

**UNIVERSIDAD COMPLUTENSE DE MADRID**  
**FACULTAD DE FILOLOGÍA**



**TESIS DOCTORAL**

**Correspondencias de traducción de marcadores verificativos  
en inglés y en español: un estudio basado en corpus**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR**

**PRESENTADA POR**

**M<sup>a</sup> Estefanía Avilés Mariño**

Directora

**Julia Lavid López**

Madrid

**UNIVERSIDAD COMPLUTENSE DE MADRID**

**FACULTAD DE FILOLOGÍA**

**Departamento de Estudios Ingleses**



**TESIS DOCTORAL**

**DOCTORADO EN LINGÜÍSTICA INGLESA**

**CORRESPONDENCIAS DE TRADUCCIÓN DE MARCADORES**

**VERIFICATIVOS EN INGLÉS Y EN ESPAÑOL:**

**un estudio basado en corpus**

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR**

**PRESENTADA POR**

**M<sup>a</sup> Estefanía Avilés Mariño**

Directora

**Julia Lavid López**

**Madrid, 2019**



**UNIVERSIDAD COMPLUTENSE DE MADRID**

**FACULTAD DE FILOLOGÍA**

**Departamento de Estudios Ingleses**

**PHD ENGLISH LINGUISTICS**



**TRANSLATION CORRESPONDENCES OF VERIFICATIVE MARKERS IN**

**ENGLISH AND SPANISH: A CORPUS-BASED STUDY**

**DOCTORAL THESIS**

Submitted by

*Dña. M<sup>a</sup> Estefanía Avilés Mariño*

Thesis advisor

*Professor Dr. Julia Lavid López*



*A mi hermana Laura, ella es toda una inspiración para mí.*



UNIVERSIDAD  
**COMPLUTENSE**  
MADRID

## DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

D./Dña. Mª Estefanía Avilés Mariño,  
estudiante en el Programa de Doctorado Doctorado en lingüística inglesa,  
de la Facultad de Filología de la Universidad Complutense de  
Madrid, como autor/a de la tesis presentada para la obtención del título de Doctor y  
titulada:

Translation correspondences of Verificative Markers in English and Spanish: a corpus-based study.

y dirigida por: Dña, Mª Julia Lavid López

### DECLARO QUE:

La tesis es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, de acuerdo con el ordenamiento jurídico vigente, en particular, la Ley de Propiedad Intelectual (R.D. legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, modificado por la Ley 2/2019, de 1 de marzo, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), en particular, las disposiciones referidas al derecho de cita.

Del mismo modo, asumo frente a la Universidad cualquier responsabilidad que pudiera derivarse de la autoría o falta de originalidad del contenido de la tesis presentada de conformidad con el ordenamiento jurídico vigente.

En Madrid, a 27 de junio de 2019

**MARÍA  
ESTEFANÍA  
AVILÉS  
MARIÑO**  
Fdo.: MARIÑO

Firmado digitalmente  
por MARÍA  
ESTEFANÍA AVILÉS  
MARIÑO  
Fecha: 2019.06.27  
16:46:28 +02'00'

# Acknowledgements

First, I would like to express my deepest gratitude to my thesis advisor PhD Dr. M<sup>a</sup> Julia Lavid López. Her continual support and encouragement over these three years has been essential for the development of my research project and this thesis. I am grateful for all our discussions in each stage of the project, for her generous support, understanding and patience.

Furthermore, Professor Dr. M<sup>a</sup> Julia Lavid has shared with me invaluable knowledge and encouraged me to get familiar with up-to-data applications such as LF Aligner (Farkas, 2016) and Overleaf (Hammersley and Lees, 2014) among others presented in the various workshops and conferences she has had the kindness to introduce me to around Europe and Spain; namely, Toulouse, Paris, Zürich, Salamanca, Cáceres, Santiago de Compostela and Madrid.

Thank you also for enhancing my soft-skills, motivating me to be critical and to improve my work, for orientating my process with rigorousness, accuracy and dedication and for showing disposition and tremendous ability to carry out efficient meetings, even when the time slots were a bit tight, since they had to fit into my working timetables outside the faculty. In particular, I appreciate the chance Professor Dr.M<sup>a</sup> Julia Lavid gave me to participate as doctoral student in the European ISCH COST Action IS1312 Research Network project “Textlink: Structuring Discourse in Multilingual Europe”<sup>1</sup> .

---

<sup>1</sup>See the project web page at: <http://textlink.ii.metu.edu.tr>

My participation in this project has been a milestone in my career and has given me the chance to meet international experts working on corpus analysis and annotation of discourse markers (DMs) in different languages, the focus of my dissertation. Within these outstanding experts I would like to mention PhD Liesbeth Degan from Louvain university.

Also, thanks to Julia's expert advice and joint work, I have been able to co-author and present several research papers at different international events at prestigious Universities (Zürich, Paris, Toulouse, Salamanca, Cáceres, Santiago de Compostela and Madrid), all of which have widened my research perspective and my international contacts as well as boosting my communication skills and providing me with invaluable feedback from expert researchers.

Second, I would also like to thank other expert researchers who were committed or involved in some stages of my research, among them, PhD Dr. Paloma Tejada, the coordinator of the Doctoral Program in English Linguistics, has always been diligently attentive, solving doubts and providing invaluable information regarding all sorts of courses and contests, as well as support on the bureaucratic stage of the thesis, showing closeness, patience and deep understanding, thank you very much.

Thus, to mention two of the most significant doctoral events in which I have participated due to Dr Paloma Tejada's expert advice: the PhD Day and thesis in 3 Minutes (3 Minutes Thesis). These contests gave me necessary skills to support my ideas in a reduced time and in front of students and experts from a variety of fields, therefore, having to make my thoughts comprehensive and concise out of my comfort zone. This could not have been possible without the pioneering vision of the PhD school, I would like to thank PhD Emilio Javier Peral Vega and PhD Dr. Jorge Arús Hita.

I would also like to thank David Pacios, from the Computer Faculty at Complutense University, for his generous and invaluable help with the Latex formatting of this dissertation. Without his help and eagerness to solve all my editing doubts, even providing video tutorials, the final presentation would not have been as professional as I would like it to be and I hope it is now.

Finally, I would strongly like to offer my gratefulness towards my family and friends. Without their motivation, understanding, encouragement and support, this project would not have been possible. Devotedly, they have always trusted my force of will and commitment, and held my hands while facing all the challenges which a PhD student has to overcome during these awesome but hard years and that is priceless.



# Abstract

The literature on discourse markers (henceforth DMs) from a cross-linguistic perspective has been expanding in the last ten years, both from theoretical and applied perspectives. Their study is interesting and challenging not only for their semantic-pragmatic variation but also for their ambiguity or polyfunctionality.

This thesis aims at contributing to this expanding field by focusing on the so-called ‘verificative markers’ (henceforth VMs) by scholars in Systemic-Functional Linguistics (SFL). These VMs have received less attention than other DMs from a contrastive perspective but are very frequent in different discourse genres.

The English VMs studied are ‘indeed’, ‘in fact’ and ‘actually’ and the Spanish ones are ‘realmente’, ‘en realidad’, ‘de hecho’ and ‘la verdad’. The selection of these VMs is motivated by their high frequency in English and Spanish reference corpora across different registers and genres in both languages.

The final goal of this dissertation is to provide a corpus-based translation study of the most frequent VMs in English and Spanish through their translation correspondences, investigating their core meanings, as well as other possible peripheral ones or the pragmatic implicatures which can be derived from them, and constructing contrastive (English-Spanish) semantic fields or translation lexical domains of VMs on the basis of their translation equivalents (Hummel, 1995).

A further aim is also to use corpus analysis to reveal possible genre-specific preferences in the use of VCs in original and translated texts and to extract corpus-based features of these VMs which can be used for large-scale annotation of parallel corpora for computational applications such as Machine Translation (MT) Multilingual Text Generation (MTG).

*The research questions addressed in this study are the following:*

- 1) What are the translation correspondences between English and Spanish VMs as revealed by the analysis of parallel corpora? This is investigated through the analysis of the Spanish translation equivalents of each of the tree VMs in all parallel corpora used in this study.
- 2) What are the translation domains revealed by such correspondences? This is investigated through the quantitative analysis of the frequency distributions of the Spanish translation equivalents (Hummel, 1995) of the English VMs complemented with the frequency distributions of the English back-translations of the Spanish equivalents.
- 3) Are there any genre-specific preferences in the use of VMs in original and translated texts in both languages? This is investigated through the quantitative analysis of the frequency distributions of the English VMs, their Spanish translation equivalents in the different genres of the bilingual parallel corpora.

With respect to the first research question, the corpus analysis revealed that the VM ‘in fact’ is mostly translated into the Spanish marker ‘de hecho’, which coincides with the translation preference for ‘indeed’; concluding that ‘in fact’ and ‘indeed’ are the most central VM’s in English (Halliday, 2014). Whereas, ‘Actually’ has as translation preference for ‘realmente’.

These results for VM’s in English confirm the strong relationship between English VM’s (Halliday, 2014) and Spanish ‘operadores de refuerzo argumentativo’ (Portolés and Zorraquino, 1999). Moreover, ‘actually’ appears to be more peripheral than ‘indeed’ and ‘in fact’.

In the case of the Spanish Markers (Portolés and Zorraquino, 1999) the answer to question one is as follows: ‘de hecho’ has as translation preference ‘indeed’ which represents a unique bilateral coincidence for the two markers; furthermore, ‘en realidad’ prefers as translation equivalent (Hummel, 1995) ‘in fact’, whereas ‘realmente’ prefers ‘truly’ and ‘la verdad’ prefers ‘the truth’.

These two last Spanish ‘operadores de refuerzo argumentativo’ (Portolés and Zorraquino, 1999) are to be the most peripheral among the others due to their translation equivalents differing from the pattern of VM’s suggested for ‘de hecho’ and ‘en realidad’.

The English functions found through translation and back translation of the English VMs are corroboration, particularization, reformulation-identification, reformulation-emphasis and reformulation-explanation. The additional functions found for the Spanish markers are reformulation-modification, reformulation-reassessment, reformulation-rectification, reformulation-explanation, discourse organization, purpose, contrast and emphasis.

Research question three is analysed checking which are the translation preferences for each genre based on the corpora under study and which is the translation genre preferred according to the most frequent translation equivalent (Hummel, 1995) for each marker. In this sense, formal expository and argumentative genera are preferred in both languages, though Spanish markers (Portolés and Zorraquino, 1999) avoid analytical texts and dialogues, whereas English markers (Halliday, 2014) do not hesitate on these genera.

The analysis of results reveals that through parallel corpora analysis in bilateral translation it is possible to encounter additional functions to existent VMs (Halliday, 2014) and, consequently, enhance taxonomies such as Halliday’s (2014) or Portolés and Zorraquino (1999). There are functional discrepancies between the two languages, for example, functions such as correction and corroboration are seen for English VMs (Halliday, 2014), while purpose and condition are seen for Spanish ‘operadores de refuerzo argumentativo’ (Portolés and Zorraquino, 1999).

It is expected that the reported work opens the way for future research on the creation of bilingual resources for applications in Translation Studies (TS), Machine Translation (MT) and Statistically-Based Machine Translation (SBMT) (Newmark, 1988).

The latter is expected to be achieved by providing a list of suitable translation equivalents according to context and function (Van Dijk, 2006), avoid literal translations which do not investigate the communicative purpose of discourse and speech while working with SMT (Statistical Machine translation).

Keywords: English, Spanish, operadores de refuerzo argumentativo, corpus-based translation, verificative markers.

# Resumen

Los estudios basados en los marcadores discursivos (MDs) desde una perspectiva contrastiva han ido creciendo en los últimos diez años, tanto desde una perspectiva teórica como desde una perspectiva práctica. Su estudio resulta interesante y complejo no sólo por su perspectiva semántico-pragmática, sino también por su ambigüedad y polifuncionalidad. Esta tesis tiene como objetivo contribuir a expandir este campo centrándose en los llamados ‘marcadores verificativos’ (MVs), tal y como se los clasifica en la lingüística sistémico-funcional (SFL).

Los MVs han recibido menos atención que otros MDs desde una perspectiva contrastiva, pero resultan muy frecuentes en diversos géneros discursivos. Los estudiados en esta tesis son ‘indeed’, ‘in fact’ y ‘actually’ para el inglés y ‘realmente’, ‘en realidad’, ‘de hecho’ y ‘la verdad’ para el español. La elección de este subtipo de marcadores VMs frente a otros se debe a que presentan una frecuencia mayor a la de otros en varios corpus de referencia del español y del inglés.

Como muestra de estudio se utilizan los corpus Multinot (Lavid et al., 2015), Europarl7 (Koehn, 2001), OpenSubtitles (Tiedemann, 2007) y EMEA (Tiedemann, 2007); los cuales abarcan tanto el registro formal como informal, oral y escrito, y los géneros argumentativo, expositivo, literario, conversacional y analítico. Es importante notar que, en el caso de Multinot (Lavid et al., 2015) sólo se toman como muestra 50 EO (textos originales en inglés), 50 SO (textos originales en español), 50 ET (textos traducidos al inglés) y 50 ST (textos traducidos al español). Además, los corpus Opensubtitles y EMEA (Tiedemann, 2007) son en realidad sub-corpus dentro del corpus Opus2 (D. Dobrovolskij & L. Pöppel, 2016).

*El estudio abarca las siguientes preguntas de investigación:*

- 1) ¿Cuáles son las correspondencias en traducción inglés-español y viceversa en los corpus paralelos? Esta pregunta se responde mediante el análisis de las equivalencias en traducción de ambos grupos de marcadores verificativos, en ambos idiomas, dentro de los corpus Multinot (Lavid et al. 2015), Europarl7 (Koehn, 2001), OpenSubtitles (López, 2016) y EMEA (López, 2016).
- 2) ¿Cuáles son los ‘dominos’ de traducción encontrados? Esta pregunta se responde con los resultados obtenidos en la primera pregunta y la retro-traducción en ambos idiomas.
- 3) ¿Existen preferencias en cuanto al género en el que aparece cada marcador? La tercera pregunta se centra en encontrar preferencias en traducción dependiendo del género textual escogido. Esto se obtiene analizando los resultados obtenidos en la resolución de la primera pregunta aplicada a los corpus Multinot (Lavid et al., 2015), Europarl7 (Koehn, 2001), OpenSubtitles y EMEA (Tiedemann, 2007).

Los resultados obtenidos para la primera pregunta son los siguientes: ‘in fact’ tiene preferencia por ‘de hecho’ en su traducción al español; en el caso de ‘indeed’, ‘de hecho’ coincide como preferencia en su traducción al español. Por otra parte, ‘actually’ prefiere en español la traducción hacia ‘realmente’. Estos resultados sustentan la premisa de que existe una fuerte conexión entre lo que Halliday (2014) denomina marcadores verificativos y lo que Portolés y Zorraquino (1999) consideran como ‘operadores de refuerzo argumentativo’. Adicionalmente, ‘actually’ se muestra como más periférico que ‘indeed’ e ‘in fact’.

En el caso de los marcadores en español (Portolés and Zorraquino, 1999) la respuesta es la siguiente: ‘de hecho’ prefiere ser traducido al inglés por ‘indeed’, ‘en realidad’ por ‘in fact’, ‘la verdad’ por ‘the truth’ y ‘realmente’ por ‘truly’. ‘La verdad’ y ‘realmente’ se muestran como los marcadores en español más periféricos dentro de los cuatro analizados en esta tesis, mientras que ‘de hecho’ y ‘en realidad’ serían los marcadores más centrales.

La respuesta a la segunda pregunta planteada se basa en los resultados obtenidos en la primera pregunta junto con la información obtenida en la traducción inversa de cada marcador. A esto hay que sumar las funciones adicionales encontradas tras la traducción en corpus paralelos de cada marcador. En el caso de los marcadores verificativos (Halliday, 2014), las funciones encontradas son: corroboración, particularización, reformulación, identificación, énfasis y explicación.

Las funciones adicionales encontradas para los marcadores en español tras el proceso de traducción en corpus paralelo (Portolés y Zorraquino, 1999) son las siguientes: reformulación, evaluación, modificación, rectificación, explicación, organización del discurso, propósito, contraste y énfasis. Se infiere que las funciones de énfasis, reformulación, evaluación y explicación son coincidentes en ambos idiomas, inglés y español.

La tercera pregunta se analiza de dos maneras, la primera trata de identificar cuál es la traducción preferida por cada marcador en cada género-corpus en ambos idiomas, español e inglés, y la segunda, identificar en que género el marcador en cuestión es más frecuente.

En este sentido, los géneros expositivo y argumentativo son los preferidos por ambos idiomas, inglés y español, no obstante, los marcadores en español (Portolés y Zorraquino, 1999) no se ven reflejados en los géneros dialogados ni analíticos mientras que los marcadores en inglés (Halliday, 2014) si lo hacen.

Funcionalmente hablando, ambos idiomas discrepan en algunos casos; las funciones de corrección y corroboración se muestran en los marcadores en inglés (Halliday, 2014), mientras que las funciones de objetivo y condición se observan para los marcadores en español (Portolés and Zorraquino, 1999). Se espera que los resultados de este proyecto sean de utilidad en todos los campos de la traducción automática (MT, SMT, SBMT) y los estudios generales en traducción (Newmark, 1988).

Estos resultados se pretenden obtener mediante la elaboración de una lista de dominios de traducción para cada marcador que permita una traducción automática más comunicativa y contextual y menos semántica y literal (Newmark, 1988).

Palabras clave: inglés, español, operadores de refuerzo argumentativo, traducción basada en corpus, marcadores verificativos.

# LIST OF ACRONYMS AND ABBREVIATIONS

**SFL:** Systemic functional linguistics.

**DRD:** discourse relational devices.

**EC:** elaborative connectives or conjunctions of elaboration or markers of elaboration or connectives of elaboration.

**ECV:** elaborative connectives of verification, verificative connectives, verificative conjunctions, elaborative conjunctions of verification, verificative markers.

**DM:** Discourse Marker.

**PDM:** Pragmatic Discourse Market.

**MT:** Machine translation.

**SMT:** statistic/ statistical Machine translations.

**EXPE:** expository texts.

**EO:** English original text.

**SO:** Spanish original text.

**ET:** English translation.

**ST:** Spanish translation.

**PM:** Per million/ per billion.

**Adj:** adjective/ adjetivo.

**AdjPh:** adjective phrase.

**Conj:** conjunction.

**Vb:** verb/verbo.

**VbPh:** Verbal phrase/ perífrasis verbal.

**N:** noun.

**Prep:** preposition/ preposición.

**PrepPH:** prepositional phrase.

**Interj:** interjection/ interjección.

**S:** Sustantivo.

**Loc:** locución.

**Adv:** adverb/ adverbio.

**AdvPH:** adverbial phrase.

**Loc.Adv:** locución adverbial.

**UKWaC:** English corpus from the .uk domain

**MULINOT:** Multidimensional Annotation of English-Spanish comparable and parallel texts for linguistic and computational applications.

**CREA:** corpus de la Real Academia Española.

**CORDE:** Corpus Diacrónico del Español.

**BYU:** Brigham Young University.

**EN-ES:** English-Spanish translation.

**ES-EN:** Spanish-English translation.

**DSM:** Discourse structural/ structuring markers.

**NMT:** Neural Machine translation.

**SL:** Source language.

**TL:** Target language.

**EMEA:** European Medical/ Medicine Agency.

**RAE:** Real Academia Española.

**BNC:** British National Corpus.

**UCM:** Universidad Complutense de Madrid.

**UNED:** Universidad Nacional de Educación a distancia.

**PhD:** Doctor.

**TS:** translation string.

**SBMT:** system-based machine translation.

**EDM:** elaborative discourse marker.

**CEO:** chief executive officer.

**<s>:** beginning of the string.

**</s>:** ending of the string.

**(...):** sentence omission.

**TRM:** text relation markers.

**CDM:** contrasting discourse markers.

**S1:** sentence one.

**S2:** sentence two.

**IDM:** inferential discourse markers.

**E-S:** English into Spanish translation.

**S-E:** Spanish into English translation.

**Discuss:** discursive.

**Doc:** document.

**EUROPARL:** European Parliament Proceedings.

**Conv:** convention.

**Pub:** public.

**Admin:** Administrative/ administration.

**Edu:** educational.

**Engine:** Engineering.

**Univ:** University.

**Ac:** accurate.

**Misc:** miscellaneous.

**Demonstratn:** demonstration.

**Nat:** Nature/ Natural.

**Brdsh:** broadsheet.

**Pop:** popular.

**Loc:** Locución.

**L1:** Language one (source).

**L2:** Language two (target).

**Alignment m:n:** Alignment and Alignment Score. (Here "alignment score" is "matching score" used in the definition of the sequence identity in problemE.html.

**U.T.C:** usado también como.

**COCA:** Corpus of Contemporary American English.

**CORPES:** El Corpus del Español del Siglo XXI.

**NEH:** National Endowment for the Humanities.

**NED:** Neduét University.

**RSS:** RSS (originally RDF Site Summary; later, two competing approaches emerged, which used the backronyms Rich Site Summary and Really Simple Syndication respectively) is a type of web feed which allows users and applications to access updates to online content in a standardized, computer-readable format. (<https://en.wikipedia.org/wiki/RSS>).

**VMs:** verificative marker.

**Ms:** Markers.

**LDMs:** lexical domains.



# List of Figures

Figure 1	Discourse particles Vs DM's, Fischer (2006). . . . .	28
Figure 2	LF Aligner (Farkas, 2016) parallel text spreadsheet result with Multinot texts (Lavid et al., 2015). . . . .	50
Figure 3	Halliday's taxonomy on verificative markers (2014). . . . .	51
Figure 4	Distribution of VMs in English reference corpora, based on VM's (Halliday, 2014) and VM's found in parallel corpora translation Multinot (Lavid et al., 2015). . . . .	63
Figure 5	Distribution of Spanish (operadores de refuerzo argumentativo) in Spanish ref- erence corpora (Portolés and Zorraquino, 1999) and Multinot (Lavid et al., 2015). . . . .	64
Figure 6	Word count distribution of registers in MULTINOT, after Lavid et al., 2015. . . . .	65
Figure 7	LF Aligner (Farkas, 2016) parallel text spreadsheet result with Multinot texts (Lavid et al., 2015). . . . .	72
Figure 8	screen capture of Europarl7 (Koehn, 2001) parallel concordance output. . . . .	74
Figure 9	screen capture of Open Subtitles (Tiedemann, 2007) parallel concordance output. . . . .	74
Figure 10	screen capture of Emea (Tiedemann, 2007) parallel concordance output. . . . .	75
Figure 11	'in fact' TLDs . . . . .	88
Figure 12	'actually' TLDs . . . . .	97
Figure 13	TLDs 'indeed'. . . . .	108

Figure 14 Genre-specific translation preferences of ‘in fact’, ‘actually’ and ‘indeed’. . . . 111

Figure 15 the lexical fields in Spanish resulting from translations of ‘infact’, ‘actually’ and ‘indeed’. . . . . 114

Figure 16 The second translation image: English translations of the most frequent Spanish translations of ‘in fact’. . . . . 115

Figure 17 The second translation image: English back-translations of the most frequent Spanish equivalents of ‘actually’. . . . . 117

Figure 18 The second translation image: English translations of the most frequent Spanish translations of ‘indeed’. . . . . 118

Figure 19 LDMs ‘de hecho’. . . . . 126

Figure 20 LDMs ‘la verdad’. . . . . 130

Figure 21 LDMs ‘en realidad’. . . . . 134

Figure 22 Genre preference for ‘realmente’, ‘la verdad’, ‘en realidad’ and ‘de hecho’ based on their occurrence in the corpora Multinot (Lavid et al., 2015), Opus2 (D. Dobrovolskij and L. Pöppel, 2016) Europarl7 (Koehn, 2001). . . . . 140

Figure 23 the semantic fields/ lexical domains for Spanish Markers ‘realmente’, ‘de hecho’, ‘la verdad’ and ‘en realidad’ based on the process of back-translation of their respective translation correspondences. . . . . 151

Figure 24 Semantic fields for ‘en realidad’, ‘realmente’. ‘de hecho’ and ‘la verdad’ based of the most frequent translation equivalents (Hummel, 1995) of their back-translations in Multinot Spanish (Lavid et al., 2015), Opus2 Spanish (D. Dobrovolskij and L. Pöppel, 2016)., Europarl7 Spanish (Koehn, 2001). . . . . 152

Figure 25 EO (English original), SO (Spanish original), ET (English translation). Arrows point out translation. . . . . 153

Figure 26 Frequency of each marker according to the analysed corpora: Multinot (Lavid et al., 2015), Opus2 (D. Dobrovolskij and L. Pöppel, 2016), Europarl7 (Koehn, 2001). Pm: per million, Multinot (Lavid et al., 2015) figures are absolute. . . . .	153
Figure 27 ‘indeed’ correspondence potential. . . . .	155
Figure 28 ‘in fact’ correspondence potential. . . . .	161
Figure 29 ‘actually’ correspondence potential. . . . .	169
Figure 30 Spanish LDMs . . . . .	175
Figure 31 Spanish LDMs in detail. . . . .	179
Figure 32 ‘en realidad’ correspondence potential. . . . .	182
Figure 33 ‘de hecho’ correspondence potential. . . . .	182



# List of Tables

Table 1	Del Saz (2002), Quirk et al (1985) and Bibber et al., (2006) sample outline. . . . .	23
Table 2	second level markers (Dirk, 2004). . . . .	43
Table 3	tools for automatic and manual analysis of text and corpora, Halliday (2014). . . . .	49
Table 4	Classification of some authors given to ‘indeed’, ‘in fact’ and ‘actually’, in con- trast to Halliday’s (2014). . . . .	53
Table 5	Del Saz (2003): Reformulation and Cuenca (2001): Reformulation and exempli- fication. . . . .	57
Table 6	Multinot (Lavid et al., 2015) word text number distribution based on a sample of 200 texts, 100 in Spanish and 100 in English from which 50 are EO and 50 are ET, 50 are SO and 50 ST. . . . .	66
Table 7	Translation correspondences of ‘in fact’ in Spanish translation corpora. . . . .	83
Table 8	Additional examples of ‘in fact’ translation equivalents (Hummel, 1995) in prag- matic functions (Aijmer, 2004). . . . .	87
Table 9	Translation correspondences of ‘Actually’ in Spanish translation corpora. . . . .	93
Table 10	Pragmatic additional examples for ‘actually’. . . . .	96
Table 11	Translation correspondences of ‘indeed’ in Spanish translation corpora. . . . .	101
Table 12	translation equivalents (Hummel, 1995) for ‘de hecho’. . . . .	124
Table 13	translation equivalents (Hummel, 1995) of ‘la verdad’. . . . .	128

Table 14 ‘en realidad’ translation equivalents (Hummel, 1995). . . . .	132
Table 15 translation equivalents (Hummel, 1995) of ‘realmente’. . . . .	136
Table 16 Corpus Spanish occurrences ‘la verdad’. . . . .	144
Table 17 Corpus Spanish occurrences ‘de hecho’. . . . .	145
Table 18 Corpus Spanish occurrences ‘en realidad’. . . . .	147
Table 19 Corpus Spanish occurrences ‘realmente’. . . . .	149
Table 20 ‘indeed’ correspondence degree. . . . .	155
Table 21 ‘in fact’ correspondence degree. . . . .	161
Table 22 ‘actually’ correspondence degree. . . . .	168
Table 23 ‘realmente’: correspondence potential. . . . .	181
Table 24 ‘la verdad’: correspondence potential. . . . .	181
Table 25 ‘en realidad’ correspondence degree. . . . .	182
Table 26 ‘de hecho’ correspondence degree. . . . .	182

# Contents

<b>Acknowledgements</b>	<b>VII</b>
<b>Abstract</b>	<b>IX</b>
<b>Resumen</b>	<b>XIV</b>
<b>List of Figures</b>	<b>XXIII</b>
<b>List of Tables</b>	<b>XXVII</b>
<b>I INTRODUCTION AND THEORETICAL BACKGROUND</b>	<b>XXXVII</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation and goals . . . . .	1
1.2 Research questions . . . . .	5
1.3 Thesis organization . . . . .	7
<b>2 Theoretical background and general approach</b>	<b>13</b>
2.1 Previous work on DMs . . . . .	13
2.2 Bruce Fraser’s account of DMs . . . . .	19
2.3 Fischer’s approach to discourse particles . . . . .	27
2.4 Dirk’s account of DMs across languages . . . . .	40

2.5	Multi-word discourse markers in translation . . . . .	45
<b>3</b>	<b>General Approach: Language in use theory of meaning</b>	<b>47</b>
3.1	The systemic-functional perspective . . . . .	47
3.2	The concept of ‘translation correspondence’ . . . . .	54
<b>4</b>	<b>Verificative Markers within the Semantic Space of Elaboration and Reformulation</b>	<b>55</b>
4.1	Previous work: del Saz (2002) and Cuenca’s (2001) . . . . .	57
4.2	Data used for the study . . . . .	59
4.2.1	Reference corpora . . . . .	60
4.2.2	English-Spanish parallel corpora outline . . . . .	64
4.2.3	The MULTINOT Corpus (Lavid et al., 2015) . . . . .	65
4.2.4	The Europarl7 Corpus (Koehn, 2001) . . . . .	66
4.2.5	The Open subtitles Corpus (Tiedemann, 2007) . . . . .	67
4.2.6	The EMEA Corpus (Tiedemann, 2007) . . . . .	67
<b>II</b>	<b>DATA ANALYSIS AND RESULTS</b>	<b>69</b>
<b>5</b>	<b>Data analysis procedure</b>	<b>71</b>
5.1	Identification of translation correspondences, study of differences between VMs from the same lexical domain and preferred genre . . . . .	72
5.2	Study of the differences between VMs from the same lexical domain . . . . .	77
5.3	Analysis of the genre translation preferences . . . . .	78
<b>6</b>	<b>Analysis of Results</b>	<b>79</b>
6.1	Translation correspondences of English verificative connectives and semantic fields .	81
6.1.1	‘In fact’ . . . . .	81
6.1.1.1	<i>Previous studies:</i> . . . . .	81

6.1.1.2	<i>Spanish translation correspondences of ‘in fact’</i> . . . . .	81
6.1.1.3	<i>Semantic groupings of ‘in fact’: Lexical domains</i> . . . . .	88
6.1.1.4	Genre preferences ‘in fact’ . . . . .	89
6.1.2	‘Actually’ . . . . .	90
6.1.2.1	<i>Previous studies</i> . . . . .	90
6.1.2.2	<i>Spanish translation correspondences of ‘actually’</i> . . . . .	91
6.1.2.3	Examples of translation equivalents for ‘actually’ based on table 9	93
6.1.2.4	Lexical domains ‘actually’ . . . . .	97
6.1.2.5	Genre-specific preferences of ‘actually’ . . . . .	97
6.1.3	‘Indeed’ . . . . .	98
6.1.3.1	<i>Previous studies</i> . . . . .	98
6.1.4	<i>Spanish translation correspondences of ‘indeed’</i> . . . . .	99
6.1.4.1	Examples of translation occurrences for ‘indeed’ . . . . .	103
6.1.4.2	Translation lexical Domains ‘indeed’ . . . . .	108
6.1.4.3	Genre-specific preferences ‘indeed’ . . . . .	109
6.1.5	Summary of genre-specific preferences of English VMs . . . . .	110
6.1.6	Summary of Genre-specific preferences ‘in fact’, ‘actually’ and ‘indeed’ . .	113
6.1.7	<b>Lexical domains for VM’s ‘in fact’, ‘indeed’, ‘actually’</b> . . . . .	114
6.1.8	Second-mirroring . . . . .	115
6.1.9	Second-mirroring ‘in fact’ . . . . .	115
6.1.9.1	Second-mirroring ‘actually’ . . . . .	117
6.1.9.2	Second-mirroring ‘indeed’ . . . . .	118
6.2	Spanish connectives: ‘la verdad’, ‘de hecho’, ‘en realidad’, ‘realmente’ . . . . .	120
6.2.1	Monolingual analysis of Spanish markers . . . . .	120
6.2.2	‘De hecho’ . . . . .	123

6.2.2.1	Monolingual analysis on ‘de hecho’ . . . . .	123
6.2.2.2	Translation correspondences of ‘de hecho’ . . . . .	124
6.2.2.3	<b>Lexical domains for ‘de hecho’</b> . . . . .	126
6.2.3	‘La verdad’ . . . . .	127
6.2.3.1	Monolingual analysis ‘la verdad’ . . . . .	127
6.2.3.2	Translation correspondences ‘ la verdad’ . . . . .	128
6.2.3.3	Translation lexical domains ‘la verdad’ . . . . .	130
6.2.4	‘En realidad’ . . . . .	131
6.2.4.1	Monolingual analysis of ‘en realidad’ . . . . .	131
6.2.4.2	Translation occurrences for ‘en realidad’ . . . . .	132
6.2.4.3	Lexical domains for ‘en realidad’ . . . . .	134
6.2.5	‘Realmente’ . . . . .	135
6.2.5.1	Monolingual study ‘Realmente’ . . . . .	135
6.2.5.2	Translation correspondences of ‘realmente’ . . . . .	136
6.2.5.3	Genre-specific preferences Spanish Markers . . . . .	139
6.2.5.4	<b>Part two of the process for Spanish markers explained: research question 2): back-translation</b> . . . . .	144
6.2.5.5	Lexical domains Spanish connectives . . . . .	150
6.2.5.6	Analysis English-Spanish: . . . . .	153
6.2.5.7	Frequency of each marker according to the analysed corpora: . . . . .	153
6.2.6	Summary of findings extension . . . . .	154
6.2.6.1	<b>Indeed:</b> . . . . .	154
6.2.6.2	<b>In fact:</b> . . . . .	160
6.2.6.3	<b>Syntax-punctuation discrepancies and similarities between ‘indeed’ and its translation counterparts:</b> . . . . .	165

---

6.2.6.4	<b>Actually:</b>	167
6.2.7	<b>Spanish ECs:</b>	174
<b>7</b>	<b>Discussion and Concluding remarks</b>	<b>187</b>
7.1	Introduction	187
7.2	Review of findings	189
7.3	Pointers for the future	191
	<b>References</b>	<b>191</b>
	<b>APPENDICES</b>	<b>211</b>
	<b>Appendix A. Curriculum Vitae</b>	<b>215</b>
	<b>Appendix B. Annotated examples in MULTINOT</b>	<b>223</b>



# **Part I**

## **INTRODUCTION AND THEORETICAL BACKGROUND**



# Chapter 1

## Introduction

### 1.1 Motivation and goals

Translation is ‘one of the very few cases where speakers evaluate meaning relations between expressions not as part of metalinguistic, philosophical or theoretical reflection, but as a normal kind of linguistic activity’ (Dyvik 1998:51). Moreover, the study of discourse markers (DMs) in the context of translation appears as crucial, mainly and foremost, due to the idiomatic nature of these structures (Aijmer 2007, Beeching 2013).

Besides, although an extensive literature has already reported language-specific traits of these events (Fraser 1990, 1999; Beeching and Detges 2014; Fisher 2000; Ghezzi and Molinelli 2014, *inter alia*), there are no systematic studies which address their cross-language behavior in the context of translation between English and Spanish. This is probably due to the wide range of meanings, terms, and the diverging typologies that have been proposed to study DMs in different languages.

In the area of Machine translation (MT), and more specifically, Statistical Machine translation (SMT) (Koehn, 2001), despite the relevance of DMs in discourse and speech and their essential role while working on translation, primarily if an accurate communicative or contextual translation is pursued; the role of DMs and even the role of discourse itself, have been disregarded or diminished in favor of segment or chunk translation, a process very close to lexical translation, which leaves aside

essential variables such as: genre, context, social background, schema, pragmatic force, pragmatic effect, politeness, emotional state and context cultural language preferences; among other variables that could determine the way a text or speech is conceived and received (Van Dijk, 2006).

According to Karim Sim Smith (2017):

Recently increasing amounts of effort have been going into addressing discourse explicitly in MT, with research covering lexical cohesion (Wong and Kit, 2012; Xiong et al., 2013b, a; Gong et al., 2015; Mascarell et al., 2015), discourse connectives (Cartoni et al., 2012, 2013; Meyer and Popescu-Belis, 2012; Meyer, 2011; Meyer et al., 2011; Steele, 2015; Steele and Specia, 2016), discourse relations (Guzmán et al., 2014), pronoun prediction (Guillou, 2012; Hardmeier et al., 2013b; Guillou, 2016) and negation (Fancellu and Webber, 2014; Wetzel and Bond, 2012). (Pg 110).

Neural Machine Translation (NMT) models are now the most performant, and although deeper than the linguistically superficial SMT, to evaluate progress we need to be able to measure the extent to which these models successfully integrate discourse. One of the issues preventing progress is a lack of understanding of the purpose of translation. (Pg 110).

While MT cannot be expected to assess the pragmatics, in terms of the intended effect on the target audience of the Source Language (SL) and ensuring a corresponding effect on the target audience of the Target Language (TL), there is a basic communicative intent in terms of the semantics which has to surely be taken into account in evaluation (...). (Pg 110).

(...) The standard metrics are still based on comparison to a single reference translation, which is inflexible (requiring a professional translation for every text automatically translated) and is also unrealistic as a text can be translated many ways, all of them valid. (...) it does not incentivize the integration of deeper linguistic elements. (Pg 110).

When referring to the specific traits used to parse DMs (markers) in the process of translation Karim Sim Smith (2017) makes the following point:

Current MT systems often fail to properly handle discourse connectives for various reasons, such as incorrect word alignments, the presence of multiword expressions as discourse markers, and the prevalence of ambiguous discourse markers. These can be incorrect or missing (Meyer and Poláková, 2013; Steele, 2015; Yung et al., 2015). (Pg 113).

Where discourse connectives are ambiguous, (...) the MT system may choose the wrong connective translation, which distorts the meaning of the text. It is also possible that the discourse connective is implicit in the source and thus needs to be inferred for the target. While a human translator can detect this, an MT system cannot. (Pg 113).

The focus of this thesis is to shed light into one specific subtype of DMs within the semantic space of Elaboration, the so-called ‘Verificative markers’ (VMs) in the systemic-functional linguistics literature (Halliday, 2014). The term ‘verificative’ is used by Halliday (2014) to refer to a group of EMs which includes markers such as ‘actually’, ‘in fact’, ‘as a matter of fact’ and ‘indeed’.

Their Spanish counterparts include markers such as ‘en realidad’, ‘de hecho’, ‘en efecto’ and ‘la verdad’ and have been described by Spanish linguists as ‘operadores de refuerzo argumentativo’ (Portolés and Zorraquino, 1999). The motivation for this selection of markers lies in the fact that these markers are very frequent in discourse but have been studied basically from a monolingual perspective (Garcés Gómez, 2005) or comparing English and Spanish (Murillo, 1999).

To our knowledge, there are no contrastive English-Spanish studies which focus on the translation correspondences of these specific markers, a much-needed task not only for descriptive linguistic and translation studies, but also for computational applications such as Machine Translation (MT) (Sim Smith, 2017) or Statistical Machine translation (SMT) (Fancellu, F. and Webber, B.; 2014):

The present thesis, therefore, is an attempt at filling a gap in this area by providing a functional-contrastive study of the most frequent VMs in English and Spanish by looking at their translation correspondences. With this general research goal in mind, the present study has the following specific aims:

- Investigate the core meanings of VMs in English and Spanish, as well as other possible peripheral ones or the pragmatic implicatures which can be derived from them.
- Construct contrastive (English - Spanish) semantic fields of VMs based on their translation equivalents.
- Investigate possible genre-specific preferences in the use of VMs in original and translated texts.
- Extract corpus-based features of VMs in English and Spanish which can be used for large-scale annotation of parallel corpora for computational applications such as Machine Translation (MT) and use the analysis for the creation of annotated parallel corpora (English-Spanish) for linguistic and computational tasks such as Multilingual Text Generation (MTG) (Karim Sim Smith, 2017).

The novelty and originality of these goals lies not only on the focus on these VMs, which have received little or no attention from the contrastive perspective, but also on the analysis of English-Spanish parallel corpora for the investigation of these VMs and on the applicability of the results for corpus annotation studies in the fields of MT and MTG (Sim Smith, 2017).

## 1.2 Research questions

In line with the aims presented above, the thesis addresses the following research questions:

*1.2.1 What are the translation correspondences between English and Spanish VMs as revealed by the analysis of parallel corpora?*

This means, which Spanish markers (Portolés and Zorraquino, 1999) appear as translation equivalents (Hummel, 1995) for each English verificative marker (Halliday, 2014) in each of the research corpora Multinot (Lavid et al., 2015), Europarl7 (Koehn, 2001), OpenSubtitles and EMEA (Tiedemann, 2007).

Each English marker (indeed, in fact, actually) will be traced using the tool SketchEngine (Kilgariff, 2019); for the cases of Europarl7 (Koehn, 2001), OpenSubtitles and EMEA (Tiedemann, 2007); and manually scanned in the case of the corpus Multinot (Lavid et al., 2015); the translation equivalents (Hummel, 1995) in Spanish will be listed through the parallel corpora mentioned before. In the case of Multinot (Lavid et al., 2015) the parallel alignment will be carried out using the tool LF Aligner (Farkas, 2016).

This process will end up with several tables of frequent translation equivalents (Hummel, 1995) for each English marker in Spanish in each corpora and their proportion per million will be compared in order to find the mostly used equivalents and the possible lexical domains based on new functions encountered for each marker in the process.

In the case of the Spanish markers; ‘la verdad’, ‘en realidad’, ‘de hecho’ and ‘realmente’; the process will be equivalent but comparing Spanish source and English target equivalents. This will also help to support the hypothesis of the strong relationship between the English VMs markers (Halliday, 2014) and the Spanish ‘marcadores de refuerzo argumentativo’ (Portolés and Zorraquino, 1999).

*1.2.2.- What are the translation domains revealed by such correspondences?*

This refers to the lexical domains into which the English and the Spanish VMs can be grouped based on their translation preferences. As will be shown, some VMs are semantically closer while others are more peripheral. The analysis of the translation equivalents of the English VMs into Spanish and the backtranslations are the basis of the creation of these translation domains for VMs in English and Spanish.

In sum, it is a process in which close equivalents in the target language are connected to the source language based on their proportional frequency in the parallel corpora, revealing additional functions for the source markers which might not have been addressed in monolingual studies.

*1.2.3- Are there any genre-specific preferences in the use of VMs in original and translated texts in both languages?* What are the typical contexts (grammatical, lexical, textual) in which these translation correspondences occur? This means, which is, proportionally speaking, the most frequent Spanish Marker (Portolés and Zorraquino, 1999) in each genre as suggested by research corpora for each verificative English Marker (Halliday, 2014).

The use of ‘genra’ refers to the different corpora used in the analysis, so this question is based on checking which translation equivalent (Hummel, 1995) is more frequent in each corpus, in the case of English source and Spanish target and Spanish source and English target. The genra which are considered are the following:

- a) Multinot essay (Lavid et al. 2015): formal, written, argumentative, expository.
- b) Multinot fiction (Lavid et al. 2015): formal, written, rhetorical.
- c) Multinot web (Lavid et al. 2015): formal, spoken, argumentative, expository.
- d) Multinot speech (Lavid et al. 2015): formal, spoken, argumentative, expository.
- e) Europarl7 (Koehn, 2001): formal, spoken, argumentative, expository.
- f) Emea (Tiedemann, 2007): formal, written, analytical.
- g) Opensubtitles (Tiedemann, 2007): informal, spoken, dialogues.

### 1.3 Thesis organization

After going through the necessary acknowledgements and the abstract in English and its translation into Spanish the thesis comprises two parts each consisting of different chapters.

Part I is dedicated to the presentation of the thesis in the introduction and theoretical background for the thesis as well as the general approach adopted for the investigation of Markers and DMs in translation; it consists of chapters 1, 2, 3 and 4 as follows:

Chapter 1 introduces the motivation and goals of the thesis, which can be summarized as finding translation equivalents (Hummel, 1995) in English and in Spanish for VMs, translation domains and genre preferences which can be useful in further translation studies and in MT (Newmark, 1988).

In chapter 2 the approach is based on the analysis of DMs carried out by several authors which will illustrate the polyfunctionality and relevance for translation of English markers and more specifically English Discourse markers, drawing a line in their relevance for translation studies and pragmatic analysis among other fields.

The theoretical background is the starting point, where the works under analysis will be that of Halliday (2014) under the label of Elaborative conjunctions as well as that of Jucker and Ziv (1988) analysis on the boundaries of DMs, this will bring light under the different points of view adopted towards the understanding of these linguistic elements. To the latter, some additional studies will be analysed, namely, Fraser's (2009) sentence analysis of DM's, Diane Blakemore (1996) concept of DMs encoding concepts and providing coherence, and what Redeker understands as 'discourse operators' (1991) as providing linkage to the discourse.

Related work by other authors is also reviewed, such as the work of Stede and Schmitz's on 'discourse particles' (2000) considered as units of talk and Howard J. R. (2010) 'signaling devices' as subordinators and conjuncts which work as 'relating networks', which provides an insight to the possible syntactic and pragmatic properties of these elements, as necessary for the cohesion and coherence of discourse and speech.

To this, the interest drawn on DM's in pragmatics will be illustrated by Del Saz, (2002) and Aijmer (2004) with several examples such as 'pragmatic markers of evidentiality', these markers are necessary to signal the truthfulness of a statement and its relevance for both the speaker/writer and the discourse/ speech itself.

In a second section of this chapter, Fraser's (2009) contribution towards providing an account of DM's will be analysed, focusing on the case of pragmatic markers which do not have connection with the discourse itself but with the context such as mental models or politeness (Van Dijk, 2006); afterwards, a classification of pragmatic markers will be provided based on Fraser's study (2009).

Fraser's (2009) account of DM's will be put into contrast with Halliday's (2014) classification of elaborative markers in the form of connectives and adverbial connectives. Giving an insight to the two sub-categories inside elaboration which are apposition and clarification (Halliday, 2014), it is precisely inside clarification where verificative markers 'indeed', 'actually' and 'in fact' are to be found.

Halliday's (2014) and Fraser's (2009) views on elaboration are set in contrast. Additional classifications for DM's will be the illustrated in a table, contrasting Del Saz (2002), Quirk et al (1985) and Biber (2006) to illustrate concepts such as reformulation, enhancement, addition, contrast, concession or corroboration, among others. Consequently, Fraser's (2009) Discourse structure markers are analysed as well as the characteristics he established for a linguistic marker to be a DM in both the Spanish (*en realidad, de hecho, la verdad, realmente*) and the English (*indeed, in fact, actually*) cases.

The next section in this chapter moves onto the analysis of Fisher's (2006) approaches to discourse particles, which includes not only DMs but also modalizers, adverbials, conjunctions and interjections, putting emphasis on the concept of poly- functionality.

A further section in this chapter moves onto Dirk's (2004) analysis of DMs and their linkage properties and how this can appear in the shape of multiple lexical and syntactical forms which can be carried out by the same linguistic element, in this case, DMs.

The fourth part of this chapter focuses on DM's in translation based on the works of D. Siepmann and L. Sprachen (2001), following Dirk's (2004) idea of second level DM's and how to use translation as a method to find additional functions for this DM's.

Chapter 3 presents the general conceptual approach followed in this thesis, based on a language in use theory of meaning, represented by the systemic-functional perspective (Halliday, 2014).

It also presents and explains the concept of "translation correspondence" as the basic methodological concept used for the current investigation and the basis of translation domains elaboration.

Chapter 4 reviews the results of monolingual studies on some VMs in English and Spanish, namely del Saz (2002) and Cuenca's (2001) studies on reformulation and exemplification in contrast with Halliday's concept of elaboration (2014).

Additionally, the choice of VMs are the point of study for this thesis will be justified. This means, the criteria by which 'indeed', 'actually', 'in fact', 'en realidad', 'de hecho', 'realmente' and 'la verdad'; have been considered as elements of analysis in this thesis, moreover, the choice of Elaborative markers (Halliday, 2014) and its relationship with Spanish 'marcadores de refuerzo argumentativo' (Portolés and Zorraquino, 1999) will be explained.

This section presents the data used for the study which is divided into reference corpora and English-Spanish corpora; accordingly, the British National Corpus (University of Oxford, 2009), the iWeb corpus (Davies, 2018), the UKWaC (Baroni et al., 2010) and the Timestamped JSI web corpus (2014-2019); as reference corpora. Dinzikova, I., and Paolini, M. (2018).

As research corpora, the ones used for this thesis are: Multinot (Lavid et al., 2015), Europarl7 (Koehn, 2001) and Opus2 (D. Dobrovolskij and L. Pöppel, 2016); as parallel corpora. Basing some premises on proportions counted in reference corpora such as BNC (2019) for English and RAE (2019) for Spanish; or the corpora analysed by professor Mark Davies (2011-2019).

Part II is focused on the study of VMs in English (indeed, actually, in fact) and Spanish (en realidad, de hecho, realmente, la verdad), putting forward the necessary data, tools and processes required in the collection of results.

Chapter 5 describes the data analysis procedure and annotation studies, explaining the process of identification of translation correspondences and lexical domains, the study of the difference between VMs from the same lexical domain and the analysis of the translation correspondences.

This chapter introduces the data and tools necessary for the research and collection of results in this thesis and it frames the context in which of the research will be carried out, illustrating the variables affecting the realization of the thesis.

Chapter 6 presents the analysis results, focusing first on the translation correspondences of the English VMs (section 6.2) and the latter on translation correspondences of the Spanish VMs (section 6.3). It illustrates the data analysis procedure explaining the process of identification of translation correspondences and lexical domains, the study of the difference between VMs from the same lexical domain and the analysis of translation correspondences.

More specifically, it would deal with the analysis of the specific English verificative markers ‘indeed’, ‘in fact’ and ‘actually’ (Halliday, 2014); answering the three research questions; and the same procedure for the Spanish ‘operadores de refuerzo argumentativo’ ‘en realidad’, ‘de hecho’, ‘la verdad’ and ‘realmente’ (Portolés and Zorraquino, 1999). Additionally, contribution to other fields and limitations of the study carried out in this thesis are also commented.

Additionally, contribution to other fields and limitations of the study carried out in this thesis will also be commented, such as the possible applications to other parallel corpora analysis or some application to SMT (Fancellu, F and Webber, B.; 2014) if the annotational process is automatized and an onomasiological dictionary is elaborated through the lexical domains found in both English (in fact, actually, indeed) and Spanish (en realidad, realmente, la verdad, de hecho).

Chapter 7 focuses on the discussion of results as well as the conclusions derived from the analysis carried out in chapter 6 for each of the VMs and for VMs as a whole, as well as the conclusions regarding the utility of translation in the analysis of VMs functions and the inclusion of new functions which have not been addressed in monolingual analysis.

After providing an account of references, some additional sections, such as appendices on symbols, figures and tables are provided as well as some screen captures of annotational examples in Multinot (Lavid et al., 2015). These screen captures illustrate the further steps to be carried out after this thesis, such as the elaboration of an onomasiological dictionary with lexical domains found in parallel translation or the collection of annotated texts from several parallel corpora.



# Chapter 2

## Theoretical background and general approach

English markers have a very strong polysemic functionality, they can word as modalizers, as adverbials, conjunctions, linkers, discourse markers or interjections (Aijmer, 2004), just to mention some. In order to provide a general overview on these markers, some of their most relevant contributions to discourse will be outlined in this section.

### 2.1 Previous work on DMs

DMs have been studied from different perspectives and under different labels in the literature. Typical labels include ‘cue phrases’ (Knott and Dale, 1994), ‘discourse connectives’ (Blakemore, 1987, 1992), ‘discourse operators’ (Redeker, 1990, 1991), ‘discourse particles’ (Schourup, 1999), ‘discourse signalling devices’ (Polanyi, 1983).

To these should be added ‘phatic connectives’ (Bazanella, 1990), ‘pragmatic connectives’ (van Dijk, 1979; Stubbs, 1983), ‘pragmatic expressions’ (Erman, 2001), ‘pragmatic formatives’ (Fraser, 1996), ‘pragmatic markers’ (Fraser, 1996, 1990; ‘pragmatic operators’ (Ariel, 1998), ‘pragmatic particles’ (Östman, 1995), ‘semantic conjunctions’ (Quirk et al., 1985), ‘sentence connectives’ (Halliday and Hasan, 1976).

As Jucker and Ziv (1988) explain:

It is therefore hard to decide which words should be included and which should not. If one takes a functional approach, there can be no finite and exclusive list of lexical items, since many words can be used both propositionally and as discourse markers. Often intuition seems to be involved, as researchers tend to be native speakers of the language under discussion. (Pg 223-259).

One classification strategy is to contrast discourse markers with other categories, such as conjunctions, but this dichotomy has been criticized. In sum, while scholars disagree about which criteria to use, they tend to agree that the boundaries of the discourse-marker category are fuzzy: it is unclear where exactly the category, however, defined, ends. (1988: 223-259).

