NO SÉ AS A DISCOURSE MARKER IN SPANISH: A CORPUS BASED APPROACH TO A CROSS-DIALECTAL COMPARISON

Javier Rivas y Esther Brown[1]
University of Colorado at Boulder

Abstract

The purpose of this study is to determine the form(s) and function(s) of *no sé* through the analysis of 820 tokens that we extract from three corpora of spoken Spanish in New Mexico, Puerto Rico and Madrid, Spain. The results of our quantitative analysis suggest that, like *I don´t know* (Bybee & Scheibman 1999), *no sé* is stored in the lexicon as a chunk in these three dialects. This analysis is reinforced by rates of phonological reduction of word-initial /s/ in New Mexico. Our results also reveal that *no sé* has similar discourse functions in the three varieties. However, other factors such as frequency of occurrence and use of first-person *yo* show that there are important dialectal differences. As a discourse marker, *no sé* has a limited use in New Mexico, whereas Madrid is the most advanced variety. The Puerto Rican variety holds an intermediate position.

Key words: *no sé,* discourse marker, discourse frequency, cross-dialectal comparison, subjectivity.

1. INTRODUCTION

In the last decades, there has been a growing interest in academic research on discourse markers (also known as *muletillas, pragmatic markers, discourse connectives, discourse operators*, *discourse particles,* among others). Research on these items has been undertaken from different perspectives by scholars such as Cortés Rodríguez (1991), Fraser (1990, 1999), Martín Zorraquino (1992), Portolés (1998), and Schiffrin (1987). Although the nature and scope of the category of *discourse marker* is still a matter of debate between scholars, they all agree that these items connect portions of discourse by imposing some type of relationship between them concerning the speaker's communicative intent.

The class of words that makes up discourse markers is heterogeneous. Markers include invariable classes such as adverbs (*bien*) prepositions (*hasta*) and conjunctions (*pero*), but they also derive from nouns (*¡hombre!*) and verbs (*mira*). In their development, they lose morphological and syntactic autonomy, and typically behave as independent elements. Concurrently their word frequency increases and the meaning becomes more general without fully losing its original semantic content. In those cases in which discourse markers come from more than one word (*o sea*), they form lexically cohesive elements.

---

[1] **Address for correspondence: Javier Rivas**: Spanish and Portuguese Department, McKenna Languages Building, 278 UCB, Boulder, CO 80309-0278, USA. Phone number: + 1-303-492-6268 / + 1-303-735-6170. Fax number: + 1-303-492-3699. E-mail: rivasrod@colorado.edu / brownel@colorado.edu. **Esther Brown**: Spanish and Portuguese Department, McKenna Languages Building, 278 UCB, Bulder, CO 80309-0278, USA. Phone number: + 1-303-492-6268 / + 1-303-735-6170. Fax number: + 1-303-492-3699. E-mail: rivasrod@colorado.edu / brownel@colorado.edu

An example of a discourse marker in English is *I don't know*. Bybee and Scheibman (1999) and Scheibman (2000), find that when *I don't know* behaves as a discourse marker, it undergoes phonological reduction of the vowel, the consonant, or both. They attribute this reduction to the status of the word-pair as a 'chunk' in lexical representation due to the frequent co-occurrence of the individual lexical items in the construction. In line with the Exemplar Model of Lexical Representation (Bybee 2001, Pierrehumbert 2001), these authors suggest that the three words that make up the discourse marker *I don't know* are, in reality, stored in the lexicon as any morphologically complex word would be.

Although *no sé*, the translation equivalent of *I don't know,* has been regarded as a discourse marker in the literature (see, for example, Bravo Cladera 2003, Cuenca and Marín, in press, and Davidson 1996), to the best of our knowledge there are no large-scale quantitative analyses on this item. The purpose of the present study is to determine the form(s) and function(s) of *no sé* through a quantitative analysis. We analyze 820 tokens that we extract from three corpora of spoken Spanish in New Mexico, Puerto Rico, and Madrid, Spain.

2. BACKGROUND

Discourse markers have been widely studied yielding great insight into language variation and use[1]. They are considered to be "sequentially dependent elements which bracket units of talk" (Schiffrin 1987: 31). Fraser (1999: 938) elaborates that discourse markers "impose a relationship between some aspect of the discourse segment they are a part of […] and some aspect of a prior discourse segment [...]". Typically, as noted by Portolés (1998: 26), the function of discourse markers is "guiar, de acuerdo con sus distintas propiedades morfosintácticas, semánticas y pragmáticas, las inferencias que se realizan en la comunicación [guide the inferences that take place in communication according to their different morphosyntactic, semantic, and pragmatic properties]".

