

## Genre effects on subject expression in Spanish: Priming in narrative and conversation

CATHERINE E. TRAVIS

*University of New Mexico, Albuquerque*

### ABSTRACT

Structural priming refers to the process whereby the use of a syntactic structure in an utterance functions as a prime on a subsequent utterance, such that that same structure is repeated. This article investigates this phenomenon from the perspective of first-person singular subject expression in Spanish. Two dialects and two genres of spoken Spanish are studied: New Mexican narratives and Colombian Spanish conversation. An analysis of 2,000 verbs occurring with first-person singular subjects reveals that subject expression undergoes a priming effect in both data sets, but that the effect is more short-lived in the Colombian data. This is found to be attributable to the interactional nature of these data, showing that the need to deal with interactional concerns weakens the priming effect. As the first study to compare priming of subject expression across distinct genres, this article makes an important contribution to our understanding of this effect, and in particular, of factors that play a role in its maintenance or dissipation in discourse.

Structural priming refers to the process whereby the use of a certain structure in one utterance functions as a prime on a subsequent utterance, such that that same structure is repeated. This phenomenon has been observed in experimental settings in psycholinguistic research (Bock, 1986; Branigan, Pickering, & Cleland, 2000a; Hartsuiker, Pickering, & Veltkamp, 2004, *inter alia*), and in spontaneous discourse in work from the fields of sociolinguistics (Cameron & Flores-Ferrán, 2003; Poplack, 1980; Scherre & Naro, 1991, *inter alia*) and corpus linguistics (Gries, 2005, to appear; Szmrecsanyi, 2005; 2006, *inter alia*). These studies have found that speakers have a strong tendency to repeat structures that they have recently produced or heard. For example, the use of a passive instead of an active clause favors subsequent use of the passive in English (Bock, 1986; Estival, 1985; Weiner & Labov, 1983); the use of one form of the English future (*will* or *going to*) favors repeated use of that same form (Szmrecsanyi, 2005, 2006); in language varieties with variable plural expression, such as Puerto Rican Spanish

I would like to thank Rena Torres Cacoullós for her many invaluable comments on this article and Jill Morford and Benedikt Szmrecsanyi for helpful discussion on the notion of priming. I also gratefully acknowledge Jessi Aaron and Matt Alba for help with the coding and analysis; Neddy Vigil and Garland Bills for making the New Mexican data available to me; Ana Aurora Medina Murillo for help with the transcription of the NM data; and María Elena Rendón, Marianne Dieck, and Rocío Amézquita for the collection and transcription of the Colombian data.

and Brazilian Portuguese, the expression of a plural morpheme favors (and the lack of expression disfavors) its subsequent expression (Poplack, 1980; Scherre, 2001). This has profound implications for our understanding of grammar, because it demonstrates that each clause is not constructed independently, but is patterned on the preceding discourse. It suggests that speakers orient to structures that occurred previously in the discourse, and use them as partial models on which to base the morphosyntax of subsequent utterances. This supports a model of grammar as emerging through discourse (cf. Bybee, Perkins, & Pagliuca, 1994; Hopper, 1987, 1998; Ono & Thompson, 1995), rather than as an abstract entity fully contained in the mind of speakers and accessed independently for each utterance.

In this article I investigate the role of priming in first-person singular subject expression in two dialects and two genres of spoken Spanish: Colombian Spanish conversation and New Mexican (NM) Spanish narratives. In both data sets a priming effect is observed, such that a preceding coreferential unexpressed (or implicit) subject tends to lead to a subsequent unexpressed subject and a preceding coreferential expressed (or explicit) subject tends to lead to a subsequent expressed subject. However, a number of differences were identified in the two data sets that provide valuable insights into the nature of the priming effect, the configuration of subject expression, and the effect of genre on patterns of language use.

First, it was found that first-person singular subjects are much more likely to be expressed in the Colombian data than in the NM data. This is precisely the opposite of what we might expect, given the intense contact NM Spanish has with English, which has near obligatory subject expression. In fact, this finding is consistent with that of Silva-Corvalán (1994), who observed no increase in rate of subject expression in East Los Angeles across generations of speakers with different degrees of proficiency in English. Silva-Corvalán did, however, find that even though rates of use did not change, the pragmatics of use did, with the third-generation speakers losing some of the constraints that were applied by first- and second-generation speakers in her data (1994:162). For the NM and Colombian data under study here, no such corresponding variation in pragmatics of use emerged, as the constraints on subject expression were found to pattern identically in both data sets. The analysis of the linguistic conditioning of subject expression across the two corpora presents strong evidence that far from being dialectal, the different rates of expression are entirely attributable to the different genres being studied, with greater subject continuity in the NM personal narratives allowing for more subjects to be left unexpressed.

Second, the priming effect is found to be much longer lasting in the NM data, where it is maintained at a distance of up to ten intervening clauses between the two coreferential mentions, that is, between the prime and target utterances. In the Colombian data, the effect is only statistically significant in cases in which the prime and the target are adjacent, or when there is one intervening clause between them. Analysis of the linguistic conditioning of priming shows that it is favored in contexts in which the same tense is maintained, and that continuity of tense can lead to more long-term priming. Thus, the divergent results for the two corpora

regarding the duration of priming are also attributable to the genres being studied. In the conversational Colombian data, there is a high degree of shifting of tense, as interaction between the speakers leads to frequent changes in the topic of conversation. This shortens the lifespan of the priming. The narrative NM data, on the other hand, exhibit continuity of tense, as the interviewees recount their life story in largely uninterrupted narratives, and this allows for long-term priming. Although the duration of priming has been investigated in some detail in experimental settings (Bock & Griffin, 2000:178; Boyland & Anderson, 1998; Branigan, Pickering, Stewart, & McLean, 2000b; Saffran & Martin, 1997), the same is not so for more spontaneous discourse, and neither the effect of continuity of tense nor that of the need to deal with interactional concerns has been dealt with in detail in the literature to date. This finding therefore sheds light on factors that play a role in the maintenance and dissipation of priming in discourse, and helps us understand the way it works as a mechanism in shaping discourse patterns.

Third, whereas in the Colombian data both expressed and unexpressed subjects enter into the priming in the same way, in the NM data, the priming was found to apply only to unexpressed subjects. This is not because the priming functions differently in the two data sets, but is again a genre effect, in this case related to the interaction between priming and distance. The results suggest that the favoring of unexpressed subjects in the environment of coreferentiality in adjacent clauses overrides the priming effect, such that both unexpressed and expressed subjects are followed by subsequent unexpressed subjects at low degrees of distance. This means that data with a higher level of subject continuity exhibit greater priming for unexpressed than for expressed subjects, whereas data with more shifting of subjects exhibit similar behavior for both. The former corresponds to the narrative NM data, and the latter to the conversational Colombian data; that is, this result is also attributable to genre.

Overall, then, this study shows that the disparities in the two corpora are not due to different linguistic conditioning in the varieties of Spanish under analysis, but to the two genres under consideration. In contrasting these two genres, we are also able to uncover a key factor in the duration of priming, namely, that of continuity of tense, or, more generally, that of interaction itself. This demonstrates that genre has a profound effect on language patterns, and must be taken into account in order to better understand the grammar of language in use.

#### PRIMING

While repetition may occur in discourse because of register constraints, or may be used consciously to build rapport or to create coherence, and so on (cf. Tannen, 1987), structural priming refers specifically to “the unintentional and pragmatically unmotivated tendency to repeat the general syntactic pattern of an utterance” (Bock & Griffin, 2000:177). Research on structural priming has been carried out in three main areas: psycholinguistics, sociolinguistics, and corpus linguistics.

The leading research in structural priming in the field of psycholinguistics is that of Bock and colleagues (cf. Bock, 1986; Bock & Griffin, 2000; Loebell & Bock, 2003, *inter alia*). Bock has examined two types of constructions, passive versus active and prepositional versus double-object dative constructions, in experimental settings involving picture-description tasks. She has consistently found that participants tend to repeat the structure they have just used to a statistically significant degree. Furthermore, she has found that this occurs independently of lexical repetition. Thus, prepositional-object constructions with the preposition *to* (e.g., “A rock star sold some cocaine to an undercover agent”) and with *for* (e.g., “The governor poured a cup of tea for the princess”) were found to prime to a similar degree a prepositional-object construction with the preposition *to* (e.g., “The girl handed a paintbrush to the boy”) (1986:367). This demonstrates that it is not the form of the preposition that leads to priming, but the syntactic construction itself.

Although structural priming does not require lexical repetition, Pickering and Branigan (1998) have found that it can be enhanced by lexical repetition. In their study of dative priming in a written sentence-completion task, they observed that while priming was evident regardless of whether the same verb was used in the prime and target, the effect was stronger if the verb was repeated. Thus, the prime “the architect gave the engineer . . .” (a sentence fragment participants tended to complete with a double-object construction, e.g., with “the latest plans”) had a stronger effect on the target utterance “the teacher gave . . .” than it did on the target utterance “the teacher sent . . .” (1998:641). That is, “the teacher gave . . .” was more likely to be completed with a double-object construction than was “the teacher sent . . .” in this context, and likewise for the prepositional-object construction. Interestingly, the strength of the priming was not affected by repetition of the tense (1998:643) nor by repetition of number of the subject (1998:645). These results suggest that priming is enhanced by lexical repetition, but is not affected by structural repetition. Lexical enhancement of priming has been observed in natural discourse by Gries (2005:373) and Szmrecsanyi (2006:192), but in contrast to Pickering and Branigan (1998), these scholars also observed enhancement of priming due to structural repetition, though the effect was weaker than that of lexical repetition.

The notion of lexical enhancement is relevant here, because one of the variants being studied, that of an expressed subject, involves lexical repetition (namely, *yo* ‘I’), whereas the other, that of an unexpressed subject, is strictly a structural phenomenon. Thus, we may expect a stronger effect for the expressed subject. This does not in fact turn out to be the case, but this is due to other constraints on subject expression, as will be discussed further on. Interestingly, repetition of tense, but not of the verb type, was found to enhance the priming in these data, suggesting that the priming of subject expression in these data undergoes structural, but not lexical, enhancement, in contrast to the results of Pickering and Branigan (1998), and partially contrasting with those of Gries (2005:373) and Szmrecsanyi (2006:192), who found that lexical repetition had a stronger effect

than structural repetition. The significance of these contradictory findings will be discussed further later in this article.

Turning now to research in sociolinguistics, the first study to observe a priming effect was that by Poplack (1980) on plural expression in Puerto Rican Spanish. Basing her analysis on a series of sociolinguistic interviews conducted with Puerto Ricans residing in the United States, Poplack found that in this dialect with variable (s) expression, one factor that affects its realization in plurals in the Noun Phrase is its expression on preceding elements in the same Noun Phrase: "Presence of a plural marker before the token favors marker retention on that token, whereas absence of a preceding marker favors deletion" (1980:63).

Scherre and Naro (1991, 1992) and Scherre (2001) found something similar for subject/verb agreement and subject/predicate adjective agreement in Brazilian Portuguese. They also studied plural marking, which, as in Puerto Rican Spanish, is variably expressed. Using a corpus of sociolinguistic interviews, they found that plural is more likely to be morphologically marked on both verbs and predicate adjectives if it is marked on preceding elements in the clause, or in the preceding clause. Thus, while Poplack (1980) identified priming effects at the clausal level, Scherre and Naro (1991, 1992) found such effects both at the clausal and at the discourse level.

These studies investigated priming in morphology. Similar findings have been made for syntactic variables. In Weiner and Labov's (1983) study of the use of the agentless passive in sociolinguistic interviews in English, they found that one of the strongest factors to account for the use of a passive was the occurrence of another passive anywhere in the preceding five clauses, demonstrating that priming effects do not dissipate immediately, but can be maintained over intervening material (1983:52).