As stated by del Saz (2002), interest on DMs has grown not only in Pragmatics and Discourse Analysis research, but also in other fields such as Sociolinguistics or Second Language Acquisition, not only for their versatility but also as “tools for testing and confirming hypotheses in new frameworks of analysis” (2002: 3-4).

This is the case of the Theory of Language Argumentation (Anscombe and Ducrot 1983), Relevance Theory (Sperber and Wilson 1986), and approaches from Text Linguistics (Van Dijk 1977, Halliday and Hasan 1976) or Discourse Analysis (Sinclair and Coulthard 1975, Stubbs 1983, Schiffrin 1987, McCarthy 1993a), to mention a few. According to Fraser (1999), discourse markers are:

Lexical expressions which signal a relationship between the interpretation of the segment they introduce, S2, and the prior segment, S1. Their meaning is procedural, and their more specific interpretation is 'negotiated' by the context. (Pg 223-259).

In the case of ‘Discourse Connectives’, Diane Blakemore (1996) points out that:

‘There are discourse connectives which, although they do not contribute to truth conditions, nevertheless encode concepts (cf. Wilson and Sperber 1993; Blakemore 1996, 1997; Ifantidou-Trouki 2001).’ (Pg 325-347).

(...) ‘Neither the relevance-theoretic reassessment of coherence relations (...) nor the reanalysis of discourse phenomena in section (...) should be taken as an argument that we should simply replace talk of coherence relations by talk of “relevance relations.” (Pg 325-347).

‘Coherence is a property of an object external to the human mind and is defined in terms of structural relations between subunits of that object. Relevance is a property of a mentally represented interpretation of the evidence a communicator provides for the thought(s) she or he intends to communicate, and is defined in terms of a function of the effects this interpretation has on the hearer’s overall representation of the world and the effort that is needed for its derivation.’ (Pg 325-347).

It seems that the point Blakemore is trying to make is that DMs, unlike the description given for ‘Cue phrases’, not only convey discourse connection but also what she calls ‘relevance’, this is, there seems to be a pragmatic implicature (Lewis, 2006) suggested by the choice of one DM or another.

It could be assumed that, as coherence and cohesion are ‘external organizers of the speech’, when trying to convey functions such as ‘emphasis’, ‘clarification’, ‘verification’, ‘exemplification’, ‘correction’, ‘summary’, and so on, DMs are working ‘internally to the discourse or speech’, not as simple organizers, but also as communicative vehicles which check that the message is properly received by the reader/listener with the intentionality the speaker/writer posted on it. In this sense, DMs include as part of their properties ‘Cue phrases’.

Discourse operators (Redeker, 1990, 1991): ‘a word or phrase (...) that is uttered with the primary function of bringing to the listener’s attention a particular kind of linkage of the upcoming utterance with the immediate discourse context (Van Dijk, 2006). Utterance considered as a clausal intonational unit’. The concept of ‘clausal intonational unit’ is very interesting, especially when the target text has no translation for the source or vice versa, but the intentionality is in either case understood.

Namely, DOs can be thought as a case of DMs, the same as CPs (Cue phrases) where thought to refer to one of the functions of DMs. According to the latter, one may conclude that DMs work both externally and internally to the text, providing both coherence and cohesion and making sure that the message is well processed, even making use only of intonation.

The previous assumptions are well summarized in Stede and Schmitz’s definition of Discourse Particles (2000), which, can be thought of as synonyms for DMs:

The class of discourse particles is a heterogeneous one and very difficult to demarcate in syntactic or other formal ways. Schiffrin (1987: 31ff) develops a lengthy definition of “units of talk” and then treats discourse particles (“markers” in her terminology) as bracketing such units. (Pg 125).

Fischer (1998) characterizes the class of German discourse particles as comprising both “interjections” and “segmentation markers”. Interjections “signal the spontaneous expression of a cognitive state” and subsume the classes of modal particles (occurring utterance medially, integrated into the phrasal intonation contour) and hesitation markers. (Pg 125).

For segmentation markers, Fischer refers to Schiffrin (1987), who lists these criteria: they bear no grammatical relationship to other elements in the sentence; they are not inflectable; they may be phonologically ill-formed; they connect utterances as a kind of “discourse glue”. (Pg 125).

Finally, a useful criterion for separating sentential usage from discourse usage is the deletion test: if a particle is removed from the utterance, and the translation is still a truthful rendering of the source utterance (even if it possibly differs in pragmatic force), then a discourse particle is at hand.” (Pg 125).

James Reggie Howard (2010): provides an explanation of the concept of Discourse signaling devices, referring to Hoey’s work on signaling:

Many clause relations in discourse are signaled through what Hoey refers to as “subordinators” and “conjuncts.” “It will be seen throughout this book that lexical signals are common and their recognition essential to successful discourse analysis.” These signals offer helpful interpretive clues. (pg 217).

Hoey (2010) points out that “Lexical signals may spell out a relation before, during, or after the event.” The other signaling device that Hoey deals with is repetition. “Winter (...) notes that sentences are unable to carry all the information that might be given on a subject; they are by their nature selective. Repetition is accordingly a way of ‘opening out’ a sentence so that its lexical uniqueness may be used as the basis for providing further, related information.” (pg 217).

(...) Two ways of uncovering or clarifying the relations. This is done through paraphrase and asking questions. He states that discourses should be recognized as “relation networks” rather than being strictly or primarily hierarchical. (pg 218).

He states, “Any sentence (or part of a sentence or group of sentences) may be in a relationship with any other sentence (or part of a sentence or group of sentences) within the same discourse.” Patterns of discourse the Hoey deals with are problem- solution1, Matching Patterns and 2, General – Particular Patterns 3.” (Pg 218).

The previous analysis on ‘discourse signals’ (DSs) brings about both ‘linguistic elements’ and ‘discourse functions.’ Conjunctions and connectives are considered as internal signalers (lexical level), but other more external relations such as ‘repetition’, ‘inquiry’ and ‘paraphrasing’ are also highlighted.

In the level of discourse, some additional functions are mentioned, namely: uncovering, clarifying, problem-solution (explanation), matching (expository) and general-particular (particularizing). With this definition of DSs, a more specific depiction of DMs is provided, getting closer to Halliday and Hassan’s 2014 taxonomy on markers relations of elaboration (apposition and clarification), there are others such as semantic conjunctions (Quirk et al., 1985), sentence connectives (Halliday and Hasan, 1976).

However, as stated by del Saz (2002), interest on DMs has grown not only in Pragmatics and Discourse Analysis research, but also in other fields such as Sociolinguistics or Second Language Acquisition, not only for their versatility but also as “tools for testing and confirming hypotheses in new frameworks of analysis” (2002: 3- 4).

This is the case of the Theory of Language Argumentation (Anscombe and Ducrot 1983), Relevance Theory (Sperber and Wilson 1986/1995), and approaches from Text Linguistics (Van Dijk 1977, Halliday and Hasan 1976) or Discourse Analysis (Sinclair and Coulthard 1975, Stubbs 1983, Schiffrin 1987, McCarthy 1993), to mention a few.

In the context of the present thesis, Discourse Markers (henceforth DMs) will be considered as ‘Nexus’ elements in the text (written o spoken) which have pragmatic, lexical, syntactic and ideational functions (del Saz, 2002), with external and/ or internal relations (Thomson, 2005) that involve two or more sentences and / or paragraphs in a text with the aim of causing a specific perlocutionary effect in the listener/ reader.

This function of DM will be considered among the other functions studied for English markers of verification (Halliday, 2014), namely, adverbials, modalizers, conjunctions, interjections or pragmatic markers (Aijmer, 2004); adding to these ones the functions obtained through the process of parallel corpora marker translation analysis, this is, having a look to the function carried out by the translation equivalent (Hummel, 1995) found for each marker in parallel corpora.

## 2.2 Bruce Fraser's account of DMs

Fraser develops a definition of Discourse Markers starting with what he calls 'pragmatic markers' (1996), which he assumes to have no connection with the 'propositional content' of the sentence or utterance where they occur but with what Austin would consider as perlocutionary force and illocutionary effect (1975), this is, with the contextual and intentional pragmatic aspects of the discourse or speech act (Austin, 1975).

Fraser distinguishes four types of pragmatic markers: basic, commentary, discourse markers and discourse structure markers. As for the present project, the two last types would have strong interest:

Discourse Marker: Fraser 2009:

'Typically signal a relation between the discourse segment which hosts them and the prior discourse segment, perhaps produced by another speaker.' He distinguishes three types:

- a) Contrastive: 'but', 'on the contrary'.
- b) Elaborative: 'and', 'anyway'.
- c) Inferential: 'so', 'as a result'.

As mentioned in the introduction, this thesis will be focused on Elaborative Markers on the grounds of Halliday's (2014) taxonomy. In the specific case of verificative Markers, it could be understood the, that the Discourse Markers which Fraser (2009) addresses as 'Elaborative Discourse

Markers' would have some similarities with Halliday's Elaborative markers when working as DMs, however, Halliday (2014) presents the following categorisation for Elaborative Markers:

#### APPOSITION

##### 1) Expository:

- In other words
- That is (to say)
- To put it another way
- I mean

##### 2) Exemplifying:

- For example
- For instance
- Thus
- To illustrate

#### CLARIFICATION

##### 1) Corrective:

- Or rather
- At least
- To be more precise

##### 2) Distractive

- By the way
- Incidentally

## 3) Dismissive

- In any case
- Anyway
- Leaving that aside

## 4) Particularizing

- In particular
- More especially

## 5) Resumptive

- As I was saying
- To resume
- To get back to the point

## 6) Summative

- In short
- To sum up
- In conclusion
- Briefly

## 7) Verificative

- Actually
- As a matter of fact
- In fact
- Indeed

As it is seen in the previous outline, the concept of elaboration differs between Halliday (2014) and Fraser (2009), Halliday (2014) defines elaboration as follows:

- 1) Apposition: in this type of elaboration some element is represented, or re-stated, either by ex-position or by example.
- 2) Clarification: the elaborated element is not simply restated but re-instated, summarized, made more precise or in some other way clarified for purposes of the discourse where there are seven subtypes:
  - a) Corrective (or rather, at least, to be more precise).
  - b) Distractive (by the way, incidentally).
  - c) Dismissive (in any case, anyway, leaving that aside).
  - d) Particularizing (in particular, more especially).
  - e) Resumptive (as I was saying, to resume, to get back to the point).
  - f) Summative (in short, to sum up, in conclusion, briefly).
  - g) Verificative (actually, as a matter of fact, in fact)'.

Whereas Fraser (2009) looks into elaboration as the construction of the discourse of the speech act, indeed, only anyway is included into elaboration by Halliday (2014) as 'dismissive'; the other markers, contrastive, elaborative and inferential appear to be seen by Halliday (2014) as extending-adversative (but, on the contrary), causal (so, as a result), extending (and).

Moreover, some other authors present alternative classifications based on the concepts of re-formulation, apposition and enhancement, among others, for English markers. The following table would illustrate some of the classifications that Del Saz (2002), Quirk et al., (1985) and Biber (2006) have provided for English markers and to which reference will be made in section 7 of this thesis 'analysis':

Del Saz (2002)	Quirk et al.'s(1985)	Bibber et al.,(2006)
<p><b>EXPLANATION</b>  <b>1)Clarification:</b>                      that is to say                      that is                      in other words                      in more technical terms                      technically speaking                      in plainer terms                      in simpler terms                      to put it more simply  <b>2)Identification:</b>                      Namely                      viz. videlicet                      to wit  <b>3)Illustration</b>                      (say) for example                      Instance  <b>RECTIFICATION</b>  <b>1)Neutral:</b>                      or rather                      on second thoughts  <b>2)Improvement</b>                      (or) better yet                      better still                      (or) better said  <b>3)Fine-tuning</b>                      (or) more precisely                      (or) more accurately                      (or) more specifically                      (or) more particularly                      (or) more exactly  <b>CONCLUSION</b>                      in conclusion                      to conclude                      to cap it off                      to top it off  <b>SUMMARY</b>  <b>1)Recapitulation:</b>                      to recap                      to recapitulate  <b>2)Summary:</b>                      in sum                      to summarize                      to sum up                      in short-brief                      in a nutshell                      in a word                      in a few words</p>	<p><b>EQUIVALENCE</b>  <b>1)Appellation:</b>                      namely                      or                      that is                      that is to say                      in other words  <b>2)Identification:</b>                      namely                      that is to say  <b>3)Designation:</b>                      that is to say                      in other words  <b>4)Reformulation:</b>                      in other words                      that is to say                      that is                      more simply                      in simpler words                      in simpler terms                      put more simply                      to put it more simply                      in more difficult language                      in scientific terminology                      in more technical terms                      technically speaking                      in words of one syllable                      or rather                      I mean  <b>THE LEAST APPOSITIVE</b>  <b>1)Exemplification:</b>                      for example                      for instance  <b>2)Particularization</b>                      Notably                      Particularly                      in particular                      chiefly                      mainly                      mostly                      in simpler words                      in simpler terms  <b>CONJUNCTS</b>                      put more simply                      to put it more simply                      or rather                      I mean  <b>1)Summative</b>                      in conclusion                      to conclude                      in sum                      to sum up                      to summarize  <b>2)Appositive</b>                      in other words                      namely                      that is to say                      that is                      more simply                      in other words</p>	<p>numeration/addition:                      a)Enumeration: first, second, third, fourth, firstly, secondly, thirdly, fourthly, in the first/second/third/fourth, place, first of all, for one thing, for another thing, to begin with, to start with, next, lastly.                      b)Addition: in addition, further, similarly, also, by the same token, furthermore, likewise, moreover, at, the same time, what is more as well too, and, add                      Summation: in sum, to conclude, all in all, in conclusion, overall, to summarize, in a nutshell                      Apposition:                      a)Restatement: which is to say, in other words, that is, i.e., namely, specifically                      b)Example: for instance for example e.g.                      Result/ inference: therefore, thus, consequently, as a result, hence, in consequence, so                      Contrast/ concession:                      a)Contrast: on the one hand, on the other hand, in contrast, alternatively, conversely, instead, on the contrary, in contrast, by comparison                      b)Concession: though, anyway, however, yet, anyhow besides, nevertheless, still, in any case, at any rate, in spite of that, after all                      Transition: by the way, incidentally, by the way</p>

Table 1: Del Saz (2002), Quirk et al (1985) and Bibber et al., (2006) sample outline.

According to the above table, the discourse Markers pointed out by Fraser (2009) seem to appear in a various of subtypes, namely: ‘on the contrary’ as reformulation-clarification (del Saz, 2002) and a appositive-contrast (Bibber, 2006), whereas ‘and’ is seen as appositive-addition (Bibber, 2006), ‘anyway’ is seen as appositive-concession(Bibber, 2006) and ‘so’ as appositive-inference (Bibber, 2006).

What could be inferred due to the discrepancies in classification is that, how a connective or marker is identified and featured depends on the perspective assumed by the author who elaborates the taxonomy, this is, whether the focus is on syntactic function, semantic or lexical domain or pragmatic force (Austin, 1962). As mentioned before, Fraser (2009) is putting the focus on the pragmatic force (Austin, 1962).

Apart from the above-mentioned DM, Fraser (2009) also mentions what he calls ‘Discourse structure Markers (DSM)’: ‘signal an aspect of the organization of the ongoing discourse’ (Fraser, 2009). He divides these markers into Discourse Management Markers (In summary, add), Topic Orientation Markers (returning to my previous topic, I want to return to), Attention Markers (look, now).

With respect to these DSM, the above table authors would classify them as follows: ‘in summary’ is considered both summative-reformulator (del Saz, 2002) and apposition-summative-conjunct (Quirk, 1985); on the other hand ‘add’ is considered as a liking adverbial of addition (Quirk, 1985).

Halliday (2014), by contrast, would classify these markers as follows: ‘in summary’ as elaborative conjunction summative and ‘add’ as extending positive additive conjunction. As said before, taxonomies rely on the linguistic feature which prevails for the author who outlines it, which means that all these functions will be complementary within each other.

Fraser (2009) establishes some conditions for a linguistic marker to be considered a DM, those conditions are the following:

- a) It must be a lexical expression.
- b) In a sequence of two sentences or propositions, a DM must occur as part of the second sentence or proposition.
- c) A DM does not add semantic meaning but points out a relationship between the first sentence or proposition and the second sentence or proposition.

It is important to note that here, Fraser (2009) is excluding from the concept of DM some specific grammatical items such as modal particles, sentence adverbs, interjections and evidential adverbs which had been indeed considered by other authors such as Halliday (2014).

This is very significant because as Aijmer (2004) suggests, the main functions of the markers that comprise this study – indeed, in fact, actually – are: sentence adverbial, evidential adverb, discourse marker and pragmatic marker of evidentiality; with the pragmatic functions of addition, contrast and enhancement.

Moreover, as it will be seen in the process of this study, modality and interjections would also appear as possible function of the VM's analysed in this thesis – the same applies to the Spanish ones: 'en realidad', 'de hecho', 'realmente', 'la verdad'. These points out the strong polyfunctionality of markers in both English and Spanish and the difficulty of enclosing them in a specific category.

With respect to the more specific classes of DM's Fraser (2009) addresses, the English markers core of this thesis: 'indeed', 'actually', 'in fact'; do not seem to enter any of his categories:

- Contrastive markers (CDMs), where a CDM signals a direct or indirect contrast between S1 and S2 ( but, alternatively, although, contrariwise, contrary to expectations, conversely, despite ( this/that ), even so , however, in spite of ( this/that ), in comparison ( with this / that ), in contrast ( to this/that ), instead ( of this / that ), nevertheless, nonetheless , ( this/that point ), notwithstanding , on the other hand , on the contrary, rather ( than this/that ), regardless ( of this/that ), still , though , whereas , yet ...). (Pg 87).
- Elaborative markers (EDMs), where an EDM signals an elaboration in S2 to the information contained in S1 ( and , above all, after all, also, alternatively, analogously, besides, by the same token, correspondingly, equally, for example, for instance, further ( more ) , in addition, in other words, in particular, likewise, more accurately, more importantly, more precisely, more to the point, moreover, on that basis, on top of it all, or, otherwise, rather, similarly ,...). (Pg 87).
- Inferential markers (IDMs), where an IDM signals that S1 provides a basis for inferring S2 ( so , all things considered, as a conclusion, as a consequence ( of this/that ), as a result ( of this/that ), because ( of this/that ), consequently, for this/ that reason, hence, it follows that, accordingly, in this/that/any case, on this/that condition, on these/those grounds, then, therefore, thus). (Pg 87).

Even though it has been said Fraser (2009) excluded adverbials as DM's:

DMs constitute a functional class, a heterogeneous syntactic group. They are drawn primarily from (28) a. conjunctions (and, but, or, nor, so, yet, although, whereas, unless, while ...) b. adverbials (anyway, besides, consequently, furthermore, still, however ,...), c. prepositional phrases ( above all, after all, as a consequence (of that), as a conclusion, as a result (of that), on the contrary, on the other hand, in other words, rather than that, regardless of that ,...) and very seldom from nouns, adjectives, verbs, or prepositions. (Pg 87).

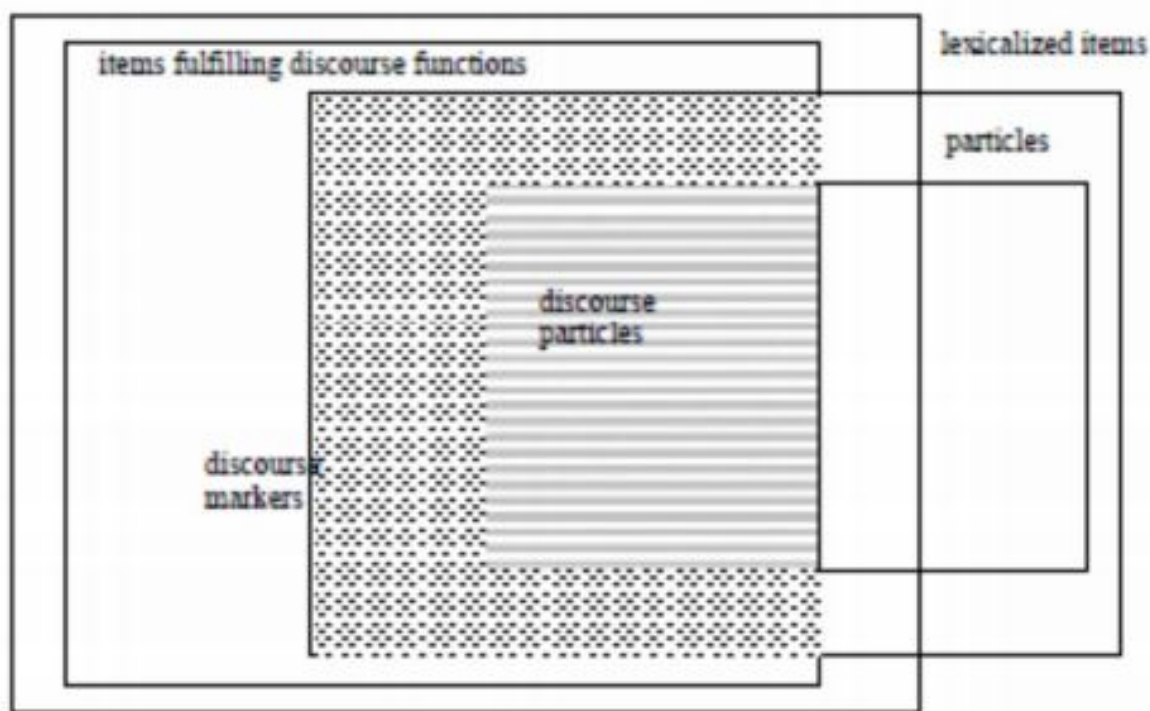
In this sense, ‘actually’ would enter as adverbial marker, ‘indeed’ as adverbial marker and ‘in fact’ as prepositional phrase marker. In the case of the Spanish markers, ‘en realidad’ would enter as prepositional phrase marker, ‘de hecho’ as prepositional phrase marker, ‘realmente’ as adverbial marker and ‘la verdad’ as an exceptional nominal marker (an issue which will be discussed in the analysis section).

To these properties, Fraser adds the concept of procedural and conceptual information (2009) this is, how the DM affects the second sentence or proposition, and what meaning, or force is intended (Austin, 1975). In this sense, when Aijmer speaks about adverbial segments, epistemic adverbs, discourse marker and pragmatic markers of evidentiality (2004), the reference would be on the procedure, whereas the concepts or functions of addition, contrast and enhancement would enter the area of concept. These two properties would appear, as illustrated in the analysis section of the present thesis.

### **2.3 Fischer’s approach to discourse particles**

The following figure illustrates the scope of functionality of discourse particles, which appear to be a type of DM which only refer to items fulfilling discourse functions. DM’s seem to go beyond items fulfilling discourse functions and include some functional particles. Finally, lexicalized items would include items fulfilling discourse functions as well as some particles (Fischer, 2006).

The relevance of this figure is that of framing the scope of functionality of markers, which would include both lexical items such as adverbials, modalizers or nominal phrases as well as discourse particles such as DM’s, conjunctions and interjections (Aijmer, 2004).



**Figure 1: Discourse particles Vs DM's, Fischer (2006).**

The previous figure provided by Fischer (2006) suggests that lexicalized items; such as those with adverbial function, nominal function, adjectival function or verbal function would have inside some specific linguistic elements which convey discourse functions. Functions such as contrast, addition or enhancement, inside which there is a group called ‘discourse markers’ that would embrace some of these lexical items with discourse function as well as other discourse particles which are not considered lexical items; such as connectives, conjunctions, prepositions or interjections.

Fischer (2006) supports the idea that one discourse item cannot be restricted into a single function, this is, for example the functions addressed by Aijmer (2004); such as addition, contrast, enhancement and evidentiality; could occur simultaneously in one discourse item depending on the utterance or discourse.

In this sense, Fischer (2006) quotes other authors to support this idea of polysemy and/or poly-functionality:

First, representatives of the polysemy approach in the narrow sense are Hansen (1998a), Lewis (2002), and Waltereit (2001). Interestingly, all three defendants of the polysemy approach combine a diachronic view with the synchronic perspective (Pg 15):

- Maj-Britt Mosegaard Hansen (1998a) proposes an approach to discourse markers, illustrated with the French discourse marker *toujours*, in which no core semantic meaning is assumed to connect the different readings of an item. Instead, the different senses are related by metaphoric or metonymic extension. (Pg 15).

Defining discourse marker meanings in networks such as radial categories or family resemblances has in her view the advantage of accounting for both the relatedness of the meanings and for the possibility of semantic change. Her model is thus designed to allow a particularly dynamic perspective on discourse marker meanings. (Pg 15).

- Diana Lewis (2002) defines discourse markers as form-function mappings, combining the semantics of discourse-relational predications, including speaker-attitudinal, relational meaning, with certain types of linguistic realisations that are syntactically dependent on their host (thus excluding interjections). (Pg 15).

The functional spectrum of discourse particles in her view comprises rhetorical management, information structuring, and thematic organisation. She also defends a polysemy approach to discourse particles (in the narrow sense) which she supports with diachronic analyses, thus presenting the synchronic state of polysemy as a result of diachronic developments. (Pg 15).

- Richard WALTERIT (1999, 2001) takes a diachronic perspective to explain the functional variability of discourse markers as well. By asking what turns a linguistic item into a discourse marker, he isolates rhetorical strategies by means of which speakers manipulate the structure of the discourse or the interaction. (Pg 15).

In this way, he explains the development of both new distributional contexts and new functions. The functional spectrum observable is in his account a result of the coexistence of uses developed in different diachronic stages in order to fulfil communicative needs. (Pg 15).

Similarly, PONS BORDERÍA (2005) holds that a polysemy approach in which the individual meanings are related by means of family resemblances would account best for his functional approach. However, in contrast to the other authors, he argues that it is just far too early to propose models of polyfunctionality. (Pg 15).

- Salvador PONS BORDERÍA (pg 15) defines discourse markedness as a hyperonym of three different functions: interactive, modal, and connective. He illustrates these functions, drawing on a wide range of descriptive methods and approaches, with the discourse markers of colloquial Spanish.

He argues that a model whose only purpose is to explain what discourse particles/markers are is of not much use. Instead, what is needed is a model of discourse that provides a place to locate these markers within. With the example of the definition of the units of discourse, he illustrates how an investigation of discourse markers can be used to develop aspects of a theory of discourse.

Then there are some approaches that assume a relationship between the different readings of an item, but which allow this relationship to be either specified by common meaning components or by conceptual or rhetorical links. This is particularly argued for by Aijmer, Foolen, and Simon-Vandenberg, and it is implicit in Roulet's and Zeevat's papers (pg 15):

- Karin Aijmer, Ad Foolen, and Anne-Marie Simon-Vandenberg (2003) analyse mainly English discourse particles, but they also look at the translation of the English discourse particles into other languages, such as Swedish. (Pg 16).

Their model of the polyfunctionality of discourse particles allows both a core meaning representation of the item under consideration as well as the description of several related senses. The core functions of discourse particles they hold to be epistemic or affective. The pragmatic interpretations of a discourse particle in their account is related to the indexicality of discourse particles and the heteroglossia of dialogues. (Pg 16).

- Eddy Roulet (1980) outlines not only a model whose purpose is to explain what text relation markers (TRMs) are but a model of discourse which provides a place to locate these markers within (cf. Pons Bordería, 2005).

In his view, accounting for the different functions TRMs can fulfil means providing a model of text units, text relations, and the role of TRMs therein. The TRMs themselves are considered to indicate illocutionary or interactive relations between a text constituent and information stored in discourse memory and to provide instructions for the reader/ hearer to facilitate access to the relevant information. The different readings of each TRM are conceptualized as distinct but related. (Pg 16).

- Henk Zeevat (2001-2002) defines discourse particles as context markers, relating the current sentence to aspects of common ground. He develops a formal semantic and pragmatic account for several English, Dutch, and German particles that have previously been discussed in connection with presupposition. In Zeevat's model, polysemy arises if a discourse particle marks two different relations to the context. (Pg 16).

The broader framework for his model could be a theory of speech act marking. The other approaches in this volume generally assume a single meaning for each phonological/orthographic form, that is, they take a monosemy approach. (Pg 16).

First, we find approaches in which general mechanisms account for the polyfunctionality of monosemous items. Ler employs the relevance theoretic framework, whereas Nyan uses the argumentation theory approach:

- Ler Soon Lay Vivien (2006) takes a relevance theoretic approach to discourse particles, investigating clause-final discourse particles in Singapore Colloquial English. (Pg 16).

The polyfunctionality of discourse particles she addresses by postulating a single instruction that is applied to different contexts. The inherent meaning of a discourse particle therefore accounts for the relatedness of the different readings, while the polyfunctionality results from the fact that each discourse particle use must be interpreted against the background of a new context. (Pg 16).

- Thanh Nyan's (2006) analysis of French discourse markers is based on Anscombe and Ducrot's Argumentation Theory (AT), augmented by neurolinguistic considerations of pragmatic processing. (Pg 16).

With AT, she holds discourse markers to have an instructional core meaning, and she is mainly concerned with their argumentative and cohesive functions. Taking a monosemy approach, AT possesses a conceptual apparatus susceptible of providing a general account of how the core meaning descriptions create different interpretations in context. (Pg 16).

Fraser and Weydt (2004) then do not specify any mechanism, but they describe those factors that determine the interpretation and that may cause pragmatic effects, such as politeness:

- Bruce Fraser (1988-2009) outlines a typology of pragmatic markers of which discourse markers constitute one subclass. Concentrating on those items that fulfil connecting functions, he attempts a systematisation of their formal and semantic properties. (Pg 17).

His model of the polyfunctionality of discourse markers comprises the core meaning description of the prototype, the interpretation and inferences of the utterances connected by the discourse marker, and linguistic and extralinguistic context factors. (Pg 17).

- Harald Weydt (1983) approaches the nature of the meanings and functions of particles by taking the question as to what particles are good for as a starting point. First, he reports on a study (Weydt et al., 1983) that shows that particles contribute to the impression of naturalness and friendliness of a dialogue. (Pg 17).

He then develops his model of how particles come to fulfil this kind of function step by step. Defending a monosemic approach to particles, he draws a clear distinction between the encoded semantic meaning of particles, holding this to be the same across word classes, and the individual uses, which may be listed in a dictionary. (Pg 17).

The pragmatic function to create a harmonious, friendly atmosphere is an aspect of their usage: they show that the speaker “is aware of what B, the other, thinks and believes.” Travis goes a step further by spelling out a set of specific readings that include the core meaning component related to the lexeme:

- Catherine Travis (1998, 2003) argues that a monosemantic view does not exclude a polysemy (in the narrow sense) approach to discourse particles. By using Wierzbicka’s Natural Semantic Metalanguage (Wierzbicka, 1994), she shows that a core meaning description can indicate how the different readings of a discourse particle are related, but at the same time different senses can be defined as extensions of the core meanings. She exemplifies her model on the Spanish discourse particle *bueno*, showing furthermore how structural position and intonation contour contribute to the disambiguation of the different senses. (Pg 17).

Like Travis' disambiguation of discourse particle interpretations by means of contextual factors, there are several approaches that take structural information to function as parameters for the selection of the right interpretation, only one of which is the invariant meaning component. (Pg 17).

Whereas Gupta concentrates on the contribution of sentence types, Yang and Rossari analyse the contribution of very complex contexts:

- Anthea Fraser Gupta (1992) proposes a definition of discourse particles as a very particular word class of Singlish, Singapore Colloquial English. This word class is directly associated with the function of signalling epistemic modality, which explains the syntactically and semantically peripheral status of discourse particles. The concrete functions of discourse particle use arise in interaction with the sentence types in which they occur. These sentence types must be considered as functional, as much as syntactic, categories. (Pg 17).
- Li-Chiung Yang (2001) attempts to integrate prosody in a model of the polyfunctionality of Mandarin Chinese discourse markers. In her model, the four functions of discourse markers she assumes (signalling phrase relationship, interactive relationship, as well as cognitive and emotional relationship) are signaled by discourse context, phrase position, lexical meaning, and prosody (including the prosodic context) to differing degrees. (Pg 17-18).

Prosody, particularly duration and intonation contour, is taken to play an important part in the disambiguation of the functions of discourse markers. The author shows how different functions of discourse markers are correlated with contours and durational features across speakers and despite the lexical tone assigned to each marker, and most importantly, also across different markers. (Pg 17-18).

- Corinne Rossari (1990) approaches the problem of the polyfunctionality of a subset of discourse markers: she takes connectives to be such items that propose restrictions for the right as well as the left context. (Pg 18).

The functional spectrum of (monosemous) connectives arises in her model due to different types of discourse configurations because of which the same operation of a connective gives rise to different semantic values. Thus, different discourse configurations are responsible for the polyfunctionality of the class. (Pg 18).

Discussing the French examples *alors* and *après tout*, she illustrates how connectives fulfil their main functions regarding the processing of information states. Another set of approaches attributes the polyfunctionality of (monosemous) discourse particles to their reference to discourse domains. Among these are Schiffrin, Redeker, and Frank-Job:

- Deborah Schiffrin's (1987) model of the polyfunctionality of discourse markers includes two scenarios: on the one hand, there may be polyfunctionality on the lexical level, on the other, discourse markers in toto perform multiple functions. Individual discourse markers, such as *and*, for which she provides an exemplary analysis, may select from a range of possible meanings depending on the domain that serves as a point of reference. (Pg 18).

While *and* in her view has only a single meaning and a single basic function, how to continue a cumulative set, what constitutes this set, and what provides the textual anchor for *and* may vary greatly. In her model, discourse markers are thus characterized as indexicals referring to different domains of discourse. (Pg 18).

- Gisela Redeker (1986-1992) develops a model of discourse coherence with three components: ideational, rhetorical, and sequential structures, regarding which discourse markers simultaneously mark semantic, rhetorical, and sequential relations respectively. The polyfunctionality of discourse markers results from their functioning in the three domains. (Pg 18).

Furthermore, Redeker demonstrates in the analysis of transcripts of Dutch and English discourse markers, as well as in a psycholinguistic experiment, that the markers of transitions in discourse function as attentional cues for the listener/reader. (Pg 18).

- Barbara Frank-Job (2006) argues that for an account of the polyfunctionality of (Italian) discourse markers a combination of a synchronic and a diachronic perspective is necessary. (Pg 18).

In a pragmatization process, originally deictic elements, indicating persons, times, or locations, may come to be used to fulfil metalinguistic functions. The synchronic polyfunctionality of these items results from their reference to three different levels of conversational structure: the turn-taking system, macrostructure, and superstructure. Regarding these three levels discourse markers fulfil their functions. Nemo takes a construction-based perspective. (Pg 18).

That is, a complex set of general form-meaning pairs, constructions that directly interact with the meaning of the respective particle morpheme, explain how the individual interpretations are created. (Pg 18).

- François Nemo (1986) suggests an approach to the polyfunctionality (and polycategoriality) of discourse particles, which he exemplifies on English but his model rests on morphemic meanings that interact with constructions, general form-meaning pairs, advocating a strict distinction between the morpheme's encoded meanings and the connective, functional, or categorial interpretations it may receive when it is inserted in specific (connective or non-connective) constructions. (Pg 19).

He also argues for a distinction between an utterance-type level and a contributory level in order to account for two subclasses of discourse particles, a large and very diverse class of utterance modifiers (illocutionary adverbs, evidentials, illocutionary particles, etc.) on the one hand and a class of contribution modifiers (discourse connectives, etc.) on the other. (Pg 19).

Diewald and Fischer then combine invariant meaning components, constructional, i.e., grammatical, information with reference to domains of discourse:

- Gabriele Diewald (1998) argues that the pragmatic functions of particles are genuine grammatical functions, indispensable for the organisation and structuring of spoken discourse. She holds that particles can be identified by their grammatical function, which is basically indexical. The apparent homonymy between the different word classes she explains by different elements indexed by the particle. At the same time, she argues for a monosemantic view of discourse particles. (Pg 19).

The abstract semantic content of a particle accounts for the relatedness of the different senses of a particle morpheme, while the reference to different domains, the application of the same semantic template to different reference elements in the sense of Sweetser (1990), explains the differences between the readings. She supports her model with current theories on the diachronic development of particles. (Pg 19).

- Kerstin Fischer (2006) proposes a model of the polyfunctionality of discourse particles that rests upon the interaction of three components: the invariant meaning of the respective particle morpheme, constructions modelling the general structural contexts in which discourse particles may occur, and communicative background frames, models of the communicative situation at hand. (Pg 19).

She develops her model by discussing the functions of the English discourse particle *okay* in human-to-human and in human-computer dialogues. The comparison of the two different corpora allows her to identify the background frames as speaker models of the respective communicative situations. (Pg 19).

Finally, Bazzanella combines reference to discourse planes with parameterization:

- Carla Bazzanella (1990) distinguishes three macrofunctions for discourse markers, which she exemplifies mainly on spoken Italian: interactional, metatextual, and cognitive functions. (Pg 19-20).

The choice of the intended reading that speaker and interlocutor/reader make in each text is, according to her model, activated based on the cooccurrence of several contextually and contextually relevant variables. (pg 19-20). These variables provide a parameterization of the meaning retrieved. Semantic correspondence provides thus only one clue to the interpretation of a discourse marker in use. (Fischer, 2006). (pg 19-20).

The latter extensive quotation on Fisher's (2006) analysis on the views of a great number of authors shows the complexity of describing those linguistic markers which mark the discourse performance. Some authors understand them as linkers, others as functional markers, others as epistemic markers, others as connectives and so on.

Moreover, there is no agreement on whether these particles have semantic/ lexical function-domain, pragmatic-domain, grammar-syntactic domain, morphological-phonetical or pragmatic one. Furthermore, some authors see them as part of the sentence or utterance, whereas others as part of the discourse as a unit or the speech act (Austin, 1975), meanwhile others understand them as being out of the discourse or speech.

As may be seen in this thesis, all these views appear to be valid and are strongly determined by context and illocution (Austin, 1975). One linguistic marker could work as either lexical or functional one; as discourse marker, sentence adverbial, pragmatic marker, epistemic adverbial, sentence conjunction or even interjection (Aijmer, 2004).

Through the process of corpus bilateral translation and annotation, additional functions and lexical domains are to be obtained, added to those of other authors, not only the above outlined but also Aijmer (2004) – addition, contrast, enhancement – Halliday (2014) Elaboration; Garcés Gómez

(2005), del Saz (2002), Cuenca (2001) – reformulation – or Portolés and Zorraquino’s (1999) – argumentative reinforcement.

This referring to the specific case of the following markers which Halliday (2014) address as Elaborative Markers of verification: ‘in fact’, ‘indeed’, ‘actually’, as well as those which Portolés and Zorraquino address as Argumentative Reinforcement operators: ‘en realidad’, ‘de hecho’, ‘realmente’, ‘la verdad’.

It is expected that the focus on these seven markers would be applicable to the study of any connective, DM or marker, in order to establish functions and domains and which would enhance existing taxonomies such as the one proposed by Halliday (2014).

As to conclude this section on Fischer (2006), I would include some of his comments on the English markers, which will be the focus of this project (pg 45-55):

- a) In fact: ‘The discourse marker in fact indicates that what follows is an elaboration of the previous idea.’ (Fischer, 2006).
- b) Indeed: ‘The particle indeed (or the Dutch *immers*, roughly ‘as you know’) presupposes the sentence in which it occurs and thus has arbitrary amounts of descriptive content.’ (Fischer, 2006).
- c) Actually: ‘Other expressions that can signal the RETREAT relation are actually, (...)’. (Fisher, 2006).

## 2.4 Dirk's account of DMs across languages

'Second-level discourse markers are restricted medium frequency collocations composed of two or more printed words and having a definable pragmatic function. They act as single units establishing local linkage between adjacent elements, sequences or text segments and/or global linkage between text segments further apart. (Dirk, 2004)' (Pg 10).

This definition of second level markers would be applicable to some of the markers studied in the present thesis, and these are: 'in fact', 'en realidad', 'de hecho', 'la verdad'. In the case of the three first ones because they conform prepositional phrases with a preposition as head and a noun as complement, whereas the latter, 'la verdad', is a nominal phrase with a noun as head and an article as adjunct.

As will be seen in this thesis, they can work as functional items, such as adverbs or nominal phrases (in the case of 'la verdad'); lexical items, such as connectors, prepositions, conjunctions or interjections; or as discourse markers, connecting two sentences, providing coherence to the discourse or speech, or external to the discourse or speech as pragmatics of evidentiality (Aijmer, 2004).

As it is now illustrated in some examples obtained from the corpus Europarl7 (Koehn, 2001):

In fact:

- There is, **in fact**, a risk of a military coup in the future. </s> *Evidentiality*.
- Gender has **in fact** been included in the compulsory introductory course for new employees, but then with only one to two hours devoted to it on each course. </s> *Adverbial, emphasis*.
- <s> It is, **in fact**, an international agreement and a trade agreement. </s> *Sentence connector*.

Indeed:

- <s> **Indeed**, it is quite in keeping with the positions this House has always adopted. </s>  
*Sentence or discourse connector.*
- <s> Madam President, I would firstly like to compliment you on the fact that you have kept your word and that, during this first part-session of the new year, the number of television channels in our offices has **indeed** increased considerably. </s> *Adverbial enhancement.*
- <s> This, **indeed**, is the purpose of the sixth Steel Aid Code. </s> *Evidentiality.*

Actually:

- <s> This time, however, I **actually** agree with her. </s> *Evidentiality, correction, truth.*
- <s> Mrs Lynne, you are quite right and I shall check whether this has **actually** not been done.  
</s> *Adverbial.*
- <s> **Actually**, I do not entirely share Mr Haarder ' s view. </s> *Sentence or discourse connector.*

En realidad:

- <s> La mayoría de las PYME tienen, **en realidad**, menos del 30 % de la cuota de mercado.  
</s> *Evidentiality, clarification.*
- <s> Señora Presidenta, ¿podría decirme por qué este Parlamento no cumple la legislación de sanidad y seguridad que **en realidad** ha aprobado? </s> *Adverbial.*
- <s> Pero **en realidad** quisiera destacar otra cuestión fundamental. </s> *Sentence connector.*

Realmente:

- <s> En determinadas localidades de mi país se produjo **realmente** una situación caótica. </s>  
*Adverbial.*
- <s> Señor Presidente, ahora la Comisión obtiene la aprobación de su gestión durante 1997; no obstante, **realmente** no se la merece. </s> *Evidentiality, clarificative.*
- <s> **Realmente**, la eficiencia ha sido notable. </s> *Sentence or discourse connector.*

De hecho:

- <s> La mayoría de ellas traducen **de hecho** las diferencias políticas que existen entre nosotros respecto a la oportunidad y a la eficacia de las ayudas estatales, habida cuenta de las insuficiencias, reconocidas o no, de las únicas inversiones privadas, market failures o insuficiencias del mercado. </s> *Asbervial.*
- <s> Se trata de la armonización de los requisitos de los exámenes, pero **de hecho** estamos hablando de los requisitos mínimos. </s> *Sentence connector.*
- <s> Otros han dicho que se estigmatizaba al pueblo austriaco y, **de hecho**, que se hacía un proceso de intenciones a un futuro gobierno. </s> *Evidentiality, factuality.*

La verdad:

- <s> Señor Presidente, **la verdad** es que la noticia de la suspensión o del retraso sine die de las conversaciones entre sirios e israelíes no es una buena noticia, como tampoco es una buena noticia el hecho de que se haya producido anteayer, nuevamente, un atentado con bomba en el que resultaron heridas 16 personas. </s> *Evidentiality, reformulation.*
- <s> Está claro que, por una vez, los Estados Unidos no han conseguido desbloquear la reanudación de estas negociaciones, y **la verdad** es que los obstáculos son difíciles: los sirios pretenden tener nuevamente bajo su soberanía y jurisdicción los Altos del Golán y recuperar las fronteras anteriores al 4 de junio de 1967, mientras que los israelíes pretenden, por considerar que se aviene mejor a su situación, tener las fronteras fijadas en el año 1923. </s> *Sentence connector.*
- Era **la verdad** del juego, pero a veces no hay suerte. </s> *Adverbial/ Adjective.*

Dirk (2004) suggests a classification of what he calls ‘second level markers’ as illustrated in the following table (pg 11):

Category	Examples
Polyword (invariable phrases which function like individual lexical items)	strictly speaking in other words at any rate what on earth?
Institutionalized Expression (invariable proverbs, aphorisms, formulaics)	get a life be that as it may nice meeting you
Phrasal Constraint (variable short-to-medium-length phrases)	as far as I (know/can tell) as a result of (...)
Sentence Builders (items which ‘provide a framework for whole sentences’)	it seems to me that my point here is (...) there’s no doubt that I think

**Table 2: second level markers (Dirk, 2004).**

The previously mentioned markers – ‘in fact’, ‘de hecho’, ‘en realidad’ and ‘la verdad’ – would correspond to the category of ‘polyword’, whereas some of the translation equivalents (Hummel, 1995) found for both the English and Spanish markers studied on this thesis would be either classified as belonging to ‘phrasal constraints’, ‘sentence builders’, or simply ‘markers’ namely:

- a) In fact: ‘en los hechos’, ‘lo cierto es que’, ‘en este sentido’, ‘por el contrario’.
- b) Indeed: ‘en el fondo’, ‘habida cuenta de’, ‘al fin y al cabo’, ‘lo cierto es que’, ‘a mi juicio’, ‘por el contrario’, ‘por consiguiente’, ‘bien mirado’.
- c) Actually: ‘a pesar de que’, ‘sin lugar a dudas’.
- d) En realidad: ‘in point of fact’, ‘it is true that’, ‘to tell the truth’.
- e) De hecho: ‘in point of fact’, ‘it is true that’, ‘to tell the truth’.
- f) Realmente: ‘in point of fact’, ‘it is true that’, ‘to tell the truth’.
- g) La verdad: ‘in point of fact’, ‘the truth is’, ‘it is true that’, ‘to tell the truth’.

These, as described by Dirk (2004), are a set of incomplete sentences or expression chunks which do perform some of the functions ideally addressed to DMs, in this thesis as one of the functions addressed to VMs (verificative markers or markers of verification (Halliday, 2014)).

## 2.5 Multi-word discourse markers in translation

Siepmann (2001), includes a chapter called: ‘Straddling cultures: three types of second-level discourse markers in contrastive perspective’. These ‘second-level’ discourse markers are divided into three groups: ‘exemplifiers’, ‘reformulators and resumers’, ‘inferrers’ (Siepmann, chapter 4, 2001).

In exemplifiers, he includes markers such as ‘for example’, ‘for instance’ or ‘namely’. In the case of reformulators and resumers he includes examples such as ‘in other words’, ‘to recapitulate’ or ‘to put it another way’. Finally, in the case of inferrers he includes examples such as ‘it follows’, ‘from here it follows’, ‘the foregoing that’.

As in the case of S.Dirk (2004), Siepmann provides examples of pragmatic markers which can constitute a sentence or a proposition, the ones called as ‘inferrers’, categories such as reformulation and resumptive markers have already been considered by other authors such as Cuenca (2001) and del Saz (2002) or Garcés Gómez (2005) as well as for the exemplifiers type. The case of inference is not that frequent but could be linked to that of pragmatic markers of evidentiality mentioned by Aijmer (2004).

One of the most significant issues is the fact that Siepmann (2001) uses translation as a method to ascribe functions, he works on intrinsic and extrinsic translation, this is, semantic domain or contextual domain (Van Dijk, 2006), moving from a functional view towards a more pragmatic one.

It is precisely through translation how, in this thesis, I will extract new functions and genre preference for the markers ‘indeed’, ‘in fact’, ‘actually’, ‘en realidad’, ‘de hecho’, ‘realmente’ and ‘la verdad’. So, it seems that corpus annotational processes in translation can help to enhance existing taxonomies and create lexical-semantic domains or families.

Some examples of the way translation can help to find additional functions could be found through the work on parallel corpora using the online tool Sketch Engine (Timestamped, 2019) are shown above (examples come from the corpus Europarl 7 (Koehn, 2001)):

In fact: Discourse Marker

- <s> There is, **in fact** , a risk of a military coup in the future. </s>
- <s> **De hecho**, existe el riesgo de un golpe militar en el futuro. </s>

Indeed: Pragmatic commentator of evidentiality

- <s> Equally, unemployment levels in the worst affected regions barely fell at all, **indeed** they rose in some cases. </s>
- <s> Igualmente, en las regiones más fuertemente afectadas apenas se han podido reducir las cifras de desempleados; **incluso** éstas han aumentado, en parte. </s>

Actually: Confirmation of truth adverbial

- <s> The Commission must track down the illegal aid and the aid which **actually** hinders the internal market. </s>
- <s> La Comisión debe perseguir las ayudas ilegales y aquellas que **realmente** ponen cortapisas al mercado interior. </s>

# Chapter 3

## General Approach: Language in use theory of meaning

### 3.1 The systemic-functional perspective

Halliday (2014) provides a detail explanation of the utility of corpora as a tool to obtain syntactic, morphologic, grammatical or pragmatic functions, as he explains (2014, pg 69):

The corpus was originally conceived as a tool for the study of grammar: Quirk referred to his Survey as ‘an NED of English usage’, and it played a fundamental part in the preparation of *A comprehensive grammar of the English language* (Quirk, Greenbaum, Leech and Svartvik, 1985).

But as it evolved into its present computerized form, the corpus was taken over by lexicologists (particularly lexicographers), and it is still thought of mainly in this connection today<sup>7</sup> although the recent *Longman grammar of spoken and written English*, developed by Bibber et al. (1999), is a corpus-based grammar.

(...) If we want to draw on the full potential of the corpus, we have to deploy computational tools to explore the corpus (see Teich, 2009; Wu, 2009) These tools are both enabling and constraining.

They are enabling because they make it possible to process much greater volumes of text than is possible by hand, allowing us to see features of language that were previously hidden from view.

They are constraining because the kind of analysis they can perform automatically is still a long way away from being as informative and rich as manual analysis.

Broadly speaking, automatic analysis gets harder the higher up we move along the hierarchy of stratification: it can handle any patterns that are stated in terms of orthographic words and it can handle certain low- ranking patterns within lexicogrammar, but is not able to handle full- fledged systemic functional analysis of clauses, and semantic analysis is also beyond its reach.

But analysis involving function structures and systemic features is much harder to automate for a free flow of text (cf. O'Donnell, 1994; Teich, 2009).

So, we have a trade-off between volume of analysis and richness of analysis: low-level analysis can be automated to handle large volumes of text, but high-level analysis has to be carried out by hand for small samples of text. (Pg 69).

In general terms, what Halliday (2014) suggests is that corpora can be used as a tool in the identification of functions in the process of translation, though, when automated, it cannot address complex functions which might be linked to either pragmatics of discourse/ speech context. The two possible processes of corpora analysis are the following shown in table 3 (pg 90):

Highest rank of analysis:	Highest axis of analysis:	Automatic analysis	Manual analysis
Clause	paradigmatic + syntagmic		<b>SysFan</b> [Wu, 2000]; WAG Coder [O'Donnell, 1994]; UAM CorpusTool [O'Donnell, 2011]
	syntagmatic only: function structure		Functional Grammar Processor [Webster, 1993]; Systemics [Kay O'Halloran & Kevin Judd]
	syntagmatic: syntagm	Standard parsers, e.g. Helsinki functional dependency	
Word		Taggers; concordancing programs, e.g. MonoConc, WordSmith, ConcGram [Graves, 2009], <b>SysConc</b> [Wu, 2000], <b>COBUILD tools</b>	

**Table 3: tools for automatic and manual analysis of text and corpora, Halliday (2014).**

In the specific case of the present thesis, the process of corpus analysis chosen has been that of manual analysis, where the specific verificative markers (Halliday, 2014) is highlighted, and his translation spotting is pointed out, annotating the possible functions to be assumed as a result of that specific translation choice as shown in the following figure which illustrates the parallel spreadsheet obtained after the use of LF Aligner (Farkas, 2016).

