10-17-2013

Lexical Frequency and Syntactic Variation: A Test of a Linguistic Hypothesis

Robert Bayley
University of California, Davis, rjbayley@ucdavis.edu

Kristen Greer
University of California, Davis, kaware@ucdavis.edu

Cory Holland
University of California, Davis, clmessing@ucdavis.edu

Lexical Frequency and Syntactic Variation: A Test of a Linguistic Hypothesis

Abstract

The role of lexical frequency in language variation and change has received considerable attention in recent years. Recently Erker and Guy (2012) extended the analysis of frequency effects to morphosyntactic variation. Based on data from 12 Dominican and Mexican speakers from Otheguy and Zentella's (2012) New York City Spanish corpus, they examined the role of frequency in variation between null and overt subject personal pronouns (SPP). Their results suggest that frequency either activates or amplifies the effects of other constraints such as co-reference. This paper attempts to replicate Erker and Guy's study with a data set of Mexican immigrant and Mexican American Spanish. Analysis of more than 8,600 tokens shows that frequency has only a small effect on SPP use. In separate analyses of frequent and non-frequent verb forms, fewer constraints reach significance with frequent verb forms only than with non-frequent forms only. Moreover, in cases where constraints reach significance in both analyses, effects are stronger with non-frequent than with frequent forms. Finally, when all verb forms are combined in a single analysis, non-frequent forms are significantly more likely than frequent forms to co-occur with overt SPPs. We conclude that claims about frequency effects in SPP variation should be treated with caution and that further analyses are needed to establish whether models incorporating frequency can be extended to this area of the grammar.

Lexical Frequency and Syntactic Variation: A Test of a Linguistic Hypothesis

Robert Bayley, Kristen Greer, and Cory Holland*

1 Introduction

In recent years, lexical frequency has become an increasingly prominent explanation for a range of linguistic phenomena, particularly in studies of language variation and change. Bybee (2000, 2001, 2002, 2010), for example, proposed a usage-based exemplar model of language in which grammar emerges through language use, and lexical frequency drives variation and change. Bybee's exemplar model has several key components. First is the notion that linguistic experience affects linguistic structure. In this view, the lexicon is not a mere list of lexical items, but rather a highly interconnected network of lexical exemplars. Each exemplar corresponds to a lexical unit (a word or phrase) that is connected in a network to other exemplars. Connections between exemplars are strong when there is a high degree of phonological and semantic similarity, and less so when there is less similarity. Moreover, individual exemplars themselves can vary in lexical strength. Bybee argues that exposure to linguistic tokens strengthens their mental representation. In this regard, token frequency directly bears on emergent linguistic structure. Specifically, Bybee contends that highly frequent words have stronger mental representations that are in turn more easily accessed during production. As a corollary, lower frequency words are more difficult to access and can even become so weak that they fade and are eventually forgotten.

In the exemplar model, lexical token frequency plays a pivotal role in language change. Bybee (2001) argues that two mechanisms of language change are based on token frequency: phonological reduction in high frequency words and analogical leveling in low frequency words. Reductive sound change results from the automation of linguistic production (Bybee 2002). In speech, repeated articulatory patterns become more efficient through language use. That is, the magnitude of extreme gestures is diminished and articulatory transitions are smoothed and overlapped. In Bybee's argument, it follows that highly frequent words, because they are used more often, have more opportunities for reduction. Every reduced token contributes to the strength of its lexical representation. This in turn contributes to the shifting of the exemplar cluster (i.e., a word and all its variants). Since the lexicon is a tightly interconnected network of exemplars and exemplar clusters, this reductive change slowly diffuses through the lexicon eventually affecting even low frequency words. This, Bybee (2001) contends, is why "phonetic processes as they first appear in a language tend to affect high-frequency and highly automated sequences and only later extend to the whole lexicon of words and phrases" (66).

Although high frequency is the catalyst for phonological change, it also serves a conserving function in regards to analogical change. Bybee (2001) argues that morphophonemic change is based on comparison of forms. Form similarity in the lexicon functions as a schema or pattern of production that gets applied when a specific exemplar is difficult to access. This analogical process tends to regularize low frequency forms by applying a more productive pattern on the word rather than accessing the lexically weak irregular form. On the other hand, high frequency protects words from the same effects of regularization. Because highly frequent words have stronger lexical representations that are easier to access during on-line production they are less prone to analogical leveling. This dual process, Bybee (2002) claims, is why low-frequency verbs such as weep/wept, leap/leapt, and creep/crept are regularizing to the past tense to form -ed. This is also why the equivalent high-frequency verbs keep/kept, sleep/slept, and leave/left show no such analogical leveling. In more general terms, this effect helps explain why morphological irregularity tends to co-occur with more frequent words.

