



Parallel Corpora across Disciplines:  
New Challenges Ahead

4<sup>th</sup> International Symposium PaCor 2023

León, November 22-24, 2023

# BOOK OF ABSTRACTS

## PLENARY SPEAKERS

### **Translation is changing, and so should corpora: complex parallel corpus designs for translation research**

Silvia Bernardini

University of Bologna (Italy)

Advances in artificial intelligence have recently brought language technologies to the attention of society. The anxieties and excitement that translators and translation scholars have experienced since 2018, faced with the dramatic improvements in machine translation, are now suddenly shared by politicians, philosophers and professionals from all walks of life, worrying over the impact of large language models. The discussions and reflections that have taken place within translation studies for the past five years have given translation scholars an edge on the rest of society: (most) translation scholars are well aware that translation in the traditional sense is going to be reduced in scope and importance, but also that it is going to be replaced by a wide range of activities involving translation-like tasks, characterized by varying degrees of professionalism, flexible multilingualism and human-computer collaboration. This opens up a wide range of possibilities for empirical research on translation with corpora, as long as we adapt our research paradigms to the new normal.

This means first of all redefining and delimiting our object of study, by encompassing non-prototypical forms of translation, and revising the methods we use to investigate them. In this talk I will reflect on these topics, referring to two settings that epitomize the complex nature of translation studies today. First, I will focus on multilingual news writing by global news agencies, a complex discourse production setting in which translation plays a fundamental, though often unacknowledged, role. Building corpora for the study of multilingual news writing is a challenging task, from the identification of parallel segments to the set-up of hybrid *comparallel* corpora. Second, I will consider research frameworks in which different target texts are available for a single source. These *multiple translation* corpora are likely to become more and more central to our research, thanks, among others, to the availability of machine translation engines and to growing interest in translation learner corpora. Yet current corpus query tools struggle with these textual constellations: this is no coincidence, as these constructs stretch the definition of ‘corpus’ to its limits. Both research settings point to the centrality of parallel corpora as a research infrastructure, but at the same time signal the need for more flexibility in how we conceive of well-established translation and corpus categories and assumptions.

## Trends in Teen Talk from a Crosslinguistic Perspective

Ignacio Palacios

Universidad de Santiago de Compostela

Teen talk has been studied extensively in the last decades (Stenström & Jørgensen 2009, Tagliamonte 2016, Palacios 2018, Groff et al. 2022, to mention just a few) for a number of reasons: (i) teenagers represent an important sector of society who deserve great attention; (ii) teenagers are often regarded as language innovators; (iii) the language of teenagers constitutes a code of its own; and (iv) it is important to know how language develops and changes over time and across different generations.

This paper will be concerned with some of the main features of the grammar and lexis of teen talk in different European languages. The purpose is to come up with some general trends which may apply to the expression of teenagers in different languages. These will include the following:

- (i) High occurrence of sound and onomatopoeic or non-lexical words;
- (ii) A special quotative system different in part from that of adults with a high preference for zero or null quotatives;
- (iii) High use of vague language;
- (iv) The language of intensification;
- (v) Common use of taboo vocatives;
- (vi) Orthography. It is quite common that young speakers spell words differently from the standard; and,
- (vii) Invariant tags. They are very common in the discourse of young speakers of all languages.

I will be using a corpus-based methodology, which will include the analysis of some comparable corpora and some multilingual resources apart from general corpora from the languages object of study. A group of informants also provided me with useful information on the different languages studied.

The analysis of these findings will allow me to come to some conclusions that will hopefully contribute to a better description and understanding of teen talk. Some methodological issues will also be tackled.

### References

- Groff, Cynthia, Andrea Hollington, Ellen Hurst-Harosh, Nico Nassenstein, Jacomine Nortier, Helma Pasch and Nuzia Yannuar. (eds.). 2022. *Global Perspectives on Youth Language Practices*. Berlin: Mouton.
- Palacios Martínez, Ignacio M. 2018. "Lexical innovation in the language of teenagers. A cross-linguistic perspective". In Arne Ziegler, ed. *Jugendsprachen: Aktuelle Perspektiven internationaler Forschung*. Berlin: Gruyter, 363-390
- Stenström, Anna-Brita and Annette Myre Jørgensen. 2009. *Youngspeak in a Multilingual Perspective*. Amsterdam: Benjamins.
- Tagliamonte, Sali. 2016. *Teen Talk. The Language of Adolescents*. Cambridge: CUP.

## **Do we think populists actually do something? A discursive analysis of populists as agents or patients in French, Spanish and Belgian tweets**

Barbara De Cock  
Université Catholique de Louvain (Belgium)

In this presentation, I will look into how the term populists is used in a comparable corpus of French, Spanish and Dutch tweets. This research is part of the broader project Tracking the uses of populism in media and political discourse (TrUMPo) (<https://www.discourseonpopulism-uclouvain.com/>), which adopts an interdisciplinary perspective to the analysis of discourse on populism (as opposed to discourse by politicians considered as populists). The data collected concern Spain, France and Belgium (both the French-speaking and Dutch-speaking parts) during 2019, a year of European and local elections in the countries involved. Though not all parts of society are equally represented on Twitter, it does offer a wider variety of contributors than studying political or media discourses, making it particularly relevant to analyze discourse on a phenomenon of broader societal and political relevance.

I will focus more specifically on what kind of actions are typically attributed to populists through a combination of corpus linguistics methods, agentivity analysis and discourse analysis. Indeed, analyses of the representation of other collectives have shown that presenting a certain group as agentive or not, as well as the type of action associated with them, have an impact on the conceptualization of this group (De Cock & Michaud 2014, Van Leeuwen 2015). This analysis then reveals whether the populists are presented as agents or patients of an action, and the type of actions with which they are associated.

The contrastive analysis will take into account both linguistic specificities and information concerning the political and electoral context. Indeed, in terms of contrastive linguistics, the languages included display differences as to the morphosyntactic behavior of *populistas*, *populistes*, *populisten*, which can be both adjective and noun in Spanish and French, but only a noun in Dutch. Moreover, for all data involved, constructional features such as modifier or article use may contribute to an interpretation as more specific or more generic. Also the construction of agentivity is subject to language-specific features, especially in the domain of depersonalizing constructions. The findings from the comparable corpus study will be presented against the background of the on-going government coalition, the electoral campaign and the position of parties considered by some as populist in the countries involved in this study.

Through this analysis, I will then show whether populists are presented in tweets as acting intentionally, as well as what kind of behavior is associated with them, focusing on how linguistic and contextual specificities influence the representation in the different comparable corpora involved. At a broader methodological level, I will point at routes for carrying out comparable corpora studies in interaction with interdisciplinary approaches to the political context in which the discourse has been produced.

### **References**

- De Cock, Barbara & Michaud Maturana, Daniel. 2014. La expresión de la agentividad en el Informe Rettig (Chile, 1991). In: *Revista Internacional de Linguística Iberoamericana*, Vol. XII, no.23, p. 123-140.
- Van Leeuwen, Maarten. 2015. Taalkundig-stilistische analyse: de casus Wilders/Pechtold, *Tijdschrift voor Taalbeheersing*, 37:1, 33-78.

## **Compiling corpora of theatrical retranslations: the impact of non-verbal agency**

Jorge Braga Riera  
Universidad Complutense de Madrid

The notion of retranslation has recently been the subject of study and review in a large number of publications (Cadera and Walsh 2017, 2022; van Poucke 2017; van Poucke and Sanz Gallego 2019; Berk Albachten and Tahir Gürçaglar 2019, 2020). Oddly, the theoretical discussion regarding literary retranslation has largely neglected drama, and the existing analyses are both far and few between or do not consider “the actions to which the translation texts are linked, and the institutional relations of the production and use of theatrical repertoires for cultural goals” (Zurbach 2009: 280). This apparent lack of interest sits uneasily with the fact that theatrical texts are especially age-sensitive, requiring rewritings and constant modernisation (Van Poucke 2017: 97).

Research in this particular field should undoubtedly benefit from wide-range projects that go beyond particular case studies and take into account the extralinguistic traits specific to the performing arts. These, however, may impede the compilation of coherent, solid corpora that incorporate play texts at different stages in the (re)translation process (e.g. the literals, play script, dramaturgy, rehearsals, performance). Thus, this lecture attempts to shed light on the non-verbal agents (personal, performative, commercial, paratextual) that bear specifically on the different phases of the retranslation and reception of a foreign play, and the extent to which they impose multiple limitations on corpus and method. To this aim, three original Golden Age Spanish plays and their corresponding English (re)translations — staged in Britain and/or America in the twentieth and twenty-first centuries — will be taken as object of study. The chosen exemplification will demonstrate how anyone between the translation and its user, or an entity (an institution, for instance) can exert power (sociocultural or professional) in the exchange process, proving the impact of the social specificities of drama translation and the methodological issues resulting from a corpus-based approach.

### **References**

- Berk Albachten, Özlem and Sehnaz Tahir Gürçaglar, eds. 2020. *Multimodality in Translation Studies and Retranslation*. Special issue of *The Translator*, 26 (1).
- Cadera, Susanne M. and Andrew S. Walsh, eds. 2017. *Literary Retranslation in Context*. Oxford and New York: Peter Lang.
- Cadera, Susanne M. and Andrew S. Walsh, eds. 2022. *Retranslation and Reception: Studies in a European Context*. Leiden: Brill.
- Van Poucke, Piet. 2017. “Aging as a motive for literary retranslation.” *Translation and Interpreting Studies* 12 (1), 91-115.
- Van Poucke, Piet and Guillermo Sanz Gallego, eds. 2019. *Retranslation in Context*. Special issue of *Translation and Interpreting Studies*, 12 (1).
- Zurbach, Christine. 2009. “The theatre translator as a cultural agent: A case study.” In *Agents of Translation*, edited by John Milton and Paul Bandia, 280-299. Amsterdam: John Benjamins.

# FULL PAPERS

## **Parallel Corpora and Translation History: in search of ideology in translated texts via the use of anchor terms**

Olaia Andaluz-Pinedo<sup>1</sup> and Cristina Gómez Castro<sup>2</sup>

<sup>1</sup>Universidad de Cantabria

<sup>2</sup>Universidad de León

Research on translation history today is almost unthinkable without the aid that digital tools provide to the researcher: corpora and the several linguistic analyses that can be carried out within them are of great benefit when mapping different aspects of translation throughout a specific period (Gómez Castro 2021). With the idea of exemplifying this, in this contribution we take as our starting point the methodology followed by the studies carried out inside the TRACE research group. The principal objective of their investigation has been to trace the translator's behaviour in a period characterized by official censorship (Franco's dictatorship 1939-1975 and beyond) and, in doing so, establish the translation norms operative at the time to verify if they were in some way related to the prevailing ideology. The descriptive-comparative analyses of the texts chosen by each of the researchers in the project have been done mainly considering the most controversial topics for the dictatorial regime in charge by then, i.e., politics, religion, sexual morals and foul language. The task of locating in those textual corpora the different examples where to examine the translator's decisions and uncover the ideology underneath them has been an arduous one, especially considering the extension of the textual material at hand. Therefore, using a search tool which would help locate the prototypical instances necessary for the analysis has come to improve the methodology used so far. This search tool has been materialized in the form of a list of "anchor terms" similar to the "anchor words" used in parallel corpora: a list of terms which are the physical representation of the problematic topics for the authorities of Franco's regime has been compiled, i.e., those taboo areas have been translated into terms which allow us to search the aligned texts with the software available for that means. This search tool has already proved of use in the case of fiction texts (Gómez Castro 2009) and in the case of drama texts (Andaluz Pinedo 2022) which suffered from censorship during Franco's period. These analyses based on anchor terms have unveiled several censoring and self-censoring shifts regarding controversial topics in fiction and drama translations. Therefore, with this work we aim to (1) illustrate the procedure of using anchor terms to retrieve and analyse (self-)censorship through case studies of narrative and theatre translations and (2) to extend this analysis to the retranslations of those genres, thus proving it useful in identifying the ideological controversial aspects present in current renditions of the texts. Finally, we also seek to (3) broaden the fields covered by the list to the use of drugs, following Lobejón Santos' classification (2013). At the time of its creation, the establishment of such a list meant a step ahead in the study of translated discourse for this kind of research, and now it has proved its usefulness in mapping translation history beyond contextual and manual analysis.

**Keywords:** parallel corpus, translation history, ideology, anchor terms

## **References**

- Andaluz-Pinedo, Olaia. 2022. Traducciones teatrales (inglés-español) desde la censura franquista hasta el siglo XXI: análisis del corpus TEATRAD. Vitoria: Universidad del País Vasco. Unpublished PhD dissertation.
- Gómez Castro, Cristina. 2009. Traducción y censura de textos narrativos inglés-español en la España franquista y de transición TRACEni (1970–1978). León: Universidad de León. Unpublished PhD dissertation.
- Gómez Castro, Cristina. 2021. "The Use of Corpora and other Electronic Tools in Translation History", *The Routledge Handbook of Translation History*. London: Routledge: 38-53.
- Lobejón Santos, Sergio. 2013. Traducción y censura de textos poéticos inglés-español en España: TRACePi (1939-1983). León: Universidad de León. Unpublished PhD dissertation.

## Effective stance in English and Spanish newspaper opinion articles on illegal immigration and humanitarian crises: 'The Independent' versus 'El Mundo'

Marta Carretero and Lara Moratón-Gutiérrez  
Universidad Complutense de Madrid

This paper presents a quantitative analysis of expressions of effective stance in English and Spanish opinion articles from two newspapers of liberal ideology, 'The Independent' and 'El Mundo', on illegal immigration and humanitarian crises involving refugees. Effective stance concerns the speaker/writer's striving for control concerning the realization of events and how they aim at determining or influencing the course of reality (Langacker 2009, 2013; Marín-Arrese 2013, 2021a, 2021b). Effective stance is divided into subcategories according to factors such as orientation towards self or others, degree of control on the realization of the event and agreement with social norms. For this research, the subcategories are listed as follows:

- Intentionality, which concerns intention and commitment to perform actions (e.g. some occurrences of English WILL and BE GOING TO, and Spanish future tense and IR A);
- Directivity, i.e. the speaker / writer's decision with respect to the event (e.g. imperative mood or clauses with verbs such as ORDER, INSIST or SUGGEST and Spanish equivalents);
- Volitionality, which concerns expressions of willingness or inclination that some event or situation should occur (e.g. the verbs WANT, HOPE, EXPECT and WISH in the first person, and Spanish equivalents);
- Normativity, which pertains to social desirability (e.g. 'it is + ADJECTIVE + that/to + INFINITIVE clauses' with adjectives such as 'essential', 'important', 'mandatory', 'necessary'... and Spanish equivalents);
- Deonticity, which covers the expression of obligation, permission and prohibition (e.g. most of the occurrences of 'must', 'should', 'ought to' and some cases of 'can', and Spanish PODER and DEBER (de) + INFINITIVE);
- Potentiality, which concerns the capability or feasibility of the realisation of events (e.g. some occurrences of 'can' in English and PODER + INFINITIVE in Spanish).