LF Aligner (Farkas, 2016) is an application in which the source and the target texts are put together in an excel document, source text in column 'A', target text in column 'B' and the annotation is explained in column 'C'. The marker in the ST is highlighter and its translation equivalent (Hummel, 1995) is highlighted in the TT. This process is illustrated in the following figure:

A	B
	1. CÓMO CONSTRUIR UN UNIVERSO
you will never be able to grasp just how tiny, how spatially unassuming, is of an atom, which is itself of course an insubstantial thing.	Por mucho que te esfuerces, nunca serás capaz de hacerte cargo de qué peq insignificante es un protón: sencillamente demasiado pequeño. Un protón es una parte infinitesimal de un átomo, que es en sí mismo, por sup
dib of ink like the dot on this i can hold something in the region of r more than the number of seconds contained in half a million years.	Los protones son tan pequeños l que una pizquita de tinta, como el punto de e unos 500.000 millones de ellos, o bastante más del número de segundos neces millón de años.
roscopic, to say the very least.	Así que los protones son extraordinariamente microscópicos, por decir algo.
course you can't) shrinking one of those protons down to a billionth of its ill that it would make a proton look enormous.	Ahora, imagínate, si puedes -y no puedes, claro-, que aprietas uno de esos pn milmillonésima parte de su tamaño normal en un espacio tan pequeño que un su lado.
ce about an ounce of matter.	Introduce después, en ese minúsculo espacio, una onza de materia. Muy bien.
3.	Ya estás en condiciones de poner un universo en marcha.
wish to build an inflationary universe.	Estoy dando por supuesto, obviamente, que lo que quieres construir es un uni
more old-fashioned, standard Big Bang universe, you'll need additional	Si en vez de eso prefirieses construir un universo clásico más anticuado, tipo t materiales suplementarios.
ip everything there is every last mote and particle of matter between here ueeze it into a spot so infinitesimally compact that it has no dimensions at	Necesitarías, <b>en realidad</b> , agrupar todo lo que hay (hasta la última mota y part hasta el límite de la creación) y apretarlo hasta reducirlo a un punto tan infinite no tuviese absolutamente ninguna dimensión.
ally big bang.	A eso es a lo que se llama una singularidad.
to a safe place to observe the spectacle.	En cualquier caso, prepárate para una explosión grande de verdad.
to retire to because outside the singularity there is no where.	Querrás retirarte a un lugar seguro para observar el espectáculo, como es nat
pand, it won't be spreading out to fill a larger emptiness.	Por desgracia, no hay ningún lugar al que retirarse, porque no hay ningún lugz
space it creates as it goes.	Cuando el universo empieza a expandirse, no lo hará para llenar un vacío may
te the singularity as a kind of pregnant dot hanging in a dark, boundless	El único espacio que existe es el que va creando al expandirse.
ssa.	Es natural, pero erróneo, visualizar la singularidad como una especie de puntc vacío ilimitado y oscuro.
around it. There is no space for it to occupy, no place for it to be.	Pero no hay ningún espacio, no hay ninguna oscuridad.
as been there—whether it has just lately popped into being, like a good re forever, quietly awaiting the right moment.	La singularidad no tiene nada a su alrededor, no hay espacio que pueda ocupi
from.	Ni siquiera cabe preguntar cuánto tiempo ha estado allí, si acaba de brotar a li buena idea, o si ha estado allí siempre, esperando tranquilamente el momentc
	El tiempo no existe.
	No hay ningún pasado del que surja.

Figure 2: LF Aligner (Farkas, 2016) parallel text spreadsheet result with Multinot texts (Lavid et al., 2015).

This picture shows the result of a process of alignment by means of the application LF aligner (Farkas, 2016) in which an original text, in this case, in English, is displayed next to its translation into another language, in this case, Spanish; with a third column were annotations according function are shown.

The marker is traced in the source, in this case ‘in fact’, and its translation equivalent (Hummel, 1995) is signaled, in this case, ‘en realidad’. Finally, there is another column added to the excel spreadsheet in which the function that ‘in fact’ has as a result to its translation into ‘en realidad’ is for that specific context and genre.

Halliday (2014) in his study of sentence behavior, found the following main functions: elaboration, extension and enhancement, and has reflected these functions in a classification of markers understood as those elements which provide either sentence/ proposition connection, discourse/ speech connection or pragmatic support. The following figure illustrates this taxonomy (pg 631):

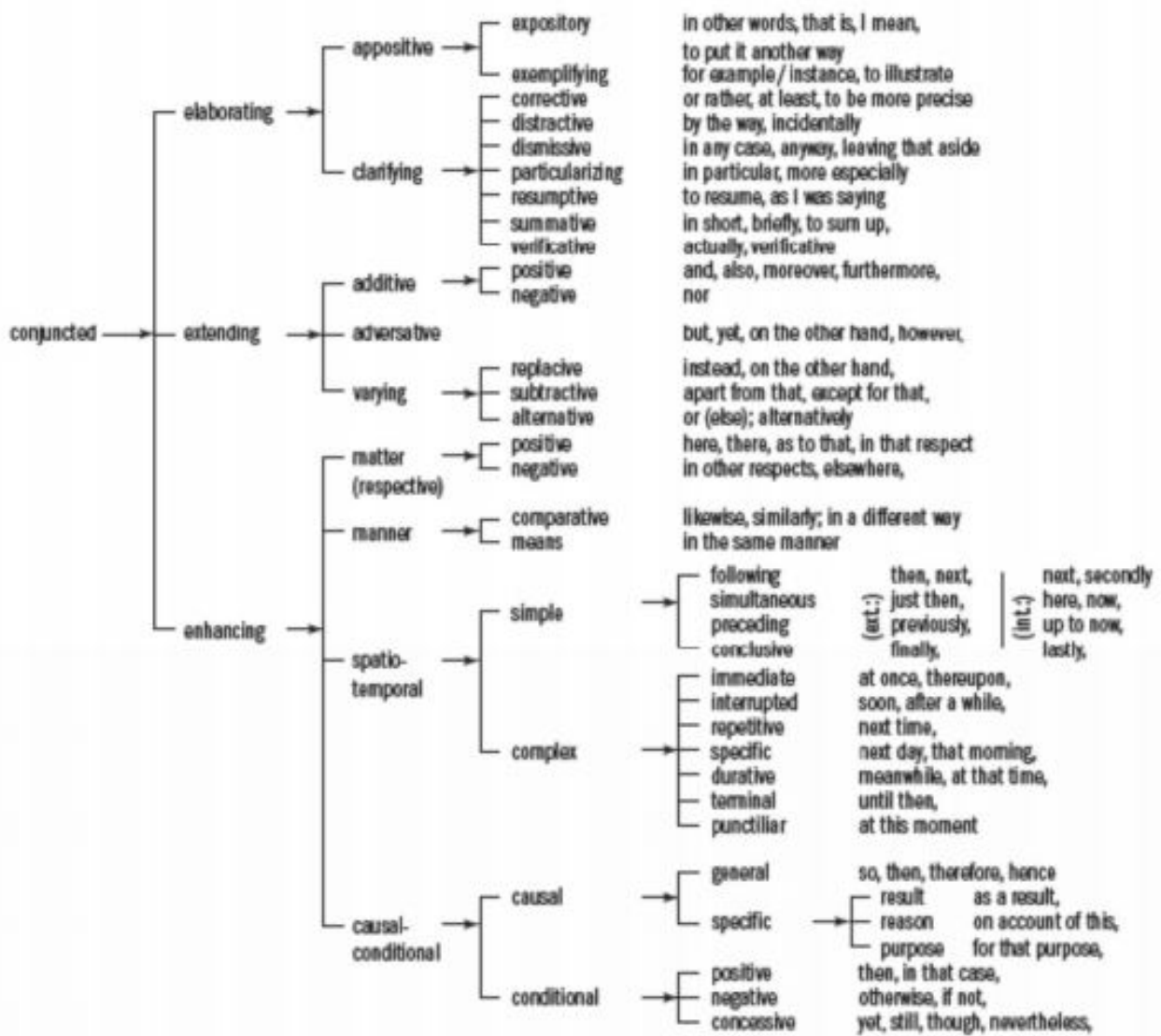


Figure 3: Halliday’s taxonomy on verificative markers (2014).

The present thesis is going to focus on the function of ‘elaboration’, and more specifically in three of the markers he addresses as ‘verificative’: ‘indeed’, ‘in fact’, ‘actually’. Halliday (2014) understands elaboration in two ways:

a) **Apposition**. In this type of elaboration some element is represented, or restated, either by exposition or by example:

- Expository/explanatory, the i.e. relation: in other words, that is, I mean, to put it another way.
- Exemplifying, the e.g. relation: for instance, for example, to illustrate.

b) **Clarification**: the elaborated element is not simply restated but re-instated, summarized, made more precise or in some other way clarified for purposes of the discourse. There are seven subtypes:

- Corrective (or rather, at least, to be more precise).
- Distractive (by the way, incidentally).
- Dismissive (in any case, anyway, leaving that aside).
- Particularizing (in particular, more especially).
- Resumptive (as I was saying, to resume, to get back to the point).
- Summative (in short, to sum up, in conclusion, briefly).
- Verificative (actually, as a matter of fact, in fact, indeed).

The present thesis focuses on the function of verification and on the three markers above mentioned, which appear in italics. The reason for choosing these markers and this specific function lies on the fact that Halliday (2014) is one of the only authors who refers to ‘verification’, the closest function in meaning is that addressed by Seyed Ali (2012) as corroboration.

The choice of the specific markers ‘indeed’, ‘in fact’ and ‘actually’ is based on their proportional occurrence per million in some reference corpora which will be illustrated in the introductory section. To illustrate the classification, some authors, give to ‘indeed’, ‘in fact’ and ‘actually’, in contraposition to Halliday’s (2014) one, the following table is provided:

Halliday, 2014	Quirk et al, 1990	Seyed Ali et al (2012)
ELABORATION CLARIFICATION Verificative Actually As a matter of fact In fact Indeed	CONJUNCTIONS: AND a)Listing: •Reinforcement and confirmation: above all, indeed, actually, in addition, additionally, moreover, again not only . . . but also . . ., also notably, as well (as) obviously, besides particularly, especially specifically, further then, furthermore too, what is more. OR a)Contrast: by (way of) contrast, conversely, in comparison, in fact, in reality, instead, on the contrary, (on the one hand) . . . on the other hand . . ., then	LINKERS Contrastive DCs a) Concessive: anyhow, anyway, anyways, besides, else, however, nevertheless, nonetheless (none the less), notwithstanding, only, still, though, yet, in any case, in any event, at any rate, at all event, for all that, in spite of that, in spite of it all, after all, at the same time, all the same, admittedly, still and all, that said, despite that, then again, whereas, while, whilst, in fact, actually, as a matter of fact, Emphasising DCs above all, after all, indeed, as a matter of fact, the main issue is, chiefly, especially, actually, the most significant, the chief characteristic, the major point, the most necessary, extremely, to emphasize, to highlight, to stress, by all means, undoubtedly, more importantly, most important of all, most of all, moreover, furthermore, significantly, without a doubt, certainly, to be sure, surely, absolutely, obviously, more and more, of major interest, to culminate, in truth, the climax of, to add to that, without question, unquestionably, as a result, probability, basically, indeed. Corroborative DCs actually, as a matter of fact, in fact, as it happens, at any rate, in actual fact, in any case, in either case, in reality, to tell the truth, that is to say, of course, apparently, well, surely, frankly, honestly, I assume, I suppose, no doubt, I am afraid, to tell the truth, in my opinion, I believe, to be truthful, unfortunately.

**Table 4: Classification of some authors given to ‘indeed’, ‘in fact’ and ‘actually’, in contraposition to Halliday’s (2014).**

The function of verification, as it will be seen in the section of analysis, is a function of checking and corroborating through elaboration and clarification of S1 (sentence 1) in S2 (sentence 2) by means of a marker (Halliday, 2014).

### 3.2 The concept of ‘translation correspondence’

A translation correspondence is a linguistic marker in the target language which appears as the equivalent of another linguistic trait in the source language when the process of translation is carried out.

a) Translation correspondences can occur in the same form and with the same function:

- **In fact**, we didn't mention anything about it.
- **De hecho**, no comentamos nada al respecto.

The source presents a connector working as discourse marker and providing enhancement in the form of a prepositional phrase followed by a comma. S2 is identical to S1 in the terms just mentioned.

b) Translation correspondences can occur with the same function but different position in the sentence and different form:

- **Indeed**, it is a shame she didn't come.
- Es una pena, **la verdad**, que no haya venido.

S1 and S1 work as discourse markers but S1 uses an adverb fronted and followed by a comma, whereas S2 uses a nominal phrase embedded in the sentence.

## **Chapter 4**

# **Verificative Markers within the Semantic Space of Elaboration and Reformulation**

Del Saz (2002) deals with “the notion of reformulation and with a group of lexical units by means of which the activity of reformulation is codified and explicitly signaled in the language. These lexical units are referred to as DMs of reformulation or reformulators.” (Del Saz, 2002).

The notion Del Saz (2002) addresses to “reformulation” is indeed very similar to that of “elaboration” addressed by Halliday (2014), there is coincidence in taxonomies for both authors with the exception of “verificative markers of elaboration” which Del Saz (2002) does not include in her taxonomy and motivates the necessity of analysing the relevance of “verificative markers of elaboration.” (Halliday, 2014).

As in the case of Halliday’s “markers of elaboration” (2014), Del Saz (2002) considers that “the area of reformulation and of English reformulators is still an uncharted and unexplored territory.” (Del Saz, 2002). Del Saz (2002) carries out a punctilious analysis of DMs according to their syntactic, distributional and pragmatic functions as well as their domain of usage.

In this thesis there is a study on the functional, syntactic and pragmatic properties found in parallel translation and compared to that of monolingual analysis, being Del Saz (2002) analysis on DMs of reformulation an excellent example to follow in this analysis.

Del Saz (2002) also analysis the domains of usage of DMs of reformulation, this analysis is a model to follow for the description of the different genra in which “verificative markers” show higher frequency of occurrence and which is shown in chapter 6 of this thesis.

Cuenca (2001) carries out a similar analysis on DMs of reformulation and explanation, comparing their usage between English, Spanish and Catalan. This represents a model of comparative analysis in translation of DMs and how translation can help to obtain more information about DMs, as it is concluded in Chapter 7 of this thesis with respect to “verificative markers” (Halliday2014).



Looking at these two taxonomies, the first thing which outstands is the difference in classification in comparison with Halliday (2014). The case of exemplification is very interesting as instead of being part of apposition, as in Halliday (2014), it is a type of reformulation inside which there is a sub-type called ‘clarification’ and another called ‘illustration’ which coincides with Halliday’s (2014) exemplification.

Moreover, ‘exposition’, Halliday (2014), is included in the sub-type of exemplification labelled as ‘clarification’. Furthermore, del Saz (2002) adds to new categories ‘conclusion’ and ‘identification’ which seem not to appear in Halliday’s (2014) taxonomy where he addresses elaboration, enhancement and extension.

Neither del Saz (2002) nor Cuenca (2001) focus on any of the three functions determined by Halliday (2014), moreover, some subtypes ascribed by Halliday (2014) seem not to be addressed by del Saz (2002); these are ‘distractive’, ‘dismissive’, ‘particularizing’ and ‘verificative’. As to the verificative category, neither del Saz (2002) nor Cuenca (2001) seem to address it and, to my knowledge, only some works use the concept of ‘verification’<sup>1</sup>.

As Halliday (2014) sees it, the function of verification is that of checking, not only clarifying, and, in order to check the listeners/ hearer’s understanding the speaker-writer might have to either illustrate, exemplify, explain, corroborate, emphasize or reformulate. All these functions could be understood as different means of verification, a premise which will be checked in the parallel corpora translation analysis carried out in chapters 6 and 7.

---

<sup>1</sup>The three English VMs— ‘indeed’, ‘in fact’, ‘actually’ – are only mentioned in the work of Quirk, Greenbaum, Leech and Svartvik (1990) and of Rezvani et al. (2012).

## 4.2 Data used for the study

Two main types of corpora were used for this study:

- a) Reference corpora for the English and the Spanish languages publicly available. These were used to select the most frequent VMs in English and Spanish. A brief description is provided in section 4.2.1 below.
- b) English-Spanish parallel corpora for the corpus analysis phase. These are described in detail in section 4.2.2 below.

Corpora main sources:

- a) Sketch Engine (Timestamped, 2019).
- b) Multinot (Lavid et al., 2015).

Tools used for the corpora analysis:

- a) Sketch Engine (Timestamped, 2019).
- b) LF Aligner (Farkas, 2016).

Specialized studies on English (in fact, indeed, actually) and Spanish markers (La verdad, en realidad, realmente, de hecho):

- a) Monolingual.
- b) Bilingual.

### 4.2.1 Reference corpora

For English, the reference corpora consulted were the British National Corpus (University of Oxford, 2009), the iWeb corpus (Davies, 2018), the UKWaC (Baroni et al, 2010) and the Timestamped JSI web corpus 2014-2019. Dinzikova I., and Paolini M. (2018). These are described in detail below:

#### **The British National Corpus (BNC)**

The BNC is a 100-million-word collection of samples of a written and spoken language of British English from the later part of the 20th century. The BNC consists of the bigger written part (90 %, e.g. newspapers, academic books, letters, essays, etc.) and the smaller spoken part (remaining 10 %, e.g. informal conversations, radio shows, etc.). The total number of words is 96,134,547. (BNC, 2019)

#### **The iWeb corpus**

The iWeb corpus (2018): this is one of the several corpora that professor Mark Davies (2018) includes in his web corpora collection. It represents a compendium of texts driven from several webs. It is the largest corpus created by Mark Davies (2018) and his group until now and consists of 14 billion words, including multimedia resources such as images, pronunciation, videos and even translations.

It allows users to create sub-corpora based on any topic of choice. Moreover, for each of the 60,000 lemmas, any of 95,000 web pages is available for the user, to trace the source of each word. Dictionaries, collocations, clusters are accessible too. (Davies, 2019).

**The UKWaC corpus** (Baroni et al, 2010)

The UKWaC (Baroni et al, 2010) is a text corpus of British English using medium-frequency words from the British National Corpus as source. It is a corpus of mainly British English with a total number of words of 1,313,058. Baroni et al (2010) developed this corpus as part of a project called WaCky (Web as Corpus Kool ynitiative) (Baroni, 2010).

**The Timestamped JSI web corpus 2014-2019:** (Dinzikova I, and Paolini M.,2018)

The Timestamped JSI web corpus is an English corpus made up of news articles gained from their RSS feeds. The corpus consists of news articles gained from newsfeed created by Jozef Stefan Institute, Slovenia (Trampus et al 2004).

The JSI Timestamped English corpus is a clean, continuous, real-time aggregated stream of semantically enriched news articles from RSS- enabled sites across the world. Dinzikova I, and Paolini M. (2018). The corpus is updated with new texts daily and grows by ca 800 million words each month. The English corpus contains over 37 billion words and it is the biggest English corpus in Sketch Engine. (Timestamped, 2019).

For Spanish, the reference corpora consulted were the Corpus de Referencia del Español Actual (CREA, 2008) and the Corpus Diacrónico del Español (CORDE, 1974), the Web dialects corpus (Davies, 2016), and the Google books corpus (Davies, 2011).

The CREA and CORDE corpora were developed at the Spanish Royal Academy ('Real Academia española').

**The CREA corpus (2008)**

The CREA corpus (2008) is a compendium of texts driven from various sources and Spanish dialectal variations, from both Iberian Spanish and Hispanic American Spanish. The topics range from standard to educational, including transcriptions from TV and radio. There has been too an annotated version of CREA (0.1) created in 2015. (CREA, 2019).

**The CORDE corpus (1974)**

The CORDE 1974 is a diachronic corpus based on text collection of Spanish speaking countries from the origin on the language until 1974. The current number of texts is 250 million registers, the genra vary from literary to scientific or legal. (RAE, 2019).

**The Web Dialects Corpus is inside the project of professor Mark Davies (2016)**

The Wed Dialects Corpus contains about two billion words of Spanish, taken from about two million web pages from 21 different Spanish- speaking countries from the past three to four years.

The corpus has been funded by the US National Endowment for the Humanities, and it has allowed to update the original Corpus del Español (2002), which was also funded by the NEH. (Davies, 2019).

**The Google Books from ‘corpus del español’ is also part of professor’s Mark Davies project (2011):**

The Google Books corpus is based on Google Books data, but it is not an official product of Google or Google Books. Rather it was created by Mark Davies, Professor of Linguistics at Brigham Young University, and it is related to other large corpora that they have created. Number of words: 45 billion words. (Davies, 2019).

The distribution per million of English VMs (as per Halliday’s taxonomy, 2014) in the English reference corpora is shown in figure 4 below:

	BNC	iweb	UKWaC
<b>Verificative (Halliday, 2014):</b>			
Actually	25 pm	30.481 pm	0.75 pm
As a matter of fact	4 pm	0.2276 pm	1.06 pm
In fact	16.1 pm	10.5696 pm	83.86 pm
Indeed	18.3 pm	62.721 pm	1.19 pm
<b>Found in parallel corpora translation Multinot (Lavid et al., 2015):</b>			
Naturally	4.1 pm	3.992 pm	0.44 pm
In practice	3.6 pm	0.901 pm	18.35 pm
In effect	2 pm	0.902 pm	8.28 pm
To verify	0.17 pm	0.955 pm	0.04 pm
In point of fact	0.03 pm	0.019 pm	0.3 pm
In truth	0.4 pm	0.288 pm	0.05 pm
Almost certainly	1.475 pm	0.481 pm	0 pm

**Figure 4: Distribution of VMs in English reference corpora, based on VM's (Halliday, 2014) and VM's found in parallel corpora translation Multinot (Lavid et al., 2015).**

As shown above, the most frequent VMs are 'in fact' (110.53), 'indeed' (82.211) and 'actually' (56.231). Therefore, the corpus analysis phase will focus on these VMs based on their higher proportion and, thus, relevance. In the case of Spanish markers, the focus has been of what Portolés and Zorraquino describe as 'operadores de refuerzo argumentativo' (1999), due to the close functional and semantic relation with the previously mentioned English markers. Their frequency distribution is displayed in figure 5 below:

<b>VM's Halliday (2014):</b>	<b>Web dialects</b>	<b>CREA</b>	<b>Google Books</b>
En realidad	89.32 pm	0.068 pm	0.117 pm
En el fondo	40.15 pm	0.030 pm	0.028 pm
De hecho	40.59 pm	0.057 pm	0.110 pm
La verdad	93.34 pm	0.092 pm	0.231 pm
<b>Multinot (Lavid et al., 2015):</b>			
En los hechos	3.67 pm	0.005 pm	0.003 pm
Lo cierto	27.01 pm	0.017 pm	0.029 pm
Realmente	77.24 pm	0.053 pm	0.221 pm
En efecto	51,51 pm	0.080 pm	0.027 pm
Bien mirado	0,87 pm	0.0005 pm	0.000023 pm

**Figure 5: Distribution of Spanish (operadores de refuerzo argumentativo) in Spanish reference corpora (Portolés and Zorraquino, 1999) and Multinot (Lavid et al., 2015).**

As shown above, the most frequent VMs are ‘en realidad’ (89.505), ‘realmente’ (77.514) and ‘de hecho’ (40.757) and ‘la verdad’ (93.664). Therefore, the corpus analysis phase will focus on these Spanish markers.

#### 4.2.2 English-Spanish parallel corpora outline

For the corpus analysis phase, it was necessary to find available English-Spanish parallel corpora which could be searched for automatically:

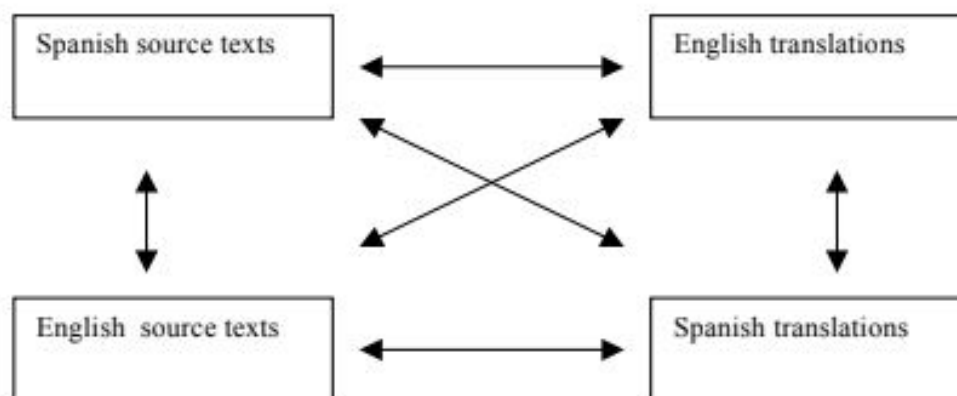
- a) The Multinot English Spanish Parallel Corpus (Lavid et al., 2015).
- b) The Europarl7 Corpus (Koehn, 2001).
- c) The Open Subtitles Corpus (Tiedemann, 2007).
- d) The EMEA Corpus (Tiedemann, 2007).

Each of these parallel corpora will be described in detail in the following subsections, including their main features, uses and composition. Some images are included in order to illustrate the output provided while working in these corpora.

### 4.2.3 The MULTINOT Corpus (Lavid et al., 2015)

The Multinot corpus (Lavid et al., 2015) was created by the FUNCAP research group at Complutense University of Madrid (UCM), as a result of the project entitled ‘Multidimensional Annotation of English-Spanish comparable and parallel texts for linguistic and computational applications.’

The following figure represents the ‘architecture of Multinot Corpus’ (Lavid et al., 2015):



**Figure 6: Word count distribution of registers in MULTINOT, after Lavid et al., 2015.**

Within MULTINOT (Lavid et al., 2015), the samples used for the corpus analysis were:

- Speech: transcriptions of parliamentary discourse and legal context.
- Expe: expository text of technical category.
- Web: uploadings of parliamentary and legal speech.
- Fiction: extracts from literary pieces.

With these sub-corpora the genera available would be exposition, argumentation, rhetoric; formal language; written and spoken language. The sample extracted from the Multinot corpus consists of two hundred texts, divided into two directional pairs from five different domains of the bilingual English-Spanish MULTINOT Corpus (Lavid et al., 2015) i.e.: fiction, essays, expository, legal procedures from webpages and speeches:

- 50 English original texts/ 50 Spanish translations.
- 50 Spanish original texts / 50 English translations.

The word text number is the one described in the following table:

GENRE	EO	STrans	SubTotal	SO	ETrans	SubTotal	TOTAL
GENRE	EO	STrans	SubTotal	SO	ETrans	SubTotal	TOTAL
ESSAY	18.532	19.883	38.415	20.036	20.500	40.536	78.951
EXPE	40.643	44.163	84.806	19.807	20.572	40.379	125.185
FICTION	10.459	11.338	21.797	13.164	13.985	27.149	48.946
WEB	16.553	18.893	35.446	14.830	13.848	28.678	64.124
SPEECH	36.675	38.610	75.285	307.625	275.856	583.481	658.766
TOTAL	122.862	132.887	255.749	375.462	344.761	720.223	975.972

**Table 6: Multinot (Lavid et al., 2015) word text number distribution based on a sample of 200 texts, 100 in Spanish and 100 in English from which 50 are EO and 50 are ET, 50 are SO and 50 ST.**

#### 4.2.4 The Europarl7 Corpus (Koehn, 2001)

The Europarl7 corpus (Koehn, 2001):

It is parallel corpus created from the European Parliament Proceedings in the official languages of the EU.

It includes 21 European languages: Romanic (French, Italian, Spanish, Portuguese, Romanian), Germanic (English, Dutch, German, Danish, Swedish), Slavik (Bulgarian, Czech, Polish, Slovak, Slovene), Finni-Ugric (Finnish, Hungarian, Estonian), Baltic (Latvian, Lithuanian), and Greek.

The corpus was repeatedly expanded with the final size around 60 million words per language. Texts are from the period January 2007 – November 2011. (Europarl, 2019).

In this thesis the both the Spanish-English and English-Spanish part of Europarl7 was used totaling a 10,000,000 for the latter out of a total number of 53,837,625 words in the whole corpus.

#### **4.2.5 The Open subtitles Corpus (Tiedemann, 2007)**

The Open Subtitles (Tiedemann, 2007) provides a collation of subtitles, constituting a sample of spoken informal language. It comprises transcriptions of English subtitles from all sorts of films. It is part of the parallel Opus2 Corpus (D. Dobrovolskij and L. Pöppel, 2016) which includes a great variety of sub-corpora and round 40 different languages.

Open Subtitles (Tiedemann, 2007) comprises several 550.4 tokens aligned between English and Spanish and several 548.9 tokens for Spanish-English. Opus2 corpus (D. Dobrovolskij and L. Pöppel, 2016) contains 1,139,515,048 words where Open Subtitles (Tiedemann, 2007) represents 815,779,086 words.

#### **4.2.6 The EMEA Corpus (Tiedemann, 2007)**

This is the European Medicines Agency documents (v.0.3) and it is available through the SketchEngine tool. It records fragments of medical technical documents as part of the corpus Opus2 (Dobrovolskij and Pöppel, 2016).

It is part of the parallel Opus2 Corpus (Dobrovolskij and Pöppel, 2016) which includes a great variety of sub-corpora and round 40 different languages. Open Subtitles (Tiedemann, 2007) comprises several 550.4 tokens aligned between English and Spanish and several 548.9 tokens for Spanish-English. Opus2 corpus (Dobrovolskij and Pöppel, 2016) contains 1,139,515,048 words where EMEA represents 15,408,620 words (Tiedemann, 2007).



## **Part II**

# **DATA ANALYSIS AND RESULTS**



# Chapter 5

## Data analysis procedure

### Data analysis procedure

The data analysis procedure of the VMs in English and Spanish consists of three phases:

- i) Identification of translation correspondences and lexical domains: answering to research questions one and two.

This means searching into parallel corpora to spot the possible translation equivalents (Hummel, 1995) of the source markers (both in English and in Spanish).

- ii) Study of differences between VMs from the same lexical domain: inside the study of research question one.

This means check the possible additional functions and proportional occurrences comparing monolingual studies of each marker with translation spottings functions.

- iii) Analysis of distribution of genre translation preferences: answering to research question three. Certain translation equivalences are used predominantly in certain genres but not in others. This is investigated in this phase both for the English VMs and their Spanish back-translations.

- iv) Checking which type of text presents a higher frequency of markers of a certain type proportionally speaking.

These phases are described in detail in the subsections below with some screen captures to illustrate the resulting outputs.

### 5.1 Identification of translation correspondences, study of differences between VMs from the same lexical domain and preferred genre

First, an VM from English is selected, and its Spanish corresponding item(s) is/are searched for in each of the bilingual corpora used in this study: MULTINOT (Lavid et al., 2015), Europarl7 (Koehn, 2001), Opus2 (D. Dobrovolskij and L. Pöppel, 2016). The procedure consists of three steps.

- 1) First an VM from English is randomly selected, and its common Spanish corresponding items are searched for. For example:

- a) Multinot (Lavid et al., 2015):

A	B
you will never be able to grasp just how tiny, how spatially unassuming, is of an atom, which is itself of course an insubstantial thing.	1. COMO CONSTRUIR UN UNIVERSO Por mucho que te esfuerces, nunca serás capaz de hacerte cargo de qué pequeño insignificante es un protón: sencillamente demasiado pequeño. Un protón es una parte infinitesimal de un átomo, que es en sí mismo, por supuesto
dib of ink like the dot on this I can hold something in the region of more than the number of seconds contained in half a million years.	Los protones son tan pequeños I que una pizquita de tinta, como el punto de unos 500.000 millones de ellos, o bastante más del número de segundos necesarios
microscopic, to say the very least.	Así que los protones son extraordinariamente microscópicos, por decir algo.
course you can't) shrinking one of those protons down to a billionth of its size that it would make a proton look enormous.	Ahora, imagínate, si puedes -y no puedes, claro-, que aprietas uno de esos protón milmillonésima parte de su tamaño normal en un espacio tan pequeño que un su lado.
ce about an ounce of matter.	Introduce después, en ese minúsculo espacio, una onza de materia. Muy bien.
3. wish to build an inflationary universe.	Ya estás en condiciones de poner un universo en marcha.
more old-fashioned, standard Big Bang universe, you'll need additional	Estoy dando por supuesto, obviamente, que lo que quieres construir es un universo. Si en vez de eso prefirieses construir un universo clásico más anticuado, tipo I materiales suplementarios.
ip everything there is every last mote and particle of matter between here and there, squeeze it into a spot so infinitesimally compact that it has no dimensions at all.	Necesitarías, <b>en realidad</b> , agrupar todo lo que hay (hasta la última mota y partícula hasta el límite de la creación) y apretarlo hasta reducirlo a un punto tan infinitesimal que no tuviese absolutamente ninguna dimensión. A eso es a lo que se llama una singularidad.
ally big bang.	En cualquier caso, prepárate para una explosión grande de verdad.
to a safe place to observe the spectacle.	Querás retirarte a un lugar seguro para observar el espectáculo, como es natural.
to retire to because outside the singularity there is no where.	Por desgracia, no hay ningún lugar al que retirarse, porque no hay ningún lugar.
pand, it won't be spreading out to fill a larger emptiness.	Cuando el universo empiece a expandirse, no lo hará para llenar un vacío mayor.
space it creates as it goes.	El único espacio que existe es el que va creando al expandirse.
te the singularity as a kind of pregnant dot hanging in a dark, boundless	Es natural, pero erróneo, visualizar la singularidad como una especie de punto vacío ilimitado y oscuro. Pero no hay ningún espacio, no hay ninguna oscuridad.
ness.	La singularidad no tiene nada a su alrededor, no hay espacio que pueda ocupar.
around it. There is no space for it to occupy, no place for it to be.	Ni siquiera cabe preguntar cuánto tiempo ha estado allí, si acaba de brotar a la vida.
as been there—whether it has just lately popped into being, like a good idea, or re forever, quietly awaiting the right moment.	El tiempo no existe.
from.	No hay ningún pasado del que surja.

Figure 7: LF Aligner (Farkas, 2016) parallel text spreadsheet result with Multinot texts (Lavid et al., 2015).

An original English-Spanish pair is chosen from the Speech sample of the MULTINOT Corpus (Lavid et al., 2015); for example, an original English text from the Speech genre (EO-Speech01) and its corresponding translation into Spanish (ST- Speech01). Using the tool LF Aligner (Farkas, 2016), the two texts are aligned and displayed in nearby columns in an Excel spreadsheet.

The one of the English VMs under study is chosen; namely ‘indeed’; and it is highlighted in the English original text (EO), afterwards, looking at the Spanish translation text on the right, its Spanish equivalent spotted is highlighted too; namely ‘en realidad’.

The same VM spotted in Multinot Speech01 (Lavid et al., 2015), is traced not only in all the other Speech sub-corpora under study but also in the sub-corpora Fiction, EXPE, WEB and ESSAY.

b) Europarl 7:

The same VM traced in Multinot (Lavid et al., 2015) is now checked in the parallel corpus Eurparl7 (Koehn 2001) by means of the Sketch Engine tool (Kilgariff, 2019). Firstly, the option ‘parallel concordance is selected’, afterwards the option ‘advanced search’ is chosen, then the English VM; namely ‘indeed’; is written on the right space provided by the application specifying the type of word (lemma, simple, phrase, etc...).

Finally, on the left section, the language must be chosen; in this case, Spanish. After the previous procedure, the Sketch Engine (Kilgarif, 2019) gives all the possible translations for the chosen English VM; namely ‘indeed’; in yellow color:

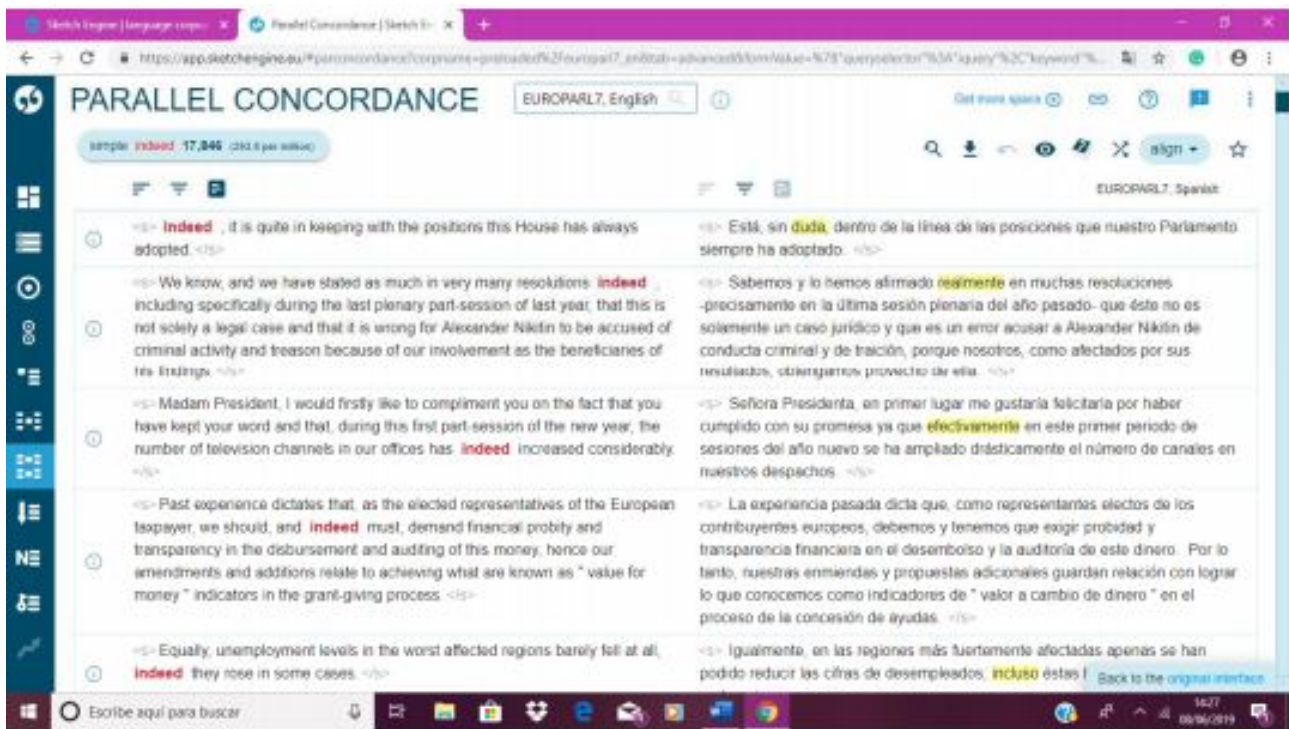


Figure 8: screen capture of Europarl7 (Koehn, 2001) parallel concordance output.

c) Opus2 (D. Dobrovolskij and L. Pöppel, 2016):

The same procedure carried out for Europarl7 (Koehn, 2001) is the one applicable to Opus2; in this case, there are two sub-corpora chosen inside Opus2 (D. Dobrovolskij and L. Pöppel, 2016) as seen in figures 9 and 10:

- OpenSubtitles (Tiedemann, 2007):

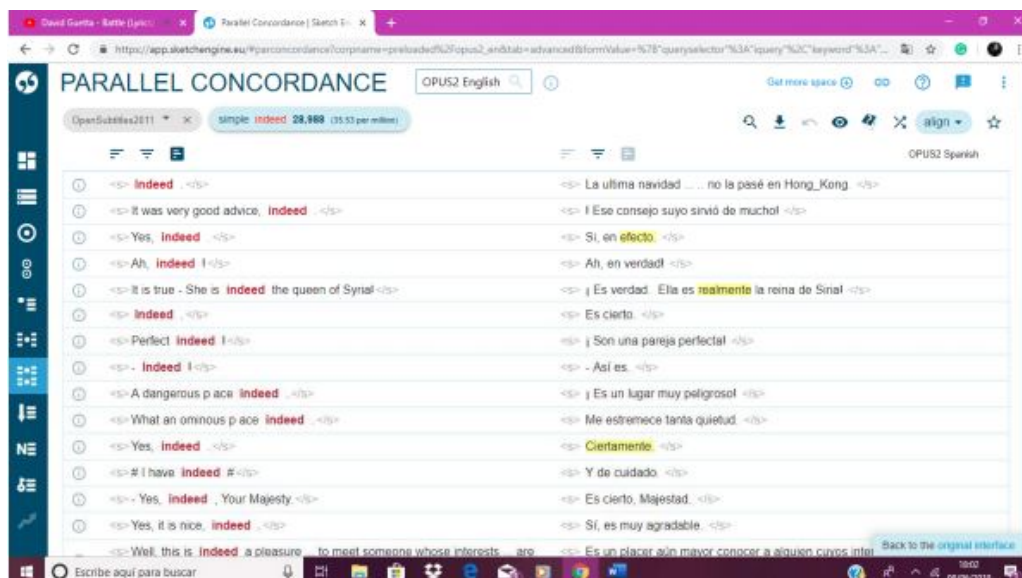
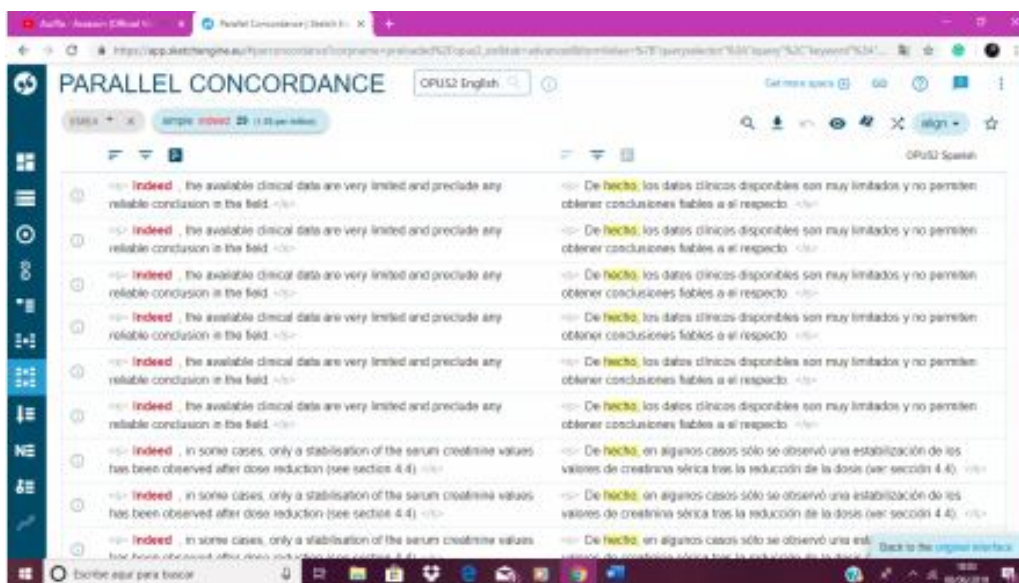


Figure 9: screen capture of Open Subtitles (Tiedemann, 2007) parallel concordance output.

- EMEA (Tiedemann, 2007)1.



**Figure 10: screen capture of Emea (Tiedemann, 2007) parallel concordance output.**

For the cases of the VMs ‘in fact’ and ‘actually’, Sketckengine (Baroni, Bernardini, Ferraresi and Zanchetta, 2009) would consider the former a phrase and the later as ‘simple’.

- d) The Spanish lexical items that occur in the same contexts are classified into the same substitution sets.

These substitution sets are created based on their frequency, this is, which is the most frequent Spanish equivalent in the target text for each VM in each of the given corpora. Based on the results substitution sets will be as follows:

- a) The most frequent translation equivalent (Hummel, 1995) for the VM in formal written expository and argumentative texts. These texts correspond to the most frequent occurrence in the corpora Multinot essay and multinot expe (Lavid et al., 2015).
- b) The most frequent translation equivalent (Hummel, 1995) in formal spoken argumentative and expository texts, these are the most frequent translation equivalents for the VM in the corpora Europarl7 (Koehn, 2001), Multinot Web and Multinot speech (Lavid et al., 2015).

- c) The most frequent translation equivalent (Hummel, 1995) in formal written analytical texts. This correspond to the maximum frequency in the corpus EMEA (Tiedemann, 2007).
- d) The most frequent translation equivalent (Hummel, 1995) in informal spoken texts. This correspond to the maximum frequency in the corpus Open Subtitles (Tiedemann, 2007).
- e) The most frequent translation equivalent (Hummel, 1995) in formal rhetorical texts. This correspond to the maximum frequency in the corpus Multinot Functions (Lavid et al., 2015).

An example of this process will be the case of ‘in fact’ and ‘de hecho’. ‘De hecho’ is the most frequent translation equivalent (Hummel, 1995) for ‘in fact’ proportionally speaking. What makes ‘de hecho’ to belong to the same translation domain as ‘in fact’.

Moreover, ‘de hecho’, for the case of ‘in fact’, occurs with significant frequency in formal argumentative and expository written and spoken texts (as mentioned above in a) to e)), which means that ‘in fact’ occurs more frequently in these type of genera/contexts (see sections 6 and 7).

- 2) The identified Spanish correspondences are translated back into English in order to find out if there are any other items which, along with the initial English VM, create a substitution set. Substitution sets containing the lexical items from two languages are then considered to form corresponding **translation lexical domains**. This method should be applicable to any type of VM.
- 3) After that, new back-translations from English into Spanish and vice versa are carried out until no new item appears in the lexical domains or sub-domains in any of the two languages.

- 4) The process described in 1, 2, 3 and 4 is then carried out with ‘de hecho’, ‘en realidad’, ‘la verdad’, ‘realmente’ (Portolés and Zorraquino, 1999) to reflect a similar analysis for Spanish markers are the one carried out for ‘indeed’, ‘in fact’ and ‘actually’ (Halliday, 2014).
- 5) The results from all the processes carried out in 5 is compared with the results obtained in 3 and 4, enhancing the possible lexical domains in both languages.

As mentioned in the introduction and theoretical framework, these four Spanish markers hold strong connection with English VMs, and this strong connection is expected to be seen by carrying out this analysis among them too.

## **5.2 Study of the differences between VMs from the same lexical domain**

The proportion per million of each translation correspondence according to each VMs is placed in a table. This table shows the proportional frequency of each translation equivalent (Hummel, 1995) in each of the parallel corpora mentioned in the previous sections, namely, Multinot (Lavid et al., 2015), Europarl 7 (Koehn, 2001) and Opus 2 (Tiedemann, 2007). Both in English and in Spanish.

Central and peripheral translation equivalents are identified according to higher and lower frequency, respectively. Then, these translation equivalents (Hummel, 1995) are analysed in context, this is, extracting an example from each of the parallel corpora where both the VMs in the source and the translation equivalent can be spotted.

Using the literature based on monolingual studies for both the VMs in the source and the translation equivalent (Hummel, 1995), compared with the function carried out by the translation equivalent in the target, existent functions and new functions for VMs are found. Then, VMs are classified according to their lexical domain and the function spotted in the translation correspondence.

### 5.3 Analysis of the genre translation preferences

Based on the results driven from the translation equivalent tables mentioned in section 5.2, the most frequent translation equivalent (Hummel, 1995) in each corpus determines the preferred genre for each VMs. The genres according to corpora are the ones described here:

- a) Multinot Essay (Lavid et al., 2015): written, formal, expository-argumentative text.
- b) Europal 7 (Koehn, 2001), Multinot Speech, Multinot Web (Lavid et al., 2015): spoken, formal, expository-argumentative texts.
- c) Open Subtitles (Tiedemann, 2007): spoken, informal texts (dialogues).
- d) EMEA (Tiedemann, 2007): written, analytical texts.
- e) Multinot Fiction (Lavid et al., 2015): written, formal, rhetorical texts.

a) to e) represent the genre, the most frequent translation equivalent (Hummel, 1995) of each VMs in each group determines which equivalent (Hummel, 1995) for each VMs is more suitable according to the source (written-spoken), context-politeness (formal- informal) and style (argumentative-expository, dialogues, analytical, rhetorical).

Additionally, among the genres mentioned in a) to e), the frequency of occurrence of each VMs is checked and compared, obtaining which is the genre where each VMs appears with higher frequency, this is, which is the preferred genre for each VMs, both in English and in Spanish.

# Chapter 6

## Analysis of Results

The results of the data analysis described in chapter 5 above are presented in detail in this chapter for both the English (section 6.1) and the Spanish VMs (section 6.2), respectively; with the goal of answering research questions one, two and three; as well as obtaining relevant data such as lexical domains or the correspondence potential.

As commented in previous sections research questions focus on the finding of additional functions for VMs to those described in monolingual studies, identifying genre preferences among VMs as well as translation preferences, this is, which is the most frequent translation correspondence for each VMs.

1) The data regarding the frequency of occurrence of a translation correspondence for a given VMs is displayed in several tables distributed according the corpora described in previous sections.

2) The preferred genre is illustrated through figures and diagrams which show the preferred VMs per genre as well as the preferred genre for VMs occurrence.

3) The lexical domains are shown in figures, illustrating the preferred translation occurrence according to the function required in context. It is noted by TLD.

4) The correspondence potential is a number calculated on the basis of the occurrence in corpus and the amount of times found within corpora. It is a method which helps to reinforce the hypothesis of whether one VMs is more central than another.

The structure of each section is the same: first, a revision of previous studies is outlined, analysing monolingual studies for each VMs in English and in Spanish; this is followed by the presentation of the translation correspondences<sup>1</sup> of each English VM.

These translation correspondences provide additional functions which are added to the ones found in monolingual studies for each VMs in both English and Spanish, finally, the creation of the semantic groupings or translation lexical domains (TLDs) based on these correspondences is done. In addition to the latter, some back- translations<sup>2</sup> are shown as well as punctuation discrepancies.

Finally, the genre-specific preferences of each English marker are outlined (actually, indeed, in fact) as well as the Spanish ones in further sections (la verdad, en realidad, realmente, de hecho). Section 6.1 and its subsections focus on the analysis results of the English VMs and section 6.2 focuses on the Spanish ones.

---

<sup>1</sup>According to Dyvik (1998), a translation correspondence is a lexical item from a target language that occurs in the same context as a corresponding item from a source language.

<sup>2</sup>According to B.A. Hatim and J. Munday (2004): 'A back-translation is a translation that is very close to the lexical and syntactic patterning of the source text, where the target equivalents are used as source to check if their translation correspondences coincide with the initial source elements'

## 6.1 Translation correspondences of English verificative connectives and semantic fields

### 6.1.1 'In fact'

#### 6.1.1.1 *Previous studies:*

One of the most interesting and thorough analysis of 'in fact', 'indeed' and 'actually' as pragmatic markers has been provided by Aijmer and Vandenberg (2004).

They first review the three different meanings posited by Schwenter and Traugott (2000), namely:

- 1) Epistemic sentence adverbial, meaning 'in practice, as far as can be told from evidence, in actuality';
- 2) Adversative marker with primarily epistemic meaning.
- 3) Discourse marker "*signalling that what follows is a stronger argument than what precedes, with respect to the speaker's rhetorical purpose at that point in the discourse.*" (Traugott(2000)).

Then they propose more types of uses of 'in fact' according to their data:

'Besides *in fact*<sub>2</sub> (which we will refer to for the sake of convenience as the 'but in fact') and *in fact*<sub>3</sub> (which we will refer to as the 'and in fact'), there are clear instances of another *in fact*, a 'for in fact' (Aijmer and Vandenberg 2004: 1788)'

They also argue that the adversative ('but') type, the additive ('and') type, and the enhancing ('for') type are "pragmatic implicatures which have become 'preferred meanings' through frequent occurrence".