Recently, Erker and Guy (2012), noting the relative absence of studies of frequency effects above the level of phonology, examined the role of frequency in morphosyntactic variation. Based on data from six Dominicans and six Mexicans from the Otheguy and Zentella (2012) corpus of New York City Spanish, they sought to determine whether the alternation between overt and null pronouns that is found in all Spanish dialects is "an aspect of language [that] can profitably be re-examined in light of important frequency effects" (Bybee, 2002:220).

Erker and Guy's (2012) results suggest that frequency does not have an independent effect. Rather, it activates or amplifies other constraints that have been identified in the literature. Their results suggest that frequency, defined as forms consisting of more than one percent of the verb forms in their corpus, activates the following constraints: morphological regularity, semantic content, and person/number. In their data, these constraints only had a significant effect in the analysis of frequent verb forms. Their results also suggest that tense/mood/aspect and switch reference effects are amplified with frequent verbs, although they also could be seen among the non-frequent forms.

In this article, we attempt to replicate Erker and Guy's results regarding the role of frequency in subject personal pronoun (SPP) variation, using a larger data set representing speakers of a single national background: Mexican immigrants and Mexican Americans in south Texas and northern California. The remainder of the article is organized as follows. First, we briefly review research on Spanish SPP variation and describe the data and the methods of the current study. Next, we present the results of multivariate analysis. Finally we discuss the results in relationship to claims about frequency. Our analysis suggests that Erker and Guy's results for frequency cannot be replicated and therefore should be treated with caution.

2 The Study of Subject Personal Pronouns

In Spanish, a subject may be expressed overtly or as null, as illustrated in (1) and (2), taken from the data for the present study:

(1) Yo/Ø le digo, 'Háblales en español.'
    'I tell him, 'Speak to them in Spanish.'')
(2) Sí nosotros/Ø hemos placticado de eso….
    'Yes, we've spoken about this….'

In recent decades, this alternation has received considerable attention in sociolinguistics. Studies of dialects in many areas, including northern and southern California, Madrid, New Jersey, New York, New Mexico, Colombia, Puerto Rico, and Andalusia, have shown that SPP alternation is subject to multiple constraints (see e.g., Bayley and Pease-Alvarez 1997, Cameron 1993, 1996, Cameron and Flores-Ferrán 2004, Flores-Ferrán 2004, 2007a, 2007b, Otheguy and Zentella 2012, Ranson 1991, Silva-Corvalán 1994, 1996-97, Travis 2007). Indeed, considering the number of studies that have been accomplished, SPP variation, like -t,d deletion in English, has become a showcase variable in quantitative sociolinguistics.

In general, the linguistic constraints on Spanish SPP variation are well established and many extend across a wide range of dialects. In fact, Erker and Guy (2012) used SPP variation to test their hypotheses because it is generally well understood. For example, numerous studies have reported that subjects that are co-referential with the subject of the preceding tensed verb are less likely to be realized overtly than subjects that are not co-referential. In addition, singular SPPs, particularly 1 sg *yo*, are more likely to be realized overtly than plurals (e.g., Cameron 1993, Cameron and Flores-Ferrán 2004, Flores-Ferrán 2004, Silva-Corvalán 1994). Other constraints that have been investigated include discourse connectedness, or the agreement of a tensed verb with the preceding verb in person, number, tense, and mood (Bayley and Pease-Alvarez 1997), and morphological ambiguity of the tensed verb of which the pronoun is a subject, e.g. *estaba*, which may be translated as 'I was' or 's/he was' and, in dialects where /s/ is frequently deleted, as in Caribbean Spanish, 'you (sg) were' (Cameron 1993, Hochberg 1986).

3 Methods

3.1 Speakers

The data for this study come from an extensive series of interviews with Mexican immigrant and Mexican American parents conducted by the first author and Sandra Schecter for an ethnographic study of language maintenance and home language use in Mexican-descent families in the San Francisco Bay area, California and San Antonio, Texas (Schecter and Bayley 2002). The interviews covered a range of topics, including participants' experiences as children, the age at which they or their parents immigrated, their own and their children's language use, their ideas about schooling and the place of Spanish and English in their own and their children's lives.