Priming has also been observed for subject expression in Spanish in research by Travis (2005b) for first-person singular subjects in Colombian Spanish conversation, and for all persons in sociolinguistic interviews by Cameron (1994), for Madrid, Spain, and San Juan, Puerto Rico, and by Flores-Ferrán (2002), for Puerto Ricans living in New York City. Travis (2005b) examined coreferential first-person singular subjects, and found that the form of the preceding coreferential form tended to be repeated at low degrees of distance, namely across adjacent clauses or if there was one clause separating the prime and target. Cameron looked at adjacent clauses regardless of coreferentiality, and found that "pronouns lead to pronouns and null subjects lead to null subjects" (1994:40).<sup>1</sup> While the tendency was evident in Madrid and San Juan for both coreferential and non-coreferential subjects, it was weaker for noncoreferential subjects in both data sets, and did not reach significance in this environment in the Madrid data. That is, where there was a switch in subject, the realization of the subject of the preceding clause did not have a significant priming effect on the realization of the subject of the subsequent clause in these data. As Cameron noted, this may be due to the tendency of switched subjects to be expressed (1994:40), which appears to override the tendency for the form to be repeated. Flores-Ferrán looked at priming between coreferential subjects at a distance of up to ten clauses, and found

that pronouns were most likely to be used in contexts in which the preceding mention of that same subject anywhere in the last 10 clauses was also pronominal, and unexpressed subjects were most likely found where the preceding mention was unexpressed (2002:69–71).

It is only recently that priming has been investigated from the perspective of corpus linguistics, through the work of Gries (2005; to appear) and Szmrecsanyi (2005, 2006). These studies involve the use of computational tools to conduct quantitative analyses of variable syntactic phenomena in corpora comprising several million words and various registers and genres (specifically, the International Corpus of English and the British National Corpus). Priming has been examined in such corpora in terms of: the future (*will / be going to*) (Szmrecsanyi, 2005, 2006); particle placement (*John picked up the book / John picked the book up*) (Gries, 2005, to appear; Szmrecsanyi, 2005, 2006); the dative construction (*he gave the book to Mary / he gave Mary the book*) (Gries, 2005, to appear), and so on. On the basis of detailed statistical testing of large numbers of tokens (e.g., over 35,000 in Szmrecsanyi's (2005) analysis of the future), priming was found to be one of a set of factors that have a significant effect on variant choice for all of the phenomena examined.

Branigan et al. (1995:492) and Pickering and Branigan (1999:136) have questioned whether priming can be unequivocally identified in corpora, due to the fact that it is not possible to control for the many other factors that may lead to repetition (such as register constraints, rapport building, coherence, etc.). Indeed, in order to allow for these other possible motivations, Szmrecsanyi used the broader term “persistence” rather than the more specific “priming” (2005:144). Gries (2005:385–387), however, argued that his data demonstrate that priming can be identified as an independent factor in natural discourse, firstly because his findings mirror those from experimental research, and secondly with such a large-scale analysis other factors are able to be taken into account. Following Gries, I will use the term “priming” to refer to the phenomenon under examination here. While recognizing that other potentially interfering factors may not have been exhaustively ruled out, the claim that this is indeed structural priming does have a firm basis, given that (a) only cases where there is no observable semantic or pragmatic difference between the expressed and unexpressed subject are included, and (b) when the range of factors I have identified as potentially affecting subject expression are considered simultaneously in multivariate analyses, the form of the previous mention of the coreferential subject emerges as among those that exert the strongest influence on this choice.

#### SUBJECT EXPRESSION IN SPANISH

Subject expression is one of the most widely studied features of Spanish syntax, yet it remains one of the least understood; factors that are argued to affect subject expression in one study are argued not to do so in others. The most robust and consistent finding across a range of different studies and dialects is in relation to

subject continuity. Subjects are more likely to be unexpressed when they are also the subject of the preceding clause, and are more likely to be expressed when there is a switch in subject from that of the preceding clause. This is the single factor that has been found to affect subject expression in all dialects studied and for all persons (cf. Ávila-Shah, 2000; Bayley & Pease-Alvarez, 1997; Bentivoglio, 1987; Cameron, 1994, 1995; Enríquez, 1984; Flores-Ferrán, 2002, 2004; Hochberg, 1986; Morales, 1986; Silva-Corvalán, 1982, 1994; Travis, 2005b, *inter alia*). This pattern is illustrated in the following examples from the two data sets, where the first mention is expressed, and subsequent mentions are unexpressed. (First-person singular verbs are double underlined, and unexpressed subjects are given in parentheses in the English translation).

- (1) *Y yo los bañaba, y los vestía, les daba de comer, los ponía a dormir,*  
 ‘And I would bathe them, dress them, feed them, put them to sleep,’  
 [NMCOS, 117–1A3: 248–251]<sup>2</sup>
- (2) *Yo la pongo encima de la mesa? ... Se la pele, y le quito así el piquito,*  
 ‘I put it on top of the table? ... (I) peel it for her, and (I) pull out the tip,’  
 [Colombia, almuerzo: 1494–1496]

Despite the strong findings in relation to coreferentiality, this only accounts for a portion of the data. In the Colombian data, even in cases in which there is continuity of subject, first-person subjects are still explicit close to 40% of the time, and in the NM data they are explicit close to one quarter of the time. Similar results have been obtained in other studies, and it is generally found that subjects are explicit between 20% and 40% of the time when they are coreferential with the subject of the preceding clause (Bentivoglio, 1987:55; Cameron, 1995:25; Flores-Ferrán, 2004:63; Silva-Corvalán, 1994:158). The following examples illustrate expressed subjects in this context.

- (3) *Y luego ya no pude ir más, porque yo iba muy lejos pa’ agarrar el bus,*  
 ‘And then (I) couldn’t go anymore, because I had to go really far to catch the bus,’  
 [NMCOS, 76–1A1: 228–229]
- (4) *Mañana voy. Yo dejé diez paquetes allá.*  
 ‘(I) will go tomorrow. I left ten packets there.’  
 [Colombia, cooking: 100–101]
- (5) *Ahí estaba yo, comenzaba yo a hablar en inglés.*  
*Yo me comencé en el nineteen fifty two.*  
 ‘There I was, I started to speak in English. I started in 1952.’  
 [NMCOS, 76–1A1: 88–90]
- (6) *Yo no sabía cuánto yo quiero a mi hermano.*  
 ‘I didn’t know how much I love my brother.’  
 [Colombia, Calima2: 642]

Note that while (1) and (2) represent what might be considered the canonical pattern, an expressed subject followed by a string of unexpressed subjects, examples (3)–(6) represent a deviation from this. In (3) and (4), we have an unexpressed subject followed by an expressed subject, and in (5) and (6), an expressed subject

followed by subsequent expressed subjects. These patterns do not have equal distribution in the data: that observed in (3) and (4) occurs only marginally, and what is much more common is a clustering together of expressed, or unexpressed, subjects as illustrated in the other examples.

This suggests that coreferentiality is just one of many factors that affect Spanish subject expression, as has been shown in the large body of research in this area. Other linguistic factors that have been found to affect subject expression (and were also found to do so in the data under consideration here) include the semantics of the verb (Bentivoglio, 1987; Enríquez, 1984; Silva-Corvalán, 1994); tense (Bayley & Pease-Alvarez, 1997; Bentivoglio, 1987; Cameron, 1994; Enríquez, 1984; Flores-Ferrán, 2002; Hochberg, 1986; Ranson, 1991; Silva-Corvalán, 1994); continuity not just of subject, but also of verb type and of tense across clauses (Ávila-Shah, 2000; Bayley & Pease-Alvarez, 1997; Paredes Silva, 1993); the form of the preceding subject, that is, priming (Cameron, 1994; Cameron & Flores-Ferrán, 2003; Flores-Ferrán, 2002), and so on. These factors will be investigated in more detail later.

Prior research in this area suggests that subject expression is affected more by linguistic factors such as those just mentioned, than by extralinguistic factors, such as gender, age, socioeconomic status, and so on (e.g., Bentivoglio, 1987; Silva-Corvalán, 2001). Regional dialect has, however, been found to be a relevant factor in subject expression. In particular, Caribbean dialects are reported to demonstrate higher rates of subject expression than other varieties (cf. Cameron, 1992, 1993; Lipski, 1994; Otheguy & Zentella, *in press*; Otheguy, Zentella, Erker, & Livert, 2005). Nevertheless, despite divergent rates of expression, no study has reported different grammatical patterning across those dialects, and instead remarkably similar results have been found. Cameron (1993, 1994), for example, compared subject expression in interview data from San Juan, Puerto Rico, and Madrid, Spain, and found that despite differences in rates of expression, the grammatical patterning of subject expression was very similar across the two dialects. This is precisely what I have found here, as will be presented in the following discussion.

One extralinguistic factor that has been sorely overlooked is that of genre, which is discussed in more detail in the following section.

#### GENRE EFFECTS

##### *Genre effects on subject expression*

The vast majority of research on Spanish subject expression is based on spoken narratives collected via sociolinguistic interviews, with some work also being done on written narratives (Bayley & Pease-Alvarez, 1997), and as far as I am aware, one study based on spontaneous conversation (Travis, 2005b). In the light of Biber's findings on the enormous variation that exists across registers (1986, 1988, 1992, *inter alia*) and his proposal that "most functional descriptions of a grammatical feature will not be valid for the language as a whole" (2001:104), the homogeneous nature of the data for which subject expression has been studied



gives us a very limited understanding of how it patterns, as has been pointed out by Silva-Corvalán (2001:163).

There is some evidence that genre may have an effect on subject expression in Spanish, from work done comparing narrative style. In her study of Spanish spoken in Valladolid, Mexico, Solomon (1999) found a much higher rate of subject expression in conflict narratives, and in particular in narratives that involved a conflict between the narrator and another character or characters. She proposed that this is because speakers have high personal stakes in such narratives, and therefore use overt subjects to highlight their role in the discourse. Flores-Ferrán also found that narratives that recounted some personal conflict have a significantly higher rate of expressed subjects than nonconflict narratives in her New York Puerto Rican data (2002:93).

A genre effect has been identified for subject expression in Mandarin Chinese in the work of Jia and Bayley (2002), who examined data drawn from telephone conversations and classroom discourse. These scholars found a significant difference in the rate of use of the second-person singular and plural depending on the setting. For the classroom data, the second-person singular pronoun tended to be expressed and the plural unexpressed, whereas for the telephone conversations, it was the plural form that tended to be expressed and the singular unexpressed. They proposed that the form that is most expected in each setting (the plural in the classroom and the singular in the telephone conversations) is that which is left unexpressed, whereas what is likely to be the less frequent form is expressed (2002:112). A similar proposal is put forward by Paredes Silva, for different rates of first- and second-person subject expression in personal letters in Brazilian Portuguese (1993:45–46).

It is unclear how a notion of “expectedness” might apply to the two genres under study here. We might anticipate a high proportion of first-person singular subjects in both, as speakers assert their role in the interaction in conversation and recount events of their past lives in personal narratives. Nevertheless, such studies are important for the issues under consideration here because they demonstrate that the discourse setting can have a significant effect on subject expression.