Drawing on previous related research (Mora-López and Ferrer-Navas 2023; Carretero and Domínguez-Romero 2023), the paper takes as point of departure the following hypotheses:

- The expression of effective stance will be more frequent in English than in Spanish.
- The subcategories of effective stance will display distributional differences in the two languages between the subcorpora of immigration and humanitarian crises, due to factors such as perception of ethnic distance, perception of 'good' and 'bad' sides or feeling of guilt.

This research is geared at (dis)confirming these hypotheses, considering the limitations inherent to the data analyzed. The corpus encompasses a total of 120,000 words, with each newspaper contributing 30,000 words in articles pertaining to illegal immigration and 30,000 words relating to humanitarian crises. This last part is divided into three sets of 10,000 words, each on a concrete humanitarian crisis involving refugees: Syria (2015), Afghanistan (2021), and Ukraine (2022 onwards). For the annotation of effective stance in the corpus, an automatic search of frequent expressions will be combined with a manual analysis aimed at detecting other expressions with effective stance meanings through reading the articles. The quantitative results derived from the annotation will reveal crosslinguistic and cross-topic distributional differences in the expression of effective stance, thus shedding light on the motivations leading journalists' choices as strategies for striving for control in discourse.

**Keywords:** effective stance, newspaper opinion articles, deontic modality, dynamic modality, directives, illegal immigration, refugees

### References

- Carretero, M. and Domínguez-Romero, E. 2023. Effective stance in conservative newspaper opinion articles about immigrants and refugees. Paper presented at the 2º Congreso Internacional de Lingüística Digital (CILIDI), University of Alicante, 4-5 May 2023.
- Langacker, R. W. 2009. *Investigations in Cognitive Grammar*. Berlin: Mouton de Gruyter.

- Langacker, R. W., 2013. Modals: striving for control. In: Marín-Arrese, J.I., Carretero, M., Arús, J., van der Auwera, J. (Eds.), *English Modality: Core, Periphery and Evidentiality*. Berlin: Mouton de Gruyter. 3-55.
- Marín-Arrese, J. I., 2013. Stancetaking and inter/subjectivity in the Iraq inquiry: Blair vs. Brown. In: Marín-Arrese, J., Carretero, M., Arús, J., van der Auwera, J. (Eds.), *English Modality: Core, Periphery and Evidentiality*. Berlin: Mouton de Gruyter. 411-445.
- Marín-Arrese, J. I. 2021a. Stance, emotion and persuasion: Terrorism and the Press. *Journal of Pragmatics* 177, 135-148.
- Marín-Arrese, J.I. 2021b. Winds of war: Epistemic and effective control in political discourse. *Cultura, lenguaje y representación / Culture, language and representation* xxvi, 289-307.
- Mora-López, N. and Ferrer-Navas, S. 2023. An English-Spanish contrastive analysis of effective stance in newspaper and political discourse. In J.I. Marín-Arrese, L. Hidalgo-Downing and J.R. Zamorano-Mansilla (eds.), *Stance, Inter/Subjectivity and Identity in Discourse*. Bern: Peter Lang. 57-74.

## Translation errors and (in)formality in English-Spanish non-fiction: The case of subject pronouns *tú* and *usted*

Sara Chamosa<sup>1</sup> and Camino Gutiérrez Lanza<sup>2</sup>

<sup>1</sup>Universidad del País Vasco

<sup>2</sup>Universidad de León

Translation as a third code (Frawley, 1984) has recurrently been viewed as the product of interference and standardization (Toury, 1995/2012). However, there are few studies on their treatment in quality assessment. Research mostly focuses on the formal aspects of errors, linking them to a widespread negative view of interference (e.g., Presada & Badea, 2014). We believe, therefore, that errors should be redefined as misapplied techniques which fail to meet the parameters of the target text and culture (Rabadán & Fernández Nistal, 2002; Rabadán & Gutiérrez-Lanza, 2020). Thus, it would be necessary to study anchor phenomena (Rabadán, 2010), which ACTRES members have identified by analyzing translations. They have discovered over and underuse trends, determined the impact of anchors on contextual parameters (De Beaugrande & Dressler, 1981), and how these can aid translation quality assessment and post-editing tool development. Since their investigations into En-Es subject personal pronoun use mainly focus on fiction (Rabadán et al., 2007; Rabadán & Ramón, 2008; Ramón & Gutiérrez-Lanza, 2018), our paper will examine subject personal pronouns in En-Es non-fiction translations, namely second person singular informal *tú* and formal *usted*, both equivalents of English *you*. We aim at: i) determining over and underuse tendencies of translated pronouns, ii) analyzing errors and their impact on the target text, and iii) contrasting trends and effects in informal and formal texts.

With this in mind, we have carried out a comparative study with data extracted from the non-fiction subcorpora of CETRI (Corpus del Español Traducido del Inglés) and CORPES XXI (Corpus del Español del Siglo XXI). Data were compiled using a two-stage sampling process. Pronoun instances were categorized using pragmatic cross-linguistic labeling: neutral, (non-)optional emphasis, contrastive, formulaic, narrative discourse marker, and generic role (Ramón & Gutiérrez-Lanza, 2018). Data were processed through hypothesis testing to assess whether subject personal pronouns are used similarly or not in translated and non-translated non-fiction. Moreover, translated pronoun concordances were qualitatively analyzed through error (Rabadán & Gutiérrez-Lanza, 2020) and technique (Molina & Hurtado Albir, 2002) classifications to see how they occur in translations. Their impact was examined by focusing on textuality parameters (Rabadán & Fernández Nistal, 2002; Rabadán & Gutiérrez-Lanza, 2020).

Quantitative findings have unveiled significant differences for the following pronoun functions. Contrastive and generic *tú* are overused, while non-optional emphasis and narrative discourse marker *tú* are underused. *Usted* overuses optional emphasis and underuses its non-optional emphasis, narrative discourse marker, and formulaic functions.

Qualitative results have revealed that overused *tú* produces unidiomatic and less intelligible discourse, negatively affecting acceptability, intentionality, situationality, and cohesion. The overuse of *usted* favors redundancy, negatively affecting acceptability, informativity and situationality. In contrast, the underuse of *tú* reduces cohesion, undermining acceptability and situationality; while underused *usted*, follows the same pattern as overused *tú*.

In conclusion, translation quality is the determinant factor of interference. To make pronoun use more acceptable in translations, special attention ought to be paid to revision, where tools such as error checklists (Rabadán & Gutiérrez-Lanza, 2020) or automated quality assessment tools (e.g., PETRA 1.0©) become indispensable.

**Keywords:** translation error, interference, English-Spanish subject personal pronouns, informal-formal register.

### References

ACTRES (2013). PETRA 1.0©. A semi-automated aid for the evaluation of the grammatical quality of English-into-Spanish translations. <https://actres.unileon.es/wp/petra-1-0-evaluation-of-translations/>



- De Beaugrande, R., & Dressler, W. (1981). *Introduction to Text Linguistics*. Longman.
- Frawley, W. (1984). Prolegomenon to a Theory of Translation. In W. Frawley (Ed.), *Translation: Literary, Linguistic, and Philosophical Perspectives* (pp. 159–178). Associated University Presses.
- Molina, L., & Hurtado Albir, A. (2002). Translation techniques revisited: A dynamic and functionalist approach. *Meta*, 47(4), 498–512. <https://doi.org/10.7202/008033ar>
- Presada, D., & Badea, M. (2014). The effectiveness of error analysis in translation classes. A pilot study. *Porta Linguarum*, 22, 49–59.
- Rabadán, R. (2010). Applied Translation Studies. In Y. Gambier & L. van Doorslaer (Eds.), *Handbook of Translation Studies* (pp. 7–11). John Benjamins.
- Rabadán, Rosa., & Fernández Nistal, P. (2002). *La traducción inglés-español: fundamentos, herramientas, aplicaciones*. Universidad de León & ITBYTE. <http://hdl.handle.net/10612/5868>
- Rabadán, R., Gutiérrez, C., & Ramón, N. (2007, September 5). Exploring Translation Research Applicability: Description for Assessment (ACTRES/TRACE) [Conference contribution]. 5th International EST Congress “Why Translation Studies Matters”, University of Ljubljana, Slovenia. <http://hdl.handle.net/10612/9157>
- Rabadán, R., & Gutiérrez-Lanza, C. (2020). Developing awareness of interference errors in translation: An English-Spanish pilot study in popular science and audiovisual transcripts. *Lingue e Linguaggi*, 40, 379–404. <https://doi.org/10.1285/i22390359v40p379>
- Rabadán, R., & Ramón, N. (2008). Teaching English-Spanish Contrastive Analysis Through Translation. In F. Boers, J. Darquennes, K. Kerremans, & R. Temmerman (Eds.), *Multilingualism and Applied Comparative Linguistics* (Vol. 2, pp. 278–301). Cambridge Scholars Publishing.
- Ramón, N., & Gutiérrez-Lanza, C. (2018). Translation description for assessment and post-editing: The case of personal pronouns in translated Spanish. *Target*, 30(1), 112–136. <https://doi.org/10.1075/target.15098.ram>
- Toury, G. (1995/2012). *Descriptive Translation Studies - and beyond*. John Benjamins.

## **The Role of Epistemic Stance in Shaping Public Opinion on Refugees and Immigration: Evidence from Conservative Newspaper Opinion Articles in English and Spanish**

Elena Domínguez Romero and Juan Rafael Zamorano Mansilla  
Universidad Complutense de Madrid

This paper aims to examine the type of epistemic stance employed in discussing refugee crises and illegal immigration within two conservative newspapers from the United Kingdom and Spain. To accomplish this, we compiled a corpus of 120,000 words consisting of opinion articles from *The Telegraph* and *El Confidencial*. Among these, 60,000 words focused on recent refugee crises (Syria in 2015, Afghanistan in 2021, and Ukraine in 2022), while 60,000 words discussed illegal immigration during comparable time periods.

Epistemic stance, which refers to how individuals express their knowledge and certainty when conveying information, can reveal communicative purposes such as attitude display, evaluation, persuasion, alignment, or distancing in journalistic discourse (Marín-Arrese 2015, 2016, 2017a, 2017b, 2021; Carretero et al., 2017; Domínguez Romero and Martín de la Rosa 2023). In this paper, three primary categories are explored: evidentiality, epistemic modality, and factivity. Evidentiality is the indication of the source or type of evidence for a proposition, such as direct observation, hearsay, or inference (Willett, 1988; Aikhenvald, 2004; Wiemer & Stathi, 2010; Boye, 2012, among others), while epistemic modality assesses the likelihood of a proposition's truth or falsehood (Nuyts, 2001; Carretero & Zamorano-Mansilla, 2013). Factivity refers to the property of expressions that presuppose the truth of their complement clauses (Kiparsky & Kiparsky, 1970).

The results reveal distributional differences in the use of epistemic expressions across languages, thus demonstrating that English is more epistemic than Spanish in this specific context. However, considering the ideological proximity of the two newspapers under analysis, similar distributions of the epistemic expressions used to articulate stances towards refugees and illegal immigrants in newspaper opinion discourse are found in both languages.

**Keywords:** epistemic stance, refugee crises, illegal immigration, journalistic discourse, evidentiality, epistemic modality, factivity

### **References**

- Aikhenvald, A. Y. (2004). *Evidentiality*. Oxford: OUP.
- Boye, K. (2012). *Epistemic Meaning: A Crosslinguistic and Functional-cognitive Study*. (Empirical Approaches to Language Typology 43). Berlin: De Gruyter.
- Carretero, M., J.I. Marín-Arrese and J. Lavid-López (2017). Adverbs as evidentials: An English-Spanish contrastive analysis of twelve adverbs in spoken and newspaper discourse. *Kalbotyra* 70, 32–59.
- Carretero, M., and Zamorano-Mansilla, J. R. (2013). Annotating English adverbials for the categories of epistemic modality and evidentiality. In J. I. Marín-Arrese, M. Carretero, J. Arús Hita, and J. van der Auwera (eds.), *English Modality: Core, Periphery and Evidentiality*, pp. 317–355. Berlin: Mouton de Gruyter.
- Domínguez Romero, E. and Martín de la Rosa, V. (2023). Epistemic Stance in Opinion Newspaper Articles and Political Speeches: An English/Spanish Contrastive Approach. In Marín Arrese, J. I., Hidalgo-Downing, L. and Zamorano, J.R. (eds.), *Stance, Inter/Subjectivity and Identity in Discourse*. Frankfurt: Peter Lang.
- Kiparsky, P. and Kiparsky, C. (1970). Fact. In M. Bierwisch and K. E. Heidolph (eds). *Progress in Linguistics*, pp. 143–173. The Hague: Mouton.
- Marín-Arrese, J. I. (2015). 'Epistemicity and stance: A cross-linguistic study of epistemic stance strategies in journalistic discourse in English and Spanish'. *Discourse Studies*, 17 (2), 210–225.
- Marín-Arrese, J. I. (2016). 'Epistemicidad y posicionamiento discursivo: Un estudio interlingüístico de la evidencialidad en el discurso periodístico en castellano y en inglés'. In R. González Ruiz, D. Izquierdo Alegría, and O. Loureda Lamas (eds.), *La Evidencialidad en español: Teoría y descripción*, pp. 329–350. Iberoamericana, Madrid and Vervuert: Frankfurt am Main.

- Marín-Arrese, J. I. (2017a). 'Stancetaking and Inter/Subjectivity in journalistic discourse: The Engagement system revisited'. In R. Breeze and I. Olza (eds.), *Evaluation in Media Discourse: European perspectives*, pp. 21–48. Bern: Peter Lang.
- Marín-Arrese, J. I. (2017b). 'Multifunctionality of evidential expressions in discourse domains and genres: Evidence from cross-linguistic case studies'. In J. I. Marín-Arrese, G. Hassler and M. Carretero (eds.), *Evidentiality Revisited: Cognitive Grammar, Functional and Discourse-pragmatic perspectives*, pp. 195–223. Amsterdam/Philadelphia: Benjamins.
- Marín-Arrese, J. I. (2021). 'Stance, emotion and persuasion: Terrorism and the Press'. *Journal of Pragmatics* 177, 135–148.
- Nuyts, J. (2001). *Epistemic Modality, Language and Conceptualization: A Cognitive-pragmatic Perspective*. Amsterdam: John Benjamins.
- Wiemer, B. & Stathi, K. (2010). The database of evidential markers in European languages. A bird's eye view of the conception of the database (the template and problems hidden beneath it). *STUF* 63(4), 275–289.
- Willett, T. (1988). 'A Cross-Linguistic Survey of the Grammaticalization of Evidentiality'. *Studies in Language* 12(1), 51–97.