Participants in the larger study represented a variety of social backgrounds and immigrant generations. However, all who chose to conduct their interviews in Spanish were either immigrants or had parents who had immigrated. Of the 29 speakers who were interviewed in Spanish, 14 resided in Texas and 15 in California, and 19 were women and 10 were men. All except two, both born in Texas, were born in Mexico. Eleven speakers claimed to be bilingual in Spanish and English and 18 claimed to speak Spanish only or to be strongly Spanish dominant. Speakers varied greatly in the extent to which they used overt SPPs, ranging from a low of 18.2 percent for one northern California man to a high of 50.2 percent for a northern California woman.

3.2 Transcription and Coding

All interviews were fully transcribed in standard orthography. The interviews analyzed here yielded 8,676 tokens of possible SPP use that were coded for a range of internal and external factors that previous research has shown to influence a speaker's choice between an overt and a null pronoun. Internal factors included (a) co-reference with the preceding tensed verb, (b) tense, mood, and aspect, (c) person and number, (d) ambiguity, (e) verb semantics, and (f) lexical aspect. External factors included (a) geographic region, (b) gender, (c) bilingualism, and (d) individual speaker.

3.2.1 Co-reference

Co-reference with the subject of the preceding tensed verb is widely agreed to be one of the most important factors conditioning SPP variation. When coding for this variable, we considered not only straightforward cases of switched (as in (3) below) v. non-switched reference but also cases of partial overlap, in which the subject of the preceding tensed verb is either a super- or a subset of the subject of the coded verb (shown in (4)), and cases of resumed reference, wherein the subject is co-referential with the subject of a tensed verb two or three clauses before it.

(3) Él va a la escuela, pos pa' … donde viven ellos, no Ø **sé** ni cómo se llama la escuela….
     'He goes to school … where they live, no I don't know the name of the school….'
(4) Él se vino primero, luego me vine yo con él, … y luego después ya Ø **fuimos** … por mi mamá y los- mis hermanos.
     'He came first, then I came with him … and then afterwards we went for my mother and my brothers.'

3.2.2 Tense, mood, and aspect

In coding for tense, mood, and aspect, we distinguished present forms from two other groupings. In the first, we combined preterit forms and all forms of the verb *ser*, and in the second, we combined subjunctive, imperfect, and conditional verb forms.

3.2.3 Person and number

As is common in studies of SPP variation, we also coded for person and number, and, based on Cameron (1992, 1996), we distinguished between specific and non-specific 2 sg *tú*. Specific *tú*,

where the referent is explicitly identifiable in the discourse, occurred in indirect or direct speech, when the speaker would address the interviewer (although *usted* was used more often in these rare cases), or when the speaker would address his/her spouse (several of the interviews from our corpus were pair interviews), as in (5):

(5) Speaker (husband): ¿Ø **hablas** español con Alex?
     'Do you speak Spanish with Alex?'
     Speaker (wife): A veces.
     'Sometimes.'
     Speaker (husband): A veces sí Ø **hablas**.
     'Sometimes yes, you speak Spanish.'

All other cases of *tú* were generally deemed non-specific. In these cases, the context was discernibly general and hence the generic reference of the pronoun was easily inferred. Linguistic indicators of this generality included (a) a lack of *digo/dice* and related forms indicating direct or indirect speech, (b) the presence of other generic pronominal forms (e.g., *uno*, *alguien*, *nadie*), and (c) the use of the impersonal *se* construction.

3.2.4 Ambiguity

Most Spanish verb forms include information about person and number. However, some forms are ambiguous with respect to person. For example, imperfect *estaba* may mean I, s/he, or you (formal) were. This factor has proven to be significant in many studies and is tested here.

Silva-Corvalán (1996–97) proposed an alternative explanation of the effect of ambiguity with respect to person in promoting a speaker's use of an overt pronoun. She noted that the preterit, where the focus is on the verb (as well as *ser* 'to be') contains no forms that would result in ambiguity of person. In Mexican Spanish, ambiguity of person may result in the imperfect, conditional, and the present subjunctive, where 1 sg and 3 sg are identical in form. Silva-Corvalán argued that features of the tense-aspect system rather than ambiguity were responsible for the observed variation. She suggested that the focus is on the verb in the case of the preterit and on the subject in the case of the imperfect, conditional, and subjunctive, with the present having an intermediate value.

