

LARA MORATÓN

JULIA LAVID

Universidad Complutense de Madrid

laramoraton@ucm.com

lavid@filol.ucm.es

STANCE AND ENGAGEMENT IN ENGLISH AND SPANISH JOURNALISTIC TEXTS: TOWARDS A RELIABLE ANNOTATION SCHEME FOR LINGUISTIC AND COMPUTATIONAL PURPOSES

Keywords: Stance, Engagement, English, Spanish, corpus annotation

Abstract

The paper describes the process of validating a reliable annotation scheme for the categories of Stance and Engagement in English and Spanish using a bilingual sample of English-Spanish journalistic texts extracted from the MULTINOT corpus (Lavid et al. 2015). The bilingual sample includes three different newspaper genres: news reports, editorials and letters to the editor. Following the generic annotation pipeline proposed by Hovy and Lavid (2010), the paper describes the different steps to validate an annotation scheme to capture the main features of Stance and Engagement and their realizations in English and Spanish. This includes the instantiation of the theoretical categories, the development of annotation guidelines and the performance of an inter-annotation agreement study on a small training corpus to measure the reliability of the proposed tags. The paper also describes the results of the annotation of a larger corpus using the validated scheme (Moratón 2015). This reveals interesting patterns of variation in the distribution of Stance and Engagement in the three newspaper genres, which can be fruitfully used for contrastive linguistic and computational purposes.



1. Introduction

The categories of Stance and Engagement in discourse have been studied from different perspectives and used for the description of different genres (see Hyland 2005; Biber, Finegan 1998; Martin, White 2005, inter alia). Hyland's characterization of these categories as part of his model of intersubjective positioning is one of the best and most comprehensive frameworks for examining the means by which interaction is achieved in written communication. In this model, Stance is the way writers express a textual "voice" or a community recognised personality; it basically refers to the ways writers present themselves and convey their judgements, opinions, and commitments, stamping their personal authority onto their arguments or stepping back and disguising their involvement. Stance concerns writer-oriented features of interaction and is comprised of four main elements: hedges, as illustrated in (1a) and (1b), boosters, as in (2a) and (2b), attitude markers, as in (3a) and (3b) and self-mentions, as in (4a) and (4b) below:

(1a) The major Western oil companies can probably stomach prices dipping temporarily to \$60 for a few months and still pay their dividends, but a prolonged slump at \$50 per barrel would lead to the scrapping of investments followed by another cycle of shortages and soaring prices

(1b) Un escándalo que ha conmocionado a la opinión pública y que puede desestabilizar al Gobierno.

Translation: A scandal that has shocked public opinion and can destabilize the Government

(2a) The same was true of Starbucks' recent announcement that it would pay £20m in corporation tax.

(2b) No cabe esperar de la ONU decisiones drásticas.

Translation: No drastic decisions can be expected from the UN.

(3a) Microsoft, the world's biggest computer company, offered a rare glimmer of hope to Wall Street after it unveiled better than expected quarterly profits.

(3b) Resulta sorprendente que un país tan aislado, pobre y proscrito internacionalmente como Corea del Norte mantenga tal empeño en romper uno tras otro los puentes que todavía le unen al exterior.

Translation: It is surprising that a country that is so isolated, poor and internationally outlawed as North Korea maintains such an effort to break the bridges that still bind it to the outside.

(4a) As a teacher, the depressing statistics do not surprise me.

(4b) Comprendo perfectamente a los miles de españoles que el pasado 14 de noviembre se han movilizado en España y han secundado la huelga general convocada por los dos sindicatos mayoritarios en contra de los recortes sociales.

Translation: I fully understand the thousands of Spaniards who on November 14 have gone to the streets in Spain and supported the general strike called by the two major unions against social cuts.

Engagement is the way writers relate to their readers with respect to the positions advanced in the text. For Hyland, engagement is “an alignment dimension where writers acknowledge and connect to others, recognizing the presence of their readers, pulling them along with their argument, focusing their attention, acknowledging their uncertainties, including them as discourse participants, and guiding them to interpretations” (Hyland 2005: 176). Engagement is typically realized by the linguistic features illustrated by the contrastive examples below, and include: questions, both real and rhetorical, realized by interrogatives (examples 5a and 5b); inclusive first person plural, indefinite, and second person pronouns and items referring to readers (examples 6a and 6b); directives, realized by imperatives – as shown in example (7a) – or obligation modals referring to actions of the reader (*must, ought, should, have to, need to*) – as shown in example (7b) – and adjectival predicates controlling a complement to clause, directing readers to a particular action; references to shared knowledge, as illustrated by examples (8a) and (8b); and asides addressed to the reader, marked off from the ongoing flow of text, as illustrated by examples (9a) and (9b) below:

(5a) How many bobbies on the beat could have been employed for the cost of this election?