Particular discourse markers previously analyzed in Spanish include *bueno* (Serrano 1999), *o sea* (Schwenter 1996), *pues* (Garcés Gómez 1992), *entonces* (Travis 2005), to name a few. Examples of these discourse markers found in our corpus follow:

Puerto Rico

(1) I: Bueno hubo dos años que estuvo bajo el mandato del penepé.

  E: ¿Dos años? ¿Cómo que dos años?

  I: **O sea** dos, dos cuatrenios, me refiero.

  ´I: Well there were two years under the rule of PNP.

  E: Two years? What do you mean two years?

  I: I mean two, two four-year periods, is what I mean´

Puerto Rico

(2)   E: ¿Y qué, y qué ha hecho por Caguas?

          I: **Bueno**, en, en Caguas se ha visto las avenidas más bonitas, ha arregla'o la entrada al pueblo está bien bonita ahí.

          ´E: And what, and what has he done for Caguas?

          I: Well, in Caguas the streets are prettier, he has fixed up the entrance to the city it is so pretty there´

Puerto Rico

(3)   I: Está bien, yo lo sé. Porque cuando la economía de Estados Unidos está mala **pues** la de aquí se, este, cuando allá tiene gripe aquí tiene pulmonía.

          ´I: It's OK, I know. Because when the US economy is bad well the one here, uh, when they have the flu there, here they have pneumonia´

These and other discourse markers have been shown to possess certain linguistic characteristics which set them apart from other non discourse / pragmatic elements. In phonological terms, they are prosodically independent. That is to say, they tend to occur either between pauses or intonation breaks, though not exclusively. In terms of morphology, they are invariable elements. They display no inflectional or derivational variation. For instance, *bueno*, as a discourse marker never reflects gender or number as the adjectival counterpart would (*bueno, buena, buenos, buenas*). Another characteristic of discourse markers is, that in terms of syntax, they are outside the syntactic structure of the clause in that they are not part of the argument structure of the verb.

       Much research has been conducted in English on the pragmatic functions of *I don't know* (Bybee & Scheibman 1999, Pichler 2007, Scheibman 2000, Tsui 1991). As a discourse marker it has been shown to convey two main functions. These uses include a subjective function and a textual function (which we will discuss below). An analysis of *no sé*, however, the translation equivalent of English *I don't know*, has not been, to our knowledge, the subject of any previous large-scale quantitative analysis.

3. DATA AND METHODS

       In order to determine the extent to which *no sé* is used as a discourse marker in Spanish, we use three separate corpora for our quantitative analyses. We relied upon transcriptions of sociolinguistic interviews conducted with native speakers from three different varieties of Spanish; New Mexican Spanish, Puerto Rican Spanish, and the Spanish of Madrid. We chose New Mexican data in order to investigate the role phonological reduction has (if any) in the occurrence of *no sé* as a discourse marker. In the traditional Spanish spoken in New Mexico and

Southern Colorado, there is considerable phonological reduction of word and syllable initial /s/ (Brown 2005), making it possible to test for reduction within the *no sé* word pair. We also chose another American variety (the Puerto Rican data) and a European variety (Madrid, Spain) to enable cross-dialectal comparison.

The New Mexico data in this analysis comes from the New Mexico and Southern Colorado Linguistic Atlas Project (Bills & Vigil 1999). The NMCOSS project initiated in 1991, documents, via interviews with 350 native speakers, the traditional language spoken throughout the state of New Mexico and the sixteen counties of southern Colorado by descendents of the first settlers to the region (Bills & Vigil 1999). The NMCOSS data was collected by trained field workers who tape-recorded interviews involving both controlled elicitation and guided conversation (Vigil 1989). The data reported upon in this study are taken solely from the free conversation portion of interviews with 18 males and 6 females, ranging in age from 36-90 years old and residing predominantly in rural regions of New Mexico (with 2 speakers coming from Albuquerque). From New Mexico we analyze *no sé* tokens in approximately 97,000 words.

The Puerto Rican data are taken from transcriptions representing roughly 27 hours of conversation with eighteen native speakers. These conversations were collected and transcribed by a native speaker in Caguas, Cayey, and San Juan, Puerto Rico in 2000 (Cortés-Torres 2005). Speakers range in age from 24-90 years old. Interviews ranged in duration from one half hour to three hours in length. The corpus of Puerto Rican Spanish consists of approximately 370,000 words.