### *Genre effects on priming*

Just as research on subject expression is lacking in cross-genre comparisons, so too is research on priming. Priming is known to take place in constrained settings, such as for picture-description and sentence-completion tasks, as have been applied in psycholinguistics (e.g., Bock, 1986; Bock & Griffin, 2000; Branigan et al., 2000a, 2000b; Hartsuiker et al., 2004; Loebell & Bock, 2003; Pickering & Branigan, 1998). It is also known to occur in primarily monologic personal narratives collected in sociolinguistic research (e.g., Cameron, 1992, 1994; Flores-Ferrán, 2002; Poplack, 1980; Scherre, 2001; Scherre & Naro, 1991; Weiner & Labov, 1983). And from research in corpus linguistics we know that priming occurs in natural data from both written and spoken sources, with varying degrees of interaction and formality (Gries, 2005, to appear; Szmrecsanyi, 2005, 2006). It is in

the latter field where there has been some investigation into genre effects on priming.

Both Szmrecsanyi (2006:202) and Gries (to appear) reported that the priming effect is stronger in spoken than written data, a finding that mirrors what has been found in written and spoken sentence-completion tasks in psycholinguistic research (Branigan et al., 2000a). Across spoken genres, Szmrecsanyi (2006) found stronger priming in the less formal data. He proposed that priming is weaker in writing and formal speech because speakers edit out such repetition in data that allows greater planning and editing. Intriguingly, Szmrecsanyi (2006) also found that, although the effect may be weaker in more formal data, it is also more long-lasting. For example, priming ceases to have an effect on choice of future form when there are 150 words intervening between the prime and target in recordings from informal encounters (taken from the demographically sampled component of the British National Corpus), but it remains in effect up to a distance of 500 words in oral histories (taken from Freiburg English Dialect Corpus) (2006:189). A similar pattern is found for the other four features he examined (particle placement, comparison strategies, genitive choice, complementation strategies)—though the specific point at which the priming effect is lost varies somewhat for each feature, priming is consistently retained for longer in the more formal data.

The duration of priming has received quite some attention in the psycholinguistic literature. Research on lexical priming has found that the effects are very short-lived, such that intervening items between the prime and target greatly weaken the priming effect (Joordens & Besner, 1992; Meyer, Schvaneveldt, & Ruddy, 1972, *inter alia*). This is because lexical priming is believed to be the result of activation, whereby exposure to one word temporarily activates that word in the brain, as is evidenced by facilitation of subsequent use (or recognition) of that same word or of a semantically related word (e.g., *bread* and *butter*) (cf. Meyer & Schvaneveldt, 1971). The priming effect is short-lived because activation necessarily weakens over time, either through the use of competing forms or simply due to the passing of time (Bock & Griffin, 2000:178; Pickering, Branigan, Cleland, & Stewart, 2000:212).

Lexical priming has been much more widely studied than structural priming, and the duration of structural priming remains under debate. There is, however, some evidence that structural priming can be long-lasting. Based on a spoken sentence-completion task using dative constructions, Branigan et al. (2000b) found that the priming effect was maintained when the participant produced one intervening sentence between the prime and target, or when there was an intervening time lag. Bock and Griffin found that priming for dative constructions was maintained to a significant degree even after ten intervening sentences were produced by the participant (2000).<sup>3</sup> Boyland and Anderson (1998) found that priming effects for passive versus active and dative constructions were maintained twenty minutes after speakers had been exposed to these constructions, specifically, after having produced multiple times one of either the double-object construction or the prepositional-dative construction, and one of either the passive or active construction. Finally, Saffran and Martin (1997) found that aphasics used dative

constructions with which they had been primed at a consistently higher rate than before the priming, even one week after the priming took place.

Based on these findings, Bock and colleagues have argued that structural priming should be understood as a different mechanism from lexical priming. They have proposed that structural priming does not function via activation but via implicit, or procedural, learning. According to a procedural-learning model, the cognitive mechanisms for producing a certain structure are tuned and strengthened through use, such that this structure becomes more readily produced, and therefore is used more than an alternative form (cf. Bock & Griffin, 2000; Chang, Dell, Bock, & Griffin, 2000). Procedural learning has long-term effects because it creates a permanent change in our cognitive processing. Thus, the finding that structural priming can be long-lasting has important implications for our understanding of how it operates as a mechanism in shaping discourse patterns.

The data presented here support the hypothesis that priming can be long-lasting in spoken discourse, and also Szmrecsanyi's (2006) finding that priming is more long-lasting in more formal genres (in this case, the narratives collected via interview). I propose that this is an epiphenomenon of the narrative genre and the high degree of continuity that this genre exhibits.

#### DATA

The NM and Colombian data come from two distinct corpora. The NM data were taken from the materials of the New Mexico Colorado Spanish Survey (NMCOS), a corpus of recorded interviews of 350 speakers conducted in 1992–1995 under the direction of Garland D. Bills and Neddy A. Vigil of the University of New Mexico (cf. Bills & Vigil, 1999). For this study, 11 interviews were used, representing a total of four and a half hours of speech or roughly 45,000 words. These interviews involve six females and five males over the age of 48 from northern New Mexico. All are native Spanish speakers who maintain proficiency in Spanish and have varying degrees of English proficiency. This corpus thus allows us to study subject expression in a variety of Spanish that is in close contact with English, but for speakers who are not undergoing language attrition. The interviews were conducted by NM graduate students from the University of New Mexico, and in most cases they did not have a personal relationship with the interviewee, though often they were from similar communities. The data consist primarily of narratives, as the interviewees tell their life stories and of their experiences in schooling and growing up in rural New Mexico.

The Colombian data were taken from a corpus of spontaneous conversation collected in the city of Cali, Colombia, in 1997 (cf. Travis, 2005a). A total of four and a half hours of conversation, or 42,500 words, were used. These four and a half hours comprise fifteen conversations of between two and four participants, and involve 22 speakers (14 women and 8 men). All speakers are middle-class native Colombians, ranging between the ages of 20 and 60. The data were collected by two participants, who recorded spontaneous conversations between

themselves and their husbands, family, and friends over a period of two months. These data are therefore as natural as possible in a situation in which participants are aware that they are being recorded.

The data from both sources were transcribed in accordance with the approach developed at the University of California, Santa Barbara (cf. Du Bois, Schuetze-Coburn, Cumming, & Paolino, 1993). The transcription conventions are given in the Appendix.

As this outline indicates, these two data sets represent distinct dialects and distinct genres. We have one variety that exists in a contact situation, and another which is largely monolingual. NM Spanish shows great influence from English at the level of the lexicon (Bills, 1997; Bills & Vigil, 1999, 2000; Vigil & Bills, 1997, 2000, 2004), and although the grammatical effects of this contact have not been investigated in depth, we may expect to find some effect on the patterns of subject expression (cf. Silva-Corvalán, 1982, 1994, who found that degree of contact with English did affect the patterns of subject expression in Spanish speakers in East Los Angeles). In terms of genre differences, the NM data consist of largely monologic narratives told by rural, older people to graduate students, whereas the Colombian data consist of highly interactional spontaneous conversation between people from the same social group. Although it is not possible to make a direct comparison of Colombian conversation and NM narratives, the results do permit us to hypothesize about the role dialect and genre each play in the differences that emerge in the two corpora. As noted earlier, and as will be outlined in more detail later, it is found that there is a great difference in terms of rate of expression, but almost no difference in terms of the grammatical patterning of expression in the two data sets. This suggests that the differences observed are a result of genre and not dialectal differences, as will be discussed further on.

## METHODOLOGY

### *First-person singular subjects*

This study analyzes subject expression for first-person singular subjects only. As prior research has already shown that priming has an effect on subject expression for all persons, and for both coreferential and noncoreferential subjects (Cameron, 1992, 1994; Flores-Ferrán, 2002), in an effort to better understand the way this phenomenon operates, I look in detail at just the first-person singular. In this way, the effects of interactional and discourse pressures that do not apply equally to all persons are controlled.

For example, unlike third-person subjects, first- and second-person subjects are not affected by issues related to information flow. They can always be considered given information because they are present in the context (cf. Chafe, 1994). In light of the robustness of the findings regarding subject continuity, a context in which the subject is necessarily given information, this distinction should not be treated lightly. First-person subjects also differ from other subjects

TABLE 1. *Coding and exclusions*

	NM	Colombia
All tensed verbs with first-person singular subjects	1210–	1182–
Exclusions (nonvariable contexts; pragmatic uses; first mentions; following truncated uses; preceding 1sg. subject = English; interviewer uses)	357	304
Total included in statistical analysis:	853	878
Expressed subjects	284 (33%)	421 (48%)
Unexpressed subjects	569 (66%)	457 (52%)

in terms of the role they play in expressing epistemicity, as it is through use of the first person that speakers can weaken or strengthen their stance towards an utterance, by using expressions such as *(yo) creo* and *(yo) pienso* ‘I think’ (cf. Bentivoglio, 1987:52; Davidson, 1996; Enríquez, 1984:236; Travis, 2006). And finally, first person singular only has two forms, namely, lack of expression, or use of the pronoun *yo*, whereas second person singular in this dialect has three different pronominal forms (*tú, vos, usted*), and third person can be expressed, not just by pronouns, but also by full Noun Phrases.

Limiting the study to first-person singular subjects provides a more homogeneous set to work with, eliminating variation in relation to these factors which are not fully understood.

### *Coding*

All finite verbs with first-person singular subjects in the two data sets were coded. A strictly syntactic definition of subject was adopted. Thus, experiencer constructions, such as *me gusta* ‘I like it’, were not included, although it is recognized that the dative NP in such constructions plays a similar function to that of subject in others.<sup>4</sup> These constructions were excluded because of the broader research question of priming. Given the fact that the first person does play a different structural role here and that it takes a different form from when it occurs as the syntactic subject (*me* vs. *yo*), it is not clear to what degree it might enter into priming relations.

Table 1 provides an overview of the coding. A total of 1,210 verbs in the NM data and 1,182 verbs in the Colombian data were extracted. Of these, a number of exclusions were made. Either because the context did not allow variation (i.e., the subject is obligatorily expressed or unexpressed); the subject was clearly playing a pragmatic role, such as being used for emphasis (e.g., if it was followed by *sí*, as in *tú no ves esas cosas en tu familia, pero yo sí las veo* ‘you don’t see those things in your family, but I do see them’ [Colombia, restaurant 663–664]); or it was not possible to determine whether a priming effect was involved (e.g., for the first mention of that first-person singular subject in the transcript and following a

truncated utterance in which there is a first-person singular subject but no verb). For the NM data, two further exclusions were made: cases in which the preceding subject was produced in English (in code-switching), in order to control for monolingual priming, and all interviewer uses. This left a total of 853 tokens for the NM data and 878 tokens for the Colombian data, which were then subjected to the statistical analysis.

As Table 1 shows, the rate of subject expression differs markedly in the corpora, with subjects being expressed one third of the time in the NM data and close to one half of the time in the Colombian data, a difference that a chi-square test revealed to be significant ( $p \leq .05$ , chi-square = 3.9744). However, these figures may be misleading, because as Poplack and Tagliamonte have noted, different rates may be a result of any number of extra-linguistic factors (2001:92). True grammatical differences are evidenced in the linguistic conditioning of variability, or the patterns of co-occurrence observed in the data. As we will see later, in both data sets, subject expression shows almost identical patterns of use, suggesting that the divergent rates observed are epiphenomenal, and are a result of the different genres under consideration.

These tokens were coded in Excel for the following factors: semantic class of the verb, tense / aspect / mood (TAM), distance from previous mention (up to 10 clauses), realization of previous mention (expressed or unexpressed), clause type (main or subordinate), relationship with previous TAM (if the same TAM was maintained or if there was a change), and position in the turn (initial or medial). The Colombian and NM data were subjected to independent variable rule analyses using the program GOLDVARB 2001 (cf. Rand & Sankoff, 1990), a program that uses multiple regression to identify the effect of individual factors on variant choice when a set of factors is considered simultaneously. Two sets of analyses will be reported on here: first, the analyses of factors affecting subject expression; and second, those of the factors affecting priming.