(5b) ¿Aquellas personas sin empleo y sin vivienda que no pueden pagar sus deudas por culpa de un Gobierno, unas empresas y una banca; o aquellas culpables de la crisis y que se están lucrando de unos recortes e impuestos sin precedente a las clases medias?
Translation: Those people without employment and housing who cannot pay their debts because of a government, companies or banks; or those guilty of the crisis who are profiting from unprecedented cuts and taxes to the middle classes?

(6a) We listened to the Home or Light programme morning noon and evening, informing ourselves about the world.

(6b) Las políticas de austeridad que nos marca Bruselas nos están empobreciendo a muchas forzadas.
Translation: The austerity policies that Brussels has set are impoverishing us.

(7a) Tell that to my parents as they huddled in the kitchen of their cramped council flat in north London in the 1950s and 1960s.

(7b) ¡Mantengan el Instituto de Cardiología!
Translation: Keep the Cardiology Institute open!

(8a) Even with the new kit, Iran’s bomb-making capability is not yet beyond control.

(8b) Cunde la sospecha de que los comportamientos irregulares en el entorno de los partidos se han desarrollado en un clima de total impunidad.
Translation: Suspicion arises that irregular behaviours in the parties have developed in a climate of total impunity.

(9a) (9a) And – as I believe many TESOL professionals will readily acknowledge – critical thinking has now begun to make its mark, particularly in the area of L2 composition.

(9b) La Ley Hipotecaria se reformó en los años noventa, creo que durante el Gobierno socialista de Felipe González.
Translation: Mortgage Law was reformed in the nineties, I think during the socialist government of Felipe González.

However, in spite of the plethora of studies which have applied Hyland's characterization of Stance and Engagement categories, there is still a lot of indeterminacy in their definitions to be able to be used in computational contexts for the development of high-quality annotated corpora. It is necessary, therefore, to empirically test the validity of these categories not only for theoretical purposes but also to be used for the annotation of large datasets, which can be later used for computational purposes such as Machine Learning, Automatic Text Classification, Multilingual Information Retrieval and Sentiment Analysis, among others.

In this paper, we undertake this task in the context of the MULTINOT project, aimed at the creation and empirical validation of discourse features in English and Spanish through corpus analysis and annotation (see Lavid et al. 2015; Lavid, Moratón 2016).

The paper is organized as follows: section 2 describes the aims of the paper and the research questions which the study tries to answer; section 3 describes the data used, consisting of a smaller training corpus of bilingual (English-Spanish) newspaper texts and a larger dataset of sixty bilingual newspaper texts; section 4 outlines the methodology used; section 5 presents and discusses the results of the annotation of the bilingual dataset. Finally, section 6 provides some concluding remarks.

2. Aims and research questions

This study has two main aims:

1. To empirically validate the categories of Stance and Engagement and their subtypes through bilingual corpus annotation.
2. To create a large corpus of bilingual newspaper texts annotated with Stance and Engagement tags and to examine their distribution patterns across three different newspaper genres and two languages, English and Spanish.

The research questions investigated in this study are the following:

1. Are Stance and Engagement, and their subtypes, valid categories for large-scale contrastive annotation of different newspaper genres?
2. If so, are there genre-specific and/or language-specific patterns of variation in the use of Stance and Engagement markers in English and Spanish newspaper texts?

3. Data

The data used for this study consisted of two datasets:

- a) A small training corpus consisting of eighteen bilingual (English-Spanish) comparable texts, divided into six news reports (3 English, 3 Spanish), six editorials (3 English, 3 Spanish) and six letters to the editor (3 English, 3 Spanish).
- b) A larger dataset of sixty bilingual texts, divided into three subcorpora: twenty news reports, twenty editorials and twenty letters to the editor.

Both the training set and the larger set were compiled from high-circulation and high-quality newspapers in the UK and Spain, namely: *Times online* (<http://www.timesonline.co.uk>), *The Independent* (<http://www.independent.co.uk/>) and *The Telegraph* (<http://www.telegraph.co.uk>), *El País* (<http://elpais.com/>), *El Mundo* (<http://www.elmundo.es/>) and *La Vanguardia* (<http://www.lavanguardia.com/>).

The reason for selecting these newspapers is their quality and seriousness, being leaders in the information market and of national circulation. Thus, *El País* has a progressist liberal editorial and laicist line (Salaverría et al. 2005: 127). *El Mundo* appeared later than *El País* but gained readers as their investigation on political corruption unveiled several cases in Spain. Its editorial line is considered to be liberal conservative. *La Vanguardia* was founded in Catalonia and has been identified as a bourgeois newspaper of liberal-moderate tendencies with a particular way of understanding Catalonia in Spain (Córdoba Hernández 2009).