## How to deal with Part-of-Speech Tagging of Bilingual Parallel Corpora

Irene Doval Reixa

Universidad de Santiago de Compostela

(Parallel) Corpora are created for a range of purposes, but their potential applications depend, in practice, on their annotation layers. One common type of annotation is part-of speech tagging (pos tagging), i.e. the addition of tags, indicating the word class to which words in a corpus belong, as well as lemmatization, the grouping of the inflected forms of the same word.

Pos-tagging and lemmatization enable searching for specific phenomena that cannot be retrieved by the simple string search. Although they are common tasks in the creation of a corpus, when dealing with multilingual data, in which different languages come into play, these tasks involve additional challenges.

In this paper, it is reviewed some challenges we faced on pos tagging two bilingual parallel corpora, the PaGeS Corpus ([www.corpuspages.eu](http://www.corpuspages.eu)) and the PaEnS Corpus ([www.corpuspaens.eu](http://www.corpuspaens.eu)), that include three languages: Spanish, English and German. Furthermore, the steps to build a consistent annotation model also valid for other language pairs with Spanish as the pivot language, are shown.

Before pos tagging, tokenization -token boundary detection- must be carried out. Here the issues are often language-specific, driven by the orthographic conventions of a particular language. In the case of German, English and Spanish the main difficulties that tokenization runs into are multiwords, contractions and mergers. Therefore, it has been decided not to use the built-in tokenizers of the taggers, but rather to carry it out beforehand and review the output.

Since the performance of a given tagger differs from one language to another, different taggers have to be applied. In our case Freeling (Padró 2012) for Spanish and Treetagger (Schmidt 1995) or RNNTagger (Schmidt 2019) for German and English were used.

On the other hand, Spanish, English and German languages each have specific grammatical features and terminology. Accordingly, each language uses annotation schemes that differ greatly between themselves in scope and detail. But to search efficiently for lexical categories through large amounts of data in two languages, corpus user need a tagging scheme with a unified format and conventions for representing the tags. This made necessary to map the specific tags of the taggers to the Universal Tagset format (Petrov et al., 2012), in order to facilitate the interoperability and corpus search.

**Keywords:** tokenization, POS tagging, multilingual corpora, tagset, universal POS tags

### References

- Schmid, Helmut. 1995. Improvements in Part-of-Speech Tagging with an Application to German. Proceedings of the ACL SIGDAT-Workshop. Dublin, Ireland. <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger2.pdf>.
- Petrov, Slav, Dipanjan Das and Ryan McDonald. 2012. A universal part-of-speech tagset. In Eighth International Conference on Language Resources and Evaluation (LREC 2012): 2089-2096.
- Padró, Lluís and Stanilovsky, Evgeny. 2012. FreeLing 3.0: Towards Wider Multilinguality. Proceedings of the Language Resources and Evaluation Conference (LREC 2012) ELRA. Istanbul, Turkey. May, 2012. <http://nlp.lsi.upc.edu/publications/papers/padro12.pdf>
- Schmid, Helmut. 2019. Deep Learning-Based Morphological Taggers and Lemmatizers for Annotating Historical Texts, DATeCH, May 2019, Brussels, Belgium.

## **La postura epistémica en gallego y español en dos corpus periodísticos ideológicamente opuestos**

Mercedes González Vázquez  
Universidad de Vigo

En este trabajo partimos de la pregunta de si existen diferencias discursivas entre la expresión de dos ideologías políticas opuestas (derecha e izquierda) en dos lenguas diferentes (el castellano y el gallego) en artículos periodísticos sobre la migración. Para responder a dicha pregunta, realizamos un análisis contrastivo español-gallego basándonos en dos corpus creados a partir cuatro periódicos gallegos: dos periódicos de izquierdas en lengua gallega y dos periódicos gallegos de derecha en español. Los dos corpus suman en total 80.000 palabras, aproximadamente 40.000 en cada lengua, y se hallan compuestos por artículos, principalmente de opinión, sobre la migración derivada de los conflictos bélicos de Siria, Afganistán y Ucrania.

Tras el análisis de los dos corpus, encontramos como resultado una diferencia en la frecuencia de uso y el modo de empleo de las expresiones de postura epistémica. La postura epistémica se define como la manera en que las distintas posiciones enunciativas del hablante/escritor muestran su actitud y valoración respecto al contenido comunicado (Heritage & Raymond 2005, Marín-Arrese 2021). Una de las manifestaciones lingüísticas de la postura epistémica es la que se plasma a través del empleo de la evidencialidad y de la modalidad epistémica. La evidencialidad indica la fuente de información sobre la que se basa el hablante al enunciar algo, al tiempo que proporciona apoyo justificativo a la proposición (Willett 1988, Boye 2012, Mushin 2013). Por su parte, la modalidad epistémica indica el grado de veracidad y de confianza del hablante en que la proposición enunciada sea verdadera, probable o improbable.

Al contrastar los dos corpus, hemos encontrado discrepancias en la presentación de la postura epistémica por parte del escritor en el uso de expresiones evidenciales y modales epistémicas. Entre las expresiones evidenciales diferenciamos los marcadores inferenciales (parece + infinitivo, parece que + infinitivo) y los reportativos (al parecer, aparentemente, por lo visto, parece ser que...), mientras que en el uso de las expresiones modales discriminamos las expresiones verbales (puede que, debe de, ha de) de las adjetivas y adverbios (es posible, es probable, es seguro, probablemente, quizás, etc.)

El análisis de los dos corpus se ha realizado mediante un sistema de anotación manual de los marcadores evidenciales y las expresiones modales epistémicas. Tras dicha anotación de los marcadores, se realizó un análisis cualitativo de los datos obtenidos. Como resultado de dicho estudio, se comprueban las diferencias tanto en la frecuencia de uso como en el contexto pragmático de empleo de la postura epistémica en las dos lenguas e ideologías.

**Keywords:** Epistemic stance, Galician and Spanish, Migration discourse, Evidentiality, Modality

### **Referencias**

- Boye, Kasper (2012): *Epistemic Meaning. A Crosslinguistic and Functional-Cognitive Study*. Berlin / Boston: Mouton De Gruyter.
- Heritage, John & Geoffrey Raymond (2005): "The Terms of Agreement: Indexing Epistemic Authority and Subordination in Talk-in-Interaction". *Social Psychology Quarterly* 68(1), 15–38.
- Marín-Arrese, Juana 2021. Winds of war: Epistemic and effective control in political discourse. *Cultura, lenguaje y representación / Culture, language and representation* xxvi, 289-307.
- Mushin, Ilana (2013): "Making knowledge visible in discourse: Implications for the study of linguistic evidentiality". *Discourse Studies* 15(5), 627–645.
- Willett, Tomas (1988): "A crosslinguistic survey of the grammaticization of evidentiality". *Studies in Language* 12 (1), 51–97.

## **Translating opinion: interactional metadiscourse in a learner corpus**

Marlén Izquierdo

Universidad del País Vasco

In today's globalized world, translation and journalism are two highly interconnected means of communication and sources of information that flow in a constant way, cross-linguistically, from culture to culture. Consequently, exposing translator trainees to journalistic discourse would satisfactorily meet a real professional need of their future career. Although modestly, over the past twenty years translation studies have been paying more and more attention to journalistic translation, a specialized way of translation that may take shape in many different text types as a result of various translational practices pursuing different communicative purposes (Hernández Guerrero 2009). Major research interests have focused on the description of translation as product, focusing on the linguistic quality of the resulting text, or the way the new text contributes to shaping the target language (TL), (ibid). Among the journalistic texts that get translated the most we find opinion articles (Hernández Guerrero 2009; Hamad Alqahtani 2019), which are argumentative types of texts where the personal involvement of the writer is essential to achieve its primarily persuasive function (van Belle 2011). From a discourse-analytic point of view, writers resort to interactional metadiscourse (MTD) to show their stance and attitude towards the text as well as to engage with readers (Hyland 2005; Hyland and Tse 2004), which is key to persuade them with the writer's point of view on the topic. Studies on the role of MTD in newspaper opinion genres, either monolingual or cross-linguistic, are relatively scarce though (Hamad Alqahtani 2019), and so is research on learner translation of news discourse (Pisanski and Zlatnar Moe 2016). The aim of this study is to describe the translational options of interactional MTD markers (MTDMKs) of an English (EN) opinion article rendered into Spanish (ES) by translator trainees. Parallel data for the study is taken from an ad hoc multiple translation corpus (Espunya 2014) belonging to the MUST (Multilingual Student Translation) corpus (Granger and Lefer 2020). The corpus-based analysis was conducted relying on previous EN-ES contrastive research (Dafouz-Milne 2008) as well as the notion of translation strategy or shift from descriptive translation studies (Hamad Alqahtani 2019). Drawing from Dafouz-Milne's classification of interactional MTD in EN-ES opinion articles (2008), I first identified in the ST all MTDMKs that conveyed both writer's stance and engagement. The ST featured a total of nineteen markers representative of four metadiscursive functions, namely, 9 attitude markers, 5 attribution resources, 3 certainty items and 2 hedging devices. I collected sixteen translational options per MTDMKs, making a sample of 304 parallel occurrences, among which I looked for recurrent patterns per MTD function. Findings reveal a tendency to use an equivalent resource that literally transfers the writer's intention to the translated text, which has been observed in 48% of the translational options. Shifts break down into modification (45%) and omission (7%); while the former is present in the translation of all functions, the latter, by contrast, is particularly frequent in the translation of attitude. Given that not all the corresponding items fulfil the same metadiscoursal function identified in the ST, pre-translation tasks tailored to developing metadiscoursal awareness may be pertinent for translator trainees to understand the importance of keeping up with the intended pragmatic effect of argumentative genres, and in particular opinion ones.

**Keywords:** learner translations, interactional metadiscourse, opinion article, multiple translation corpus

### **References**

- Dafouz-Milne, Emma. 2008. The pragmatic role of textual and interpersonal metadiscourse markers in the construction and attainment of persuasion: a cross-linguistic study of newspaper. *Journal of Pragmatics* 40(1), 95-113.
- Espunya, Anna. 2014. The UPF learner translation corpus as a resource for translator training. *Lang Resources & Evaluation* 48, 33-43.
- Granger, Sylviane, & Lefer, Marie-Aude. 2020. The Multilingual Student Translation corpus: a resource for translation teaching and research. *Language Resources and Evaluation* 54, 1183-1199.

- Hamad Alqahtani, Laila. 2019. Shifts in the Translation of Interactional Metadiscourse Markers in Arabic and English Opinion Articles. Unpublished PhD Thesis. Norwich: University of East Anglia.
- Hernández Guerrero, M<sup>a</sup> José. 2009. Traducción y periodismo. Bern: Peter Lang
- Hyland, Ken. 2005. Metadiscourse. Exploring interaction in writing. London: Continuum
- Hyland, Ken, Tse, Polly. 2004. Metadiscourse in Academic Writing: A Reappraisal. *Applied Linguistics* 25(2): 156-177.
- Pisanski Peterlin, Agnes, Zlatnar Moe, Marija (2016). Translating hedging devices in news discourse. *Journal of Pragmatics* 102, 1-12.
- van Belle, Hilde. 2011. The correlation between style and argumentation in newspaper columns. In van Haften, Ton, Jansen, Henrike, De Jong, Jaap, Koetsenruijter, Willem (Eds.) *Bending Opinion: Essays on persuasion in the public domain*. 185-199. Leiden University Press.

## **Using comparable corpora to foster economic and cultural internationalization: Corpus-based digital tools to support professional specialized communication in English**

M<sup>a</sup> Belén López Arroyo and Lucía Sanz Valdivieso  
Universidad de Valladolid

A 21st century citizen requires constant learning to adapt their knowledge, skills and behaviors to the many changes present not only in education, but in technological, cultural and social contexts (Loveless and Williamson 2012). In today's globalized and ever-changing society, efficient lifelong learners who use English for Specific Purposes (ESP) professionally are not exempt from the need of being up to date in their respective field. In this context, and regardless of area of expertise, accurate and state-of-the-art terminology is the necessary first step to achieve successful communication (Bourn 2018; Conole and Paredes 2018). This paper is based on the use and understanding of ESP in an autonomous online lifelong learning environment oriented to Spanish professionals. We focus on the wine and olive oil industry since they are internationally glooming sectors in Spain; Spain is the biggest wine exporter in terms of volume and the third in terms of value, and, regarding olive oil, Spain leads the world market, making up 45% of the world's production (Ministerio de Agricultura, Pesca y Alimentación). In spite of these figures and the need of communication in these fields in English as the international lingua franca, there are not any specialized tools to help Spanish professionals succeed communicatively in this context, that we are aware of. Considering these aspects, we have developed a series of specialized ICT tools so that these professionals have access not only to specific receptive tools such as terminology and real-life expressions (García Sánchez and Luján García 2020, 196) but also to productive tools that allow them to learn their use and writing in context. To develop these tools, comparable corpora of wine and olive oil tasting notes (TNs) were compiled using pragmatic criteria. The corpora include 760 wine and 620 olive oil TNs in English (118,846 and 20,855 words respectively and 750 wine and 330 olive oil TNs in Spanish (118,760 and 23,445 in Spanish). These corpora were annotated grammatically, semantically and rhetorically to be able to produce the resulting tools: first, an interactive bilingual Spanish-English phraseological dictionary which includes the equivalents in English, an example of use, the field and rhetorical composition strategies, i.e the move(s) (according to Swales 1990 methodology) in which the phraseological units are used; secondly, a practical suite of vocabulary exercises with different levels of complexity; finally, a semi-automatic writing tool, part of a larger project, was integrated. These tools, based on real-life instances of language use within their specialized discourse community, allow professionals to understand and practice their competence in these ESPs autonomously and ubiquitously.