The analysis of *no sé* tokens in Madrid was conducted using a portion of the Corpus del Español (Davies 2002) online. We limit our coding to the first 194 occurrences of *no sé* in 'Habla Culta de Madrid'. There are 487 number of tokens of *no sé* in the total 140,000 number of words in Habla Culta de Madrid. We do not control for cross-corpora differences of size or for topic of conversation or sociolinguistic factors.

In both the American varieties of Spanish, the frequency of occurrence of *no sé* overall is very similar. In New Mexico, *no sé* has a text frequency of approximately 1,371 per million, and in Puerto Rico, the word pair has a text frequency of approximately 1,354 per million. In contrast, in Madrid the text frequency for *no sé* is more than twice that of the non-European dialects (approximately 3,474 per million). Based on this word frequency, we could predict that the Spanish variety would be the one to be more likely to have more developed discourse / pragmatic uses.

In order to determine the discourse / pragmatic and referential functions of *no sé* in these varieties of Spanish we isolated all occurrences of the word pair. Across the three corpora of spontaneous conversational data this yielded 828 instances of *no sé*. We first identified referential uses of *no sé*, which are cases in which its use only indicates a lack of knowledge of something on the part of the speaker. To do this, we considered both grammatical and semantic content. For example, instances in which *no sé* takes either a nominal or a clausal direct object are coded as referential uses. Examples are found in (4) and (5)[2].

New Mexico (*no sé* + NP)

(4) Sí. Leo pero **no sé** *las puntuaciones*, para escribir…

´Yes. I can read but I don't know punctuation, to write…´

Madrid (*no sé* + *si* – clause)

(5) yo **no sé** *si eso hará gracia a los niños*.

´I don't know if this will make the kids laugh´

Lack of a complement did not automatically lead to a token being considered as a discourse marker. We also regarded as referential those cases in which *no sé* only indicated lack of knowledge. This can be seen exemplified in (6).

Puerto Rico

(6) B: ¿A qué hora es?

J: **No sé**.

´B: What time is it?

J: I don't know´

Overall, these methods allowed us to identify 616 referential uses of *no sé*. The remaining 212 tokens were all coded as discourse markers.

Following Bybee and Scheibman (1999), Pichler (2007), Scheibman (2000), and Tsui (1991), we classified the discourse / pragmatic functions into three distinct categories; subjective, textual, or a combination of the two (both). Tokens were coded as 'subjective' when they were used by the speaker in order to soften the strength of their propositions, to avoid face-threatening acts, or attenuate disagreeing statements. In many cases the subjective uses of *no sé* co-occur with other discourse markers of epistemic stance like *yo creo* and *pues* in (7), or *o sea* and *como,* as in (8).

New Mexico

(7) *yo creo* que la mujer, *pues* **no sé**, se le fue con otro, pero *yo creo* que tenía problemas con el alcohol también.

´I think that his wife, well I don't know, he ran away with another man, but I think that she had alcohol problems too´

Madrid

(8) por esto, *o sea*, porque **no sé**, siempre estás *como* más acompañada.

´for that reason, [I mean?], because I don't know, you're always like not alone´

Tokens were coded as 'textual' when they were used by the speaker for turn management and/or topical development. Textual functions also include cases of hesitation and repair. Examples are given in (9) and (10).

Madrid

(9)   Es que a mí me da muchísimo miedo el... el **no sé**, el... por una tontería,
'The thing is that I'm afraid of the…the I don't know, the…from something silly'

Madrid

(10)   y después, novela, pues **no sé** y... ¿ tú qué tal?
'and after that, a whole store, well I don't know and…. How are you?'

In number (9) there is clear hesitation and in example (10) there is an attempt to close a topic and leave the floor to the interlocutor.

    Not all cases are either clearly subjective or textual, but rather seem to perform both functions simultaneously, as noted by Pichler (2007: 180). Examples are provided in (11) and (12).

Puerto Rico

(11)   Bueno, no, eso es lo que dijeron ahora, porque ahora están hablando y **no sé.**
'Well, no, that is what they just said, because now they are talking and I don't know'

Madrid

(12)   el progreso, pero creo que más fundamental todavía es el... **no sé**, las ideas,
'progress, but I think that even more important yet is the…I don't know, the ideology,'

In example (11) it appears the speaker would like to end her turn, but at the same time could wish to soften the disagreement in her previous negative statement. In example (12), there is repair, and at the same time the speaker seems to want to attenuate the previous utterance (which also includes the epistemic use of *yo creo*).