3.2.5 Verb semantics

Following Travis (2007), we also coded for the effect of verb semantics. In addition to Travis' categories of psychological verbs, speech act verbs, motion verbs, and copulas (116−117), we also distinguished verbs of perception, counting all tokens that did not fall under one of these five categories as 'other'.

Verbs of emotion formed a large portion of the total verbs of the 'other' category. Although intuitively relatable to the psychological category, we sought to constrain the psychological verbs to those that relate explicitly to mental processing, cognition, and opinion forming. Emotion verbs like *querer* "to want' and *desear* 'to desire' as well as verbs denoting non-dynamic mental states (*necesitar* 'to need', *tener* 'to have', *poder* to be able', *querer* + infinitive) were thus treated as 'other'. In some cases, a polysemous verb could fall into more than one category depending on the context. A verb like *agarrar*, when understood as 'to begin to understand', counted as a psychological verb, but in cases where its literal sense of 'to capture' was implied, it was treated as 'other'. Similarly, *conocer* as 'to know (someone)' was a psychological verb but as 'to meet (someone)', it was an 'other' verb, and *fijarse* as 'to pay attention to' was a psychological verb, but as 'to notice' was coded as a verb of perception.

3.2.6 Lexical aspect

We also coded for the effects of lexical aspect on SPP use following Vendler's (1967) original categories of achievement, accomplishment, activity, and state. Andersen describes these categories in relation to the idea of 'energy': states are atelic events that continue without any energy,

activities are atelic events that require energy to continue, accomplishments are telic activities that require energy to begin and then complete, and achievements are non-durational, punctual events that require energy (Andersen, 1991: 310). To distinguish verbs from these different categories, we used several linguistic 'tests' developed in Andersen (1991) and Shirai and Andersen (1995).

For example, we treated *creer* 'to believe' (6) as a state, *ayudar* 'to help' (7) as an activity, *explicar* 'to explain' (8) as an accomplishment, and *llegar* 'to arrive' (9) as an achievement.

(6) **State:** ¿y **tú crees** que en America nomás se habla inglés?
   'and you believe that in America only English is spoken?'
(7) **Activity: Yo le ayudaba** con sus … problemas de matemáticas.
   'I used to help him with his math problems.'
(8) **Accomplishment:** /Te **voy a leer** un cuento.
   'I'm going to read you a story.'
(9) **Achievement:** …pero quando **llegan** aquí a la casa….
   '…but when they arrive here at home….'

3.2.7 Social Factors

In addition to the possible linguistic constraints discussed above, we also coded for a number of social factors including geographical location (Texas or California) and gender. Because we wished to test the possible influence of language contact with English, we also coded for whether speakers were bilingual or had only minimal or no proficiency in English. Finally, because speakers varied widely in their rate of SPP use, although not in the effect of the main constraints, we coded each individual speaker as a factor.

3.3 Atypical pronominal subjects

Diverging slightly from much of the literature on SPP variation, we included less prototypical pronominal subjects. When speaking generally, participants often used pronominal subjects such as *alguien* 'someone', *los dos* 'the two', *gente* 'people', *muchos* 'many', *nadie* 'no one', *otros* 'others', *todos* 'everyone', or, most frequently, *uno* 'one'. We treated such forms as possible contexts of SPP variation given that in these generic contexts, once the non-specific pronoun has been used, subsequent reference to the generic subject can be realized as either an overt or a null pronoun. Example (10) illustrates this patterning.

(10) Es raro cuando **alguien llega** y 'tá hablando ingles solamente. Se habla inglés con alguien que no sabe el español, pero si **Ø sabe** español, … **Ø lo habla** en español.
   'It's unusual when someone arrives and speaks only English. English is spoken with someone who doesn't know Spanish, but if s/he knows Spanish, s/he speaks in Spanish.'

3.4 Exclusions

A small number of tensed verbs in our corpus were excluded from our analysis. Such verbs occurred in contexts that arguably do *not* permit SPP variation, although in some cases, there is considerable debate about whether or not such contexts legitimately preclude variation (see, e.g. Amaral and Schwenter 2005). We excluded verbs that (a) heavily indicated a contrast in subject reference from the subject of a prior tensed verb or, alternatively, (b) indicated a heavy emphasis on the subject, (c) were embedded in dependent clauses, (d) appeared as the second conjunct in a coordinated structure, (e) were used in interjections or set phrases such as *tú sabes* 'you know', or (f) were repetitions of immediately preceding tensed verbs.

