Oral Production of Discourse Markers by Intermediate Learners of Spanish: A Corpus Perspective

Leonardo Campillos Llanos and Paula González Gómez

Abstract This study describes the oral production of discourse markers by 40 (N = 40) learners of Spanish and compares it with usage by native speakers (N = 4). Our data belong to a learner corpus of oral interviews with university learners from over nine language backgrounds at intermediate level: A2 (N = 20) and B1 (N = 20) (*Common European Framework of Reference*). Semiautomatic part-of-speech (POS) tagging was used to count and retrieve the discourse markers produced by each group of learners and the group of native speakers. Results show a slight increase in the acquisition of these particles from A2 to B1, although the production is still lower when compared with the group of native speakers. Certain groups of students (especially Chinese learners) show a poorer usage of this category in our data, which could reveal a certain difficulty acquiring fluency at the discourse level. A breakdown of the most used discourse markers in our corpus (in native and non-native speakers, and at A2 and B1) is presented, as well as a distribution across interviews of the ten most frequent markers. Results are discussed comparing the usage data in our corpus with teaching guidelines for Spanish.

Keywords Spanish language • Second language acquisition • Learner corpus research • Discourse markers • Oral production

L. Campillos Llanos (✉)
Computational Linguistics Laboratory, Department of Linguistics and Modern Languages, Facultad de Filosofía y Letras, Universidad Autónoma de Madrid, C/Francisco Tomás y Valiente, 1 Cantoblanco Campus, 28049 Madrid, Spain
e-mail: leonardo.campillos@uam.es

P. González Gómez
Language Service, Universidad Autónoma de Madrid, C/Freud, 1, Cantoblanco Campus, 28049 Madrid, Spain

J. Romero-Trillo (ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New Empirical and Theoretical Paradigms*, Yearbook of Corpus Linguistics and Pragmatics 2,
DOI 10.1007/978-3-319-06007-1_11,

1 Introduction

Interlanguage Pragmatics, as a subfield of cross-cultural pragmatics and second language (henceforth, L2) acquisition, is concerned with the study of L2 pragmatic use and L2 pragmatic learning/acquisition. Kasper (1992, 1996) points out that the study of use dominates Interlanguage Pragmatics and observes the need for more acquisitional studies to inform second language pedagogy.[1]

With this work we try to contribute to the research on Interlanguage Pragmatics by focusing on the acquisition of discourse markers (henceforth, DMs) in spoken L2 Spanish. Our study aims to analyse the non-native use of these devices by learners of Spanish at A2 and B1 levels (*Common European Framework of Reference*, henceforth *CEFR*; Council of Europe 2001) in order to address the following research questions:

- What markers are the most frequent in the oral discourse of learners with levels A2 and B1?
- What markers are the most frequent in the oral discourse of learners with different first language (henceforth, L1)?
- What differences exist in the use of DMs in native and non-native oral discourse?

The results of our data showed that there was a statistically significant difference between native and non-native speakers, who used fewer DMs at A2 and B1 levels. These outcomes support our claim that the appropriate use of DMs is crucial for the management of fluid and well-structured oral and written discourse. From a pedagogical point of view, these results lead us to suggest the explicit instruction of DMs to facilitate their acquisition, with the objective of helping students to develop their pragmatic-discourse competence.

This is a corpus-based study on foreign oral discourse and we are basically presenting a descriptive analysis, excluding variables such as the type of instruction received by learners. The reason is that, for some students, we do not know the nature of their instructional background, and to what degree it varies across learners. The work will start by explaining the theoretical framework and by introducing the research background on Interlanguage Pragmatics, especially on DMs. Then, we will focus on the existing literature regarding the acquisition of Spanish DMs, mainly corpus-based studies. After describing the methodology used,

[1] Research into acquisitional pragmatics should also be sociocognitive in orientation, just as we can put social and cognitive pragmatics together (Escandell 2004). A synthesis of the different aspects involved (learner and sociocultural context) configures the best framework for exploring how pragmatic competence is really acquired. Kasper and Rose (2002) offer a good review of the different theories of L2 pragmatic development (e.g. acculturation model, cognitive processing, and language socialization). Bardovi-Harlig (2013) also integrates social and cognitive factors in her consideration of the areas of research that are of interest for L2 pragmatics research (e.g. task design for the study of implicit and explicit knowledge, and the effect of environment on pragmatic development).

we will provide an analysis of the results. Finally, some pedagogical implications will be offered with the aim of improving the teaching of these particles in the classroom.

2 Theoretical Framework

The acquisition of pragmatic knowledge plays an essential role in the development of Communicative Competence (Hymes 1972; Canale and Swain 1980; Bachman 1990; Celce-Murcia et al. 1995). Pragmatic knowledge can be defined as the ability to use adequate language in different communicative situations. Consequences of pragmatic failure (Thomas 1983) can be worse at the social level than those produced by errors dealing with grammar or vocabulary. This mainly happens when the learner shows a high level of proficiency.

More research is still needed about pragmatic teachability. Nevertheless, in recent years it has been widely suggested that explicit pragmatic instruction improves the development of pragmatic competence (Kasper 1996; Rose and Kasper 2001; Jung 2002; Bardovi-Harlig and Mahan Taylor 2003). In order to guide instruction, we need to know how learners acquire pragmatic strategies and devices such as discourse markers. For that purpose, oral production data can provide very useful information about spoken interlanguage.

Some criticism has been made of the fact that studies in Interlanguage Pragmatics have paid more attention to the description of language use than to developmental aspects (Langer 2001), and also to the limited amount of research in Romance languages. Although it is true that research on the acquisition of the pragmatic level in Spanish is scarce, Galindo (2005) reported some references of studies regarding pragmatic transfer (none of them based on oral data). Other studies focused on speech acts and pragmatic errors in oral production (Blum-Kulka and House 1989; Koike 1989, 1996; Lorenzo-Dus 2001). These studies used listening comprehension tests or written questionnaires, which do not shed much light on the acquisition of oral skills. Nevertheless, by applying a Discourse approach, Lorenzo-Dus and Meara (2004) researched the assessment of spoken competence, and Sessarego (2009) examined the effects of instruction.

Non-native speakers have special difficulties mastering pragmatic devices that are more frequent or characteristic of speech, as is the case with discourse markers. These elements (also called *pragmatic markers* or *discourse particles*; see Blakemore 2004: 221; Aijmer and Simon-Vandenbergen 2011: 226) contribute significantly to the interpretation of the utterance due to its procedural meaning (Fraser 1999), as established in the Relevance theory (Blakemore 1987, 1992; Sperber and Wilson 1986).

A comprehensive and unified taxonomy of these particles is far from achieved. As Aijmer and Simon-Vandenbergen explain (2011: 228), Fraser (1996) distinguishes pragmatic markers, commentary markers, and parallel markers, but excluding words such as *well*. Further, some hesitators (e.g. *uhm, er*) are also sometimes

considered DMs (e.g. Gilquin 2008), but there is no general agreement on this criterion. However, some studies (e.g. Aijmer 2002) broadly identify a *textual function* and an *interpersonal function* of DMs.[2] Both functions are related to the notion of *indexicality* or *indexical function*, i.e. DMs denote the relation of an utterance to the immediate context, thereby creating cohesion (Fung and Carter 2007: 414; Aijmer and Simon-Vandenbergen 2011). While the *textual function* refers to the role these particles play in structuring the discourse, the *interpersonal function* concerns the expression of the speaker's attitudes or his/her role in the conversation. But regardless of the categorization of DMs, these devices should be present in the learners' interlanguage for them to successfully build their discourse, convey their attitude, or interact with other people in a natural way.

