

Near final version (06/05/2009):

Rivas, Javier & Esther L. Brown. 2010. "Variable development of intersubjectivity in Spanish." *A Mosaic of Corpus Linguistics. Selected Approaches*. Eds. Aquilino Sánchez & Moisés Almela. Frankfurt/Berlin: Peter Lang Frankfurt. 61–78.

Variable development of intersubjectivity in Spanish

Javier Rivas & Esther L. Brown
University of Colorado at Boulder

Abstract

As has been shown by Traugott & Dasher (2002), discourse markers proceed along the following grammaticalization cline; non-subjective > subjective > intersubjective uses. We look for evidence of this grammaticalization cline using a previously unexamined discourse marker in Spanish - *no sé* ('I don't know'). In line with Bybee & Sheibman (1999), we show that *no sé* seems to be stored in lexical representation as a single item in Spanish. We base this analysis on phonological reduction and measures of string frequency and transitional probability of the word combination. Further, following Company Company (2002), a cross-dialectal comparison of relative frequency of use of *no sé* reveals differing degrees of intersubjectification of this discourse marker.

Key words

no sé, discourse marker, discourse frequency, cross-dialectal comparison, subjectification, intersubjectification.

1. INTRODUCTION

Grammaticalization is a gradual process whereby a lexical item or construction becomes a grammatical item or a grammatical item becomes even more grammatical. Grammaticalization processes start in very specific pragmatic and syntactic contexts of use, which, by means of reanalysis, allow the lexical item to acquire grammatical functions. As a result, the grammaticizing element typically undergoes decategorialization, phonological reduction and semantic bleaching (Brinton & Traugott 2005, Lehmann 1995, Hopper & Traugott 2003). In addition, as is noted by Bybee (2003), grammaticalized forms are more frequent than their original lexical source; they increase their productivity because they undergo syntactic expansion, that is to say, they progressively co-occur with different unit types. This, in turn, increases the text frequency of the forms undergoing grammaticalization.

2 Variable development of intersubjectivity in Spanish

Lexical items participating in grammaticalization processes generally undergo a shift from a more concrete/specific meaning to a more abstract/general meaning that typically relates to the speaker's attitude towards the message. This phenomenon, known as subjectification (Traugott 1995a: 31), manifests itself in the development of grammatical forms that explicitly refer to the speaker's perspective. Subjectification plays an important role in many grammaticalization processes, such as for example, the development of discourse markers, as is shown by Traugott (1982, 1995a, 1995b, 2003) and Traugott and Dasher (2002) regarding, among others, markers such as *in fact*, *indeed*, *actually* and *well*.

Discourse markers are items that connect portions of discourse by simultaneously referring anaphorically to the previous discourse and cataphorically to the subsequent discourse by imposing some type of relationship between them. Traugott (2003: 125) points out that discourse markers are subjective because they express the speaker or writer's viewpoint towards the connectivity between clauses. Furthermore, discourse markers are also subjective in that they are concerned with the speaker's degree of knowledge of or involvement in the message s/he is conveying. The rise of these epistemic uses may eventually lead to intersubjectification, that is to say, "the development of meanings that encode speaker/writers' attention to the cognitive stances and social identities of addressees" (Traugott 2003: 124). Discourse markers acquire these intersubjective uses because they serve as hedges in the utterance and are used to soften the strength of the proposition.

Discourse markers may have different lexical and/or syntagmatic sources. *I don't know*, for example, arises from the construction 1ST PERSON SINGULAR SUBJECT PRONOUN *I* + CONTRACTED NEGATED AUXILIARY *don't* + VERB *know*. Bybee and Scheibman (1999) and Scheibman (2000) suggest that *I don't know* is a chunk in lexical representation, and is in reality, stored in the lexicon as any morphologically complex word would be. These authors based this hypothesis on the frequent repetition of the construction, which, consequently, undergoes phonological reduction. In fact, phonological reduction is almost the norm when *I don't know* is used pragmatically to express subjective and/or textual meanings.