3.5 Frequency

Because we wished to test Erker and Guy's (2012) hypotheses concerning the role of frequency, we coded for frequency using the same criterion they used. Verb forms that accounted for one percent or more of the tokens in the corpus were counted as frequent. All other verbs were classi-

fied as non-frequent. Nineteen verb forms accounted for 31.1 percent of all verb tokens in the corpus, with a mean frequency of 137.46, compared to 4.86 for the non-frequent forms. The 2,612 frequent forms are shown in table 1, which also includes a total for the non-frequent verb forms.

3.6 Analysis

Data were analyzed with Rbrul, a specialized application of logistic regression that allows the researcher to include individual speakers as random effects (Johnson 2009). We performed three separate analyses. First, we examined frequent verbs and other verbs in separate analyses. We then analyzed all 8,676 tokens together, including frequency as a factor group.

Verb form	N	% of corpus	% overt	Verb form	N	% of corpus	% overt
digo (I say)	290	3.3	42.1	sabe (s/he knows)	123	1.4	48.8
dice (s/he says)	185	2.1	37.3	voy (I go)	119	1.4	40.0
hablan (they speak)	184	2.1	25.4	está (s/he is)	115	1.3	33.9
sé (I know)	177	2.0	37.3	puedo (I can)	108	1.2	29.6
creo (I believe)	170	2.0	76.5	dije (I said)	95	1.1	45.3
tengo (I have)	164	1.9	39.6	van (they go)	94	1.1	16.0
tiene (s/he has)	139	1.6	33.1	estaba (I, s/he was)	93	1.1	39.8
estoy (I am)	130	1.5	35.4	tenemos (we have)	91	1.0	16.5
están (they are)	125	1.5	14.4	tienen (they have)	86	1.0	19.8
habla (s/he speaks)	124	1.4	41.1	All others	6064	69.9	36.6

Table 1. Frequent verb forms

4 Results

The results of multivariate analysis reveal a rich patterning of constraints. For all analyses, co-reference, person/number, semantic class, and individual speaker proved to be significant at the .05 level. In the analysis of non-frequent verbs only, two additional factor groups reached significance: ambiguity by tense/mood/aspect and lexical aspect. The overall rate of use of overt pronouns differed only minimally in the analyses of frequent and non-frequent verb forms. For frequent forms, the overall rate was 36.8 percent; for non-frequent forms, the rate was 36.6 percent. In the combined analysis, all of the factors that reached significance in the separate analyses also reached significance. Frequency, which we considered as a binary factor group following Erker and Guy (2012), did reach significance, contrary to the hypothesis that frequency has no independent effect. Frequency, as a factor, disfavors the use of overt SPPs. Finally, when speaker was included as a random variable, none of the external constraints–region, gender, or bilingualism–reached significance at the .05 level in any analysis.

In the following sections we first present the results for the separate analyses. We then present the combined results, with particular attention to the role of frequency.

4.1 Co-reference

As expected, co-reference proved to have a significant effect in all our analyses. A switch from the subject of the preceding tensed verb favored use of an overt subject, while subject continuity favored the use of a null subject. As the results in table 2 show, however, there is very little difference in the effect of this constraint when frequent and non-frequent verbs are analyzed separately. In the analysis with frequent verbs only, 43.8 percent of the tokens were used with an overt pronoun, with a weight of .575, while 29.4 percent of same reference tokens included an overt pronoun, with a weight of .425. For the infrequent verb forms, the corresponding figures are 41.4 percent of the tokens used with an overt pronoun, with a weight of .581, and 31.5 percent of the tokens used with a null subject, with a weight of .419. These results are contrary to Erker and Guy's (2012) result showing that token frequency amplifies the constraint effects that have been docu-

mented in previous work, such as switch reference. Rather, in the data presented here, frequency has almost no relationship to the co-reference factor group.

4.2 Person and number

Person/number was one of three internal constraints that proved to be significant in the analyses of both frequent and non-frequent verb forms. As in the case of co-reference, the results are consistent with other studies (see Flores-Ferrán 2007b for a review). In the analysis that included frequent forms only, the results show that 1 sg (.638) and 3 sg/Ud. (.619) favor the use of an overt pronoun, while 2/3 pl (.410) and 1 pl (.335) disfavor overt pronoun use. Several facts about the frequent forms require attention, however. First, there are no 2 sg forms among the frequent verb forms. Second, the frequent verb forms include very few 1 pl forms, which strongly disfavor overt pronoun use. In fact, *tenemos* 'we have' (N = 91) is the only 1 pl form among the frequent forms. In contrast, the frequent forms include a much higher percentage of 1 sg forms, which strongly favor overt pronoun use, than the non-frequent forms. Thus, 49.6 percent of the frequent forms are 1 sg, compared to only 33.3 percent of the non-frequent forms. Given that, we might expect the overall use of overt pronouns to be considerably higher among the frequent forms, regardless of whether frequency has any effect. However, as results in table 2 show, that is not what we found.