We acknowledge the inherent subjectivity involved with coding speaker intent. In all cases, tokens of *no sé* were coded to the best of our ability taking into the grammatical and semantic cues available from the context. The following section highlights the results of our quantitative analyses.

4. RESULTS

The results of our quantitative analyses reveal that *no sé* has discourse functions in the three varieties. This word pair is used 194 times in the Madrid data that we analyzed, with 60% of the uses being non-referential. This is summarized in Table 1. In Puerto Rico, of the 501 instances of *no sé*, 17% are used with a discourse / pragmatic function. In New Mexico, 10% of the tokens have a non-referential function.

Table 1: % of *no sé* tokens used with discourse/pragmatic functions in three varieties of Spanish

	Discourse Marker	Referential Meaning	% Discourse Marker
Madrid	116	78	60%
Puerto Rico	83	418	17%
New Mexico	13	120	10%

Bybee and Scheibman (1999) and Scheibman (2000) discuss phonological reduction and frequency of co-occurrence of the individual lexical items that make up the construction *I don't know*. They suggest this reduction is an indicator of its discourse marker status differentiating its use from other (non-pragmatic) uses of *I don't know*. Results of our analyses suggest that *no sé* behaves significantly differently than the lexical combination of *no* + other 1st person singular high frequency verbs. This can be seen summarized in Table 2.

Table 2: String frequency of high frequency verbs (1st person singular indicative)

	Madrid	Puerto Rico	New Mexico
no sé	487	501	133
no puedo	25	83	17
no quiero	8	34	8
no tengo	52	70	12
no soy	15	13	1
no voy	15	42	4
no estoy	24	21	5
no creo	37	19	5

The tokens of use for *no sé* listed in Table 2 reflect an interesting fact. In all three varieties, the string frequency (the token frequency of a multiword combination) of *no* + *sé* is quite high. In Madrid, it occurs 487 times, in Puerto Rico 501, and in New Mexico 133 times. This word combination frequency can be compared to the other verbs of high frequency listed in

the Table. In Madrid 'no + verb' combination with the next highest text frequency is *no tengo* with 52 occurrences. In Puerto Rico it is *no puedo* with 83 occurrences, and in New Mexico it is also *no puedo* with 17 occurrences. These are all significant differences in string frequency (*no sé* vs. *no tengo* Madrid p = 0.0000, Chi-square = 4321.407, *no sé* vs. *no puedo* Puerto Rico p = 0.0000, Chi-square = 299.8766, *no sé* vs. *no puedo* New Mexico p = 0.0000, Chi-square = 89.90075).

Another difference to highlight between *no sé* and the other verb combinations is the strikingly high transitional probability (Jurafsky, Bell, Gregory & Raymond 2001) of the combination. That is to say, of all the examples of *sé* in the corpus (N = 565 in Madrid, for example), 86% appear after *no*. This can be seen summarized in Table 3. A similarly high transitional probability is found for *no sé* in the Puerto Rican and New Mexican data (64% and 88% respectively). The implication of such a high transitional probability, and high string frequency as discussed in the preceding paragraph, is the increased likelihood of lexical storage as a chunk in all three dialects (Bybee 2001:109, Bush 2001: 269).

Table 3: % of high frequency verbs (1st person singular indicative) directly preceded by lexical item *no*

	Madrid	Puerto Rico	New Mexico
no sé	487 / 565: **86%**	501 / 776: **64%**	133 / 151: **88%**
no puedo	25 / 77: 32%	83 / 181: 46%	17 / 50: 34%
no quiero	8 / 44: 18%	34 / 173: 20%	8 / 41: 20%
no tengo	52 / 280: 19%	70 / 498: 14%	12 / 98: 12%
no soy	15 / 90: 17%	13 / 123: 11%	1 / 10: 10%
no voy	15 / 133: 11%	42 / 493: 9%	4 / 43: 1%
no estoy	24 / 157: 15%	21 / 318: 7%	5 / 23: 22%
no creo	37 / 512: 7%	19 / 358: 5%	5 / 127: 4%

As a chunk, does *no sé* exhibit phonological reduction as *I don't know* was shown to do in Bybee and Scheibman (1999)? This is testable in the New Mexican data in which syllable and word-initial /s/ has been shown to variably aspirate or delete. Word-initial /s/ reduces in this variety at a rate of 16% overall and at a rate of 30% overall in word-medial, syllable-initial position (Brown 2005: 815). If *no sé* were stored as two distinct words, we might expect reduction of the /s/ in *sé* to reflect rates found for other word-initial /s/ tokens. However, reduction in the word-initial /s/ of the form *sé* in the combination *no sé* is 37% (N = 115). This reduction rate more closely reflects reduction of the word-medial, syllable-initial /s/ tokens. Further, reduction of the /s/ of *sé* outside the combination *no sé* (in combinations such as *yo sé*, *sí sé*, etc.) is significantly lower at 7% (N = 30, p = .0021, X^2 = 10.49417). This result is suggestive that *no sé* has lexical status as a chunk.