#### 6.1.1.2 *Spanish translation correspondences of 'in fact'*

When analyzing the translations of 'in fact' in the Spanish translation corpora used in this study, the equivalents found are manifold, which indicates that this DM is highly polysemic. Their frequencies, both absolute and per million words (in parentheses) are displayed in table 7 below:

SPANISH Translation In fact	Corpus frequency								
	Corpus total Number of Words	Europarl 7 English 53,837,625	Open subtitles 1,000,000	Emea 5,455,537	Multinot fiction 10,459	Multinot essay 18,538	Multinot Expe 40,643	Multinot Speech 36,675	Multinot Web 16,553
<b>Realmente</b>	235 (3.87)	37 (0.05)	<b>10 (0.65)</b>	0	0	0	0	0	282 (4.57)
<b>En realidad</b>	226 (3.72)	<b>274 (0.34)</b>	2 (0.13)	0	1 (53.94)	<b>5 (123.02)</b>	<b>1 (27.27)</b>	0	510 (208.42)
<b>De hecho</b>	<b>429 (7.06)</b>	219 (0.27)	1 (0.06)	0	<b>4 (215.77)</b>	3 (73.81)	<b>1 (27.27)</b>	0	<b>657 (350.97)</b>
En los hechos	5 (0.08)	0	0	0	0	0	0	0	5 (0.08)
Verdaderamente	51 (0.84)	1 (<0.01)	0	0	0	0	0	0	51 (0.84)
De verdad	12 (0.2)	8 (0.01)	0	0	0	0	0	0	20 (0.21)
En verdad	12 (0.2)	18 (0.02)	0	0	0	0	0	0	30 (0.22)
La verdad	15 (0.25)	21 (0.03)	0	0	0	0	0	0	36 (0.28)
Efectivamente	409 (6.73)	12 (0.01)	0	0	0	1 (24.60)	0	0	422 (31.34)
En efecto	295 (4.86)	24 (0.03)	0	0	0	0	0	0	319 (4.89)
Por ejemplo	119 (1.96)	0	0	0	0	0	0	0	119 (1.96)
A saber	37 (0.61)	1 (<0.01)	0	0	0	0	0	0	37 (0.61)
Ya	242 (3.98)	52 (0.06)	0	0	0	0	0	0	294 (4.04)
Puesto que	99 (1.63)	0	0	0	0	0	0	0	99 (1.63)
Pues	194 (3.19)	11 (0.01)	0	0	0	0	0	0	205 (3.20)
Justamente	45 (0.74)	3 (<0.01)	0	0	0	0	0	0	48 (0.74)
Indudablemente	17 (0.28)	1 (<0.01)	0	0	0	0	0	0	18 (0.28)
Sin duda	68 (1.12)	4 (<0.01)	0	0	0	0	0	0	72 (1.12)
Ciertamente	53 (0.87)	1 (<0.01)	0	0	0	0	0	0	54 (0.87)
Lamentablemente	24 (0.4)	4 (<0.01)	0	0	0	0	0	0	28 (0.4)
Sinceramente	10 (0.16)	0	0	0	0	0	0	0	10 (0.16)
En el fondo	36 (0.59)	2 (<0.01)	0	0	0	0	0	0	38 (0.59)
En absoluto	35 (0.58)	3 (<0.01)	0	0	0	0	0	0	37 (0.58)
Naturalmente	58 (0.95)	0	0	0	0	0	0	0	58 (0.95)
Fundamentalmente	14 (0.23)	0	0	0	0	0	0	0	14 (0.23)
Supuestamente	19 (0.31)	7 (0.01)	0	0	0	0	0	0	26 (0.32)
Evidentemente	27 (0.44)	0	0	0	0	0	0	0	27 (0.44)
Obviamente	16 (0.26)	1 (<0.01)	0	0	0	0	0	0	17 (0.26)
Por supuesto	43 (0.71)	0	1 (<0.01)	0	0	0	0	0	44 (0.71)
Habida cuenta	23 (0.38)	0	0	0	0	0	0	0	23 (0.38)
Tal vez	25 (0.41)	0	0	0	0	0	0	0	25 (0.41)
Aparentemente	17 (0.28)	7 (0.01)	2 (0.13)	0	0	0	0	0	26 (0.42)
Precisamente	211 (3.74)	9 (0.01)	0	0	0	0	0	0	220 (3.75)
Por cierto	31 (0.51)	10 (0.01)	0	0	0	0	0	0	41 (0.52)
Finalmente	32 (0.53)	1 (<0.01)	0	0	0	0	0	0	33 (0.53)
Al final	27 (0.44)	0	0	0	0	0	0	0	27 (0.44)
Así	176 (2.9)	65 (0.08)	0	0	0	0	0	0	241 (2.98)
Pues	194 (3.19)	11 (0.01)	0	0	0	0	0	No entries	205 (3.20)
Ni siquiera	37 (0.61)	3 (<0.01)	0	0	0	0	0	0	40 (0.61)
Exactamente	33 (0.54)	16 (0.02)	0	0	0	0	0	0	49 (0.56)
Prácticamente	36 (0.59)	1 (<0.01)	0	0	0	0	0	0	37 (0.62)
Es cierto que	31 (0.51)	7 (0.01)	0	0	0	0	0	0	38 (0.56)
Concretamente	45 (0.47)	0	0	0	0	0	0	0	45 (0.47)
En concreto	30 (0.49)	1 (<0.01)	0	0	0	0	0	0	31 (0.49)
Especialmente	97 (1.6)	2 (<0.01)	0	0	0	0	0	0	99 (1.6)
En especial	19 (0.31)	0	0	0	0	0	0	0	19 (0.31)
Tan sólo	17 (0.28)	1 (<0.01)	0	0	0	0	0	0	18 (0.28)
En definitiva	30 (0.49)	0	0	0	0	0	0	0	30 (0.49)
Es decir	107 (1.76)	2 (<0.01)	0	0	0	0	0	0	109 (1.76)
Sea	25 (0.41)	0	0	0	0	0	0	0	25 (0.41)
Mejor dicho	20 (0.33)	0	0	0	0	0	0	0	20 (0.33)
Bueno	342 (5.63)	85 (0.01)	0	0	0	0	0	0	427 (5.64)
(más bien)	193 (3.18)	6 (0.01)	0	0	0	0	0	0	199 (3.19)
Tal como	60 (0.99)	0	0	0	0	0	0	0	60 (0.99)
Actualmente	70 (1.15)	1 (<0.01)	0	<b>1 (95.611)</b>	0	0	0	0	71 (1.15)
<b>Es más</b>	49 (0.81)	15 (0.02)	0	0	0	0	0	0	65 (96.44)
Sobre todo	124 (2.04)	1 (<0.01)	0	0	0	0	0	0	125 (2.04)
En la práctica	71 (1.17)	1 (<0.01)	0	0	0	0	0	0	72 (1.17)
Lo cierto es que	21 (0.35)	2 (<0.01)	0	0	0	0	0	0	23 (0.35)
Más que	106 (1.75)	21 (0.03)	0	0	0	0	0	0	127 (1.78)
Por lo demás	76 (1.25)	0	0	0	0	0	0	0	76 (1.25)
Dado que	63 (1.04)	0	0	0	0	0	0	0	63 (1.04)
Acaso	19 (0.31)	0	0	0	0	0	0	0	19 (0.31)
Todavía	96 (1.58)	11 (0.01)	0	0	0	0	0	0	107 (1.59)
Aún	153 (2.52)	18 (0.02)	0	0	0	0	0	0	171 (2.54)
En ese/este sentido	42(0.51)	1 (<0.01)	0	0	0	0	0	0	43 (0.51)
Suficientemente	24 (0.40)	1 (<0.01)	0	0	0	0	0	0	25 (0.40)
Pero	305 (5.02)	208 (0.25)	0	0	0	0	0	0	24 (0.40)
Por el contrario	43 (0.71)	0	0	0	0	0	0	0	513 (5.27)
Al contrario	42 (0.69)	0	0	0	0	0	0	0	43 (0.71)
Sin embargo	85 (1.40)	3 (<0.01)	4 (0.26)	0	0	0	0	0	92 (1.66)
En cambio	21 (0.35)	2 (<0.01)	0	0	0	0	0	0	23 (0.35)
No obstante	19 (0.31)	0	0	0	0	0	0	0	19 (0.31)
Apenas	27 (0.44)	0	0	0	0	0	0	0	27 (0.44)

Por lo demás	76 (1.25)	0	0	0	0	0	0	19 (0.31)
Solamente	71 (1.17)	2 (<0.01)	0	0	0	0	0	27 (0.44)
Aunque	238 (3.92)	42 (0.05)	0	0	0	0	0	76 (1.25)
A pesar de	89 (1.47)	0	0	0	0	0	0	73 (1.17)
Por consiguiente	24 (0.40)	0	0	0	0	0	0	280 (3.97)
También	275 (4.53)	24 (0.03)	0	0	0	0	0	89 (1.47)
Tal	126 (2.07)	8 (0.01)	0	0	0	0	0	24 (0.40)
Por tanto	76 (1.25)	0	0	0	0	0	0	296 (4.56)
Por ello	37 (0.61)	0	0	0	0	0	0	134 (2.08)
Además	195 (3.21)	1 (<0.01)	0	0	0	0	0	76 (1.25)
Incluso	231 (3.80)	13 (0.02)	4 (0.26)	0	0	0	0	37 (0.61)
<b>En rigor</b>	<b>1 (0.02)</b>	0	0	0	0	0	0	196 (3.21)
								248 (4.08)
								<b>1 (0.02)</b>
	<b>7,406</b>	1,306	20	1	5	9	2	<b>8,749</b>
	(124.59)	(1.44)	(2.06)	(95.611)	<b>(269.71)</b>	(221.43)	(54.54)	<b>(769.381)</b>

**Table 7: Translation correspondences of 'in fact' in Spanish translation corpora.**

As shown in table 7, the most frequent translation equivalent (Hummel, 1995) for 'in fact' is '**de hecho**' according to the column 'total' with the following proportion: 657 (350.97). Consequently, the answer to research question two for 'in fact' would be that its preferred translation equivalent (Hummel, 1995) in Spanish is 'de hecho', being the most frequent translation equivalent in the corpus Europarl7 (Koehn, 2001) with a proportion of 429 (7.06), in Multinot Essay and speech (Lavid et al., 2015) with a proportion of 4 (215.77) and 1 (27.27) respectively.

This is followed by '**en realidad**' 510 (208.42), this being the most frequent translation equivalent (Hummel, 1995) in the corpus OpenSubtitles (Tiedemann, 2007) with a proportion of 274 (0.34) in this corpus. It is also the most frequent translation in Multinot Expe and Speech (Lavid et al., 2015) with a proportion of 5 (123, 2) and 1 (27, 27) respectively.

'Realmente' is the most frequent translation equivalent (Hummel, 1995) for 'in fact' in the corpus EMEA (Tiedemann, 2007) with a proportion of 10 (0.55), but it is important to mention that the column 'total' does not show 'realmente' as the translation equivalent (Hummel, 1995) for 'in fact' following in proportion to 'de hecho' and 'en realidad'. This issue will be discussed in the following sections.

‘Es más’ (95.611) is the most frequent translation equivalent (Hummel, 1995) for ‘in fact’ only in the corpus Multinot Fiction (Lavid et al., 2015). But it is important to mention, as in the case of ‘realmente’, that the column ‘total’ does not show ‘es más’ as the translation equivalent for ‘in fact’ following in proportion to ‘de hecho’ and ‘en realidad’. This issue will be discussed in the following sections.

According to the column ‘total’, the translation equivalents (Hummel, 1995) which would follow ‘de hecho’ and ‘en realidad’ are the following:

- a) Efectivamente 422 (31,34): ‘operador de refuerzo argumentativo’ explicative (Portolés and Zorraquino, 1999).
- b) En efecto 319 (4.89): ‘operador de refuerzo argumentativo’ explicative (Portolés and Zorraquino, 1999).
- c) También 296 (4.56): consecutive connector (Portolés and Zorraquino, 1999).
- d) Ya 294 (4.04): metadiscursive connector (Portolés and Zorraquino, 1999).
- e) Aunque 280 (3.97): contra-argumentative connector adversative (Pons Bordería, 2005).

These translations correspondences confirm the most central meanings of ‘in fact’ pointed out in previous studies. Thus, the translation of ‘in fact’ by ‘de hecho’ and ‘efectivamente’- ‘en efecto’ point to the central meaning of ‘in fact’ proposed by Aijmer and Vandenberg (2004), i.e.: “to signal the speaker’s attitude that something is in reality or in truth the case.”

With respect to the other translation equivalents (Hummel, 1995) the following functions are to be noted:

‘En realidad’ and ‘realmente’: Portolés and Zorraquino (1999) consider ‘en realidad’ and ‘realmente’ into the category of argumentative reinforcement, which reinforces the premise of the functional closeness between English VM’s (Halliday, 2014) and Spanish ‘Operadores de refuerzo argumentativo’ (Portolés and Zorraquino, 1999).

Moreover, El Meassaudi Rifi (2018), points out correction as a function to be addressed to the maker ‘en realidad’. This function is additional to those addressed by Halliday (2014) and Aijmer (2004) and confirm the premise that through parallel corpora translation it is possible to encounter additional functions for VMs.

Other authors such as Violeta de Monte (1999), add some additional comments with respect to the so called ‘operadores de refuerzo argumentativo’ and considers that ‘en realidad’ and ‘realmente’ emphasize the contrast between the truth of the statement where it appears. Emphasis is mentioned by Aijmer (2004) as enhancement, though confirming the centrality of ‘in fact’ among other VMs (Halliday, 2014).

‘Es más’:

Cortés and Camacho (2005), see ‘es más’ with the function of reformulation-enhancement, which confirm Aijmer’s (2004) enhancement function:

<s> It was a shock, **in fact** , to me </s> OPUS 2- OPEN SUBTITLES (Tiedemann, 2007.

<s> **Es más**, para mí fue un golpe </s> OPUS 2- OPEN SUBTITLES (Tiedemann, 2007.

With respect to ‘ya’, ‘aunque’ and ‘también’ they add new functions to the VMs (Halliday, 2014) ‘in fact’ as mentioned before: adversative (also suggested by Aijmer, 2004):

Informative about the ending of the action described in S1:

<s> Several goals had **in fact** been set under the Portuguese Presidency, in Lisbon and Feira. </s> Europarl7 (Koehn, 2001).

<s> se han tomado **ya** varias citas durante la Presidencia portuguesa, en Lisboa y en Feira. </s> Europarl7 (Koehn, 2001).

Informative about the continuity of the action started in S1 as well as adversative:

<s> This is the final round of the war of accountants, it is **in fact** still on. </s> Europarl7 (Koehn, 2001).

<s> Es la última ronda de la guerra de los contables, **aunque** aún continúa. </s> Europarl7

(Koehn, 2001).

<s> This last aspect is, **in fact** , an unambiguous indication of commitment and political determination. </s> Europarl7 (Koehn, 2001).

<s> **también** este último aspecto es una señal de identidad y de determinación política de interés unívoco. </s> Europarl7 (Koehn, 2001).

Likewise, the Spanish translations also show the **three pragmatic uses** of in fact pointed out by Aijmer and Vanderbergen (2004):

- a) Those equivalents which express a contrast (i.e., equivalents of the 'but' type of in fact):

This is represented by the Spanish translations of 'in fact' by the 'en realidad', often collocating with the 'pero' in Spanish, thus showing the same pattern as in English, as illustrated here with 'aunque' too.

<s> One very much better, **in fact** . </s> OPUS 2- OPEN SUBTITLES (Tiedemann, 2007).

<s> Una mucho mejor, **en realidad**. </s> OPUS 2- OPEN SUBTITLES (Tiedemann, 2007).

<s> This is the final round of the war of accountants, it is **in fact** still on. </s> Europarl7 (Koehn, 2001).

<s> Es la última ronda de la guerra de los contables, **aunque** aún continúa. </s> Europarl7 (Koehn, 2001).

- b) Those which express a stronger reformulation (i.e., equivalents of the 'and' type), This is represented by the Spanish translations of 'in fact' by the 'es más', as illustrated here.

<s> It was a shock, **in fact** , to me </s> OPUS 2- OPEN SUBTITLES (Tiedemann, 2007).

<s> **Es más**, para mí fue un golpe </s> OPUS 2- OPEN SUBTITLES (Tiedemann, 2007).

- c) And those which express a reason (i.e., equivalents of the 'for' type). As illustrated with the beyond example.

<s> To present them to Parliament as a single package would, in my view, be extremely unwise since they are **in fact** separate subjects. </s> (Europarl7, Koehn, 2001).

<s> Sería muy desacertado, a mi parecer, presentarlos al Parlamento dentro de un mismo paquete, **puesto que** se trata de dos temas separados. </s> (Europarl7, Koehn, 2001).

With respect to the least frequent translation equivalents (Hummel, 1995) of ‘in fact’, these can be grouped following the three pragmatic uses of in fact pointed out by Aijmer and Vanderbergen (2004): a) Contrast between S1 and S1, usually expressed as adversative meaning. b) Reinforcement: or enhancement, which aims at emphasizing S1 in S2. c) Addition: which tries to provide more information or characteristics of S1 in S2. d) Reason: which explains or justifies the truthfulness of S1 in S2.

Contrast:	Reinforcement	Addition	Reason
Pero			
Sino			
Por el contrario	Sobre todo		
Al contrario	En la práctica	Además	
Sin embargo	Es sabido que	A su vez	Puesto que
En cambio	Lo cierto es que	Incluso	Pues
No obstante	Ahora bien	De paso	
Apenas	Por lo demás		
Por lo demás	Dado que		
Solamente			
A pesar de			

**Table 8: Additional examples of ‘in fact’ translation equivalents (Hummel, 1995) in pragmatic functions (Aijmer, 2004).**

6.1.1.3 *Semantic groupings of ‘in fact’: Lexical domains*

The translation preferences of ‘in fact’ and the contexts where they occur can be grouped into translation lexical domains (TLDs), as graphically shown in Figure 11 below, the translation occurrences are grouped according to the context function they perform for ‘in fact’:

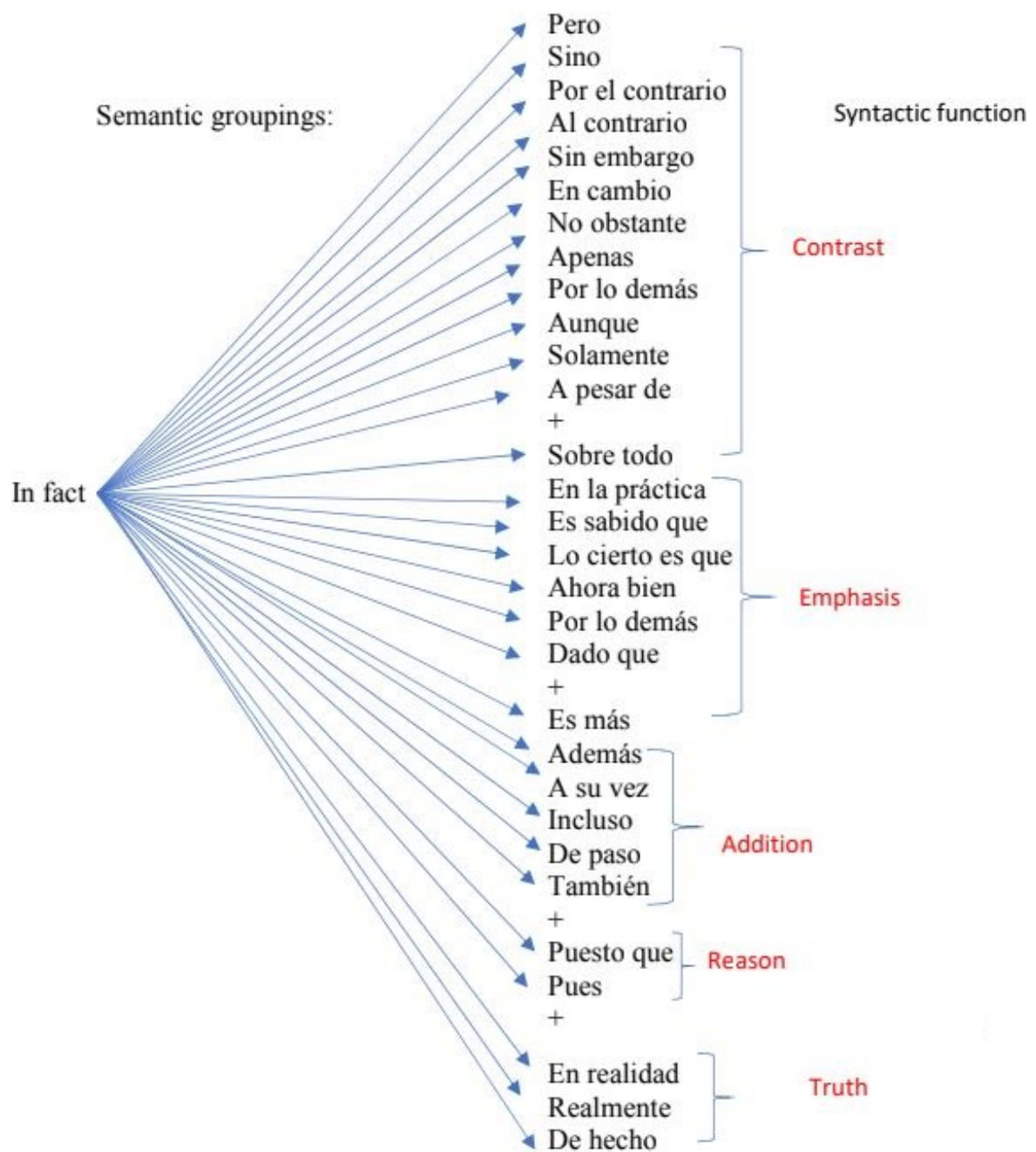


Figure 11: ‘in fact’ TLDs

#### 6.1.1.4 Genre preferences ‘in fact’

With respect to the genre-specific preferences, the corpus analysis yields the following results:

- 1) ‘De hecho’ is the most frequent translation equivalent (Hummel, 1995) in Europarl7 (Koehn, 2001), Multinot Essay and Speech (Lavid et al., 2015) (7.06, 215.77 and 27.27 per million respectively).
- 2) ‘En realidad’ is the most frequent translation equivalent (Hummel, 1995) in the Open Subtitles (Tiedemann, 2007) corpus, Multinot- Expe and Speech (Lavid et al., 2015) (0.34, 121.02 and 27.27 per million respectively).
- 3) ‘Realmente’ is the most frequent translation equivalent (Hummel, 1995) in EMEA (Tiedemann, 2007) (0.65 per million).
- 4) ‘Es más’ is the most frequent translation equivalent (Hummel, 1995) in Multinot Fiction (Lavid et al., 2015) (95.611 per million).

## 6.1.2 ‘Actually’

### 6.1.2.1 *Previous studies*

As in the case of ‘in fact’, the item ‘actually’ has been studied by Aijmer and Vandenberg (2004), as well as by authors such as Oh (2000), Smith and Jucker (2000) and Powell (1992, among others). These authors distinguish several functions: Smith and Jucker’s study focus is on how conversationalists use such expressions to introduce repairs to the common ground. They refer to these items as “markers of an apparent discrepancy between propositional attitudes of conversational partners” (2000: 207).

Powell’s analysis shows that really, actually, among other adverbs, have developed “**highly emphatic uses**” (1992: 101) and concludes that all these words have “great practical utility for a speaker who wishes to stipulate and thereby control the ways in which a message is to be understood and accepted” (1992: 105).

Aijmer and Vandenberg (2004) emphasize the importance of rhetorical functions of these items over and above purely epistemic ones and indicate “any attempt to plot markers in a lexical field must finally reckon with more central elements (those which signal expectation more saliently or more frequently) versus more marginal ones, some elements being closer together than others”.

Halliday (2014) considers ‘actually’ a conjunction which conveys a function of restatement verifying the clarification of a characteristic, a comment or a circumstance (Halliday 2014: pg 612). Furthermore, Quirk et al (1990), consider ‘actually’ as a connective providing addition in the form of emphasis and confirmation. ‘Actually’ connects S1 with S2 by means of adding a comment or characteristic of S1 in S1 emphasizing its relevance or confirming its truth (Quirk, 1990).

Seyed Ali et al (2012), consider ‘actually’ as a discourse connector and sees three possible functions for ‘actually’, contrast-concessive, emphasis and corroboration. Which means that, for the former, ‘actually’ would provide a contrast by means of expressing a condition or circumstance of S1 in S2; for the second function, a comment or characteristic or the truth of S1 will be enhanced on S2;

the latter is close to Halliday's (2014) idea of verification, a characteristic of S1 or the truth of S1 is confirmed in S1 (Seyed Ali et al, 2012).

### 6.1.2.2 *Spanish translation correspondences of 'actually'*

The following table shows the absolute frequency and the frequency per million, in parenthesis, of occurrence of each translation equivalent in each of the corpora under analysis. Accordingly, 'Actually' in the Spanish translation corpora shows that the equivalents are centered around the following items: '**realmente**' (addition of the figures in each corpora: 575 (520.711) is the most frequent translation followed by '**en realidad**' (addition of the figures in each corpora: 765 (115.51)), '**de verdad**' (addition of the figures in each corpora: 589 (99.021)), '**de hecho**' (addition of the figures in each corpora: 902 (11.41)) and '**así**' (addition of the figures in each corpora: 1052 (4.89)) , as shown in table 9 (checking the values per million in each column).

SPANISH Translation Actually	Corpus frequency								TOTAL
	Corpus total Number of Words	Europarl 7 English 53,837,625	Open subtitles 1,000,000	Emea 5,455,537	Multinot fiction 10,459	Multinot essay 18,538	Multinot Expe 40,643	Multinot Speech 36,675	
<b>Realmente</b>	320 (5.27)	227 (0.98)	<b>10 (0.65)</b>	<b>1 (95.611)</b>	0	<b>17 (418.27)</b>			<b>575 (520.781)</b>
<b>En realidad</b>	397 (6.54)	364 (0.45)	2 (0.13)	0	<b>2 (107.9)</b>	3 (73.81)			<b>768 (188.83)</b>
En la realidad	24 (0.4)	11 (0.01)	0	0	1 (53.94)	0			36 (54.35)
<b>De hecho</b>	<b>600 (9.98)</b>	301 (0.37)	1 (0.06)	0	0	0			<b>902 (10.41)</b>
En los hechos	3 (0.05)	25 (0.02)	0	0	1 (53.94)	0			29 (54.01)
Verdaderamente	183 (3.01)	45 (0.06)	0	0	0	0			228 (3.07)
<b>De verdad</b>	176 (2.9)	412 (0.51)	0	<b>1 (95.611)</b>	0	0			<b>589 (99.021)</b>
A decir verdad	4 (0.07)	54 (0.07)	0	0	0	0			58 (0.14)
En verdad	28 (0.46)	249 (0.31)	0	0	0	0			277 (0.77)
La verdad	33 (0.54)	200 (0.25)	0	0	0	2 (49.2)			235 (49.99)
Efectivamente	456 (7.51)	24 (0.03)	0	0	0	0			480 (7.54)
En efecto	118 (1.94)	18 (0.02)	0	0	0	0			136 (1.96)
Por ejemplo	174 (2.86)	0	0	0	0	0			174 (2.86)
A saber	52 (0.86)	2 (<0.01)	0	0	0	0			54 (0.87)
Ya	205 (3.37)	210 (0.26)	0	0	0	0			215 (3.63)
Puesto que	105 (1.73)	1 (<0.01)	0	0	0	0			106 (1.74)
Pues	214 (3.52)	223 (0.27)	0	0	0	0			437 (3.79)
Justamente	29 (0.48)	19 (0.02)	0	0	0	0			48 (0.5)
Sin duda	65 (1.07)	2 (<0.01)	0	0	0	0			67 (1.08)
Ciertamente	43 (0.71)	9 (0.01)	0	0	0	0			52 (0.72)
Lamentablemente	32 (0.53)	7 (0.01)	0	0	0	0			39 (0.54)
En el fondo	36 (0.59)	15 (0.02)	0	0	0	0			51 (0.61)
En absoluto	58 (0.95)	10 (0.01)	0	0	0	0			68 (0.96)
Naturalmente	48 (0.79)	3 (<0.01)	0	0	0	0			51 (0.80)
Evidentemente	26 (0.43)	1 (<0.01)	0	0	0	0			27 (0.45)
Por supuesto	57 (0.94)	0	0	0	0	0			57 (0.94)
Tal vez	48 (0.79)	0	0	0	0	0			48 (0.79)
Precisamente	167 (2.75)	35 (0.04)	0	0	0	0			202 (2.79)
Por cierto	25 (0.41)	32 (0.01)	0	0	0	0			57 (0.42)
Finalmente	78 (1.28)	33 (0.04)	0	0	0	0			111 (1.32)
Al final	57 (0.94)	0	0	0	0	0			57 (0.94)
<b>Así</b>	236 (3.89)	<b>816 (1.00)</b>	0	0	0	0			<b>1,052 (4.89)</b>
Ni siquiera	44 (0.72)	50 (0.06)	0	0	0	0			94 (0.78)
Exactamente	74 (1.22)	64 (0.08)	0	0	0	0			138 (1.30)
Es cierto que	39 (0.64)	32 (0.04)	0	0	0	0			71 (0.68)
Concretamente	60 (0.99)	8 (0.01)	0	0	0	0			68 (1.00)
En concreto	44 (0.72)	0	0	0	0	0	No entries	No entries	44 (0.72)
Especialmente	104 (1.71)	18 (0.02)	3 (0.19)	0	0	0			125 (1.92)
En definitiva	19 (0.31)	21 (0.01)	0	0	0	0			40 (0.32)
Es decir	166 (2.73)	10 (0.01)	0	0	0	0			176 (2.74)
sea	21 (2.73)	0	0	0	0	0			21 (2.73)
Bueno	94 (1.55)	204 (0.25)	0	0	0	0			298 (1.80)
(más bien)	261 (4.3)	64 (0.08)	2 (0.01)	0	0	0			327 (4.39)
Aún	177 (2.91)	147 (0.18)	0	0	0	0			324 (3.09)
Tal como	64 (1.05)	0	0	0	0	0			64 (1.05)
Actualmente	134 (2.21)	202 (0.25)	0	0	0	0			336 (2.46)
Es más	49 (0.81)	75 (0.09)	0	0	0	0			124 (0.90)
Sobre todo	118 (1.94)	6 (0.01)	0	0	0	0			124 (1.95)
En la práctica	169 (2.78)	3 (<0.01)	0	0	0	0			172 (2.79)
Por lo demás	13 (0.21)	75 (0.09)	0	0	0	0			88 (0.30)
Dado que	54 (0.89)	0	0	0	0	0			54 (0.89)
Acaso	18 (0.3)	28 (0.03)	0	0	0	0			46 (0.33)
Todavía	134 (2.21)	129 (0.16)	0	0	0	0			263 (2.37)
En este sentido	39 (0.66)	0	0	0	0	0			39 (0.66)
Suficientemente	38 (0.63)	18 (0.02)	0	0	0	0			56 (0.65)
Pero	370 (6.09)	241 (0.3)	0	0	0	0			611 (6.12)
Por el contrario	37 (0.061)	0	0	0	0	0			37 (0.061)
Al contrario	212 (0.2)	0	0	0	0	0			212 (0.20)
Sin embargo	105 (1.73)	14 (0.02)	4 (0.26)	0	0	0			203 (1.99)
No obstante	36 (0.59)	0	0	0	0	0			36 (0.59)
Apenas	32 (0.53)	20 (0.02)	0	0	0	0			52 (0.55)
A pesar de que	130 (2.14)	0	0	0	0	0			130 (2.14)
Solamente	113 (1.86)	19 (0.02)	0	0	0	0			132 (1.88)
Aunque	332 (5.47)	194 (0.24)	02 (0.13)	0	0	0			528 (5.84)
También	298 (4.91)	12 (0.01)	0	0	0	0			210 (4.92)
Tal	132 (2.17)	51 (0.06)	0	0	0	0			183 (2.23)
Por tanto	66 (1.09)	11 (0.01)	0	0	0	0			77 (1.10)
Por ello	41 (0.67)	5 (0.01)	0	0	0	0			46 (0.68)
Además	234 (3.85)	11 (0.01)	0	0	0	0			245 (3.86)
Incluso	326 (5.37)	98 (0.12)	4 (0.26)	0	0	0			428 (5.75)
Desde luego	23 (0.38)	0	0	0	0	0			23 (0.38)
En serio	40 (0.66)	115 (0.14)	0	0	0	0			155 (0.80)
De veras	6 (0.1)	47 (0.06)	0	0	0	0			53 (0.16)
									34 (0.42)

Propiamente	15 (0.25)	2 (<0.01)	0	0	0	0	0	No entries	17 (0.26)
Literalmente	9 (0.15)	18 (0.02)	0	0	0	0	0		27 (0.17)
La realidad	10 (0.30)	17 (0.02)	0	0	0	0	0		98 (1.13)
Lo cierto	18 (0.80)	14 (0.02)	0	0	0	1 (24.60)	0		28 (25.92)
Aparentemente	120 (1.98)	68 (0.08)	0	0	0	1. (24.60)	0		33 (24.92)
Más que			0	0	0	1 (24.60)	0		189 (26.66)
	8,826 (144.221)	5581 (7.6602)	28 (1.69)	2 (181.22)	4 (215.78)	25 (615.09)			14,466 (1,165.6612)

**Table 9: Translation correspondences of ‘Actually’ in Spanish translation corpora.**

The figures commented before, table 9, show discrepancies between absolute results and proportion per million. According to the absolute figures the frequency of preference in translation equivalents (Hummel, 1995) for ‘actually’ is the following (from the most frequent to the least): ‘así’, ‘de hecho’, ‘en realidad’, ‘de verdad’, ‘realmente’.

However, according to the figures per million the frequency of preference in translation equivalents (Hummel, 1995) for ‘actually’ is the following (from the most frequent to the least): ‘realmente’, ‘en realidad’, ‘de verdad’, ‘de hecho’, ‘así’. As each corpus has a different number of words, the focus will be put in the proportion per million as the point of reference, a million of words, is the same for all the corpora.

### 6.1.2.3 Examples of translation equivalents for ‘actually’ based on table 9

These most frequent translation equivalents (Hummel, 1995) of ‘actually’, i.e., ‘realmente’, ‘de hecho’, ‘de verdad’, ‘en realidad’ point to its central meaning as an epistemic marker, very close to the central meaning of ‘in fact’:

1) Statement of truth: <s> But his portfolio of information technology and tourism **actually** go very well together. </s> EUROPARL7.

<s> Pero sus carteras de informática y turismo, **de hecho** , casan muy bien juntas. </s> EUROPARL7.

2) Reinforcement of truth: <s> What I learned is that, **actually**, there is sufficient legislation in place but the problem is the lack of supervision. </s> EUROPARL7.

<s> Lo que deduje de todo aquello es que **en realidad** hay suficiente regulación, pero el problema estriba en que no se controla. </s> EUROPARL7.

3) Addition and reinforcement of true qualities: <s> FOSAVANCE not only prevents the loss of bone but **actually** helps to rebuild bone you may have lost and reduces the risk of bones breaking in the spine and hip. </s> OPUS 2- EMEA.

<s> FOSAVANCE no sólo previene la pérdida de hueso sino que **realmente** ayuda a reconstruir el hueso que puede haber perdido y reduce el riesgo de fractura de huesos en la columna vertebral y en la cadera. </s> OPUS 2- EMEA.

Other two translation equivalents (Hummel, 1995) to be mentioned for ‘actually’ as found in parallel corpora:

1) De verdad: enhancement:

<s>What **actually** happens when you die is that your brain stops working and your body rots, like Rabbit did when he died and we buried him in the earth at the bottom of the garden. </s> Multinot fiction (Lavid et al., 2015).

<s> Lo que **de verdad** pasa cuando te mueres es que tu cerebro deja de funcionar y el cuerpo se pudre, como el de Conejo cuando se murió y lo enterramos al fondo del jardín. </s> Multinot fiction (Lavid et al., 2015).

The function here is strongly emphatic, in the English source the sentence starts with a wh-pronoun, whose equivalent is the combination of a definite article ‘lo’ and an interrogative pronoun ‘que’; both ‘what’ and ‘lo que’ are substituting something previously said or written in the discourse/speech. The use of ‘actually’ and ‘de verdad’ draws the reader’s/ listener’s attention towards the relevance of the explanation provided afterwards.

Moreover, ‘de verdad’ is a ‘marcador de refuerzo argumentativo’ (Portolés and Zorraquino, 1999), a fact that illustrates the connection between English VMs (Halliday, 2014) and Spanish ones.

The case of ‘de verdad’:

- a) Pons and Bordería (2005) would point out that ‘de verdad’, due to its contrastive nature in this sentence, has an anti-oriented argumentative function.
- b) Correction (El Messaoudi Rifi, 2018) and emphasis of truth or factuality (Violeta de Monte, 1999) are functions to be considered here for ‘actually’ and ‘de verdad’.

2) Así: Purpose-Explanation:

<s> **Actually**, no. </s> Tiedemann (2007).

<s> **Así**, no. </s> Tiedemann (2007).

‘Así’ is an adverb, which is used here as part of a utterance to explain how it should not be done what has previously been mentioned. Adverb is one of the functions suggested for ‘actually’ in monolingual analysis, epistemic adverb indeed. Here the main functions are correction and illustration.

The case of ‘así’:

- a) Portolés and Zorraquino (1999) consider for ‘así’ the function of consecutive connector, this is, expressing a consequence or result.
- b) Cuenca (2001) points out the function of exemplifying reformulation for ‘Así’,
- c) Núñez et al (2006) consider reformulation too, but with a function of recapitulation for ‘así’.
- d) Cortés y Camacho (2005) consider reformulation too, but with a function of concretion for ‘así’.

Based on the previous analysis, more specially on the frequency occurrence table which introduces this section, the most frequent translation equivalents (Hummel, 1995) for ‘actually’ are ‘realmente’, ‘en realidad’, ‘de verdad’, ‘de hecho’ and ‘así’; ordered from the most preferred translation equivalent for ‘actually’, ‘realmente’, to the least.

‘En realidad’ is the most frequent translation equivalent for ‘actually’ in the corpus Multinot essay (Lavid et al., 2015), ‘de verdad’ is the most frequent translation equivalent (Hummel, 1995) for ‘actually’ in the corpus Multinot Fiction (Lavid et al., 2015), ‘de hecho’ is the most frequent translation equivalent for ‘actually’ in the corpus Europarl7 (Koehn, 2001) and ‘así’ is the most frequent translation equivalent (Hummel, 1995) for ‘actually’ in the corpus Opensubtitles (Tiedemann, 2007).

*What is the preferred translation equivalent for ‘actually’ in Spanish after the analysis of parallel corpora?* ‘**realmente**’ is, proportionally speaking, the preferred translation equivalent (Hummel, 1995) for ‘actually’ as found in the analysis of its equivalents in parallel corpora.

Analysis of other frequent translations for ‘actually’:

Taking into consideration the pragmatic functions pointed out by Aijmer (2004) addition, contrast, reinforcement and result; the least frequent translations for ‘actually’ will be shown and grouped in table 10 bellow:

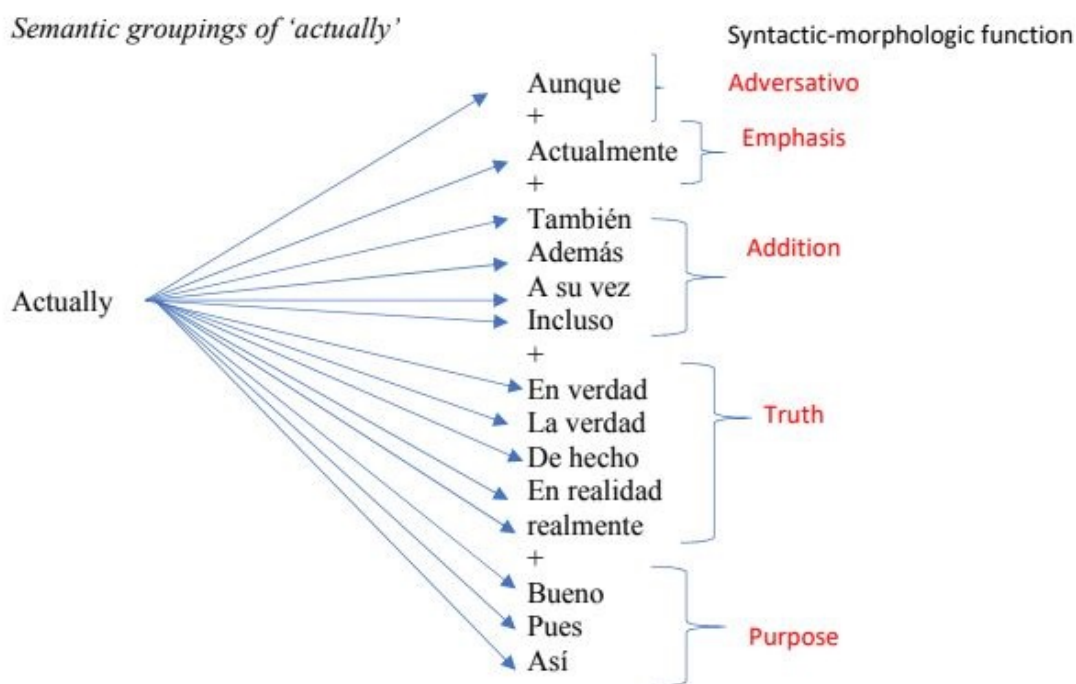
- a) Contrast: between S1 and S1, mainly adversative meaning.
- b) Reinforcement: emphasis about some element of S1 in S2.
- c) Addition: of information about S1 in S2.
- d) Truth: truthfulness of S1 defined in S2.
- e) Reason: justification or explanation of some S1 in S2.

Contrast:	Reinforcement	Addition	Truth	Reason
Aunque	Actualmente	También Además A su vez Incluso	En verdad La verdad	Bueno Pues

**Table 10: Pragmatic additional examples for ‘actually’.**

#### 6.1.2.4 Lexical domains ‘actually’

The following figure illustrates the ‘translation lexical domains’, in this case, this refers to the translation equivalent for ‘actually’ according to the function developed in the discourse of speech act:



**Figure 12: ‘actually’ TLDs**

#### 6.1.2.5 Genre-specific preferences of ‘actually’

With respect to the genre-specific preferences, the corpus analysis yields the following results:

- 1) ‘De hecho’ is the most frequent translation equivalent in Europarl7 (Koehn, 2001): 9.98 per million.
- 2) ‘En realidad’ is the most frequent translation equivalent in Multinot essay and speech (Lavid et al., 2015): 107.9 per million.
- 3) ‘Realmente’ is the most frequent translation equivalent for ‘actually’ in EMEA (Tiedemann, 2007), Multinot Fiction and expe (Lavid et al., 2015): 0.65 per million, 95.611 per million, 418.27 per million, respectively.

- 4) 'De verdad' is the most frequent translation equivalent for 'actually' in Multinot Fiction (Lavid et al., 2015): 95.611 per million.
- 5) 'Así' is the most frequent translation equivalent for 'actually' in Open Subtitles (Tiedemann, 2007) 1 per million.

### **6.1.3 'Indeed'**

#### **6.1.3.1 *Previous studies***

Halliday categorises (2014) 'indeed' as either a conjunction; inside elaboration as a verificative type of clarification, or as an adverbial conjunct. It conveys a function of restatement verifying the clarification of a characteristic, a comment or a circumstance (Halliday 2014).

This author also points out that 'some of the items that are used as conjunctions with verificative senses also have assessment senses and serve as mood or comment adjuncts.' This means that 'indeed' as an adverbial can also comment or express a circumstance.

Other authors consider 'indeed' as a discourse connector and mention two possible functions for 'indeed', appositive, emphasis and corroboration (Seyed Ali, et al, 2012). Which means that, for the former, 'indeed' would provide a strong referential connection between S1 and S2 and for the latter a comment or characteristic or the truth of S1 will be enhanced on S2 (Seyed Ali et al, 2012).

#### 6.1.4 *Spanish translation correspondences of 'indeed'*

Our corpus analysis of the translation equivalents (Hummel, 1995) of 'indeed' in the Spanish translation corpora are graphically displayed in table 11. The figures represent the absolute occurrence of each translation spotting in each corpus and the proportion per million in parenthesis.

SPANISH Translation Indeed	Corpus frequency								
	Corpus total Number of Words	Europarl 7 English 53,837,625	Open subtitles 1,000,000	Emea 5,455,537	Multinot fiction 10,459	Multinot essay 18,538	Multinot Expe 40,643	Multinot Speech 36,675	Multinot Web 16,553
<b>Realmente</b>	399 (6.57)	182 (0.22)	0	0	0	3 (73.81)	1 (27.27)		<b>585 (107.87)</b>
En realidad	274 (4.51)	53 (0.06)	0	0	0	0	0		327 (4.57)
<b>De hecho</b>	269 (4.43)	222 (0.27)	<b>2 (0.13)</b>	0	<b>4 (215.7)</b>	<b>6 (147.62)</b>	<b>1 (27.27)</b>		<b>504 (395.42)</b>
Verdaderamente	137 (2.26)	32 (0.04)	0	0	0	0	0		169 (2.30)
De verdad	65 (1.07)	83 (0.1)	0	0	0	0	0		148 (1.17)
En verdad	28 (0.46)	107 (0.13)	0	0	0	0	0		135 (0.59)
La verdad	33 (0.54)	16 (0.02)	0	0	0	0	0		49 (0.56)
<b>Efectivamente</b>	231 (3.80)	146 (0.18)	0	0	0	0	<b>1 (27.27)</b>		<b>378 (31.25)</b>
<b>En efecto</b>	419 (6.90)	99 (0.12)	<b>2 (0.13)</b>	0	0	3 (73.81)	<b>1 (27.27)</b>		<b>524 (108.23)</b>
Por ejemplo	174 (2.86)	0	0	0	0	0	0		174 (2.86)
A saber	36 (0.59)	2 (<0.01)	0	0	0	0	0		38 (0.60)
<b>Ya (que)</b>	239 (3.93)	<b>223 (0.27)</b>	0	0	0	2 (49.2)	0		464 (53.40)
Puesto que	126 (2.07)	0	0	0	0	0	0		126 (2.07)
Pues	233 (3.84)	51 (0.06)	0	0	0	0	0		284 (3.90)
Justamente	28 (0.46)	1 (<0.01)	0	0	0	0	0		29 (0.47)
Indudablemente	25 (0.41)	3 (<0.01)	0	0	0	0	0		28 (0.42)
No cabe duda	21 (0.35)	0	0	0	0	0	0		21 (0.35)
Sin duda	342 (5.63)	87 (0.11)	0	0	0	0	0		429 (5.74)
Ciertamente	320 (5.27)	0	1 (0.06)	0	0	0	0		321 (5.33)
Lamentablemente	31 (0.51)	113 (0.14)	0	0	0	0	0		144 (0.65)
Sinceramente	30 (0.49)	0	0	0	0	0	0		30 (0.49)
En el fondo	9 (0.15)	4 (<0.01)	0	0	0	0	0		13 (0.16)
En absoluto	45 (0.74)	6 (0.01)	0	0	0	0	0		51 (0.75)
Naturalmente	86 (1.42)	6 (0.01)	0	0	0	0	0		92 (1.43)
Fundamentalmente	15 (0.25)	6 (0.01)	0	0	0	0	0		21 (0.26)
Evidentemente	38 (0.63)	5 (0.01)	0	0	0	0	0		43 (0.64)
Definitivamente	11 (0.18)	1 (<0.01)	0	0	0	0	0		12 (0.19)
Obviamente	24 (0.40)	7 (0.01)	0	0	0	0	0		31 (0.41)
Por supuesto	201 (3.31)	3 (<0.01)	0	0	0	0	0		204 (3.32)
Habida cuenta	21 (0.35)	0	0	0	0	0	0		21 (0.35)
Tal vez	48 (0.79)	0	0	0	0	0	0		48 (0.79)
Precisamente	175 (2.88)	13 (0.02)	0	0	0	1 (24.6)	0		189 (26.50)
Por cierto	47 (0.77)	72 (0.09)	0	0	0	0	0		119 (0.86)
Finalmente	66 (1.09)	1 (<0.01)	0	0	0	0	0		67 (1.10)
Al final	40 (0.66)	0	0	0	0	0	0		40 (0.66)
Así	327 (5.38)	203 (0.25)	0	0	0	0	0		330 (5.63)
Ni siquiera	63 (1.04)	4 (<0.01)	0	0	0	0	0	No entries	67 (1.05)
Exactamente	48 (0.79)	11 (0.01)	0	0	0	0	0		59 (0.80)
Prácticamente	28 (0.46)	0	0	0	0	0	0		28 (0.46)
Es cierto que	119 (1.96)	33 (0.04)	0	0	0	0	0		152 (2.00)
Por cierto	47 (0.77)	72 (0.09)	0	0	0	0	0		119 (0.86)
Concretamente	53 (0.87)	0	1 (0.06)	0	0	0	0		54 (0.93)
En concreto	50 (0.82)	0	0	0	0	0	0		50 (0.82)
Especialmente	179 (2.95)	2 (<0.01)1 (<0.01)	0	0	0	0	0		181 (2.96)
En especial	44 (0.72)	0	0	0	0	0	0		45 (0.73)
En definitiva	23 (0.38)	0	0	0	0	0	0		23 (0.38)
Al fin y al cabo	15 (0.25)	0	0	0	0	0	0		15 (0.25)
<b>Es decir</b>	113 (1.86)	0	<b>2 (0.13)</b>	0	0	0	0		<b>115 (1.99)</b>
O sea	15 (0.25)	2 (<0.01)	0	0	0	0	0		15 (0.25)
Mejor dicho	22 (0.36)	25 (0.03)	0	0	0	0	0		24 (0.37)
Bueno/a/en	731 (12.03)	162 (0.2)	0	0	0	0	0		756 (12.06)
(más) bien	377 (4.89)	21 (0.03)	0	0	0	0	0		539 (4.91)
Aún	201 (3.31)	0	0	0	0	1 (24.6)	0		223 (27.94)
Tal como	89 (1.47)	0	0	0	0	0	0		89 (1.47)
Actualmente	105 (1.73)	7 (0.01)	0	0	0	0	0		105 (1.73)
Es más	66 (1.09)	3 (<0.01)	<b>2 (0.13)</b>	0	0	0	0		73 (1.10)
<b>Sobre todo</b>	230 (3.79)	0	0	0	0	0	0		<b>235 (3.93)</b>
En la práctica	40 (0.66)	1 (<0.01)	0	0	0	0	0		40 (0.66)
Lo cierto es que	17 (0.28)	12 (0.01)	0	0	0	0	0		18 (0.29)
Más que	112 (1.84)	0	0	0	0	0	0		134 (1.85)
Por lo demás	54 (0.89)	0	0	0	0	0	0		54 (0.84)
Dado que	68 (1.12)	0	0	0	0	0	0		68 (1.12)
En cualquier caso	26 (0.43)	5 (0.01)	0	0	0	0	0		191 (3.06)
Todavía	185 (3.05)	0	0	0	0	0	0		62 (1.02)
En ese/este sentido	62 (1.02)	0	0	0	0	0	0		13 (0.21)
A mi juicio	13 (0.21)	200 (0.25)	0	0	0	0	0		228 (0.71)
Suficientemente	28 (0.46)	6 (0.01)	<b>2 (0.13)</b>	0	0	0	0		<b>313 (5.16)</b>
<b>Pero</b>	305 (5.02)	0	0	0	0	0	0		27 (0.44)
Por el contrario	27 (0.44)	0	0	0	0	0	0		26 (0.43)
Al contrario	26 (0.43)	0	0	0	0	0	0		85 (1.40)
Sin embargo	85 (1.40)	0	0	0	0	0	0		11 (0.18)
En cambio	11 (0.18)	9 (0.01)	0	0	0	0	0		32 (0.39)
No obstante	23 (0.38)	0	0	0	0	0	0		21 (0.35)
Apenas	21 (0.35)	0	0	0	0	0	0		54 (0.89)

Por lo demás	54 (0.89)	0							88 (1.43)
Solamente	86 (1.42)	2 (<0.01)	0	0					406 (6.20)
Aunque	374 (6.16)	32 (0.04)	0	0	0	0	0		124 (2.04)
A pesar de	124 (2.04)	0	0	0	0	0	0		30 (0.49)
Por consiguiente	30 (0.49)	28 (0.03)	0	0	0	0	0		72 (7.34)
También	44 (7.31)	11 (0.01)	0	0	0	0	0		305 (3.47)
Tal	293 (3.40)	0	0	0	0	0	0		90 (1.48)
Por tanto	90 (1.48)	0	0	0	0	0	0		56 (0.92)
Por ello	56 (0.92)	9 (0.01)	1 (0.06)	0	0	0	0	No entries	220 (3.48)
Además	211 (3.47)	1 (<0.01)	0	0	0	0	0		16 (0.26)
A su vez	15 (0.25)	14 (0.01)	0	0	0	0	0		<b>852 (13.81)</b>
<b>Incluso</b>	<b>838 (13.80)</b>	14 (0.01)	0	0	0	0	0		133 (56.15)
Claramente	128 (2.11)	4 (<0.01)	0	0	0	0	0		168 (2.77)
Desde luego	168 (2.77)	0	0	1 (53.94)	0	0	0		1 (0.02)
Pues bien	1 (0.02)	0	0	0					401 (4.24)
Claro	246 (4.05)	155 (0.19)	0	0					0
Bien mirado	0	0							
	1,092 (186.75)	2,574 (3.69)	13 (0.83)	0	5 (269.64)	<b>16</b> <b>(393.62)</b>	4 (108.08)	0	<b>3,704</b> <b>(962.61)</b>

**Table 11: Translation correspondences of ‘indeed’ in Spanish translation corpora.**

As shown in table 11 the translation equivalent (Hummel, 1995) for ‘indeed’ with the highest frequency in all parallel corpora is ‘**de hecho**’ (504 (395.6)) followed by ‘en efecto’ (379 (31.25)), ‘en realidad’ and ‘efectivamente’ (108.23), ‘incluso’, ‘ya’; es decir, sobre todo and pero (53.4).

Some examples of each of these translation equivalents (Hummel, 1995) are presented below:

‘Realmente’, and ‘de hecho’ are especially frequent as translation equivalents (Hummel, 1995) for ‘indeed’ Multinot speech (Lavid et al., 2015); ‘es decir’, ‘pero’, ‘en efecto’, ‘de hecho’ and ‘sobre todo’ are especially frequent as translation equivalents (Hummel, 1995) for ‘indeed’ in EMEA (Tiedemann, 2007).

‘De hecho’ is the most frequent translation equivalent for ‘indeed’ in the corpora Multinot essay and expe (Lavid et al., 2015); ‘incluso’ is the most frequent translation equivalent (Hummel, 1995) for ‘indeed’ in the corpus Europarl7 (Koehn, 2001) and ‘ya (que)’ is the most frequent translation equivalent for ‘indeed’ in the corpus Opensubtitles (Tiedemann, 2007).

