactive resources (French keywords).*Hedge:*

aurai* | certain* | devrai* | donnerai* | environ | paraît* | peut | pouvons | peuvent | pouvez | peut-être | possible* | pourrai* | semble* | serai*

Boosters:

bel et bien | bien sûr | clairement | démontre* | effectivement | entièrement | évidemment | extrêmement | incontestablement | parfaitement | pleinement | plus que jamais | sans aucun doute | surtout | totalement | très | véritable* | véritablement | vrai+ +

Attitude markers:

adéquat+ + | ambitieux+ + + | ambition | approprié+ + + | attrayant+ + + | beau+ | - belle+ | bon+ + + | chaleureusement | chaleureux+ + + | confiance | confiant+ + + | considérable+ | convaincu+ + + | conviction | craint* | de qualité | efficace | essentiel+ + + | excellent+ + + | exceptionnel+ + + | excessi+ + + | exclamation | extraordinaire+ | fantastique+ | fier+ d* | formidable+ | grâce à | grâce au+ | grand+ + | grave+ | heureux+ + + | heureusement | il faut | important+ + + | inapproprié+ + | ingrat+ + | innovateur+ + + | irréprochable* | malheureusement | mieux | meilleur+ + + | nécessairement | nous intéress* | intéress* | m'intéret* | optimal | optimiste+ | optimisme+ | paradoxalement | persuadé+ + | précieusement | préférable | principale+ | prometteu+ + + | raisonnable+ | rapide | redout* | regrettois | regrettable | remarquable* | remarquablement | remerci* | évé* | rigoureux* | satisfaisant+ + | significatif+ + | significativement | simple | souci | soucieux+ + + | souhaitable+ | souhaitons | spectaculaire+ | substantiel+ + + | surprenant+ + + | terrible+ | vivement | voulons

Self mentions:

nos | notre | nôtre | nous

Engagement markers:

*ez [imperativo] | ? [interrogación] | mesdames | votre | vous

Figure 6. Interactional resources (French keywords).*Endophoric markers:*

antérieurement | cf. | cuadro+ | est+ + | apartado+ | est+ + | artículo+ | est+ + | capítulo+ | est+ + | cuadro+ | est+ + | diagrama+ | est+ + | gráfico+ | est+ + | párrafo+ | esta+ | figura+ | esta+ | parte+ | esta+ | secci+ + + | gráfico | l+ + | apartado+ | l+ + | artículo+ | l+ + | capítulo+ | l+ + | cuadro+ | l+ + | diagrama+ | l+ + | ejemplo+ | l+ + | gráfico+ | l+ + | párrafo+ | la figura | la parte | la sección | las figuras | las líneas siguientes | las partes | las próximas líneas | las secciones | las siguientes líneas | más abajo | más arriba | p. | página+ | pie de página | próxima+ | parte+ | próxima+ | secci+ + + | próximo+ | apartado+ | próximo+ | artículo+ | próximo+ | capítulo+ | próximo+ | párrafo+ | siguiente+ | apartado+ | siguiente+ | artículo+ | siguiente+ | capítulo+ | siguiente+ | párrafo+ | siguiente+ | parte+ | siguiente+ | sección+ | table+ | v. | véa+ + se

Frame markers:

a continuación | a modo de conclusión | al margen de | cabe preguntarse si | ánimo de | ánimo es | ante todo | antes de todo | así mismo | asimismo | cabe concluir que | cambiando de tema | centraré en | centraremos en | comencemos con | como breve conclusión | como dijimos al principio | con respecto a+ | con todo y lo anterior | concluyamos | continuaremos | de entrada | de igual forma | de igual manera | de igual modo | de manera global | de momento | de otra parte | de un lado | deseamos | desearía | desearíamos | deseo | después | en conclusión | en conjunto | en cuanto a+ | en cuarto lugar | en definitiva | en el presente trabajo | en esta lección | en esta lección | en este estudio | en este trabajo | en líneas generales | en lo que respecta a | en primer lugar | en quinto lugar | en resumen | en resumidas cuentas | en segundo lugar | en séptimo lugar | en sexto lugar | en síntesis | en suma | en tercer lugar | en último lugar | en un sentido general | esto nos lleva a finalmente | globalmente | hasta ahora | hasta aquí | lo | hasta el momento | hasta este momento | hay otro aspecto que | igualmente | intención de | intención es | llegado a + este | nos conduce | nos lleva | objetivo de | objetivo es | para comenzar | para concluir | para empezar | para terminar | pasemos a | por el momento | por lo demás | por lo que se refiere a+ | por otra parte | por otro lado | por su parte | por último | por una bien | pretendemos | pretendería | pretenderíamos | pretendo | previamente | primeramente | propósito de | propósito es | prosiguiendo con el tema | pues bien | queremos | queremos | querriamos | quiero | recapitulando | recapitulamos | respecto a+ | respecto de+ | resumiendo | si | volvemos atrás | sintetizando | volvamos a | volviera a