Interesting, however, is the result obtained by just examining the phonological reduction of the very few tokens of *no sé* as a discourse marker in New Mexico. Following Bybee and Scheibman (1999), we might predict higher reduction when used with a non-referential meaning

as these authors find for *I don't know*. However, our findings do not bear this out, as Table 4 summarizes[3]. Initial /s/ reduction is greater (36%) in referential uses than in discourse / pragmatic uses (14%). This may suggest that phonological reduction is not characteristic of this particular discourse marker in Spanish (and/or may simply reflect the 'weak' status of *no sé* as a discourse marker in New Mexico which we outline in the discussion).

Table 4: Phonological reduction of *no sé* in New Mexican Spanish

	Referential Uses	Discourse / Pragmatic Uses
Non-reduced ([s])	75	6
Reduced ([h], [ø])	43	1
Percent reduction	**36%**	**14%**

The item *no sé* appears to have chunk status based upon its usage (Tables 2 and 3) and on the higher than average reduction rate in the word combination (37%). If *no sé* is stored as a single lexical unit, this implies it is not necessarily analyzed as a negative particle (*no*) + verb (*sé*), but rather is a prime candidate to be adopted as a discourse marker (with no morphological variation and lack of complementation).

As a discourse marker, we could predict that *no sé* would demonstrate different linguistic behavior from the referential uses (for example, regarding co-occurrence with subject pronouns) because it no longer performs true verbal functions. We analyze co-occurrence of *no sé* with the subject pronoun *yo* for both referential and non-referential uses (as summarized in Table 5). On the left we list uses of *yo* with the discourse / pragmatic function of *no sé* and on the right uses of the subject pronoun with the referential function of *no sé*.

Table 5: Use of subject pronoun *yo* with discourse/pragmatic vs. referential uses of *no sé* in three varieties of Spanish

	Discourse Markers			Verbs		
	yo no sé	*no sé*	*% yo*	*yo no sé*	*no sé*	*% yo*
Madrid	5	111	**4 %**	13	65	**17 %**
Puerto Rico	30	53	**36 %**	211	206	**51 %**
New Mexico	9	4	**69 %**	49	71	**41 %**

In Madrid, when used with a discourse function, the subject pronoun appears in just 4% of the cases. Conversely, when used with a referential function in Madrid, subject pronoun usage is significantly higher at 17% (p = .0036, X^2 = 8.459556). Puerto Rico is a variety of Spanish with high overt subject pronoun usage (Cameron 1993: 306, Hochberg 1986: 615, Morales 1997: 155), and rates of *yo* usage are, perhaps not surprisingly, higher. A similar pattern of *no sé* usage is found in this variety however. When *no sé* functions as a discourse marker, the subject pronoun co-occurs at a rate of 36%. With true verbal function, the subject pronoun is used significantly more, in 51% of the cases (p = .0161, X^2 = 5.792954).

In the New Mexican data, we see a different pattern of pronominal usage[4]. Despite frequency per million of *yo* being lower in New Mexico (10,927) than in Puerto Rico (15,056), when *no sé* functions as a discourse marker, overt *yo* usage is quite high (69%). In fact, subject pronoun usage for the discourse marker is significantly higher than the usage with referential meaning ($p < .05$, $X^2 = 3.846318$). We will comment on this in our discussion below.

Further results of our quantitative analyses are summarized in Table 6. Through an analysis of *no sé* tokens that perform a non-referential function, it is apparent that it fulfills the same discourse functions in the three varieties of Spanish we analyze. This can be seen by the fact that in all three dialects, *no sé* performs the three functions that we examined; subjective, textual, and both. In all three varieties, the most frequent use of *no sé* is the subjective function (Madrid 59%, New Mexico 85%, Puerto Rico 76%). In both New Mexico and Puerto Rico, however, the number of textual uses (NM 8%, PR 6%) is considerably lower than that found in Madrid (28%). Subjective~textual uses are the least frequent in all three dialects as Table 6 makes evident.