Although some studies (for example, Bravo Caldera 2003, Cuenca & Marín 2009, and Davidson 1996) identify *no sé*, the translation equivalent of *I don't know*, as an example of a discourse marker, there are no studies dedicated to the analysis of this construction exclusively. The purpose of the present paper is to determine the forms and functions

3 Variable development of intersubjectivity in Spanish of *no sé* through a large-scale quantitative analysis. We analyze 828 tokens that we extract from three corpora of spoken Spanish in New Mexico, Puerto Rico, and Madrid, Spain. As is the case with *I don't know*, we will show that *no sé* seems to have lexical status as a chunk in Spanish. We will base this analysis on phonological reduction and measures of the frequency of occurrence of *no sé*, such as the string frequency and the transitional probability of the combination. In addition, the analysis of the three varieties of Spanish mentioned above will allow us to show that the extent of intersubjective uses of *no sé* as a discourse marker varies according to dialect.

2. BACKGROUND

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 (1994), Portolés (1998), Schiffrin (1987), Traugott (1982, 1995), and Travis (2005), *inter alia*. Discourse markers 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]”.

The morphosyntactic sources of discourse markers are heterogeneous. They include invariable classes such as adverbs (*bien* ‘well’) prepositions (*hasta* ‘until’) and conjunctions (*pero* ‘but’), but also nouns (*¡hombre!* ‘man!’) and verbs (*mira* ‘look’). 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 (Hopper 1991, Schwenter 1996: 858). In those cases in which discourse markers come from more than one word — for example, *o sea*, which consists of the adversative conjunction *o* ‘or’ and the third person present subjunctive of *ser* ‘be’ — they form lexically cohesive elements.

Particular discourse markers previously analyzed in Spanish

4 Variable development of intersubjectivity in Spanish include *o sea* (Schwenter 1996), *bueno* (Serrano 1999), and *pues* (Garcés Gómez 1992), to name a few. Examples of these discourse markers found in our corpus follow:

- (1) Puerto Rico
 I: Danny Rivera a cada rato va a allí.
 E: Pero esos son, **o sea**, son permisos especiales, digo de turista.
 T: Danny Rivera goes there all the time
 E: But those are, I mean, they are special permits, I mean, tourist permits´
- (2) Puerto Rico
 E: ¿Y quiénes quedan en la compañía, papi, que tú conozcas?
 I: **Bueno**, todavía quedan dos o tres de los muchachos. Jaime creo que se va a retirar ahora.
 E: And who is still in the company, daddy, that you know?
 I: Well, there are still two or three chaps. Jaime, I think he´s going to retire now´.
- (3) Puerto Rico
 E: Yo tengo que saber de dónde viene mi familia.
 I: **Pues**, pregúntale a Ángel cuando lo veas a ver si él sabe más que yo de eso.
 E: I have to know where my family comes from.
 I: Well, ask Angel when you see him. He may know about that more than I do´

These and other discourse markers have been shown to possess certain linguistic characteristics which set them apart from other non-discourse / pragmatic elements. They are prosodically independent. That is to say, they tend to occur either between pauses or intonation breaks, though not exclusively. Furthermore, 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 they lie outside the limits of the clause. In fact, their scope often includes several clause combinations.

Much research has been conducted in English on the forms and functions of *I don't know* (Bybee & Scheibman 1999, Pichler 2007, Scheibman 2000, Tsui 1991). These studies have shown that *I don't know* conveys two main functions as a discourse marker: subjective and textual. In these uses, the construction departs from its original verbal meaning of 'lacking knowledge about something' to express subjective

5 Variable development of intersubjectivity in Spanish meanings relating to the epistemic stance of the speaker and textual meanings regarding turn management, topic development, and repair, *inter alia*. In addition, Bybee and Scheibman (1999: 303) find that, when *I don't know* performs a discourse function, it systematically undergoes phonological reduction; the vowel of the auxiliary *don't* is typically reduced to schwa, and, furthermore, in most cases the initial consonant may also be reduced to a flap. This phonological reduction is due to the high frequency of co-occurrence of the construction. This evidence suggest that *I don't know* constitutes a chunk in lexical representation. In line with the Exemplar Model of Lexical Representation (Bybee 2001, Pierrehumbert 2001), these authors suggest that this construction is stored in the lexicon as if it were a morphologically complex word.