4. Methodology

On the basis of what is now considered to be the standard methodology in corpus annotation projects, we performed the following steps:¹

- a) We first designed a preliminary annotation scheme identifying the categories to be used as tags for annotation. The annotation scheme included a core and an extended tagset inspired on the theoretical definitions provided by Hyland for Stance and Engagement (2005), but modified the latter on the basis of the proposal for Engagement features proposed by Martin and White (2005).
- b) We then performed a pilot agreement study to validate our initial annotation scheme.
- c) Once the initial scheme was validated, a larger bilingual set was annotated and the distribution of Stance and Engagement markers in the three newspaper genres in English and Spanish was compared.

These steps are described in detail in the following subsections.

4.1. Annotation scheme: core and extended tags

Our annotation scheme consists of a core and an extended tagset capturing the general categories of Stance [SM] and Engagement markers [EM] as core tags, and a number of subtypes as part of the extended tagset, as shown in Table 1.

As shown in Table 1, the annotation scheme is organized to allow annotators to choose from the coarser to the more specific tags. The abbreviated form of each tag is between square brackets next to the full form. Examples extracted from the corpus annotated with each of these tags are provided below; a suggested translation into English is provided for the Spanish examples:

¹ See the work of Hovy and Lavid (2010) and Lavid (2012) for more details on the different steps in the corpus annotation process.

Core Tagset	Extended Tagset	
Stance Markers [SM]	Hedges [HE]	
	Boosters [BO]	
	Attitude [AT]	
	Self-mention [SE]	
Engagement Markers [EM]	Questions [QU]	
	Inclusive 1st person plural [1P]	
	Indefinite 2nd person [2P]	
	Directives & Deontic [DI]	
	Attribution [A]	Acknowledge [AC]
		Distance [D]
		Concur [C]
	Proclaim [P]	Pronounce [Pr]
		Endorse [E]
	Entertain [EN]	Evidentials [EV]

Table 1. Core and Extended Tagsets for Stance and Engagement in English and Spanish

- ◆ Stance Markers [SM]:
 - Hedge [HE]:
 - a) With Mopti lost, it *might* [HE] have proved impossible to stop Mali's collapse.
 - b) Pero en este caso *considero* [HE], desde mi profesión de médico, que están dejando fuera aspectos tan importantes como no valorar que muchas personas, y muy preferentemente los parados, (...)
Translation: But in this case, I consider [HE], from my profession as a doctor, that they are leaving aside important aspects as not appreciating that many people, and most preferably the unemployed.
 - Booster [BO]:
 - a) *True* [BO], Mr Cameron sought to quell some of the more fevered speculation when he stressed today that the blame for what had happened rested firmly with the terrorists and no one else.
 - b) *Una de las virtudes que con razón* [BO] se le atribuyeron al debate sobre el estado de la nación (también en este periódico) fue el grado de concreción de las propuestas económicas del presidente del Gobierno.
Translation: One of the virtues that rightfully [BO] was attributed to the debate on the state of the nation (also in this newspaper) was the degree of concreteness of the economic proposals of the President of the Government.
 - Attitude [AT]:
 - a) As a teacher, the *depressing* [AT] statistics do not surprise me.
 - b) *Me quedo estupefacto* [AT] al saber que, prácticamente sin aviso previo, cierran el Instituto de Cardiología, del que soy (¡era!) afortunado paciente.

Translation: I am astonished [AT] to know that, almost without prior notice, they are closing the Institute of Cardiology, of which I am (was!) lucky patient.

