Haya vs. *Haiga*: An Analysis of the Variation Observed in Mexican Spanish Using a Mixed Effects Model

Mary Johnson and Sonia Barnes
The Ohio State University

1. Introduction

In Spanish there exists an alternation between the present subjunctive forms of the verb *haber*. In the standard conjugation, the stem of the verb ends in the palatal fricative /ʝ/, as in *haya*, *hayas*, etc. However, together with this variant, forms in which the stem ends in the voiced velar plosive /g/, such as *haiga*, *haigas*, etc., are frequently heard in spoken Spanish. Table 1 shows the conjugation of both variants.

Form	Standard conjugation with /ʝ/	Non-standard conjugation with /g/
1st pers. sing	*haya*	*haiga*
2nd pers. sing	*hayas*	*haigas*
3rd pers. sing	*haya*	*haiga*
1st pers. pl.	*hayamos*	*haigamos*
2nd/3rd pers. pl	*hayan*	*haigan*

Table 1. Standard and non-standard conjugation of *haber* in the present subjunctive.

The non-standard forms with /g/ are the result of the analogical extension of a phonological phenomenon that took place in the development from Latin to Spanish (Malkiel 1974, Lloyd 1987, Elson 1988, Penny 2002, Fondow 2010, Kania 2011). An etymological stem-final velar is found in the first person present indicative and in all the forms of the present subjunctive of the verbs *hacer*, *decir* and *yacer: hago* (<*fac(i)o*)*, digo* (<*dico*) and *yago* (<*iac(e)o*). Because these were high frequency verbs, Spanish speakers began to associate the presence of /g/ with first person singular in the present indicative and all the present subjunctive forms and the sound was added to verb forms whose stem did not originally end in a velar plosive, such as *venio* and *teneo*, resulting in *vengo* and *tengo*. It should be noted that this regularization only affected verbs that belonged to the *–er* and *–ir* conjugations. The first verbs to undergo analogical extension were those whose stem ended in /n/ (as in *poneo* > *pongo*), followed by those ending in /ɾ/, /s/, /l/ and /ʝ/, such as *habeam* (>*haya* ~ *haiga*). Even though it is accepted that this phenomenon is the result of analogical extension, there exists controversy regarding the sources and the motivation for the analogy, which go beyond the scope of this paper. Part of the discussion is centered around the status of the inflectional yod present in *venio* and *teneo* (Elson 1988, Lenfest 1993) and the role that verbs with the Latin ending *–ngere*, such as *tañer, ceñir* and *plañir*, had in the extension of the velar insert. While some scholars propose *veno* and *teno* (without yod) as the form that was used before the insertion of the velar occurred (e.g. Lenfest 1993), most believe that the yod survived long enough to palatalize the nasal and propose *veño* and *teño* as the starting point (e.g. Lloyd 1987). Lloyd also maintains that the alternation observed between *taño* and *tango* (from *tangere*

* We would like to acknowledge Scott Schwenter for his critiques and insights. We would also like to thank Kathryn Campbell-Kibler and the members of SoMean discussion group for their input in the development of this project. We also wish to thank the attendees of WSS6 and the anonymous reviewers for their thoughtful comments and suggestions.

> *tañer*) provided the model for the velar insertion in *tengo* and *vengo*. However, Malkiel (1974:328) rejects this hypothesis based on the fact that the frequency of verbs like *tañer* was too low to influence the development of *teneo* and *venio*. Instead, he proposes that the palatalization of the nasal resulted in an alternation between /n/ and /ŋ/ in the paradigm (*teño* vs. *tienes*), which was abandoned in favor of a stronger contrast between /n/ and /ng/.

According to Fondow (2010:47) some of the verbs that underwent analogical extension maintained the velar from the beginning. Such is the case for *caigo, pongo, traigo* and *valgo* among others. In other cases, the alternation that was registered in the Middle Ages was resolved in favor of the variant without the velar, as in *atribuigo* and *destruiga*, where the only forms used in modern Spanish are *atribuyo* and *destruya*, respectively. Finally, some verbs continued to show variation until present. Fondow reports the existence of dialectal variation for the following verbs: *bullir, concluir, coser, creer, dar, doler, haber, huir, ir, moler, parir, reír, soler, ver* and *zambullir,* with *haber, doler, ir*, *huir* and *ver* being the most commonly cited verbs showing synchronic variation with regards to the velar insert. For example, Penny (2004:220) points out that the "use of *haiga, vaiga, huiga,* and occasionally *veiga*, as present subjunctive forms of *haber, ir, huir* and *ver*, is found all over the Spanish-speaking world, at a rural level, and sometimes at other non-standard levels". Cortés Rodríguez (1994:60) mentions the alternation between *haya* and *haiga* as part of a group of linguistic phenomena that are found in all "sociolects" of Spanish. Except for the verbs *roer* and *raer,* where both forms are accepted as prescriptively correct, the variants with the velar insert are normally considered non-standard. Table 2 shows the absolute frequencies of the standard forms for the first and third person singular of the present subjunctive of the most frequently used verbs in which the alternation has been reported. The frequencies were determined using the *Corpus de Referencia del Español Actual* (*CREA*). The absolute frequency reflects the total number of occurrences of each form in all dialectal varieties of Spanish and only in spoken data.