### **References**

- Bourn, Douglas. 2018. *Understanding Global Skills for 21st Century Professions*. Cham: Palgrave Macmillan.
- Conole, Gráinne, and Pascual Pérez Paredes. "An Analysis of Adult Language Learning in Informal Settings and the Role of Mobile Learning." In *Mobile and Ubiquitous Learning*, edited by Shengquan Yu, Mohamed Ally, Avgoustos Tsinakos, 45–57. Singapore: Springer.
- García-Sánchez, Soraya, and Carmen Luján-García. "Ubiquitous Knowledge and Experiences to Foster EFL Learning Affordances." *Computer Assisted Language Learning* 29 (7): 1169–80.
- Loveless, Avril, and Ben Williamson. 2013. *Learning Identities in a Digital Age: Rethinking Creativity, Education and Technology*. Abingdon, New York: Routledge.
- Ministerio de Agricultura, Pesca y Alimentación, Gobierno de España. 2022. *Producciones Agrícolas. Aceite de oliva*. <https://www.mapa.gob.es/es/agricultura/temas/producciones-agricolas/aceite-oliva-y-aceituna-mesa/aceite.aspx>
- Ministerio de Agricultura, Pesca y Alimentación, Gobierno de España. 2022. *Producciones Agrícolas. Vitivinicultura*. <https://www.mapa.gob.es/es/agricultura/temas/producciones-agricolas/vitivinicultura/>
- Swales, John. 1990. *Genre Analysis*. Cambridge: Cambridge University Press.



## **Effective Stance in Political Discourse on Migration and Refugees: A contrastive corpus study (English and Spanish)**

Juana I. Marin-Arrese

Universidad Complutense de Madrid

Stance resources may be said to index speaker/writers' epistemic, evaluative or effective positionings in the discourse (Biber & Finegan 1989; DuBois 2007; Englebretson 2007, *inter alia*). The paper draws on the Epistemic-Effective Stance (EES) framework (cf., Marin-Arrese 2011, 2021), and focuses on effective stance as the way we construct subject positions and strive for control of relations at the level of reality (cf., Langacker 2009, 2013), and for persuasive coercion with respect to plans of action and the realization of events.

From the interdisciplinary perspective of Critical Discourse Studies, this paper explores ideological polarisation and stance of politicians and political parties in the discourse on migration and refugees (Bennett 2019; Breeze 2011; Chilton 2017; Reyes 2011; Richardson et al. 2016; Van Dijk 1993). The paper presents results of a contrastive corpus study on the deployment of effective stance expressions in political speeches (UK: Conservative vs. Labour; Spain: PP vs. PSOE), and party manifestos in recent elections (UK: 2015, 2017, 2019; Spain: 2015, 2016, 2019). Effective stance resources include a variety of lexico-grammatical elements which may be grouped into the categories of deonticity, directivity, normativity, intentionality and potentiality. The tokens of effective stance were manually annotated according to these categories, and the retrieval and quantification of the data was carried out using Monoconc. For the analysis of the significance of the quantitative results, a set of statistical tests were used: Chi-square test, SIGIL Corpus Frequency Test Wizard, and UCREL's Log likelihood wizard.

The paper addresses the following research issues: (a) variation in the deployment of effective stance resources in relation to ideological leanings of the political parties, and languages; (b) variation in the use of effective expressions in two genres of political discourse, political speeches vs. party manifestos. Preliminary results point to significant differences in the deployment of effective stance expressions in political speeches across languages, with Spanish favouring the categories of deonticity and directivity. With respect to the ideological variable, in both languages we find a preference for the category of intentionality in the left-wing parties, and for potentiality in the right-wing parties.

Political discourse appears to be characterised by a distinct preference for effective stance, in particular expressions of deontic modality (DM), intentionality (INT), and potentiality (POT). It will be argued that these resources serve the strategic functions of (de)legitimation and coercion, that is, support for the validity of their ideological agenda and persuasive intent regarding proposed plans of action and preferred events and social action (Chilton 2004; Marin-Arrese 2011, 2021, 2023).

**Keywords:** stance, effectivity, legitimisation strategies, political discourse, migration, refugees

### **References**

- Bennett, S. (2019) 'Crisis' as a discursive strategy in Brexit referendum campaigns. *Critical Discourse Studies* 16 (4): 449–464.
- Biber, D. and E. Finegan. (1989). Styles of stance in English: lexical and grammatical marking of evidentiality and affect. *Text* 9.1 (Special Issue on the Pragmatics of Affect): 93-124.
- Breeze, R. (2011). Variations on a theme: Party Manifesto discourses in the UK 2010 Election. *Culture, Language and Representation* 9 (9): 9–30.
- Chilton, P. (2004). *Analysing Political Discourse*. London: Routledge.
- Chilton, P. (2017). "The People" in Populist Discourse: Using Neuro-cognitive Linguistics to understand political meanings. *Journal of Language and Politics* 16 (4): 582–94.
- DuBois, J.W. (2007). The stance triangle. In: R. Englebretson (Ed.), *Stancetaking in discourse: Subjectivity, evaluation, interaction*. Amsterdam/Philadelphia: John Benjamins. 139-182.
- Englebretson, R. (2007). Introduction. In R. Englebretson (Ed.), *Stancetaking in Discourse*. Amsterdam: John Benjamins. 1-26.
- Langacker, R.W. (2009). *Investigations in Cognitive Grammar*. Berlin: Mouton de Gruyter.

- Langacker, R.W. (2013). Modals: Striving for Control. In: Marín-Arrese, J.I.; Carretero, M.; Arús, J. and van der Auwera, J. (Eds.) *English Modality: Core, Periphery and Evidentiality*. Berlin, Mouton de Gruyter. 3-55.
- Marín-Arrese, J.I. (2011). Effective vs. Epistemic Stance and Subjectivity in Political Discourse: Legitimising Strategies and Mystification of Responsibility. In: Hart, Christopher (ed.) *Critical Discourse Studies in Context and Cognition*. Amsterdam, John Benjamins. 193-223.
- Marín-Arrese, J.I. (2021). Winds of War: Epistemic and effective control in political discourse on war. *Cultura, Lenguaje y Representación/Culture, Language and Representation*, Vol. xxv: 1-19.
- Marín Arrese, J.I. (2023). Epistemic and Effective Stance: Legitimation Strategies and the Expression of Inter/Subjectivity in Discourse. In: J.I. Marín-Arrese, L. Hidalgo-Downing & J.R. Zamorano-Mansilla (eds), *Stance, Inter/Subjectivity and Identity in Discourse*. Bern: Peter Lang. 29-56.
- Reyes, A. (2011). Strategies of legitimization in political discourse: From words to actions. *Discourse & Society* 22 (6): 781–807
- Richardson, J.E., Krzyżanowski, M., Machin, D. and R. Wodak. (2016). *Advances in critical discourse studies*. New York: Routledge.
- van Dijk, T.A. (1993) *Elite Discourse and Racism*. London: Sage Publications.

## **Annotating a parallel corpus Old English-English with Universal Dependencies. From database to CoNLL-U format**

Javier Martín Arista

Universidad de La Rioja

Universal Dependencies (de Marneffe et al. 2021) is a model of morphological and syntactic annotation devised for the compilation of computerised data sets aimed to natural language processing (Nivre 2015) and to areas of applied linguistics like language acquisition and translation (MacDonald et al. 2013; Nivre 2016). The annotation framework of Universal Dependencies includes categories and relations. On the categorial part, universal part-of-speech tags, language-specific part-of-speech tags, universal morphological features and lemmas are distinguished, while, on the relational part, dependency heads are related to dependents through dependency relation labels (Nivre et al., 2016). The 2021 release of Universal Dependencies comprises 183 treebanks over 104 languages (Nivre et al., 2020). Against this background, this paper delves into the application of the Universal Dependencies framework to Old English, the historical stage of the English language spoken in England between approximately the 7th and the 12th centuries (CE). Such an application is based on a parallel corpus Old English-English comprised of 300,000 words in parallel texts aligned at word, sentence and fragment level (Martín Arista et al. 2021). Given that the parallel corpus has been implemented in database format, this paper addresses the question of how to migrate the data from the database and to export them to the format required by Universal Dependencies, namely CoNLL-U format, a revised version of the CoNLL-X format in which annotations are encoded in plain text files (UTF-8 with LF character as line break and end of file, as indicated at <https://universaldependencies.org/format.html>). The CoNLL-U format specifies ID (word index), FORM (word form), LEMMA (lemma of word form), UPOS (universal part-of-speech tag), XPOS (language specific part-of-speech tag), FEATS (morphological features), HEAD (head of the current word), DEPREL (universal dependency relation), DEPS (enhanced dependency) and MISC (any other annotation). The migration and exportation of the data raise issues of tokenisation (including the word-token distinction drawn in the Universal Dependencies framework), categorial tagging (the inventory of category tags of Universal Dependencies being more extensive than traditional ones) and morphological feature annotation (by means of explicit sets as posited by Universal Dependencies). The conclusion is reached that the tasks involved in the migration and the exportation of the data can be largely automated. It is also a conclusion of this paper that two specific fields should be added to the Universal Dependencies annotation scheme, to wit, a gloss field, given the historical character of the language; and a morphological relatedness field, in order to account for its associative lexicon.

**Keywords:** Parallel Corpus, Old English, Universal Dependencies, Database, CoNLL-U

### **References**

- de Marneffe, M., Manning, C., Nivre, J., & Zeman, D. (2021). Universal Dependencies. *Computational Linguistics*, 47 (2), 255–308.
- Martín Arista, Javier, Sara Domínguez Barragán, Laura García Fernández, Esaúl Ruíz Narbona, Roberto Torre Alonso & Raquel Veá Escarza (comp.). 2021. *ParCorOEv2*. An open access annotated parallel corpus Old English-English. Nerthus Project, Universidad de La Rioja, [www.nerthusproject.com](http://www.nerthusproject.com).
- McDonald, R., Nivre, J., Quirnbach-Brundage, Y., Goldberg, Y., Das, D., Ganchev, K., Hall, K., Petrov, S., Zhang, H., Täckström, O., Bedini, C., Bertomeu, N. & Lee, J. (2013). Universal Dependency Annotation for Multilingual Parsing. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (pp. 92–97).
- Nivre, J. (2015). Towards a universal grammar for natural language processing. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing* (pp. 3–16). New York: Springer.
- Nivre, J. (2016). Universal Dependencies: A Cross-Linguistic Perspective on Grammar and Lexicon. *Proceedings of the Workshop on Grammar and Lexicon: Interactions and Interfaces* (pp. 38-40).

Nivre, J., de Marneffe, M., Ginter, F., Hajič, J., Manning, C., Pyysalo, S., Schuster, S., Tyers, F., & Zeman, D. (2020). Universal Dependencies v2: An evergrowing multilingual treebank collection. Proceedings of the Twelfth International Conference on Language Resources and Evaluation (LREC 2020) (pp. 4027–4036).

## **Automatic parsing of a parallel corpus Old English-English with Universal Dependencies**

Javier Martín<sup>1</sup>, Ana Elvira Ojanguren López<sup>1</sup>, Sara Domínguez Barragán<sup>1</sup> and Mario Barcala<sup>2</sup>

<sup>1</sup>Universidad de La Rioja

<sup>2</sup>NLP GO

The aim of this paper is to assess the results of the the automatic parsing of a parallel corpus Old English-English with the morpho-syntactic framework of Universal Dependencies (Nivre et al. 2016). The parallel corpus under analysis is ParCorOE<sub>v2</sub>. An open access annotated parallel corpus Old English-English (Martín Arista et al. 2021). The task of automatic parsing comprises three training corpora, three configurations and three training procedures. Three training corpora, based on or related to ParCorOE<sub>v2</sub>, are tested in this task (1,000, 5,000, 10,000, and 20,000 words). The training methods are a SpaCy pipeline with default configuration, pre-training of tok2vec stage and a model of language based on transformers. The main result of the assessment is that larger training corpora turn out better performance in all the stages of the pipeline, especially in part-of-speech and dependency annotation. The main conclusion is that the performance could be improved by processing the Old English and the English parallel text.

**Keywords:** Parallel Corpus, Old English, Universal Dependencies, Automatic Parsing

### **References**

- Martín Arista, J., Domínguez Barragán, S., García Fernández, L., Ruíz Narbona, E., Torre Alonso, R., and Vea Escarza, R. (comp.). 2021. ParCorOE<sub>v2</sub>. An open access annotated parallel corpus Old English-English. Nerthus Project, Universidad de La Rioja, [www.nerthusproject.com](http://www.nerthusproject.com).
- Nivre, J., de Marneffe, M., Ginter, F., Goldberg, Y., Hajič, J., Manning, C., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R., & Zeman, D. 2016. Universal Dependencies v1: a multilingual treebank collection. Proceedings of the 10th International Conference on Language Resources and Evaluation, pages 1659-1666.

## **Representation of Syrian and Ukrainian refugees in the discourse of mainstream media: a contrastive corpus study**

Victoria Martín de la Rosa and Juana I. Marín-Arrese  
Universidad Complutense de Madrid

The aim of this paper is to study the strategies of representation of two distinct population groups within the category of refugees, Syrian and Ukrainian displaced people, in a corpus of opinion columns compiled from two British and two Spanish newspapers with left and right-wing ideologies (The Guardian and The Daily Telegraph, against El País and El Mundo). The timeframe will range from 2015-2016 for the Syrian group (Pew Research Center 2016), and 2022 for the Ukrainian group.

The paper draws on the conceptual frameworks of Critical Discourse Studies (Baker et al. 2008; KhosraviNik 2010; van Dijk 1993; Wodak & Meyer 2015) and Critical Metaphor Analysis (Hart 2008; Charteris-Black 2014; Musolff 2015). The study posits the following research questions: (i) What are the main strategies of representation used for both groups? (ii) Can we identify any significant differences in the representation of both groups of refugees across languages (English vs. Spanish media) and ideologies (right vs. left-wing)?

The study will proceed in a contrastive fashion while examining data qualitatively and quantitatively. The analysis will combine manual annotation and the use of software programmes for corpus analysis and comparison (ATLAS.ti and Wmatrix), and will focus on features of representation involving predicational strategies and metaphor scenarios.

It is expected that the most significant differences will not necessarily be related to the language or the ideology of the newspaper, but rather to the target group of refugees under discussion. This means that, as the Syrian refugees are geographically and culturally more distant than the Ukrainians, the former are likely to be portrayed more distinctly as ‘THEM’, and not as close to the host community, that is, ‘US’ (be it British or Spanish)(van Dijk 1998). The Syrian are thus more likely to be placed outside the margins as an undifferentiated group of people, and as a burden to the host community, whereas the latter are likely to be represented as individualized actors joining and contributing to the community.