Finally, the preferred translation equivalent (Hummel, 1995) for ‘indeed’ is ‘de hecho’ which appears with high frequency in the corpora Multinot essay, speech and expe (Lavid et al., 2015) as well as in EMEA (Tiedemann, 2007).

The equivalences and additional functions found in parallel corpora translation with ‘de hecho’ and ‘realmente’; have already been commented for the case of ‘in fact’, given the synonymic relationship between ‘indeed’ and ‘in fact’, it is assumed that these additional functions and relations are applicable to ‘actually’.

It is thus interesting to see what happens in the cases of the translation equivalents (Hummel, 1995) of ‘indeed’; ‘efectivamente’, ‘en efecto’, ‘ya (que)’, ‘es decir’, ‘sobre todo’, ‘pero’ and ‘incluso’.

*What is the preferred translation equivalent for ‘indeed’ in Spanish after the analysis of parallel corpora?* ‘**de hecho**’ is the preferred translation equivalent (Hummel, 1995) for ‘indeed’ as found in the analysis of its equivalents in parallel corpora, coinciding with ‘in fact’ which suggests a strong equivalence between the two verificative markers, being ‘actually’ more peripheral.

In order to check the features pointed out by Halliday (2014) and Aijmer (2004) for what in this thesis are referred to as ‘verificative markers’ (see figure ‘Functions and forms found by Halliday (2014) and Aijmer (2004) in monolingual analysis’), I will provide an analysis of the translation spottings found for ‘indeed’ in several tables, for the cases of ‘efectivamente’, ‘en efecto’, ‘ya (que)’, ‘es decir’, ‘sobre todo’, ‘pero’ and ‘incluso’.

With translation spottings I refer to the Spanish marker found as equivalent of ‘indeed’ while observing its translation from English into Spanish in parallel corpora (see table 11).

#### 6.1.4.1 Examples of translation occurrences for ‘indeed’

The following is an example of ‘indeed’ translated as ‘efectivamente’ in Multinot speech (Lavid et al., 2015):

<s> We can thereby guarantee that the queso manchego comes **indeed** from La Mancha and not from any other part of the world. </s>

<s> Con ello se garantiza que el queso manchego proceda, **efectivamente**, de La Mancha y no de cualquier otra parte del mundo. </s>

- a) According to Halliday (2014) this is an adverbial of circumstance or an adverbial of mood and confirmation. Aijmer (2004) would consider this as an epistemic adverbial where the information is enhanced.
- b) As the information is enhanced Pons Bordería (2005) would see here a function of co-oriented argumentation.
- c) According to the RAE (2019), ‘efectivamente’ is an adverbial which expresses confirmation and truth statement, thus ‘indeed’ would have a mixture of confirmation and emphasis function.

‘En efecto’ Multinot speech (Lavid et al., 2015) as translation correspondence for ‘indeed’:

<s> **Indeed**, some have doubts about the European project.</s> <s> <s> **En efecto**, hay quien duda de nuestro proyecto europeo.

- a) Aijmer (2004) would see in this example a discourse marker, Halliday (2014). would see an adverbial of comment and a function of reinstatement, while Aijmer (2004) would remark enhancement function.
- b) Pons Bordería (2005) would remark enhancement as linked to a co-oriented argumentative function.

- c) The Rae (2019) considers ‘en efecto’ as an adverbial of purpose, which can be extraposed to ‘indeed’ as a function.

‘En efecto’ EMEA (Tiedemann, 2007) as translation correspondence for ‘indeed’:

<s> b) The Applicant **indeed** considered that it is the amount of respirable particles <4.7 µm that will ultimately determine bronchodilator efficacy, particularly for stages 3/ 4/ 5 (i.e. particle size 1.1-4.7 µm) which correspond to bronchial deposition where the airway smooth muscle beta-2 adrenoceptors are located. </s>

<s> b) El solicitante consideró, **en efecto**, que es la cantidad de partículas respirables «Fz 1 4,7 µm la que en última instancia determina la eficacia broncodilatadora, sobre todo en las fases 3/ 4/ 5 (es decir, tamaño </s>

- a) According to Halliday (2014) this is an adverbial of circumstance or an adverbial of mood and confirmation. Aijmer (2004) would consider this as an epistemic adverbial where the information is enhanced.
- b) As the information is enhanced Pons Bordería (2005) would see here a function of co-oriented argumentation.
- c) According to the RAE (2019), ‘en efecto’ is an adverbial which expresses confirmation and truth statement, thus ‘indeed’ would have a mixture of confirmation and emphasis function.

‘Ya (que)’ Open Subtitles (Tiedemann, 2007) as translation correspondence of ‘indeed’:

<s> Yes, **indeed**. </s> <s> **-Indeed**. </s> <s> Stepmother, **indeed!** </s>

<s> - Sí, **ya** lo creo. </s> <s> **-Ya** veo. </s> <s> ¡**Ya**, mi madrastra! </s>

- 1) Example one illustrates a locution, a speech marker, which Aijmer (2004) would see as a pragmatic marker of evidentiality. This example shows a common Spanish idiom ‘ya lo creo’. The Rae (2019) sees this expression as an illocution of confirmation.

- 2) This example two illustrates another Spanish idiomatic expression ‘ya veo’, which entails a pragmatic marker of evidentiality (Aijmer, 2004) and, apart from confirmation (RAE, 2019), expresses the speaker’s attitude or mood.
- 3) Portolés and Zorraquino (1999) see ‘ya’ as a meta-discursive marker, which is, then, the function ‘indeed’ is conveying here.

‘Es decir’ EMEA (Tiedemann, 2007) as translation correspondence of ‘indeed’:

<s>**Indeed**, we should end up drafting guidelines. </s>

<s> Es decir, también deberíamos poder llegar a adoptar directrices. </s>

- a) Aijmer (2004) would point out these as discourse markers with the function of contrast, consequently, Pons Bordería (2005) would emphasize the anti-oriented argumentative function.
- b) Portolés and Zorraquino (1999) consider ‘es decir’ as a reformulation marker with the function of explanation as well as Garcés Gómez (2005) does, which is what ‘indeed’ is doing in the source.
- c) Del Saz (2016) considers the following functions for ‘es decir’ inside reformulation: expansion, presupposition, entailment and felicity condition. What means that ‘indeed’ in S2 is expanding S1, presupposing the information expressed in S1 and assuming the truth of S1 in S2.
- d) Núñez et al. (2006) not only see reformulation as explanation but also as exemplification, rectification and resumptiveness. Which means that ‘indeed’, apart from the functions addressed in a), b) and c) also reformulates S1 in S2 providing an example based on S1, rectifying some element of S1 and summarizing S1 into its core message.

'Sobre todo' EuroParl7 (Koehn, 2001) as translation correspondence for 'indeed':

<s> Mr President, at the outset I should like to congratulate Mr Garosci on the preparation and **indeed** on the presentation of his report. </s>

<s> Señor Presidente, desearía ya de entrada felicitar a el Sr. Garosci por la preparación y, **sobre todo**, por la presentación de este informe. </s>

- a) Halliday (2005) would point out the clarification degree of the marker here and how it provides information of an additional characteristic of S1 by means of an adverbial. Aijmer (2004) would emphasize the additional purpose and Pons Bordería, based on the additional function, would point out the co-oriented argumentative purpose.
- b) Quirk (1990) would consider 'sobre todo' as a connector of reinforcement and confirmation, while Seyed Ali et al (2012) would focus on the emphasis function. Thus, in this example, 'indeed' reinforces S1 putting emphasis on one of its elements and adding information about S1 in S2 in order to emphasize a consequence or feature of S1.
- c) It is important to note the collocation 'and indeed' which is very frequent (y sobre todo) and but indeed (pero sobre todo).

'Incluso' Open Subtitles (Tiedemann, 2007) as translation correspondence for 'indeed':

<s> Throwing away a document that might be of significance to Tom and within future generation of readers as the basis of a novel **or indeed** a trilogy, was not an act he was so stupid to commit, although he had to admit that in a moment of weakness he might have said he would. </s>

<s> Deshacerse de un documento que podría ser de importancia para Tom ... y para futuras generaciones de lectores como base para una novela ... **o incluso** una trilogía ... no era un acto que él fuese tan estúpido de cometer ... aunque tuvo que admitir que un momento de debilidad... tal vez hubiera dicho que podría. </s>

- a) The example above shows one of the most frequent collocations of ‘incluso’ which is ‘o incluso’, which in the source text is reflected as ‘or indeed’.
- b) The function is both adversative, due to the collocation with ‘o’/ ‘or’, and emphatic (Aijmer, 2004) with a clarification purpose towards the addition of some features of S1 in S2 (not a novel but a trilogy) (Halliday, 2014).
- c) Due to the adversative nature of the markers Pons Bordería (2005) would point out the anti-oriented argumentative function.
- d) Due to the emphatic nature of the markers Pons Bordería (2005) would point out the co-oriented argumentative function.
- e) Portolés and Zorraquino (1999) as well as Fuentes (1998) refer to the marker ‘incluso’ as an additive connector, so ‘indeed’ would also have an additive nature apart from the functions addressed in a) to d).

The comparative study of ‘indeed’ and its translation equivalents (Hummel, 1995) has revealed additional functions to the ones considered for the ones found in the translation equivalents of ‘in fact’ and ‘actually’:

- a) Purpose and confirmation (Quirk, 1990).
- b) Meta-discursive marker (Portolés and Zorraquino, 1999).
- c) Reformulation, explanation, presupposition (del Saz, 2016).

In this section, through the comparative analysis of verificative markers in parallel corpora and their translation equivalents (Hummel, 1995) in Spanish, new functions have been found as to be added to the ones regarded in some monolingual studies such as Halliday (2014) and Aijmer (2004).

These results reveal that translation can be used as a tool to understand and spot the functions of a specific marker through the features and functions conveyed by its translation equivalent (Hummel, 1995) in another language.

### 6.1.4.2 Translation lexical Domains ‘indeed’

Translation then goes beyond finding the translation spotting in the target (translation occurrence in context), it is also a tool which provides information about function and about preference towards one translation equivalent (Hummel, 1995) among other within its context-function, helping to ensure the accuracy of the target output, as it is shown in figure 13 illustrating the Translation lexical domains (TLDs).

In the following figure the ‘translation domains’ of ‘indeed’ are shown, where the translation correspondences are grouped according to the most suitable function displayed within the context where the source and the target are found:

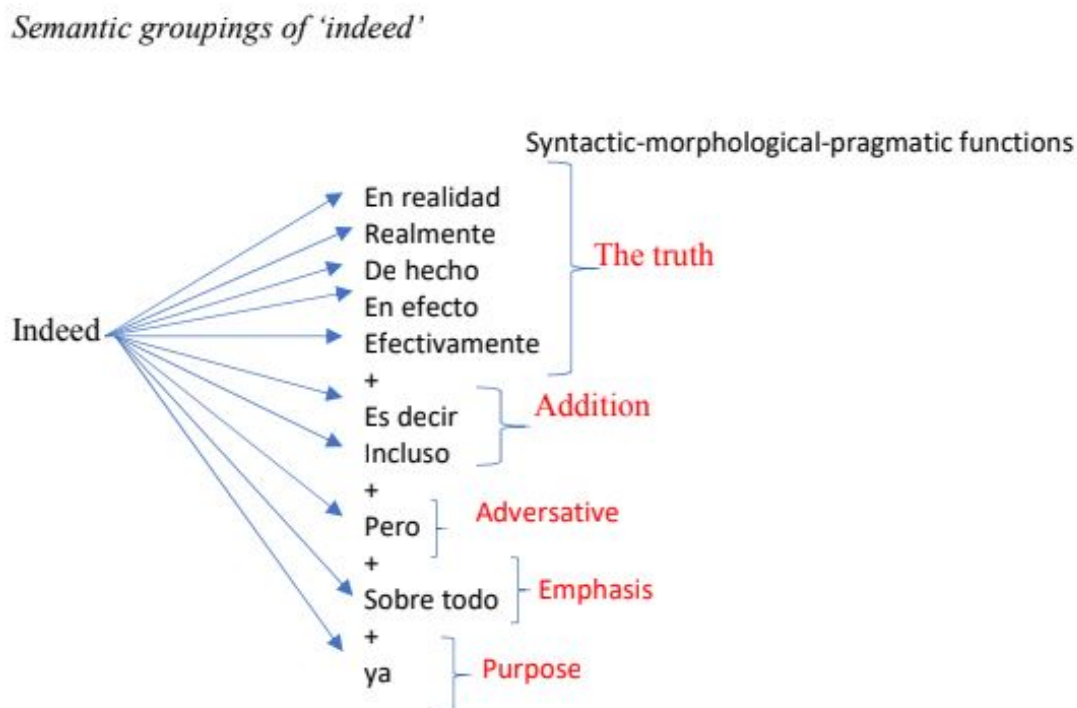


Figure 13: TLDs ‘indeed’.

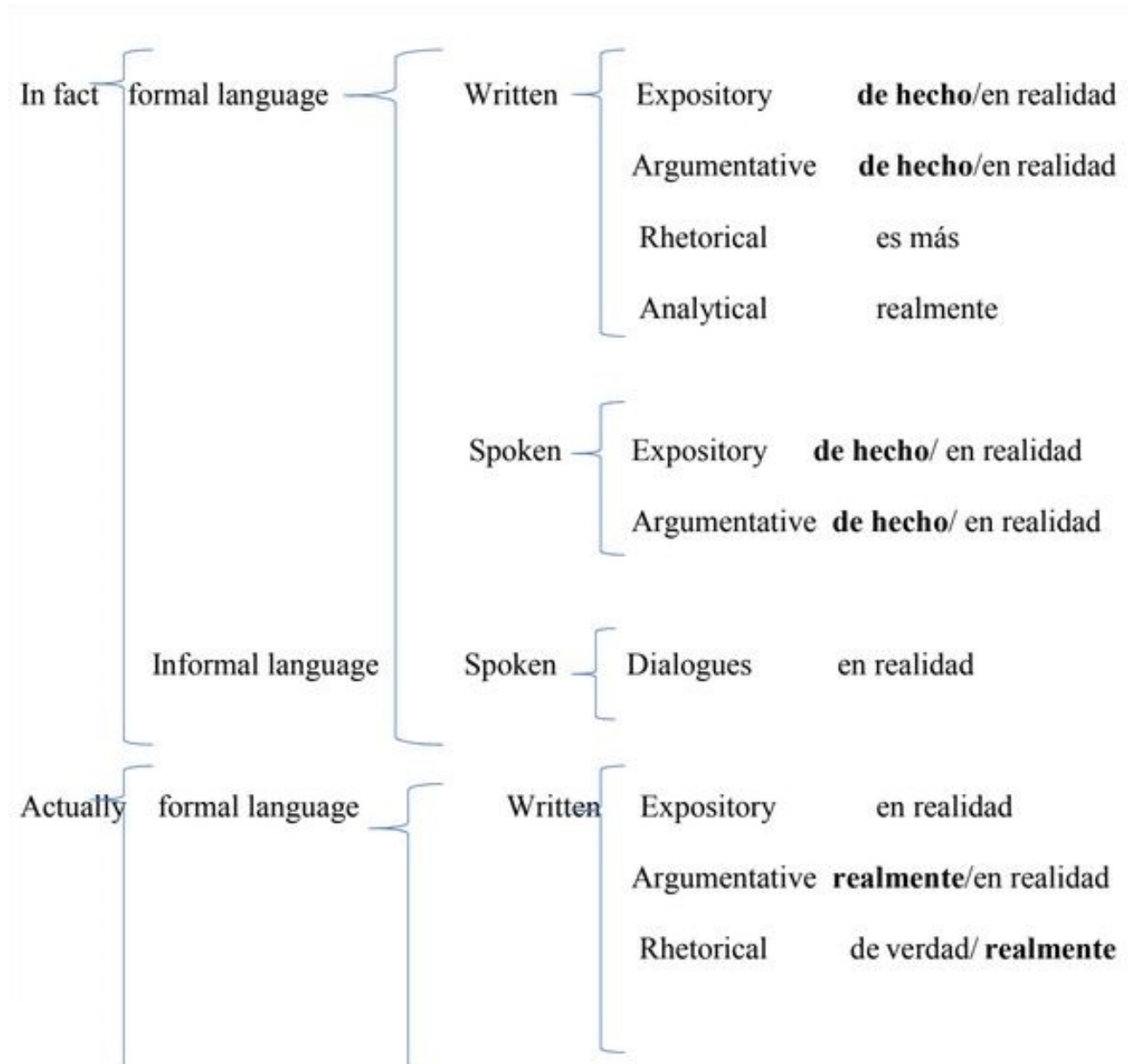
### 6.1.4.3 Genre-specific preferences ‘indeed’

With respect to the genre-specific preferences, the corpus analysis yields the following results:

- a) ‘De hecho’ is the most frequent translation equivalent (Hummel, 1995) for ‘indeed’ in the following corpora: Emea (Tiedemann, 2007); Multinot essay, expe and speech (Lavid et al., 2015): 0.13, 215.7, 147.62 and 27.27 per million, respectively.
- b) ‘Realmente’: most frequent translation equivalent (Hummel, 1995) for ‘indeed’ in Multinot speech (Lavid et al., 2015): 27.27 per million.
- c) ‘Efectivamente’ is the most frequent translation equivalent (Hummel, 1995) for ‘indeed’ in Multinot speech (Lavid et al., 2015): 27.27 per million.
- d) ‘En efecto’ is the most frequent translation equivalent (Hummel, 1995) for ‘indeed’ in EMEA (Tiedemann, 2007) and Multinot speech (Lavid et al., 2015): 0.13 per million and 27.27 per million respectively.
- e) ‘Ya (que)’ is the most frequent translation equivalent (Hummel, 1995) for ‘indeed’ in Open Subtitles (Tiedemann, 2007): 0.27 per million.
- f) ‘Es decir’, ‘sobre todo’ and ‘pero’ some of the most frequent translation equivalents (Hummel, 1995) for ‘indeed’ in EMEA (Tiedemann, 2007): 0.13 per million in the three cases.
- g) ‘Incluso’ is the most frequent translation (Hummel, 1995) counterpart for ‘indeed’ in the corpus Europarl7 (Koehn, 2001): 13.8 per million.

### 6.1.5 Summary of genre-specific preferences of English VMs

Based on the characteristics which define each of the corpora presented the genre-specific translation preferences of ‘in fact’, ‘actually’ and ‘indeed’ are displayed in figure 14 below.



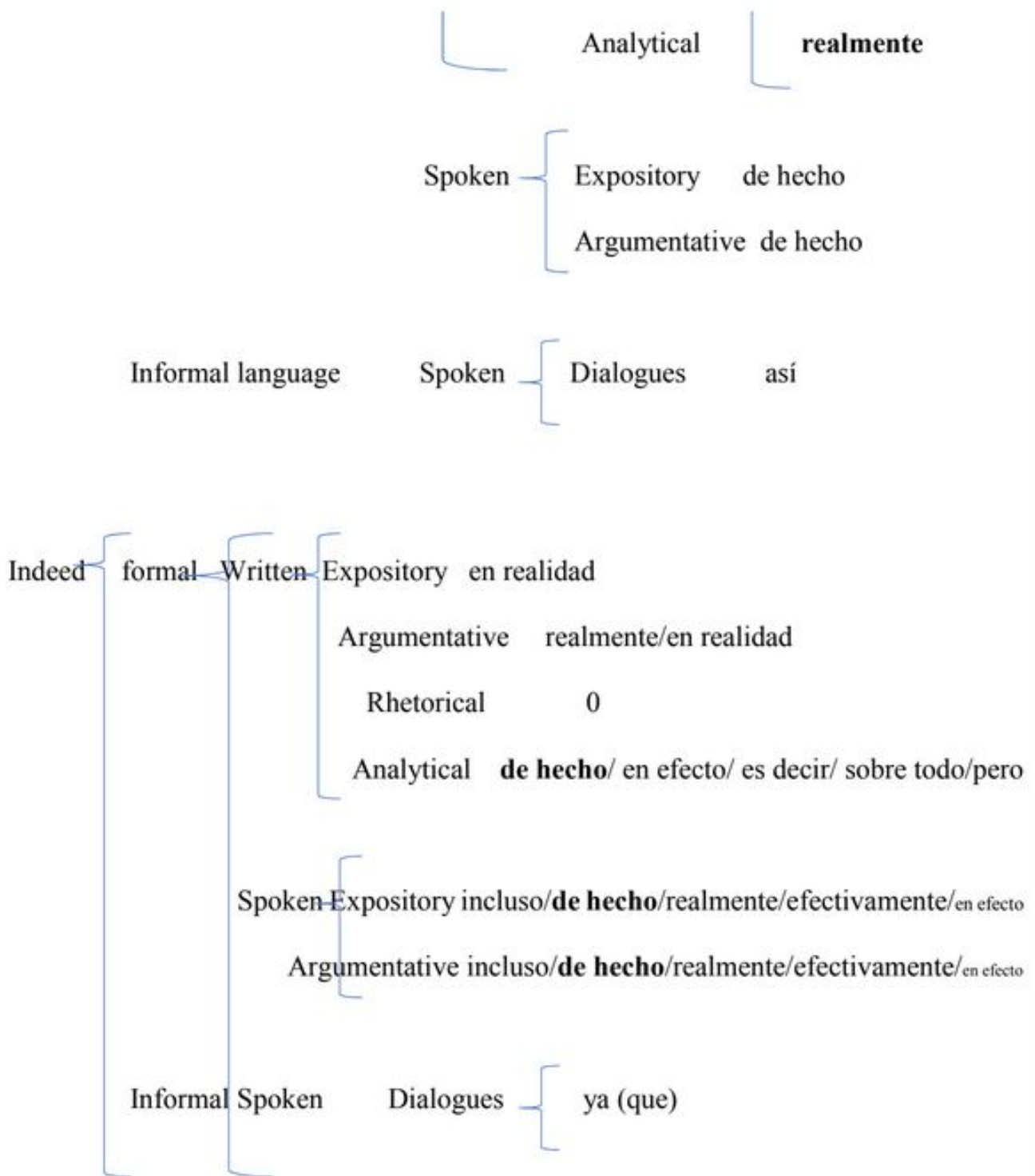


Figure 14: Genre-specific translation preferences of ‘in fact’, ‘actually’ and ‘indeed’.

The classification of each genre has been based on the characteristics of the corpora used in the analysis of translation equivalence as follows:

- a) Europarl7 (Koehn, 2001): it represents a formal spoken argumentative and expository genre as it offers transcriptions of legal and parliamentary discourse.
- b) EMEA (Tiedemann, 2007): it represents analytical genre as it reproduces medical texts.
- c) Open Subtitles (Tiedemann, 2007): it represents informal spoken language as it corresponds to the transcriptions of the subtitles based on film dialogues.
- d) Multinot Expe and Essay (Lavid et al., 2015): it represents formal written argumentative and expository language, the former as expository texts and the latter as scientific and educational essays.
- e) Multinot Speech and Web (Lavid et al., 2015): it represents a formal spoken argumentative and expository genre as it offers transcriptions of legal and parliamentary discourse.
- f) Multinot Fiction (Lavid et al., 2015): it represents formal written rhetorical texts as it is based on extracts from English language literature.

The analysis reveals that:

- 1) Written formal expository argumentative texts prefers ‘de hecho’ as translation equivalent (Hummel, 1995) for ‘in fact’, and ‘realmente’ and ‘en realidad’ for both ‘actually’ and ‘indeed’.
- 2) Spoken formal expository argumentative texts prefer ‘de hecho’ as translation equivalent (Hummel, 1995) for ‘in fact’, ‘actually’ and ‘indeed’. But in the case of ‘indeed’ additional equivalents will be valid: ‘incluso’, ‘efectivamente’, ‘en efecto’ and ‘realmente’.
- 3) Written formal analytical texts prefer ‘realmente’ as translation equivalent (Hummel, 1995) for ‘in fact’, ‘de hecho’ as translation equivalent of ‘actually’ and a significant number of translation equivalents for ‘indeed’: ‘en efecto’, ‘es decir’, ‘sobre todo’, ‘pero’.

- 4) Written rhetorical texts prefer ‘es más’ as translation equivalent (Hummel, 1995) for ‘in fact’; ‘de verdad’ and ‘realmente’ as translation equivalents (Hummel, 1995) for ‘actually’; and does not offer any occurrence for ‘indeed’.
- 5) Dialogues prefer ‘en realidad’ as translation equivalent for ‘in fact’, ‘así’ as translation equivalent (Hummel, 1995) for ‘actually’ and ‘ya’ as translation equivalent for ‘indeed’.

As mentioned in the answers provided to the first research question ‘which are the translation preferences of the VM’, ‘in fact’ prefers ‘de hecho’, ‘actually’ prefers ‘realmente’ and ‘indeed’ prefers ‘de hecho’.

### 6.1.6 Summary of Genre-specific preferences ‘in fact’, ‘actually’ and ‘indeed’

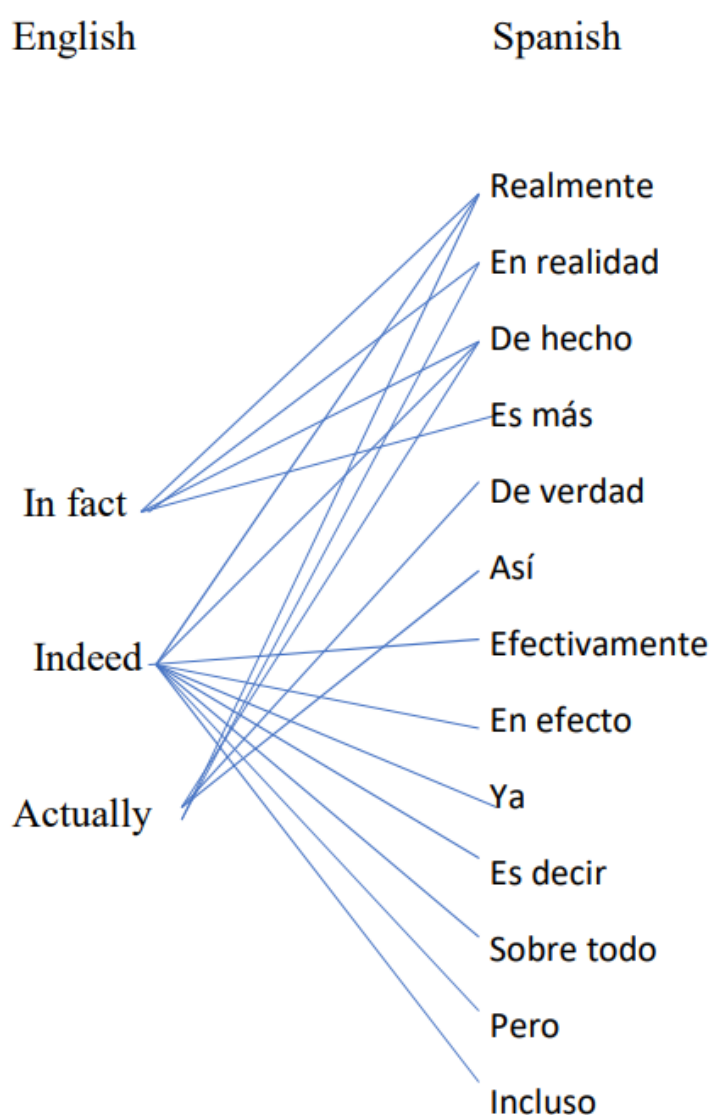
*What are the genre specific preferences of each VM?*

- 1) ‘In fact’ and ‘indeed’ coincide in translation preference so they seem to be more central than ‘actually’, which would be more peripheral.
- 2) ‘De hecho’ is the translation equivalent mostly preferred in each genre.
- 3) ‘In fact’ is more frequent in both formal written and spoken expository and argumentative texts, whereas ‘actually’ would prefer formal spoken expository argumentative texts only; and ‘indeed’ would be more likely in formal written analytical texts and in formal spoken expository argumentative texts.
- 4) The three verificative markers coincide in their preference towards formal spoken expository and argumentative texts.
- 5) ‘Indeed’ is not frequent in formal written rhetorical texts.
- 6) ‘Indeed’ presents the highest rate of translation equivalents (Hummel, 1995).
- 7) The three VM are preferred in formal contexts.

**6.1.7 Lexical domains for VM’s ‘in fact’, ‘indeed’, ‘actually’ .**

In the following figure, the lexical domains of ‘in fact’, ‘indeed’ and ‘actually’ will be presented as follows: the English VMs (Halliday, 2014) are the point of departure and their corresponding Spanish equivalents found in parallel corpus translation, as seen previously in this section, are the target.

With this figure, it is possible to see where the three English VMs (Halliday, 2014) coincide being those coincidences the lexical fields for these VMs. The following figure 15 illustrates the lexical fields found in Spanish for ‘in fact’, ‘indeed’ and ‘actually’:



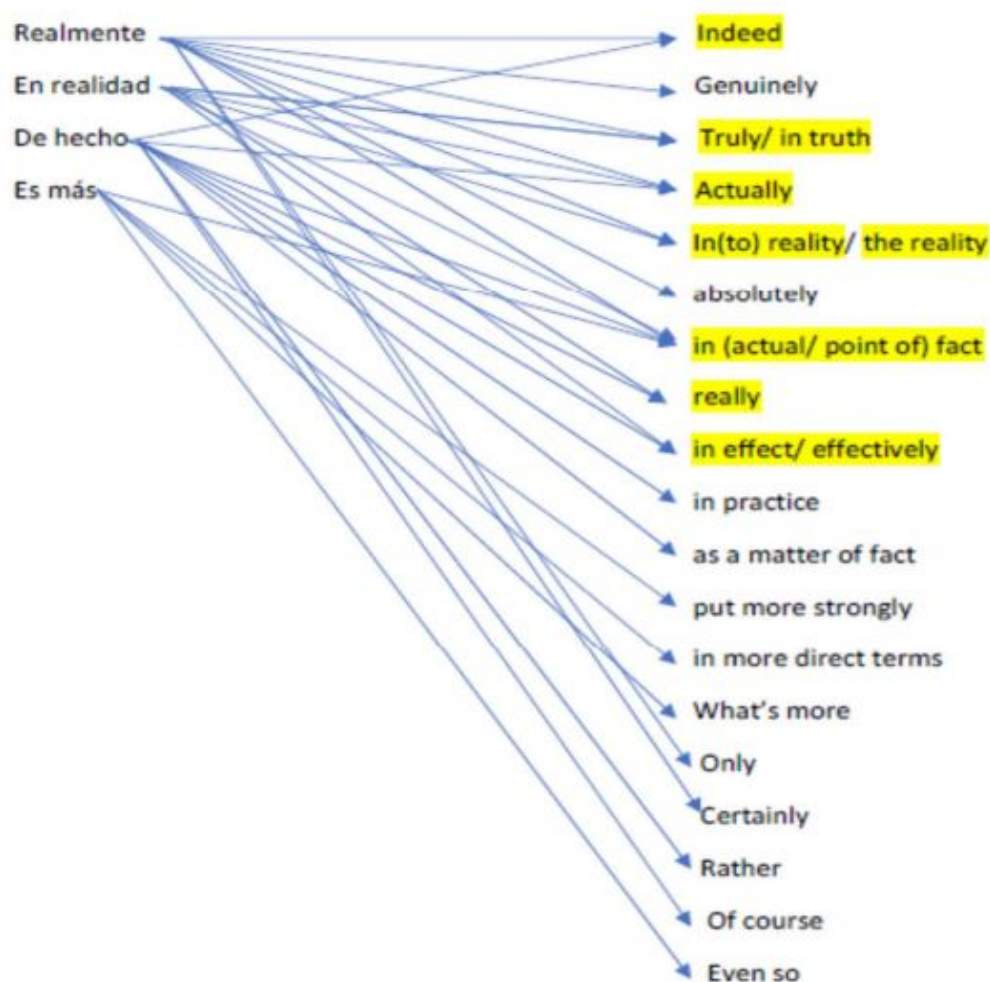
**Figure 15: the lexical fields in Spanish resulting from translations of ‘infact’, ‘actually’ and ‘indeed’.**

### 6.1.8 Second-mirroring

In this section, the most frequent translation equivalents for each VMs in English will be ‘back-translated’ in order to check whether the equivalence is bilateral and the correctness of that equivalence.

### 6.1.9 Second-mirroring ‘in fact’

The following figure 16 represents the English back translations of the most frequent Spanish equivalents of ‘in fact’:



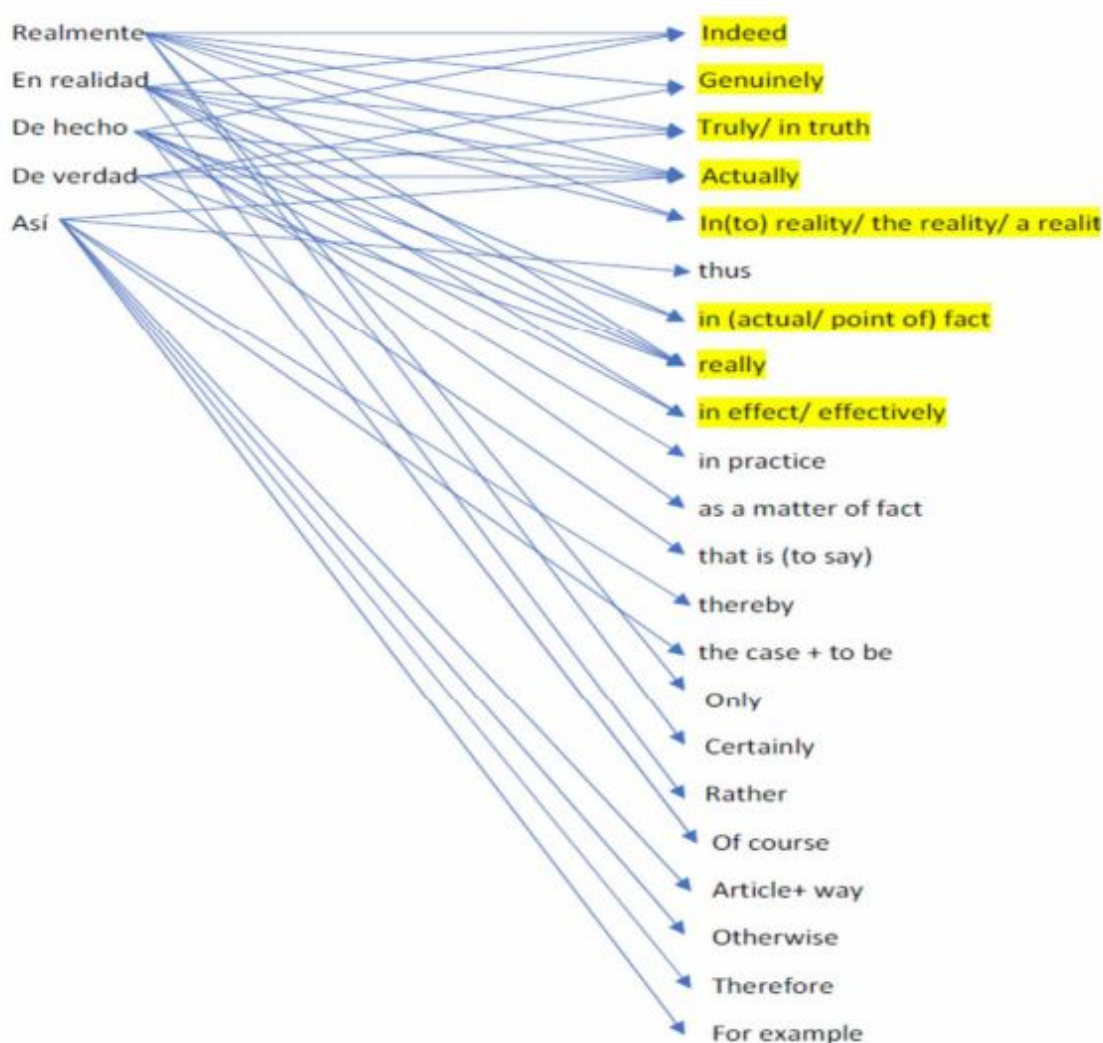
**Figure 16: The second translation image: English translations of the most frequent Spanish translations of ‘in fact’.**

The information figure 16 provides is the following:

- a) There is bilaterality between VMs (Halliday, 2014), as ‘actually’ and ‘indeed’ are strongly chosen.
- b) ‘In fact’ appears to be central due to its strong relationship with ‘indeed’ and ‘actually’.
- c) ‘In effect’, ‘effectively’ and ‘rely’ are confirmed as having a strong link with the function of verification (Halliday, 2014).
- d) There is strong relationship between English VMs (Halliday, 2014) and ‘Marcadores de refuerzo argumentativo’ (Portolés and Zorraquino, 1999): ‘de hecho’ and ‘en realidad’.

### 6.1.9.1 Second-mirroring ‘actually’

The following figure 17 represents the second translation image: English back-translations of the most frequent Spanish equivalents of ‘actually’:



**Figure 17: The second translation image: English back-translations of the most frequent Spanish equivalents of ‘actually’.**

Information obtained in figure 17:

a) ‘Actually’ is strongly connected with the truth verification, which is seen in: ‘truly’ or ‘genuinely’.

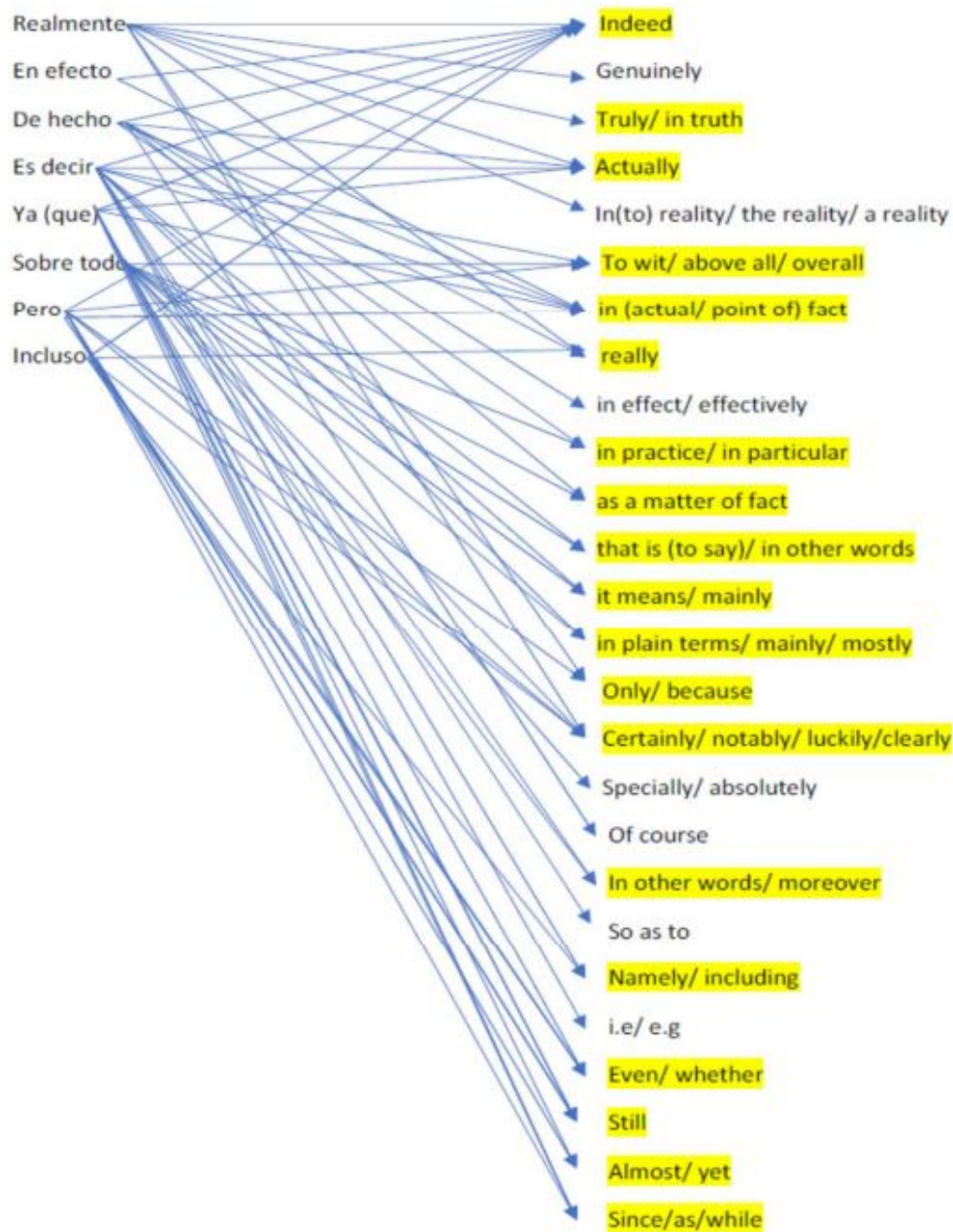
b) As in the case of ‘in fact’, ‘really’, ‘effectively’ and ‘in effect’ are confirmed as possible VMs.

c) The four ‘Marcadores de refuerzo argumentativo’ (Portolés and Zorraquino, 1999) appear as

having strong relationship with ‘actually’: ‘realmente’, ‘de hecho’ and ‘de verdad’.

### 6.1.9.2 Second-mirroring ‘indeed’

The following figure 18 represents the second translation image: English back-translations of the most frequent Spanish equivalents of ‘indeed’:



**Figure 18: The second translation image: English translations of the most frequent Spanish translations of ‘indeed’.**

Information obtained in figure 18:

- a) 'Indeed' offers the broadest example of translation mirroring.
- b) There is confirmation about the relationship between 'verification' Halliday (2014) and 'reformulation' (de Saz, 2003): 'namely' explanation-identification, 'to wit' explanation-identification.
- c) 'Indeed' has a strong adverbial tendency: 'mainly', 'certainly', 'only'.
- d) 'Indeed' relates to translation occurrences of 'actually' and 'in fact': 'truly'-'in truth', for the former, 'really' for the latter.
- e) The Spanish connectives as most frequent correspondences of 'in fact' not only include 'marcadores de refuerzo argumentativo' (Portolés y Zorraquino, 1999), but also adversatives (pero) or emphasizees (incluso).

## 6.2 Spanish connectives: ‘la verdad’, ‘de hecho’, ‘en realidad’, ‘realmente’

After presenting the results of the English VMs, their translation correspondences in Spanish, their semantic groupings or TLDs and their genre-specific preferences in the parallel corpora used in this study, this second part of Chapter 6 presents the analysis results of the Spanish VMs which are the most frequent back-translations of these English connectives. This includes the following markers: ‘de hecho’, ‘en realidad’, ‘realmente’, and ‘la verdad’, among the most frequent.

The structure of each section is the same as for English VMs:

First, a revision of previous studies regarding these Spanish connectives is outlined with monolingual and bilingual analysis. This is followed by the presentation of the translation correspondences of each Spanish connective in frequency table as has been shown for English VMs. The the translation preferences and genre preference are explained. Finally, the creation of the translation lexical domains (TLDs) on the basis of these correspondences is illustrated on several figures which are analysed.

### 6.2.1 Monolingual analysis of Spanish markers

It seems that the categorization of these Spanish connectives, as it happens with the English ones, does not appear to be homogeneous, according to Garcés Gómez (2005:70):

Según Portolés y Zorraquino (1999) los **marcadores de reformulación** forman parte de lo que en esta tarea denominamos los **marcadores discursivos**.

Cabe decir que ya sobre el nombre que se atribuye a estas unidades no existe conformidad. Otros lingüistas suelen utilizar términos como **conectores** (Pons, 1998), **operadores discursivos** (Casado, 1993), **muletillas** (Cortés Rodríguez, 1998) o enlaces **extraoracionales** (Gili Gaya, 1961).

(...) No todos los lingüistas utilizan el término “**reformulación**” para referir a este fenómeno, Schegloff, Jefferson y Sacks (1977) por ejemplo siempre hablan de “**reparación**” pero se trata del mismo fenómeno.

The main difference between a ‘marcador’ and an ‘operador’ is whether they work as discourse connectors or sentence connectors, but for the essence of the present project both will be englobed into the term connectives. With respect to the various classifications addressed to these Spanish connectives I will illustrate some of the most frequent:

According to some authors such as Portolés and Zorraquino (1999), Catalina Fuentes (2010) and Mar Garacha Camarero the classification would be as follows:

- 1) ‘Operadores’: these are a type of discourse markers which do not connect directly two units but that determine the condition of the proposition that follows (Portolés and Zorraquino, 1999): *en realidad, de hecho, la verdad, verdaderamente* (among others).
- 2) ‘Marcadores’: DMs which directly connect two units of discourse.

Regarding Pons Bordería (2005) and de Briz, DM’s (1998) can also be classified into meta-discursive ones and argumentative ones. The Spanish connectives I am analyzing in this paper correspond to the second group and are defined as (El Messaoudi Rifi, 2018):

Those markers which stablish discursive relationships with a logic-argumentative function between two or more elements of the discourse as connectors or that express extra-discursive connections with the context as operators. Inside the argumentative markers, there are two subtypes, the co-oriented ones and the anti-oriented ones, the former lead to a similar discourse conclusion, whereas the latter suggests contraposition.

Catalina Fuentes (2016), for example, has other categorizations for these markers, namely, she considers ‘*en realidad*’ as a modal and informative operator, ‘*de hecho*’ is considered to be a connector and in some contexts a modal, ‘*la verdad*’ is considered as a modal, ‘*realmente*’ is considered as a modal too.

Finally, other authors such as Violeta de Monte (1999), add some additional comments with respect to the so called ‘operadores de refuerzo argumentativo’:

- a) En realidad: emphasizes the contrast between the truth of the statement where it appears and the hypothesis of the argument which precedes it.
- b) De hecho: it contrasts the certainty of a statement with the likeability or unlikeability of another argument.
- c) Realmente: emphasizes the factuality of the statement in which it appears.
- d) La verdad: it shows the evidentiality of the statement that follows with respect to the speaker’s/writer’s thoughts.

## 6.2.2 ‘De hecho’

In this section, as it has been done for English VMs (Halliday, 2014), the three research questions suggested in this thesis are answered. The study of ‘de hecho’ starts with a short analysis of monolingual categorizations of it by several researchers, then and based on a frequency table, translation preferences are shown with some examples, afterwards genre preference, once more based on the previous frequency table, is analysed. Finally, some additional comments regarding TLDs and correspondence potential are given.

### 6.2.2.1 Monolingual analysis on ‘de hecho’

These are the categorizations for ‘de hecho’:

- a) Statement of Truth (El Messaudi Rifi, 2018). This is, checking the truthfulness of the statement.
- b) Modal connector (Catalina Fuentes, 2010). As quality of circumstance.
- c) Probability and contrast (Violeta de Monte, 1999). This is, conditional and adversative.
- d) Reinforcement (Portolés and Zorraquino, 1999). This is, emphasis or enhancement.

‘Indeed’ and ‘in fact’ are considered as VMs by Halliday (2014), as seen in the previous sections, these two VMs have as correspondence ‘de hecho’. As Aijmer (2004) considers several pragmatic functions for these two VMs, addition, contrast, emphasis and purpose, as well as statement of truth; these functions must be adhered to ‘de hecho’ as well. The only function which is not reflected in the Spanish monolingual analysis outline is ‘addition’, which should be added to the functions of ‘de hecho’ as part of the parallel translation analysis carried out in this thesis.

6.2.2.2 Translation correspondences of ‘de hecho’

Answering the first research question, and based on the results shown in table 12 above, the most frequent translation correspondence in English for ‘de hecho’ is ‘indeed’ 263 (207.95), which appears with the strongest frequency in the corpora Multinot fiction and web (Lavid et al., 2015). Meanwhile, ‘in fact’ is the most frequent translation equivalent (Hummel, 1995) for ‘de hecho’ in the corpus Europarl7 (Tiedemann, 2007) 420 (6.9) and in Multinot essay (Lavid et al., 2015) 2 (99.82).

It is true, though, that, if looking at the absolute numbers the results seem to be the way around, but, as mentioned for the English VMs, as the absolute figures are based on the particular number of words of each corpus and the proportion per million is a common standard for all the corpora, the proportion per million is the one taken as reference. See the following table 12:

SPANISH ORIGINAL	ENGLISH TRANSLATION	Corpus frequency								TOTAL Per million
		Corpus total Number of Words	Europarl 7 English 53,837,625	Open subtitles 1,000,000	Emea 5,455,537	Multinot fiction 13,164	Multinot essay 20,036	Multinot Expe 40,643	Multinot Speech 307,625	
<b>De hecho</b>	<b>Indeed</b>	245 (4.03)			2 (151.92)	0			16 (52)	263 (203.92)
	Genuinely	14 (0.23)			0	0			0	14 (0.23)
	Truly	18 (0.3)			0	0			0	18 (0.30)
	Actually	161 (2.65)			0	0			0	161 (2.65)
	Really	30 (0.49)			0	0			0	30 (0.49)
	Absolutely	11 (0.18)			0	0			0	11 (0.18)
	<b>In fact</b>	<b>420 (6.9)</b>			0	<b>2 (99.82)</b>			15(48.75)	<b>437(155.47)</b>
	In reality	26 (0.43)			0	0			0	26 (0.43)
	In point of fact	6 (0.1)			0	0			0	6 (0.10)
	The truth	8 (0.13)			0	0			0	8 (0.13)
	It is true that	8 (0.13)			0	0			0	8 (0.13)
	Nevertheless	11 (0.18)	No entries	No entries	0	0	No entries	No entries	0	11 (0.18)
	On truth	0			0	0			0	0
	But	147 (2.42)			0	0			0	147 (2.42)
	In effect	61 (1.00)			0	0			1 (3.25)	62 (4.25)
	Into reality	0			0	0			0	0
	In practice	24 (0.39)			0	0			0	24 (0.39)
	To tell	1 (0.02)			0	0			0	1 (0.02)
	<b>De facto</b>	225 (3.7)			0	<b>2 (99.82)</b>			3 (9.75)	10 (109.67)
	In point of fact	6 (0.1)			0	0			0	6 (0.10)
	Rather	23 (0.38)			0	0			1 (3.25)	24 (3.63)
	Effectively	41 (0.67)			0	0			0	41 (0.67)
	Of course	27 (0.44)			0	0			1 (3.25)	28 (3.69)
<b>TOTAL Per million</b>		1542 (24.97)	2(151.92)	4 (199.64)					37 (120.25)	1585 (496.78)

Table 12: translation equivalents (Hummel, 1995) for ‘de hecho’.

Examples of ‘de hecho’ and its common translations:

a) Sentence connector, contrast. Fronted:

<s> **De hecho**, el primero de dichos Protocolos se re-fleja en la declaración inequívoca de mi Gobierno en el sentido de que la edad mínima para el reclutamiento en las fuerzas armadas es de 18 años. </s> MULTINOT SPEECH 4

<s> **Indeed**, the first Protocol is reflected in my Government unequivocal declaration that the minimum age for recruitment into the armed forces is. </s> MULTINOT SPEECH 4

b) Sentence connector, illustration and rectification. Fronted:

<s> **De hecho**, en 1996 se celebraron en Brunei Darussalam varios seminarios y talleres, incluida una reunión de consulta regional conjunta con el UNICEF. </s> MULTINOT SPEECH 4

<s> **In fact**, several seminars and workshops, including a regional consultative meeting, were held jointly with UNICEF in Brunei Darussalam in 1996. </s> MULTINOT SPEECH 4

c) Sentence connector, embedded, enhancement:

<s> En teoría, si esto puede responder a las exigencias de concentración y, por lo tanto, de eficacia de las intervenciones, surge espontánea la pregunta de si la situación actual de las regiones del objetivo 1 es tal que permite, **de hecho**, una política de intervenciones integradas, para cuya consecución es necesario un instrumento operativo de coordinación de las ayudas. </s> EUROPARL 7

<s> While in theory this may meet the need for concentration and therefore the efficiency of interventions, we cannot help wondering whether the current situation of the Objective 1 zones allows, **in practice**, a policy of integrated interventions, for which an operational instrument to coordinate the appropriations is necessary. </s> EUROPARL 7

d) Sentence connector, embedded, certainty and evidentiality:

<s> Señalo bien, “ sin razón ” ya que, **de hecho**, va a reducir esos derechos. </s> EUROPARL 7

<s> This Charter is really misnamed because it will **actually** reduce these rights. </s> EUROPARL 7

6.2.2.3 Lexical domains for ‘de hecho’

Figure 19 presents the lexical domains found for ‘de hecho’, it shows the most frequent translation occurrences for ‘de hecho’ and the function they perform in discourse when this VM is translated into each of these correspondences.