## RESULTS FOR SUBJECT EXPRESSION

Table 2 presents the results of the independent variable rule analyses for subject expression in the NM and Colombian data. It is by comparing these results that we are able to identify underlying differences and similarities in the linguistic conditioning of subject expression across the two corpora, in order to determine whether the contrasting rates observed are representative of grammatical differences, or whether they should be attributed to other factors.

These results provide three pieces of evidence to help us understand the conditioning of variation in subject expression (cf. Bayley, 2002; Poplack & Tagliamonte, 2001:92–95).

1. Those groups of factors that have a significant effect on variant choice are distinguished from those that do not. As Table 2 shows, the NM and Colombian data behave in precisely the same way, with verb class, distance from previous mention,

TABLE 2. *Two independent variable rule analyses of the contribution of factors selected as significant to the probability of expressed subjects in NM and Colombian data*

	NMCOS			Colombia		
Total N	853			878		
% expressed S	33%			48%		
Corrected mean	.31			.48		
	Weight	%	% of data	Weight	%	% of data
<b>Verb class</b>						
psychological	.70	55	18	.68	67	20
copula	.55	39	7	.63	60	6
speech	.53	35	20	.53	51	16
other	.43	25	41	.42	39	47
motion	.35	20	12	.36	31	9
Range	35			32		
<b>Distance</b>						
5+ clauses	.63	46	23	.58	55	38
2-4 clauses	.58	39	16	.51	49	17
1 clause	.53	34	16	.45	44	13
0 (subject continuity)	.39	23	44	.42	38	30
Range	24			16		
<b>Previous realization</b>						
expressed	.67	50	35	.57	55	49
unexpressed	.41	23	64	.43	40	50
Range	26			14		
<b>TAM</b>						
ambiguous TAM	.62	37	22	.62	60	10
unambiguous TAM	.47	32	77	.48	46	89
Range	15			14		

NMCOS:  $p = .003$ ; log likelihood = -476.040; chi-square/cell = 1.1082. Factor groups not selected as significant: clause type, relationship with previous TAM; position in turn.

Colombia:  $p = .013$ ; log likelihood = -563.736; chi-square/cell = 1.2097. Factor groups not selected as significant: clause type, relationship with previous TAM; position in turn.

realization of previous mention, and TAM having a significant effect, as opposed to clause type, relationship with previous TAM and position in the turn, which do not.

2. The hierarchy of constraints, or the ranking of the factors within each factor group, is identified. The percentage of subjects that are expressed in co-occurrence with each factor (given in the second column) are translated into probability weights in the multivariate analysis (given in the first column), which indicate the probability that the subject will be expressed with each factor, independently of the other factors with which each may co-occur. Note that here also the same results are achieved in the two corpora, that is, the ranking of factors that most favor (with

weights above .50), to those that least favor (with weights below .50) subject expression is precisely the same in each factor group.

3. The magnitude of effect of the factor groups is determined. This is captured in the range, which represents the difference between the factor that most favors realization of the variant (with the weight closest to 1), and that which least favors its realization (with the weight closest to 0). Once again, we get similar results across the two data sets, with verb class having the strongest effect (with a range of 35 in the NM data and 32 in the Colombian data), and TAM having one of the weakest effects (with a range of 15 and 14, respectively). However, the ranking of distance and previous realization (i.e., priming) differs, with previous realization having the second strongest effect in the NM data (with a range of 26, compared to distance, with a range of 24), and the third strongest effect in the Colombian data (with a range of 14, compared to distance, with a range of 16). Note that this is the only difference in patterning between the two data sets, and this will be explored in detail later.

In summary, these results show that the linguistic conditioning of subject expression in the two data sets is identical, with the exception of a relatively stronger priming effect in the NM than the Colombian data. These results thus fail to account for the varying rates of subject expression, that is, for the fact that subjects are expressed just one third of the time in the NM data and close to half the time in the Colombian data, and therefore we must seek an explanation elsewhere. Before moving on to this, we will consider the results for each of the factor groups in more detail.

### *Verb class*

A number of studies have found that subject expression interacts with verbal semantics. Bentivoglio (1987:60), Enríquez (1984:240), and Silva-Corvalán (1994:162) noted that verbs that express the opinion of the speaker, such as *creer* ‘think, believe’ and *suponer* ‘suppose’, favor explicit subjects more than other verb classes. In the case of the first person, the high use of explicit subjects with these verbs may be related to the epistemic role such constructions play (cf. Scheibman, 2001; Thompson, 2002).

The categories applied here are adapted from Bentivoglio (1987:50) and Enríquez (1984:151–153), with some modifications to better suit the data. Table 3 lists the categories used with examples of the most frequently occurring verbs in each category.

Returning now to the results of the variable rule analysis presented in Table 2, note that psychological verbs and copulas most favor explicit subjects, with probability weights of .70 and .55, respectively in the NM data, and .68 and .63 in the Colombian data, followed by speech act verbs (.53 for both) and “other” (.43 and .42, respectively), with motion verbs most disfavoring explicit subjects (.35 and .36). This result is consistent with previous findings (cf. Bentivoglio, 1987; Enríquez, 1984; Silva-Corvalán, 1994). The inherently epistemic nature of cognitive verbs may account for their high rate of subject expression. Because they



TABLE 3. *Categorization of verb classes*

Psychological:	<i>saber</i> 'know', <i>creer</i> 'believe', <i>pensar</i> 'think', <i>acordarse</i> 'remember', <i>imaginarse</i> 'imagine', <i>entender</i> 'understand'
Speech act:	<i>decir</i> 'say', <i>llamar</i> 'call', <i>pedir</i> 'ask, request', <i>contar</i> 'tell', <i>preguntar</i> 'ask'
Copula:	<i>ser</i> , <i>estar</i> 'be', <i>quedar(se)</i> 'be, stay'
Motion:	<i>ir</i> 'go', <i>venir</i> 'come', <i>llevar</i> 'take', <i>traer</i> 'bring'
Other:	all verbs that do not fit into the above categories; <i>tener</i> 'have', <i>hacer</i> 'do', <i>querer</i> 'want', <i>trabajar</i> 'work', <i>dar</i> 'give'

are used to express speaker opinion, the speaker asserts their role in the utterance with an expressed subject.

The finding that copulas favor subject expression is in accordance with Enríquez (1984:240), who found that stative verbs in general favor explicit subjects. It is also in accordance with Ashby and Bentivoglio (1993:65), who noted that copula subjects behave differently from subjects of other intransitive verbs, in that they tend not to occur as full Noun Phrases in both Spanish and French. Note that motion verbs (the majority of which are intransitive in these data) disfavor expressed subjects, though why this should be the case is not clear.

The fact that speech act verbs favor explicit subjects slightly more than "other" verbs may be attributable to the use of the verb *decir* 'say'. In both data sets *decir* accounts for over two thirds of the speech act verbs (80% in the NM data and 69% in the Colombian data), and in fact it is overwhelmingly the single most frequent verb to occur with first-person singular subjects in both data sets.<sup>5</sup> A closer examination of this class reveals that whereas *decir* favors subject expression, the remaining speech act verbs pattern similarly to the "other" verb class, and disfavor expression. The high rate of explicit subjects with *decir* may be related to its use as a marker of epistemicity (e.g., in the expression *yo digo que* ... literally, 'I say that', used to introduce an opinion), and perhaps this use of *decir* should be considered independently from other speech act verbs in future studies, and should possibly be treated as part of the class of psychological verbs.

A final point worthy of mention about this factor group is that not only do we get the same constraint hierarchies and a similar magnitude of effect in both corpora, but we also get remarkably similar distribution across the verbal categories, reflected in the third column in Table 2 labeled "percentage of data." In both data sets, the "other" category makes up close to half of the verbs occurring with first-person singular subjects, psychological and speech act verbs approximately 20% each, and copula and motion verbs approximately 10% each. A chi-square test reveals that there is no significant difference in the distribution observed ( $p \leq 1$ ; chi-square = 1.4643). Thus, though we may expect to find varying rates of use of certain types of verbs across different genres, this is not found to be the case here.

*Tense/aspect/mood*

One factor that has been widely tested in the literature on subject expression is that of potential ambiguity in the verb form. It is often assumed that unexpressed subjects are allowed in Spanish because verbs carry person and number marking, and therefore in many contexts an explicit subject is redundant (cf. discussion in Toribio, 1996:409–411). There are, however, some cases in which the verb form is ambiguous. For example, for the conditional, imperfect, and subjunctive, first-person and third-person singular take the same form. It has been proposed that explicit subjects may be used to resolve this ambiguity (Hochberg, 1986). A number of quantitative studies have found a correlation between ambiguous verb forms and expressed subjects (Bayley & Pease-Alvarez, 1997; Cameron, 1994; Hochberg, 1986; Silva-Corvalán, 1994), but others have found no such correlation (Bentivoglio, 1987; Enríquez, 1984; Ranson, 1991). It has also been noted that cases of true ambiguity are rare in natural discourse, as even with unexpressed subjects the morphological ambiguity is generally resolved by context (Ávila-Shah, 2000:242; Bentivoglio, 1987:45). This suggests that the function of the subject is something other than to resolve the ambiguity of the verb, as has been argued from within functionalist (Silva-Corvalán, 1997, 2001) as well as formalist (Toribio, 1996) frameworks.

TAM was included in the current study to investigate whether ambiguous and nonambiguous forms behaved differently. Thus, a broad-based, two-way distinction was made between ambiguous TAMs (conditional, imperfect, pluperfect, and subjunctive) and unambiguous TAMs (future, present indicative, present perfect, and preterit). As seen in Table 2, ambiguous TAMs favor expressed subjects with a probability weight of .62 in both data sets, and unambiguous TAMs show little effect, with a weight of .47 in the NM data and .48 in the Colombian data. These results thus partly support the notion that ambiguous TAMs favor subject expression. They must, however, be interpreted with caution for two reasons. Ambiguous verb forms are only a small proportion of the total number of verbs in each data set, accounting for less than one quarter of the NM data and just 10% of the Colombian data. Furthermore, this factor group has one of the lowest magnitudes of effect, with a range of just 15 in the NM data and 14 in the Colombian data. Thus, the results for the effect of ambiguity of TAM remain inconclusive, and would need to be investigated in a larger data set.

In terms of TAM distribution, a chi-square test of the difference between the two data sets finds this to be significant ( $p \leq .025$ , chi-square = 5.3675). Thus, the NM data show a significantly higher use of the ambiguous tenses (primarily the imperfect indicative) than the Colombian data. This is no doubt also a genre effect, with the regular use of the imperfect in narratives (cf. Biber, 1988, who noted the greater use of the past tense in narrative).

Silva-Corvalán (1997, 2001) has proposed that it is not the ambiguity but the discourse function of the different TAMs that motivates their use with expressed or unexpressed subjects. She observed that those TAMs that happen to be morphologically ambiguous (such as the conditional, imperfect, and subjunctive) are

nonfactual, nonassertive, and mark backgrounded events, whereas those that are not ambiguous (such as the present and the preterit) are factual and assertive, and that the preterit specifically marks foregrounded events. Explicit subjects are more likely to occur with the conditional, imperfect, and subjunctive, because of their backgrounding nature, and are less likely to occur with the preterit, because of its foregrounding nature, whereas the present tense is expected to show little effect (2001:161–163). This is supported by Bayley and Pease-Alvarez's study, in which they argued that the discourse function of these verb forms better accounts for their data than does the notion of morphological ambiguity (1997:363). In an independent variable rule analysis using this three-way breakdown based on the discourse function of the TAM (instead of the two-way breakdown based on ambiguity), the backgrounded TAMs were found to most favor subject expression in both data sets. However, whereas in the NM data the preterit disfavored subject expression and the present showed little effect, in the Colombian data the present slightly disfavored subject expression, but in fact neither the present nor the preterit showed a strong effect.<sup>6</sup> Thus, the NM narrative data follow the pattern predicted by Silva-Corvalán (2001), whereas the Colombian conversational data do not. As Silva-Corvalán's (2001) and Bayley and Pease-Alvarez's (1997) studies are both based on narrative data, this suggests that this tendency may be typical of narrative data, and further testing is required to determine how far it can be generalized beyond that. This is difficult to do for these data because, once again, there were problems with distribution, in that there were many fewer tokens of the imperfect and the preterit than of the present tense.