Transition markers:

a no ser que | a pesar de | además | ahora bien | antes bien | así mismo | así pues | así que | así y todo | asimismo | aun así | aunque | con todo | contrariamente | de ahí que | de donde se sigue | de ello resulta que | de hecho | de igual forma | de igual manera | de igual modo | de manera que | de modo que | de suerte que | de todas formas | de todas maneras | de todos modos | del mismo modo | después de todo | en cambio | en consecuencia | en cualquier caso | en efecto | en lugar de | en tanto que | en vez de | excepto si | igualmente | mientras que | no obstante | pese a | por consiguiente | por dicha causa | por dicha razón | por dicho motivo | por el contrario | por ello | por ende | por esa causa | por esa razón | por ese motivo | por lo que | por lo tanto | por tal causa | por tal motivo | por tal razón | por tanto | sea como sea | sin embargo | sino | tal como | tal y como | y con eso

Evidentials:

(????) | (???) | (???) | (???) | (???) | como aconseja+ | como advierte+ | como apunta+ | como constata+ | como demuestra+ | como menciona+ | como muestra+ | como observa+ | como recomienda+ | como señala+ | como sugiere+

Code glosses:

a saber | calificad* de | como el | como l+ + | concretamente | conocid+ + + | como | de hecho | de la siguiente manera | forma | modo | definid+ + + | como | denominad+ + | dicho con otras palabras | dicho de otra forma | dicho de otra manera | dicho de otro modo | efectivamente | en concreto | en cuatro palabras | en dos palabras | en otras palabras | en particular | en pocas palabras | en realidad | entendid+ + | como | es decir | específicamente | esto es | esto quiere decir | esto significa | llamad+ + + | lo que quiere decir | lo que quiere significar | mejor dicho | o sea | p.ej. | particularmente | por ejemplo | queremos decir | quiero decir | realmente | significa | sirva de ejemplo | verdaderamente

Figure 7. Interactive resources (Spanish keywords).*Hedge:*

daría | debería* | en torno a | habría | habrían | parec* | podría* | posible+ | posiblemente | puede | pueden | puedes | puedo | quizá+ | serí* | supondría | tendría

Boosters:

absolutamente | altamente | claramente | completamente | dem+ + + str* | efectivamente | evidentemente | extremadamente | más que nunca | muy | notablemente | de forma notable | de manera notable | perfectamente | plen+ + | plenamente | por supuesto | significativamente | sin lugar a dudas | sin duda | sobre todo | totalmente | verdader+ + + | verdaderamente

Attitude markers:

!|adecuad++|afortunadamente|por fortuna|agradec*|ambici*|apropiad++|atractiv++|bella+|buen++|-conf+*a*|considerable+|conveniend*|convicción|cualificado+|de calidad|desafortunadamente|deseable+|deseamos|eficaz+++|eficazmente|esencial++|especial++|espectacular++|esperamos|excelente+|excesiv++|extraordinari++|fantástic+++|favorablemente|feliz|félices|formidable+|gracias a+|gran++|grave+|hay que|importante+|inapropiad++|ingrat++|innovador++|interesamos|mos interés|nuestro+inter+*+*|mis+|inter+*+*|lament*|mejor+*|necesariamente|óptim++|optimist++|optimismo|orgullo* de+|paradójicamente|precios++|preferible+|preocupa*|pretende-mos|principal++|principalmente|prometedor+*|queremos|rápid++|razonable+|riguros+|satisfactori++|significativ++|simple+|simplemente|so ña*|sucñ*|sorprendente+|sustancial++|teme*|temor|terrible+

Self mentions:

*emos|*amos|*imos|nos|nosotros|nuestr++

Engagement markers:

2ªesc|2ªate|no 2ªe|no 2ªes|*ás [futuro]? [interrogación]? señoras|señores|te|tenga|permita|recuerde|de-je|busque|compre|encuentre|haga|tú|tu|us|usted++

Figure 8. Interactional resources (Spanish keywords).