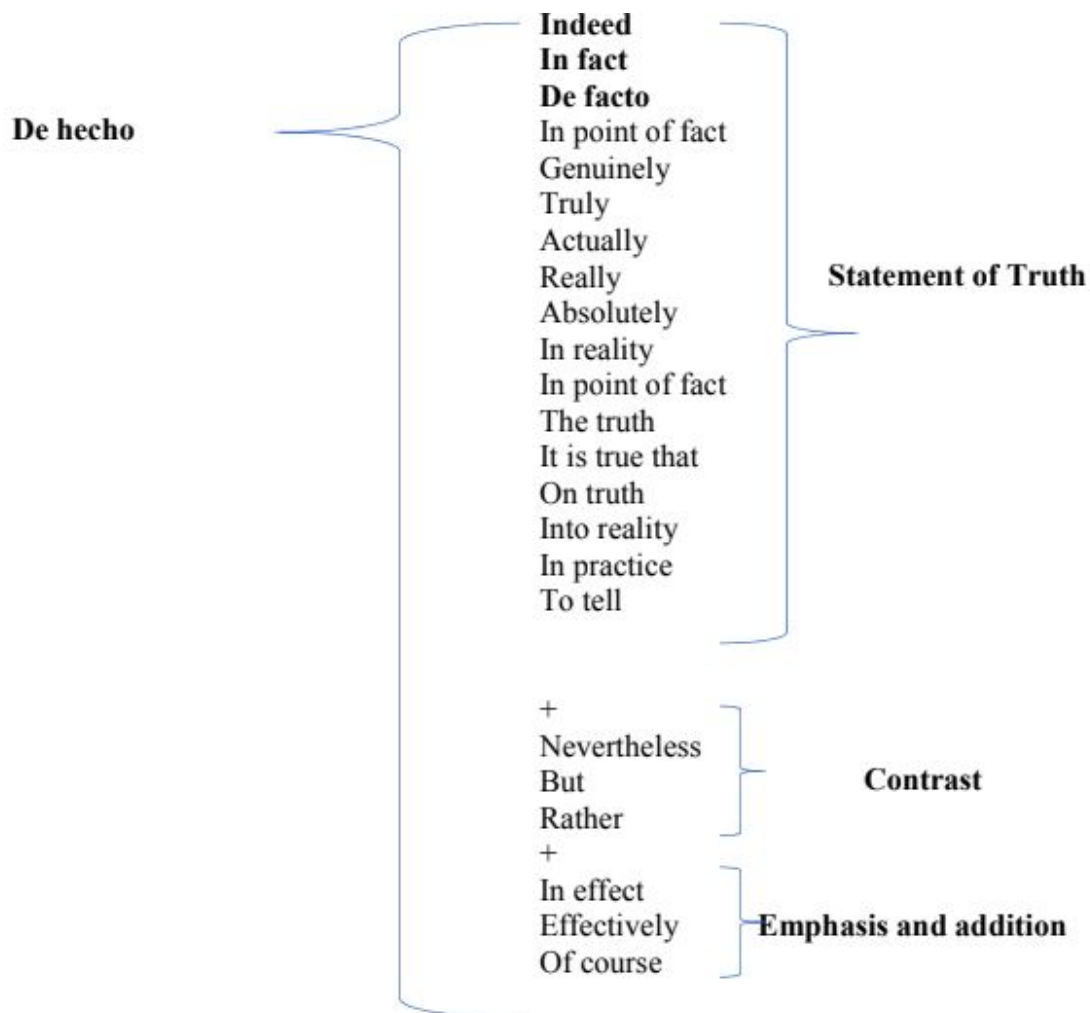


Figure 19: LDMs ‘de hecho’.

It is important to point out ‘de hecho’ as central, it is strongly polysemic and presents bilateral equivalence with the three English VMs (Halliday, 2014) (indeed, in fact, actually) presented in previous sections, the three of them with the same function ‘statement of truth’.

### 6.2.3 ‘La verdad’

In this section, as it has been done for English VMs (Halliday, 2014), the three research questions suggested in this thesis are answered. The study of ‘la verdad’ starts with a short analysis of monolingual categorizations of it by several researchers, then and based on a frequency table, translation preferences are shown with some examples, afterwards genre preference, once more based on the previous frequency table, is analysed. Finally, some additional comments regarding TLDs and correspondence potential are given.

#### 6.2.3.1 Monolingual analysis ‘la verdad’

These are the categorizations for ‘la verdad’:

- a) Confirmation (El Messaudi Rifi, 2018). This is, ensuring the truth of the previous statement.
- b) Modal connector (Catalina Fuentes, 2010). Expressing attitude or circumstance.
- c) Probability and contrast (Violeta de Monte, 1999). Expressing a condition or adversative tone.
- d) Evidentiality (Portolés and Zorraquino, 1999). Pragmatic attitude of the speaker or writer.

As in the case of ‘de hecho’, the only function which is not coincidental with English VMs (Halliday, 2014) functions found in previous sections, is the function addition. Due to the strong connection between English VMs (Halliday, 2014) and Spanish ‘Operadores de refuerzo argumentativo’ (Portolés and Zorraquino, 1999), and ‘la verdad’ corresponds to this last group. In conclusion, ‘addition’ is a function to be added to ‘la verdad’ analysis.

6.2.3.2 Translation correspondences ‘la verdad’

Answering the first research question, the most frequent VMs in English for ‘la verdad’ is ‘the truth’ 489 (415.03), which appears with the strongest frequency in the corpora Multinot fiction, essay, expe and speech (Lavid et al., 2015) and in Europarl7 (Tiedemann, 2007).

As shown in table 13, ‘the truth’ is the most frequent translation equivalent (Hummel, 1995) it all the corpora so this is a unique correspondence between markers in Spanish and English.

SPANISH ORIGINAL	ENGLISH TRANSLATION	Corpus frequency								TOTAL Per million
		Europarl 7 English 53,837,625	Open subtitles 1,000,000	Emea 5,455,537	Multinot fiction 13,164	Multinot essay 20,036	Multinot Expe 40,643	Multinot Speech 307,625	Multinot Web 16,553	
<b>La verdad</b>	Indeed	14 (0.23)			0	0		0	0	14 (0.23)
	Genuinely	0			0	0		0	0	0
	Truly	11 (0.18)			0	0		0	0	11 (0.18)
	Actually	21 (0.35)			0	0		0	0	21 (0.35)
	Really	30 (0.49)			0	0		0	0	30 (0.49)
	Absolutely	8 (0.13)			0	0		0	0	8 (0.13)
	In fact	14 (0.23)			0	0		0	0	14 (0.23)
	In reality	3 (0.05)			0	0		0	0	3 (0.05)
	In point of fact	1 (0.02)			0	0		0	0	1 (0.02)
	<b>The truth</b>	<b>466 (7.66)</b>			<b>2 (151.9)</b>	<b>3 (149.73)</b>		<b>1 (50.48)</b>	<b>17 (55.26)</b>	<b>489 (415.03)</b>
	It is true that	8 (0.13)			0	0		0	0	8 (0.13)
	Nevertheless	8 (0.13)	No entries	No entries	0	0	No entries	0	0	8 (0.13)
	On truth	9 (0.15)			0	0		0	0	9 (0.15)
	But	121 (1.99)			0	0		0	0	121 (1.99)
	In effect	0			0	0		0	0	0
	Into reality	1 (0.02)			0	0		0	0	1 (0.02)
	In practice	6 (0.1)			0	0		0	0	6 (0.1)
	To tell	33 (0.54)			0	0		0	0	33 (0.54)
	De facto	1 (0.02)			0	0		0	0	1 (0.02)
	In point of fact	1 (0.02)			0	0		0	0	1 (0.02)
Rather	15 (0.23)			0	0		0	0	15 (0.23)	
Effectively	7 (0.12)			0	0		0	0	7 (0.12)	
Of course	10 (0.16)			0	0		0	0	10 (0.16)	
<b>TOTAL Per million</b>		788 (12.95)			2 (151.9)	3 (149.73)		1 (50.48)	17 (55.26)	801 (420.32)

Table 13: translation equivalents (Hummel, 1995) of ‘la verdad’.

Translation examples of ‘la verdad’: Examples from Europarl7:

a) Nominal adverbial fronted phrase, explanation and clarification:

<s>**La verdad** es que el mercado único aún no está concluido. </s>

<s>**The truth** is that the single market is not yet complete. </s>

b) Sentence connector, correction and adversative:

<s>Voy a dedicarle poco tiempo, porque es, **la verdad**, un buen informe. </s>

<s>I will not take long, because it is **in fact** a very good report. </s>

c) Nominal sentence adverbial, fronted, enhancement:

<s>**La verdad** es que ya he indicado que sería mejor votar sobre mi informe lo antes posible porque -como soy uno de los que participan en las cuestiones relativas a la aviación, estoy más al corriente que muchos de lo que sucede- no hay más que un vuelo para regresar a mi país, y por lo tanto estoy entre la espada y la pared. </s>

<s>**Indeed** I indicated earlier that this needed to be taken as early as possible in the voting because - and as one of those involved in aviation, I know this better than most - I have only got one flight to get me home. </s><s>And so I am stuck! </s>

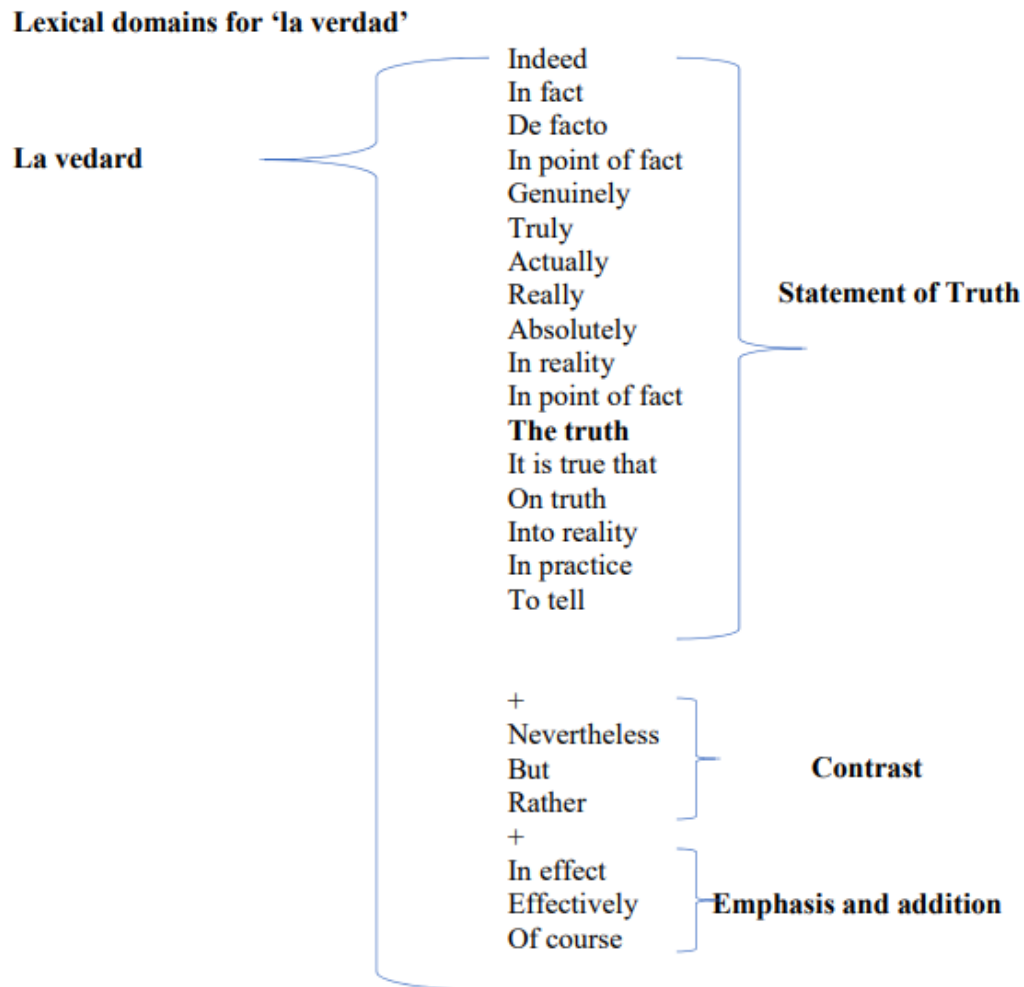
d) Sentence adverbial, evidentiality:

<s>**La verdad** es que no veo ningún motivo para aprobar la recomendación si no hay intención de aplicarla. </s>

<s>I **certainly** do not see the point in passing the recommendation if it was not intended to be implemented. </s>

**6.2.3.3 Translation lexical domains ‘la verdad’**

Figure 20 presents the lexical domains found for ‘la verdad’, it shows the most frequent translation occurrences for ‘la verdad’ and the function they perform in discourse when this VM is translated into each of these correspondences.



**Figure 20: LDMs ‘la verdad’.**

As seen in the previous figure, the lexical domains for ‘la verdad’ are the same as those found for ‘de hecho’, though there is just one strong translation equivalent ‘the truth’, which shows that ‘la verdad’ is more peripheral than ‘de hecho’.

#### 6.2.4 ‘En realidad’

In this section, as it has been done for English VMs (Halliday, 2014), the three research questions suggested in this thesis are answered. The study of ‘en realidad’ starts with a short analysis of monolingual categorizations of it by several researchers, then and based on a frequency table, translation preferences are shown with some examples, afterwards genre preference, once more based on the previous frequency table, is analysed. Finally, some additional comments regarding TLDs and correspondence potential are given.

##### 6.2.4.1 Monolingual analysis of ‘en realidad’

These are the categorizations for ‘en realidad’:

- a) Statement of Truth (El Messaudi Rifi, 2018). This is, checking the truthfulness of the statement.
- b) Modal connector (Catalina Fuentes, 2010). As quality of circumstance.
- c) Probability and contrast (Violeta de Monte, 1999). This is, conditional and adversative.
- d) Reinforcement (Portolés and Zorraquino, 1999). This is, emphasis or enhancement.

‘Indeed’ and ‘in fact’ are considered as VMs by Halliday (2014), as seen in the previous sections, these two VMs have as correspondence ‘en realidad’. As Aijmer (2004) considers several pragmatic functions for these two VMs, addition, contrast, emphasis and purpose, as well as statement of truth; these functions must be adhered to ‘en realidad’ as well. The only function which is not reflected in the Spanish monolingual analysis outline is ‘addition’, which should be added to the functions of ‘de hecho’ as part of the parallel translation analysis carried out in this thesis.

6.2.4.2 Translation occurrences for ‘en realidad’

Answering the first research question, the most frequent VMs in English for ‘en realidad’ is ‘in fact’ 263 (207.95), which appears with the strongest frequency in the corpora Multinot fiction and web (Lavid et al., 2015). Meanwhile, ‘in fact’ is the most frequent translation equivalent (Hummel, 1995) for ‘de hecho’ in the corpus Europarl7 (Tiedemann, 2007) 420 (6.9) and in Multinot essay (Lavid et al., 2015) 2 (99.82).

It is true, though, that, if looking at the absolute numbers the results seem to be the way around, but, as mentioned for the English VMs, as the absolute figures are based on the particular number of words of each corpus and the proportion per million is a common standard for all the corpora, the proportion per million is the one taken as reference. See the following table 14:

SPANISH ORIGINAL	ENGLISH TRANSLATION	Corpus frequency								TOTAL Per million
Corpus total Number of Words		Europarl 7 English 53,837,625	Open subtitles 1,000,000	Emea 5,455,537	Multinot fiction 13,164	Multinot essay 20,036	Multinot Expe 40,643	Multinot Speech 307,625	Multinot Web 16,553	
<b>En realidad</b>	<b>Indeed</b>	64 (1.05)			0	0		0	<b>18 (58.51)</b>	8 (259.56)
	Genuinely	11 (0.18)			0	0		0	0	11 (0.18)
	Truly	25 (0.41)			0	0		0	1 (3.25)	26 (3.66)
	<b>Actually</b>	376 (6.18)			0	0		<b>1 (50.48)</b>	1 (3.25)	378 (59.91)
	Really	158 (2.6)			0	0		0	0	158 (2.6)
	Absolutely	17 (0.28)			0	0		0	0	17 (0.28)
	<b>In fact</b>	221 (3.63)			<b>2 (151.92)</b>	<b>1 (49.91)</b>		0	4 (13)	<b>228 (116.46)</b>
	<b>In reality</b>	<b>493 (8.1)</b>			1 (75.96)	0		0	11 (35.75)	505 (114.81)
	In point of fact	6 (0.1)			0	0		0	0	6 (0.1)
	The truth	14 (0.23)			1 (75.96)	0		0	0	14 (0.23)
	It is true that	11 (0.18)			0	0		0	0	11 (0.18)
	Nevertheless	19 (0.31)	No entries	No entries	0	0	No entries	0	0	19 (0.31)
	On truth	0			0	0		0	1 (3.25)	0
	But	220 (3.61)			0	0		0	0	220 (3.61)
	In effect	43 (0.71)			0	0		0	0	43 (0.71)
	Into reality	37 (0.61)			0	0		0	0	37 (0.61)
	In practice	15 (0.25)			0	0		0	0	15 (0.25)
	To tell	3 (0.05)			0	0		0	0	6 (0.1)
	De facto	6 (0.1)			0	0		0	0	6 (0.1)
	In point of fact	6 (0.1)			0	0		0	0	18 (0.3)
	Rather	18 (0.3)			0	0		0	1 (3.25)	11 (0.18)
	Effectively	11 (0.18)			0	0		0	2 (6.50)	13 (0.21)
Of course	13 (0.21)			0	0		0	0		
<b>TOTAL Per million</b>		1787 (29.28)			<b>4 (303.84)</b>	1 (49.91)		1 (50.48)	39 (126.76)	1832 (560.27)

Table 14: ‘en realidad’ translation equivalents (Hummel, 1995).

The results for ‘en realidad’ are like those found for ‘de hecho’, functionally speaking and regarding LDMS. See translation examples below:

a) Sentence connector, correction:

<s>En realidad, después de nuestra llegada tuvimos que esperar 45 minutos al autobús. </s>

Europarl7

<s>In fact, it took 45 minutes after we arrived before the bus appeared at the airport. </s>

Europarl7

b) Sentence connector, adversative:

<s>En realidad, esto se oye cada vez menos. </s> Europarl7

<s>Actually, I hear less and less about this. </s> Europarl7

c) Sentence adverbial, fronted, corroboration:

<s>En realidades algo que, de facto, se considera imposible. </s> Europarl7

<s>Certainly it is perceived de facto to be impossible. </s> Europarl7

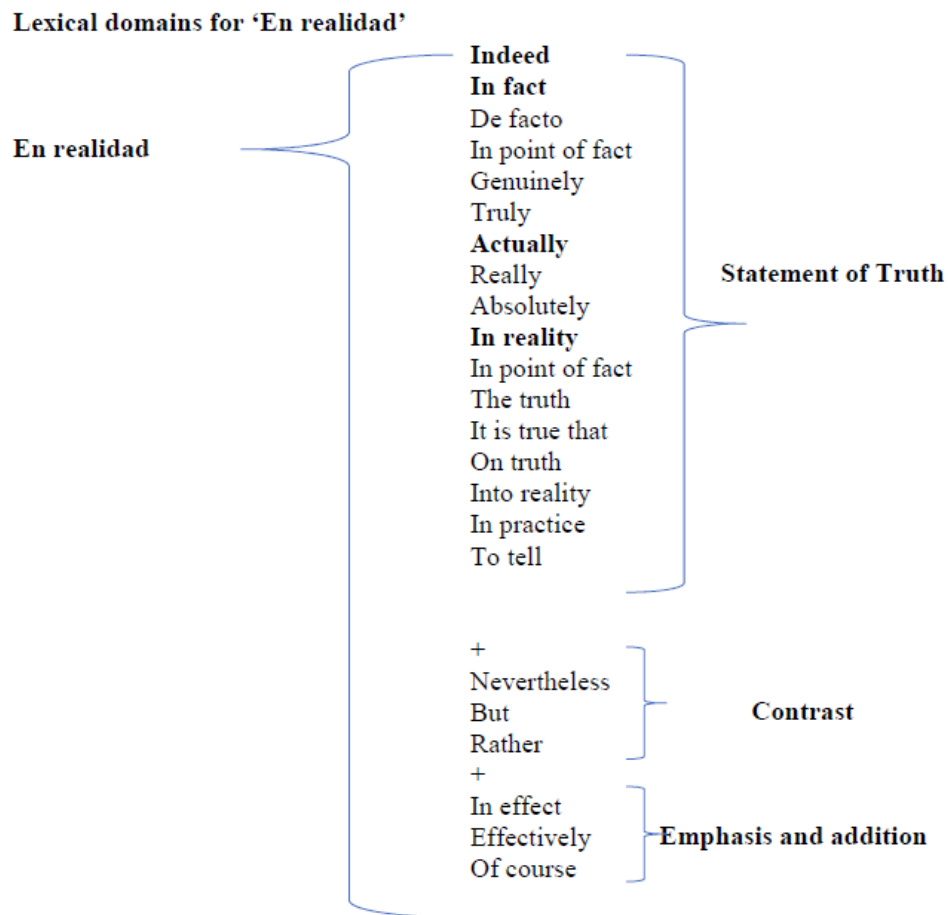
d) Sentence connector, enhancement:

<s> En realidad, con la adopción de la Declaración sobre la supervivencia, la protección y el desarrollo del niño y del Plan de Acción para su aplicación para los años 90, que constituye el documento final de la Cumbre Mundial en favor de la Infancia, quedaron claramente establecidas nuestras responsabilidades individuales y colectivas. </s> MULTINOTSPPECH 003

<s> Indeed, the adoption of the Declaration and Plan of Action on the Survival, Protection and Development of Children for the 1990s – the outcome document of the World Summit for Children – clearly sets out our individual and collective responsibilities. </s> MULTINOTSPPECH 003

**6.2.4.3 Lexical domains for ‘en realidad’**

Figure 21 presents the lexical domains found for ‘en realidad’ are strongly similar to those of ‘de hecho’, being these two ‘operadores de refuerzo argumentativo’ (Portolés and Zorraquino, 1999) the most central.



**Figure 21: LDMs ‘en realidad’.**

As seen in the previous figure, the lexical domains for ‘en realidad’ are the same as those found for ‘de hecho’.

### 6.2.5 ‘Realmente’

In this section, as it has been done for English VMs (Halliday, 2014), the three research questions suggested in this thesis are answered. The study of ‘realmente’ starts with a short analysis of monolingual categorizations of it by several researchers, then and based on a frequency table, translation preferences are shown with some examples, afterwards genre preference, once more based on the previous frequency table, is analysed. Finally, some additional comments regarding TLDs and correspondence potential are given.

#### 6.2.5.1 Monolingual study ‘Realmente’

These are the categorizations for ‘realmente’:

- a) Statement of Truth (El Messaudi Rifi, 2018). This is, checking the truthfulness of the statement.
- b) Modal connector (Catalina Fuentes, 2010). As quality of circumstance.
- c) Probability and contrast (Violeta de Monte, 1999). This is, conditional and adversative.
- d) Reinforcement (Portolés and Zorraquino, 1999). This is, emphasis or enhancement.

‘Indeed’ and ‘in fact’ are considered as VMs by Halliday (2014), as seen in the previous sections, these two VMs have as correspondence ‘realmente’. As Aijmer (2004) considers several pragmatic functions for these two VMs, addition, contrast, emphasis and purpose, as well as statement of truth; these functions must be adhered to ‘en realidad’ as well. The only function which is not reflected in the Spanish monolingual analysis outline is ‘addition’, which should be added to the functions of ‘realmente’ as part of the parallel translation analysis carried out in this thesis.

6.2.5.2 Translation correspondences of ‘realmente’

Based on table 15 below and answering the first research question, the most frequent VMs in English for ‘realmente’ is ‘truly’ 541 (156.7) in all the corpora so this is a unique correspondence between markers in Spanish and English.

SPANISH ORIGINAL	ENGLISH TRANSLATION	Corpus frequency								
		Europarl 7 English	Open subtitles Spanish	Emea Spanish	Multinot fiction SO	Multinot essay SO	Multinot Web SO	Multinot Expe SO	Multinot Speech SO	TOTAL Per million
Corpus Total number of Words										
	Indeed	102 (1.68)				0			10 (32.50)	112 (34.18)
	Genuinely	454 (7.46)				1 (49.91)			2 (6.50)	457 (63.87)
	<b>Truly</b>	<b>494 (8.12)</b>				<b>2 (99.82)</b>			<b>15 (48.76)</b>	<b>513 (156.70)</b>
	Actually	330 (5.42)				0			3 (9.75)	333 (15.17)
	Really	12 (0.2)				0			4 (13)	16 (13.2)
	Absolutely	28 (0.46)				0			0	28 (0.46)
	In fact	99 (1.63)				0			0	99 (1.63)
	In reality	76 (1.25)				0			0	76 (1.25)
	In point of fact	5 (0.08)				0			0	5 (0.08)
	The truth	30 (0.49)				0			0	30 (0.49)
	It is true that	15 (0.25)				0			0	15 (0.25)
<b>Realmente</b>	Nevertheless	20 (0.33)	No entries	No entries	No entries	0	No entries	No entries	0	20 (0.33)
	On truth	0				0			0	0
	But	166 (2.73)				0			0	166 (2.73)
	In effect	16 (0.26)				0			0	16 (0.26)
	Into reality	3 (0.05)				0			0	3 (0.05)
	In practice	34 (0.56)				0			0	34 (0.56)
	To tell	0				0			0	0
	De facto	8 (0.13)				0			0	8 (0.13)
	In point of fact	5 (0.08)				0			0	5 (0.08)
	Rather	41 (0.67)				0			0	41 (0.67)
	Effectively	39 (0.64)				0			0	39 (0.64)
	Of course	31 (0.51)				0			0	31 (0.51)
<b>TOTAL Per million</b>		2013 (33)				<b>3 (149.73)</b>			34 (110.51)	2050 (283.24)

Table 15: translation equivalents (Hummel, 1995) of ‘realmente’.

The case of ‘realmente’ is very similar to that of ‘la verdad’ and thus the functions found are very similar. See the translation examples below:

a) Sentence adverbial, evidentiality:

<s>Para que el mercado de transportes común **realmente** funcione no sólo es importante que existan reglas, sino que también esas reglas, en la medida de lo posible, sean comunes. </s> EU-ROPARL7

<s>If we are to get a common transport market **genuinely** up and running, it is important that we should not only have regulations but that these regulations should also, as far as possible, apply to every country. </s> EUROPARL7

b) Sentence adverbial, certainty, truthfulness:

<s>Cualquier análisis coste-beneficio, que contabilice **realmente** todo el daño a las personas y al medio ambiente de estos accidentes, nos dará la razón.</s> EUROPARL7

<s>Any cost-benefit analysis which **truly** takes into account all the harm occasioned to persons and the environment by these accidents, is bound to support this approach. </s> EUROPARL7

c) Sentence adverbial, confirmation:

<s> Mi pregunta es: ¿ está **realmente** dispuesta la Comisión a cumplir con las directrices aprobadas para el gender mainstreaming en la Dirección General para las Ayudas al Desarrollo? </s> EUROPARL7

<s>My question is this: is the Commission **really** prepared to comply with the adopted guidelines for gender mainstreaming at the Directorate-General for Development? </s> EUROPARL7

d) Sentence adverbial, statement:

<s> Esto es **realmente** lo que constituye una alianza.</s> MULTINOT-SPEECH 001

<s> That is what partnership **truly** is all about.</s> MULTINOT-SPEECH 001

e) Connector, embedded, enhancement:

<s> En comparación con otros países, los niños de mi país, Islandia, son **realmente** afortuna-

dos puesto que gozan de la bendición de la paz y la seguridad, de una buena atención médica y de educación.</s> MULTINOT-SPEECH 002

<s> By international comparison, the children of my country, Iceland, are **indeed** fortunate, as they are blessed with peace and security, good health care and education. We must, however, continue to acknowledge </s> MULTINOT-SPEECH 002

### 6.2.5.3 Genre-specific preferences Spanish Markers

Research question3): *'are there any genre-specific preferences for the source Spanish markers and their target translation?'*

Based on the characteristics which define each of the corpora Europarl7 (Koehn, 2001), Opus2 (D. Dobrovolskij and L. Pöppel, 2016) and Multinot (Lavid et al., 2015)'; the following figure will provide an analysis of the genre specific preference of the Spanish markers 'realmente', 'la verdad', 'en realidad' and 'de hecho' according to the English translation equivalents (Hummel, 1995) found in the above mentioned corpora.

De hecho	formal language	Written	Expository	in fact	
			Argumentative	in fact	
			Rhetorical	indeed	
			Analytical	0	
Informal language	Spoken	Spoken	Expository	indeed/ in fact	
			Argumentative	indeed/ in fact	
		Spoken	Dialogues	0	
En realidad	formal language	Written	Expository	in fact/ actually	
			Argumentative	in fact/ actually	
			Rhetorical	in fact	
			Analytical	0	
	Informal language	Spoken	Spoken	Expository	indeed/ in reality
				Argumentative	indeed/ in reality
			Spoken	Dialogues	0
La verdad	formal language	Written	Expository	in truth	
			Argumentative	in truth	
			Rhetorical	in truth	
			Analytical	0	
	Informal language	Spoken	Spoken	Expository	in truth
				Argumentative	in truth
			Spoken	Dialogues	0
Realmente	formal language	Written	Expository	truly	
			Argumentative	truly	
			Rhetorical	0	
			Analytical	0	
	Informal language	Spoken	Spoken	Expository	truly
				Argumentative	truly
			Spoken	Dialogues	0

Figure 22: Genre preference for ‘realmente’, ‘la verdad’, ‘en realidad’ and ‘de hecho’ based on their occurrence in the corpora Multinot (Lavid et al., 2015), Opus2 (D. Dobrovolskij and L. Pöppel, 2016) Europarl7 (Koehn, 2001).

The previous figure outlines the genre preference for each Spanish marker translation based on their most frequent translation equivalent (Hummel, 1995) found in each corpus. The classification of each genre has been based on the characteristics of the corpora used in the analysis of translation equivalence as follows:

- a) Europarl7 (Koehn, 2001): it represents a formal spoken argumentative and expository genre as it offers transcriptions of legal and parliamentary discourse.
- b) EMEA (Tiedemann, 2007): it represents analytical genre as it reproduces medical texts.
- c) Open Subtitles (Tiedemann, 2007): it represents informal spoken language as it corresponds to the transcriptions of the subtitles based on film dialogues.
- d) Multinot Expe and Essay (Lavid et al., 2015): it represents formal written argumentative and expository language, the former as expository texts and the latter as scientific and educational essays.
- e) Multinot Speech and Web (Lavid et al., 2015): it represents a formal spoken argumentative and expository genre as it offers transcriptions of legal and parliamentary discourse.
- f) Multinot Fiction (Lavid et al., 2015): it represents formal written rhetorical texts as it is based on extracts from English language literature.

The analysis based on the previous figure which represents genre preference for ‘realmente’, ‘la verdad’, ‘en realidad’ and ‘de hecho’ based on their occurrence in the corpora Multinot (Lavid et al., 2015), Opus2 (D. Dobrovolskij and L. Pöppel, 2016), Europarl7 (Koehn, 2001); reveals that:

- a) Written formal expository argumentative texts would prefer ‘in fact’ as translation equivalent (Hummel, 1995) of ‘de hecho’; ‘actually’ and ‘in fact’ as translation equivalents of ‘en realidad’; ‘the truth’ as translation equivalent (Hummel, 1995) of ‘la verdad’ and ‘truly’ as translation equivalent of ‘realmente’.
- b) Spoken formal expository argumentative texts would prefer ‘indeed’ and ‘in fact’ as translation equivalents (Hummel, 1995) of ‘de hecho’; ‘indeed’ and ‘in reality’ for ‘en realidad’; ‘the truth’ for ‘la verdad’ and ‘truly’ for ‘realmente’.
- c) Written formal analytical texts have no occurrence for these Spanish markers.
- d) Written rhetorical texts would prefer ‘indeed’ as translation equivalent (Hummel, 1995) (Hummel, 1995) of ‘de hecho’, ‘in fact’ as translation equivalent (Hummel, 1995) of ‘en realidad’ and ‘the truth’ as translation equivalent of ‘la verdad’ (Hummel, 1995). In this genre ‘realmente’ does not have occurrences.
- e) Dialogues have no occurrence for these Spanish markers.

As mentioned in the answers provided to the first research question ‘which are the translation preferences of the Spanish Markers’, ‘realmente’ prefers ‘truly’, ‘la verdad’ prefers ‘in truth’, ‘en realidad’ prefers ‘in fact’ and ‘de hecho’ prefers ‘indeed’.

*What are the genre specific preferences of each Spanish VM?*

- 1) 'De hecho' prefers formal written rhetorical texts as well as formal spoken argumentative and expository texts.
- 2) 'En realidad' prefers formal written rhetorical texts as well as formal written expository and argumentative texts.
- 3) 'La verdad' is found in all formal genera except for analytical and in no informal genera dialogues.
- 4) 'Realmente' is found in all formal genera except for analytical and rhetorical and in no informal genera dialogues.

**6.2.5.4 Part two of the process for Spanish markers explained: research question 2): back-translation**

A process of back translation will be carried out. This process of back translation consists of tracing the Spanish equivalents in translation for the English translation equivalents (Hummel, 1995) found for ‘de hecho’, ‘en realidad’, ‘la verdad’ and ‘realmente’.

The goal of this ‘back translation’ process is to find lexical fields or semantic groupings for these four markers, answering to the research question ‘2) what are the translation domains revealed?’. These back translations will be illustrated by the following tables:

Back-translation ‘la verdad’:

English Back translation	Corpus Spanish occurrences La verdad							
	Europarl 7 English	Open subtitles English	Emea English	Multinot fiction - EO	Multinot essay - EO	Multinot Expe-EO	Multinot Speech-EO	Multinot WEB- EO
<b>The truth</b>	La verdad En realidad Los hechos De hecho Exactamente Es más Verdaderamente Es decir	La verdad Los hechos Exactamente Sinceramente Ciertamente	No entries	No entries	No entries	La verdad	La verdad	No entries

**Table 16: Corpus Spanish occurrences ‘la verdad’.**

- a) ‘La verdad’, ‘en realidad’, ‘los hechos’, ‘de hecho’, ‘verdaderamente’: these examples are to be expected as they correspond to ‘operadores de refuerzo argumentativo’ (Portolés and Zorraquino, 1999).
- b) ‘Es más’: this was found as one of the most frequent translation equivalents (Hummel, 1995) for ‘in fact’, and it is categorized as reformulation-enhancement (Cortés and Camacho, 2005). So, there is connection between ‘in fact’ and ‘la verdad’.
- c) ‘Exactamente’, ‘sinceramente’ and ‘ciertamente’: del Saz (2016) considers these markers with a function of modification and reformulation, a function which has already been addressed to Spanish markers of argumentative reinforcement (Portolés and Zorraquino, 1999).

- d) ‘Es decir’: this is one of the most frequent translation equivalents for ‘indeed’ (Hummel, 1995) as explained in section 6, whilst, there is connection between ‘la verdad’ and the verificative marker (Halliday, 2014) ‘indeed’.

#### Back-translation ‘de hecho’

English Back translation	Corpus Spanish occurrences de hecho							
	Europarl 7 English	Open subtitles English	Emea English	Multinot fiction - EO	Multinot essay - EO	Multinot Expe-EO	Multinot Speech-EO	Multinot WEB- EO
<b>Indeed</b>	En efecto Realmente Sin duda Ciertamente Así Más bien Incluso	Ya que De hecho Así Realmente En verdad (más) bien Lamentablemente	De hecho En efecto Ciertamente Concretamente Es decir Sobre todo Pero Tal	No entries	De hecho Claramente	Realmente De hecho En efecto Ya (que) Precisamente Aún	Realmente De hecho Efectivamente En efecto	
<b>In fact</b>	De hecho Efectivamente Bueno Pero	En realidad De hecho	Realmente Aparentemente En realidad De hecho Sin embargo Incluso	Es más	En realidad De hecho	En realidad Realmente La verdad Lo cierto Aparentemente Más que	En realidad De hecho	
<b>In point of fact/ de facto</b>	En la práctica De hecho En realidad A decir verdad En efecto Realmente En realidad Además La verdad Ciertamente	Al fin al cabo De hecho Precisamente Realmente En realidad Efectivamente	No entries	No entries	No entries	No entries	No entries	No entries

**Table 17: Corpus Spanish occurrences ‘de hecho’.**

- a) ‘De hecho’, ‘a decir verdad’, ‘en realidad’, ‘realmente’ and ‘la verdad’: these Spanish markers are to be expected as they belong to the ‘operadores de refuerzo argumentativo’.
- b) ‘Claramente’, ‘lamentablemente’, ‘precisamente’, ‘concretamente’, ‘aparentemente’, ‘sin duda’, ‘en la práctica’, ‘al fin y al cabo’ and ‘ciertamente’: del Saz (2016) considers these markers with a function of modification, reformulation and or hypothesis-condition and or conclusion a functions which have already been addressed to Spanish markers of argumentative reinforcement (Portolés and Zorraquino, 1999).

- c) ‘Es decir’: this is one of the most frequent translation equivalents (Hummel, 1995) for ‘indeed’, whilst, there is connection between ‘de hecho’ and the verificative marker (Halliday, 2014) ‘indeed’.
- d) ‘En efecto’ and ‘efectivamente’: these are very linked to ‘indeed’ and thus will be ‘de hecho’, the Rae (2019) considers ‘en efecto’ as an adverbial of purpose, which can be extraposed to ‘de hecho’ as a function.
- e) ‘Ya’, ‘pero’, ‘sin embargo’, ‘incluso’ and ‘sobre todo’ are very frequent translation equivalents (Hummel, 1995) of ‘indeed’ which reinforces the premise of the strong link between ‘indeed’ and ‘de hecho’. These adverbials have a strong adversative, contrastive additive and emphatic functions (Quirk, 1990).
- f) ‘Así’: this is one of the most frequent translation equivalents (Hummel, 1995) of ‘actually’, consequently, not only there is connection between ‘indeed’ and ‘de hecho’ but also with ‘actually’. Portolés and Zorraquino (1999) consider for ‘así’ the function of consecutive connector, this is, expressing a consequence or result.
- g) ‘Más bien’, ‘aún’, ‘bueno’ and ‘tal’: del Saz (2016) considers ‘más bien’ as modification reformulator, ‘aún’ as reassessment reformulator; Cortes y Camacho (2005) she ‘bueno’ as a reformulator of rectification; finally, ‘tal’ is a discourse organizer. Reformulation is a function addressed for ‘de hecho’, these specific functions of modification, reassessment, organisation are to be added to both argumentative markers (Portolés and Zorraquino, 1999) and Verificative markers (Halliday, 2014).

Back-translation ‘en realidad’

English Back translation		Corpus Spanish occurrences de hecho						
	Europarl 7 English	Open subtitles English	Emea English	Multinot fiction - EO	Multinot essay - EO	Multinot Expe-EO	Multinot Speech-EO	Multinot WEB- EO
<b>Indeed</b>	En efecto Realmente Sin duda Ciertamente Así Más bien Incluso	Ya que De hecho Así Realmente En verdad En verdad (más) bien Lamentablemente De verdad En realidad	De hecho En efecto Ciertamente Concretamente Es decir Sobre todo Pero Tal	No entries	De hecho Claramente	Realmente De hecho En efecto Ya (que) Precisamente Aún	Realmente De hecho Efectivamente En efecto	
<b>Actually</b>	De hecho Efectivamente Incluso Aunque Pero En realidad Realmente	Realmente De hecho En verdad La verdad Ya Pues Así Actualmente Pero	Realmente En realidad De hecho Especialmente (más) bien Sin embargo Aunque Incluso	Realmente De verdad	En realidad En la realidad En los hechos	Realmente De verdad	No entries	
<b>In fact</b>	De hecho Efectivamente Bueno Pero	En realidad De hecho	Realmente Aparentemente En realidad De hecho Sin embargo Incluso	Es más	En realidad De hecho	En realidad Realmente La verdad Lo cierto Aparentemente Más que	En realidad De hecho	
<b>In reality</b>	En realidad Realmente (En)la realidad De facto	En realidad La verdad De verdad Realmente	No entries	No entries	En realidad	Lo cierto En realidad	No entries	No entries

**Table 18: Corpus Spanish occurrences ‘en realidad’.**

- a) ‘De hecho’, ‘a decir verdad’, ‘en realidad’, ‘realmente’ and ‘la verdad’: these Spanish markers are to be expected as they belong to the ‘operadores de refuerzo argumentativo’.
- b) ‘Claramente’, ‘lamentablemente’, ‘precisamente’, ‘concretamente’, ‘aparentemente’, ‘sin duda’, ‘en la práctica’, ‘al fin y al cabo’ and ‘ciertamente’: del Saz (2016) considers these markers with a function of modification, reformulation and or hypothesis-condition and or conclusion a functions which have already been addressed to Spanish markers of argumentative reinforcement (Portolés and Zorraquino, 1999) .

- c) ‘Es decir’: this is one of the most frequent translation equivalents (Hummel, 1995) for ‘indeed’ as explained in section 6, whilst, there is connection between ‘en realidad’ and the verificative marker (Halliday, 2014) ‘indeed’.
- d) En efecto and efectivamente: these are very linked to ‘indeed’ and thus will be ‘en realidad’, the Rae (2019) considers ‘en efecto’ as an adverbial of purpose, which can be extraposed to ‘de hecho’ as a function.
- e) ‘Ya’, ‘pero’, ‘sin embargo’, ‘incluso’ and ‘sobre todo’ are very frequent translation equivalents (Hummel, 1995) of ‘indeed’ which reinforces the premise of the strong link between ‘indeed’ and ‘en realidad’. These adverbials have a strong adversative, contrastive additive and emphatic functions (Quirk, 1990).
- f) ‘Así’: this is one of the most frequent translation equivalents (Hummel, 1995) of ‘actually’, consequently, not only there is connection between ‘indeed’ and ‘en realidad’ but also with ‘actually’. Portolés and Zorraquino (1999) consider for ‘así’ the function of consecutive connector, this is, expressing a consequence or result.
- g) ‘Más bien’, ‘aún’, ‘bueno’ and ‘tal’: del Saz (2016) considers ‘más bien’ as modification reformulator, ‘aún’ as reassessment reformulator; Cortes y Camacho (2005) she ‘bueno’ as a reformulator of rectification; finally, ‘tal’ is a discourse organizer. Reformulation is a function addressed for ‘en realidad’, these specific functions of modification, reassessment, organisation are to be added to both argumentative markers (Portolés and Zorraquino, 1999) and Verificative markers (Halliday, 2014).

Back-translation ‘realmente’

English Back translation	Corpus Spanish occurrences de hecho							
	Europarl 7 English	Open subtitles English	Emea English	Multinot fiction - EO	Multinot essay - EO	Multinot Expe-EO	Multinot Speech- EO	Multinot WEB- EO
<b>Truly</b>	Verdaderamente Realmente En realidad Con certeza Auténticamente Efectivamente Sinceramente Ciertamente Naturalmente	Sinceramente Verdaderamente De verdad Realmente En verdad La verdad	Realmente Ciertamente	No entries	No entries	La verdad Realmente de verdad	Realmente verdaderamente	No entries

**Table 19: Corpus Spanish occurrences ‘realmente’.**

- a) ‘La verdad’, ‘en realidad’, ‘los hechos’, ‘de hecho’, ‘verdaderamente’: these examples are to be expected as they correspond to ‘operadores de refuerzo argumentativo’ (Portolés and Zorraquino, 1999).
- b) ‘Exactamente’, ‘sinceramente’, ‘con certeza’, ‘auténticamente’, ‘naturalmente’ and ‘ciertamente’: del Saz (2016) considers these markers with a function of modification and reformulation, a function which has already been addressed to Spanish markers of argumentative reinforcement (Portolés and Zorraquino, 1999).

### 6.2.5.5 Lexical domains Spanish connectives

Research question two ‘are there any lexical domains?’. This question will be answered through the analysis of the following figure obtained as a result of the analysis of the results obtained in the previous tables which display the backtranslation occurrences for the Spanish Markers ‘de hecho’, ‘en realidad’, ‘la verdad’ and ‘realmente’:

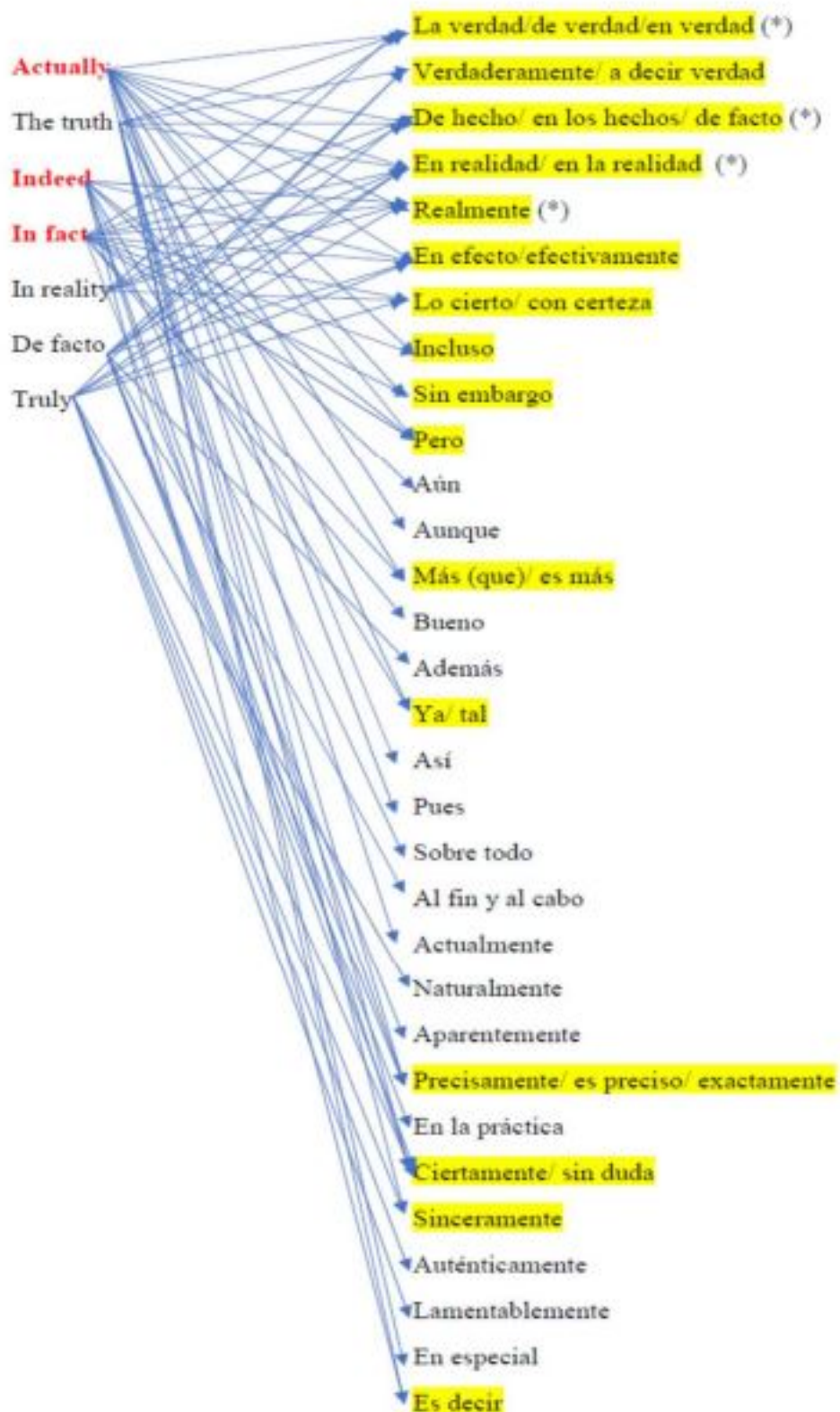


Figure 23: the semantic fields/ lexical domains for Spanish Markers ‘realmente’, ‘de hecho’, ‘la verdad’ and ‘en realidad’ based on the process of back-translation of their respective translation correspondences.

According to the previous figure which represents the semantic fields/ lexical domains for Spanish Markers ‘realmente’, ‘de hecho’, ‘la verdad’ and ‘en realidad’ based on the process of back-translation of their respective translation correspondences, an analysis of the semantic fields according to meaning an function will be displayed in the following figure:

---

**Semantic fields:** ‘En realidad’, ‘De hecho’, ‘Realmente’ and ‘La verdad’:

**Operador de refuerzo argumentativo:**

La verdad  
 De verdad  
 En verdad  
 De hecho  
 En los hechos  
 De facto  
 En realidad  
 En la realidad  
 Realmente

**Modification and reformulation:**

Precisamente  
 Es preciso  
 Exactamente  
 Ciertamente  
 Sin duda  
 Sinceramente  
 Más bien

**Reassessment and reformulation:**

Aún

**Rectification and reformulation:**

Bueno

**Explanation and reformulation:**

Es decir

**Discourse organizer:**

Tal

**Purpose:**

En efecto  
 Efectivamente  
 Adversative:  
 Pero  
 Sin embargo

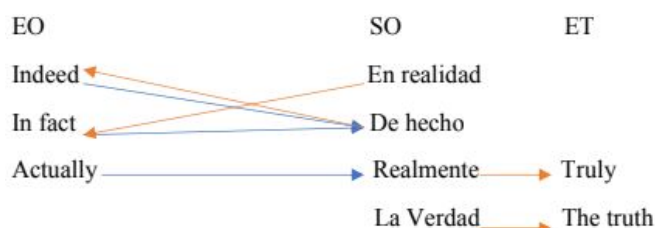
**Emphasis**

Ya  
 Tal  
 Es más  
 Más bien  
 Incluso

**Figure 24: Semantic fields for ‘en realidad’, ‘realmente’, ‘de hecho’ and ‘la verdad’ based of the most frequent translation equivalents (Hummel, 1995) of their back-translations in Multinot Spanish (Lavid et al., 2015), Opus2 Spanish (D. Dobrovolskij and L. Pöppel, 2016)., Europarl7 Spanish (Koehn, 2001).**

### 6.2.5.6 Analysis English-Spanish:

The following figure shows the strong relationship between ‘en realidad’-‘de hecho’ and ‘indeed’-‘in fact’, they represent the most central verificative and argumentative markers (Halliday, 2014; Portolés and Zorraquino, 1999). Please note that ‘EO’ refers to ‘English original’, ‘SO’ refers to ‘Spanish original’ and ‘ET’ refers to ‘English translation’.



**Figure 25: EO (English original), SO (Spanish original), ET (English translation). Arrows point out translation.**

### 6.2.5.7 Frequency of each marker according to the analysed corpora:

The following figure represents the most frequent markers per genre or corpus in English and in Spanish, ‘indeed’ is the most frequent verificative marker (Halliday, 2014), whereas ‘realmente’ is the most frequent ‘operador de refuerzo argumentativo’ (Portolés and Zorraquino), it is significant to note that whereas English prefers a central marker Spanish chooses a more peripheral one, thus the contrast between the two languages.

	Europarl 7	EMEA	Open Subtitles	Multinot
Indeed	293.8 pm	1.88 pm	35.53 pm	25 pm
Actually	22.66 pm	5.58 pm	248.73 pm	4 pm
In fact	237.18 pm	0.71 pm	14.83 pm	22 pm
En realidad	92.49 pm	0pm	0pm	24 pm
De hecho	121.91 pm	0pm	0pm	20 pm
Realmente	314.84 pm	0pm	0pm	5 pm
La verdad	32.07 pm	0pm	0pm	1pm

**Figure 26: Frequency of each marker according to the analysed corpora: Multinot (Lavid et al., 2015), Opus2 (D. Dobrovolskij and L. Pöppel, 2016), Europarl7 (Koehn, 2001). Pm: per million, Multinot (Lavid et al., 2015) figures are absolute.**

### 6.2.6 Summary of findings extension

In this sections I include additional comments and analysis regarding the findings in the process of this thesis regarding research questions: additional functions, translation preference, genre preference and lexical domains. Both English VMs (Halliday, 2014) (indeed, in fact, actually) and ‘operadores de refuerzo argumentativo’ (Portolés and Zorraquino, 1999) (la verdad, en realidad, de hecho, realmente).