#### *Distance from previous mention*

This factor group measures the number of clauses since the previous coreferential first-person subject, that is, a first-person singular subject produced by the same speaker.<sup>7</sup> It is here that the effects of coreferentiality and distance between mentions are tested, and in considering the interaction between this factor group and that of the form of previous realization, or priming, we are also able to investigate the duration of priming in the data.

Distance was counted up to ten clauses, with no distinction made beyond this. That is, 11 categories were used, from no intervening clauses to ten (or more) intervening clauses. Preliminary GOLDVARB results revealed natural breaks in the data, with certain degrees of distance patterning similarly, and therefore the 11 categories were collapsed into four groups: (1) no intervening clause (i.e., cases in which the subject is coreferential with that of the preceding clause); (2) one intervening clause; (3) between two and four intervening clauses; and (4) five or more intervening clauses.

Examples (1) through (6) above illustrate coreferential subjects with no intervening clauses, and the examples below give coreferential subjects at greater degrees of distance. Example (7) shows coreferential subjects at a distance of one clause, (8) at a distance of two clauses and (9) at a distance of three clauses.

First-person singular verbs have been double underlined and the verbs in the intervening clauses have been single underlined and numbered.

- (7) y yo tenía cuidado de ellos,  
 (H) ellos se iban<sub>1</sub> a la escuela y yo me quedaba alzando la casa y,  
 ‘And I took care of them, they would go to school and I would stay maintaining the house and, ...’  
 [NMCROSS, 117–1A3: 237–238]
- (8) A: Yo no creo, pues, que eso sea<sub>2</sub> lo mejor.  
 ‘I don’t think that that is the best.’  
 S: ... (2.0) Yo no sé<sub>1</sub> mi amor.  
 ‘I don’t know, my love.’  
 A: No=, no= ... De verdad. yo quiero que hable=mos, negro.  
 ‘No, no. Really. I want us to talk, sweetheart.’  
 [Colombia, restaurant: 1149–1157]
- (9) aprendí a hacer to=do. ... Todo. ... Me enseñaron<sub>3</sub>. Me enseñó<sub>2</sub> ella.  
No me dejaba<sub>1</sub> salir mucho a jugar porque, yo era la mayor,  
 ‘(I) learnt how to do everything. Everything. (They) taught me. (She) taught me.  
 (She) wouldn’t let me go out much to play because I was the oldest.’  
 [NMCROSS, 117–1A3: 180–185]

As these examples illustrate, main and subordinate clauses were included, as were clauses produced both by the same speaker and by the interlocutor. Excluded from the clause count were fixed expressions such as *es que* ‘it’s that’, *será que* ‘could it be that’ and *mira (que)* ‘look’ because they function as discourse markers, and thus have partly lost their verbal status (cf. Company Company, 2006).

The results for distance from previous mention were precisely as expected, that is, the greater the distance from the previous coreferential subject, the greater the probability that the subject will be explicit (cf. Table 2). However, even in cases in which there are no intervening clauses, that is, in contexts of subject continuity, the rate of expressed subjects is still 23% in the NM data and 38% in the Colombian data. Also note that the range for this factor group is only 24 points for the NM data and 16 points for the Colombian data, giving it a magnitude of effect much weaker than that of verb class (with ranges of 35 and 32 in the respective data sets), and making it close to the range of previous realization in each data set (26 and 14, respectively). This indicates that, despite the consistency of the findings in the literature regarding the effect of coreferentiality, it cannot be considered a defining feature of subject expression for these data.

#### *Realization of previous mention*

Each subject was coded for realization of the previous coreferential first-person subject, as either expressed or unexpressed, in order to test for a priming effect. Only identical, fully coreferential subjects were included. Thus, partially coreferential forms (such as first-person plural subjects, or second-person singular subjects produced by another speaker) were not considered, nor were formally

TABLE 4. *Independent variable rule analyses of the contribution of factors to the probability of expressed subjects in NM and Colombian data at different degrees of distance*

	0 clauses		1 clause		2–4 clauses		5+ clauses	
	NM	Col.	NM	Col.	NM	Col.	NM	Col.
N	374	265	141	118	141	157	197	338
Previous realization	✓	✓	✓	✓	✓	×	✓	×
Verb class	✓	✓	✓	×	×	×	✓	✓
TAM	×	×	✓	×	×	✓	×	×
Relationship w/ previous TAM	✓	×	×	×	×	×	×	×
Position in turn	×	×	×	×	✓	✓	×	×
Clause type	×	×	×	×	×	NA	×	×

The constraint hierarchies in both data sets where the factor groups were selected as significant are: for previous realization, verb class, and TAM, as indicated in Table 2; for relationship with previous TAM, a change in TAM favors and maintenance of the same TAM disfavors subject expression; for position in the turn, turn-initial position favors and turn-medial position disfavors subject expression.

identical but noncoreferential mentions (such as first-person singular subjects produced by another speaker, cf. example (8)). Although such partially coreferential forms may enter into the priming, in order to maintain a maximally homogeneous data set they were excluded.

The results demonstrate a clear priming influence (cf. Table 2), in that we are more likely to get an explicit subject in contexts in which the previous subject was also explicit (with weights of .67 in the NM data and .57 in the Colombian data), and are less likely to get an explicit subject in contexts in which it was not (with weights of .41 and .43, respectively). However, note that the effect is much stronger in the NM data, in which the previous realization factor group has the second highest range (26 points), than in the Colombian data, in which this factor group has one of the smallest ranges (14 points, equal to that of TAM). Analysis reveals that this is because of the way in which priming interacts with distance, which we will now consider.

*Priming and distance*

To test the duration of the priming in these data, I conducted independent analyses on factors affecting subject expression at the different categories of distance that were applied in the two data sets. That is, what factors affect subject expression just in those cases in which the subject is coreferential with the subject of the immediately preceding clause; what factors play a role when there is one intervening clause; when there are between two and four intervening clauses; and when there are five or more intervening clauses. Table 4 lists the factor groups selected as significant in the two data sets at each of the degrees of distance.

Although these results reveal much interesting information about the patterning of subject expression, we will just concentrate here on the priming effect. As can be seen, in the NM data, realization of the previous coreferential subject is selected as significant regardless of the distance between the prime and target. In the Colombian data, previous realization only has a statistically significant effect when there are no, or when there is only one, intervening clause. That is, in the Colombian data, but not in the NM data, the priming effect dissipates very rapidly.

This result explains why the NM data present a stronger result for priming overall than the Colombian data (cf. Table 2). In the Colombian corpus, priming is only significant at low degrees of distance, and therefore affects less than half of the data (43%), whereas in the NM corpus, priming affects the entirety of the data. This naturally gives rise to a stronger overall effect for the NM data.

The difference in duration of priming raises the question of whether priming actually functions differently in the two data sets, or whether there is some other explanation for this disparity. This is particularly relevant given current research into the duration of priming effects, and the discussion about the implications of this for our understanding of the way in which priming operates. In order to determine whether indeed the priming does function differently in the two corpora, I conducted variable rule analyses on the two data sets to identify those factors that affect priming, that is, in what contexts a subject tends to repeat the form of the preceding coreferential subject and in what contexts it does not. The results for this are reported in the following section.

## RESULTS FOR PRIMING

In this set of analyses, “repetition of form” was treated as the dependent variable, and the following factor groups were tested to see whether they affected such repetition: realization (expressed or unexpressed); verb class (as applied earlier); TAM (ambiguous or not); relationship with previous TAM (same or different); distance from previous mention (up to 10 clauses);<sup>8</sup> and the presence of an intervening clause produced by an interlocutor. For the Colombian data, I also tested for clause type (main vs. subordinate), but for the NM data it was necessary to restrict the analysis to main clauses only, as clause type interacted with other factors. The results are presented in Table 5.

Note that the overall rate of priming is higher in the NM data, in which 69% of the subjects have the same realization as the preceding coreferential subject, as opposed to 57% for the Colombian data. This difference is, however, not significant ( $p \leq .1$ , chi-square = 3.0888). The results obtained for the two data sets are very similar. In neither data set were verb class,<sup>9</sup> TAM, or an intervening clause by another interlocutor<sup>10</sup> selected as significant, and in the Colombian data, clause type was also not selected. In both data sets, the relationship with the previous TAM and the degree of distance between the prime and target (0–1 clause vs. 2 or more clauses) were found to be significant. The key difference is that in the NM

TABLE 5. *Two independent variable rule analyses of the contribution of factors selected as significant to the probability of priming taking place in NM and Colombian data*

	NMCOS			Colombia		
Total <i>N</i>	712*			878		
% primed subjects	69%			57%		
Corrected mean	.70			.58		
	Weight	%	% of data	Weight	%	% of data
<b>Realization</b>						
unexpressed	.56	75	66	[.50]**	58	52
expressed	.38	57	34	[.50]	57	48
Range	18					
<b>Relationship w. previous TAM</b>						
same TAM	.54	73	63	.56	64	42
different TAM	.43	62	37	.45	52	57
Range	11			11		
<b>Distance</b>						
0–1 clause	.54	73	60	.54	62	43
2+ clauses	.45	62	39	.47	53	56
Range	9			7		

NMCOS: based on main clauses only;  $p = .042$ ; log likelihood =  $-422.447$ ; chi-square/cell =  $0.9914$ . Factor groups not selected as significant: verb class, TAM, presence of intervening turn from other speaker.

Colombia:  $p = .028$ ; log likelihood =  $-589.593$ ; chi-square/cell =  $0.8564$ . Factor groups not selected as significant: realization, verb class, TAM, clause type, presence of intervening turn from other speaker.

\*The NM data is based on 712 verbs for the analysis of priming due to the exclusion of the 141 subordinate clauses.

\*\*Factor groups not selected as significant are given in square brackets.

data only, the realization itself was found to be significant. I will now discuss each of these results individually.

*Realization*

There is an important difference in the two data sets in terms of realization. In the NM data, unexpressed subjects favor priming, whereas expressed subjects do not. That is, unexpressed subjects tend to be followed by more unexpressed subjects, but expressed subjects also tend to be followed by unexpressed subjects. Thus, the priming in these data is attributable to the unexpressed subjects alone, while in the Colombian data it is attributable to all subjects.

Although this appears on the surface to be a major disparity in priming in the two data sets, it can be accounted for by the complex interaction that exists between priming and distance and the way this relates to the two genres being considered

TABLE 6. *Interaction between priming and distance (+: priming and distance function synergistically, -: priming and distance function as opposing tendencies)*

Distance	Priming	
	Preceding expressed	Preceding unexpressed
Greater	+	-
Lesser	-	+

here. In terms of the “distance effect”, we saw earlier that unexpressed subjects are favored at low degrees of distance between coreferential mentions, and expressed subjects are favored at high degrees of distance. In terms of the priming effect, by definition a preceding expressed subject favors a subsequent expressed subject and a preceding unexpressed subject favors an unexpressed subject. This means that in some environments, the priming and distance effects work synergistically, and in others, they work as counteracting tendencies. This is summarized in Table 6. At greater degrees of distance following an expressed subject, the priming and the distance effect work together (represented by a plus sign), favoring another expressed subject. These two effects also converge at lesser degrees of distance following an unexpressed subject, in which case both favor an unexpressed subject. However, at lesser degrees of distance following an expressed subject, these two tendencies work against each other (represented by a minus sign), with the priming effect favoring an expressed subject, and the distance effect favoring an unexpressed subject. The same is so at greater degrees of distance following an unexpressed subject, though in this case the priming effect favors unexpressed subject and the distance effect an expressed subject.