In the last three decades research on DMs has proliferated since the initial studies (e.g. Schiffrin 1987), and most research has been performed within a corpus-based approach, and in particular for English (e.g. Aijmer 2002; a full list of references is collected in O'Keeffe et al. 2007: 171–172). A recent methodology for the study of discourse markers has even proposed the use of translations to analyse the correspondences and differences between particles in the same semantic and pragmatic field (Aijmer and Simon-Vandenbergen 2003a, b, 2004; Simon-Vandenbergen and Willems 2011; see also the volume edited by Aijmer and Simon-Vandenbengen 2006).

Even though there has been less research devoted to the acquisition of these particles, the last decade has seen a growing interest in the application of corpus methods to address this question, especially in spoken interlanguage (e.g. Romero-Trillo 2002, 2004; Fuller 2003; Aijmer 2004, 2011; Müller 2005; Ramírez Verdugo and Romero-Trillo 2005; Llinares-García and Romero-Trillo 2006, 2008; Buysse 2007, 2011, 2012; Fung and Carter 2007; Gilquin 2008; Hellermann and Vergun 2007; Liao 2009; Mukherjee 2009; Polat 2011; Wei 2011; Zhao 2013; see the volumes edited by Romero-Trillo 2008, 2013).

3 Studies on Acquisition of Spanish Discourse Markers

Although there are already detailed studies on Spanish DMs (e.g. Briz 1993a, b; Casado Velarde 1998; Martín Zorraquino and Portolés 1999; Pons Bordería 2006), we are only aware of a few studies related to the acquisition of these particles. Fernández, in her error analysis of written texts (1990, 1997), included some DMs along with cohesive devices such as conjunctions, though other conversational markers were not covered (e.g. *bueno* or *hombre*). More recent research has gathered spoken data. For example, Díez Domínguez (2008) studied discourse

[2] Other studies such as Fung and Carter (2007) further distinguish a *structural function* and a *cognitive function* (e.g. to express hesitation or to indicate that the speaker is thinking).

markers by means of interviews with three learners who were studying in Spain and who had an advanced level (C1, *CEFR*).

Nogueira (2011) analysed the oral production of more than 20 Brazilian learners from upper-intermediate level (B2, *CEFR*), who performed a narrative task. The aim of Nogueira's study was to compare their use of DMs with those introduced in textbooks for learning Spanish. De la Fuente (2009) researched the influence of the type of instruction (explicit and implicit focus on form) in the acquisition of Spanish DMs by 24 English university learners. Hernández (2008, 2011) conducted similar research on the influence of explicit instruction and input flood on the acquisition of DMs by English learners of Spanish. Results from both studies suggested the positive influence of the explicit instruction on the acquisition of these markers. Finally, there is a study by Domenech (2008) on coordinate and subordinate connectors by Arabic learners (in oral and written tasks). Given that this study departed from a syntactic approach, we cannot strictly relate it to the pragmatic level.

A difficulty for research on discourse markers is the heterogeneity of methods used by each researcher in gathering the data: e.g. narrative tasks, interviews, consciousness raising tasks, reading aloud protocols, or questionnaires. This fact can make it awkward in comparing results among different studies.

Moreover, approaches for the analyses differ among investigators. While most studies are performed using a quantitative approach, other pieces of research are undertaken within a more descriptive or qualitative framework (Walsh 2013, for example, discusses this subject). It may be that the combination of both quantitative and qualitative approaches proves to be more beneficial than either one on its own.

Another hindrance for research in this area is the lack of a common taxonomy of DMs. Among the different classifications of Spanish DMs, we will mainly follow Martín Zorraquino and Portolés' classification (1999), a five-class typology that includes most of the DMs analysed in our study. In fact, it seems that discourse markers are sometimes considered a *wild card* category where different particles meet depending on the interests of each researcher (see, for example, the discussion by Fung and Carter 2007: 411; and by Aijmer and Simon-Vandenbergen 2011).

Thus, the research literature published on this topic should be read cautiously regarding the generalisation of results to different settings or type of learners.

4 Research Design

Our research is framed in the Learner Corpus Research (LCR) (Granger 2012). In particular, we followed a corpus-based approach (McEnery et al. 2006), since we used part-of-speech (henceforth, POS) annotation and frequency lists as research methods. As we have previously pointed out, corpus methodology is used more and more to tackle questions from the L2 acquisition research. Improvements in computing technology have made corpora progressively more available and easy to use. For English there already exist important learner corpus projects: e.g. the

International Corpus of Learner English (ICLE; Granger et al. 2009) and the Louvain International Database of Spoken English Interlanguage (LINDSEI; Gilquin et al. 2010). With regard to the Spanish language, there are at present resources such as the CEDEL corpus of written texts (Lozano and Mendikoetxea 2011). As for spoken skills, there is a longitudinal corpus of interviews (Díaz Rodríguez 2007), and the Spanish Learner Language Oral Corpus (SPLLOC; Mitchell et al. 2008).

While many research projects in the field of Spanish L2 have used both written and oral empirical data to test a hypothesis (Lafford and Salaberry 2003), corpus methods aim at developing a balanced and representative collection that could be re-employed for different topics of research. That is the reason why a key issue in the corpus paradigm is the design stage of the data collection (Granger 2008). Below, we will first describe the participants in our project; secondly, the design of the corpus; thirdly, the data elicitation methods; and finally, the procedure of the data analysis.

4.1 Participants and Corpus Design[3]

The participants were learners studying in Madrid thanks to international exchange programs (e.g. Erasmus), and almost all of them were between 19 and 26 years old, studying at an undergraduate or postgraduate level.[4] Our framework is limited to the use of Spanish in an academic context.

Unlike other research studies, our aim was to obtain spoken samples from several language backgrounds. A total of 40 (N = 40) subjects were interviewed, with four learners for every mother tongue. The participants were native speakers of nine different languages, which range from the Romance languages (Italian, French and Portuguese), the Germanic languages (English, German and Dutch) and the Slavic languages (Polish), to other Asian languages such as Chinese or Japanese. There was also another mixed group of four learners with other language backgrounds (one Finnish, one Korean, one Turkish and one Hungarian). Regarding the students' proficiency level, half of the learners (N = 20) were registered at the elementary level (A2), and the other half (N = 20) at the threshold level (B1) according to the *CEFR*.

Separately, a group of four native speakers (N = 4) with a similar educational level and age were interviewed with the same elicitation methods (control group).

[3] Further information about the corpus can be found at: http://cartago.lllf.uam.es/corele/index.html

[4] With respect to their learning background, some of the participants in our study had received formal instruction in Spanish in their countries, and others started in Madrid, combining formal instruction with natural-context acquisition. All of them had been in an immersion context for several months before being interviewed.