In contrast to *I don't know*, its translation equivalent *no sé* has not been the subject of any previous analysis. In what follows, we will account for the forms and functions of *no sé* through the analysis of three Spanish dialects; New Mexico, Puerto Rico, and Madrid.

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 2008). The NMCROSS 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 NMCROSS 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

6 Variable development of intersubjectivity in Spanish

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 133 examples of *no sé* 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, which consists of approximately 370,000 words, contains 501 examples of *no sé*.

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 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 verbal functions of *no sé* in these varieties of Spanish we isolated all occurrences of the words *no* and *sé* contiguously. Across the three corpora of spontaneous conversational data this yielded 828 instances of *no sé*. We first identified lexical 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 verbal uses. Examples are found in (4) and (5).¹

¹ Thompson (2002) argues against the subordinate status of so-called complement clauses such as "si le pedí su nombre" in (5). Rather, examples such as (5) are "combinations of (i) CTP [Complement-Taking Predicate]-

7 Variable development of intersubjectivity in Spanish

(4) Madrid (*no sé* + NP)

A. - Y es cierto, que...pero, la verdad, que yo **no sé** las dificultades que habrá, pero que no se hace.

´A: And it is true, that...but, to be honest, I don't know about the difficulties that we will have, but that is not acceptable´

(5) New Mexico (*no sé* + *si* – clause)

E: Y **no sé** si le pedí su nombre, ¿cómo era, cómo se llamaba ella?

´E: And I don't know if I asked her name, what was it, what was her name?´

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

(6) Puerto Rico

E: ¿Fue a todas las casas?

I: Yo **no sé**. Por lo menos vino a esta.

´E: Did he visit all the houses?

J: I don't know. He at least came to this one´

Overall, these methods allowed us to identify 616 lexical 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).

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.

8 Variable development of intersubjectivity in Spanish

(7) New Mexico

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

‘M: 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’

(8) Madrid

I: por esto, *o sea*, porque **no sé**, siempre estás *como* más acompañada.

‘I: ‘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).

(9) Madrid

I: Es que a mí me da muchísimo miedo el... el **no sé**, el... por una tontería, porque realmente cuando pasan esas cosas suelen ser por una tontería

‘I: The thing is that I’m afraid of the...the I don’t know, the...for something silly because when things like that happen it is generally for something silly’

(10) Madrid

B: Bueno, pues mira es que [...] a mí, la narrativa me gusta...estoy leyendo, pero eso ya no es novela, sino que es ensayo, es un libro de, de John Wilson, "El lenguaje y la búsqueda de la verdad" y después, novela, pues **no sé** y... ¿tú qué tal?

A: Claro, lo que pasa es que yo veo, que a ti te interesa mucho más el ensayo...

‘B: Well, look, the thing is... I like narrative...I’m reading, but this is not a novel but an essay, it’s a book by, by John Wilson, "Language and the search of truth", and then novel, well I don’t know and.... how are you?’

A: Sure, the thing is that I can see that you are much more interested in essays...’

In excerpt (9) *no sé* acts as a “filler” repair (Fox, Hayashi and Jasperson 1996: 204). The speaker placed *no sé* in the structural slot of the head noun of the subject NP of *dar miedo* ‘frighten’ as a means to delay the production of a next noun due during what appears to be a word search.

9 Variable development of intersubjectivity in Spanish

In excerpt (10) *no sé* is used in 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).