**Keywords:** media discourse, refugees, representation, predicational strategies, metaphor scenarios

### **References**

- Baker, Paul, Gabrielatos, Costas, Khosravini, Majid, Krzyzanowski, Michal, McEnery, Tony & Ruth Wodak. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society* 19 (3): 273–306.
- Charteris-Black, Jonathan (2014). *Analysing political speeches: Rhetoric, discourse and metaphor*. Basingstoke: Palgrave Macmillan.
- Hart, Chris. (2008). Critical Discourse Analysis and metaphor: toward a theoretical framework. *Critical Discourse Studies* 5 (2): 91-106.
- KhosraviNik, Majid (2010) The representation of refugees, asylum seekers, and immigrants in British newspapers. *Journal of Language and Politics* 9 (1): 1-28.
- Musolff, Andreas (2015) Dehumanizing metaphors in UK immigrant debates in press- and online media. *Journal of Language Aggression and Conflict* 3 (1): 41-56.
- Pew Research Center (2016). Number of Refugees to Europe Surges to Record 1.3 Million in 2015. <<https://www.pewresearch.org>>
- van Dijk, Teun A. (1993) *Elite Discourse and Racism*. London: Sage Publications.
- van Dijk TA (1998) *Ideology: A Multidisciplinary Approach*. London: Sage.
- Wodak, Ruth & Michael Meyer (2015). *Critical Discourse Studies: History, agenda, theory and methodology*. In R. Wodak & M. Meyer (Eds.) *Methods of Critical Discourse Studies*, 3rd ed. London: Sage. 2-22.

## Effective stance in opinion articles: the Ukraine migrant crisis as a case study

Natalia Mora-López<sup>1</sup> and Alfonso Sánchez-Moya<sup>2</sup>

<sup>1</sup>Universidad Complutense de Madrid

<sup>2</sup>University of Massachusetts (USA)

Following Marín-Arrese (2021a), effective stance refers to the ways in which the speaker/writer carries out a stance act aimed at determining or influencing the course of reality itself, that is, the positioning of the speaker/writer with respect to the realisation of events. Effective stance is especially relevant in the media, since it may be a resource that helps the author to shape discourse in a way that increases its persuasive and compelling effect. Previous research has shown how it may serve to legitimise knowledge about events and plans, not only in newspaper but also in political discourse, and that its use may vary depending on political slant and language (Martín de la Rosa et al. 2020; Marín-Arrese, 2009, 2011a, 2011b, 2013, 2021a, 2021b). New social and political situations occur where authors from different sides and opinions intend to persuade their audience to see events as they do. This is the case of the Ukraine war from 2022 and how Ukrainian refugees are addressed in the media. With this in mind, this paper seeks out to explore linguistic traces of effective stance in newspaper discourse from a contrastive English-Spanish perspective. More specifically, this study dives into different strategies used to convey effective stance in opinion articles around the recent migrant crisis in Ukraine. The main research questions this research sheds light on are worded as follows: (RQ1) Which categories within effective stance are more salient in English and Spanish centre-left newspaper opinion articles about the migrant crisis in Ukraine? (RQ2) Which are the most frequent realisations of effective stance in these texts? (RQ3) Are there any differences in the use of effective stance in English and Spanish in these opinion articles? In the attempt to provide answers to these questions, this paper is based upon a corpus of circa 20,000 words consisting of opinion articles about the migrant crisis in Ukraine published in 2022 and 2023 in the online edition of the British newspaper *The Guardian* and the Spanish online newspaper *elDiario*. Both newspapers share their political slant, which can be classified as centre-left. The corpus was manually annotated following the classification put forward by Marín-Arrese (2021a), which includes the Effective stance categories Deonticity (related to the necessity of the realization of an event), Directivity (directive illocutionary force towards the realization of an event), Intentionality (personal inclination or volition regarding the realization of an event), Normativity (social desirability and appropriateness of the realization of an event) and Potentiality (ability, capacity or possibility of the realization of an event). Results show interesting qualitative and quantitative remarks on the preferred effective stance strategies employed in this newspaper genre and languages, drawing parallels with similar cross-linguistic explorations across newspaper genres (Mora-López & Ferrer-Navas, 2023).

**Keywords:** effective stance, opinion articles, migrant crisis, Ukraine

### References

- Marín-Arrese, J.I. (2009). Effective vs. epistemic stance and subjectivity/ intersubjectivity in political discourse: A case study. In A. Tsangalidis, R. Facchinetti (Eds.), *Studies on English modality. In honour of Frank R. Palmer*. Berlin: Peter Lang. pp. 23-52.
- Marín-Arrese, J.I. (2011a). Effective vs. Epistemic Stance and Subjectivity in Political Discourse: Legitimising Strategies and Mystification of Responsibility. In C. Hart (ed.), *Critical Discourse Studies in Context and Cognition*. Amsterdam: Benjamins. pp. 193–224.
- Marín Arrese, J.I. (2011b). Epistemic Legitimising Strategies, Commitment and Accountability in Discourse. *Discourse Studies* 13 (6): 789-797.
- Marín Arrese, J. (2013) Stancetaking and Inter/Subjectivity in the Iraq Inquiry: Blair vs. Brown. In J. Marín Arrese, M. Carretero, J. Arús & J. van der Auwera (eds.) *English Modality: Core, Periphery and Evidentiality*. Berlin: Mouton de Gruyter. Pp. 411-445.
- Marín-Arrese, J. I. (2021a). Stance, Emotion and Persuasion: Terrorism and the Press. *Journal of Pragmatics* 177: 135–148.
- Marín-Arrese, J.I. (2021b). Winds of war: Epistemic and effective control in political discourse. *Culture, language and representation XXVI*: 289-307.

- Martín de la Rosa, V., Domínguez Romero, E., Pérez-Blanco, M. & Marín Arrese, J.I. (2020) Epistemic and effective stance in political discourse: The European refugee crisis. In L. Berlin (ed.), *Positioning and Stance in Political Discourse*. Delaware (USA): Vernon Press. pp. 141-156.
- Mora-López, N. and Ferrer-Navas, S. (2023). An English-Spanish contrastive analysis of effective stance in newspaper and political discourse. In J.I. Marín-Arrese, L. Hidalgo-Downing and J.R. Zamorano-Mansilla (Eds.), *Stance, Inter/Subjectivity and Identity in Discourse*. Bern, Berlín: Peter Lang. pp. 57-74.



## **What corpora can tell about translation competence: a case study on a trade agreement**

Leticia Moreno Pérez

Universidad de Valladolid

Although a good command of translation competences in general is necessary to perform the translation of any type of document, each field of expertise has specific characteristics and standards that directly affect the required skills for a successful transfer. That is the case of business genres, whose idiosyncratic nature including hybridity at various levels, different degrees of term specialization, or persuasive intention (Alcaraz Varó, 2000; Román Mínguez, 2012; Moreno-Pérez, 2018; Carrió-Pastor, 2019), makes it necessary to observe their specific translation competences from a more restrictive standpoint.

The aim of this study is to determine the translation competences required to translate a representative business genre, the distributor agreement, through corpus analysis. To do so, we will first set the theoretical background on what translation competence means, and will overview different approaches that have been explored in literature (Hatim & Mason, 1997; Koby & Melby, 2013; PACTE, 2020), focusing on those tailored to translation in business contexts (Román Mínguez, 2012; Lyu, 2020; Andrienko & Shpeniuk, 2021). This will serve as foundation to search for the specific abilities that a translator in this field should ideally have according to the specific characteristics of business genres, which will be analyzed next; although different classifications have been proposed from multiple points of view (participants, thematic field...), scholars agree in considering them as a separate area with specific characteristics and translation needs, what will directly influence the skills that a translator in this field will require.

After this introductory part, we will develop the methodology used to reach our aim: we will analyze a 10,000-word parallel corpus (limited, but within the adequate size range proposed for LSP studies by scholars like Bowker and Pearson, 2002) consisting of a 484-word distributor and manufacturer agreement as source text (ST), written in American English by a legal expert, and twelve target texts (TTs) of translations into Spanish carried out by translation trainees; as results will be of interest for training, the use of this profile will make it possible to specifically analyze trainees' strategies and needs. The first stage of the analysis will be to determine the most characteristic features of the ST, focusing on terminology (representative terms of the genre), style (specific drafting strategies), and meaning (key elements that set the purpose of the text); these features will be extracted and analyzed manually. After this initial step, we will retrieve the translation equivalents of those features from the TTs in order to contrast the differences in the performance of the translators; this process will involve computerized one-to-many alignment through the use of a translation memory. The assessment of the different levels of adequacy of the equivalents will shed light on what skills are necessary to meet the expectations of the expert target audience, and on how the (in)sufficient command of such skills influences the success of the TT.

Results show that linguistic, terminological and genre-related competences are the key skills that need to be deployed for the translation difficulties posed by the ST, along with skills related to translation standards in the business context and thematic knowledge. As each translation problem requires the use of certain translation competences to solve them, the thorough analysis of each competence also reveals that translators experience more difficulties when dealing with genre-related terminology and business thematic knowledge.

Despite the limitations of this study, corpus analysis has resulted a useful methodology to obtain relevant data on translation competences in business genres, and we wish to refine and broaden our study in the future.

**Keywords:** translation competence, trade agreement, business translation, corpus analysis, parallel corpus.

### **References**

- Alcaraz Varó, E. (2000). *El inglés profesional y académico*. Madrid: Alianza Editorial.
- Andrienko, T. and Shpeniuk, I. (2021). Strategic competence as a core of business translator's qualification. *SHS Web of Conferences* 105, 05001. doi: 10.1051/shsconf/202110505001

- Bowker, L. and Pearson, J. (2002). *Working with Specialized Language: a Practical Guide for Using Corpora*. London: Routledge.
- Carrió-Pastor, M. L. (2019). "Authorial engagement in business emails: a cross-cultural analysis of attitude and engagement markers". In Sancho Guinda, C. (Ed.). *Engagement in Professional Genres*. Amsterdam/Philadelphia: John Benjamins Publishing Company, pp. 47-66.
- Hatim, B., and Mason, I. (1997). *The translator as communicator*. London, New York: Routledge.
- Koby, G. S., and Melby, A. K. (2013). Certification and job task analysis (JTA): Establishing validity of translator certification examinations. *The International Journal of Translation and Interpreting Research*, 5(1), 174–210. doi:10.12807/ti.105201.2013.a10
- Lyu, J. (2020). The Modularized Construction on Translation Competence for Business English Majors in China. *English Language Teaching*; Vol. 13, No. 7, 124-129.
- Moreno-Pérez, L. (2018). "A case of hybrid text translation: secondary-field terminology in purchase agreements". In Gallego Hernández, D. and Poirier, E. (eds.), *Business and institutional translation: new insights and reflections*. Newcastle upon Tyne: Cambridge Scholars Publishing, pp. 73-88.
- PACTE Group. Amparo Hurtado Albir, Anabel Galán-Mañas, Anna Kuznik, Christian Olalla-Soler, Patricia Rodríguez-Inés and Lupe Romero. (2020). Translation competence acquisition. Design and results of the PACTE group's experimental research, *The Interpreter and Translator Trainer*, 14:2, 95-233. doi: 10.1080/1750399X.2020.1732601
- Román Mínguez, V. (2012). "Algunas peculiaridades de la traducción inglés-español del discurso de los negocios". In Martino Alba, P. and Lebsanft, Ch. (eds.), *Telar de traducción especializada*. Madrid: Dykinson, pp. 91-103.

## **How economists modify their claims: A study of epistemic and attitudinal stance devices in the diachronic LexEcon corpus**

Maria Teresa Musacchio

University of Trieste (Italy)

In the investigation of economic discourse, a major role has been played by the way economists express arguments (Merlini Barbaresi 1983, 1996; Swales 1993; Donohue 2006). The linguistic study of economic argumentation has focused on hedging (Bloor and Bloor 1993) in the discourse of predicting and forecasting (Merlini 2005) as specific devices economists use to position their (hybrid) discipline as a soft science increasingly exploiting methods and tools typical of hard sciences. However, argumentation has been regarded as typical of academic research papers while textbooks avoid presenting critical reading since they are supposed to transmit a canon (Swales 1993: 224). As a research unit within the national project LexEcon – The Economic Teacher: A translational and diachronic study of treatises and textbooks of economics (18th to 20th century). Intra- and interlingual corpus-driven and corpus-based analysis with a focus on lexicon and argumentation – funded by the Italian Ministry of University and Research, we investigate economic discourse as evidenced by an English, French and Italian corpus of theoretical treatises, textbooks and popular science handbooks published between 1900 and 1970. In this paper, we present a pilot study of epistemic and attitudinal stance markers – adverbials, complement clauses, nouns + prepositional phrases, premodifying stance adverbs – as hedging features pragmatically expressing degrees of epistemic certainty as opposed to attitudinal inferencing in the construction of economic arguments in a 6.5m-word subcorpus of English textbooks and handbooks (1900-1929). Triangulating data from previous research (Hyland 2005, Martin and White 2005, Biber et al. 2021) we have extracted a list of epistemic and attitudinal stance markers to probe our corpus and investigate how predicting and forecasting are presented in textbooks to train students in economic argumentation as part and parcel of education in becoming economists. Preliminary results suggest that epistemic and attitudinal stance is mainly expressed through a core set of adverbs and complement clauses with different frequencies compared to general language corpora such as the BNC and COCA and special language corpora to indicate the central role played by pragmatically hedging claims in economic predicting and forecasting. In terms of epistemic certainty the most frequent adverbs or expressions are *be possible/possibly/the possibility that*, *be likely* and *be (un)certain that/to*, while *be obvious/obviously*, *be/become clear/clearly*, *necessarily* and *be sure/surely* ranked most frequent as inferencing devices. Future research will extend to cover both the remaining English component and the Italian component of the 20th century LexEcon corpus to provide a more exhaustive diachronic view of predicting and forecasting as expressed through epistemic and attitudinal stance markers in economics up to 1970s.

**Keywords:** diachronic economics corpus, epistemic stance devices, attitudinal stance devices, economic predicting, economic forecasting

### **References**

- Biber, D., Johansson, S., Leech, G.N., Conrad, S., Finegan E. (2021). *Grammar of Spoken and Written English*. Benjamins.
- Bloor, M. and Bloor, T. (1993) How economists modify propositions. In W. Henderson, T. Dudley-Evans and R. Backhouse (eds.), *Economics and Language* (pp. 153-169). Routledge.
- Donohue, J.P. (2006). How to support a one-handed economist: The role of modalisation in economic forecasting. *English for Specific Purposes*, 25, 200-216.
- Hyland, K. (2005). *Metadiscourse*. Continuum.
- Martin, J.R. and White, P.R.R. (2005). *The Language of Evaluation. Appraisal in English*. Palgrave MacMillan.
- Merlini Barbaresi, L. (1983). *Gli atti del discorso economico: la previsione. Status illocutorio e modelli linguistici nel testo inglese*. Edizioni Zara.
- Merlini Barbaresi, L. (1996). *Traduzione e pragmatica del discorso*. In G. Cortese (ed.), *Tradurre i linguaggi settoriali* (pp. 73-85). Edizioni Libreria Cortina.