Table 6: Discourse / pragmatic functions of *no sé* in three varieties of Spanish

	Madrid	New Mexico	Puerto Rico
Subjective	69 (59 %)	11 (85%)	63 (76%)
Textual	32 **(28 %)**	1 (8%)	15 (18 %)
Subjective~Textual	15 (12%)	1 (8%)	5 (6%)
	N = 116	N = 13	N = 83
	% Subjective use Madrid vs. PR: **$p < .05$**, $X^2 = 5.841602$		

5. DISCUSSION

The results of our large-scale quantitative analysis of usage of *no sé* in three varieties of Spanish has yielded interesting results. As a discourse marker, this construction has limited use in the New Mexican variety. This is evident in Table 1, where we can see that out of the 133 tokens of *no sé* we found in the corpus, only 10% (N = 13) are used discursively. Despite the low number of examples, we nevertheless see a similar pattern of usage to the other two dialects (Table 2, Table 3, Table 6). It appears, however, that when compared to the other varieties of Spanish, *no sé* is a less preferred discourse marker in New Mexico.

This lack of subjective strength of *no sé* as a discourse marker is perhaps reflected in the result we obtained regarding subject pronoun usage (Table 5). Contrary to the other varieties of Spanish, in New Mexico when *no sé* is used non-referentially, there is a strikingly high occurrence of *yo*. As has been noted by Davidson (1996: 551), the subject pronoun when used explicitly adds 'pragmatic weight' to an utterance, and specifically can trigger non-referential readings of certain verbs. We interpret this result in Table 5, therefore, as indicative of the fact

that *no sé* does not have a strong subjective or textual use in this variety without the added weight of the subject pronoun.

The Puerto Rican variety of Spanish appears to hold an intermediary position between the other two varieties we study. As seen in Table 1, the number of uses as a discourse marker is higher than in New Mexico, at a rate of 17%, yet does not reach the high usage seen in Madrid. As in New Mexico, the preferred non-referential use of *no sé* is subjective (as seen in Table 6). The number of textual uses is higher than in New Mexico - 18%. It may be the case that the textual function is taken up by other discourse markers in Puerto Rico and New Mexico [e.g. *so* or *entonces* (Aaron 2004)].

The results of our analysis on the Madrid data reveal an important difference between the American dialects and the European variety. Firstly, in the Peninsular variety, the proportion of *no sé* tokens that are non-referential as opposed to referential is significantly higher. In Madrid, 60% of uses have a discourse / pragmatic function (Table 1). Further, just looking at non-referential uses of *no sé*, in Madrid it is used non-subjectively (with textual functions) significantly more than in the Puerto Rican variety (Table 6), revealing significant differences in usage between the varieties[5]. The dialectal differences we find between America and Spain may reveal similar, but divergent, diachronic pathways.

Traugott and Dasher (2002) have argued that the development of discourse markers depends upon changes towards subjectivity. There is a pathway of change in which items that become discourse markers change from a more objective, referential use to a more subjective, pragmatic use. In the case of *no sé*, we have evidence of this subjective stance in cases in which *no sé* co-occurs with other epistemic markers such as *yo creo, me parece, pues, bueno, o sea,* etc. The function of these uses o *no sé* is to attenuate disagreeing statements, avoid face-threatening acts, and soften the propositional stance. In so doing it acquires *inter*subjective uses, not just subjective (Traugott & Dasher 2002: 155) in that they take into consideration the interlocutor.

Intersubjectivity is also a key element of textual uses in that these uses encompass issues such as turn management. As we mentioned previously, textual instances of *no sé* include functions such as closing of a topic and turn-closing. This seems to suggest that *no sé* could proceed along the following grammaticalization cline (adapted from Traugott & Dasher 2002: 187):

referential / propositional uses > epistemic / subjective uses > intersubjective uses

As we noted in our analyses, there are more textual uses in Madrid than in other the other varieties we analyze (Table 6) suggesting that this discourse marker has extended its use further in Spain than in America.

The dialectal division we find between America and Spain in the use of *no sé* may reflect a diachronic pathway, or, as was found in Company Company (2002), simply different grammaticalization pathways determined by the weight given to specific functions in the distinct

varieties of Spanish. We lack, in our current analysis, historical data that will help determine if the synchronic data we present, are in fact representative of a unified diachronic pathway of change, with Madrid leading the way, or rather whether each variety basically selects different discourse markers to perform these textual functions.