Recall from Table 2 that the NM and Colombian data show a marked difference in distribution according to the degree of separation between coreferential subjects. In particular, notice that 60% of the NM data have a coreferential subject in the immediately preceding clause or in the clause preceding that one, whereas just 43% of the Colombian data fall into this category. A chi-square test reveals this distribution difference to be significant ( $p \leq .025$ , chi-square = 6.4904). Here we can find a likely explanation for why the priming effect is not observed for the expressed subjects in the NM data. At lower degrees of distance, the priming and the distance effect reinforce each other for the unexpressed subjects, but counteract each other for the expressed subjects. As a large proportion of the NM data occurs with zero or one intervening clause, this greatly weakens the effect for expressed subjects in the NM data. For the Colombian data, however, as a smaller proportion of the data fall into this category, the effect is maintained for both expressed and unexpressed subjects.



TABLE 7. *Priming at different degrees of distance according to continuity of TAM*

Distance	Same TAM		Shift in TAM		Total	
	N	%	N	%	N	%
New Mexico						
0-1 clause						
Priming	224	76	93	68	317	74
No priming	70	24	43	32	113	26
Total	294		136			
2+ clauses						
Priming	105	67	70	56	175	62
No priming	51	33	56	44	107	38
Total	156		126			
Total	450		262		712	
Colombia						
0-1 clause						
Priming	132	69	108	56	240	63
No priming	59	31	84	44	143	37
Total	191		192			
2+ clauses						
Priming	109	60	157	50	266	54
No priming	74	40	155	50	229	46
Total	183		312			
Total	374		504		878	

This different distribution can therefore account for why realization is selected as significant in the NM data and not in the Colombian data. Furthermore, it demonstrates that genre affects not only the rate of subject expression, but also the way in which priming manifests itself. However, it does not explain why the effect is more long-lasting in the NM data. This can be explained by the role of continuity of TAM in priming.

*Relationship with previous TAM*

In both data sets, the relationship with the previous TAM has an almost identical effect, as seen in Table 5. Repetition of the TAM favors priming (with weights of .54 in the NM data and .56 in the Colombian data) and a change in TAM disfavors priming (with weights of .43 and .45, respectively). Furthermore, this effect is evident regardless of the distance between the prime and target, as is shown in Table 7 and captured graphically in Figure 1, which present a cross-tabulation of priming with distance and continuity of TAM. Note that in both data sets, the same TAM leads to a higher percentage of primed subjects than does a shift in TAM at the two degrees of distance applied. That is, at a distance of zero or one clause, in the NM data, priming is evident 76% of the time when the TAM is

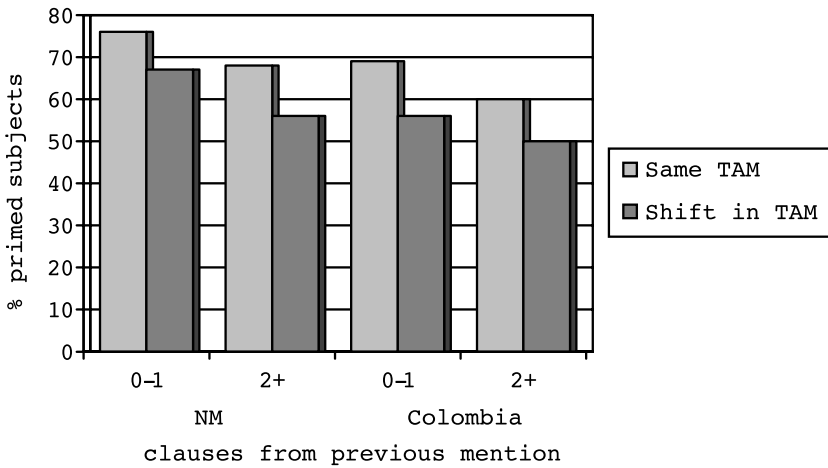


FIGURE 1. Percentage of primed subjects at different degrees of distance according to continuity of TAM.

repeated, a figure that drops to 68% when there is a shift in TAM. The corresponding figures for the Colombian data are 69% and 56%. At a distance of over two clauses, in the NM data, priming is evident 67% of the time when the TAM is repeated and only 56% of the time when there is a shift in TAM. And in the Colombian data, the figures are 60% and 50%.<sup>11</sup>

The result that priming is strengthened by continuous use of the same TAM is particularly interesting because it is contrary to that of Pickering and Branigan (1998) noted earlier, who found that lexical, but not morphological, repetition enhances the priming effect. In order to test whether lexical repetition had an effect I considered cases where the verb type was repeated across coreferential mentions (regardless of whether the same TAM occurred), but this was not found to be statistically significant. The only case in which repetition of verb type was found to be significant was if it was considered together with repetition of TAM. Thus, it appears that repetition of verb type does have an effect, but this is too weak for it to emerge as significant on its own, and it only gains significance when considered together with TAM.

Two major differences need to be borne in mind in comparing Pickering and Branigan's (1998) study and this one. First, theirs was based on a written sentence-completion task, so the nature of the data being studied is very distinct. Notice that Gries (2005) and Szmrecsanyi (2006), who included both spoken and written corpus data, found lexical and morphological enhancement of priming, which may suggest that lexical priming is more evident in written language, though why this should be the case is unclear. Second, in Pickering and Branigan's (1998) study the prime and target were always in adjacent clauses, whereas here coreferential subjects regardless of distance were considered. It may be that at low

degrees of distance, lexical effects play a role, whereas at higher degrees of distance, morphosyntactic effects come into play. As Gries' (2005) and Szmrecsanyi's (2006) data included primes and targets at varying degrees of distance, their findings that morphological repetition enhances priming neither confirms nor refutes this hypothesis.

It is interesting to note that the experimental studies referred to earlier that have found long-term priming (Bock & Griffin, 2000; Boyland & Anderson, 1998; Saffran & Martin, 1997) have been based on picture-description tasks. These tasks make use of different verbs, but there would seem to be little tense variation. The primes, for example, tend to be given in the progressive or in the past tense (cf. Bock & Griffin, 2000:192), and that same tense is likely to be repeated in the target utterances (itself a priming effect). It may be that the nature of the task itself is therefore contributing to the duration of the priming. This notion gains some support from Szmrecsanyi's (2006) finding that priming is more long-lasting in "formal" data. Although Szmrecsanyi did not directly compare the effects of maintaining the same TAM across the different genres, it may be that more formal data show greater maintenance of tense than do less formal (and more interactional) data, and therefore more long-lived priming. This is certainly corroborated for the data under consideration here, where in the NM data, 63% of the subjects shared the same TAM as the preceding coreferential subject, whereas in the Colombian data just 42% did (cf. Table 5), a difference which a chi-square test revealed to be significant ( $p \leq .01$ , chi-square = 8.9419). This maintenance of TAM across clauses is thus playing a major role in the strength of the priming in the narrative data.

Examples (10) and (11) illustrate this patterning for the two corpora. In example (10), which comes from one of the personal narratives from the NM data, the preterit is used consistently up to and including line 10, and then the speaker shifts to the imperfect, which is used in the three verbs in line 11. In example (11), from the conversational Colombian data, considering just the tense of the coreferential first-person singular mentions, we see that Angela (A) switches from the present tense in line 1 to the present perfect in line 8 (double underlining), and Nury (N) from the present in line 3 to the future in line 5 (dotted underlining). Note that these switches correspond to greater distance between mentions, and interaction with the interlocutor, namely, to features typical of spontaneous conversation between friends.

(10)

1. S: (H) *en el sesenta y cuatro, me fui pa'trás, y trabajé por seis meses.*  
'In '64, (I) went<sub>(PRET)</sub> back, and (I) worked<sub>(PRET)</sub> for 6 months.'
2. D: *Mhm.*
3. S: (H) *con la floresta. Luego vine pa' atrás, luego me agarró el draft.*  
'With the forest service. Then (I) came<sub>(PRET)</sub> back, then the draft got<sub>(PRET)</sub> me.'
4. D: *Mhm.*
5. S: *cuando el tiempo de la guerra de Vietnam.*

6. ... *Y puse dos años. . . en active duty.*  
 ‘during the time of the Vietnam War.  
 ... And (I) did<sub>(PRET)</sub> two years. In active duty.’
7. D: *Mhm=.*
8. S: *Los pasé en XX Lousiana. . . Ahí me estuve. ahí hice la . . . basic training.*  
 ‘(I) spent<sub>(PRET)</sub> them in XX Lousiana. There (I) was<sub>(PRET)</sub>.  
 I did<sub>(PRET)</sub> my basic training there.’
9. D: *Hm=.*
10. S: *de ahí me fui pa’ Alaska. Fort Wainright.*
11. *Y ahí era -- ahí tenía un= -- . . . Era bombero no?*  
 ‘from there (I) went<sub>(PRET)</sub> to Alaska. Fort Wainwright.  
 And there (I) was<sub>(IMPF)</sub> -- there (I) had<sub>(IMPF)</sub> a -- (I) was<sub>(IMPF)</sub> a fireman,  
 right?’

[NMCROSS, 102–2B1: 141–161]

(11)

1. A: ... *Una copia, y cuándo se la traigo?*  
 ‘... A copy, and when will (I) bring<sub>(PRES)</sub> it to you?’
2. N: *No=. para que me la dé en disquete. . . No una copia en ho=jas.*  
 ‘No=. So that you give<sub>(PRES SIV)</sub> me a disk. Not a paper copy.’
3. ... *Yo le doy un [disquete].*  
 I will give<sub>(PRES)</sub> you a disk.’
4. A: [Ah=], *bueno. Entonces, deme el disquete.*  
 ‘Oh, OK. So, give<sub>(IMPERATIVE)</sub> me the disk.’
5. N: *Sí=, eso es lo que le voy a dar. [un disquete].*  
 ‘Yes, that is<sub>(PRES)</sub> what (I) am going to give<sub>(FUT)</sub> you. A disk.’
6. A: [Que no tenga] *ningún virusito.*  
 ‘May (it) not have<sub>(PRES SIV)</sub> any little virus.’
7. N: *No=, está nuevo.*  
 ‘No, (it) is<sub>(PRES)</sub> new.’
8. A: .. *No le he metido ni un disquete, todavía a ese computador.*  
 ‘(I) haven’t<sub>(PRES PERFECT)</sub> put a single disk, in that computer yet.’

[Colombia, estudios: 59–75]

In sum, because continuity of TAM enhances the priming effect, the greater continuity of TAM in the NM narrative data leads to a more long-lasting priming effect, and the greater shifting of TAM in the Colombian conversational data leads to a more short-term effect. That is, the duration of the priming is an epiphenomenon of the genre being considered.

### Distance

To test the effect of distance on priming, I broke the data into two categories: those with no clauses or one clause intervening between the prime and target, and those with two or more clauses intervening. I used these two categories because the data revealed a natural break at this point—when comparing the different degrees of distance individually, priming was favored only at zero and one clause of separation for both data sets, whereas all other degrees of distance achieved a weight below .50. Thus, although the form of the previous realization has a significant effect on subject expression at high degrees of distance for the NM data,

in comparison to the lower degrees of distance these higher degrees show a disfavoring of priming. As Table 5 illustrates, priming is favored at the lower degrees of distance (with a weight of .54 in both the NM and Colombian data), and disfavored at higher degrees of distance (with weights of .45 and .47, respectively). This is evidence that priming effects do indeed dissipate over time, and, importantly, that this occurs in the same way in both data sets. Thus, although Table 4 suggests that priming interacts differently with distance in the two data sets, this is in fact not the case, and distance has the same effect in both data sets.