Table 1 Summary of data

Group	Level	N	DMs	% out of LUs	Lexical units	Mean	SD	Min	Max	Median	Range
Learners	A2	20	726	2.76	26,317	36.300	26.57	9	128	27	119
	B1	20	955	3.62	26,371	47.750	28.70	11	127	38	116
	Total:		**1,681**	**3.19**	52,688	**42.025**	27.91				
Native		4	**582**	**6.76**	8,610	**145.500****	69.73	82	241	129	159

N number of subjects, *DMs* frequency of discourse markers, *% out of LUs* ratio per lexical unit, *SD* standard deviation, *Min* minimum frequency, *Max* maximum frequency
**Results are significant at $p = 0.0031$

These interviews were used as a benchmark to compare the use of DMs between native and non-native speakers and to unveil certain phenomena that were more frequent in the spoken discourse in both groups.

Each interview lasted about 15–20 min, so approximately more than 1 h was recorded for every group of participants. The recordings were made at the Universidad Complutense and the Universidad Autónoma de Madrid during the academic courses 2007–2008, 2008–2009, and 2009–2010. On the whole, more than 13 h were collected for the learner corpus, which comprised a total of over 55,000 tokens, whereas the control group exceeded 9,000 tokens (in this count, a token was considered as the word between two white spaces, so the discourse marker *es decir,* 'that is to say', counted as two tokens). With a view to normalising the data and achieving more precision when analysing the lexis, we performed another count by lexical unit (with multiwords such as *por otra parte,* 'on the other hand', or *o sea,* 'that is', counting as one unit). Following these criteria, the learner corpus amounted to 52,688 lexical units, and the native speakers' group, to 8,610 (Table 1).

4.2 Elicitation Methods

The interview, carried out by the researcher for the learners and the native speakers, was voluntary and, learners had the opportunity to receive an explanation of their errors after its completion. The interview began with a brief introduction (e.g. providing information such as their studies, years studying Spanish, time in a Spanish-speaking country). Subsequently, in order to obtain comparable data, they all accomplished the same tasks, which were similar to those in foreign language examinations:

- Providing a description of two photographs (in this project, related to food).
- Retelling a story from pictures, to make learners use the past tense and to ask them a question involving two speech acts (a suggestion and a request).

The last part of the interview was more spontaneous, and the researcher asked the learners to give opinions about non-personal topics (in our interview, related to

food; e.g. changes in food habits in present-day society, or differences between food in Spain and that from the learner's country).

4.3 Procedure and Data Analysis

After the recording, interviews were manually transcribed, and the transcriptions were subsequently POS-tagged with a view to studying and describing the learners' spoken production of the different morphological categories. POS tagging was partially automated with GRAMPAL, a morphological processor of Spanish that was adapted for spoken data (Moreno and Guirao 2006). GRAMPAL handles multiword units (such as *gracias a,* 'thanks to'), and it is also able to tag discourse markers, which are often multiword units: e.g. *por otro lado* ('on the other hand'), *quiero decir* ('I mean'), or *por eso* ('because of that').

A manual revision of the tagging was performed so as to correct ambiguities concerning the automatic assignment of categories. Some ambiguities were due to homonymy: e.g. *vale,* which can be a noun ('voucher') or a discourse marker ('OK'). Other ambiguities were due to incorrect categorization: e.g. *bueno*, which can be an adjective in *hombre bueno* ('good man') or a discourse marker in *¡Hombre! Bueno…* ('Hey! Well…'). Finally, other ambiguities arose in phrase chunking: e.g. *es decir* ('that is'), which is not a discourse marker in *Lo que hace es decir tonterías* ('What he is doing is saying silly things').

After the POS-tagging process, a list of DMs was obtained for each group of learners, as well as for the group of native speakers. In this count, we did not take into consideration markers from other languages that were found in learner's discourse as a result of interference phenomena.[5] For example, in the following utterance, the conversational marker *bem* (from Portuguese) was used instead of *bueno,* 'well' (we show a simplified version of the transcriptions):

JUS: ***bom*** *yo creo que esta é una paella*	(PORWA2_2)
'well I think that this is a paella'	

Among the learners from the A2 level, nine DMs interfered while they were speaking in Spanish: *bom* (twice), *like* ('como', twice), *por exemplo* ('for example', twice), *ben* ('right', instead of Spanish *bien*, once) and *é que* ('it is just that', once). Among the learners at the B1 level, this interference decreased slightly, since only four non-Spanish DMs were registered: *like* (two times), *I mean* ('quiero decir'), and *bon* ('right').

[5] We decided to include *OK* in our study because it is used more frequently with the same meaning as 'de acuerdo' or 'vale' by native speakers, due to the influence of English. Notwithstanding that fact, *OK* is not found among the twenty-five most frequent discourse markers used by native speakers (Table 3).

Similarly, we did not include in this count misformations due to interference or false hypotheses about the right form. For example, in the following utterance the right connector should have been *por otro lado*:

EVE: creo que ***en el otro lado*** *es un costumbre bueno*	(DUTWB1)
'on the other hand, I think that it is a good habit'	

Only six non-existent DMs were documented in our data (three at the A2 level and three at the B1 level).

5 Results

Considering the above-mentioned criteria, a total of 1,681 DMs were registered in the learner group. The learners at the A2 level produced 726 DMs, with a mean of 36.30 markers and about 2.76 markers for every one hundred lexical units. At the B1 level this amount slightly increased to 955 in our data, with a mean of 47.75 markers per interview and approximately 3.62 every one hundred lexical units (Fig. 1 and Table 1). In turn, the native speakers produced 582 DMs, with a higher mean per interview (145.50) when compared to that of the learners', and a higher ratio of DMs per lexical unit (6.76 %).

The native speakers' production of DMs varied among the four participants interviewed, as the standard deviation measure and the range was greater than that of the groups of learners. This would imply that the mean value of DMs produced by the native speakers was not as reliable as the mean of the learners. In order to gain further insight into the difference in the use of discourse markers, we performed a statistical test comparing their frequency among the three groups. As the distribution of the data did not pass any test of normality (Kolmogorov-

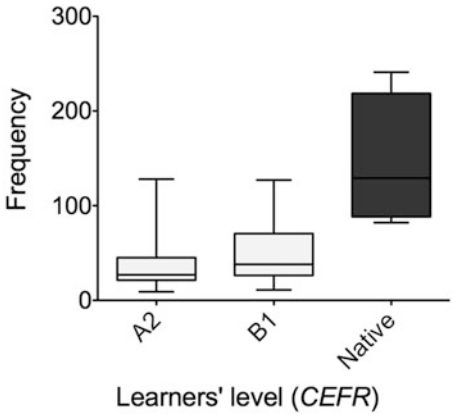

Fig. 1 Distribution of markers across levels and comparison with native speakers. The *outer lines* of the plot represent the maximum and minimum values in each group

Table 2 Summary of data

Group	L1	N	Disc. mark.	% out of LUs	Lexical units	Mean	SD	Min	Max	Range
Learners	Portuguese	4	194	2.64	7,329	48.500	22.43	25	71	46
	Italian	4	195	3.18	6,133	48.750	31.56	21	79	58
	French	4	349	5.44	6,413	87.250	46.54	44	128	84
	English	4	186	3.93	4,731	46.500	27.29	11	77	66
	German	4	133	2.91	4,573	33.250	15.96	17	55	38
	Dutch	4	153	2.95	5,186	38.250	12.12	26	51	25
	Polish	4	109	2.06	5,300	27.250	8.61	22	40	18
	Chinese	4	72	2.00	3,592	18.000	5.47	14	26	12
	Japanese	4	163	3.49	4,664	40.750	27.74	11	78	67
	Other	4	127	2.66	4,767	31.750	15.69	9	45	36
	Total	40	**1,681**	**3.19**	52,688	**42.025**	27.89	9	128	118
Native		4	**582**	**6.76**	8,610	**145.500**	69.92	82	241	159

L1 learners' mother tongue, *N* number of subjects, *Disc. mark.*, frequency of discourse markers, *% out of LUs* ratio of discourse marker per lexical unit (in each group), *SD* standard deviation, *Min* minimum frequency, *Max* maximum frequency

Smirnoff and D'Agostino), a non-parametric test was chosen. Notwithstanding the aforementioned variance, significant differences among the A2, the B1 and the group of native speakers were revealed (Kruskal-Wallis: $H = 11.57$, 2df, $p = 0.0031$). In contrast, we did not find significant differences between the A2 and the B1 group (Mann-Whitney, one-tailed: $p = 0.0648$).