Endphoric markers:

in chapter *|in part *|in section *|in the * chapter|in the * part|in the * section|in this chapter|in this part|in this section|example|Fig|Figure|P|Page|Table|above|before|below|earlier|later

Frame markers:

chapter|part|section|the * chapter|the * part|the * section|this chapter|this part|this section|finally|first|first of all|firstly|last|lastly|(a)|a)|a. |next|(1)|1)|1.|second|secondly|subsequently|then|third|thirdly|to begin|to start with|all in all|at this point|at this stage|by far|for the moment|in brief|in conclusion|in short|in sum|in summary|now|on the whole|overall|so far|thus far|to conclude|to repeat|to sum up|to summarize|this chapter|this part|this section|aim|desire|to focus|goal|intend|to intention|objective|purpose|seek|to want|to wish|to would like|to|back|to|digress|in regard|to|move on|now|resume|return|to|revisit|shift|to|so|to look more closely|turn|to|well|with regard to

Transition markers:

accordingly|additionally|again|also|alternatively|although|and|as a consequence|as a result|at the same time|because|besides|but|by contrast|by the same token|consequently|conversely|equally|even though|further|furthermore|hence|however|in addition|in contrast|in the same way|leads to|likewise|moreover|nevertheless|nonetheless|on the contrary|on the other hand|rather|result in|similarly|since|so|so as to|still|the result is|thereby|therefore|though|thus|whereas|while|yet

Evidential:

(*)|cite|quote|[*]|according to|cited|quoted

Code glosses:

|(*)|as a matter of fact|called|defined as|e.g. |for example|for instance|I mean|i.e. |in fact|in other words|indeed|known as|namely|or|put another way|say|specifically|such as|that is|that is to say|that means|this means|viz|which means

Figure 9. Interactive resources (English keywords).

Hedges:

about|almost|apparent|apparently|appear|appeared|appears|approximately|argue|argued|argues|around|assume|assumed|broadly|certain amount|certain extent|certain level|claim|claimed|claims|could|couldn't|doubt|doubtful|essentially|estimate|estimated|fairly|feel|feels|felt|frequently|from my perspective|from our perspective|from this perspective|generally|guess|indicated|indicates|in general|in most cases|in most instances|in my opinion|in my view|in this view|in our opinion|in our view|largely|likely|mainly|may|maybe|might|mostly|often|on the whole|ought|perhaps|plausible|plausibly|possible|possibly|postulate|postulated|postulates|presumably|presumably|probable|probably|quite|rather|relatively|roughly|seems|should|sometimes|somewhat|suggest|suggested|suggests|suppose|supposed|supposes|suspect|suspects|tend|tended|to|tends to|to my knowledge|typical|typically|uncertain|uncertainly|unclear|unclearly|unlikely|usually|would|wouldn't|in general terms

Boosters:

actually|always|believe|believed|believes|beyond doubt|certain|certainly|clear|clearly|conclusively|decidedly|-definite|definitely|demonstrate|demonstrated|demonstrates|doubtless|establish|established|evident|evidently|find|finds|found|in fact|incontestable|incontestably|incontrovertible|incontrovertibly|indeed|indisputable|indisputably|know|known|must|never|no doubt|obvious|obviously|of course|prove|proved|proves|realize|realized|realizes|really|show|showed|shown|shows|sure|surely|think|thinks|thought|truly|true|undeniable|undeniably|undisputedly|undoubtedly|without doubt

Attitude markers:

!|admittedly|(we/I) agree|agrees|agreed|amazed|amazing|amazingly|appropriate|-appropriately|astonished|astonishing|astonishingly|correctly|curious|curiously|desirable|disappointing|disappointingly|disagree|disagreed|disagrees|dramatic|dramatically|essential|essentially|even|expected|expectedly|fortunate|fortunately|hopeful|hopefully|important|importantly|inappropriate|-inappropriately|interesting|interestingly|prefer|preferable|preferably|preferred|remarkable|remarkably|shocked|shocking|shockingly|striking|strikingly|surprised|surprising|surprisingly|unbelievable|unbelievably|understandable|understandably|unexpected|unexpectedly|unfortunate|unfortunately|unusual|unusually|usual

Self mentions:

I|we||me|my|our|mine|us|the author|the author's|the writer|the writer's

Engagement markers:

(|?)|reader's|add|allow|analyse|apply|arrange|assess|assume|by the way|calculate|choose|classify|compare|connect|consider|consult|contrast|define|demonstrate|determine|do not|develop|employ|ensure|estimate|evaluate|find|follow|go|have|to|imagine|incidentally|increase|input|insert|integrate|key|let|let us|let's|look at|mark|measure|mount|must|need|to|note|notice|observe|one's|order|ought|our|pay|picture|prepare|recall|recover|refer|regard|remember|remove|review|see|select|set|should|show|suppose|state|take|think about|think of|turn|us|use|we|you|your

Figure 10. Interactional resources (English keywords).