- Merlini Barbaresi, L. (2005). Il discorso economico/argomentativo: marcatezza e complessità della previsione. In L. Schena, C. Preite, & S. Vecchiato (eds.), *Gli insegnamenti linguistici nel nuovo ordinamento: Lauree triennali e specialistiche dell'area economico-giuridica* (pp. 301–324). Egea.
- Swales, J. (1993). “The paradox of value: Six treatments in search of the reader”. In W. Henderson, T. Dudley-Evans and R. Backhouse (eds.), *Economics and Language* (pp. 223-239). Routledge.

## **La terminología del reino fungi y su traducción del español al inglés: un estudio contrastivo basado en corpus**

María-Teresa Ortego-Antón and Rosa Pascual-Contreras  
Universidad de Valladolid

La internacionalización del sector agroalimentario en España ha propiciado un aumento exponencial de las necesidades de los servicios de redacción y de traducción del español al inglés. El valor estratégico de la investigación en el procesamiento del lenguaje natural (PLN) para el desarrollo de herramientas que automaticen esta demanda queda fuera de toda discusión. En el marco de esta necesidad y de esta parcela de I+D se encuadra nuestra propuesta, consistente en el estudio del comportamiento morfosintáctico de léxico especializado y de su fraseología en el campo específico de la industria de la trufa y de las setas. Nos hemos propuesto, a partir del estudio de un corpus virtual paralelo (P-FUNGICORP), compuesto por fichas técnicas de trufa y setas en español y de su traducción a la lengua inglesa, detectar las estrategias de equivalencia de dicho léxico y de su fraseología para definir los patrones de comportamiento de esta terminología agroalimentaria. Basándonos en un protocolo de compilación de corpus previamente avalado por autores como Seghiri (2017), Ortego Antón (2019) o Sánchez Carnicer (2022), entre otros, para garantizar que la compilación de nuestro corpus reúna todos los requisitos que permitan una investigación de calidad y contrastada, hemos detectado, extraído y analizado con la asistencia de ParaConc (Moropa, 2007) los patrones de comportamiento de los principales términos en español y sus equivalentes en lengua inglesa, así como de la fraseología especializada. A continuación, hemos contrastado dichos patrones con los resultados que se desprenden del análisis de C-FUNGICORP, un corpus virtual comparable español-inglés de fichas técnicas de trufa y setas, que hemos explotado con el analizador de concordancias de SketchEngine. Los resultados de dicho contraste nos han permitido, en primer lugar, detectar patrones de comportamiento morfosintáctico y fraseológico típicos de los textos traducidos y comprobar si dichos patrones se reproducen en los textos redactados originalmente en lengua inglesa. Por último, estos resultados pueden ser aplicados para la generación de una base de datos terminológica que alimente futuras aplicaciones basadas en la IA y en el PLN, así como en el aula de traducción especializada.

**Keywords:** corpus, fraseología, setas, trufa, inglés, español

### **References**

- Moropa, K. 2007. Analysing the English-Xhosa parallel corpus of technical texts with Paraconc: a case study of term formation processes. *Southern African Linguistics and Applied Language Studies* 25(2): 183-205.
- Ortego Antón, M. T. 2019. *La terminología del sector agroalimentario (español-inglés) en los estudios contrastivos y de traducción especializada basados en corpus: los embutidos*. Berlín: Peter Lang.
- Sánchez Carnicer, J. 2022. *Traducción y discapacidad. Un estudio comparado de la terminología inglés-español en la prensa escrita*. Berlin: Peter Lang.
- Seghiri, M. 2017. Metodología de elaboración de un glosario bilingüe y bidireccional (inglés-español/español-inglés) basado en corpus para la traducción de manuales de instrucciones de televisores. *Babel* 63(1): 43-64.

## Exploring text complexity in translation through readability and automatic textual profiling

Alberto Pasquetto

University of Padova (Italy)

This paper shows the results of a preliminary study on text complexity in Italian considering traditional readability indexes and innovative automatic textual profiling perspectives. Readability has been an established field of study since the late 19th century (DuBay 2004), and readability indexes have long been the reference method to assess the complexity of a variety of texts, finding applications in many different fields to address the issues of both interlingual and intralingual linguistic barriers. Readability is generally regarded as the ease with which a text is comprehensible to a reader. Its assessment is usually achieved through an estimation of text complexity by means of simple formulae that are based on textual attributes such as number of characters, number of words, syllable counts and sentence length. To this day, the formulae introduced in the 20th century are still largely used in academia (Lee, Lee 2023). However, the notion of complexity in linguistics can also refer to different aspects of a language and has also been extensively studied since the mid-20th century (Szmrecsányi 2004). Thanks to the new findings in computational linguistics, it is now possible to automatically assess text complexity through automatic text profiling, which carries out in-depth linguistic analysis of texts through highly efficient methodologies. Despite its potential, to this day this approach has somewhat been neglected. The main objective of this study is to compare traditional and innovative approaches to readability and study their compatibility, in order to establish which method is most suitable to give indication of complexity of a text. This research is part of a larger study that considers translation and easy language in crisis communication in order to produce texts that are understandable to people with poor linguistic proficiency or linguistic minorities. Readability and complexity in emergencies can be more challenging compared to other fields, as information should be accessible to most of the target audience, also in a variety of different languages (Temnikova et al. 2015).

A corpus consisting of 68 Italian texts about emergencies published by different Italian institutions has been created and used for the evaluation. The Gulpease Readability Index (Lucisano, Piemontese 1988) was used to compute readability of each text. Then, other measures of complexity (Type-Token Ratio, Mean Word Frequency, Herdan's and Guiraud's Indexes) were calculated with the concordancer AntConc (Anthony 2022). Finally, a state-of-the-art linguistic tool for textual profiling and analysis – Profiling-UD (Brunato et al. 2020) – was used to retrieve data from the deep structure of the texts. Statistical evaluations showed that eight out of the thirty language complexity features considered had statistical significance in the variation of readability. These features are related to lexis, deep syntax, and superficial syntax. In particular, readability is hindered by extensive use of subordination, by clauses complexity both in terms of length and depth of dependency links, by lexical density, and distribution of verbal heads. However, results show that estimating complexity through traditional readability indexes can still be a valid strategy to have a quick insight of the difficulty of a text in Italian.

**Keywords:** readability, complexity, textual complexity, automatic textual profiling

### References

- Anthony, L. (2022). AntConc (4.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>.
- Brunato, D., Cimino, A., Dell'Orletta, F., Venturi, G., Montemagni, S. (2020). Profiling-UD: a Tool for Linguistic Profiling of Texts. In: Proceedings of the Twelfth Language Resources and Evaluation Conference, pp. 7145–7151.
- DuBay, W. H. (2004). *The Principles of Readability*. USA: ERIC Clearinghouse.
- Lee, B. W., Lee, J. H. J. (2023). Traditional Readability Formulas Compared for English. arXiv:2301.02975
- Lucisano, P., Piemontese, M. E. (1988). GULPEASE: Una formula per la predizione della difficoltà dei testi in lingua italiana. *Scuola e città*, 3, 110-124.

- Mondorf, B. (2003). Support for more-support. In: G. Rohdenburg, B. Mondorf (Eds.). *Determinants of Grammatical Variation in English*, pp. 251-304.
- Szmrecsányi, B. M. (2004). On operationalizing syntactic complexity. In: *Proceedings of the 7th International Conference on Textual Data Statistical Analysis*, 1032-1039.
- Temnikova, I., Vieweg, S., Castillo, C. (2015). The case for readability of crisis communications in social media. *Social Web for Disaster Management Conference*. In: *Proceedings of the 24th International Conference on World Wide Web*, 1245–1250.

## **La traducción automática de las expresiones multiverbales del turismo gastronómico: una propuesta metodológica de análisis basada en corpus**

Isabel Peñuelas Gil

Universidad de Valladolid

En la última década, las tecnologías de la traducción han experimentado grandes cambios, los cuales no solo son aparentes en el entorno profesional, sino también en el social. Es especialmente llamativo el crecimiento sin precedentes en lo referido al uso de traductores automáticos, pues, gracias a la aparición de motores de traducción automática (TA) en línea, es cada vez más fácil encontrar los resultados de estas herramientas en ámbitos muy variados de la sociedad. Este crecimiento, propiciado por una atmósfera que se decanta por la globalización, se ha hecho también aparente en el sector turístico, donde cada vez es más común ver cómo empresas, especialmente aquellas de pequeño y mediano tamaño, así como algunas organizaciones e instituciones se decantan por usar estas herramientas para vender sus servicios y productos a un público internacional más amplio. No obstante, se trata de herramientas que, en la actualidad, proporcionan resultados limitados en cuanto a calidad y adecuación que pueden derivar en problemas de comunicación con los potenciales turistas internacionales. La presente investigación se centra en el ámbito del turismo gastronómico y analiza la efectividad de la TA (ES-EN) al lidiar con uno de los elementos lingüísticos más característicos de esta rama de conocimiento y una de las mayores dificultades para estas herramientas: las expresiones multiverbales (EMV) con un trasfondo altamente cultural. Con este propósito, se ha compilado un corpus monolingüe (ES) compuesto por 30 folletos y guías sobre el turismo gastronómico de distintas ciudades y regiones españolas, siguiendo las bases del protocolo de compilación propuesto por Seghiri (2017) y que posteriormente se ha utilizado en diversos trabajos (Ortego Antón, 2019, 2020, en prensa; Sánchez Carnicer, 2022; Peñuelas Gil, en prensa). Las EMV empleadas en este proceso se seleccionaron haciendo uso de las herramientas AntConc (Anthony, 2020) y Sketch Engine (Kilgarriff et al., 2004), que permitieron ver estos elementos dentro de su contexto el cual se tendría también en cuenta en la fase de traducción. Los traductores automáticos empleados (DeepL, Google Translate, Yandex y Microsoft Translator) representan los paradigmas más comunes a día de hoy dentro de la TA para fines específicos, cuentan con una popularidad significativa entre el público general y, al haber sido entrenados con diferentes corpus, proporcionan resultados diferentes. Los resultados obtenidos de este proceso se analizaron y categorizaron siguiendo una versión ampliada del modelo MQM propuesto por Ortiz (2016), prestando atención primero a las EMV en sí mismas y luego a los contextos. Durante el análisis se han identificado patrones de traducción literal que resultan incompatibles, en la mayoría de los casos, con la idiomática de las expresiones analizadas, así como errores relacionados con, por ejemplo, la estructura gramatical de las oraciones al incluir los elementos fraseológicos, tratamiento incorrecto de la persona gramatical o sugerencias de modificación del original no pertinentes. La combinación de los hallazgos, pone de relieve los obstáculos de comunicación que el uso de TA puede acarrear y, por lo tanto, las carencias que los textos de este sector presentados a los potenciales turistas de lengua no hispana, interfiriendo así en el correcto desarrollo de una de las actividades económicas más importantes de España.

**Keywords:** turismo gastronómico, traducción automática, expresiones multiverbales, corpus, español, inglés.

### **Referencias**

- Anthony, L. (2020). AntConc (Versión 3.5.9) [Computer Software]. Tokio, Japón: Waseda University. Disponible en <https://www.laurenceanthony.net/software/antconc/>.
- Kilgarriff, A.; Rychlý, P.; Smrž, P.; Tugwell, D. (2004). «The Sketch Engine». Proceedings of the 11th EURALEX International Congress, pp. 105-116. Disponible en <http://www.sketchengine.eu>.
- Ortego Antón, M. T. (2019). La terminología del sector agroalimentario (español-inglés) en los estudios contrastivos y de traducción especializada basados en corpus: los embutidos. Berlín: Peter Lang.



- Ortego Antón, M. T. (2020). Las fichas descriptivas de embutidos en español y en inglés: un análisis contrastivo de la estructura retórica basado en corpus. *Revista Signos*, 53(102), 170-194.
- Ortego Antón, M. T. (en prensa). «The Design of Torrezno TRAD: the Semiautomatic Spanish-English Writing and Translation Aid Tool», en Peñuelas Gil, I. y Ortego Antón, M. (en prensa), *Interpreting and Translation for Agri-food Professionals in the Global Marketplace*. Berlín: DeGruyter.
- Peñuelas Gil, I. (en prensa). «From corpora to a semi-automatic tool: the use of model lines for the development of TorreznoTRAD», en Peñuelas Gil, I. y Ortego Antón, M. (en prensa), *Interpreting and Translation for Agri-food Professionals in the Global Marketplace*. Berlín: DeGruyter.
- Sánchez Carnicer, J. (2022). *Traducción y discapacidad. Un estudio comparado de la terminología inglés-español en la prensa escrita*. Berlín: Peter Lang.
- Ortiz, C. (2016). *Implementing Machine Translation and Post-Editing to the Translation of Wildlife Documentaries through Voice-over and Off-screen Dubbing* [Tesis doctoral, Universitat Autònoma de Barcelona]. Cora TDX. <http://hdl.handle.net/10803/400020>
- Seghiri, M. (2017). Metodología de elaboración de un glosario bilingüe y bidireccional (inglés-español/español-inglés) basado en corpus para la traducción de manuales de instrucciones de televisores. *Babel*, 63(1), 43-64.

## Towards the development of a near-parallel corpus of news texts: an experiment in using multilingual sentence embeddings

Piotr Pęzik<sup>1</sup> and Łukasz Grabowski<sup>2</sup>

<sup>1</sup>University of Łódź (Poland)

<sup>2</sup>University of Opole (Poland)

Acquisition of bilingual corpus data for a parallel corpus is a complex task and the challenges include data availability, copyright, data quality, linguistic variation or domain-specificity, among others. Capitalizing on recent developments in parallel corpus mining using multilingual sentence embeddings (e.g., Guo et al. 2018; Artexte & Schwenk 2019a, 2019b), we undertake an attempt at extending the English-Polish parallel corpus Paralela (Pęzik 2016) with domain-specific publicly available news texts covering the years 2015-2021. Thanks to news circulation, whereby the news is first published in English and then it is reported by press agencies in other languages in various online channels of communications (online newspapers, news portals etc.), it is possible to obtain valuable up-to-date bilingual data that convey the same or similar message in two languages.

After extracting the English and Polish seed texts via API from Monco corpora: monco.frazeo.com and monco.frazeo.pl (Pęzik 2020), we obtained bilingual news data by using multilingual sentence embeddings (LaBERT, i.e. Language-agnostic Bidirectional Encoder Representation from Transformers), that is, by calculating a cosine distance between the source and target sentence vectors. More precisely, each English source language sentence was aligned with one best scoring (by absolute cosine similarity) Polish target sentence, an approach called an absolute forward method. Additionally, as suggested by Artexte and Schwenk (2019b), we applied the margin between the cosine score of a given candidate target sentence as well as the average cosine score of its *k* nearest neighbours. This method called margin-based scoring (Artexte and Schwenk 2019b) enabled us to avoid cases when due to hubness problem an incorrectly aligned poor bilingual sentence pair has a higher cosine similarity than a good translation pair.