#### 6.2.6.1 Indeed:

What follows is an in-depth analysis of the resulting proportions per million:

- a) ‘De hecho’ is the most frequent translation equivalent (Hummel, 1995) for ‘indeed’ in the following corpora: Emea (Tiedemann, 2007); Multinot essay, expe and speech (Lavid et al., 2015): 0.13, 215.7, 147.62 and 27.27 per million, respectively.
- b) ‘Realmente’: most frequent translation equivalent (Hummel, 1995) for ‘indeed’ in Multinot speech (Lavid et al., 2015): 27.27 per million.
- c) ‘Efectivamente’ is the most frequent translation equivalent (Hummel, 1995) for ‘indeed’ in Multinot speech (Lavid et al., 2015): 27.27 per million.
- d) ‘En efecto’ is the most frequent translation equivalent (Hummel, 1995) for ‘indeed’ in EMEA (Tiedemann, 2007) and Multinot speech (Lavid et al., 2015): 0.13 per million and 27.27 per million respectively.
- e) ‘Ya (que)’ is the most frequent translation equivalent (Hummel, 1995) for ‘indeed’ in Open Subtitles (Tiedemann, 2007): 0.27 per million.
- f) ‘Es decir’, ‘sobre todo’ and ‘pero’ some of the most frequent translation equivalents (Hummel, 1995) for ‘indeed’ in EMEA (Tiedemann, 2007): 0.13 per million in the three cases.

g) ‘Incluso’ is the most frequent translation equivalent (Hummel, 1995) for ‘indeed’ in the corpus Europarl7 (Koehn, 2001): 13.8 per million.

**‘Indeed’: calculating the correspondence degree:**

The correspondence degree illustrates which is the rate of equivalence between one source marker, in this case ‘indeed’, and a target one, in this case ‘incluso’, ‘ya’, ‘de hecho’, ‘en efecto’, ‘sobre todo’, ‘pero’, ‘realmente’ and ‘efectivamente’.

**‘Indeed’: calculating the correspondence degree:**

$$\text{Correspondence potential} = \text{number of correspondences} + \text{correspondence degree}$$

Europarl 7 English	Open subtitles	Emea	Multinot fiction	Multinot essay	Multinot expe	Multinot speech	Multinot web
Incluso	ya (que)	de hecho En efecto Sobre todo Pero	0	de hecho	realmente de hecho efectivamente en efecto	0	

**Table 20: ‘indeed’ correspondence degree.**

**‘Indeed’: calculating the correspondence potential:**

The correspondence potential is the likeability each translation equivalent has to occur as correspondence of the source marker, in this case ‘indeed’.

**De hecho** (highest correspondence potential).

Incluso.

Ya (que).

En efecto.

Realmente.

Efectivamente.

**Sobre todo** (smallest correspondence potential).

**Por ejemplo** (smallest correspondence potential).

**Figure 27: ‘indeed’ correspondence potential.**

Different functionality found according to monolingual translation, sentence occurrence, monolingual analysis, translation correspondences: (Aijmer, 2004)

- ‘Epistemic adverb of certainty’, this can be seen in the translation spottings found in the latter corpora as for the following cases: efectivamente, justamente, indudablemente, ciertamente, sinceramente, naturalmente, fundamentalmente, supuestamente, evidentemente, obviamente, aparentemente, precisamente, exactamente, verdaderamente.
- This function of ‘certainty’ is also conveyed by other translation counterparts such as: de verdad, en verdad, en efecto, en rigor, sin duda – prepositional phrases, la verdad -nominal phrases, tal vez -adverbial phrases, es cierto que – nominal verbal phrases.
- Particularizing functions may also be conveyed by the following translation spottings: por ejemplo, por cierto, en concreto, en especial, en este/ese sentido – prepositional phrases, exactamente, prácticamente, concretamente, solamente, especialmente, suficientemente- adverbs, más que, tan solo, tal, tal como, apenas – quantifiers.
- Additive functions could be illustrated as well by the following translation spottings: también, además, incluso – adverbs.
- Whereas rectification functions might be seen in the following spottings: en absoluto, en el fondo, en la práctica – prepositional phrases, mejor dicho, Bueno, más bien – adjuncts, lo cierto es que – propositions.
- Enhancement and verification could also be conveyed by some additional spottings: ya, así, lamentablemente, aún, todavía- adverbs, por supuesto, ni siquiera, sobre todo- prepositional phrases.
- Adversative functions are to be seen in the following translation spotting cases: pues, es decir, sea, acaso, pero, por el contrario, al contrario, sin embargo, en cambio, no obstante, aunque, a pesar de que, which are a combination of conjunctions, prepositional phrases and adverbs.

Noting the functions addressed by some authors such as Portolés and Zorraquino (1999), Cuenca (2001), del Saz (2003) or Garcés Gómez (2005), just to mention some, the following functions are to be noted: ‘marcadores de refuerzo argumentativo’ (argumentative reinforcement markers) and markers of reformulation or corroborative ones.

Which are usually addressed to markers such as ‘en efecto’, ‘efectivamente’, ‘de hecho’ and ‘realmente’, whereas ‘es decir’ is considered by Portolés and Zorraquino (1999) as exemplifier and ‘ya (que)’, ‘sobre todo’, ‘pero’ and ‘incluso’ will usually be included into the category of adversatives.

- 1) These functions are to be added to the ones commented when analyzing lexical domains, moreover the categories of ‘quantifier’, ‘conjunction’, ‘epistemic adverb’ and adjunct are to be added to grammar types.
- 2) Speaker’s/ writer’s attitude and desired effect: certainty, illustration, explanation- exposition and contraposition.

**Syntax-punctuation discrepancies and similarities between ‘indeed’ and its translation counterparts**

It is also interesting to see how, in the process of translation, the punctuation criteria would strongly differ between the English source and the Spanish counterpart. This particularity ought to be due to the Germanic constraints of English and the Latin ones of Spanish, for example:

1) <s> **Indeed**, it is quite in keeping with the positions this House has always adopted. </s>

<s> Está, **sin duda**, dentro de la línea de las posiciones que nuestro Parlamento siempre ha adoptado. </s>

2) <s> The fact that administrative expenditure on INTERREG III is - in my opinion - still excessive is something which we can and **indeed** always must criticize, although it almost seems to me as though this is unavoidable with the Commission ' s support programmes. </s>

<s> Podemos criticar siempre que en INTERREG sigue siendo excesivo - como creo yo también- el trajín administrativo, **pero** también hay que criticarlo siempre, aunque esto casi me parece inevitable en el caso de los programas de fomento de la Comisión. </s>

3) <s> We know, and we have stated as much in very many resolutions **indeed** , including specifically during (...).

<s> Sabemos y lo hemos afirmado **realmente** en muchas resoluciones (...).

4) <s> Equally, unemployment levels in the worst affected regions barely fell at all, **indeed** they rose in some cases. </s>

<s> Igualmente, en las regiones más fuertemente afectadas apenas se han podido reducir las cifras de desempleados; **incluso** éstas han aumentado, en parte. </s>

5) <s> **Indeed** , it is not just about modernisation of Community law, (...)

<s> **En efecto**, no sólo es importante una modernización del Derecho comunitario (...).

- (1) Reflects a case in which the English original is shown as connective followed by comma, whereas the Spanish target is an adverbial embedded cluster.
- (2) In this case the English original appears as sentence adverbial, meanwhile the Spanish target is defined as an adverbial connector preceded by comma.
- (3) The English source displays a conclusive adverbial followed by comma; however, the Spanish target prefers no punctuation marks.
- (4) The English source is juxtaposed preceded by comma and the Spanish target is juxtaposed by means of a semi-column.
- (5) Finally, in this last example, the English source and the Spanish target coincide in punctuation as sentence connectives followed by comma.

As random as the choice of punctuation may seem, the translator adapts the construction so as not to sound awkward in the target language, always keeping both illocutionary force and perlocutionary effect (Austin J.L, 1975), as it is aimed in a communicative contextual translation.

### 6.2.6.2 In fact:

The translation preferences and can be summarized as follows in the different translation corpora:

- 1) ‘De hecho’ is the most frequent translation equivalent (Hummel, 1995) in Europarl7, Multinot Essay and Speech (7.06, 215.77 and 27.27 per million respectively).
- 2) ‘En realidad’ is the most frequent translation equivalent (Hummel, 1995) in the Open Subtitles corpus, Multinot- Expe and Speech (0.34, 121.02 and 27.27 per million respectively).
- 3) ‘Realmente’ is the most frequent translation equivalent (Hummel, 1995) in EMEA (0.65 per million).
- 4) ‘Es más’ is the most frequent translation equivalent (Hummel, 1995) in Multinot Fiction (95.611 per million).

Adding up the proportions per million of ‘de hecho’ and ‘in fact’, the results are the following: ‘de hecho’ results in a proportion of 324.24 per million if its proportional figures per million found in in Eurparl7 (Koehn, 2001), Multinot Essay and Speech (Lavid et al., 2015) are added up (as these are the corpora where ‘de hecho’ is found with higher frequency).

Whereas ‘en realidad’ results in a proportion of 154.69 per million if its proportional figures per million found in Open subtitles (Tiedemann, 2007) Multinot expe and speech (Lavid et al., 2015) are added up (as these are the corpora where ‘en realidad’ is found with higher frequency).

*What is the preferred translation equivalent for ‘in fact’ in Spanish after the analysis of parallel corpora? ‘de hecho’, proportionally speaking, is the preferred translation equivalent (Hummel, 1995) for ‘in fact’ as found in the analysis of its equivalents in parallel corpora, it is followed by ‘en realidad’, ‘es más’ and ‘realmente’ (form the most frequent proportionally speaking to the least).*

In question three, it will be specified which genre, according to the proportion found in each corpus, prefers mostly which of the four previous translation counterparts (see section ‘research question 3’).

### ‘In fact’: calculating the correspondence degree:

The correspondence degree illustrates which is the rate of equivalence between one source-marker, in this case ‘in fact’, and a target one.

Europarl 7 English	Open subtitles	Emea	Multinot fiction	Multinot essay	Multinot expe	Multinot speech	Multinot web
De hecho	en realidad	realmente	es más	de hecho	en realidad	en realidad de hecho	0

**Table 21: ‘in fact’ correspondence degree.**

### Correspondence potential ‘in fact’

The correspondence potential is the likeability each translation equivalent has to occur as correspondence of the source marker, in this case ‘in fact’.

Correspondence potential = number of correspondences + correspondence degree

**De hecho** (highest correspondence potential).  
**En realidad** (highest correspondence potential).  
 Es más.  
**Realmente** (smallest correspondence potential).

**Figure 28: ‘in fact’ correspondence potential.**

Different functionality found according to monolingual translation, sentence occurrence, monolingual analysis, translation correspondences (Aijmer, 2004; Cuenca, 2001; Portolés and Zorraquino 1999; Halliday, 2014; Garcés Gómez, 2005; del Saz, 2003):

Emphasis, explanation, illustration, particularization, rectification, clarification, exemplification and commentator (attitudinal). Concerning its syntactic function and part of speech, it seems to work as connector, adverb or prepositional phrase or even speech interjection. Adding unknown information or reflecting a true action. With respect to the syntactic and semantic fields apart from adverbial, prepositional, conjunction and connective features; phrase, expression and idiom should be added.

Clarification-verification, contrast-adversative, resultive-concessive and corroborative must be added due to translation spottings. To these, the previously mentioned functions addressed by Aijmer (2004) plus the ones found at the beginning of this section are added:

Enhancing, explanation, illustration, exemplification commentator would correspond to the additive function, commentator, particularization, rectification as well as the pragmatic expectation-evidentiality (Chafe, 1986) function.

'In fact' when translated into 'en realidad', 'es más', 'de hecho' or 'realmente' can work as any of its common semantic counterparts (adverbial of contrast, sentence adverbial, pragmatic marker of evidentiality) and with most of its functions especially when working as connective, as it is the main focus of this paper: addition, contrast, enhancement, rectification, explanation, illustration and particularization and Halliday's clarification- verificative function.

It should be added that 'in fact' seems to be preferred in written argumentative or expository texts with a strong preference towards 'de hecho', which, as 'en realidad', presents similar form, in the sense that the three of them correspond to prepositional phrases – leaving aside the fact that in Spanish, as romantic language, the noun which modifies the preposition displays gender and number, unlike English as Germanic language -.

The case of 'es más' differs in the form with 'in fact' as it corresponds to a verbal periphrasis comprised by a nominal verb, the verb to be in the third person singular, and a quantifier, this last one in comparative form. 'Realmente' also differs in the form with 'in fact' as it is a lexical adverb

whereas ‘in fact’ is a prepositional phrase.

It has previously been mentioned that ‘in fact’ also displays a function as ‘epistemic adverb of certainty’, this can be seen in the translation spottings found in the latter corpuses as for the following cases: efectivamente, justamente, indudablemente, ciertamente, sinceramente, naturalmente, fundamentalmente, supuestamente, evindentemente, obviamente, aparentemente, precisamente, exactamente. Verdaderamente.

This function of ‘certainty’ is also conveyed by other translation counterparts such as de verdad, en verdad, en efecto, en rigor, sin duda – prepositional phrases, la verdad - nominal phrases, tal vez -adverbial phrases, es cierto que – nominal, verbal phrases.

Particularizing functions may also be conveyed by the following translation spottings: por ejemplo, por cierto, en concreto, en especial, en este/ese sentido – prepositional phrases, exactamente, prácticamente, concretamente, solamente, especialmente, suficientemente- adverbs, más que, tan solo, tal, tal como, apenas – quantifiers.

Additive functions could be illustrated as well by the following translation spottings: también, además, incluso – adverbs. Whereas rectification functions might be seen in the following spottings: en absoluto, en el fondo, en la práctica – prepositional phrases, mejor dicho, Bueno, más bien – adjuncts, lo cierto es que – propositions.

Enhancement and verification could also be conveyed by some additional spottings: ya, así, lamentablemente, aún, todavía- adverbs, por supuesto, ni siquiera, sobre todo- prepositional phrases. Finally, adversative functions are to be seen in the following translation spotting cases: pues, es decir, sea, acaso, pero, por el contrario, al contrario, sin embargo, en cambio, no obstante, aunque, a pesar de que, which are a combination of conjunctions, prepositional phrases and adverbs.

Noting the functions addressed by some authors such as Portolés and Zorraquino (1999), Cuenca (2001), del Saz (2003) or Garcés Gómez (2005), to mention some, the following functions are to be noted: argumentative reinforcement (marcador de refuerzo argumentativo), reformulation. These two functions, reinforcement and reformulation, are also conveyed by some of the translation spottings found for the case of 'in fact'.

### 6.2.6.3 Syntax-punctuation discrepancies and similarities between ‘indeed’ and its translation counterparts:

It is also interesting to see how, in the process of translation, the punctuation criteria would strongly differ between the English source and the Spanish counterpart. This particularity ought to be due to the Germanic constraints of English and the Latin ones of Spanish, for example:

1) <s> She gets very angry **in fact** , she’ s furious. </s> Opus2 – Open Subtitles

<s> Ella se enfada. Se pone furiosa. </s>

2) <s> Be king **in fact** as well as name. </s> Opus2- Open Subtitles

<s> ¡ Sed rey **de hecho**, tanto como de palabra! </s>

3) <s> There is, **in fact**, a risk of a military coup in the future. </s> - Europarl7

<s> **De hecho**, existe el riesgo de un golpe militar en el futuro. </s>

4) <s> I was **in fact** shocked to read in the report </s> - Europarl7

<s> **En realidad**, me ha atemorizado la noticia, </s>

5) <s> **In fact** , renewable energies unquestionably provide a way of achieving these objectives.

</s> - Europarl7

<s> **Ahora bien**, las energías constituyen un medio indiscutible para lograr esos objetivos.

</s>

(1) Reflects a case in which the source uses subordination whereas the target uses full stop juxtaposition or simply independent sentences. This makes the Spanish translation to force a stronger stop and consequently omits *in fact* in turn.

(2) Illustrates an interesting case where ‘in fact’ is used as an adjective defining the type of ‘king’, the English source uses *in fact* as a lexical item whereas the Spanish target forces a stop through juxtaposition.

- (3) In this case, the English source prefers embedding whereas the Spanish target goes for fronting, both cases create emphasis but at different stages of the sentence.
- (4) The English source uses ‘in fact’ a lexical item whereas the Spanish target prefers to use a fronted adverbial phrase, in this case the Spanish target creates stronger emphasis than the English source.
- (5) This example illustrates punctuation similarity, in both the source and the target fronting is preferred.

As random as the choice of punctuation may seem, the translator adapts the construction so as not to sound awkward in the target language, always keeping both illocutionary force and perlocutionary effect (Austin, 1975).

#### 6.2.6.4 Actually:

Translation correspondences between English and Spanish as revealed by the analysis of parallel corpora: proportionally speaking the translation equivalent (Hummel, 1995) with the highest frequency is ‘realmente’ followed by ‘en realidad’, ‘de verdad’, ‘de hecho’ and ‘así’.

- a) ‘De hecho’ is the most frequent translation equivalent (Hummel, 1995) in Europarl7 (Koehn, 2001): 9.98 per million.
- b) ‘En realidad’ is the most frequent translation equivalent (Hummel, 1995) in Multinot essay and speech (Lavid et al., 2015): 107.9 per million.
- c) ‘Realmente’ is the most frequent translation equivalent (Hummel, 1995) for ‘actually’ in EMEA (Tiedemann, 2007), Multinot Fiction and expe (Lavid et al., 2015): 0.65 per million, 95.611 per million, 418.27 per million, respectively.
- d) ‘De verdad’ is the most frequent translation equivalent (Hummel, 1995) for ‘actually’ in Multinot Fiction (Lavid et al., 2015): 95.611 per million.
- e) ‘Así’ is the most frequent translation equivalent (Hummel, 1995) for ‘actually’ in Open Subtitles (Tiedemann, 2007): 1 per million.

The previous results commented in 1) to 5) reveal that the most frequent translation equivalents for ‘actually’ are ‘realmente’, ‘en realidad’, ‘de verdad’, ‘de hecho’ and ‘así’; ordered from the most preferred translation equivalent for ‘actually’, ‘realmente’, to the least.

‘En realidad’ is the most frequent translation equivalent (Hummel, 1995) for ‘actually’ in the corpus Multinot essay (Lavid et al., 2015), ‘de verdad’ is the most frequent translation equivalent (Hummel, 1995) for ‘actually’ in the corpus Multinot Fiction (Lavid et al., 2015), ‘de hecho’ is the most frequent translation equivalent (Hummel, 1995) for ‘actually’ in the corpus Europarl7 (Koehn, 2001).

‘Así’ is the most frequent translation equivalent for ‘actually’ in the corpus Opensubtitles (Tiedemann, 2007). Finally, as mentioned before, the preferred translation equivalent (Hummel, 1995) for ‘actually’ is ‘realmente’ which appears with high frequency in the corpora Multinot Fiction and Expe (Lavid et al., 2015).

The equivalences and additional functions found in parallel corpora translation with ‘de hecho’, ‘en realidad’ and ‘realmente’ have already been commented for the case of ‘in fact’. Given the synonymic relationship between ‘actually’ and ‘in fact’, it is assumed that these additional functions and relations are applicable to ‘actually’. It is thus interesting to see what happens in the cases of the translation equivalents (Hummel, 1995) of ‘actually’, ‘de verdad’ and ‘así’.

*What is the preferred translation equivalent for ‘actually’ in Spanish after the analysis of parallel corpora? ‘realmente’ is, proportionally speaking, the preferred translation equivalent (Hummel, 1995) for ‘actually’ as found in the analysis of its equivalents in parallel corpora.*

**‘Actually’: calculating the correspondence degree:**

The correspondence degree illustrates which is the rate of equivalence between one source-marker, in this case ‘actually’, and a target one

Correspondence potential = number of correspondences + correspondence degree
--

Europarl 7 English	Open subtitles	Emea	Multinot fiction	Multinot essay	Multinot expe	Multinot speech	Multinot web
De hecho	así	realmente	de verdad realmente	en realidad	realmente	0	0

**Table 22: ‘actually’ correspondence degree.**

**‘Actually’: calculating the correspondence potential:**

The correspondence potential is the likeability each translation equivalent has to occur as correspondence of the source marker, in this case ‘actually’

<b>Realmente:</b> (highest correspondence potential).
De hecho.
En realidad.
De verdad.
<b>Así.</b> (smallest correspondence potential).

**Figure 29: ‘actually’ correspondence potential.**

Different functionality found according to monolingual translation, sentence occurrence, monolingual analysis, translation correspondences (Aijmer, 2004; Cuenca, 2001; Portolés and Zorraquino 1999; Halliday, 2014; Garcés Gómez, 2005; del Saz, 2003):

Emphasis, explanation, illustration, particularization, rectification, clarification, exemplification and commentator (attitudinal). Concerning its syntactic function and part of speech, it seems to work as a connector, adverb or even speech interjection.

Strong attitudinal functions – surprise, verification of truth, politeness, introducing a new topic, contrast, emphasis and addition. Concerning frequent collocations, the following might be illustrated: and actually (addition), but actually (contrast), well actually (rectification/correction), if actually (condition/ truth).

Aijmer (2004): the adversative (‘but’) type, the additive (‘and’) type, and the enhancing (‘for’): emphasis, explanation, illustration, particularization, rectification, clarification, exemplification and commentator (attitudinal), adding unknown information or reflecting a true action, strong attitudinal functions – surprise, verification of truth, politeness, introducing a new topic.

Enhancing would be what I addressed as ‘emphasis’, explanation-illustration- clarification-exemplification-commentator would correspond to the additive function, commentator-clarification-

particularization-rectification would enter the adversative type. As well as the functionality of ‘actually’ as a pragmatic marker of expectation- evidentiality (Chafe, 1986).

Emphasis, explanation, illustration, particularization, rectification, clarification, exemplification and commentator (attitudinal), adding unknown information or reflecting a true action, strong attitudinal functions – surprise, verification of truth, politeness, introducing a new topic, addition, contrast and emphasis as well as the pragmatic expectation-evidentiality (Chafe, 1986) function.

Moreover, Halliday (2014) also makes mention to ‘actually’ in the following terms: proving modal assessment in domains of realization and domains of manifestation as clause-modal adjunct-adverb or prepositional phrase showing counter-expectancy.

‘Actually’ when translated into ‘en realidad’, ‘así’, ‘de hecho’. ‘de verdad’ or ‘realmente’ can work as any of its common semantic counterparts (adverbial of contrast, sentence adverbial, pragmatic marker of evidentiality) and with most of its functions especially when working as connective, as it is the main focus of this paper: addition, contrast, enhancement, rectification, explanation, illustration, factual truth, surprise, epistemic operator and particularization, among others, and Halliday’s clarification- verificative function.

There is the function as ‘epistemic adverb of certainty’, this can be seen in the translation spottings found in the latter corpora as for the following cases: efectivamente, justamente, indudablemente, ciertamente, sinceramente, naturalmente, fundamentalmente, supuestamente, evindentemente, obviamente, aparentemente, precisamente, exactamente, verdaderamente.

This function of ‘certainty’ is also conveyed by other translation counterparts such as: de verdad, en verdad, en efecto, en rigor, sin duda – prepositional phrases, la verdad - nominal phrases, tal vez -adverbial phrases, es cierto que – nominal verbal phrases.

Particularizing functions may also be conveyed by the following translation spottings: por ejemplo, por cierto, en concreto, en especial, en este/ese sentido – prepositional phrases, exactamente, prácticamente, concretamente, solamente, especialmente, suficientemente- adverbs, más que, tan solo, tal,

tal como, apenas – quantifiers.

Additive functions could be illustrated as well by the following translation spottings: también, además, incluso – adverbs. Whereas rectification functions might be seen in the following spottings: en absoluto, en el fondo, en la práctica – prepositional phrases, mejor dicho, Bueno, más bien – adjuncts, lo cierto es que – propositions.

Enhancement and verification could also be conveyed by some additional spottings: ya, así, lamentablemente, aún, todavía- adverbs, por supuesto, ni siquiera, sobre todo- prepositional phrases.

Finally, adversative functions are to be seen in the following translation spotting cases: pues, es decir, sea, acaso, pero, por el contrario, al contrario, sin embargo, en cambio, no obstante, aunque, a pesar de que, which are a combination of conjunctions, prepositional phrases and adverbs.

Noting the functions addressed by some authors such as Portolés and Zorraquino (1999), Cuenca (2001), del Saz (2003) or Garcés Gómez (2005), just to mention some, the following functions are to be noted: ‘marcadores de refuerzo argumentativo’ (argumentative reinforcement markers) and markers of reformulation or corroborative ones, which are usually addressed to markers such as ‘en realidad’, ‘de verdad’, ‘de hecho’ and ‘realmente’.

**Syntax-punctuation discrepancies and similarities between ‘indeed’ and its translation counterparts:**

It is also interesting to see how, in the process of translation, the punctuation criteria would strongly differ between the English source and the Spanish counterpart. This particularity ought to be due to the Germanic constraints of English and the Latin ones of Spanish, for example:

1) <s> But his portfolio of information technology and tourism **actually** go very well together.

</s> Europarl7

<s> Pero sus carteras de informática y turismo, **de hecho** , casan muy bien juntas.

2) <s> What I learned is that, **actually**, there is sufficient legislation in place but the problem is the lack of supervision. </s>- Europarl7

<s> Lo que deduje de todo aquello es que **en realidad** hay suficiente regulación, pero el problema estriba en que no se controla. </s>

3) <s> It’ s kind of hard to explain, **actually** . </s>

<s> Es difícil de explicar, **la verdad**. </s> Opus2- Open subtitles

4) <s> Actually I don’ t like junior at all. </s> Opus2-Open subtitles

<s> No me gusta Júnior para nada, **de hecho**. </s>

5) <s> **Actually** went shopping without a bodyguard. </s> Opus2-Open subtitles

<s> **En realidad**, fui de compras sin guardaespaldas. </s>

(1) Reflects a case in which the English original appears like a sentence adverbial, whereas the Spanish translation prefers to be embedded in between commas. It seems that English intonation is conveyed in Spanish through ‘comma’ stop.

(2) In this case, it is the English original the one which appears embedded between commas, whereas the Spanish translation is used as sentence adverbial and the ‘comma’ is delayed in the sentence, these underlies on the intonation differences between English and Spanish.

(3) This example shows a speaker’s commentary, and in both the source and the target, it is reflected through ending comma juxtaposed item.

(4) English source uses a fronted adverbial, whereas Spanish target prefers final juxtaposition; both strategies are ways of showing emphasis.

(5) In both the source and the target, the adverbial is fronted, though in the Spanish target this fronted adverb is followed by a comma to cause a stop.

As random as the choice of punctuation may seem, the translator adapts the construction so as not to sound awkward in the target language, always keeping both illocutionary force and perlocutionary effect (Austin, 1975).

### 6.2.7 Spanish ECs:

In this section the connectives in Spanish will be the focus, the study will be based on their characteristics, their functions, their translations into English and the back translation into Spanish to find semantic fields as it has been done with the English connectives. The connectives chosen for this section are the following: ‘realmente’, ‘en realidad’, ‘la verdad’, ‘de hecho’.

#### **Possible lexical domains:**

The following graph suggest the possible semantic or lexical fields for ‘en realidad’, ‘realmente’, ‘de hecho’, ‘la verdad’; which correspond to: es decir, sinceramente, ciertamente/ sin duda, precisamente (et al), ya/tal, más que/ es más, pero, sin embargo, incluso, lo cierto, con certeza, en efecto, efectivamente, verdaderamente/ a decir verdad.

What means that ‘en realidad’, ‘realmente’, ‘la verdad’, ‘de hecho’; have connection with other argumentative operators/connectives such as ‘verdaderamente’, with reformulators such as ‘es decir’, adversative or disjunctive items such as ‘pero’/ ‘sin embargo’, epistemic adverbs or adverbials such as ‘ciertamente’/ ‘con certeza’, enhancers such as ‘es más’/ ‘incluso’, and some temporal commentators such as ‘ya’ and ‘tal’.

Moreover, it seems to be coincidence with the English connectives studied as well in the present paper, namely: ‘indeed’, ‘in fact’ and ‘actually’. So it can be concluded that there is, indeed, equivalence between what is called in English ‘Verificative markers’ and what is called in Spanish ‘Marcadores/ operadores argumentativos’.

And that there would be, indeed, a connection between elaboration in English and argumentation in Spanish, to which, the additional functions found in the English and Spanish section are to be added. Figure 30 in the following page shows the LDMs of Spanish markers.

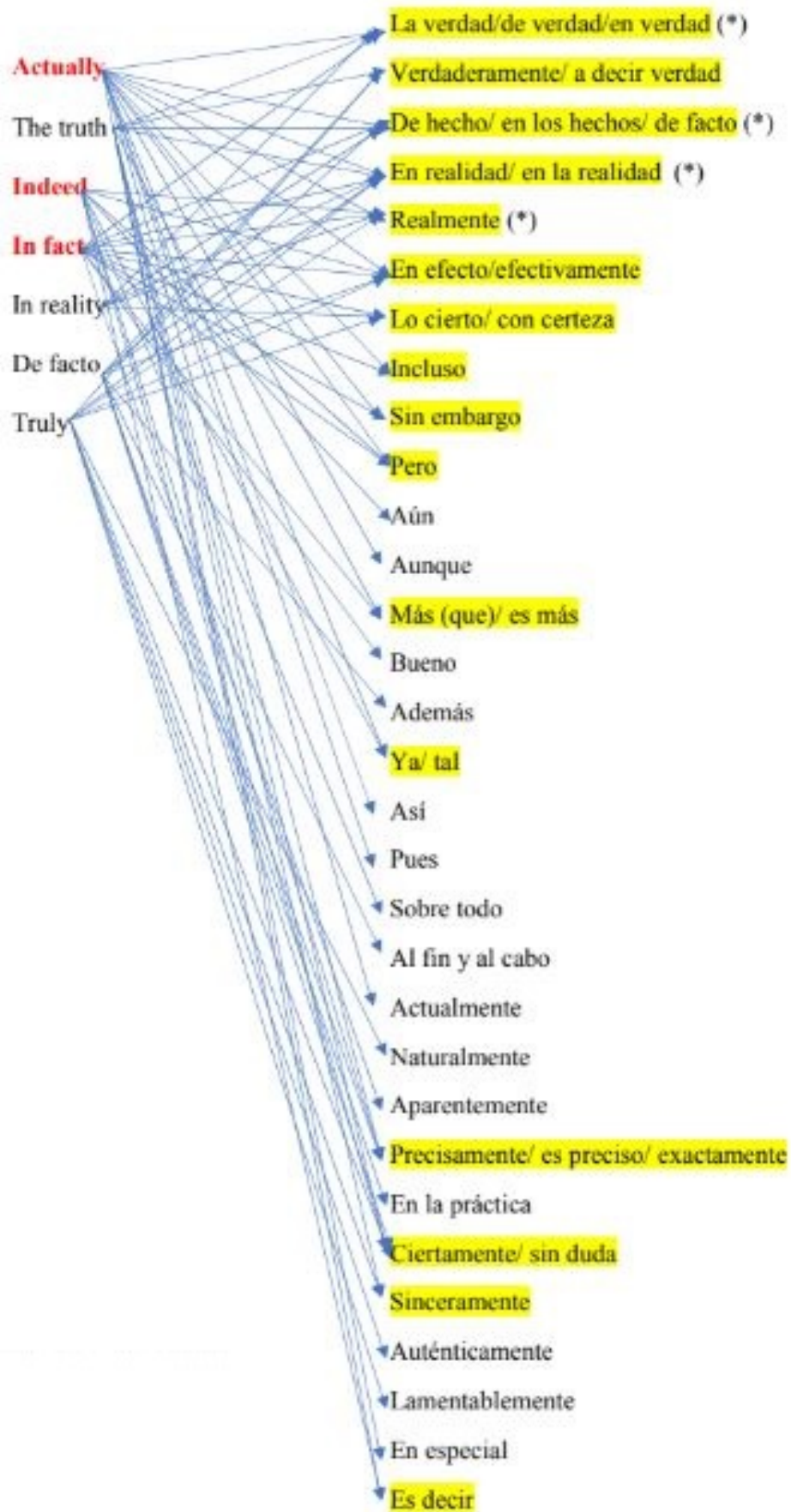


Figure 30: Spanish LDMs

As seen in the subsection devoted to Spanish markers or connectives, the words which would share semantic field with ‘en realidad’, ‘de hecho’, ‘la verdad’, realmente are those which show highest frequency in the tables found in that same section (see Spanish connectives). To the generic function of argumentative operators, these additional linguistic items found in back translation would add the functions of reformulation, explanation, evaluation, emphasis, corroboration, contrast and truth statement.

With respect to the least frequent backtranslation spottings found, these are the functions and grammatical characterisations:

- a) Además: conjunction-additive, additive connector (Portolés and Zorraquino, 1999).
- b) Así: adverb- concessive, consecutive connector (Portolés and Zorraquino, 1999).
- c) Pues: conjunction-adversative, consecutive connector (Portolés and Zorraquino, 1999).
- d) Sobre todo: prepositional phrase, conjunction, emphasis, argumentative reinforcement (Portolés and Zorraquino, 1999).
- e) Lamentablemente: adverb, commentator, negative.
- f) Auténticamente: adverb, evaluator, commentator.
- g) En especial: prepositional phrase, emphasis, particularization.
- h) En la práctica: prepositional phrase, clarification, resumptive reformulator (Portolés and Zorraquino, 1999).
- i) Aparentemente: adverb, hypothesis, commentator.
- j) Actualmente: adverb, time, time marker.
- k) Al fin y al cabo: adverbial phrase, resulting, reformulation, resumptive (Portolés and Zorraquino).

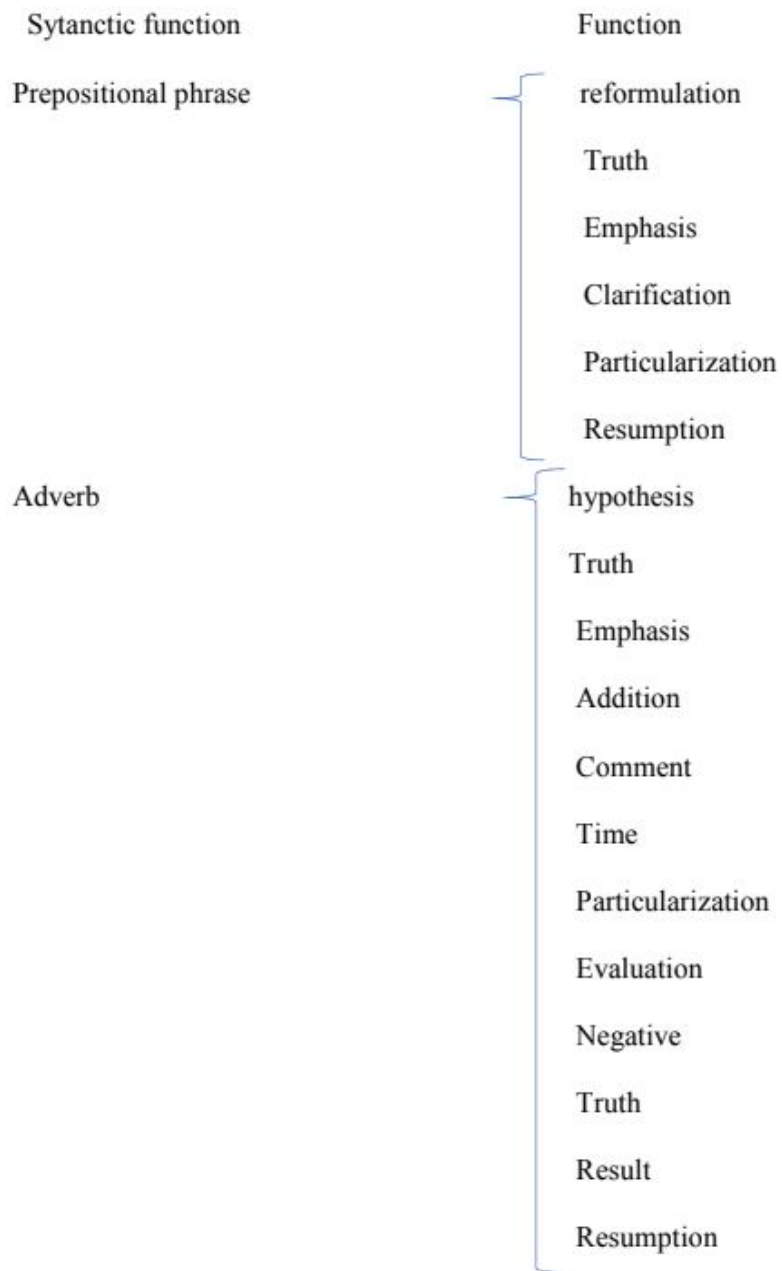
l) Bueno: adjective-conjunction, evaluator, deontic marker (Portolés and Zorraquino, 1999).

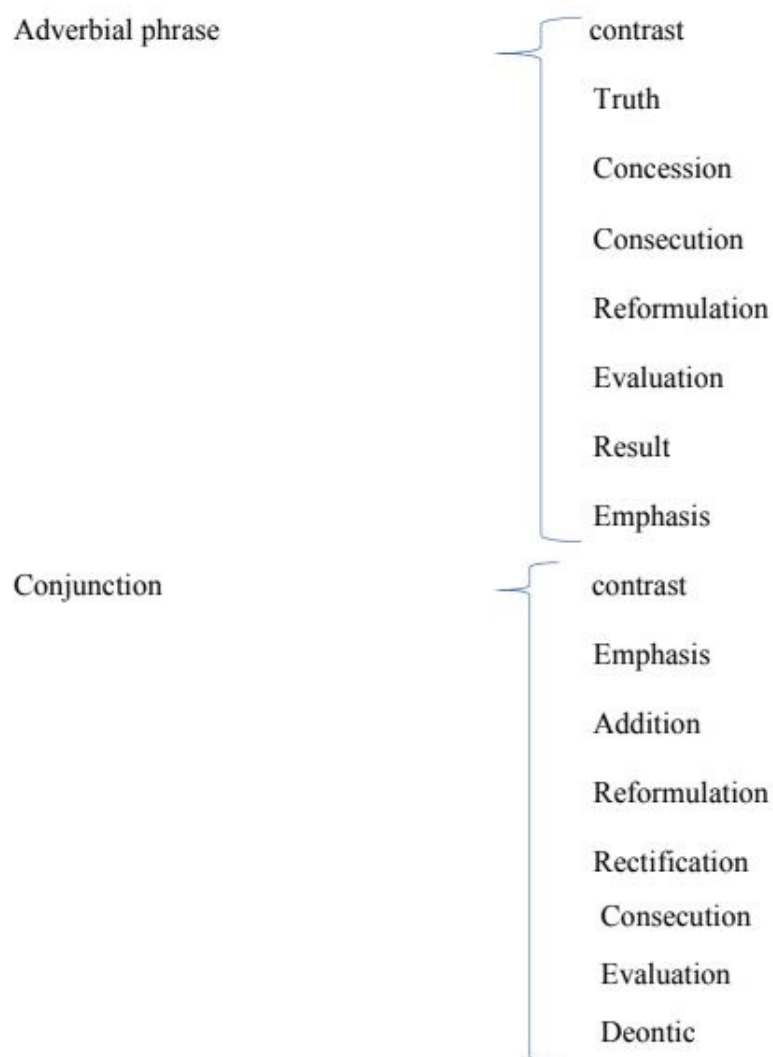
m) Aunque: conjunction, adversative, emphasis reformulator (Cortés y Camacho 2005).

n) Aún: Adverb, time, reformulation, rectification (Portolés and Zorraquino, 1999).

To the previous functions, these additional markers provide the following: reformulation, contrast, rectification, emphasis, evaluation, hypothesis, comment, time, clarification, resumption, concession, consecution and addition.

Lexical domains based on both most frequent and least frequent markers:



Lexical domains based on both most frequent and least frequent markers (cont.):**Figure 31: Spanish LDMs in detail.**

The speakers/ writer's attitude regarding discourse or speech act (Austin, 1975) can be of: hypothesis, conclusion, emphasis, corroboration, addition, rectification, evaluation (negative or positive), contrast, truthfulness, illustration or comment.

Translation preference according to the following criteria:

- Genra (argumentative, explanation, exposition, analysis, rhetoric, dialogue).
- Style (formal, informal).
- Format (spoken, written).

The following conclusions are to be drawn:

- 1) The preferred counterpart for ‘realmente’ is truly.
- 2) The preferred counterpart for ‘la verdad’ is ‘the truth’.
- 3) The preferred counterpart for ‘en realidad’ is ‘in fact’, followed by: ‘in reality’, ‘actually’ and ‘indeed’.
- 4) The preferred counterpart for ‘de hecho’ is ‘indeed’, followed by: ‘in fact’ and ‘de facto’.

With respect to the preferred genre:

- a) Europarl7 Spanish: it is the most various, but not the most preferred in any of the four cases.
- b) Multinot Fiction: it is the most preferred in the cases of ‘en realidad’ and ‘la verdad’, which means that these items are most likely to be found in rhetorical formal written texts.
- c) Multinot essay: it is the most preferred in the cases of ‘de hecho’ and ‘realmente’, which means that these items are more likely to be found in formal expository and/or argumentative texts.
- d) Multinot expe and speech: they appear with some significance in all the cases, specially in all the cases with greater frequency than Europarl7 Spanish. This means that these connectives could also be found with certain frequency in formal expository written or spoken texts.

Correspondence degree and potential of ‘realmente’, ‘la verdad’, ‘en realidad’ and ‘de hecho’

The correspondence degree illustrates which is the rate of equivalence between one source marker and a target one and the correspondence potential is the likeability each translation equivalent has to occur as correspondence of the source marker, in this case ‘realmente’ and ‘la verdad’.

$$\text{Correspondence potential} = \text{number of correspondences} + \text{correspondence degree}$$

**Realmente**

Europarl 7 English	Open subtitles	Emea	Multinot fiction	Multinot essay	Multinot web	Multinot expe	Multinot speech
Truly	0	0	0	truly	0	0	truly

**Table 23: ‘realmente’: correspondence potential.**

Truly is the most frequent translation counterpart in the three corpora in which it is found, the correspondence potentials would be, accordingly: 31.12; 101.82; 53.76. So, the highest correspondence potential corresponds to the sub-corpus Multinot essay (Lavid et al., 2015).

**La verdad**

Europarl 7 English	Open subtitles	Emea	Multinot fiction	Multinot essay	Multinot web	Multinot expe	Multinot speech
The truth	0	0	The truth	The truth	0	The truth	the truth

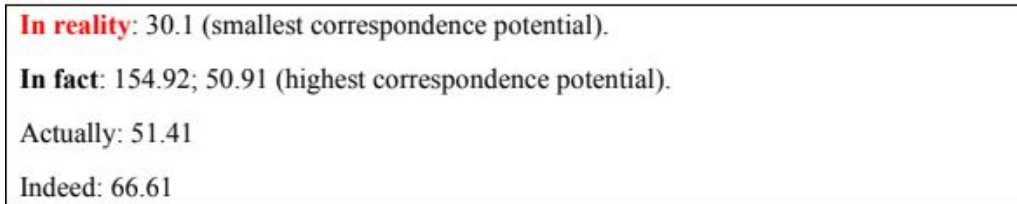
**Table 24: ‘la verdad’: correspondence potential.**

The truth is the most frequent translation counterpart in the three corpora in which it is found, the correspondence potentials would be, accordingly: 28.76; 52.9; 150.73; 51.78; 56.26. The highest correspondence potential corresponds to the sub-corpus Multinot Essay (Lavid et al., 2015).

**‘En realidad’ correspondence degree:**

Europarl 7 English	Open subtitles	Emea	Multinot fiction	Multinot essay	Multinot web	Multinot expe	Multinot speech
In reality	0	0	in fact	in fact	0	actually	indeed

**Table 25: ‘en realidad’ correspondence degree.**

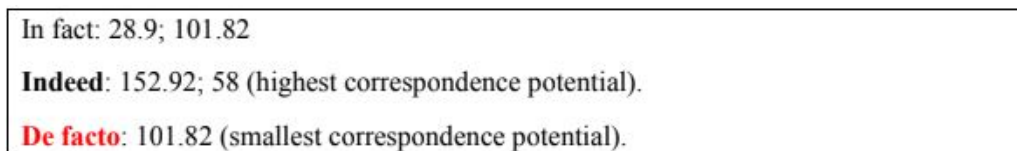


**Figure 32: ‘en realidad’ correspondence potential.**

**‘De hecho’ correspondence degree:**

Europarl 7 English	Open subtitles	Emea	Multinot fiction	Multinot essay	Multinot web	Multinot expe	Multinot speech
in fact	0	0	indeed	in fact De facto	0	0	indeed

**Table 26: ‘de hecho’ correspondence degree.**



**Figure 33: ‘de hecho’ correspondence potential.**

**Additional functions to be commented:**

The grammatical functions of ‘realmente’, according to the previous examples, might be sentence adverbial, epistemic adverbial, connector, conjunction. With respect to the semantic and pragmatic functions it seems that, for the case of ‘realmente’, and based on the latter examples the following are found correction, contraposition, verification, rhetoric’s, addition, mood, confirmation, truthfulness, explanation, condition, particularization, emphasis and surprise.

Syntactically ‘en realidad’ can work as sentence adverbial, epistemic adverbial, conjunction and connector. Concerning its semantic and pragmatic functions, it shows evidentiality, confirmation of truth, illustrates surprise or disbelief, assurance of truth, illustration-addition and explanation, contraposition and correction and emphasis.

It appears as either embedded, fronted or juxtaposed, followed by comma or between commas, but it may be assumed that it could ideally be found with semi-columns, columns, parenthesis or hyphens as well as full stops. It can collocate with conjunctions such as ‘y’, ‘aunque’, ‘pero’ and there seem to be examples of conditional phrases such as ‘y si en realidad...’.

Syntactically speaking, ‘la verdad’ could be conjunction, sentence adverbial, adjunct, subject, and connector. It can be followed by comma, between commas or preceded by full stop, ideally collocated with semi-column, column, hyphen or between parenthesis, it can be fronted, embedded or backed to the end of the sentence it can collocate with adversative conjunctions ‘pero la verdad’ or additive ones ‘y la verdad es que’, ideally with conditional ones such as ‘y si la verdad es que’.

Functionally and pragmatically speaking, it shows contrast, reformulation, addition, condition, verification, explanation, illustration, surprise, truthfulness, contrast and matification. ‘De hecho’ has very similar functions, forms and characteristics to the previously mentioned lexical items – realmente, la verdad, en realidad – with the exception of the possibility of being a subject as the case of ‘la verdad’ but the possibility of being an adjective as ‘en realidad’.

Some additional functions to those highlighted in the previous sections are sarcasm or irony, truthfulness, evidentiality, factual, of high social status, evidentiality, emphasis, factuality, hypothesis, contrast, truthfulness, modality, commentators, correction, meta-discursive, reformulation and rectification. (As found y monolingual dictionaries, see analysis section).

Moreover, it seems to be coincidence with the English connectives studied as well in the present paper, namely: ‘indeed’, ‘in fact’ and ‘actually’. So it can be concluded that there is, indeed, equivalence between what is called in English ‘Verificative markers’ and what is called in Spanish ‘Marcadores/ operadores argumentativos’, and that there would be, indeed, a connection between elaboration in English and argumentation in Spanish, to which, the additional functions found in the English and Spanish section are to be added and pointed out in the conclusion. (As mention at the beginning of this section).

**Syntax-punctuation discrepancies and similarities between ‘indeed’ and its translation counterparts:**

Discrepancies in sentence position or punctuation between the source of the language, this is due to the linguistic constraints which rule Germanic (English) and Latin (Spanish) languages. Some examples would be:

(1) <s> Señalo bien, “ sin razón ” ya que, de hecho, va a reducir esos derechos. </s>

<s> This Charter is really misnamed because it will actually reduce these rights. </s> (Europarl7 Spanish)

(2) <s> Voy a dedicarle poco tiempo, porque es, **la verdad**, un buen informe. </s>

<s> I will not take long, because it is **in fact** a very good report. </s> (Europarl7 Spanish).

(Source embedded between commas, target adverb in a nominal predicate).



# Chapter 7

## Discussion and Concluding remarks

### 7.1 Introduction

In this chapter some of the most important findings are shown, namely:

- a) Lexical domains for English markers.
- b) Lexical domains for Spanish markers.
- c) New functions found for each English and Spanish marker.
- d) Translation correspondence preferences.
- e) Genre preference.

With all the previous results, translation can be confirmed as a necessary tool in discourse and speech analysis which provides essential information regarding verificative markers, with the potential of being useful for the study of other discourse and speech elements in parallel corpora analysis.

This thesis has investigated the core meanings of a subset of VMs in English and Spanish, as well as other possible peripheral ones based on their translation equivalents (Hummel, 1995). VMs have shown to be a highly polysemous subset of DMs with numerous translations representing different senses and implicatures, very much in line with other DMs belonging to the Elaborating subtype (see Lavid and Avilés 2019).

The translation correspondences of VMs revealed by the analysis of the parallel corpora used in this study have made explicit certain meanings which are not so obvious from a monolingual analysis. For example, the corpus analysis has revealed that in addition to the ‘in fact’ is a highly polysemic VM with various meanings, revealed through the translation analysis: a core epistemic meaning (translated by “de hecho”, and ‘en efecto’, but also some derived pragmatic uses, namely: a) expressing a contrast (translated by ‘en realidad’); b) expressing a stronger reformulation (translated by ‘es más’; c) expressing a reason (translated by ‘puesto que’).

These meanings, however, had already been spotted in previous monolingual and crosslingual analysis (see Aijmer and Vandenberg, 2004), but the translation analysis has revealed other meanings which has not been detected before. These include purpose, concession and correction.

Similarly, the analysis of ‘de hecho’ one of the backtranslations of ‘in fact’ reveals similar additional functions to the ones explained in the previous paragraph. The analysis has also shown that there are genre-specific preferences in the use of VCs in original and translated texts. Indeed, some uses of both the English and the Spanish VMs differ depending on the genre where they occur.

Based on the characteristics which define each of the corpora the classification of each genre has been based on the characteristics of the corpora used in the analysis of translation equivalence as follows:

1. Europarl7 (Koehn, 2001): it represents a formal spoken argumentative and expository genre as it offers transcriptions of legal and parliamentary discourse.
2. EMEA (Tiedemann, 2007): it represents analytical genre as it reproduces medical texts.
3. Open Subtitles (Tiedemann, 2007): it represents informal spoken language as it corresponds to the transcriptions of the subtitles based on film dialogues.

4. Multinot Expe and Essay (Lavid et al., 2015): it represents formal written argumentative and expository language, the former as expository texts and the latter as scientific and educational essays.
5. Multinot Speech and Web (Lavid et al., 2015): it represents a formal spoken argumentative and expository genre as it offers transcriptions of legal and parliamentary discourse.
6. Multinot Fiction (Lavid et al., 2015): it represents formal written rhetorical texts as it is based on extracts from English language literature.

## 7.2 Review of findings

1. Written formal expository argumentative texts prefers ‘de hecho’ as translation equivalent (Hummel, 1995) for ‘in fact’, and ‘realmente’ and ‘en realidad’ for both ‘actually’ and ‘indeed’.
2. Spoken formal expository argumentative texts prefer ‘de hecho’ as translation equivalent for ‘in fact’, ‘actually’ and ‘indeed’. But in the case of ‘indeed’ additional equivalents (Hummel, 1995) will be valid: ‘incluso’, ‘efectivamente’, ‘en efecto’ and ‘realmente’.
3. Written formal analytical texts prefer ‘realmente’ as translation equivalent for ‘in fact’, ‘de hecho’ as translation equivalent (Hummel, 1995) of ‘actually’ and a significant number of translation equivalents for ‘indeed’: ‘en efecto’, ‘es decir’, ‘sobre todo’, ‘pero’.
4. Written rhetorical texts prefer ‘es más’ as translation equivalent (Hummel, 1995) for ‘in fact’; ‘de verdad’ and ‘realmente’ as translation equivalents (Hummel, 1995) for ‘actually’; and does not offer any occurrence for ‘indeed’.
5. Dialogues prefer ‘en realidad’ as translation equivalent (Hummel, 1995) for ‘in fact’, ‘así’ as translation equivalent for ‘actually’ and ‘ya’ as translation equivalent for ‘indeed’.

As mentioned in the answers provided to the first research question ‘which are the translation preferences of the VM’, ‘in fact’ prefers ‘de hecho’, ‘actually’ prefers ‘realmente’ and ‘indeed’ prefers ‘de hecho’.

*What are the genre specific preferences of each VM?*

- a) ‘In fact’ and ‘indeed’ coincide in translation preference so they seem to be more central than ‘actually’, which would be more peripheral.
- b) ‘De hecho’ is the translation equivalent mostly preferred in each genre.
- c) ‘In fact’ is more frequent in both formal written and spoken expository and argumentative texts, whereas ‘actually’ would prefer formal spoken expository argumentative texts only; and ‘indeed’ would be more likely in formal written analytical texts and in formal spoken expository argumentative texts.
- d) The three verificative markers coincide in their preference towards formal spoken expository and argumentative texts.
- e) ‘Indeed’ is not frequent in formal written rhetorical texts.
- f) ‘Indeed’ presents the highest rate of translation equivalents (Hummel, 1995).
- g) The three VM are preferred in formal contexts.