(11) Puerto Rico

E: Yo le dije que yo no tenía, yo no tenía ganas de salir porque a mí no me gusta salir en despedida de año, pero, no me gusta, porque aquí la gente es muy salvaje guiando, no me quiero imaginar el día de despedida de año con dos o tres palos encima la gente guiando por ahí.

B: Pues hasta ahora-

E: Así que **no sé, no sé, no sé**. ¿Y qué hicieron el día de Navidad?

┐E: I told him that I didn't feel, I didn't feel like going out because I don't like going out on New Year's Eve, because people here drive very aggressively, I can't imagine on New Year's Eve, people driving after having two or three drinks under their belt

B: Well, up to now-

E: So I don't know, I don't know, I don't know. And what did you do on Christmas day?'

(12) Madrid

I: Es verdad que la preparación técnica es fundamental, es decir, el avance de... técnico creo que es fundamental para... para el progreso , pero creo que más fundamental todavía es el... **no sé**, las ideas, el mundo de las ideas

┐I: It is true that technical training is essential, that is to say, technical advancement I think it is essential for progress, but I think that even more important yet is the...I don't know, ideology, the world of ideas'

In excerpt (11) it appears the speaker would like to close a topic, but at the same time could wish to soften the strong statement that she made in her previous turn, and hence the repetition of *no sé*. Excerpt (12) has another example of *no sé* as a filler repair (cf. with (9)) but at the same time the speaker seems to want to attenuate the previous utterance (which also includes the epistemic use of *yo creo*).

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 pragmatic. 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 pragmatic function.

Table 1: Percent of pragmatic uses in three dialects

	N	%
Madrid	116/194	60
Puerto Rico	83/501	17
New Mexico	13/133	10

As we mention before, 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
<i>no sé</i>	487	501	133
<i>no puedo</i>	25	83	17
<i>no quiero</i>	8	34	8
<i>no tengo</i>	52	70	12
<i>no soy</i>	15	13	1
<i>no voy</i>	15	42	4
<i>no estoy</i>	24	21	5
<i>no creo</i>	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

11 Variable development of intersubjectivity in Spanish

combination frequency can be compared to the other verbs of high frequency listed in the Table. In Madrid, the ‘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. The transitional probability (or conditional probability) of a particular target word (*sé* in this case) given a previous word (*no*) is calculated from our corpus by calculating the number of occurrences of *no sé* and dividing by the number of times the word *no* occurs. This calculation controls for the frequency of the previous word. Some word dyads can have a high string frequency simply as a result of being two highly frequent words (e.g.; *es el*). As Jurafsky, Bell, Gregory & Raymond (2001: 272) note, however, unlike string frequency, “the conditional probability would be high only if the second word was particularly likely to follow the first”.

The transitional probability of *no sé* is quite high, as can be seen summarized in Table 3. Of all the tokens of *sé* in the corpus (N = 565 in Madrid, for example), 86% appear after *no*, yielding the construction *no sé*. A similarly high transitional probability is found for *no sé* in the Puerto Rican and New Mexican data (64% and 88% respectively).

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

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

The implication of such a high transitional probability, and high string frequency as discussed in the preceding paragraph, is the increased

12 Variable development of intersubjectivity in Spanish
 likelihood of lexical storage as a chunk in all three dialects (Bybee 2001:163, Bush 2001: 269).

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 36% (N = 116), as can be seen summarized in Table 4. 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 6% (N = 32, p = .0021, X² = 10.49417). This result is suggestive that *no sé* has lexical status as a chunk.

Table 4: Percent of syllable-initial /s/ reduction in New Mexico

	N	% Reduction
word-initial /s/	415/2594	16
word-medial, syllable initial /s/	928/3048	30
/s/ reduction in <i>no sé</i> dyad	43/116	36
/s/ reduction in <i>sé</i> (outside of dyad)	2/32	6

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 pragmatic meaning as these authors find for *I don't know*. However, our findings do not bear this out, as Table 5 summarizes.²

Table 5: Phonological reduction of *no sé* in New Mexican Spanish

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

² We have very low token numbers for this analysis and all conclusions are thus tentative.