Next, a custom-designed procedure for manual assessment of the degree of equivalence was developed and applied to the data, where a distinction between “translation equivalence” and “news equivalence” was made. The former one means that the bilingual data represents a source text and its translation in the target language, while the latter means that the bilingual data represents a pair of texts in the source and target language is not in fact an accurate translation but conveys, in approximation, a similar message in two languages. In short, we used a random sample of 150 news headlines and assessed them qualitatively and quantitatively by calculating the cosine distance between LaBERT vectors, a procedure undertaken to set the “translation equivalence” thresholds using conditional inference trees. Bilingual pairs of news headlines falling below the translation equivalence threshold, yet above the news equivalence threshold, can be referred to as “near-parallel data”.

We obtained 1,016,367 pairs of aligned English-Polish segments, which will be made available on the Paralela corpus website (paralela.clarin-pl.eu). We call the compiled sub-corpus of bilingual news texts as a “near-parallel” corpus as it contains high-quality, up-to-date translational sentence-level equivalents as well as thematically similar bilingual pairs of headlines, the latter providing a valuable insight into cross-cultural news circulation. We believe that the corpus offers valuable bilingual data useful for translators, translator trainers and translation researchers, among others.

**Keywords:** parallel corpus, corpus mining, multilingual sentence embeddings, translation equivalence, news equivalence.

### References

- Artetxe, M. & Schwenk, H. (2019a). Massively Multilingual Sentence Embeddings for Zero-Shot Cross-Lingual Transfer and Beyond (<https://arxiv.org/abs/1812.10464>)
- Artetxe, M. & Schwenk, H. (2019b). Margin-based Parallel Corpus Mining with Multilingual Sentence Embeddings. Proceedings of the 57th Annual Meeting of the Association for

- Computational Linguistics, pages 3197–3203 (<https://www.aclweb.org/anthology/P19-1309>)
- Guo, M., Shen, Q, Yang, Y., Ge, H., Cer, D., Abrego, G., Stevens, K., Constant, N., Sung, Y-H., Strope, B. & Kurzwel, R. (2018). Effective Parallel Corpus Mining using Bilingual Sentence Embeddings. Proceedings of the Third Conference on Machine Translation (WMT), Volume 1: Research Papers, pages 165–176 (<https://www.aclweb.org/anthology/W18-6317>)
- Lai, G., Dai, Z. & Yang, Y. (2020). Unsupervised Parallel Corpus Mining on Web Data. (<https://deepai.org/publication/unsupervised-parallel-corpus-mining-on-web-data>)
- Pęzik, P. (2016). Exploring Phraseological Equivalence with Paralela. In: E. Gruszczyńska & A. Leńko-Szymańska (Eds), Polish-Language Parallel Corpora, pages 67–81. Warsaw: Instytut Lingwistyki Stosowanej UW, 2016.
- Pęzik, P. (2020). Budowa i zastosowania korpusu monitorującego MoncoPL. Forum Lingwistyczne (7), 133-150. (doi: 10.31261/fl.2020.07.11)

## **Headlines and leads in English-Spanish news reports: Examining the evolution of writing and translation practices through conjunctions and lexical features**

Isabel Pizarro-Sánchez

Universidad de Valladolid

News reports can be classified as a genre according to Swales' definition (1990) because they share the common communicative aim of conveying current events to their audience. This is reflected in their rhetorical structure, consisting of moves and steps. Among them, headlines and leads are crucial moves as they capture the reader's attention and provide a concise summary of the story.

The aim of this study is to provide insight into the composition of headlines and leads and their translations. More specifically, it aims to analyze their composition across two distinct time periods to determine whether there are significant differences, and to examine their corresponding translations from English into Spanish to establish if there has been an evolution in the translation practices over time. The analysis focuses on two main elements to uncover patterns and tendencies in the writing and translation processes: (i) conjunctions, and (ii) lexical density and lexical diversity (Nasseri & Thompson, 2021).

We used a parallel corpus of English and Spanish news reports from *The Economist*, *Actualidad Económica* and *La Razón*. It includes texts dealing with a variety of topics, published in two distinct periods (1993 and 2021), and totaling 168 samples, 42 per language and period. The corpus was aligned to facilitate contrastive analysis. The SketchEngine tool was used to generate lists of types and tokens, with their respective frequencies, for calculating lexical diversity and density. Lists of conjunctions were also obtained as they reflect syntax complexity.

Our analysis revealed interesting patterns in the way news is reported and translated in these publications. The *Economist* headlines tend to be more concise and use simpler syntactic structures compared to their corresponding translations in *Actualidad Económica* and *La Razón*. This trend was observed in both periods of time under investigation. Furthermore, English headlines showed higher lexical density and lower lexical diversity. Spanish Leads are generally longer and provide more context than those in *The Economist*, adhering to the genre conventions of each language. However, this may also be attributable to the explicitation translation universal (Baker, 1993). Additionally, our findings indicate that the Spanish leads exhibit a greater syntax complexity, as evidenced by a normalized higher frequency of conjunction usage, a higher diversity, and more frequent use of subordinate conjunctions. Regarding the evolution over time, the most significant difference was observed in the length of the headlines, with those in the 2021 subcorpus being longer than those in the 1993 subcorpus, for both source and translated news reports. The reasons for this difference are unclear but may be due to changes in editorial policies or shifts in news consumption.

The findings can be applied to teaching professional writing to L1 journalists and L2 learners of English for Specific Purposes and developing guidelines for human translators. By analyzing the differences between English and Spanish news reports at various linguistic levels, we can provide valuable insights for language learners and educators.

Author keywords: news reports, headlines, leads, parallel corpus, contrastive analysis, translation, writing, conjunctions, lexical density, lexical diversity

### **References**

- Baker, M. (1993). *Corpus Linguistics and Translation Studies: Implications and Applications*. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and Technology: In Honour of John Sinclair* (pp. 233-250). John Benjamins Publishing Company.
- Nasseri, M., & Thompson, P. (2021). Lexical density and diversity in dissertation abstracts: Revisiting English L1 vs. L2 text differences. *Assessing Writing*, 47, 100511. <https://doi.org/10.1016/j.asw.2020.100511>
- Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge University Press.

## **“Let us Begin by Supposing That”: An Analysis of the Verbs Expressing Prediction and Forecast in the LexEcon Corpus**

Carla Quinci

University of Padova (Italy)

In the long-lasting debate around the divide between hard and soft sciences, Economics has been generally considered a social – i.e. soft – discipline due to its relative inability to predict (cf. Shapin 2022). Yet, this has not impeded economists from analysing and interpreting economic facts, formulating hypotheses and scenarios, and making predictions, which constitute precisely the distinguishing feature of the economic discourse (Merlini Barbaresi 2005, 309; Musacchio 1995, 17–18). Depending on their functions, economic predictions and forecasts were found to be (a) interpretative, when they stem from the economist’s interpretation of past events, (b) illustrative, when they aim to exemplify complex economic phenomena, (c) applied, when they predict future evolutions based on the analysis of the current economic situation, and (d) instrumental, when they serve to back up the economist’s proposals for measures to be implemented (Merlini Barbaresi 1983). These can also vary with reference to their epistemic gradient, i.e. the level of certainty they imply, and inferential continuum, i.e. their level of subjectivity. Both the predictive nature of economic texts and the economist’s stance in predictions and forecasts can be observed through the analysis of verbs by considering, for instance, their semantics, tense (Musacchio 2017, 57), and modality (Donohue 2006; cf. the role of will and may in Merlini Barbaresi 1984). Drawing on these considerations, this paper seeks to explore the role of verbs in formulating predictions and forecasts in 46 English economics textbooks and handbooks published between 1900 and 1929 (6.5 million tokens), which represent a subcorpus of the much larger LexEcon corpus, a diachronic and multilingual collection of 18th-, 19th- and 20th-century economic texts in six different languages, i.e. Italian, French, English, German, Spanish and Portuguese. This investigation adopts a primarily corpus-based approach and observes the frequency and use of 32 lexical and (semi-)modal verbs (Biber et al. 2021, 482–96) in the selected subcorpus. The list was obtained by drawing on previous research on economic prediction and expanded by including synonyms and the verbal instances related to prediction/forecasting and modality that were found among the first 200 items of the wordlist. In an attempt to cover the epistemic gradient and inferential continuum as much as possible, the list includes assertive (e.g. believe, think, find), predictive (e.g. assume, predict, expect, forecast), and modal verbs, as well as the semi-modals be going to and need to. The data revealed the preponderance of specific predictive verbs (i.e. suppose and assume) over the others, which are generally used in a non-scientific fashion, and confirmed the key role of prediction (i.e. will, would, shall, be going to) over other modalities, including that of possibility (i.e. may, might, could), which ranked second. The analysis also explored the connections between the most frequent lexical and (semi-)modal verbs and specific types of predictions and their position on the epistemic gradient and inferential continuum, confirming for instance the different stances implied in the modals may and will and their synonyms.

**Keywords:** prediction, forecast, economics, corpus linguistics, LexEcon

### **References**

- Biber, Stig, Geoffrey Johansson, Susan Leech, Edward Conrad, and Douglas N. Finegan. 2021. *Spoken English Grammar of and Written*. Amsterdam/Philadelphia: John Benjamins.
- Donohue, James P. 2006. “How to Support a One-Handed Economist: The Role of Modalisation in Economic Forecasting.” *English for Specific Purposes* 25 (2): 200–216. <https://doi.org/10.1016/j.esp.2005.02.009>.
- Merlini Barbaresi, Lavinia. 1983. *Gli atti del discorso economico: status Illocutorio e modelli nel testo inglese: la previsione*. Parma: Edizioni Zara.
- Merlini Barbaresi, Lavinia. 1984. “Will, may e l’espressione della supposizione e dell’inferenza.” In *Letteratura e deduzione & Discourse Analysis. Atti del VI Convegno Nazionale dell’Associazione Italiana di Anglistica, Pavia 22-23-24 Ottobre 1983*, edited by Tomaso Kemeny, Lia Guerra, and Anthony Baldwin, 123–29. Fasano: Schena.

- Merlini Barbaresi, Lavinia. 2005. "Il discorso economico/argomentativo: marcatezza e complessità della previsione." In *Gli insegnamenti linguistici nel nuovo ordinamento: lauree triennali e specialistiche dell'area economico-giuridica*, edited by Leandro Schena, Chiara Preite, and Sara Vecchiato, 301–24. Milano: Egea.
- Musacchio, Maria Teresa. 1995. *La traduzione della lingua dell'economia dall'inglese all'italiano*. Trieste: Edizioni LINT.
- Musacchio, Maria Teresa. 2017. *Mediating across Languages and Cultures: Economics and Finance as Popular Science in Translation*. Padova: CLEUP.
- Shapin, Steven. 2022. "Hard Science, Soft Science: A Political History of a Disciplinary Array." *History of Science* 60 (3): 287–328. <https://doi.org/10.1177/00732753221094739>.

## **Going to / Ir a + inf and Be/Estar + gerund: Mutual correspondence and cross-linguistic distribution**

Rosa Rabadán

Universidad de León

English going to + inf is frequently used to express the future, but also indicates fulfillment of present intention (Huddleston and Pullum 2002) and prediction (Budts and Petré 2016). Ir a + inf means imminent future, and deontic (intention) and epistemic modality (Gómez Torrego 1999, Fernández de Castro 1999, Melis 2006: 921-22). Be + gerund marks progressive aspects, including iteration, temporary habit, arrangement, or plan. Estar + gerund (Yllera 1999) marks progressive aspect (Gómez Torrego 1999), temporary habit, and iteration (Gómez Torrego 1988:143-44). This paper sets out to answer two questions: Do these correspondent constructions have the same distribution in English and Spanish? And, what can translated data offer to gain a better understanding of how they relate cross-linguistically?

Data come from the bi-directional parallel corpus P-ACTRES 2.0. This paper uses a bi-directional cross-linguistic double contrast to find the translation solutions given in the corresponding target language and account for meanings marking using the tertium comparationis outlined above. The mutual correspondence value (Altenberg 1999; Ebeling and Ebeling 2015, Ebeling 2017) and the calculation of statistically significant differences help to unveil the degree of overlap between the forms in English and Spanish. The qualitative contrast adds information on meaning distribution in the cross-linguistic map.

Bilingual findings indicate that the going to + inf form occurs 322.62 times per million words (pmw), while ir a + inf happens 912.20 pmw. The MC value comes to 57.34%, which may indicate that over 40% of the cases have alternative interpretations. According to our corpus data, the be + gerund construction occurs at 2700.64 pmw and estar + gerund at 43.96 pmw. The MC value is 23.34%, indicating a low degree of overlap.

Going to + inf is translated by its formal correspondents in 64.60% of cases, with nearly 40% adopting other translation solutions. Ir a + inf is translated almost 50% into its correspondent and other options. Concerning going to + inf meanings, results indicate that 91.35 % of cases mark futurity combined with prediction, while nearly 5% are deontic (intention) readings. In 4% of cases analyzed, the epistemic reading (prediction) prevails over the basic futurity meaning. Ir a + inf findings show that 61.3% of the cases mark the future, while 25.3% mean prediction and 13.4% signal intention.

Be + gerund favors simple tenses as translation solutions (62.25%) over formal correspondents (18.31%). Other options include nominalizations (6.19%). The less frequent estar + gerund is translated into the English progressive tenses in 75 % of the cases. Other options include modal constructions (8.3%) and non-finite constructions (5.5%). Over 97% of the cases of be + gerund convey progressive aspect, arrangement/ plan cases come up to 1.61% while temporary habit and iterative aspect make 0.8% and 0.53%, respectively. Findings for estar + gerund show 94.4% progressive cases, and 5.55% epistemic readings, linked to Spanish future and conditional tenses. Regularities form-meaning in the translation solutions are inconclusive, except for those cases translated by alternative means, which tend to indicate intention and prediction. These results suggest different views of aspectual content cross-linguistically and noticeably other distributions. Contrasting them with non-translated materials will help identify more clearly differences in distribution and possible interference in the translations.

**Keywords:** Going to / Ir a + inf, Be/Estar + gerund, bi-directional parallel corpus, mutual correspondence, cross-linguistic distribution, translation solutions

### **References**

- Altenberg, B. 1999. "Adverbial Connectors in English and Swedish: Semantic and Lexical Correspondences." In H. Hasselgård and S. Oksefjell (eds). *Out of Corpora: Studies in Honour of Stig Johansson*. Amsterdam: Rodopi. 249–268
- Berglund, Y. 2000. *Gonna and going to in the spoken component of the British National Corpus*. In C. Mair and M. Hundt (eds) *Corpus linguistics and linguistic theory*. Brill. 35-49.