### 7.3 Pointers for the future

This thesis is a preliminary step in the creation of an onomasological dictionary where lexical domains and translation preferences will be included. It is expected that this dictionary will be enriched with more entries of other VMs.

The analysis on VMs carried out in this thesis is expected to be applicable not only to any marker, but also to any linguistic element which is studied through parallel corpora analysis in the context of translation.

It is expected that the study shown in this thesis is applicable to the comparative analysis of any pair of languages, not only English and Spanish, or even two groups of more than two languages at a time. The compilation of results in this thesis is expected to be useful in SMT in order to improve the output provided by the machine, enhancing the appropriateness and correctness of the translation provided.



# References

- Aijmer, K and Simon-Vandenberg, A-M. (2003). "The discourse particle well and its equivalents in Swedish and Dutch". *Linguistics* 41 (6): 1123-1161.
- Aijmer, K, Foolen, Ad, Simon-Vandenberg, A-M, in press. (2003). "Discourse particles in the perspective of heteroglossia". In: Fischer, K. (Ed.), "Approaches to Discourse Particles".
- Aijmer, K and Simon-Vandenberg, A-M. (2004). "A Modal and Methodology for the study of pragmatic markers: the semantic field of expectation". *Journal of Pragmatics* 33 1787-1805.
- Aijmer, K. (2004). "Pragmatic Markers in Spoken Interlanguage". *Nordic Journal of English Studies*, 3, 173-190.
- Aijmer, K. (2007): "Translating discourse particles: A case of complex translation, in Incorporating Corpora". *The Linguist and the Translator. Multilingual Matters*. Clevedon, page 95–116.
- Anscombe, J. C.; Ducrot, O. (1983). "L'argumentation dans la langue. Liège: Mardaga". - Ducrot, O. (1994). *La argumentación en la lengua*, ed. rev. y aum. Madrid: Gredos.
- Ariel, M. (1998). "Discourse markers and form-function correlations". In A. Jucker and Y. Ziv (eds.), *Discourse Markers. Descriptions and Theory*, 223 - 259. Amsterdam: John Benjamins.
- Austin, J. (1975). *How to Do Things with Words*: "Saying something will often, or even normally, produce certain consequential effects upon the feelings, thoughts, or actions of the audience, or of the speaker, or of other persons: and it may be done with the design, intention, or purpose of producing them." Oxford: Oxford University Press, p. 101.

- Baroni, M; Bernardini, S; Ferraresi, A and Zanchetta, E. (2009). “The wacky wide web: A collection of very large linguistically processed web-crawled corpora.” *Language Resources and Evaluation*, 43(3):209–226.
- Baroni, et al. (2010). “Distributional Memory: a general framework for corpus-based semantics,” in *Computational Linguistics* 36: 4, 673–721.
- Bazzanella, C. (1990). “Phatic connectives as interactional cues in contemporary spoken Italian”. *Journal of Pragmatics*. 14: 629 - 647.
- BBC MMXI: “Grammar, Vocabulary & Pronunciation: Learning English: Linking Adverbs”:  
<http://www.bbc.co.uk/worldservice/learningenglish/grammar/learnit/learnitv288.shtml>  
(10/2018).
- Biber, D. (2006). “University Language: A Corpus-Based Study of Spoken and Written Registers”. Amsterdam: John Benjamins.
- Beeching, K. (2013): “A parallel corpus approach to investigating semantic change, Advances in corpus-based contrastive linguistics”. *Studies in honour of Stig Johansson*. Amsterdam: John Benjamins. Pages 103–125.
- Beeching, K. and Detges, U editors (2014). “Discourse Functions at the Left and Right Periphery. Crosslinguistic Investigations of Language Use and Language Change”. Brill, 2014.
- Blakemore, D. (1987). “Semantic Constraints on Relevance”. *Semantics in Philosophy of Language*. Philosophy Documentation Center. 2019. Canada.
- Blakemore, D. (1988). “So as a constraint on relevance”. In R.M. Kempson (ed.). *Mental Representations. The interface between language and reality*. Cambridge: Cambridge University Press, 183-196.
- Blakemore, D. (1989a). “Linguistic form and pragmatic interpretation: the explicit and the implicit”. In L. Hickey (ed.). *The Pragmatics of Style*. London: Routledge, 28-51.
- Blakemore, D. (1989b). “Denial and contrast: a relevance analysis of but”. *Linguistics and Philosophy*

12, 12-37.

Blakemore, D. (1992). "Understanding Utterances". London. Blackwell.

Blakemore, D. (1993). "The relevance of reformulations". *Language and Literature* 2 (2), 101- 20.

Blakemore, D. (1994). "Relevance, poetic effects and social goals: a reply to Culpeper". *Language and Literature* 3 (1), 49-59.

Blakemore, D. (1995). "On so-called discourse connectives". Unpublished lecture.

Blakemore, D. (1996). "Are apposition markers discourse markers?", *Journal of Linguistics* 32, pages 325-347.

Blakemore, D. (1997). "Restatement and exemplification: a relevance theoretic reassessment of elaboration". *Pragmatics and Cognition* 5, pages 1-19.

British National Corpus: <https://www.sketchengine.eu/british-national-corpus/> (2019).

Briz Gómez, A. (1993a). "Los conectores pragmáticos en español coloquial (I): su papel argumentativo". *Contextos* XI, 21/22, páginas 145-188.

Briz Gómez, A. (1993b). "Los conectores pragmáticos en español coloquial (II): su papel metadiscursivo". *Español Actual* 59, páginas 39-56.

Briz Gómez, A. (1995). "Comentario argumentativo de un texto coloquial. La incidencia de los conectores pragmáticos". *Review of Applied Linguistics*. Páginas 107-108, 113- 142.

Briz Gómez, A. & A. Hidalgo. (1998). "Conectores pragmáticos y estructura de la conversación". In M. A. Martín Zorraquino & E. Montolio, (eds.). *Marcadores discursivos: Teoría y Práctica*. Madrid: Arco, páginas 122-142.

Burton-Roberts. N. (1975). "Nominal apposition", *Foundations of Language* 13, 391-419.

Roberts, N. 1994. "Apposition". In R.E. Asher J.M.Y. Simpson (eds.). *The Encyclopaedia of Language and Linguistics*. Pergamon Press and Aberdeen University Press, 184-187.

Cambridge University press. *Cambridge Grammar and Dictionary*.

<https://dictionary.cambridge.org/es/diccionario/> (October, 2018).

- Carlson, L. (1984). "Well' in dialogue games: A discourse analysis of the interjection 'well' in idealized conversation". Amsterdam: Benjamins.
- Carston, R. (1997). "Pragmatics and the Explicit-Implicit Distinction". PhD Thesis, University College London.
- Carter, R. A., & M. J. McCarthy. (1997). "Exploring Spoken English". Cambridge: CUP.
- Cartoni, B. (2012). "Improving MT Coherence Through Text-Level Processing of Input Texts: the COMTIS Project". <http://lodel.irevues.inist.fr/tralogy/index.php>.
- Cartoni, B. (2013). "Annotating the Meaning of Discourse Connectives by Looking at their Translation: The Translation Spotting Technique". *Dialogue & Discourse* 4(2): pages 65–86.
- Casado Velarde, M. (1991). "Los operadores discursivos: es decir, esto es, o sea y a saber: valores de lengua y funciones textuales". *Lingüística Española Actual* 13, páginas 87-116.
- Casado Velarde, M. (1993). Introducción a la gramática de/ texto en español. Madrid, Arco.
- Chafe Wallace, J. (1982). "Integration and involvement in speaking, writing and oral literature". In D. Tannen (ed.). *Spoken and written language*. Norwood, NJ: Ablex, 35:53.
- Collins online dictionary: <https://www.collinsdictionary.com/dictionary/english/conjunction> (10/2018).
- Cortés, L y Camacho, M. (2005). *Unidades de Segmentación y Marcadores del Discurso*. Madrid. Arco Libros.
- Cortés Rodríguez, L. (1998). "Marcadores del discurso y análisis cuantitativo". En Martín Zorraquino, M.A. & Montolío, E. (1998a): páginas 143-160.
- Cuenca, M. J. (2001) "Anàlisi Contrastiva dels marcadors de reformulació i exemplificació". *Caplletra* 30 (Primavera, 2001), pp. 47-72.
- CREA and CORDE <http://www.rae.es/recursos/banco-de-datos/crea-version-annotada>. Beta 1.0. (2019).
- Davies, M. (2018). iWeb corpus: <http://davies-linguistics.byu.edu/personal/> (2019)

- Davies, M. Corpus del español: Web dialects: <https://www.corpusdelespanol.org/web-dial/> (2019).
- Davies, M. (2018). Corpus del español: Google Books: <https://www.corpusdelespanol.org/web-dial/> (2019).
- Del Saz, M (2002): “An analysis of Discourse Markers of Reformulation”. Valencia.
- Dermot, M. (2009). Biggerplate. <https://www.biggerplate.com/mindmaps/n587689/conjunction> (10/2018).
- Dictionary.com: VOX, LARROUSE. [www.dictionary.com](http://www.dictionary.com) (2019).
- Diewald, G., Fischer, K., (1998). “Zur diskursiven und modalen Funktion der Partikeln aber, auch, doch und ja in Instruktionsdialogen”. *Linguistica*, 38, pages 75-99.
- Dinzikova I, and Paolini M. (2018). “Contrastive analysis of Terminology Phrasemes with colour components in English, Frech, Italian and Slovak. Field of Economy and Finance.” *Ekonomické Rozhl'ady- Economic Review*. Volume 47. Slovak Republic.
- Dirk, S. (2004). “Discourse Markers Across Languages: A Contrastive Study of Second-Level Discourse Markers in Native and Non-Native Text with Implications for General and Pedagogic Lexicography”. Routledge. London.
- Dobrovolskij, D and Ludmila, P. (2016). Stockholm University, Faculty of Humanities, Department of Slavic and Baltic Studies, Finnish, Dutch, and German.
- Ducrot, O. (1983). “Operateurs argumentatifs et visee argumentative”. *Cahiers de Linguistique Française* 5: “Connecteurs pragmatiques et structure du discours” (actes du weme Colloque de Pragmatique de Geneve). Geneve, pages 7-36.
- Dyvik, H. (1998). “A translational basis for semantics.” In S. Johansson and S. Oksefjell (1998), pages 51-86.
- Education First ltd. ‘Viewpoint and Commenting Adverbs’: <https://www.ef.com/wwen/english-resources/english-grammar/viewpoint-and-commenting-adverbs/> (10/2018).

English Language Smart Words: “Linking Words: also called connecting words – Overview”:

<https://www.smart-words.org/linking-words/> (10/2018).

Erman, B. (2001). “Pragmatic markers revisited with a focus on you know in adult and adolescent talk”.

*Journal of Pragmatics*. 33: 1337 - 1359.

Errazuriz Cruz, M. C. (2012). *Scielo: Perfiles Educativos*: “Análisis del uso de los marcadores discursivos en argumentaciones escritas por estudiantes universitarios/ Analysis of one of the discourse

markers in argumentations written by higher education students” 2012. vol.34 no.136.

Europarl7 English corpus: <https://www.sketchengine.eu/europarl7-english-corpus/> (2019).

Europarl7 Spanish corpus: <https://www.sketchengine.eu/europarl7-spanish-corpus/> (2019).

Fischer, K. (2000): *From Cognitive Semantics to Lexical Pragmatics*: “The Functional Polysemy of

Discourse Particles”. Mouton de Gruyter, Berlin/New York, 2000.

Fischer, K. (2006). “Approaches to Discourse Particles”. *Studies in Pragmatics*, 1 (pp. 339 - 358).

Amsterdam: Elsevier.

Fancellu, F and Webber, B. (2014). “Applying the semantics of negation to SMT through n-best list

re-ranking”. *In Proceedings of the 14th Conference of the European Chapter of the Association*

*for Computational Linguistics*. Gothenburg, Sweden, EACL, pages 598–606.

<http://aclweb.org/anthology/E/E14/E14-1063.pdf>.

Frank-Job, B. (2006). “A Dynamic-international approach to discourse markers”. In K. Fischer 2006.

Elsevier Ltd.

Fraser, B. (1988). “Types of English discourse markers”. *Acta Linguistica Hungarica*, 38, pages 19-33.

Fraser, B. (1990): An approach to discourse markers. *Journal of Pragmatics*, 14: pages 383– 395, 1990.

Fraser, B. (1996): ‘Pragmatic Markers’, *International Pragmatics Association*, 6: 2. pages 167- 169,

1996.

Fraser, B. (1998). “Contrastive Discourse Markers in English”. In: Jucker, A.H., Ziv, Y. (Eds.), *Discourse*

*Markers: Descriptions and Theory [Pragmatics and Beyond 57]*. Benjamins, Amsterdam.

- Fraser, B. (1999): “What are discourse markers?”, *Journal of Pragmatics*, 31: 931–952, 1999.
- Fraser, B. (2009): “Topic Orientation Markers”, *Journal of Pragmatics*, 41, 2009.
- Fox Tree, J.E., & J.C. Schrock. (1999). “Discourse markers in spontaneous speech: Oh what a difference an “oh” makes”. *Journal of Memory and Language* 40, pages 280- 295.
- Fuentes Rodriguez, C. (1987). “Enlaces extraoracionales”. Sevilla, Alfar.
- Fuentes Rodriguez, C. (1990). “Procedimientos intradiscursivos: decir y los explicativos”. *Sociolingüística andaluza* 5. Sevilla: Universidad de Sevilla, 103-123.
- Fuentes Rodriguez, C. (1993). “Conclusivos y reformulativos”. *Verba* 20, 171-196.
- Fuentes Rodriguez, C. (1998). *Diccionario de conectores y operadores del español*. Arco Libros. Madrid.
- Garacha Camarero, Mar. (2001-2002). “Gramaticalización y cambio sintáctico”. *Anuari de filologia*. Secció F, Estudios de lengua y literatura españolas, No. 11-12, 2001-2002, págs. 9-16.
- Garcés Gómez, M. (2005): “Reformulación y marcadores de reformulación”, en: Casado Velarde, Manuel/-González Ruiz, Ramón/Loureda Lamas, Óscar (eds.): *Estudios sobre lo metalingüístico en español*. Frankfurt: Peter Lang, 47-66.
- Ghezzi, C. and Molinelli, P ed. (2014): “Discourse and Pragmatic Markers from Latin to the Romance Languages”. *Oxford Studies in Diachronic and Historical Linguistics* 9. Oxford University Press, 2014.
- Gili Gaya, S. (1983 [1961]). *Curso superior de sintaxis española*. Barcelona: VOX.
- Glossary of linguistic terms online: <https://glossary.sil.org/term/conjunction> (10/2018).
- Goddard, C., Wierzbicka, A. (1994). “Semantic and Lexical Universals: Theory and Empirical Findings”. Benjamins, Amsterdam.
- Greenbaum, S. (1969). “Studies in English Adverbial Usage”. London: Longman.
- Grice, H. Paul. (1975/1989). “Logic and Conversation” in Peter Cole J. Morgan (eds.). *Syntax and Semantics 3: Speech-Acts*. New York: Academic Press, 41-58. Reprinted in H. Paul Grice.

- Studies in the Way of Words. Cambridge, Mass: Harvard University Press (1989).
- Guillou, L. (2012). "Improving Pronoun Translation for Statistical Machine Translation." *In Proceedings of the Student Research Workshop at the 13th Conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, Stroudsburg, PA, USA, EACL '12*, pages 1–10. <http://dl.acm.org/citation.cfm?id=2380943.2380944>.
- Guillou, L. (2016). "Incorporating Pronoun Function into Statistical Machine Translation". Ph.D. thesis, University of Edinburgh.
- Gupta, A.F., (1992). "The Pragmatic particles of Singapore Colloquial English". *Journal of Pragmatics*, 18, pages 31-57.
- Guzmán, F et al. (2014). "Using Discourse Structure Improves Machine Translation Evaluation." *In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014, June 22-27, 2014, Baltimore, MD, USA, Volume 1: Long Papers. The Association for Computer Linguistics*, pages 687–698. <http://aclweb.org/anthology/P/P14/P14-1065.pdf>.
- Halliday, M and Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Halliday, M. (1991): The Notion of "Context" in language education. Chapter 13. Library of the Congress. USA.
- Halliday, M and Matthiessen, M.I.M (2014). *Cohesion in English*. Routledge. London.
- Halliday, M and Matthiessen, M.I.M (2004, 2015): *Introduction to Functional Grammar*. Hodder Headline Group. London, Great Britain.
- Hallig, R & Wartburg, W. (1963). "Begriffssystem als Grundlage für die Lexikographie. Versuch eines Ordnungsschemas". Berlin: Akademie-Verlag (2nd ed.).
- Hansen, Maj-Britt. M. (1998a). "The semantic status of discourse markers". *Lingua* 104, pages 235-260.
- Hardmeier, C. (2012). "Discourse in Statistical Machine Translation: A Survey and a Case Study". Elanders Sverige, Sweden.
- Hardmeier, C et al. (2013a). "Docent: A Document Level Decoder for Phrase-Based Statistical Machine

- Translation.” *In 51st Annual Meeting of the Association for Computational Linguistics*, 2013, Proceedings of the Conference System Demonstrations, 4-9 August 2013, Sofia, Bulgaria. pages 193–198. <http://aclweb.org/anthology/P/P13/P13-4033.pdf>.
- Hardmeier, C et al. (2013b). “Latent Anaphora Resolution for Cross-Lingual Pronoun Prediction.” *In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Seattle, Washington, USA, pages 380–391. <http://www.aclweb.org/anthology/D13-1037>.
- Hatim, B.A, Muday, J. (2004). *Translation: An advanced resource book*. [ent.taylorfrancis.com](http://ent.taylorfrancis.com)
- Howard, J. R. (2010). *On the surface of discourse*. [researchgate.net](http://researchgate.net).
- Hummel, K.M. (1995). “Translation and second language learning”. *Canadian Modern Language Review*. [utpjournals.press](http://utpjournals.press)
- Ifantidou, E. (2001) *Evidentials and Relevance*. Amsterdam: John Benjamins.
- Ilysh, B. (1969). *BRNO Studies in English*: “Prepositions and Conjunctions in present-day English”. The Herzen Pedagogical Institute, Leningrad. Volume 8. Pages 99-103.
- Jamal, Z. ResearchGate. (2017). [https://www.researchgate.net/figure/Types-of-Conjunctions\\_tbl1\\_320911525](https://www.researchgate.net/figure/Types-of-Conjunctions_tbl1_320911525) (10/2018).
- Jucker, A. H. (1993). “The discourse marker well: A relevance theoretic account”. *Journal of Pragmatics* 19, pages 435-452.
- Jucker, A. H. Y. Ziv. (1998). “Discourse markers: Introduction”. In A. H. Jucker Y. Ziv. (eds.). *Discourse markers: Description and theory*. Amsterdam: Benjamins, pages 1-12.
- Khany, R. (2011). *A Comparative Study of Metaphorical Markers in Academic Research Articles*. Article 3, Volume 2. Ilam University.
- Kilgarriff, A. (2012-2019). “Translation Tools in the Sketch Engine Lexical Computing”. UK: Ltd.
- Knott, A. (1993). “Using Cue Phrases to Determine a Set of Rhetorical Relations”. In O. Rambow (ed). *Intentionality and Structure in Discourse Relations*. Proceedings of the ACL SIGGEN Workshop,

pages 1-4.

- Koehn, P. (2001). 'Europarl: A Parallel Corpus for Statistical Machine Translation'. Scotland: School of Informatics University of Edinburgh.
- Kuhi, D Almasi, K. (2013). "The Study of Impact of Learner's Personal Constructs in Illocutionary Acts Induction," *International Journal of Enhanced Research in Educational Development. Er Publications*. Vol. 1 Issue 6, Sept–Oct 2013, pages 4–37.
- Lavid, J, Arús, J, DeClerck, B and Hoste, V. (2015). "Creation of a high-quality, register- diversified parallel corpus for linguistic and computational investigations. In Current Work in Corpus Linguistics: Working with Traditionally- conceived Corpora and Beyond. Selected Papers from the 7th International Conference on Corpus Linguistics (CILC2015)." *Procedia - Social and Behavioral Sciences*, Volume 198, 24 July 2015, pages 249–256.
- Lakoff, W. (1973). "Questionable answers and answerable questions". In B. B. Kachru, R. B. Lees, Y. Malkiel, A. Pietrangeli & S. Saporta (eds.). *Issues in Linguistics. Papers in Honor of Henry and Renee Kahane*. Urbana: University of Illinois Press, 453-467.
- Lavid, J. and Avilés, E. (2019a): "Investigating Discourse Markers in English and Spanish through translation: the case of Elaborating Connectives". In Pérez Veneros, M and I. Elorza (eds). Proceedings of the 27o Congreso europeo lingüística sistémico-funcional.
- Lavid, J. and Avilés, E. (2019b). "Annotating Discourse Markers in the MULTINOT corpus: The case of elaborating connectives". In L.M. HO-Dac & P. Müller (eds.). *Cross-linguistic discourse annotation applications perspectives*. Final Action conference. University of Toulouse. pages: 89-91.
- Lavid, J (in press). (2019). "Translation correspondences of digressive discourse markers in English and Spanish: a corpus-based study". International Europhras Conference, University of Málaga.
- Lee, H-K. (2002). "Towards a new typology of connectives with special reference to conjunction in English and Korean". *Journal of Pragmatics* 34, pages 851-866.

- Lenk, U. (1998). "Discourse markers and global coherence in conversation". *Journal of pragmatics* 30, pages 245-257.
- Ler Soon Lay, V. (2006). "A relevance-theoretic approach to discourse particles". In Singapore English". *In Approaches to Discourse Particles*. K, Fischer (2006). Amsterdam; Boston; London: Elsevier.
- Levinson. S.C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Lewis, D.M. (2002). "Rhetorical factors in lexical-semantic change: the case of at least". In: Díaz Vera, J.E. (Ed.), *A Changing World of Words: Studies in English Historical Lexicography, Lexicology and Semantics*. Rodopi, Amsterdam, pages 525-538.
- Lewis, D.M. (2006). "Discourse markers in English: a discourse-pragmatic view". *Approaches to discourse particles*. [academia.edu](http://academia.edu)
- LF Aligner: <https://lf-aligner.soft112.com/> (2018-2019).
- Lizárraga Navarro, G. (2011). "Las clases de palabras en los libros de texto". *Revista Mexicana de Investigación Educativa*, vol. 16, núm. 49.
- Logos online dictionary: [www.logosdictionary.org](http://www.logosdictionary.org) (2019).
- Logos online library: [www.logoslibrary.eu](http://www.logoslibrary.eu) (2019).
- Mann, W.C. & S.A. Thompson. (1988). "Rhetorical structure theory: A theory of text organization". *Text* 8 (3), pages 243-281.
- Martín Zorraquino, MA and Portolés, J. (1999). "Los marcadores del discurso", In I. Bosque y V. Demonte (eds.), *Gramática Descriptiva de la Lengua Española*, tomo III, Madrid: Espasa Calpe, páginas 4051-4203.
- Mascarell, L. et al. (2015). "Detecting Document-level Context Triggers to Resolve Translation Ambiguity". *In Proceedings of the Second Workshop on Discourse in Machine Translation*. Association for Computational Linguistics, Lisbon, Portugal, pages 47-51. <http://aclweb.org/anthology/W15-2506>.
- Meyer, T. (2011). "Disambiguating Temporal Contrastive Discourse Connectives for Machine Translation". *In Proceedings of the ACL 2011 Student Session*. Association for Computational Linguistics.

*tics, Stroudsburg, PA, USA, HLT-SS '11*. pages 46–51.

<http://dl.acm.org/citation.cfm?id=2000976.2000985>.

Meyer, T et al. (2011). “Multilingual Annotation and Disambiguation of Discourse Connectives for Machine Translation”. In *SIGDIAL Conference*. The Association for Computer Linguistics, pages 194–203.

Meyer, T and Popescu-Belis, A. (2012). “Using Sense-labeled Discourse Connectives for Statistical Machine Translation”. In *Proceedings of the Joint Workshop on Exploiting Synergies Between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation (HyTra)*. Association for Computational Linguistics, Stroudsburg, PA, USA, EAACL 2012, pages 129–138.

McCarthy, M.J. (1991b). “Discourse Analysis for Language Teachers”. Cambridge: CUP. McCarthy, M.J. 1993a. “Grammar, Discourse and the Fanzine”. In Ikegami, Y. & M.

McCarthy, M. J. (1993b). “Spoken Discourse Markers in Written Text”. In Fox, G., Hoey, M., & J. Sinclair (eds.). *Techniques of Description*. London: Routledge, pages 170-182.

McCarthy, M. J. 1998. “Spoken Language and Applied Linguistics”. Cambridge: C. U. P. McCarthy, M.J., 2001. *Issues in Applied Linguistics*. Cambridge: Cambridge University Press.

Macmillan online dictionary: [www.macmillandictionary.com](http://www.macmillandictionary.com) (2019).

Merriam Webster online dictionary: <https://www.merriam-webster.com/dictionary/conjunction> (10/2018).

Meyer, T and Popescu-Belis, A. (2012). “Using sense-labelled discourse connectives for statistical machine translation”. In: *EAACL Joint Workshop on Exploiting Synergies between IR and MT, and Hybrid Approaches to MT (ESIRMTHyTra)*, pages 129-138.

Meyer, T and Poláková, L. (2013). “Machine Translation with Many Manually Labeled Discourse Connectives.” In *Proceedings of the 1st DiscoMT Workshop at ACL 2013* (51st Annual Meeting of the Association for Computational Linguistics). Sofia, Bulgaria, page 8.

Multinot Corpus (Lavid et al., 2015): <https://www.ucm.es/funccap/multinot> (10/2018).

- Murillo, S. (1999). "Some notes on the Informative Meaning of English Reformulatory Discourse Markers". Proceedings of the XX/I International Conference of AEDEAN, 197-292.
- Nemo, F. (1986). "Contraintes de pertinence et argumentativité". *Semantikos*, 9, pages 21-33.
- Newmark, P. (1988). *A Textbook of Translation*. NY: Prentice Hall.
- Nicolov, N et al. (2007). *Recent Advances in Natural Language processing. Selected Papers from 2007*: Tiedemann, Jörg. "News from Opus - A collection of not included Multilingual parallel Corpora". John Benjamin Publishing. University of Stockholm.
- Núñez, P, Muñoz, A, Mihovilovic, E. (2006). "Las funciones de los marcadores de reformulación en el discurso académico en formación". *Revista Signos* [en línea] 2006, 39 (Sin mes): [Fecha de consulta: 01 de febrero de 2019] Disponible en: <<http://www.redalyc.org/articulo.oa?id=157013770007>> ISSN 0035-0451
- Murthy, K.N. (2009). "Automatic Cateorization of Telegu News Articles". Department of Computer and Information Sciences, University of Hyderabad, Hyaderabad.
- Nyan, Thahn. (2006). "From procedural meaning to processing requirement". In Singapore English". *In Approaches to Discourse Particles*. K, Fischer (2006).
- O'Donnell, M. 1994. "Sentence analysis and generation: a systemic perspective". Sydney University: Ph.D
- Opus2 English corpus: <https://www.sketchengine.eu/opus2-english-corpus/> (2019).
- Opus2 Spanish corpus: <https://www.sketchengine.eu/opus2-spanish-corpus/> (2019).
- Östman, J. (1995). "Pragmatic particles twenty years after". *Anglicana Turkuensia*. 14: pages 95 - 108.
- Oxford English Dictionary (OED), *English Oxford living Dictionaries online (EOLD)*, (2018), <https://en.oxforddict> (2018).
- Pennock Speck, B. (1994). "The rhetoric of coherence, reformulation, repetition and augmentation of given information in newspaper texts". Actas del Primer Encuentro Interdisciplinar sobre Retórica, Texto y Comunicación. Cadiz, páginas 227- 230.
- Polanyi, L and Scha, R. (1983). *The syntax of text and discourse*. Text 3: pages 261 - 270.

- Pons Borderia, S. (1998a). "Reformulación y reformuladores: A proposito de Les operations de reformulation". *Oralia* 1, 183-198.
- Pons Borderia, S. (1998b). "Conexión y Conectores: el estudio de su relación en el registro informal de la lengua". *Cuadernos de Filología*, Anejo XXVII. Facultat de Filología. Universitat de Valencia.
- Pons Borderia, S. (2000). "Los conectores". In A. Briz Grupo Val.Es. Co (eds.). *Como se comenta un texto coloquial*. Barcelona, Ariel, páginas 193-220.
- Pons Borderia, S. (2001). "Connective/Discourse Markers. An Overview". *Quaderns de Filologia*. Estudis Lingistics VI, páginas 219-243.
- Portoles Lazaro, J. (1994). "La distinción entre los conectores y otros marcadores del discurso en español". *Verba* 29, páginas 141-170.
- Portoles Lazaro, J. (1995a). "Diferencias gramaticales y pragmáticas entre los conectores discursivos, pero, sin embargo y no obstante". *BRAE*, LXXV, Mayo-Agosto, páginas 231-269.
- Portolés, J. - Martín Zorraquino (1999), "Los marcadores del discurso", en Bosque, I. y Demonte, V. (dirs.), *Gramática descriptiva de la lengua española*. Madrid: Espasa Calpe, capítulo 63.
- Quirk, R., Greenbaum, S., Leech, G. & J. Svartvik. (1972). *A Grammar of contemporary English*. London, Longman.
- Quirk, R. & S. Greenbaum. (1973). *A University Grammar of English*. London: Longman.
- Quirk, R., Greenbaum, S., Leech, G. & J. Svartvik. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
- Real Academia Española (RAE), *Diccionario online de la Real Academia (DRAE)*, (2017) <http://dle.rae.es/?id=Dg> (2018).
- Redeker, G., (1986). "Language use in informal narratives: Effects of social distance and listener involvement". Doctoral dissertation, University of California, Berkeley. Redeker, G., 1990. Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics*, 14, pages 367-381.
- Redeker, G. (1990). "Ideational and pragmatic markers of discourse structure". *Journal of Pragmatics*

14, pages 367-381.

Redeker, G. (1991). "Linguistic markers of discourse structure". *Linguistics* 29, pages 1139- 1172.

Redeker, G., (1992). 'Kleine woordjes' in spontaan taalgebruik – stoplapjes of signalen voor de lezer/luisteraar? ['Small words' in spontaneous language use – fillers or signals for the reader/hearer?]. *Toegepaste Taalwetenschap in Artikelen*, 43, pages 55-65.

Redeker, G. (2006). "Discourse markers as attentional cues at discourse transitions". In K. Fischer (Ed.), *Approaches to Discourse Particles*. Studies in Pragmatics, 1 (pp. 339 - 358). Amsterdam: Elsevier.

Robin L. (1997 – 2018): 'The Conjunctive Adverb':

<http://www.chompchomp.com/terms/conjunctiveadverb.htm> (10/2018).

Rossari, C. (1990). "Projet pour une typologie des operations de reformulation". *Cahiers de Linguistique Fran9aise* 14, pages 151-171.

RMIT English Worldwide. (2015). <https://rmitenglishworldwide.com/blog/cohesion-part-2-linking-words> (19/2018).

Roulet, E. (1980). "Interactional markers in dialogue". *Applied Linguistics*. Vol 1 (3), pages 224- 233.

Schiffrin, D. (1987). *Discourse Markers*. Cambridge University Press, Cambridge.

Schiffrin, D. (2012). *Discourse Markers*. Cambridge University Press. Georgetown University, Washington DC.

Schourup, L. (1999). *Discourse markers*. *Lingua*. 107: pages 227 - 265

Seyed Ali, Rezvani Kalajahi, Ain Nadzimah Abdullah, Jayakaran Mukundan and Dan J. Tannacito (2012). "Discourse Connectors: An Overview of the History, Definition and Classification of the Term". *World Applied Sciences Journal* 19 (11): pages 1659-1673. Maltepe University. Turkey.

Siepmann, D and Sprachen, L. (2001). "Multi-word Discourse Markers in Translation: a Corpus- based Investigation into Restrictors". *Zeitschrift für interlinguale und interkulturelle Kommunikation*. Volume 46, Issue 3. 2019. Ed. by Schmitt, Peter A. / Lee-Jahnke, Hannelore. Online. *The*

*Gruyter.*

Sinclair, J. McH., R. M. Coulthard. (1975). "Towards an Analysis of Discourse". Oxford: Oxford University Press.

Sketch Engine: <https://www.sketchengine.eu/what-can-sketch-engine-do/> (2019).

Smith, K. (2017). *Proceedings of the Third Workshop on Discourse in Machine Translation*: "On Integrating Discourse in Machine Translation". Association for Computational Linguistics. Copenhagen. pages 110–121

Sourceforge: LF Aligner: <https://sourceforge.net/projects/aligner/> (10/2018).

Stede, M. & Schmitz, B. *Machine Translation* (2000) 15: page 125.

Steele, D. (2015): "Improving Improving the Translation of Discourse Markers for Chinese into English". Proceedings of NAACL-HLT 2015 Student Research Workshop (SRW), pages 110–117, Denver, Colorado, June 1, 2015. Association for Computational Linguistics.

Steele, D and Specia, L. (2016). "Predicting and Using Implicit Discourse Elements in ChineseEnglish Translation." *In Proceedings of the 19th annual conference of the European Association for Machine Translation (EAMT)*. Riga, Latvia, pages 305–317.

Stubbs, M. 1983. *Discourse Analysis: The Sociolinguistic Analysis of Natural Language*(Language in Society, 4). USA. University of Chicago Press.

Sunday Unibi, Abraham. (2016). "Conjunctions in English: Meaning, types and uses". *International Journal of Social Science and Humanities Research* 2348-3164 (online).

Swan. M. (1985). *Practical English Usage*. Oxford: Oxford University Press.

TEI: Text Encoding Initiative: <http://www.tei-c.org/> (10/2018).

Teich, E. (2009). "Computational linguistics". In: Halliday & Webster (eds). Pages 113–127.

The Free dictionary. Com: [www.thefreedictionary.com](http://www.thefreedictionary.com) (2019).

Thompson, G. (2005). "But me some but: a multi-dimensional view of conjunction". *Text* 25.6: pages 763- 791.

- Tiedemann, J. (2007). "Parallel Data, Tools and Interfaces in OPUS". Department of Linguistics and Philology Uppsala University, Uppsala/Sweden.
- Timestamped JSI web corpus (2014-2019) English <https://www.sketchengine.eu/timestamped-english-corpus/>
- Toyota (eds.). "Aspects of English as a World Language". Tokyo: Maruzen Co., Ltd, pages 147- 161.
- Travis, C.E., (1998). "Bueno: A Spanish interactive discourse marker". Berkeley Linguistics Society, 24, pages 268-279.
- Travis, C.E. (2003). "The semantics of the Spanish subjunctive: its use in the Natural Semantic Meta-language". *Cognitive Linguistics*, 14, pages 47-69.
- Trujillo, F. (2003). "Culture in writing: Discourse in English and Spanish student writing". *Tadea Seu Liver de Amicitia*. Granada: Universidad de Granada.
- UKWaC corpus: <https://www.sketchengine.eu/ukwac-british-english-corpus/> (2019).
- UW Staff Web Server, University of Washington. Index of /marynell/grammar: logicalconnectors.html: <http://staff.washington.edu/marynell/grammar/logicalconnectors.html> (28-Feb-1998 09:02) (10/18).
- Van Dijk, T.A. (1977a). *Text and context*. London: Longman.
- Van Dijk. T.A. (1977b). "Connectives in text grammar and text logic". In T. van Dijk & J. Petofi (eds.). *Grammars and descriptions*. New York: De Gruyter Press, pages 11-63.
- Van Dijk, T.A. (1979). "Pragmatic connectives". *Journal of Pragmatics* 3, pages 447-456.
- Van Dijk. T. A. (1980). "Texto y contexto". Madrid. Catedra. Vol. 4, Issue 3, pages 202-213, Month: July - September 2016.
- Van Dijk. T. A. (2006). "Discourse, context and cognition". *Discourse studies*. [journals.sagepub.com](http://journals.sagepub.com)
- Vercruyse, Astrid. (2009). "Un estudio de algunos reformuladores en español: a saber, esto es, es decir, o sea, mejor dicho." Universiteit Gent. Netherlands.
- Yang, L.-C. (2001). "Prosodic shape and expressive meaning: the expression and recognition of emotions

- through prosody”. Proceedings of the ISCA Prosody and Speech Recognition Workshop. Red Bank, NJ.
- Yates, J. (2006). “Algunas Partículas escalares del español y su traducción al inglés”. Madrid: Universidad Autónoma de Madrid.
- Yung, F et al. (2015). “Crosslingual Annotation and Analysis of Implicit Discourse Connectives for Machine Translation.” In Proceedings of the Second Workshop on Discourse in Machine Translation. Association for Computational Linguistics, Lisbon, Portugal, pages142–152.  
<http://aclweb.org/anthology/W15-2519>.
- Waltereit, R. (1999). “Reanalyse als metonymischer Prozess”. In: Lang, J., Neumann-Holzschuh, I. (Eds.), Reanalyse und Grammatikalisierung in den romanischen Sprachen [*Linguistische Arbeiten* 410]. Niemeyer, Tübingen, pages. 19-29.
- Waltereit, R. (2001). “Modal particles and their functional equivalents: A speech-act-theoretic approach”. *Journal of Pragmatics*, 33. Pages 1391-1417.
- Warner, R. (1985). “Discourse connectives in English”. New York: Garland.
- Webster, M. (1828). <https://www.merriam-webster.com/dictionary/denominal> (2018).
- Wetzel, D and Bond, F. (2012). “Enriching Parallel Corpora for Statistical Machine Translation with Semantic Negation Rephrasing”. In *Proceedings of the Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation*. Jeju, Republic of Korea, SSST-6, pages 20–29.  
<http://dl.acm.org/citation.cfm?id=2392936.2392940>.
- Wikipedia in: “Powers”, Shelley (2003). Practical RDF. O’Reilly. <https://en.wikipedia.org/wiki/RSS> (2019).
- Wilson, D. & Sperber, D. (1986) Pragmatics and modularity. Chicago Linguistic Society 22, Parasession on Pragmatics and Grammatical Theory. 68-74. Reprinted in Davis (ed.), pages 583-95.
- Wilson, D. & Sperber, D. (1993) *Linguistic form and relevance*. *Lingua* 90. Pages 1-25.
- Wong, B. and Kit, C. (2012). “Extending Machine Translation Evaluation Metrics with Lexical Cohesion.

To Document Level”. In *Proceedings of EMNLP-CoNLL*. pages 1060–1068.

Wu, C. (2009). “Corpus-based research”. In: Halliday & Webster (eds). 128–142.

Wystra, Nioman. Wordpress. Diposkan pada 20 November (2016).

[https://nyomanwiaspding.wordpress.com/2016/11/20/logical-connectors/\(10/2018\).](https://nyomanwiaspding.wordpress.com/2016/11/20/logical-connectors/(10/2018).)

[http://sydney.edu.au/learningcentre/clearer\\_writing/module4/logical\\_relations/](http://sydney.edu.au/learningcentre/clearer_writing/module4/logical_relations/)

[logical\\_relations\\_intro.html](#) retived 5 January 2016.

Xiong. D et al. (2013a). “Modeling Lexical Cohesion for Document-Level Machine Translation.” In *Proceedings of IJCAI*.

Xiong. D et al. (2013b). “Lexical Chain Based Cohesion Models for Document-Level Statistical Machine Translation”. In *Proceedings of EMNLP*. pages 1563– 1573.

Zeevat, H. (2001). “The semantic role of modal particles”. In: van Rooy, R. (Ed.), *Proceedings of the 13th Amsterdam Colloquium*. ILLC, Amsterdam.

Zeevat, H. (2002). “Explaining presupposition triggers”. In: van Deemter, K., Kibble, R. (Eds.), *Information Sharing*. CSLI, Stanford, CA, pages 61-87.



# Appendices



## Appendix A. Curriculum Vitae

## Education

9/2015 to present	<p><b>PhD in English Linguistics (Universidad Complutense de Madrid)</b></p> <ul style="list-style-type: none"> <li>· Thesis: <i>TRANSLATION CORRESPONDENCES OF VERIFICATIVE MARKERS IN ENGLISH AND SPANISH: A CORPUS-BASED STUDY.</i></li> </ul>
28/07/2016	<p><b>Postgraduate University course in Spanish for foreign students (INESEM: instituto europeo de estudios empresariales)</b></p> <ul style="list-style-type: none"> <li>· This course was focused on the acquisition of the necessary skills for Spanish teaching to foreign students.</li> </ul>
20/10/2015	<p><b>MASTER IN ENGLISH TEACHING FOR HIGHER EDUCATION (Universidad a distancia de Madrid (UDIMA))</b></p> <ul style="list-style-type: none"> <li>· Thesis: <i>La teoría de las inteligencias múltiples aplicada a la adquisición de las competencias básicas en el aula de inglés.</i></li> </ul>
27/10/2011	<p><b>BA in <i>English Studies</i> (Universidad Complutense de Madrid)</b></p> <ul style="list-style-type: none"> <li>· Major in English Linguistics.</li> <li>· Major in English Literature.</li> </ul>
2004	<p><b>Postgraduate University course in mathematics, informatics and English for engineering purposes (Universidad Politécnica de Madrid)</b></p> <ul style="list-style-type: none"> <li>· These courses were part of the university degree in Telecommunication Engineering.</li> </ul>
May 2014	<p><b>CAMBRIDGE PROFICIENCY IN ENGLISH C2</b></p>
25/07/2017	<p><b>DEL F Alliance Française B2</b></p>

## Courses and Additional training

---

<b>2019</b>	<b>Habilitación Lingüística por la CAM</b>
<b>15/02/2017</b>	<b>ETWINNING COMO RECURSO EN EL AULA (CAM Course)</b>
<b>05/11/15</b>	<b>Grammar, Spelling and Punctuation workshop Universidad de Educación de la UCM.</b>
<b>16/08/15</b>	<b>How to teach Cambridge KET Cambridge learning</b>
<b>06/08/15</b>	<b>How to teach Cambridge PET Cambridge Learning</b>
<b>21/03/15</b>	<b>Nuevas Metodologías de Mobile Learning Fundación Tripartita</b>
<b>21/03/15</b>	<b>Curso de Inteligencias Múltiples Fundación Tripartita</b>
<b>11/12/15</b>	<b>Curso de Educación Emocional Colegio Internacional de Aravaca</b>
<b>30/12/14</b>	<b>CBL and PBL Fundación Tripartita</b>
<b>2018</b>	<b>CAPMAN TOEFL Workshop</b>
<b>2018</b>	<b>Trinity workshop Facultad de Educación de la UCM</b>
<b>2019</b>	<b>Cursos de Latex 2019 facultad de informática de la UCM</b>

## Working Experience

---

<b>09/18-present</b>	<p><b>CENP: Centro español de nuevas profesiones (University)</b></p> <ul style="list-style-type: none"> <li>· English teacher in Graphic Design degree.</li> </ul>
<b>09/18-present</b>	<p><b>Escuela Inglesa Pozuelo de Alarcón</b></p> <ul style="list-style-type: none"> <li>· Official certificates teacher for Cambridge, Trinity, iBT and Public State Competitions</li> </ul>
<b>20/10/2015</b>	<p><b>Universidad Complutense de Madrid: Lenguas Modernas y Traductores (IULMyT)</b></p> <ul style="list-style-type: none"> <li>· Terminology and document analysis teacher (Master in translation)</li> </ul>
<b>Freelance</b>	<p><b>Cambridge University Press</b></p> <ul style="list-style-type: none"> <li>· Proof Reader</li> </ul>
<b>2015-2017</b>	<p><b>CECBAF KHALIL GIBRAN,S.L</b></p> <ul style="list-style-type: none"> <li>· Languages coordinator: English, French and German.</li> <li>· English teacher: secondary education</li> </ul>
<b>2013</b>	<p><b>Fundación Juan March</b></p> <ul style="list-style-type: none"> <li>· Cultural events coordinator</li> <li>· Proof reader</li> <li>· Film subtitler</li> <li>· Editor</li> <li>· English translator</li> </ul>

## Projects and Publications

---

2018	<p><b>Textlink2018 TOULOUSE:</b>          • Contribution to text annotation and parallel corpora</p>
09/2016-present	<p><b>PhD on Multinot (Lavid et al, 2015) and parallel corpora in Sketch Engine.</b></p>
2013	<p><b>Department of Musicology of UCM</b>          • “Catálogo Colectivo de Pliegos de Villancicos, fase 5” for the Ministry of Industry</p>
Freelance	<p><b>Cambridge University Press</b>          • Proof Reader</p>
26/05/13	<p><i>Un Átomo Para la Vida y un Fotón para la Eternidad</i></p>
2018	<p><b>27ESFLC PROCEEDINGS VOLUME</b>          ·Lavid, J. and Avilés, Estefanía.          “Investigating Discourse Markers in English and Spanish:          The case of elaborating connectives”</p>

## Contributions to seminars

- Avilés, Estefanía. 2013. *Un Átomo Para la Vida y un Fotón para la Eternidad*. United PC. Madrid.
- Avilés, Estefanía. May 2017. “Translation Correspondences of Discourse Relational Devices in English and Spanish: A corpus Annotation Study”. *Dissertation Workshop*. Universidad Complutense de Madrid.
- Avilés, Estefanía. December 2017. “Conectores de elaboración/ elaborativos en inglés y en español: un estudio preliminar de traducción basada en corpus”. *Phday*. Universidad Complutense de Madrid.
- Avilés, Estefanía. Abril 2018. “LOS MARCADORES DEL DISCURSO Y SUS CORRESPONDENCIAS EN TRADUCCIÓN ENTRE EL INGLÉS Y EL ESPAÑOL: UN ESTUDIO SOBRE CASOS EN CORPUS”. *II Concurso de Tesis en 3 minutos*. Universidad Complutense de Madrid.
- Lavid, Julia and Avilés, Estefanía. June 2017. “Investigating Discourse Markers in English and Spanish: The case of elaborating connectives”. *27th ESFLC*. Universidad de Salamanca.
- Lavid, Julia and Avilés, Estefanía. January 2018. “Elaborating discourse markers in English and Spanish parallel texts: a corpus-annotation study”. *LPTS conference*. París.
- Lavid, Julia and Avilés, Estefanía. March 2018. “Annotating Discourse Markers in the MULTINOT corpus: The case of Elaborating Connectives in English and Spanish”. *TEXTLINK FINAL ACTION CONFERENCE*. TOULOUSE
- Lavid, Julia and Avilés, Estefanía. May 2018. “Bilingual annotation of elaborating connectives in English and Spanish: a corpus-based translation

study”. *INTERNATIONAL CONFERENCE ON CORPUS LINGUISTICS* (CILC2018). University of Extremadura (Cáceres, Spain).

Lavid, Julia and Avilés, Estefanía. September 2018. “Translation correspondences of elaborating connectives in English and Spanish: a contrastive corpus study”. *SdC IWoDA’ International Workshop on Discourse Analysis*. Santiago de Compostela, Spain.

Lavid, Julia and Avilés, Estefanía. November 2018. “Verificative connectives in English and Spanish: a corpus-based translation study”. *International Symposium on Parallel Corpora*. Universidad Complutense de Madrid.

### **Symposium attendance**

- *Understanding Nature, Understanding Beauty: a proposal for the study of aesthetics in language* 2019 UCM
- *BILINGUALISM AND INTERCULTURALITY: CHALLENGES, LIMITS AND SOLUTIONS* 2018 UNED
- *Bilingualism and interculturality: challenges, limits and solutions* 29-30/11/18 UCM
- *Metaphor in Gesture*: 18/05/18 UCM
- *Exploring Standard Language and Change*: 13/03/18 UCM
- *The Rise of Standard English*: 13/03/18 UCM
- *Investigado la internacionalización de la docencia universitaria desde una perspectiva interdisciplinar (INTER\_LICA)*: 6/11/17 UCM
- *Gender and Sexuality: seminario de lingüística avanzada* 25/04/2019 UCM

**Additional Merits**

- Tutor in the Master of professor training. Rey Juan Carlos University
- 01/06/17
- Tutor in the Master of professor training. Francisco de Vitoria University 03/05/16
- Tutor in the Master of professor training. Isabel I University. 06/06/16



## Appendix B. Annotated examples in MULTINOT (Lavid et al., 2015)

	A	B	C	D	E
1	Instructions:	EDM found: for example (x2): por ejemplo: at least (x1): tan como*	ES	LAV	
2	1) Review and correct the pairings. See instructions on worksheet 2.				
3	2) Write your notes (to be added to each translation unit in the TMX) in column C if you wish.				
4	3) Save and close this file, and close any other open spreadsheets.				
5	4) Return to the aligner window.				
6					
7					
8					
9	What Personal Information About Customers Does Amazon.com Gather?	¿Qué datos personales de sus clientes recopila Amazon.es?			
10	The information we learn from customers helps us personalize and continually improve your Amazon experience. Here are the types of information we gather:	La información personal que recabamos de nuestros clientes nos ayuda a personalizar y mejorar continuamente sus compras en Amazon.es. En general, utilizamos la información para gestionar pedidos, entregar productos y ofrecer servicios, procesar pagos, ponernos en contacto con usted en relación con pedidos, productos, servicios y ofertas promocionales, actualizar nuestros registros, realizar el mantenimiento de su cuenta en Amazon, mostrar contenidos como las listas de deseos y opiniones de los clientes y para recomendar productos y servicios que puedan serle de interés. También hacemos uso de esta información para mejorar nuestra tienda y nuestra plataforma, así como para prevenir o detectar fraudes o abusos en nuestra sitio web, y para permitir a terceros partes llamar a cabo soporte técnico, logístico u otras funciones en nuestro nombre.			
11	Information You Give Us: We receive and store any information you enter on our Web site or give us in any other way. Click here to see examples of what we collect. You can choose not to provide certain information, but then you might not be able to take advantage of many of our features. We use the information that you provide for such purposes as responding to your	Estos son los tipos de información que recopilamos: Información que usted nos facilita: Recibimos y almacenamos toda la información que usted introduce en nuestra página web o que nos facilita de cualquier otro modo. Haga clic aquí para ver algunos ejemplos de información que recopilamos. Puede elegir no facilitar cierta información, pero si así lo hace, se podrá disfrutar de muchas de las funciones que le			

	A	B	C	D	E
1	Instructions:	Cosectores: en realidad/ actualy; como/ as such; como/as (2)	EA		
2	1) Review and correct the pairings. See instructions on worksheet 2.				
3	2) Write your notes (to be added to each translation unit in the TMX) in column C if you wish.				
4	3) Save and close this file, and close any other open spreadsheets.				
5	4) Return to the aligner window.				
6					
7					
8					
9	PRIMERA PARTE	Part I ~			
10	LA POBREZA DEL HOMBRE COMO RESULTADO DE LA RIQUEZA DE LA TIERRA	Manfred's Poetry a is a Consequence of the Wealth of the Land			
11	1. FIEBRE DEL ORO, FIEBRE DE LA PLATA	1. Lust for Gold, Lust for Silver			
12	EL SIGNO DE LA CRUZ EN LAS EMPUÑADURAS DE LAS ESPADAS	THE SIGN OF THE CROSS ON THE HILT OF THE SWORD			
13	Cuando Cristóbal Colón se lanzó a atravesar los grandes espacios vacíos al oeste de la Escúmeria, había aceptado el desafío de las leyendas.	When Christopher Columbus headed across the great emptiness west of Christendom, he had accepted the challenge of legend.			
14	Tempestades terribles jugaban con sus naves, como si fueran cáscaras de nuez, y las arrojaban a las bocas de los monstruos; la gran serpiente de los mares tenebrosos, hambrienta de carne humana, estaría al acecho.	Terrible storms would play with his ships as if they were nutshells and had them into the jaws of monsters; the sea serpent, hungry for human flesh, would be lying in wait in the murky depths.			
15	Sólo faltaban mil años para que los fuegos purificadores del juicio final aniquinaran el mundo, según creían los hombres del siglo XV, y el mundo era entonces el mar Mediterráneo con sus costas de ambigua proyección hacia el África y Oriente.	According to fifteenth-century man, only 1,000 years remained before the purifying flames of the Last Judgment would destroy the world, and the world was then the Mediterranean Sea with its uncertain horizons: Europe, Africa, Asia.			
16	Los navegantes portugueses aseguraban que el viento del oeste trala cadáveres extraños y a veces arrastraba leñas curiosamente talladas, pero nadie sospechaba que el mundo sería, pronto, asombrosamente multiplicado.	Portuguese navigators spoke of strange corpses and curiously carved pieces of wood that floated in on the west wind, but no one suspected that the world was about to be startlingly extended by a great new land.			
17	América se solo carecía de nombre.	America not only lacked a name.			
18	Los noruegues no sabían que la habían descubierta hacía largo tiempo, y el propio Colón murió.	The Norwegians did not know they had discovered it long ago, and Columbus himself died.			