13 Variable development of intersubjectivity in Spanish

Initial /s/ reduction is greater (36%) in lexical 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 lower degree of entrenchment of *no sé* as a discourse marker in New Mexico which we outline in the discussion.

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 (36%) (Table 4). 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 lexical 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 lexical and pragmatic uses (as summarized in Table 6). 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 verbal function of *no sé*.

Table 6: Use of subject pronoun *yo* with discourse/pragmatic vs. verbal uses of *no sé* in three varieties of Spanish

	Discourse Markers			Verbs		
	<i>yo no sé</i>	<i>no sé</i>	% <i>yo</i>	<i>yo no sé</i>	<i>no sé</i>	% <i>yo</i>
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 verbal 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 ($p = .0161$, $X^2 = 5.792954$).

14 Variable development of intersubjectivity in Spanish

In the New Mexican data, we see a different pattern of pronominal usage.³ 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 lexical 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 7. Through an analysis of *no sé* tokens that perform a pragmatic 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 18%) is considerably lower than that found in Madrid (28%). Subjective~textual uses are the least frequent in all three dialects as Table 7 makes evident.

Table 7: 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

As a discourse marker, *no sé* 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 7). 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

³ This result is tentative owing to the low token number (N = 13).

15 Variable development of intersubjectivity in Spanish

usage (Table 6). Overall in New Mexican Spanish, the average rate of the subject pronoun *yo* expression is 34% (Torres Cacoullos & Travis to appear). Contrary to the other varieties of Spanish, in New Mexico when *no sé* is used pragmatically, there is a strikingly high occurrence of *yo* (69%). As has been noted by Davidson (1996: 551), the subject pronoun when used explicitly adds ‘pragmatic weight’ to an utterance, and specifically can trigger more abstract readings of certain verbs. We interpret this result in Table 6, 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 pragmatic use of *no sé* is subjective (as seen in Table 7). 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 pragmatic as opposed to lexical is significantly higher. In Madrid, 60% of uses have a discourse / pragmatic function (Table 1). Further, just looking at pragmatic uses of *no sé*, in Madrid it is used with textual functions significantly more than in the Puerto Rican variety (Table 7), revealing significant differences in usage between the varieties.⁴ The dialectal differences we find between America and Spain may reveal similar, but divergent, diachronic pathways.

Traugott (1982, 1995b) and Traugott & 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/propositional/non-subjective use to a more subjective, pragmatic use. Traugott and Dasher (2002: 155) point out that discourse markers are subjective because they “indicate the SP[speaker]/W[riter]’s rhetorical, metatextual, stance towards the cohesiveness of the discourse being developed” by means of the way the speaker chooses to deal with

⁴ 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.

16 Variable development of intersubjectivity in Spanish

the preceding information: s/he may choose to develop, qualify, background, or foreground it. Furthermore, discourse markers are subjective in that they reflect the degree of knowledge a speaker has regarding his/her message. 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 of *no sé* is to attenuate disagreeing statements, avoid face-threatening acts, and soften the propositional stance. In so doing it acquires intersubjective 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 to us that *no sé* could proceed along the following grammaticalization cline (adapted from Traugott & Dasher 2002: 187): non-subjective/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 7) suggesting that this discourse marker has extended its use further in Spain than in America. In fact, Traugott (2003: 130) points out that the development from subjectification to intersubjectification is unidirectional, that is to say, intersubjectification can only exist after some degree of subjectification.