Once again, the different distribution is important here. Whereas 60% of the NM subjects have a coreferential subject at a distance of zero or one clause, just 43% of the Colombian subjects occur in this environment. A chi-square test of this difference finds it to be significant ( $p \leq .025$ , chi-square = 5.8479). The NM data exhibit a stronger effect, not because the priming is in fact stronger, but because more of the data falls into this low-distance group.

### *Summary*

We have seen that priming functions very similarly in the two corpora, in that it is favored and disfavored to comparable degrees in the same environments. The fact that the NM and Colombian data differ in terms of the degree of separation between the prime and target at which priming has a significant effect on subject expression can be attributed to the greater continuity of TAM in the NM as opposed to the Colombian data. This disparity in continuity can readily be accounted for in terms of the different genres being studied. In interactional conversation, speakers must constantly attend to their interlocutor(s), which minimally involves responding to contributions they make. This creates greater distance between coreferential subjects, and leads to frequent shifts in subject as well as in TAM. In interview data, on the other hand, contributions from the interviewer are kept to a minimum and primarily consist of comments that encourage the speaker to go on with their stream of speech. This naturally allows for greater maintenance of the same subject and TAM, as a speaker moves through a story of past events. The contrasting results obtained for the strength of the priming at the different degrees of distance for the NM and Colombian data are therefore entirely attributable to the different genres being studied.

### RATES OF SUBJECT EXPRESSION IN NM AND COLOMBIAN SPANISH

So far we have seen that subject expression patterns very similarly in the NM and Colombian data, and we have not found any explanation for why there should be such contrasting rates of expression in the two corpora, in which subjects are expressed just one third of the time in the NM data but close to half the time in the Colombian data. A very simple explanation can be found if we reexamine the distribution of the data with a finer breakdown for the different degrees of distance. This is presented in Table 8.

TABLE 8. *Distribution of the data according to distance*

Distance	% of data	
	NM	Colombia
0 clauses	44	30
1 clause	16	13
2–3 clauses	13	17
4–9 clauses	14	13
10 <sup>+</sup> clauses	12	27

Table 8 shows a significant disparity in the distribution for the two data sets ( $p \leq .05$ , chi-square = 9.5216). This is particularly apparent at the extremes, with close to one half of the NM data compared to less than one third of the Colombian data (44% vs. 30%) occurring with a coreferential subject in the immediately preceding clause, and just over one tenth of the NM data and over one quarter of the Colombian data occurring with a coreferential subject at a distance of over 10 clauses (12% vs. 27%). Given the correlation observed in Table 2 between greater distance and explicit subjects, the lower degree of subject continuity in the Colombian data explains the higher rate of subject expression, and the greater continuity of reference in the NM data explains the lower overall rate of subject expression. That is, the rate of subject expression is due to the degree of subject continuity. To fully understand this, then, we need to account for the divergence in subject continuity.

Once again we must recognize the possibility that this is a dialectal difference. However, there is no evidence to suggest that this is the case, and furthermore, there is no reason why one dialect should have a higher level of continuity than another. This would in fact suggest a different speech style, something that is directly related to genre. The narratives of the NM data involve the interviewee's life story, a context that favors both continuity of reference and first-person subjects. This can be seen in example (10) presented earlier, in which the first-person subject is maintained across the 12 clauses presented, with just one intervening subject (in line 3). The interactive conversation making up the Colombian data, on the other hand, has more shifting of topics, and therefore less continuity of subjects. This is illustrated in example (11). Here, the double underlining represents Angela's (A's) first-person singular verbs (in lines 1 and 8, separated by seven clauses), the dotted underlining Nury's (N's) first-person singular verbs (in lines 3 and 5, separated by two clauses), and single underlining other verbs. Notice that as speakers ask and respond to questions and negotiate with each other the subject is constantly shifting, and this creates greater distance between coreferential mentions.

To reiterate, then, the low rate of subject expression in the NM data is attributable to the narrative genre, which demonstrates a great deal of subject continuity. The high rate of subject expression in the Colombian data is attributable to the more interactive nature of these data, and the resulting greater shifting of

subject across clauses. The divergent rates of expression are therefore not representative of underlying grammatical differences. That is, they are not accountable in terms of dialectal differences, but in terms of genre.

## CONCLUSIONS

This article has shown that variable subject expression in both NM and Colombian Spanish is conditioned by a priming effect, whereby a pronominal mention favors a subsequent pronominal mention, and an unexpressed mention favors a subsequent unexpressed mention to a statistically significant degree. Although the overall results suggest that the priming effect is more long-lasting in the NM data than in the Colombian data, closer analysis reveals that this is an epiphenomenon of the greater continuity of TAM in the NM narratives. There are three major conclusions that can be drawn from this.

The first relates to the duration of structural priming. It was noted above that short-term priming can be accounted for in terms of activation, whereas long-term priming is better accounted for in terms of procedural learning (Bock & Griffin, 2000; Chang et al., 2000). Whereas Bock and colleagues have found long-term priming effects in experimental settings, we have seen here that priming can also have long-term effects in natural discourse, as long as a certain degree of continuity is maintained. This kind of continuity is evident in monologic, particularly narrative, data, but not in more interactional data. It is also evident in the kind of language that is obtained in experimental settings, and this continuity may be facilitating the long-term effects observed in the psycholinguistic research reported on here, as well as the long-term effects found by Szmrecsanyi (2006) in more formal registers. The results of this study indicate that more experimental research is required to better understand the life-time of structural priming, in terms of what may affect its maintenance or dissipation.

The second point relates to the genre differences observed. We have seen that the effect of genre, and in particular, interaction, on patterns of language use, should not be underestimated. The conversational data exhibit a higher rate of subject expression and more short-term priming because the interactional discourse gives rise to regular shifts in subject and TAM. The continuity of subject and TAM in the monologic narrative data results in a lower rate of subject expression and more long-term priming. It is often taken for granted that linguistic conditioning found in any one data set holds for language use in general, but this study has shown that even spontaneous spoken discourse is highly heterogeneous. Here, the different rates of subject expression in the two data sets are due to the different distribution of conditioning factors, whereas the effect of those factors is parallel. To fully understand subject expression, priming, and language use in general, a variety of genres needs to be analyzed (cf. Biber, 2001). Although the study presented here extends previous research on subject expression by taking into account conversational and narrative data, this is just a small subset of genres that would need to be investigated to obtain a more complete understanding of the patterning of subject expression, and the mechanism of priming.

And finally, the significance of the finding that priming has an effect on language use across genres of spontaneous discourse must be highlighted. As noted in the introduction to this article, this has profound implications for our view of grammar, as it indicates that the grammar of discourse is developed on-line, as a response to and deriving from what precedes. We know from the literature on grammaticization (e.g., Bybee et al., 1994) that grammars are shaped over time by the conventionalization of repeated patterns of use. The priming effects outlined here demonstrate that, in accordance with Hopper's notion of emergent grammar (1998), this happens not only diachronically but synchronically, in real time as discourse is constructed.

## NOTES

1. Cameron did not observe a priming effect for full Noun Phrases (1994:38).
2. This information gives the corpus, the name of the transcript from which the example is drawn, and the line numbers of the excerpt.
3. The priming effect for the passive vs. active clauses, however, was found to be short-lived (Bock & Griffin, 2000:186).
4. Cf. Cameron (1992; 1994), who included such constructions on these grounds.
5. This is probably not unique to the first person. Biber (2001:107) found *say* to be among the four most frequent lexical verbs in conversational English, and the most frequent lexical verb in written English of a variety of genres including fiction, news, and academic prose.
6. For the NM data, the following results were obtained: imperfect etc. .59, present .50, preterit .43; and for the Colombian data: imperfect etc. .63, preterit .52, present .47.
7. I also tested for the number of different human subjects intervening between the prime and target, which was found to have a significant effect—as is to be expected, the greater the number of different intervening subjects, the greater the likelihood that subjects will be expressed.
8. As with the study of subject expression, here also I conducted an independent analysis to determine whether the number of different human subjects intervening had an effect, but this was not selected as significant.
9. It has been pointed out to me that certain set expressions such as *yo creo* and *yo digo* may function much like discourse markers and therefore be immune to the priming. Although I have not tested for these specific expressions, in an independent analysis of subject expression with cognitive verbs in general, priming maintained a significant role. This may be evidence that such constructions, although fixed to some degree, are not wholly grammaticized as yet.
10. In an independent analysis, I tested for position in the turn (initial or medial) instead of presence of an intervening turn by another interlocutor, but this was not found to be significant for either data set.
11. One anonymous reviewer suggested that this may be due to an interaction between expressed and unexpressed subjects and repetition of TAM. A cross-tabulation revealed that there is no significant difference in subject expression according to TAM repetition, demonstrating that these two factor groups are independent.

## REFERENCES

- Ashby, William J., & Bentivoglio, Paola. (1993). Preferred argument structure in spoken French and Spanish. *Language Variation and Change* 5:61–76.
- Ávila-Shah, Bárbara I. (2000). Discourse connectedness in Caribbean Spanish. In A. Roca (ed.), *Research on Spanish in the United States*. Somerville, MA: Cascadilla Press. 238–251.
- Bayley, Robert. (2002). The quantitative paradigm. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (eds.), *The handbook of language variation and change*. Oxford: Blackwell. 117–141.
- Bayley, Robert, & Pease-Alvarez, Lucinda. (1997). Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change* 9:349–371.
- Bentivoglio, Paola. (1987). *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela.
- Biber, Douglas. (1986). Spoken and written textual dimensions in English: Resolving the contradictory findings. *Language* 62:384–416.