- Self-mention [SE]:
 - a) *I* [SE] understand Neil March's discomfort about the BBC payout to Lord McAlpine (letter, 17 November), though what he does with the money is surely his business.
 - b) *Yo* [SE] invito encarecidamente a los responsables a que pasen revista a todo lo que, con nuestro dinero, se está costeando y que sale por más de 8,7 millones.
Translation: I [SE] strongly urged those responsible to review everything that is being paid with our money, and that goes for more than 8.7 million.
- ♦ Engagement:
 - Question [Q]:
 - a) Did all of their governments offer help, and what might have resulted if any or all had moved to intervene unilaterally? [Q]
 - b) Pero mi pregunta es la siguiente, ¿quiénes son los "morosos"? [Q]
Translation: But my question is the following: who are the defaulters? [Q]
 - Inclusive 1st person plural [1P]:
 - a) But *we* [1P] should applaud Lord McAlpine for taking action to discourage thoughtless and irresponsible tweeting.
 - b) Que como sabemos [1P] son FAES, Vera, etcétera.
Translation: As we [1P] know they are FAES, Vera, etc...
 - Indefinite 2nd person [2P]:
 - a) (...) if *you* [2P] go out to elect a councillor or MP, you may [2P] as well pick a police commissioner too.
 - b) Hasta entonces, no se podía hipotecar un bien si *uno* [2P] no era propietario del mismo.
Translation: Until then, a good could not be mortgaged if one [2P] did not own it.
 - Directives [DI]:
 - a) *Tell* [DI] that to my parents as they huddled in the kitchen of their cramped council flat in north London in the 1950s and 1960s.
 - b) ¡*Mantengan* [DI] el Instituto de Cardiología!
Translation: Keep [DI] the Institute of Cardiology!
 - Attribution [A]: propositions attributed to the citation of external sources out of the text. This includes two main types:
 - Acknowledge [AC]: Presents an external authorial voice clearly dissassociated with the author: *X said...*, *X believes...*, *according to X*, *in X's view*
 - a) "We, like most companies, saw a slowdown at the tail end of the quarter in particular", Chief Financial Officer Chris Liddell *said* [AC].
 - b) Este comando permitirá "abrir la caja negra, luchar contra la corrupción y la optimización fiscal agresiva", *asegura* [AC] Canfin.
Translation: This command would allow "to open the black box, fight against corruption and aggressive fiscal optimization", says [AC] Canfin.

- Distance [D]: it allows to express the author’s alignment of disalignment with the cited voice: *X claims that, it’s rumoured that*
 - a) He *warned* [D] that the most dangerous position this country could take in the light of changes being forced by the eurozone sovereign debt crisis would be to stand back and say we are going to do nothing about it.
 - b) El G7 *insiste* [D] además en que todos sus miembros tienen un “interés común” de que haya un sistema financiero internacional “fuerte y estable”.
Translation: The G7 also insists [D] that all its members have a “common interest”, a “strong and stable” international financial system.
- Proclaim [P]: Expresses contrary opinion or opposition to a certain point of view but not overtly. This includes:
 - Concur [C]: This formulation overtly announces the addresser position with a certain proposition.
 - a) Their proposal, for a three-month extension is *far from perfect*. [C]
 - b) O sea, *que con la está cayendo* [C], se destinan varios cientos de millones de euros que en ningún caso se justificará, ni el gasto ni los resultados de tales ayudas.
Translation: That is to say, even in the bad situation we are in [C], several hundred million euros are spent, with any justification, neither its expense nor the results of such aid.
 - Pronounce [Pr]: Covers overt intervention of the author in the text:
 - a) (...) *in the misguided belief that* [Pr] politics is a business where common sense and a spirit of co-operation automatically prevail.
 - b) *Es cierto que* [Pr] Francia tiene muchos más intereses en la región que el resto de sus socios europeos.
Translation: It is true that [Pr] France has many more interests in the region than the rest of its European partners.
 - Endorse [E]: The author refers to external sources as *valid, correct, undeniable* through the use of verbs as *show, prove, demonstrates...*
 - a) But it *shows* [E] the party has realised that politics as kamikaze warfare is self-defeating.
 - b) Ahora, como *demuestra* [E] la guerra en Mali, parece obvio que las bandas criminales que conforman AQMI tienen la fuerza necesaria para convertir el corazón del Sahel en un territorio hostil a Europa, dominado por la concepción más atrasada del islam.
Translation: Now, as the war in Mali shows [E], it seems obvious that the criminal gangs that make up AQMI (...)
- Entertain [EN]: It construes a high commitment to the proposition on the part of the speaker/writer via an assessment of high probability. It includes the so-called
 - Evidentials [EV]:
 - a) The coffee giant *sounded as though* [EV] it was making a generous donation to the UK’s collecting tin rather than discharging its obligations.

- b) *Parece ser* [EV] que todos los catalanes debemos posicionarnos a favor o en contra de la independencia, o lo que es lo mismo, a favor del Gobierno de la Generalitat o contra él, como si no importase nada más. Translation: It seems that the Catalans have to position ourselves in favour or against independence, or so to say, in favour or against the Generalitat Government, as if nothing else mattered.

4.2. Agreement study

In order to test the reliability and consistency of the core and the extended tagsets of the proposed annotation scheme, a pilot agreement study was performed by two expert annotators working on a small training corpus of eighteen comparable texts extracted from the larger set of bilingual newspaper texts contained in the MULTI-NOT corpus (Lavid et al. 2015). The training corpus contained eighteen texts, evenly divided into comparable sets of English and Spanish news reports, editorials and letters to the editor. The annotation procedure was carried out by the two annotators independently, as Lavid (2012: 5) points out: “the underlying premise of annotation is that if people cannot agree enough, then either the theory is wrong (or badly stated or instantiated), or the annotation process itself is flawed”.