The dialectal division we find between America and Spain in the use of *no sé* may reflect a diachronic pathway. An alternate explanation may be found in the argument put forth in Company Company (2002). Through an analysis of relative frequencies of four distinct syntactic forms (nominal possession, diminutives, leísmo, present perfect), Company Company (2002) outlines a dialect division between Mexican and Peninsular Spanish, noting that the:

“español de España y español de México gramaticalizan varias zonas de sus respectivos sistemas siguiendo dos pautas distintas de gramaticalización, las cuales obedecen a una distinta selección o perfilamiento de un mismo complejo semántico – los rasgos externos o referenciales el primer dialecto, los rasgos internos, relacionales o valorativos el segundo ” [Spanish of Spain and Spanish of Mexico grammaticalize different areas of the respective systems following two different grammaticalization processes which respond to a different selection or outline of a similar semantic complex – external or referential features the former, internal, relational or evaluative traits the latter.] (2002:43)

In Mexico the syntactic phenomena Company Company (2002) highlights have developed primarily subjective uses, whereas in Spain, these same syntactic phenomena are used referentially. Company Company (2002:65) proposes a dialect split, therefore, based on the fact that the four independent syntactic phenomena she outlines behave in similar ways. 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.

Acknowledgements

We are indebted to the New Mexico and Southern Colorado Linguistic Atlas directors Neddy Vigil and Garland Bills for allowing us use of their data. We would also like to express our gratitude to Mayra Cortés-Torres for sharing with us the interviews that she transcribed and collected in Puerto Rico. We would like to acknowledge helpful comments we received on an earlier version of this work by Ivo Sánchez and by audience members at CILC 09 in Murcia, Spain.

References

- Aaron, J. 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. and N. Vigil. 1999. Ashes to ashes: the historical basis for dialect variation in New Mexican Spanish. *Romance Philology* 53: 43-67.
- . 2008. *The Spanish Language of New Mexico and Southern Colorado. A Linguistic Atlas*. Albuquerque, NM: The University of New Mexico Press.
- Blecua, J. et al. eds. 1994. *Actas del Congreso de la Lengua Española. Sevilla, 7 al 10 de Octubre de 1992*. Madrid: Instituto Cervantes,.
- Bravo Cladera, N. 2003. Un estudio de los marcadores del discurso y su distribución en interacciones diádicas de jóvenes bilingües. In Sánchez Miret, F., ed. vol. II: 126-133.
- Brinton, L. and E. Traugott 2005. *Lexicalization and Language Change*. Cambridge: Cambridge University Press.
- 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 Bybee, J. and P. Hopper, eds. 255-280.
- Bybee, J. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.

18 Variable development of intersubjectivity in Spanish

- . 2003. Mechanisms of change in grammaticalization: the role of frequency. In Joseph, B. and R. Janda, eds. 602-623.
- . 2007. *Frequency of Use and the Organization of Language*. Cambridge: Cambridge University Press.
- Bybee, J. and 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. 294-311].
- Bybee, J. and P. Hopper, eds. 2001. *Frequency and the Emergence of Linguistic Structure*. Amsterdam/Philadelphia: John Benjamins.
- 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. 2005. *La Perífrasis Estar + -Ndo en el Español Puertorriqueño: Variación Dialectal o Contacto Lingüístico?* Unpublished PhD dissertation. University of New Mexico, Albuquerque, New Mexico, USA.
- Cuenca, M. and M. Marín. 2009. Co-occurrence of discourse markers in Catalan and Spanish oral narrative. *Journal of Pragmatics* 41: 899-914.
- 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>.
- Fox, B., M. Hayashi and R. Jaspersen. 1996. Resources and repair: a cross-linguistic study of syntax and repair. In Ochs, E., E. Schegloff and S. Thompson eds. 185-237.
- Fraser, B. 1990. An approach to discourse markers. *Journal of Pragmatics* 14: 383-395.
- . 1999. What are discourse markers? *Journal of Pragmatics* 31: 931-952.
- Garcés Gómez, M. 1992. El operador discursivo pues en el español hablado. *Romantistisches Jahrbuch* 43: 261-276.
- Hickey, R. ed. 2003. *Motives for language change*. Cambridge: Cambridge University Press.
- Hochberg, J. 1986. Functional compensation for /s/ deletion in Puerto Rican Spanish. *Language* 62(3): 609-621.