6. CONCLUSIONS

The analysis of 820 tokens that we extracted from three corpora of spoken Spanish in New Mexico, Puerto Rico, and Madrid, Spain allows us to arrive at the following conclusions regarding the form(s) and function(s) of *no sé*:

a) As is the case with *I don´t know* in English, the following evidence suggests that *no sé* has lexical status as a chunk, that is to say, the two words that make up the construction *no sé* are stored in the lexicon as any morphologically complex word would be:

1. The significantly high string frequency of *no + sé* in the three corpora.
2. The high transitional probability of *no sé* in the three corpora.
3. In New Mexico, where syllable- and word-initial /s/ can be aspirated or deleted, phonological reduction of /s/ in *no sé* reduces at a similar rate to word-medial, syllable-initial /s/ tokens, and at a significantly higher rate than other combinations involving *sé* (for example, *yo sé*).

b) Although *no sé* has discourse functions in the three varieties of Spanish under analysis, frequency of occurrence and use of first-person *yo* show that there are important dialectal differences.

c) In New Mexico, the following evidence suggests that *no sé* has a ´weak´ status as a discourse marker:

1. The overall number of discourse tokens is very low.
2. Phonological reduction of /s/ is greater in referential uses than in discourse uses.
3. Non-referential uses of *no sé* are generally accompanied by the subject pronoun *yo*. This fact suggests that *no sé* does not have a strong discourse / pragmatic use in this variety without the added pragmatic weight of the subject pronoun.

d) Madrid is the variety in which *no sé* is more advanced as a discourse marker. This conclusion is based on the following evidence:

1. The proportion of non-referential *no sé* tokens is significantly higher than in the non-European varieties.
2. When *no sé* is used with a discourse function, the occurrence of the subject pronoun *yo* is significantly lower than when *no sé* is used referentially. This fact suggests that *no sé* may be undergoing decategorialization.
3. As a discourse marker, the number of textual uses of *no sé* in Madrid is considerably higher than that found in Puerto Rico or New Mexico, possibly because in the non-European dialects the textual function is taken up by other discourse markers.

642

e) The Puerto Rican variety holds an intermediary position both in number of discourse tokens of *no sé*, and in co-occurrence of *yo* with non-referential uses of *no sé*.

Further research should determine whether the conclusions of this analysis reflect a unified grammaticalization pathway for *no sé* from more epistemic / subjective uses to more intersubjective uses, with Madrid leading the way, or whether the dialect division we find between America and Spain simply reflects different diachronic pathways of change.

Acknowledgements

Notes

[1] See Travis (2005: 27-61) for a useful summary.

[2] Thompson (2002) argues against the subordinate status of so-called complement clauses such as "si eso hará gracia a los niños" in (5). Rather, examples such as (5) are "combinations of (i) CTP [Complement-Taking Predicate]-phrases, serving specific epistemic, evidential, and evaluative frames for the clauses they occur with, and (ii) finite indicative clauses, both declarative and interrogative" (Thompson 2002: 136). According to this analysis, *no sé* conveys similar semantic / pragmatic functions regardless of grammatical context. However, in line with Thompson (2002: 143), we have considered *no sé* as a discourse marker only in those contexts in which it does not occur with a so-called complement clause.

[3] We have very low token numbers for this analysis and all conclusions are thus tentative.

[4] This result is tentative owing to the low token number (N = 13).

[5] It does not achieve statistical significance for the New Mexican variety perhaps as a result of the low token numbers we have for the New Mexico data.

REFERENCES

Aaron, J.E. (2004). 'So respetamos un tradición del uno al otro': So and entonces in New Mexican Bilingual Discourse. *Spanish in Context,* 1:2, 161-179.

Bills, G. & N. Vigil (1999). Ashes to ashes: The historical basis for dialect variation in New Mexican Spanish. *Romance Philology,* 53, 43-67.

Bravo Cladera, N. (2003). Un estudio de los marcadores del discurso y su distribución en interacciones diádicas de jóvenes bilingües. In F. Sánchez Miret (Ed.), *Actas del XXIII congreso internacional de lingüística y filología románica, Salamanca 24-30 de septiembre 2001*. Tubingen: Max Niemeyer Verlag, vol. II, pp. 126-133.

Brown, E. (2005). New Mexican Spanish: Insight into the Variable Reduction of "*la ehe inihial*" (/s-/) *Hispania,* 88:4, 813-824.

Bush, N. (2001). Frequency effects and word-boundary palatalization in English. In J. Bybee & P. Hopper (Eds.) *Frequency and the Emergence of Linguistic Structure*. Amsterdam / Philadelphia: John Benjamins, pp. 255-280.

Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Bybee, J. & J. Scheibman (1999). The effect of usage on degree of constituency: The reduction of *don´t* in English. *Linguistics*, 37:4, 575-596. [Reed. in Bybee, J. (2007). *Frequency of Use and the Organization of Language*. Cambridge: Cambridge University Press, pp. 294-311].