- Budts, S. and Petré, P. 2016. Reading the intentions of be going to on the subjectification of future markers. *Folia Linguistica Historica* 37: 1-32
- Ebeling, J. & S. Oksefjell Ebeling. 2015. An English-Norwegian contrastive analysis of downtoners, more or less. *Nordic Journal of English Studies (NJES)* 14 (1), 62–89.
- Ebeling, S. Oksefjell. 2017. Bringing home the bacon! A contrastive study of the cognates bring/bringe in English and Norwegian. *Kalbotyra*. ISSN 1392-1517. 70, p. 104–126. doi: 10.15388/klbt.2017.11193.
- Fernández de Castro, F. 1999. *Las perífrasis verbales en el español actual*, Madrid, Gredos.
- Gómez Torrego, L. 1988. *Perífrasis verbales*, Madrid: Arco Libros.
- Gómez Torrego, L. 1999. Los verbos auxiliares. Las perífrasis verbales de infinitivo. In I. Bosque and V. Demonte (eds) *Gramática descriptiva de la lengua española*. Madrid: Espasa. 3323-3389
- Hasselgård, H. 2015. Coming and going to the future: Future-referring expressions in English and Norwegian. In S.O. Ebeling and H. Hasselgård (eds), *Cross-linguistic Perspectives on Verb Constructions*, 88-115. Newcastle: Cambridge Scholars Publishing.
- Huddleston, R. and G. K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316423530>
- Melis, C. 2006. Verbos de movimiento. La formación de los futuros perifrásticos. In C. Company Company (dir.), *Sintaxis histórica de la lengua española*. Volume 2, México: Fondo de Cultura Económica, Universidad Nacional Autónoma de México. 873-968.
- Yllera, A. 1999. Las perífrasis verbales de gerundio y participio. In I. Bosque and V. Demonte (eds.) *Gramática descriptiva de la lengua española*. Madrid: Espasa. 3391-3442.



## **Inclusión, integración o inserción de las personas con discapacidad, ¿qué término se debería emplear?**

Jaime Sánchez Carnicer  
Universidad de Valladolid

En la actualidad, dentro de nuestra sociedad, todo lo relacionado con la discapacidad (accesibilidad, discriminación, posición dentro de la sociedad, etc.) es un asunto con una gran repercusión a nivel social y político. En esta misma situación se encuentra el lenguaje empleado para referirse a todo lo que rodea a las personas con discapacidad en diferentes aspectos de su vida diaria. De igual forma, los corpus (monolingües, bilingües, multilingües) dentro de los Estudios de Traducción, se han convertido en una herramienta esencial hoy en día para, entre otras muchas aplicaciones, el análisis del discurso y el análisis contrastivo entre lenguas. Por ello, en este trabajo aunaremos ambas disciplinas para analizar las colocaciones y connotaciones en los medios de comunicación escritos de referencia de los términos empleados para hacer referencia a la participación de las personas con discapacidad dentro de la sociedad, como por ejemplo «inclusión», «integración» o «inserción» en lengua española y sus equivalentes en lengua inglesa. Para la extracción de los resultados emplearemos el corpus DISCORP-PRESS (Sánchez Carnicer, 2022), un corpus virtual comparable bilingüe (EN/ES), compilado mediante el protocolo propuesto por Seghiri (2017), empleado con posterioridad en otros trabajos (Ortego Antón, 2021), compuesto por textos obtenidos de periódicos españoles y británicos de tirada nacional. Emplearemos herramientas como AntConc (Anthony, 2022) y Sketch Engine (Kilgarrieff et al., 2004) para realizar la extracción previamente mencionada, ya que nos permitirán estudiar las colocaciones y las connotaciones que presentan los términos que forman parte de la muestra de estudio para este trabajo dentro del corpus analizado. Los resultados que se desprendan del análisis nos permitirán ampliar el conocimiento dentro de este campo del saber y poder constatar la evolución de la terminología relativa a las personas con discapacidad. Con este fin, en este trabajo, mediante el análisis de los resultados obtenidos con la explotación del corpus DISCORP-PRESS, se comprobará el uso y la prevalencia de los términos de nuestra muestra de estudio durante los años que abarca nuestro corpus en la prensa española y británica, tanto en la lengua española como en la inglesa. De la misma manera, podremos observar la forma en la que se comportan las dos lenguas involucradas en este estudio, con el fin de detectar un uso más o menos inclusivo de cada una de ellas y cuál de las dos ha evolucionado más dentro de este ámbito de estudio. Por último, al contar con textos provenientes de medios de comunicación escritos de ideologías opuestas, se analizará si el hecho de contar con esta dualidad de opiniones se ve reflejada en la mayor presencia de alguno de los términos de estudio o alguna colocación concreta en ambas lenguas.

**Keywords:** corpus, discapacidad, traducción, inglés, español.

### **Referencias**

- Anthony, L. 2022. AntConc (Version 4.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Kilgarrieff, Adam; Rychly, Pavel; Smrz, Pavel; y Tugwell, David. 2004. «The Sketch Engine». *Proceedings of Eurlex*: 105-116.
- Ortego Antón, María Teresa. 2021. «e-DriMe: A Spanish-English Frame-based e-Dictionary about Dried Meats». *Terminology*, 27 (2): 294-321. DOI: 10.1075/TERM.20013.ort
- Sánchez Carnicer, Jaime. 2022. *Traducción y discapacidad. Un estudio comparado de la terminología inglés-español en la prensa escrita*. Berlín: Peter Lang.
- Seghiri, Míriam. 2017. «Metodología de elaboración de un glosario bilingüe y bidireccional (inglés-español/español-inglés) basado en corpus para la traducción de manuales de instrucciones de televisores». *Babel* 63 (1): 43-64. DOI: 10.1075/babel.63.1.04seg

## Proyecto PaCorES: corpus paralelos bilingües con español

M<sup>a</sup> Teresa Sánchez<sup>1</sup> Nieto and Irene Doval<sup>2</sup>

<sup>1</sup>Universidad de Valladolid

<sup>2</sup>Universidad de Santiago de Compostela

El objetivo de esta ponencia es presentar el proyecto PaCorES, una colección de corpus paralelos bilingües con español como una de sus lenguas. Se mostrarán las características principales de PaCorES y sus funcionalidades actuales sobre la base de dos de los dos corpus que lo componen y que se encuentran actualmente ya en línea: los corpus paralelos español/alemán PaGeS ([www.corpuspages.eu](http://www.corpuspages.eu)) y español/ inglés PaEnS ([www.corpuspaens.eu](http://www.corpuspaens.eu)). Además, se informará sobre el desarrollo de los dos nuevos corpus previstos para el futuro más inmediato: los corpus paralelos chino/español (PaCheS) y español/francés (PaFreS).

En primera parte de la ponencia se expondrán las reflexiones básicas que sustentan la organización, compilación y puesta a disposición de la colección de corpus paralelos PaCorES:

(a) Un corpus paralelo ha de ser multifuncional. Dado el ingente trabajo que supone su elaboración, ha de ser una herramienta en abierto que resulte útil para gran variedad de aplicaciones y capaz de satisfacer las necesidades de usuarios muy variados y con diferentes grados de experiencia en la interacción con herramientas de corpus en general y de corpus paralelos en particular.

(b) El corpus ha de ofrecer todos sus recursos textuales en línea en una interfaz única y amigable, donde el usuario pueda configurar sus búsquedas según distintos parámetros.

(c) A la hora de construir un corpus ha de procederse a la reutilización de los conocimientos adquiridos y ha de buscarse la reusabilidad de los recursos creados extendiendo el modelo a otros pares de lenguas.

(d) Un corpus ha de ser utilizado y explotado también por el propio equipo creador para distintas investigaciones y aplicaciones, según sus intereses particulares. De este modo se puede observar de primera mano la funcionalidad de la herramienta y se detectan necesidades y/o carencias.

En la segunda parte de la ponencia, tras una breve revisión de los corpus existentes en este momento relacionados con la identidad de PaCorES —esto es, corpus paralelos multilingües accesibles en abierto y que incluyan el español como una de sus lenguas—, se incidirá en las siguientes cuestiones: (1) la arquitectura de los corpus paralelos que componen PaCorES: se definirán los dos elementos que la sustentan —el corpus nuclear y los suplementos— y se explicará el tipo de recursos que integra cada uno de estos elementos; (2) cuestiones básicas relativas a la compilación del corpus, incidiendo en las tareas de preprocesado de los textos y en su postprocesado, con la anotación lingüística y extralingüística; (3) la alineación de los textos, con detalles sobre la selección de la herramienta utilizada, sobre las estadísticas básicas del alineado en los dos corpus de PaCorES actualmente en línea —PaGeS y PaEnS— y sobre los principales fenómenos observados en la revisión del alineado; (4) las herramientas de búsqueda que ofrece el motor de búsqueda a través de la interfaz web del corpus y las características de visualización de los resultados de las búsquedas y se incidirá en la optimización y ampliación de las funcionalidades de los corpus de PaCorES.

**Keywords:** corpus paralelos, corpus en línea, alemán-español, inglés-español, corpus comparables.

## References

- Doval Reixa, I., Fernández Lanza, S., Jiménez Juliá, T. E., Liste Lamas, E., Lübke, B., Doval Reixa, I. (2019). Corpus PaGeS: A multifunctional resource for language learning, translation and cross-linguistic research. In I. Doval Reixa & M. T. Sanchez Nieto (Eds.), *Parallel Corpora for Contrastive and Translation Studies: New resources and applications*, (pp. 103–121).
- Doval Reixa, I., & Sánchez Nieto, M. T. (2019). Parallel corpora in focus: An account of current achievements and challenges. In I. Doval Reixa & M. T. Sanchez Nieto (Eds.), *Parallel corpora for contrastive and translation studies: New resources and applications* (pp. 1–15). Retrieved from <https://benjamins.com/catalog/scl.90>

- Hasselgård, H. (2015). Parallel corpora and contrastive studies. Proceedings of the International Symposium on Using Corpora in Contrastive and Translation Studies 2010 Conference (UCCTS2010). <https://www.lancaster.ac.uk/fass/projects/corpus/UCCTS2010Proceedings/papers/Hasselgard.pdf>.
- McEnery, T., & Xiao, R. (2007). Parallel and Comparable Corpora: What is Happening? In G. Anderman & M. Rogers (Eds.), *Incorporating corpora. The linguist and the translator* (pp. 18–31). Clevedon/Buffalo/Toronto: Multilingual Matters.
- Mikhailov, M., & Cooper, R. (2016). *Corpus linguistics for translation and contrastive studies: a guide for research* [Book]. Londres: Routledge.
- Zanettin, F. (2012). *Translation-Driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies*. Manchester: St. Jerome.

## **Un triple corpus paralelo bilingüe para el estudio interpretación simultánea con texto en el Parlamento Vasco**

Claudia María Torralba Rubinos and Inés García Azkoaga  
Universidad del País Vasco

Las intervenciones que realizan los políticos en el seno de un parlamento pueden considerarse como pertenecientes al discurso especializado si asumimos que su desarrollo está ligado a la comunidad epistemológico-discursiva que utiliza la lengua (Swales, 1990). Dichas producciones verbales, además, comparten ciertas características y regularidades ligadas al género de texto o discurso (Bronckart, 1996). Algunas de esas particularidades han sido estudiadas, por ejemplo, por Van Dijk y Rodrigo-Mendizabal (1999) o Brenes (2016). En cualquier caso, también han sido objeto de atención en estudios del ámbito de la traducción y de la interpretación (Schäffner, 1997; Boéri, 2004, entre otros). En esta ocasión, proponemos poner el foco en una modalidad muy concreta de interpretación: la interpretación simultánea con texto. Así, una de las particularidades a tener en cuenta es que la oralización del discurso parlamentario casi siempre tiene como punto de partida el texto escrito en las intervenciones iniciales, es decir, en el primer turno de palabra de cada parlamentario, por lo que cuentan con un alto grado de planificación (Fuentes-Rodríguez & Alcaide-Lara, 2002). Considerando también las particularidades propias del contexto sociolingüístico bilingüe del País Vasco, los discursos o textos producidos en el parlamento de esa comunidad son a menudo objeto de interpretación simultánea con texto en la combinación lingüística vasco-español, por lo que a las versiones escritas y oralizadas hay que añadir también la de la interpretación. El objetivo de este trabajo es, precisamente, indagar en las posibilidades que ofrece un corpus paralelo de estas características para analizar las expresiones parentéticas. Su especificidad y novedad radica en que recoge discursos escritos y orales (transcritos para su análisis) producidos en el Parlamento Vasco, algunos íntegramente en euskera y otros bilingües, así como su interpretación al castellano, en consonancia con las prácticas discursivas monolingües y bilingües que allí se llevan a cabo. Concretamente, en el corpus se recogen el texto escrito original (texto A), la transcripción del discurso oral realizado en sede parlamentaria (texto B) y la transcripción de su interpretación simultánea al castellano (texto C). En total, se han recogido las tres versiones de 30 intervenciones, con lo que se ha confeccionado un corpus de 140.046 palabras con una duración de 5 horas 28 minutos y 58 segundos para cada una de las versiones orales. El análisis comparativo de este corpus permitirá el estudio de las particularidades que afectan a la realización de los elementos parentéticos y su reinterpretación, tanto en la versión oralizada del parlamentario como en la versión del intérprete. El análisis comparativo de este corpus nos permite constatar que la presencia de esos elementos parentéticos no está garantizada de forma paralela en las tres versiones.

**Keywords:** corpus paralelo, discurso parlamentario, bilingüismo, interpretación, simultánea con texto.

### **Referencias**

- Boéri, J., (2004). Aproximación a la interpretación simultánea de discursos políticos: discursos del Parlamento Europeo. *Jornades de Foment de la Investigació*. Universitat Jaume I.
- Brenes, E. (2016). La fraseología en el discurso parlamentario. Análisis argumentativo de las locuciones y paremias. In C. Fuentes (eds.), *Estrategias argumentativas y discurso político*. Madrid: Arco Libros, 251-272.
- Bronckart, J.-P. (1996). *Activité langagière. Textes et discours. Pour un interactionisme socio-discursif*. Lausanne: Delachaux et Niestlé.
- Schäffner, C. (1997). Strategies of Traslating Political Texts. In A. Trosborg (arg.), *Text Typology and Translation*. Amsterdam: John Benjamins, 119-144.
- Swales, J.M. (1990). *Genre analysis. English in academic and reseach setting*. Cambridge: Cambridge University Press
- Van Dijk, T. A. y Rodrigo-Mendizabal, I. (1999). *Análisis del discurso social y político*. Quito: Abya-Yala.