The data for the non-frequent forms contain two more person/number factors (2 sg +/– specific) than the corresponding group in the analysis of frequent forms, and although this has some effect, the overall constraint ranking for this factor group is quite similar to the ranking for the frequent forms. Thus, 1 sg strongly favors use of an overt pronoun (.759). In fact a higher percentage of 1 sg non-frequent forms are used with an overt pronoun (54.4 percent) than is the case among the frequent forms (44.2 percent), and this difference is reflected in the factor weights. In Erker and Guy's (2012) model, however, frequency activates the effect of person and number on SPP use. In their data, the person/number constraint has no effect on SPP use when only non-frequent verbs are analyzed. The results presented here, based on a larger data set, suggest that their conclusion with respect to this factor cannot be generalized.

Factor group	Factor	Frequent forms			Non-frequent forms		
		N	%	weight	N	%	weight
Co-reference	Switch	1344	43.8	.575	3129	41.4	.581
	Same	1268	29.4	.425	2935	31.5	.419
Person/number	1 sg	1295	44.2	.638	1900	54.4	.759
	2 sg (+SPEC)	na	na	na	245	36.3	.539
	2 sg (–SPEC)	na	na	na	148	13.5	.276
	3 sg/Ud.	736	38.3	.619	1278	36.9	.611
	1 pl	91	16.5	.355	811	14.5	.290
	2, 3 pl	490	19.8	.410	1232	26.2	.527
Semantic class	Psychological	532	50.4	.585	983	42.4	.492
	Copula	222	36.5	.507	499	39.3	.529
	Speech act	991	29.2	.483	958	34.8	.513
	Other	867	29.2	.425	3627	37.5	.466
Tense-mood-aspect X ambiguity	Imperf, subj., cond., ambig.	93	39.1	ns	834	51.6	.577
	Present	2424	36.4	ns	3230	36.5	.535
	Preterit (+ser)	95	47.0	ns	1326	33.3	.467
	Imperf, subj., cond., non-ambig.	--	na	na	674	33.1	.421
Lexical aspect	Stative	1347	39.2	ns	2486	41.1	.562
	Activity	502	43.2	ns	1708	33.8	.499
	Punctual	702	38.9	ns	849	33.5	.488
	Telic	61	27.9	ns	1021	33.1	.451
Total	Input	2612	36.8	.274	6064	36.6	.233

Table 2. Linguistic factors: Frequent and non-frequent verb forms

4.3 Semantic class

Semantic class also proved to be significant for both frequent and non-frequent verbs. However, unlike co-reference and person and number, the constraint ordering differed in the two separate analyses. For frequent verbs, the order was psychological > copula > speech act > other. Among the non-frequent verbs, the order was copula > speech act > psychological > other. We suggest that one reason for the difference in constraint ranking may be the large number of tokens of *creo* 'I believe' among the frequent verbs (n = 170). Although overt pronoun use with *creo* is by no means categorical, for some speakers, *yo creo* seems to function as a frozen expression.

4.4 Tense/mood/aspect and ambiguity

Testing for both tense/mood/aspect and ambiguity required that we combine the two factor groups, as shown in tables 2 and 3. Not surprisingly, given the predominance of 1 sg present forms among the frequent verbs, the factor group failed to reach significance in the analysis that included frequent verb forms only. TMA by ambiguity, however, was significant when only the non-frequent forms were included. As expected, overt pronoun use was more likely with an imperfect, subjunctive, or conditional ambiguous form. Interestingly, overt pronoun use was disfavored with imperfect, subjunctive or conditional non-ambiguous verb forms. However the fact that such forms are often plural, a disfavoring environment, may explain the result.

4.5 Lexical aspect

As shown in table 2, lexical aspect, defined in Vendler's (1967) terms, reached significance for the non-frequent verb forms, but not for the frequent forms. Among the non-frequent forms, statives were most likely to be used with an overt pronoun, followed by activity verbs, then punctuals, and finally telics.