Cameron, R. (1993). Ambiguous agreement, functional compensation, and nonspecific *tú* in the Spanish of San Juan, Puerto Rico and Madrid, Spain. *Language Variation and Change,* 5, 305-334.

Company Company, C. (2002). Gramaticalización y dialectología comparada. Una isoglosa sintáctico-semántica del español. *Dicenda. Cuadernos de Filología Hispánica*, 20, 39-71.

Cortés Rodríguez, L. (1991). *Sobre conectores, expletivos y muletillas en español*. Málaga: Editorial Librería Ágora.

Cortés-Torres, M.E. (2005). *La perífrasis estar + -ndo en el español puertorriqueño: variación dialectal o contacto lingüístico?* Unpublished PhD dissertation. University of New Mexico, Alburquerque, New Mexico, USA.

Cuenca, M.J. & M.J. Marín (in press). Co-occurrence of discourse markers in Catalan and Spanish oral narrative. *Journal of Pragmatics,* (2008), doi:10.1016/j.pragma.2008.08.010.

Davidson, B. (1996). ´Pragmatic weight´ and Spanish subject pronouns: The pragmatic and discourse uses of ´tú´ and ´yo´ in spoken Madrid Spanish. *Journal of Pragmatics*, 26, 543-565.

Davies, M. (2002-). *Corpus del español* (100 million words, 1200s-1900s). Available online at http://www.corpusdelespanol.org.

Fraser, B. (1990). An approach to discourse markers. *Journal of Pragmatics,* 14, 383-395.

Fraser, B. (1999): What are discourse markers? *Journal of Pragmatics,* 31, 931-952.

Garcés Gómez, M.P. (1992). El operador discursivo *pues* en el español hablado. *Romantistisches Jahrbuch*, 43, 261-276.

Hochberg, J. (1986). Functional compensation for /s/ deletion in Puerto Rican Spanish. *Language* 62:3, 609-621.

Jurafsky, D., A. Bell, M. Gregory & W. D. Raymond (2001): Probabalistic relations between words: evidence from reduction in lexical production. In J. Bybee and P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam / Philadelphia: John Benjamins, pp. 309–335.

Martín Zorraquino, M.A. (1992). Gramática del discurso: los llamados marcadores del discurso. In *Actas del congreso de la lengua española,* Madrid, Instituto Cervantes, 709-720.

Morales, A. (1997). La hipótesis funcional y la aparición de sujeto no nominal: el español de Puerto Rico. *Hispania* 80:1, 153-165.

Pichler, H. (2007). Form-function relations in discourse: The case of I DON'T KNOW" *Newcastle Working Papers in Linguistics,* 13, 174-187.

Pierrehumbert, J.B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee and P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure*. Amsterdam / Philadelphia: John Benjamins, pp. 123-136.

Portolés, J. (1998). *Marcadores del discurso*. Barcelona: Ariel.

Scheibman, J. (2000). *I dunno*: A usage-based account of the phonological reduction of *don´t* in American English conversation. *Journal of Pragmatics,* 32, 105-124.

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

Schwenter, S. (1996). Some reflections on *o sea*: A discourse marker in Spanish. *Journal of Pragmatics*, 25, 855-874.

Serrano, M.J. (1999). *Bueno* como marcador de inicio de turno y contraposición: estudio sociolingüístico. *International Journal of the Sociology of Language,* 82, 115-133.

Thompson, S.A. (2002). Object complements and conversation: Towards a realistic account. *Studies in language* 26:1: 125-163. [Reed. as Hacia una explicación de los "complementos objeto" y la conversación. In M. Sedano, A. Bolívar & M. Shiro (Eds.) *Haciendo lingüística. Homenaje a Paola Bentivoglio*. Caracas: Universidad Central de Venezuela, 2006, pp. 347-392].

Traugott, E.C. & R.B. Dasher (2002). *Regularity in Semantic Change,* Cambridge: Cambridge University Press.

Travis, C.E. (2005). *Discourse Markers in Colombian Spanish: A Study in Polysemy*. Berlin / New York: Mouton de Gruyter.

Tsui, A.B. (1991). The pragmatic functions of *I don´t know*. *Text*, 11:4, 607-622.

Vigil, N. (1989). Database for a linguistic atlas of the Spanish of New Mexico and southern Colorado, computer methods in dialectology. *Journal of English Linguistics Special Issue*, 1, 69-75.