The inter-annotator measure agreement used in our study was the Kappa coefficient (K) (Cohen 1960), a robust statistic for either interrater or intrarater reliability testing. Cohen’s Kappa ranges from -1 to $+1$, where 0 represents the amount of agreement that is obtained from random chance, and 1 represents perfect agreement between the raters. The range of Kappa results can be interpreted as shown in Table 2:

<i>k</i> -value	Agreement
<0	Chance
0.01–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Substantial
0.81–0.99	Almost perfect

Table 2. Interpretation of Kappa from Viera and Garrett (2005)

The values ≤ 0 indicate no agreement; values of 0.01–0.20 as none to slight; values of 0.21–0.40 as fair; values of 0.41–0.60 as moderate; values of 0.61–0.80 as substantial, and values of 0.81–1.00 as almost perfect agreement.

The K value result obtained in the inter-annotator agreement study on the Spanish dataset was of 0.631, which corresponds to a “substantial” agreement between

coders (0.61–0.80), as shown in Table 3, obtained after data evaluation with the SPSS statistical program:

Symmetric Measures

		Value	Asymptotic Standard Error ^a	Approximate T ^b	Approximate Significance
Measure of Agreement	Kappa	.631	.065	16.768	.000
N of valid Cases		59			

^a Not assuming the null hypothesis.

^b Using the asymptotic standard error assuming the null hypothesis.

Table 3. Inter-annotator agreement K value for Spanish texts

As to the English dataset, the K value result obtained in the inter-annotator agreement study was 0.790, which also indicates a “substantial” agreement between coders, as shown in Table 4.

Symmetric Measures

		Value	Asymptotic Standard Error ^a	Approximate T ^b	Approximate Significance
Measure of Agreement	Kappa	.790	.045	22.285	.000
N of valid Cases		91			

^a Not assuming the null hypothesis.

^b Using the asymptotic standard error assuming the null hypothesis.

Table 4. Inter-annotator agreement K value for English texts

Given the high degree of agreement obtained in the study, we consider that the proposed annotation scheme is empirically validated and can be used for the annotation of the larger bilingual dataset.

5. Annotation of the larger dataset

The annotation of the larger dataset offers interesting results regarding generic preferences in the use of Stance and Engagement markers in the three different newspaper genres. As shown in Table 5, there is a statistically-significant difference between these genres ($p < 0.05$), with a predominant use in letters to the editor, followed by editorials and a lower use in news reports.

	News Reports		Editorials		Letters to the Editor	
English	78/125	62.40%	157/212	74.06%	48/51	94.12%
Spanish	49/147	33.33%	84/137	63.50%	75/94	79.79%

Table 5. Distribution of Stance and Engagement markers in three newspaper genres

This result may be expected due to the communicative purpose of each newspaper genre. As explained elsewhere, news reports “should strive to remain objective and use neutral language while presenting a diversity of opinions, voices, and perspectives of the event, incident, or issue under discussion” (Lavid et al. 2013: 263). News reporters “avoid including explicit value judgments about the participants and the events in the news reports or confine contentious claims about causes and effects to the quotations of external sources” (Lavid et al. 2013: 263). The low frequency of expressions of Stance and Engagement markers in news reports is a reflection of this communicative purpose. In addition, the use of Attribution strategies to cite sources external to the text increase the credibility and objectivity of the information provided.

Editorials, by contrast, present the point of view of the newspaper and contribute to the formulation of certain *preferred* viewpoints about the world. They are one of the most well-written evaluative genres that reveal the position of the newspaper on a variety of current issues and thus have a vital role in the newspaper (Hynds 1984). They “offer newspaper readers a distinctive and sometimes authoritative voice that speaks to the public directly about matters of public importance” (Wang 2008: 170). This is reflected in the much higher number of Stance and Engagement markers (74.06% in the English newspapers and 63.50% in the Spanish ones), than the ones found in news reports (62.40% in the English newspapers and 33.33% in the Spanish ones).

Letters to the editor are written by individual readers that call for change or support and are usually highly subjective and passionate. This is reflected in the high proportion of Stance and Engagement markers found both in the English and in the Spanish Letters (English 94.12% and Spanish 79.79%).