With regard to the learner's L1, we observed that Chinese learners in our data produced the lowest number of discourse markers (with a mean of 18 per interview, and a ratio of two for every one hundred lexical units; Table 2). On the contrary, the French learners were the group that used more DMs (a mean of 87.25, and a ratio of 5.44 %). Nevertheless, the variance within the French group was also quite high ($SD = 46.54$, with a minimum of 44 DMs in one interview but a maximum of 128 in another). Due to the scarce number of participants in each L1 group, we did not perform further statistical analyses.

The difference between the Chinese group and the learners whose L1 is closer to Spanish (such as French, Portuguese, and Italian) could reveal certain difficulties in discourse and pragmatic skills in relation to the Spanish language. However, the typological explanation is not coherent with the high frequency of discourse markers used by other learners whose L1 is not Indo-European. An example in our data is Japanese learners, who produced a mean of 40.75 per interview (with a ratio of 3.49 %), even higher than that of the native speakers of Polish (an Indo-European language), who showed a mean of 8.61 markers (a ratio of 2.06 %).

The differences between the equivalence of discourse markers in the L1 and the L2 could better account for some of these results. For example, some researchers (Liao 2009: 1320) have confirmed the underuse of certain English discourse markers by Chinese learners due to the lack of a similar counterpart in Mandarin

(e.g. *well, I mean*).[6] This aspect deserves to be studied further by analysing the pragmatic and semantic correspondence of the markers in each language through Aijmer and Simon-Vandenbergen's model (2004).

In addition, extralinguistic factors could also explain these results. Apart from the fact that Chinese participants in our study showed a low fluency during the interview, teaching-related factors can possibly justify the underuse of these particles. Teaching methodologies of English or French language have been including the explicit instruction of discourse markers for the last decade. Interestingly, learners whose L1 is French or English showed a high rate of DMs. We are inclined to think that these learners acquired a good command of these devices when they learnt a language such as English or French, and they benefited from this experience when learning an L3 (e.g. Spanish), unlike Chinese students. Other variables such as the learning background of Spanish (e.g. in an academic setting or in a study abroad program), the time of exposure to the L2, or the discourse abilities of each learner (his/her idiolect) should also be taken into account in future studies. Given the scarce number of participants in our study, conclusions cannot be generalised.

Regarding the frequency of use of each discourse marker, Table 3 shows a breakdown of the 25 most frequent particles in our study (grouped by learners and native speakers). Among the 25 most frequent DMs, 16 were produced by both native and non-native participants. Most of these markers belong to the following types according to Martín Zorraquino and Portoles' typology (1999): connectors (*pues, entonces*), reformulators (*o sea*), argumentative operators (*por ejemplo*), and conversational markers (*bueno, vale, vamos, claro, oye, por favor, mira, oye, ¿no?*). The other markers produced by both learners and native speakers (*como, es que, por eso, a ver*) are not included in that typology. It is interesting that there are no information structurers among them (only the learners used *primero*, 'first', which introduces a set of points or ideas).

Nevertheless, the learners did not produce some conversational markers (*nada*, *hombre, vamos, efectivamente*), certain contra-argumentative connectors (*sin embargo*), specific additive connectors (*y demás, y tal, etcétera*) or reformulators (*o sea que,* though they used *o sea*). These results could be explained by the elicitation tasks employed in the study. These tasks were mainly of descriptive and narrative nature and did not favour the different functions associated with the mentioned markers. Some particles (e.g. *hombre* or *nada*[7]) are frequently used by

[6] Yao (2012) compared Spanish and Chinese discourse markers and explained some differences among them. In Chinese, modal particles do not convey a conceptual meaning, and, unlike Spanish, these elements can be used as discourse markers. In addition, most discourse markers in Chinese appear at the beginning of the utterance. In Spanish, discourse markers can occur either at the beginning, the middle or the end of the utterance. Further information on Chinese DMs can be found in Liao (1986), Miracle (1991), and Liu (2009).

[7] These conversational markers have received little attention in studies regarding their description or teaching (Gozalo 2008; Gaviño 2011; Schmer 2012).

Table 3 The twenty-five most frequent markers (Learners and native speakers)

Learners (N = 40)			Native speakers (N = 4)		
Discourse marker	Freq.	Rel. freq. (%)	Discourse marker	Freq.	Rel. freq. (%)
Vale	283	16.84	*Pues*	193	33.16
Como	192	11.42	*Bueno*	104	17.87
Pues	141	8.39	*Es que*	33	5.67
Es que	135	8.03	*O sea*	28	4.81
Bueno	133	7.91	*¿No?*	27	4.64
Entonces	117	6.96	*Claro*	26	4.47
Por eso	117	6.96	*Vale*	25	4.30
Ok	95	5.65	*Entonces*	20	3.44
Claro	94	5.59	*Pero bueno*	15	2.58
¿No?	72	4.28	*Por ejemplo*	13	2.23
Por ejemplo	60	3.57	*Como*	11	1.89
Por favor	51	3.03	*A ver*	9	1.55
¿Sabes?	37	2.20	*Hombre*	6	1.03
O sea	27	1.61	*Y tal*	6	1.03
Es como	22	1.31	*Efectivamente*	4	0.69
Bien	21	1.25	*Etcétera*	4	0.69
A ver	10	0.59	*Mira*	4	0.69
Por supuesto	10	0.59	*Por favor*	4	0.69
Primero	9	0.54	*Vamos*	4	0.69
En fin	5	0.30	*Y demás*	4	0.69
Mira	5	0.30	*Nada*	3	0.52
Luego	4	0.24	*O sea que*	3	0.52
Oye	3	0.18	*Oye*	3	0.52
Pero bueno	3	0.18	*Por eso*	3	0.52
Quiero decir	3	0.18	*Sin embargo*	3	0.52

native speakers and therefore present in the input received by learners. However, these markers are not usually found in pedagogic materials or considered in formal instruction.

If we consider the 20 most frequent markers used by the learners (see Table 4), production was similar at both the A2 and B1 levels. Additionally, comparing Tables 3 and 4, we observed that the B1 learners produced seven of the ten most frequent markers in the native group (*pues, bueno, es que, ¿no?, claro, vale, entonces*), while the A2 learners just five (*pues, bueno, es que, claro, entonces*).