- (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- (1992). Using computer-based text corpora to analyze the referential strategies of spoken and written texts. In J. Svartvik (ed.), *Directions in corpus linguistics: Proceedings of Nobel Symposium 82, Stockholm 1991*. Berlin / New York: Mouton de Gruyter. 214–252.
- (2001). Using corpus-based methods to investigate grammar and use: Some case studies on the use of verbs in English. In R. C. Simpson & J. M. Swales (eds.), *Corpus linguistics in North America*. Ann Arbor: University of Michigan Press. 101–115.
- Bills, Garland D. (1997). New Mexican Spanish: Demise of the earliest European variety in the United States. *American Speech* 72:154–171.
- Bills, Garland D., & Vigil, Neddy A. (1999). Ashes to ashes: The historical basis for dialect variation in New Mexican Spanish. *Romance Philology* 53:43–66.
- (2000). The continuity of change: Nahuatlismos in New Mexican Spanish. In A. Roca (ed.), *Research on Spanish in the United States: Linguistic issues and challenges*. Somerville, MA: Cascadilla Press. 137–153.
- Bock, J. Kathryn. (1986). Syntactic persistence in language production. *Cognitive Psychology* 18:355–387.
- Bock, J. Kathryn, & Griffin, Zeni M. (2000). The persistence of structural priming: Transient activation or implicit learning. *Journal of Experimental Psychology: General* 129:177–192.
- Boyland, Joyce Tang, & Anderson, John R. (1998). Evidence that syntactic priming is long-lasting. *Proceedings of the Annual Conference of the Cognitive Science Society* 20:1205.
- Branigan, Holly P., Pickering, Martin J., & Cleland, Alexandra A. (2000a). Syntactic co-ordination in dialogue. *Cognition* 75:B13–25.
- Branigan, Holly P., Pickering, Martin J., Liversedge, Simon P., Stewart, Andrew J., & Urbach, Thomas P. (1995). Syntactic priming: Investigating the mental representation of language. *Journal of Psycholinguistic Research* 24:489–506.
- Branigan, Holly P., Pickering, Martin J., Stewart, Andrew J., & McLean, Janet F. (2000b). Syntactic priming in spoken production: Linguistic and temporal interference. *Memory and Cognition* 28:1297–1302.
- Bybee, Joan, Perkins, Revere, & Pagliuca, William. (1994). *The evolution of grammar: Tense, aspect and modality in the languages of the world*. Chicago: University of Chicago Press.
- Cameron, Richard. (1992). *Pronominal and null subject variation in Spanish: Constraints, dialects, and functional compensation*. Doctoral dissertation, University of Pennsylvania, Philadelphia, PA.
- (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.
- (1994). Switch reference, verb class and priming in a variable syntax. *Papers from the Regional Meeting of the Chicago Linguistic Society: Parasession on variation in linguistic theory* 30:27–45.
- (1995). The scope and limits of switch reference as a constraint on pronominal subject expression. *Hispanic Linguistics* 6–7:1–27.
- Cameron, Richard, & Flores-Ferrán, Nydia. (2003). Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context* 1:41–65.
- Chafe, Wallace. (1994). *Discourse, consciousness and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
- Chang, Franklin, Dell, Gary S., Bock, J. Kathryn, & Griffin, Zeni M. (2000). Structural priming as implicit learning: A comparison of models of sentence production. *Journal of Psycholinguistic Research* 29:217–229.
- Company Company, Concepción. (2006). Subjectification of verbs into discourse markers: Semantic-pragmatic change only? In B. Cornillie & N. Delbecq (eds.), *Topics in subjectification and modalization*. Amsterdam: Benjamins. 97–121.
- Davidson, Brad. (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.
- Du Bois, John W., Schuetze-Coburn, Stephan, Cumming, Susanna, & Paolino, Danae. (1993). Outline of discourse transcription. In J. Edwards & M. Lampert (eds.), *Talking data: Transcription and coding in discourse*. Hillsdale, NJ: Erlbaum. 45–89.
- Enríquez, Emilia V. (1984). *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: Consejo Superior de Investigaciones Científicas, Instituto Miguel de Cervantes.
- Estival, Dominique. (1985). Syntactic priming of the passive in English. *Text* 5:7–21.
- Flores-Ferrán, Nydia. (2002). *Subject personal pronouns in Spanish narratives of Puerto Ricans in New York City: A sociolinguistic perspective*. Munich: Lincom Europa.
- (2004). Spanish subject personal pronoun use in New York City Puerto Ricans: Can we rest the case of English contact? *Language Variation and Change* 16:49–73.

- Gries, Stefan Th. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research* 34:365–399.
- (to appear). New perspectives on old alternations. In J. E. Cihlar, A. L. Franklin, D. W. Kaiser, & I. Kimbara (eds.), *Papers from the 39th Regional Meeting of the Chicago Linguistics Society, Vol. 2*. Chicago IL: Chicago Linguistic Society.
- Hartsuiker, Robert J., Pickering, Martin J., & Veltkamp, Eline. (2004). Is syntax separate or shared between languages. *Psychological Science* 15:409–414.
- Hochberg, Judith G. (1986). Functional compensation for /s/ deletion in Puerto Rican Spanish. *Language* 62:609–621.
- Hopper, Paul J. (1987). Emergent grammar. *Proceedings of the Berkeley Linguistics Society* 13:139–157.
- (1998). Emergent grammar. In M. Tomasello (ed.), *The new psychology of language, vol. 2: Cognitive and functional approaches to language structure*. Mahwah, NJ: Erlbaum. 155–175.
- Jia, Li, & Bayley, Robert. (2002). Null pronoun variation in Mandarin Chinese. *University of Pennsylvania Working Papers in Linguistics* 8(3):103–116.
- Joordens, Steve, & Besner, Derek. (1992). Priming effects that span an intervening unrelated word: Implications for models of memory representation and retrieval. *Journal of Experimental Psychology* 18(3):483–491.
- Lipski, John M. (1994). *Latin American Spanish*. London: Longman.
- Loebell, Helga, & Bock, Kathryn. (2003). Structural priming across languages. *Linguistics* 41(5): 791–824.
- Meyer, David E., & Schvaneveldt, Roger W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology* 90(2): 227–234.
- Meyer, David E., Schvaneveldt, Roger W., & Ruddy, Margaret G. (1972). Activation of lexical memory. Paper presented at the 13th Annual Meeting of the Psychonomic Society, St. Louis, MO.
- Morales, Amparo. (1986). Algunos aspectos de gramática en contacto: La expresión del sujeto en el español de Puerto Rico. *Anuario de Letras* 24:71–85.
- Ono, Tsuyoshi, & Thompson, Sandra A. (1995). What can conversation tell us about syntax? In P. W. Davis (ed.), *Alternative linguistics: Descriptive and theoretical modes*. Amsterdam: Benjamins. 213–271.
- Otheguy, Ricardo, & Zentella, Ana Cecilia. (in press). Apuntes preliminares sobre el contacto lingüístico y dialectal en el uso pronominal del español en Nueva York: Variación, cambio e identidad en el uso variable del pronombre en seis comunidades hispanohablantes de la Gran Manzana. In R. Cameron & K. Potowski (eds.), *Spanish in contact: Policy, social and linguistic inquiries*. Amsterdam / Philadelphia: John Benjamins. 239–257.
- Otheguy, Ricardo, Zentella, Ana Cecilia, Erker, Daniel, & Livert, David. (2005). Factores gramaticales y sociodemográficos en la evolución y continuidad de los pronombres sujetos del español de los latinoamericanos en Nueva York. Paper presented at the XIV Congreso de la Asociación de Lingüística y Filología de América Latina (ALFAL), Monterrey, México.
- Paredes Silva, Vera Lucia. (1993). Subject omission and functional compensation: Evidence from written Brazilian Portuguese. *Language Variation and Change* 5:35–49.
- Pickering, Martin J., & Branigan, Holly P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language* 39:633–651.
- (1999). Syntactic priming in language production. *Trends in Cognitive Sciences* 3:136–141.
- Pickering, Martin J., Branigan, Holly P., Cleland, Alexandra A., & Stewart, Andrew J. (2000). Activation of syntactic information during language production. *Journal of Psycholinguistic Research* 29:205–216.
- Poplack, Shana. (1980). The notion of the plural in Puerto Rican Spanish: Competing constraints on (s) deletion. In W. Labov (ed.), *Locating language in time and space*. New York: Academic Press. 55–67.
- Poplack, Shana, & Tagliamonte, Sali. (2001). *African American English in the diaspora*. Malden, MA: Blackwell.
- Rand, David, & Sankoff, David. (1990). GoldVarb: A variable rule application for the Macintosh (2.0 ed.). Montreal: Centre de recherche mathématiques, Université de Montréal.
- Ranson, Diana L. (1991). Person marking in the wake of /s/ deletion in Andalusian Spanish. *Language Variation and Change* 3:133–152.
- Saffran, Eleanor M., & Martin, Nadine. (1997). Effects of structural priming on sentence production in aphasics. *Language and Cognitive Processes* 12:877–882.
- Scheibman, Joanne. (2001). Local patterns of subjectivity in person and verb type in American English conversation. In J. Bybee & P. J. Hopper (eds.), *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins. 61–89.

- Scherre, Maria Marta Pereira. (2001). Phrase level parallelism effect on Noun Phrase number agreement. *Language Variation and Change* 13:91–107.
- Scherre, Maria Marta Pereira, & Naro, Anthony J. (1991). Marking in discourse: 'Birds of a feather'. *Language Variation and Change* 3:23–32.
- \_\_\_\_\_. (1992). The serial effect on internal and external variables. *Language Variation and Change* 4:1–13.
- Silva-Corvalán, Carmen. (1982). Subject expression and placement in Mexican-American Spanish. In J. Amastae & L. Elías Olivares (eds.), *Spanish in the United States: Sociolinguistic aspects*. New York: Cambridge University Press. 93–120.
- \_\_\_\_\_. (1994). *Language contact and change: Spanish in Los Angeles*. Oxford: Clarendon.
- \_\_\_\_\_. (1997). Variación sintáctica en el discurso oral: Problemas metodológicos. In F. Moreno Fernández (ed.), *Trabajos de sociolingüística hispánica*. Alcalá de Henares, España: Universidad de Alcalá. 115–135.
- \_\_\_\_\_. (2001). *Sociolingüística y pragmática del español*. Washington DC: Georgetown University Press.
- Solomon, Julie. (1999). *Phonological and syntactic variation in the Spanish of Valladolid, Yucatán*. Doctoral dissertation, Stanford University, Stanford, CA.
- Szmrecsanyi, Benedikt. (2005). Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 1:113–149.
- \_\_\_\_\_. (2006). *Morphosyntactic persistence in spoken English: A corpus study at the intersection of variationist sociolinguistics*. Berlin/New York: Mouton de Gruyter.
- Tannen, Deborah. (1987). Repetition in conversation: Toward a poetics of talk. *Language* 63(3): 574–605.
- Thompson, Sandra A. (2002). 'Object Complements' and conversation: Towards a realistic account. *Studies in Language* 26:125–163.
- Toribio, Almeida Jacqueline. (1996). Dialectal variation in the licensing of null referential and expletive subjects. In C. Parodi, C. Quicoli, M. Saltarelli, & M. L. Zubizarreta (eds.), *Aspects of Romance linguistics: Selected papers from the 24th Linguistic Symposium on Romance Languages, 1994*. Washington, DC: Georgetown University Press. 409–432.
- Travis, Catherine E. (2005a). *Discourse markers in Colombian Spanish: A study in polysemy*. Berlin/New York: Mouton de Gruyter.
- \_\_\_\_\_. (2005b). The yo-yo effect: Priming in subject expression in Colombian Spanish. In R. Gess & E. J. Rubin (eds.), *Selected papers from the 34th Linguistic Symposium on Romance Languages (LSRL), Salt Lake City, 2004*. Amsterdam: Benjamins. 329–349.
- \_\_\_\_\_. (2006). Subjetivización de construcciones: Los verbos 'cognitivos' en el español conversacional. In Rosa María Ortiz Ciscomani (ed.), *Serie Memorias del VIII Encuentro Internacional de Lingüística en el Noroeste*, vol. 2. Hermosillo, Sonora, Mexico: UniSon. 85–109.
- Vigil, Neddy A., & Bills, Garland D. (1997). A methodology for rapid geographical mapping of dialect features. In A. R. Thomas (ed.), *Issues and methods in dialectology*. Bangor: University of Wales. 247–255.
- \_\_\_\_\_. (2000). El español de Nuevo México: hablamos mexicano. *Encuentro Internacional de Lingüística en el Noroeste* 5:197–217.
- \_\_\_\_\_. (2004). Dialect shift in New Mexican Spanish: A turkey by any other name. *Romance Philology* 57:323–343.
- Weiner, E. Judith, & Labov, William. (1983). Constraints on the agentless passive. *Journal of Linguistics* 19:29–58.

## APPENDIX

## TRANSCRIPTION CONVENTIONS (DU BOIS ET AL., 1993)

LETTER:	speaker label	=	lengthened syllable
.	final intonation contour	..	short pause (about 0.5 sec)
,	continuing intonation contour	...	medium pause (> 0.7 sec)
?	appeal intonation contour	...(N)	long pause (of N sec)
--	truncated intonation contour	[ ]	speech uttered in overlap
(H)	in breath		