The distribution of Stance and Engagement markers in the three different newspaper genres is displayed in Table 6:

		News Reports		Editorials		Letters to the Editor	
English	Stance	29/125	23.20%	47/212	22.17%	24/51	47.06%
	Engagement	49/125	39.20%	110/212	51.89%	24/51	47.06%
Spanish	Stance	1/147	0.68%	21/137	15.33%	29/94	30.85%
	Engagement	48/147	32.65%	63/137	45.99%	46/94	48.94%

Table 6. Distribution of Stance and Engagement in three newspaper genres

The use of Engagement strategies is higher than the use of Stance in the three genres and both languages, presenting a statistically significant difference among all genres.

Figure 1 illustrates this distribution graphically:

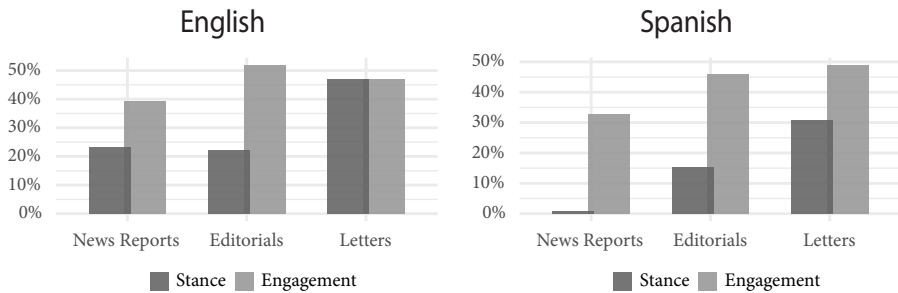


Figure 1. Distribution of Stance and Engagement markers in English and Spanish per genre

Some examples of Stance markers in the English and the Spanish editorials are provided in (1), (2), (3) and (4) below: in (1) we find an example of hedging through the use of the epistemic modal *may*; (2) is an example of a *hedge* in Spanish through the use of the epistemic construction *es poco probable que* (it is rather unlikely that). In (3) the adjective *true* acts as a *booster* and in (4) the adjective *seguro* (sure) fulfills a similar role in the Spanish sentence:

- (1) This *may* [HE] sound an attractive prospect, but it is not the direction of travel in Europe.
- (2) *Quizás* [HE] para curarse en salud, el tribunal se cuida de delimitar el objeto de su veredicto, negando que pueda extenderse “a valoraciones ajenas al objeto procesal ni a la conducta de personas no acusadas y sobre las que, en consecuencia, no se ha practicado prueba ni se ha formado convicción alguna”.
Translation: Perhaps [HE] by a way of precaution, the court is careful to delimit the object of its verdict, denying that it can be extended “to valuations that are foreign to the procedural object (...)”.

- (3) *True* [BO], Mr Cameron sought to quell some of the more fevered speculation.
- (4) (...) pero *es seguro que* [BO], ha asestado un golpe formidable a la reputación de un Parlamento considerado modélico en el exterior y honorable más allá de toda duda en el interior.
Translation: ...but for sure [BO] it has dealt a formidable blow to the reputation of a parliament considered a model abroad and honourable beyond any doubt in the interior.

As to the use of Attribution resources in news reports in both languages, this is the predominant strategy in comparison with the other Stance or Engagement markers (English 79.59% and Spanish 58.33%). Examples (5), (6), (7) and (8) make use of verbs such as *say* or *indican* (point out), *believe* or *reconocen* (acknowledge):

- (5) Economists *said* [AC] that this was a clear signal that the 15-nation bloc was sliding into recession.
- (6) “Estamos preocupados por la reciente excesiva volatilidad en el tipo de cambio del yen y sus posibles implicaciones adversas para la estabilidad económica y financiera”, *indican* [AC] los ministros de Finanzas y gobernadores de los bancos centrales del G7.
Translation: “We are worried about the recent over volatility in the exchange rate of yen and its possible negative implications for economic and financial stability” point out [AC] the G7 ministers of Finance and governors of central banks.
- (7) Mr Brown *believes* [AC] the IMF needs substantially more than the current \$250 billion set aside for struggling nations.
- (8) “Nos llevó tiempo llegar a esta situación, y llevará tiempo salir de ella”, *reconoce* [AC] el jefe del equipo de asesores económicos de la Casa Blanca, Edward Lazer, (...).
Translation: “It took us time to arrive to this situation and will also take some time to come out of it” acknowledges [AC] the chief of economic advisors to the White House, Edward Lazer (...).