Table 5 breaks down the distribution across interviews of the ten most frequent discourse markers. According to our data (Fig. 2), when compared to the A2 group, we identified a spread of use of almost every marker in the interviews of the B1 group. The only exceptions were the following: *OK* and *por eso* (with a lower distribution), and *entonces* and *pues* (with the same distribution). These results suggest a positive progress in the acquisition of DMs.

Table 4 The twenty most frequent markers used by learners (A2 and B1 level)

A2 (N = 20)			B1 (N = 20)		
Discourse marker	Freq.	Rel. freq. (%)	Discourse marker	Freq.	Rel. freq. (%)
Vale	130	17.91	*Vale*	153	16.02
Pues	100	13.77	*Como*	115	12.04
Como	77	10.61	*Bueno*	111	11.62
Entonces	62	8.54	*Es que*	74	7.75
Es que	61	8.40	*Entonces*	55	5.76
Por eso	53	7.30	*¿No?*	52	5.45
Ok	50	6.89	*Claro*	50	5.24
Claro	44	6.06	*Ok*	45	4.71
Por ejemplo	25	3.44	*Pues*	41	4.29
Por favor	24	3.31	*Por eso*	64	6.70
Bueno	22	3.03	*Por ejemplo*	35	3.66
¿No?	20	2.75	*O sea*	25	2.62
Es como	17	2.34	*Por favor*	27	2.83
¿Sabes?	16	2.20	*¿Sabes?*	21	2.20
Por supuesto	6	0.83	*Bien*	18	1.88
Bien	3	0.41	*A ver*	10	1.05
O sea	2	0.28	*Primero*	9	0.94
Pero bueno	2	0.28	*Es como*	5	0.52
Por otra parte	2	0.28	*En fin*	4	0.42
En fin	1	0.14	*Luego*	4	0.42

Table 5 Distribution across interviews from each level (*CFER*) of the ten most frequent discourse markers produced by the learners in our study

Discourse markers	A2 (N = 20)		B1 (N = 20)	
	Distribution	Frequency (% out of 20)	Distribution	Frequency (% out of 20)
Vale	17	85	19	95
Como	12	60	15	75
Entonces	12	60	12	60
Claro	11	55	12	60
¿No?	10	50	11	55
Por eso	10	50	9	45
Es que	9	45	11	55
Pues	9	45	9	45
OK	8	40	4	20
Bueno	3	15	6	30

6 Discussion and Pedagogical Implications

The types of markers produced by the learners and native speakers were very similar. The number of information structurers or contra-argumentative connectors was scarce, and the most frequent were conversational markers (*vale, bueno, ¿no?, claro*) and consecutive connectors (*por eso, pues, entonces*). This could be

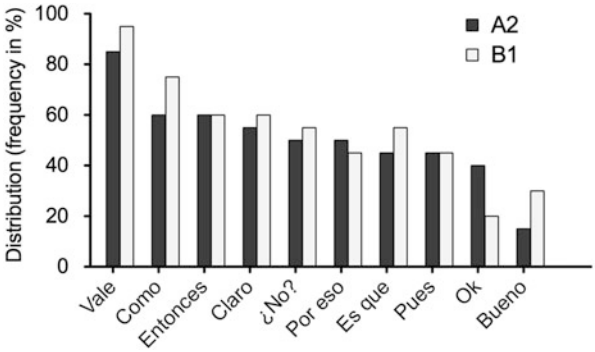

Fig. 2 Distribution across interviews (expressed as percentage) of the ten most frequent discourse markers used by learners at A2 and B1

explained by the elicitation method and also by the natural context of L2 acquisition: markers that are frequently used by native speakers configure the input received by learners.

In the distribution across interviews we observed that, among the ten most frequent markers, there were mainly connectors (*entonces, por eso, pues*) and conversational markers (*vale, claro, no, bueno, OK*). As before, we also found in these data two markers that have received less attention in theoretical studies. An example is *como,* which has traditionally been considered a hedge or an expletive. Another example is *es que,* which some scholars classify as an argumentative connector that introduces an explanation or a justification (Fernández Leborans 1992; Pons Bordería 1998; Porroche Ballesteros 1998; Santos Río 2003). These markers are not included in the classification offered in Martín Zorraquino and Portolés (1999).

As suggested above, some interesting trends in the use of discourse markers were observed when comparing both proficiency levels. For example, the conversational marker *bueno* was not one of the ten most frequent markers at level A2, but it was the third most frequent marker at level B1 (Table 4). Its use by the learners seemed to be very close to the native production in which it was the second most frequent marker (Table 3). If we consider its distribution (Table 5), *bueno* was used in six interviews at level B1, twice the rate used at level A2. Curiously, the occurrence of *bueno* was markedly reduced in the learners' production when it appeared combined with the conjunction *pero*. According to Table 3, *pero bueno* was registered 15 times among the native speakers, in contrast to only three instances among the learners.

The case of the marker *OK* is also interesting. Its occurrence in the groups of learners was greater at the A2 level and by speakers whose L1 was English. The interlinguistic interference could explain these data. At the B1 level, its use was reduced in favor of the Spanish equivalent *vale*.

The distribution of markers such as *pues, entonces,* and *por eso* across participants did not only increase (in the case of the first two particles), but was reduced at B1 (Table 5). This fact could be attributed to the learner's idiolect (especially in the case of *por eso*). Some students showed a tendency to employ some markers with an exclusively expletive function.

Table 6 Discourse markers proposed in the *PCIC* registered among the most frequent in our data

Level	Discourse markers (*PCIC*)
A1	*Por ejemplo, ¿no?*
A2	*Por eso, entonces, primero, mira*
B1	*Como, es que, pues, o sea, claro, ¿sabes?*
B2	*Por supuesto, en fin, bueno*

However, although the statistical test did not yield significant results when comparing the two levels (Table 1), frequency and distribution analyses indicated a slightly progressive increase in the production of markers from the A2 level to the B1 level.

Following this, we can compare our data with the established guidelines for the teaching of Spanish as a second language (*Plan Curricular del Instituto Cervantes,* from here on, *PCIC*; Instituto Cervantes 2006) (Table 6). According to this reference document, the marker *¿no?* should be introduced at A1, *oye*/*oiga* and *mira*/*mire* at A2, and *¿sabes?* at B1. Indeed, learners in our data did use them at these levels (excepting *oye*/*oiga*). Further, they did use most of the markers for the A1-B1 levels, despite not having produced instances of some particles established for B1. The non-native speakers even used some particles corresponding to B2.

To sum up, in general, learners in our study produced a wide range of markers established in the guidelines for Spanish language teaching. Nevertheless certain DMs, which were found in the native oral discourse, were not produced by the learners. Therefore we consider that more effort should be made in the teaching of those markers. Pragmatics is teachable and a majority of studies favor explicit approaches to the teaching of L2 pragmatics (Koike 1989; Bouton 1999; House 1996; Kasper 1997; Kasper and Rose 2002), and particularly of DMs (Yoshimi 2001; Hernández 2008, 2011; Hernández and Rodríguez-González 2012). Explicit teaching is especially important at low proficiency levels for which pedagogic materials offer exclusively implicit teaching, and in foreign language environment, because there are fewer learning opportunities outside the classroom.