Factor Group	Factor	N	%	Weight
Co-reference	Switch	4473	42.1	.578
	Same	4203	30.9	.422
Person/number	1 sg	3195	50.2	.742
	2 sg (+ spec)	245	36.3	.543
	2 sg (− spec)	148	13.5	.272
	3 sg	2464	37.2	.637
	1 pl	902	14.7	.295
	2/3 pl	1722	24.4	.515
TMA x ambiguity	Ambiguous, imperfect, cond., subj.	926	50.3	.564
	Present	5643	36.5	.536
	Preterit + ser	1426	36.1	.499
	Non-ambig., imperf, cond., subj.	681	37.7	.423
Semantic features	Speech act	1949	35.6	.550
	Psychological	1525	45.2	.526
	Copula	721	38.4	.517
	Perception	379	36.4	.504
	Other	3526	34.5	.464
	Motion	586	29.9	.439
Lexical aspect	Stative	3833	40.5	.560
	Activity	2210	32.5	.484
	Telic	1723	35.5	.450
	Punctual	910	33.1	.506
Frequency	Frequent (> 1% of verb forms)	2612	36.8	.451
	Non-frequent	6064	36.6	.549
Input	Corrected mean	8676	36.7	.193

Table 3. Combined analysis: Frequent and non-frequent forms

4.6 Combined analysis

Table 3 shows the results for the linguistic constraints for a combined analysis that included both frequent and non-frequent verb forms. Space precludes a full discussion of the combined results, but they generally agree with previous studies. For example, a switch in reference favors an overt form, while continuity of reference favors the null form. However, in the combined analysis, contrary to what Erker and Guy maintain, frequency does have a significant independent effect, although the effect is considerably less than the effect of such well-established constraints as person/number and switch reference. Frequent verb forms disfavor the overt option (.451), while frequent forms favor overt SPPs (.549).

5 Discussion and Conclusion

This study presents evidence that frequency has neither a magnifying nor activating effect on SPP use. Rather, frequency has a relatively small effect on SPP use among the speakers examined here. The only factor group where frequency seems to have a magnifying effect is the semantic features group, where the spread in factor values is greater among the frequent than among the non-frequent verb forms. Nevertheless, non-frequent verbs show a more complex array of significant linguistic constraints and, in two of the three cases where internal factor groups are significant in both analyses, the results of separate analyses of frequent and non-frequent verb forms clearly show that constraints such as switch reference and person/number have stronger effects among the non-frequent verb forms. For example, 78 percent of the frequent verb forms are singular, a factor that generally favors overt SPP use, compared to 59 percent of the non-frequent verbs. Despite this difference in distribution, the overall difference in rate of use between frequent and non-frequent forms is only .2 percent. The fact that frequency fails to activate or magnify constraint effects can be seen even more clearly in the person/number factor group. Among the frequent verb forms, 44.2 percent of 1 sg forms, a factor that all studies show favors overt pronoun use (Flores-Ferrán 2007b), are used with an overt pronoun. Among the non-frequent forms, the percentage used with overt pronouns in this environment rises to 54.4.

As a corollary to the idea that frequency activates and/or magnifies constraint effects on SPP use, Erker and Guy suggest that frequency has no independent effect. As shown in table 3, however, in our study frequency is significant, although Rbrul results indicate that the significance level is considerably less than is the case with co-reference or person/number.

The literature on frequency presents a somewhat confusing picture. In some studies of frequency and phonological variation, such as Bybee (2002), lexical frequency has a strong effect on coronal stop deletion, while Walker's (2012) recent study of the same variable finds that only phonological and morphological factors are significant once interaction and lexical effects are accounted for. With respect to the role of frequency in SPP variation, we have found a similar contradictory result. Erker and Guy (2012) suggest that frequency operates behind the scenes, either activating or magnifying the well-established constraints that other studies have found to be significant. Our results suggest that their account does not hold up. Rather, constraint effects are generally stronger for non-frequent verb forms. Moreover, in contrast to the view that frequency has no independent effect, our results show that frequency does have a significant, if relatively minor, effect on SPP variation. As a factor in studies of sociolinguistic variation, frequency appears to be something of a Cheshire cat, appearing fully in some studies, faintly in others, and not at all in still others. With respect to the role of frequency in SPP variation, Gregory Guy (personal communication 2012) has suggested that we now have two data points: Erker and Guy (2012) and the current study. Clearly we need additional studies of the role of frequency in SPP variation if we are to understand whether current theorizing about usage based models can be extended to this area of the grammar and to other cases of syntactic variation.