By contrast, editorials and letters to the editor use *deontic and directives* to direct the reader to a particular action. Addressing the reader overtly and showing the position of the writer in these genres is widely used. Examples (9), (10), (11), and (12) show the use of these strategies in editorials; examples (13) and (14) illustrate these strategies in letters to the editor:

- (9) *What Britain needs is* [AC] a robust and unambiguous tax system that requires corporations to pay their share, without the need for negotiations with HMRC, or moral urgings from politicians, or gestures of largesse by pressurised managements.
- (10) Pero *es difícil admitir que* [HE] indagar sobre la existencia y naturaleza de la misión que parece haber condicionado la conducta del principal acusado sea tarea ajena al tribunal, bien directamente, bien por deducción de testimonio.
Translation: But it is difficult to admit [HE] that to investigate the existence and nature of the mission that seems to have conditioned the behaviour of the principal accused is a task outside the court, either directly or by deduction of testimony.

- (11) The Prime Minister's promise of "no boots on the ground" *must be* [HE] adhered to, however, regardless of how the conflict plays out.
- (12) El momento *no aconseja* [HE] ni excesos retóricos ni grandes representaciones.
Translation: The moment does not advise [HE] either rhetoric excesses or great representations.

In letters to the editor *deontic and directives* are the most used strategies in English (25%) and the second most preferred in Spanish (18%).

- (13) *Tell* that [HE] to my parents as they huddled in the kitchen of their cramped council flat in north London in the 1950s and 1960s.
- (14) *Hay que estar ciego* [HE] para no darse cuenta de que esto nos conduce al abismo y a un punto sin retorno.
Translation: One must be blind [HE] not to realize that this leads us to the abyss and to a point of no return.

Regarding language-specific preferences, the bilingual annotation revealed a lower frequency of Stance and Engagement markers in Spanish than in English, although the only statistical difference is the use of Stance elements in news reports, as shown in Figure 2.

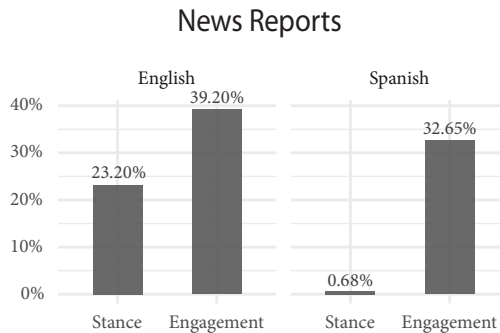


Figure 2. Distribution of Stance and Engagement markers in English and Spanish in news reports

The use of *hedges* in English editorials are the elements that mark the difference between these two genres. However, and according to the following examples, this may be due to a linguistic characteristic of English in the use of modal verbs that is not reflected in Spanish:

- (15) A deep recession *might* [HE] precipitate an oil price collapse.
- (16) The major Western oil companies *can probably* [HE] stomach prices dipping temporarily to \$60 for a few months and still pay their dividends, but a prolonged slump at \$50 per barrel would lead to the scrapping of investments followed by another cycle of shortages and soaring prices.

However, although the addition of new labels account for more types of Engagement strategies used in both languages, English editorials tend to present a higher frequency in comparison to the Spanish ones (English 51.89% vs. Spanish 45.99%). This result indicates that English editorialists tend to engage more frequently with their audience than Spanish ones, which remain more distant with respect to their audience in the high-quality press. According to Pak (1996: 68–69), “the Peninsular Spanish newspaper editorials, making attributions to a special source is thought to be unnecessary since the editorials are intended for a selected elite group of well-informed readers”.

Spanish letters to the editor present a similar frequency of Stance and Engagement expressions than the English ones (Engagement expressions in Spanish 48.94% and English 47.06%; Stance expressions in Spanish 30.85% and in English 47.06%, respectively). These results suggest that both languages follow a similar pattern, in the main: letters to the editor display a higher presence and involvement of the writers and more connection to their readers than the rest of the genres and this is reflected in the use of *self-mention* as the preferred Stance strategy for both languages (English 50% and Spanish 55%).

6. Conclusions

The annotation study presented in this paper has validated the categories of Stance and Engagement and their subtypes empirically for cross-linguistic large-scale annotation. The inter-annotator agreement study carried out on a training dataset and measured according to the Kappa coefficient revealed that the different proposed tags are robust and can be replicated by different annotators in a larger corpus.

The annotation of the larger dataset both in English and Spanish has revealed genre-specific and language-specific differences in the use of Stance and Engagement in the three newspaper genres studied. According to Swales (1990: 58), the communicative purpose “shapes the schematic structure of the discourse and influences and constrains the choice of content and style”. The *objectivity* of news reports is reflected in the scarce use of interpersonal strategies in the text’s construction while attribution strategies which make reference to external sources to the text abound in this genre.