The instruction can be carried out from the lowest levels of learning (Kasper 1997) and we suggest following the three I's methodology: Illustration, Interaction and Induction (Carter and McCarthy 1995). This methodology involves basic principles such as the exposition to authentic input, the activation of acquired knowledge, and consciousness-raising procedures. The three stages should be complemented with communicative practice and corrective feedback in order to improve the effectiveness of the teaching of DMs.

7 Concluding Remarks

In this work we have presented a corpus-based description of the use of discourse markers in native and non-native oral production at intermediate levels. Results of the analysis can be summarised as follows. With regard to our first research question—i.e. what markers are the most frequent in the oral discourse of learners at A2 and B1

levels?—we found a wide range of Spanish DMs in the oral production of learners at these levels of proficiency. However, the statistical tests did not reveal significant results when comparing the A2 and B1 groups. Even so, a progress in the acquisition was observed. Not only the number of markers increased from A2 to B1, but also the distribution of markers across participants spread as well.

Regarding the second research question—i.e. what markers are the most frequent in the oral discourse of learners with different L1?—we observed dissimilarities in the production of these particles, both in number and type of marker. The explanation of these results may be to a certain extent interlinguistic. For example, the equivalence of markers in learners' first language can influence the usage of these particles, as we found among our Chinese learners. Additionally, other factors such as the type of instruction received by each learner or his/her learning context (e.g. in an academic setting or in a study abroad program) may also be involved in the learners' use and acquisition of DMs.

As for the third research question—i.e. what differences exist in the use of discourse markers in native and non-native discourse?—the statistical tests yielded significant results when comparing the native and non-native speakers groups (the frequency of markers used by the learners is significantly lower). The types of markers used by the learners were also similar to those used by the native speakers, probably due to the elicitation task. In addition, the learners produced most of the markers proposed by the teaching guidelines for Spanish language established in the *PCIC* for the A2 and B1 levels.

Further analyses with data from more participants could confirm these results. As well, further research should be carried out regarding qualitative aspects of production. An analysis of the usage context of each marker is necessary to check if learners use these particles accurately, and what functions pose a problem in the acquisition of polyfunctional markers (e.g. *bueno* or *pues*). Likewise, the semantic and pragmatic differences between equivalent discourse markers in each language deserve to be closely studied following the model proposed by Aijmer and Simon-Vandenberger (2004). These data, along with those from further studies on the degree of influence of the explicit teaching on the acquisition of discourse markers, will certainly help guide the explicit teaching of these devices.

Acknowledgments This work was supported by the Regional Government of Madrid and the European Social Fund under a PhD grant; and the Comisión Interministerial de Ciencia y Tecnología (CICYT) [project TIN2010-20644-C03-03]. The authors gratefully acknowledge the anonymous reviewers for their valuable comments, and would like to thank T. Shipley and C. Herrero for their assistance in the linguistic revision of the manuscript.

References

Aijmer, K. (2002). *English discourse particles: Evidence from a corpus*. Amsterdam/Philadelphia: John Benjamins.

Aijmer, K. (2004). Pragmatic markers in spoken interlanguage. *Nordic Journal of English Studies, 3*(1), 173–190.

Aijmer, K. (2011). *Well I'm not sure I think*... The use of *well* by non-native speakers. *International Journal of Corpus Linguistics, 16*(2), 231–254.

Aijmer, K., & Simon-Vandenbergen, A.-M. (2003a). The discourse particle *well* and its equivalents in Swedish and Dutch. *Linguistics, 41*(6), 1123–1161.

Aijmer, K., & Simon-Vandenbergen, A.-M. (2003b). The expectation marker *of course* in a cross-linguistic perspective. *Languages in Contrast, 31*, 13–43.

Aijmer, K., & Simon-Vandenbergen, A.-M. (2004). A model and a methodology for the study of pragmatic markers: The semantic field of expectation. *Journal of Pragmatics, 36*, 1781–1805.

Aijmer, K., & Simon-Vandenbengen, A.-M. (Eds.). (2006). *Pragmatic markers in contrast*. Oxford: Elsevier.

Aijmer, K., & Simon-Vandenbergen, A.-M. (2011). Pragmatic markers. In J. Zienkowski, J.-O. Östman, & J. Verschueren (Eds.), *Discursive pragmatics* (Handbook of Pragmatic Highlights, Vol. 8, pp. 223–247). Amsterdam/Philadelphia: John Benjamins.

Bachman, L. (1990). Habilidad lingüística comunicativa. In M. Llobera (Ed.), *Competencia comunicativa* (pp. 105–129, 1995). Madrid: Edelsa.

Bardovi-Harlig, K. (2013). Developing L2 pragmatics. *Language Learning, 63*(1), 68–86.

Bardovi-Harlig, K., & Mahan Taylor, R. (Eds.). (2003). *Teaching pragmatics*. Washington, DC: United States Department of State.

Blakemore, D. (1987). *Semantic constraints on relevance*. Oxford: Blackwell.

Blakemore, D. (1992). *Understanding utterances*. Oxford: Blackwell.

Blakemore, D. (2004). Discourse markers. In L. R. Horn & G. Ward (Eds.), *Handbook of pragmatics* (Blackwell handbooks in linguistics, pp. 221–240). London: Blackwell.

Blum-Kulka, S., & House, J. (1989). Cross cultural and situational variation in requesting behavior. In S. Blum-Kulka, J. House, & G. Kasper (Eds.), *Cross-cultural pragmatics: Requests and apologies* (pp. 123–154). Norwood: Ablex.

Bouton, L. (1999). Developing non-native speaker skills in interpreting conversational implicatures in English: Explicit teaching can ease the process. In E. Hinkel (Ed.), *Culture in second language teaching and learning* (pp. 47–70). Cambridge: Cambridge University Press.

Briz, A. (1993a). Los conectores pragmáticos en español coloquial. I: Su papel argumentativo. *Contextos, 11*, 145–188.

Briz, A. (1993b). Los conectores pragmáticos en la conversación coloquial. II. Su papel metadiscursivo. *Español actual, 59*, 39–56.

Buysse, L. (2007). Discourse marker *so* in the English of Flemish university students. *Belgian Journal of English Language and Literatures,* New Series *5*: 79–95.

Buysse, L. (2011). The business of pragmatics: The case of discourse markers in the speech of students of business English and English linguistics. *ITL: International Journal of Applied Linguistics, 161*, 10–30.

Buysse, L. (2012). *So* as a multifunctional discourse marker in native and learner speech. *Journal of Pragmatics, 44*, 1764–1782.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1–47.

Carter, R., & McCarthy, M. (1995). Grammar and the spoken language. *Applied Linguistics, 16*(2), 141–158.

Casado Velarde, M. (1998). Lingüística del texto y marcadores del discurso. In M. A. Martín Zorraquino & E. Montolío Durán (Eds.), *Marcadores del discurso. Teoría y análisis* (pp. 55–70). Madrid: Arco/Libros, S.L.

Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). A pedagogical framework for communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics, 6*(2), 5–35.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

De la Fuente, M. J. (2009). The role of pedagogical tasks and focus on form on acquisition of discourse markers by advanced language learners. In R. P. Leow, H. Campos, & D. Lardiere

(Eds.), *Little words: Their history, phonology, syntax, semantics, pragmatics and acquisition* (pp. 211–221). Washington, DC: Georgetown University Press.
Díaz Rodríguez, L. (2007). *Interlengua española: estudio de casos*. Barcelona: Printulibro Intergrup.
Díez Domínguez, P. (2008). *Análisis de los marcadores discursivos más usados en el habla de tres estudiantes extranjeros de E/LE nivel C1 tras una estancia de un mes en España*. Master's thesis, University of Barcelona, Barcelona.
Domenech Val, A. (2008). *La conectividad en español no nativo: un estudio sobre hablantes arabófonos*. Master's thesis, University of Barcelona, Barcelona.
Escandell, V. (2004). Norms and principles. Putting social and cognitive pragmatics together. In R. Márquez-Reiter & M. E. Placencia (Eds.), *Current trends in the pragmatics of Spanish* (pp. 347–371). Amsterdam/Philadelphia: John Benjamins.
Fernández Leborans, M. J. (1992). La oración del tipo 'es que…'. *Verba, 19*, 223–239.
Fernández López, S. (1990). Problemas discursivos en la interlengua de aprendices de español como lengua extranjera. In S. Montesa Peydró, & A. M. Garrido (Eds.), *Proceedings of the second national ASELE conference* (pp. 267–278, 1994). Málaga: ASELE.
Fernández López, S. (1997). *Interlengua y análisis de errores en el aprendizaje del español*. Madrid: Edelsa.
Fraser, B. (1996). Pragmatic markers. *Pragmatics, 6*(2), 167–190.
Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics, 31*, 931–952.
Fuller, J. M. (2003). Discourse marker use across speech contexts: A comparison of native and non-native speaker performance. *Multilingua, 22*, 185–208.
Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics, 28*(3), 410–439.
Galindo Merino, M. (2005). La transferencia pragmática en el aprendizaje de ELE. In A. Álvarez et al. (Eds.), *Proceedings of the XVI international ASELE conference* (pp. 289–297). Oviedo: Ediciones de la Universidad de Oviedo
Gaviño, V. (2011). Operaciones metalingüísticas del marcador discursive *hombre*. *Marcoele, 12*. http://marcoele.com/descargas/12/gavino-marcador_hombre.pdf. Accessed 9 Aug 2013.
Gilquin, G. (2008). Hesitation markers among EFL learners: Pragmatic deficiency or difference? In J. Romero-Trillo (Ed.), *Pragmatics and corpus linguistics. A mutualistic entente* (Mouton Series in Pragmatics, Vol. 2, pp. 119–149). Berlin: Mouton de Gruyter.
Gilquin, G., De Cock, S., & Granger, S. (2010). *The Louvain international database of spoken English interlanguage*. Louvain-la-Neuve: Presses Universitaires de Louvain.
Gozalo, P. (2008). El marcador discursivo hombre y su tratamiento en el aula de E/LE. *REDELE, 14*. www.doredin.mec.es/documentos/00820083000393.pdf. Accessed 3 Sept 2013.
Granger, S. (2008). Learner corpora. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (Vol. 1, pp. 259–275). Berlin: Walter de Gruyter.
Granger, S. (2012). How to use foreign and second language learner corpora. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 7–29). London: Blackwell.
Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (2009). *International corpus of learner English V2*. Louvain-la-Neuve: Presses universitaires de Louvain.
Hellermann, J., & Vergun, A. (2007). Language which is not taught: The discourse marker use of beginning adult learners of English. *Journal of Pragmatics, 39*(1), 157–179.
Hernández, T. A. (2008). The effect of explicit instruction and input flood on students' use of Spanish discourse markers on a simulated oral proficiency interview. *Hispania, 91*, 665–675.
Hernández, T. A. (2011). Re-examining the role of explicit instruction and input flood on the acquisition of Spanish discourse markers. *Language Teaching Research, 15*(2), 159–182.
Hernández, T. A., & Rodríguez-González, E. (2012). Impact of instruction on the use of L2 discourse markers. *Journal of Second Language Teaching and Research, 2*(1), 3–31.
House, J. (1996). Developing pragmatic fluency in English as a foreign language: Routines and metapragmatic awareness. *Studies in Second Language Acquisition, 18*(2), 225–252.
Hymes, D. (1972). Acerca de la competencia comunicativa. In M. Llobera (Ed.), *Competencia comunicativa* (pp. 27–47, 1995). Madrid: Edelsa.

Instituto Cervantes. (2006). *Plan Curricular del Instituto Cervantes*. http://cvc.cervantes.es/ensenanza/biblioteca_ele/plan_curricular/. Accessed 3 Sept 2013.

Jung, J. Y. (2002). Issues in acquisitional pragmatics. *Working Papers in TESOL & Applied Linguistics, 2*(3), 1–13.

Kasper, G. (1992). Pragmatic transfer. *Second Language Research, 8*(3), 203–231.

Kasper, G. (1996). Interlanguage pragmatics in SLA (Introduction to thematic issue). *Studies in Second Language Acquisition, 18*(2), 145–148.

Kasper, G. (1997). *Can pragmatic competence be taught?* Honolulu: University of Hawai'i, Second Language Teaching & Curriculum Center. www.nflrc.hawaii.edu/NetWorks/NW6/. Accessed 3 Sept 2013.

Kasper, G., & Rose, K. (2002). *Pragmatic development in a second language*. Malden: Blackwell.

Koike, D. A. (1989). Pragmatic competence and adult L2 acquisition: Speech acts in interlanguage. *The Modern Language Journal, 73*, 279–289.

Koike, D. A. (1996). Transfer of pragmatic competence and suggestions in Spanish foreign language learning. In S. M. Gass & J. Neu (Eds.), *Speech acts across cultures: Challenges to communication in a second language* (pp. 257–281). Berlin: Mouton de Gruyter.

Lafford, B. A., & Salaberry, M. R. (2003). *Spanish second language acquisition: State of the science*. Washington, DC: Georgetown University Press.

Langer, B. (2001). Teaching pragmatics forms in Spanish. *Segundas lenguas e inmigración en red, 5*, 5–34.

Liao, Q. (1986). Discourse connectives in modern Mandarin Chinese. *Zhongguo Yuwen, 6*, 413–427.

Liao, S. (2009). Variation in the use of discourse markers by Chinese teaching assistants in the US. *Journal of Pragmatics, 41*, 1313–1328.

Liu, B. (2009). Chinese discourse markers in oral speech of mainland Mandarin speakers. In Y. Xiao (Ed.), *Proceedings of the 21st North American Conference on Chinese Linguistics (NACCL-21)* (Vol. 2, pp. 358–374). Smithfield: Bryant University. Available at: https://naccl.osu.edu/sites/naccl.osu.edu/files/NACCL-21_Vol._2–Binmei_Liu–pp._358-374.pdf. Accessed 27 Aug 2013.

Llinares-García, A., & Romero-Trillo, J. (2006). Discourse markers in the EFL classroom. In M. Carretero, L. Hidalgo Downing, J. Lavid, E. Martínez-Caro, J. Neff, S. Pérez de Ayala, & E. Sanchez-Pardo (Eds.), *A pleasure of life in words: A festschrift for Angela Downing* (Vol. 2, pp. 163–176). Madrid: Complutense University.