19 Variable development of intersubjectivity in Spanish

- Hopper, P. 1991. On some principles of grammaticalization. In Traugott, E. and B. Heine, eds. 17-36.
- Hopper, P. and E. Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
- Joseph, B. and R. Janda, eds. 2003. *The Handbook of Historical Linguistics*. Oxford/New York: Blackwell.
- Jurafsky, D., A. Bell, M. Gregory and W. Raymond. 2001. Probabilistic relations between words: evidence from reduction in lexical production. In Bybee J. and P. Hopper, eds. 309–335.
- Lehmann, C. 1995. *Thoughts on Grammaticalization*. Munich: LINCOM EUROPA.
- Lehmann, W. and Y. Malkiel. eds. 1982. *Perspectives on Historical Linguistics*. Amsterdam/Philadelphia: John Benjamins.
- Martín Zorraquino, M. 1994. Gramática del discurso: los llamados marcadores del discurso. In Blecua, J. et al., eds. 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.
- Ochs, E., E. Schegloff and S. Thompson, eds. 1996. *Interaction and Grammar*. Cambridge: Cambridge University Press.
- 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. 2001. Exemplar dynamics: word frequency, lenition and contrast. In Bybee, J. and P. Hopper, eds. 123-136.
- Portolés, J. 1998. *Marcadores del Discurso*. Barcelona: Ariel.
- Rivera-Mills, S. and D. Villa. forthcoming. *Spanish of the Southwest: A Language in Transition*. Frankfurt: Iberoamericana/Vervuert.
- Sánchez Miret, F. ed. 2003. *Actas del XXIII Congreso Internacional de Lingüística y Filología Románica. Salamanca 24-30 de Septiembre de 2001*. Tübingen: Max Niemeyer Verlag.
- 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.
- Sedano, M., A. Bolívar and M. Shiro, eds. 2006. *Haciendo Lingüística. Homenaje a Paola Bentivoglio*. Caracas: Universidad Central de Venezuela.
- Serrano, M. 1999. *Bueno* como marcador de inicio de turno y contraposición: estudio sociolingüístico. *International Journal of*

- 20 Variable development of intersubjectivity in Spanish
the *Sociology of Language* 82: 115-133.
- Stein, D. and S. Wright. 1995. *Subjectivity and Subjectivisation: Linguistic Perspectives*. Cambridge: Cambridge University Press.
- Thompson, S. 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 Sedano, M., A. Bolívar and M. Shiro, eds. 347-392].
- Torres Cacoullós, R. and C. Travis. forthcoming. Variable *yo* expression in New Mexico: English influence? In Rivera-Mills, S. and D. Villa eds.
- Traugott, E. 1982. From prepositional to textual and expressive meanings: some semantic-pragmatic aspects of grammaticalization. In Lehmann, W. and Y. Malkiel, eds. 245-271.
- . 1995a. Subjectification in grammaticalization. In Stein, D. and S. Wright, eds. 31-54.
- . 1995b. The role of the development of discourse markers in a theory of grammaticalization. Available at <http://www.stanford.edu/~traugott/papers/discourse.pdf>
- . 2003. From subjectification to intersubjectification. In Hickey, R. ed. 124-139.
- Traugott, E. and B. Heine, eds. 1991. *Approaches to Grammaticalization*. Amsterdam/Philadelphia: John Benjamins.
- Traugott, E. and R. Dasher. 2002. *Regularity in Semantic Change*, Cambridge: Cambridge University Press.
- Travis, C. 2005. *Discourse Markers in Colombian Spanish: A Study in Polysemy*. Berlin / New York: Mouton de Gruyter.
- Tsui, A. 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.