Editorials and letters to the editor, have a persuasive and evaluative communicative purpose and this is reflected in the high proportion of *deontic and directives* elements by which the writer evaluates a particular situation in terms of probability, obligation or necessity. It is in the use of these elements where we find a language-specific variation: English tends to use more deontic expressions than Spanish, but this is probably due to the differences in use of modality in both languages.

Finally, letters to the editor are a highly subjective genre in both languages, since they are used to express personal views or make complaints, suggestions and recommendations. Thus, Stance and Engagement markers abound in this genre, independently from the British or the Spanish newspaper where they are published.

We bring this paper to an end in the belief that the results presented shed new light on the behaviour of Stance and Engagement markers in different newspaper genres and in British and Spanish newspapers. As we hope to have shown throughout the paper these markers are not only empirically-valid interpersonal features of texts, but can also be used as rhetorical signals of the communicative purposes fulfilled by different genres. These results are to be applied to written texts, not having yet being tested to oral language. In addition, it is our hope that the created annotated corpus of bilingual newspaper texts can be fruitfully used in the computational community for the creation of machine learning algorithms which can process interpersonal discourse features such as the ones investigated in this paper.

References

- Biber D., Finegan E. 1998. Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. – *Text-Interdisciplinary Journal for the Study of Discourse* 9.1: 93–124.
- Cohen J. 1960. A coefficient of agreement for nominal scales. – *Educational and Psychological Measurement* 20.1: 37–46.
- Córdoba Hernández A.M. 2009. *La línea editorial de abc, el país, el mundo y la vanguardia frente al conflicto palestino-israelí: 1993–2004*. [Ph.D. thesis, Pamplona; available at: <http://dadun.unav.edu/bitstream/10171/5124/4/Ana%20Maria%20Cordoba%20%28tesis%29.pdf>].
- Hovy E.H., Lavid J. 2010. Towards a ‘science’ of corpus annotation: A new methodological challenge for Corpus Linguistics. – *International Journal of Translation* 22.1: 13–36.
- Hyland K. 2005. Stance and engagement: A model of interaction in academic discourse. – *Discourse Studies* 7: 173–192.
- Hynds E.C. 1984. Editorials, opinion pages still have vital roles in most newspapers. – *Journalism Quarterly* 61: 624–639.
- Lavid J. 2012. Corpus analysis and annotation in CONTRANOT: Linguistic and methodological challenges. – Moskowich I., Crespo B. (eds.). *Encoding the past, decoding the future: Corpora in the 21st century*. Cambridge: 205–220.
- Lavid J., Arús J., DeClerck B., Hoste V. 2015. Creation of a high-quality, register-diversified parallel corpus for linguistic and computational investigations. – *Current work in Corpus Linguistics: Working with traditionally-conceived corpora and beyond. Selected papers from the 7th International Conference on Corpus Linguistics (CILC2015)* 198.24: 249–256.
- Lavid J., Arús J., Moratón L. 2013. Thematic variation in English and Spanish newspaper genres: A contrastive corpus-based study. – Aijmer K., Altenberg B. (eds.). *Advances in corpus-based Contrastive Linguistics: Studies in honour of Stig Johansson*. Amsterdam: 261–286.
- Lavid J., Moratón L. 2016. Generic structures, rhetorical relations and thematic patterns in English and Spanish journalistic texts: A comparative study. [Paper presented at the 26th ESFLW].
- Martin J.R., White P.R.R. 2005. *The language of evaluation. Appraisal in English*. Basingstoke.
- Moratón L. 2015. *Thematic patterning in English and Spanish: Contrastive annotation of a bilingual newspaper corpus for linguistic and computational applications*. [Unpublished Ph.D. thesis, Madrid; available at <http://eprints.ucm.es/39741/>].
- Pak C. 1996. *Newspaper editorials from The New York Times, El País, and El Universal: A comparative applied genre analysis*. [Unpublished Ph.D. dissertation, University of Michigan]. Ann Arbor (MI).

- Salaverría R., López García X., Limia Fernández M. et al. 2005. Análisis comparativo de cibermedios: Lavanguardia.es, Elmundo.es y Elpaís.es. – Salaverría R. (ed.). *Cibermedios. El impacto de internet en los medios de comunicación en España*. Sevilla: 125–143.
- Swales J.M. 1990. *Genre analysis: English in academic and research setting*. Cambridge.
- Viera A.J., Garrett J.M. 2005. Understanding interobserver agreement: The kappa statistic. – *Family Medicine* 37.5: 360–363.
- Wang W. 2008. Newspaper commentaries on terrorism in China and Australia: A contrastive genre study. – Connor U., Nagelhout E., Rozycki W.V. (eds.). *Contrastive rhetoric: Reaching to intercultural rhetoric*. Amsterdam: 169–191.