Llinares-García, A., & Romero-Trillo, J. (2008). The pragmatic role of discourse markers of native and non-native teachers in CLIL contexts. In J. Romero-Trillo (Ed.), *Pragmatics and corpus linguistics: A mutualistic entente* (pp. 191–204). Berlin/New York: Mouton de Gruyter.

Lorenzo-Dus, N. (2001). Compliment responses among British and Spanish university students. A contrastive study. *Journal of Pragmatics, 33*, 107–127.

Lorenzo-Dus, N., & Meara, P. (2004). Role-plays and the assessment of oral proficiency in Spanish. In R. Márquez-Reiter & M. E. Palencia (Eds.), *Current trends in the pragmatics of Spanish* (pp. 79–98). Amsterdam/Philadelphia: John Benjamins.

Lozano, C., & Mendikoetxea, A. (2011). CEDEL2 (Corpus Escrito del Español como L2): A large-scale corpus for L2 Spanish acquisition research. In M. L. Carrió Pastor & M. A. Candel Mora (Eds.), *Proceedings of CILC 2011. Universidad Politécnica de Valencia (Spain)*. Valencia: Universidad Politécnica de Valencia.

Martín Zorraquino, M. A., & Portolés, M. A. (1999). Los marcadores del discurso. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (Vol. 2, pp. 4051–4214). Madrid: Espasa-Calpe.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies. An advanced resource book*. London/New York: Routledge Applied Linguistics.

Miracle, W. C. (1991). *Discourse markers in Mandarin Chinese*. PhD thesis, Ohio State University, Columbus. Available at: https://etd.ohiolink.edu/ap:10:0::NO:10:P10_ACCESSION_NUM:osu1232114951. Accessed 27 Aug 2013.

Mitchell, R., Domínguez, L., Arche, M. J., Myles, F., & Marsden, E. (2008). SPLLOC: A new database for Spanish second language acquisition research. In L. Roberts, F. Myles, & A. David (Eds.), *EuroSLA yearbook* (Vol. 8, pp. 287–304). Amsterdam/Philadelphia: John Benjamins.

Moreno, A., & Guirao, J. M. (2006). Morpho-syntactic tagging of the Spanish C-ORAL-ROM corpus: Methodology, tools and evaluation. In Y. Kawaguchi, S. Zaima, & T. Takagaki (Eds.), *Spoken language corpus and linguistic informatics* (pp. 199–218). Amsterdam/Philadelphia: John Benjamins.

Mukherjee, J. (2009). The grammar of conversation in advanced spoken learner English: Learner corpus data and language-pedagogical implications. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 203–230). Amsterdam/Philadelphia: John Benjamins.

Müller, S. (2005). *Discourse markers in native and non-native English discourse*. Amsterdam/Philadelphia: John Benjamins.

Nogueira da Silva, A. (2011). *Enseñanza de los marcadores del discurso para aprendices brasileños de E/LE: análisis de manuales de E/LE y de narraciones orales de aprendices brasileños*. PhD thesis, Universidad de Salamanca, Salamanca. Available at: http://hdl.handle.net/10366/110547

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom. Language use and language teaching*. Cambridge: Cambridge University Press.

Polat, B. (2011). Investigating acquisition of discourse markers through a developmental learner corpus. *Journal of Pragmatics, 43*(15), 3745–3756.

Pons Bordería, S. (1998). *Conexión y conectores. Estudio de su relación en el registro informal del la lengua*. Valencia: Universidad de Valencia.

Pons Bordería, S. (2006). A functional approach to the study of discourse markers. In K. Fischer (Ed.), *Approaches to discourse particles* (pp. 77–99). Oxford: Elsevier.

Porroche Ballesteros, M. (1998). Sobre algunos usos de *que*, *si* y *es que* como marcadores discursivos. In A. Mª, M. Zorraquino, & E. Montolío Durán (Eds.), *Los marcadores del discurso. Teoría y práctica* (pp. 229–242). Madrid: Arco/Libros.

Ramírez Verdugo, D., & Romero-Trillo, J. (2005). The pragmatic function of intonation in L2 discourse: English tag questions used by Spanish speakers. *Intercultural Pragmatics, 2*(2), 151–168.

Romero-Trillo, J. (2002). The pragmatic fossilization of discourse markers in non-native speakers of English. *Journal of Pragmatics, 34*(6), 769–784.

Romero-Trillo, J. (2004). Subjective and objective grounding in discourse markers: A cross-linguistic corpus-driven approach. In K. Aijmer (Ed.), *Dialogue analysis VIII: Understanding and misunderstanding in dialogue: Selected papers from the 8th IADA conference, Göteborg 2001* (pp. 185–197). Tübingen: Niemeyer.

Romero-Trillo, J. (Ed.). (2008). *Pragmatics and corpus linguistics. A mutualistic entente* (Mouton Series in Pragmatics, Vol. 2). Berlin: Mouton de Gruyter.

Romero-Trillo, J. (Ed.). (2013). *Yearbook of corpus linguistics and pragmatics 2013. New domains and methodologies*. New York: Springer.

Rose, K., & Kasper, G. (2001). *Pragmatics in language teaching*. Cambridge: Cambridge University Press.

Santos Río, L. (2003). *Diccionario de partículas*. Salamanca: Luso-Española de ediciones.

Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.

Schmer, D. (2012). Un análisis de nada como marcador del discurso en el español coloquial de Buenos Aires. *Revista del Instituto de Investigaciones Lingüísticas y Literarias Hispanoamericana (RILL), 17*(1/2), 218–227.

Sessarego, C. (2009). Pragmatic language instruction and beginner learners of Spanish: A discourse approach to pragmalinguistics. *Estudios de Lingüística Aplicada, 49*, 97–120.

Simon-Vandenbergen, A.-M., & Willems, D. (2011). Cross-linguistic data as evidence for grammaticalization. In L. Degand, & A.-M. Simon-Vandenbergen (Eds.), *Grammaticalization,*

pragmaticalization and/or (inter)subjectification: Methodological issues for the study of discourse markers. Thematic Issue of *Linguistics*, *49*(2): 333–364. Berlin: De Gruyter Mouton.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell's.

Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics, 4*(2), 91–112.

Walsh, S. (2013). Corpus linguistics and conversation analysis at the interface: Theoretical perspectives, practical outcomes. In J. Romero-Trillo (Ed.), *Yearbook of corpus linguistics and pragmatics 2013. New domains and methodologies* (pp. 37–51). New York: Springer.

Wei, M. (2011). Investigating the oral proficiency of English learners in China: A comparative study of the use of pragmatic markers. *Journal of Pragmatics, 43*(14), 3455–3472.

Yao, J. (2012). Clasificación de los marcadores del discurso en chino mandarín. In F. J. Rodríguez Muñoz (Ed.), *Los marcadores del discurso en español y en chino mandarín. Philologica Urcitana. Revista Semestral de Iniciación a la Investigación en Filología* (Vol. 7, pp. 5–20). Almería: Universidad de Almería.

Yoshimi, D. R. (2001). Explicit instruction and JFL learners. Use of interactional discourse markers. In K. Rose & G. Kasper (Eds.), *Pragmatics in language teaching* (pp. 223–244). Cambridge: Cambridge University Press.

Zhao, H. (2013). A study on the pragmatic fossilization of discourse markers among Chinese English learners. *Journal of Language Teaching and Research, 4*(4), 707–714.