References

Amaral, Patricia Matos, and Schwenter, Scott. 2005. Contrast and the (non-)occurrence of subject pronouns. In *Selected Proceedings of the 7th Hispanic Linguistics Symposium*, ed. David Eddington. 116-27. Som-

erville, MA: Cascadilla Proceeding Project.

Andersen, Roger. 1991. Developmental sequences: The emergence of aspect marking in second language acquisition. In *Crosscurrents in Second Language Acquisition and Linguistic Theories*, ed. Thom Huebner and Charles A. Ferguson, 305–24. Amsterdam: John Benjamins.

Bayley, Robert, and Pease-Alvarez, Lucinda. 1997. Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change* 9: 349–71.

Bentivoglio, Paola. 1987. *Los sujetos pronominales de primera persona en el habla de Caracas.* Caracas: Universidad Central de Venezuela, Consejo de Desarrollo Científico y Humanístico.

Bybee, Joan. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language, Variation and Change* 14: 261–90.

Bybee, Joan. 2010. *Language, Usage and Cognition.* Cambridge: Cambridge University Press.

Cameron, Richard. 1993. Ambiguous agreement, functional compensation, and nonspecific *tú* in the Spanish of San Juan, Puerto Rico and Madrid, Spain. *Language Variation and Change* 5: 304–34.

Cameron, Richard. 1996. A community-based test of a linguistic hypothesis. *Language in Society* 25: 61–111.

Cameron, Richard, and Flores-Ferrán, Nydia. 2004. Perservation of subject expression across regional dialects of Spanish. *Spanish in Context* 1: 41–65.

Erker. Danny, and Guy, Gregory R. 2012. The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language* 88: 526–57.

Flores-Ferrán, Nydia. 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.

Flores-Ferrán, Nydia. 2007a. A bend in the road: Subject personal pronoun expression in Spanish after 30 years of sociolinguistic research. *Language and Linguistics Compass* 1: 624–52.

Flores-Ferrán, Nydia. 2007b. *Los mexicanos in New Jersey*: Pronominal expression and ethnolinguistic aspects. In *Selected Proceedings of the Third Workshop on Spanish Sociolinguistics*, ed. Jonathan Holmquist, Augusto Lorenzino, and Lotfi Sayahi, 85–91. Somerville, MA: Cascadilla Proceedings Project.

Hochberg, Judith. 1986. Functional compensation for /s/ deletion in Puerto Rican Spanish. *Language* 62: 609–21.

Johnson, Daniel E. 2009. Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3: 359–83.

Otheguy, Ricardo, and Zentella, Ana Celia. 2012. *Spanish in New York: Language Contact, Dialect Leveling, and Structural Continuity.* Oxford: Oxford University Press.

Ranson, Diana. 1991. Person marking in the wake of /s/ deletion in Andalusian Spanish. *Language Variation and Change* 3: 133–52.

Schecter, Sandra R. and Bayley, Robert. 2002. *Language as Cultural Practice: Mexicanos en el norte*. Mahwah, NJ: Lawrence Erlbaum.

Shin, Naomi Lapidus. 2012. Variable use of Spanish subject pronouns by monolingual children in Mexico. In *Selected Proceedings of the 14th Hispanic Linguistics Symposium*, ed. Kimberly Geeslin and Manual Díaz-Campos, 130–41. Somerville, MA: Cascadilla Proceedings Project.

Shirai, Yasuhiro, and Andersen, Roger. 1995. The acquisition of tense-aspect morphology: A prototype account. *Language* 71: 742–63.

Silva-Corvalán, Carmen. 1994. *Language Contact and Change: Spanish in Los Angeles.* Oxford: Clarenden Press.

Silva-Corválan, Carmen. 1996–97. Avances en el estudio de la variación sintáctica. *Cuadernos del Sur* 27: 35-49.

Travis, Catherine. 2007. Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change* 19: 101–33.

Vendler, Zeno. 1967. Verbs and times. In *Linguistics in Philosophy*, ed. Zeno Vendler, 97–120. Ithaca, NY: Cornell University Press.

Walker, James A. 2012. Form, function, and frequency in phonological variation. *Language Variation and Change* 24: 397–415.

Department of Linguistics
University of California, Davis
One Shields Avenue
Davis, CA 95616
rjbayley@ucdavis.edu
kaware@ucdavis.edu
clmessing@ucdavis.edu