Verb form	Absolute frequency
haya	2538
vaya	984
dé	590
vea	300

Table 2. Absolute frequencies of the standard forms for the first and third person singular of the present subjunctive.

The frequencies in Table 2 indicate that *haya* is much more frequent than the first and third person singular forms of any other verb that has been reported to show variation with respect to velar insertion. While the forms that show velar insertion in each verb were also searched in *CREA*, we found that only *haiga* returned any results, indicating that the velar variant of *haber* is also more common than that of the rest of the verbs that show variation in modern Spanish.

Despite recognizing the common use of forms like *haiga*, which are socially stigmatized, there has been no empirical research that explores the distribution of the two variants (*haya* and *haiga*) in Spanish. The purpose of this study is to explore what linguistic and extra-linguistic constraints govern the variation between the forms with stem-final /g/ and those with stem-final /ʝ/ in the case of *haber*. The analysis of social factors will allow us to determine whether education, gender, or age have any influence on the choice of one variant over the other. Due to the current stigmatization of the velar forms, we hypothesize that their use will be more frequent in speakers with a lower socioeconomic status or educational level. Examining the role that linguistic factors have in the variation will allow us to further characterize the nature of the analogical process, elucidating whether certain features favor the extension of /g/ and perhaps even why the variation specifically between *haya~haiga* is still observed in spoken Spanish, in contrast with other verbs that underwent the same process and where one of the variants completely disappeared or is less frequently used.

2. Methodology

The data used for this study were extracted from three corpora of Mexican Spanish - one from Monterrey, Mexico (Rodriguez Alfano, 2006), and two from Mexico City (Butragueño & Lastra, 2011 and Lope Blanch, 1976). The Monterrey corpus includes interviews with 117 speakers: 55 females ranging in age from 18 to 75, and 62 males ranging in age from 19 to 76. The Mexico City corpus is composed of interviews with 108 speakers: 54 females ranging in age from 20 to 91, and 54 males ranging in age from 20 to 86. Finally, the *Habla Popular* corpus contains interviews with 45 speakers: 21 females with age ranging from 19 to 72 and 24 males with age ranging from 17 to 72. The corpora include information about education level of each subject, as well as their age and gender. The transcriptions were done by researchers in Mexican universities (Instituto Tecnológico y de Estudios Superiores de Monterrey, Universidad Autónoma de Nuevo León, Universidad Autónoma de México and Colegio de México). All the transcriptions are orthographic, but they reflect phonological and morphological alternations present in the spoken language. In the case of the Mexico City Corpus, the standard form was included next to the form used by the speaker, when this was the non-standard variant. Each of the transcriptions was reviewed by at least one person other than the transcriber. The corpora, in PDF form, were searched for all occurrences of present subjunctive *haber,* including the analogous forms *haiga, haigas, haigamos* and *haigan,* as well as the prescriptive forms *haya, hayas, hayamos* and *hayan*, yielding a total of 423 tokens from 169 speakers. All of these tokens were coded for the following independent linguistic variables: type of use, number, person, type of clause, phonological form and presence or absence of negation. For type of use, the semantic function of the verb was coded, namely whether it is used as an auxiliary (eg: *Espero que Juan haya hablado con Maria* 'I hope that Juan has talked to Maria') or a presentational verb (eg: *Espero que haya mucha gente en la fiesta* 'I hope that there are a lot of people at the party'). Number was coded by looking at the morphology on the verb (singular or plural). Because subjunctive forms appear in subordinate clauses, the tokens were also coded for whether they appeared in a noun clause, an adjective clause or an adverbial clause. Phonological form was coded to consider the fact that first and third person forms have the same phonological form (eg: *Duda que yo haya hablado con María* 'He doubts that I have talked to Maria'*; Duda que Juan haya hablado con María* 'He doubts that Juan has talked to Maria'). This resulted in four different factor levels: *haya~haiga, hayas~haigas, hayamos~haigamos* and *hayan~haigan*.

Because *haber* is used as both an auxiliary verb and as a presentational verb, different linguistic factors were coded for the different uses and an additional statistical analysis was performed for each type of use. Auxiliary uses of *haber* were further coded for person, animacy of the subject, overtness of the subject (overt vs. null) and the type of participle on the verb (*-ado*, *-ido* or irregular form). The coding for person was taken from the morphology of the verb and the surrounding context when necessary (since first and third person singular morphology are identical). Presentational verbs were further coded for the plurality of the entity being introduced (singular vs. plural). Because *haber* as a presentational verb can (and prescriptively does) lack plural morphology even when the entity being introduced is plural, this was taken from the surrounding context.

For each speaker, we also coded for four social factors. These variables included geographic variety (determined by their city of residence), age, gender and education level. The latter was divided into three groups - low education, middle education and high education. Low education includes speakers with elementary education or less. Middle education includes speakers with secondary education (obligatory school until age 16), *preparatoria* (optional pre-college preparatory school), some *preparatoria*, or technical school. High education includes college, some college, graduate school and some graduate school. Age was tested as a continuous variable and as a categorical variable, which had three different levels: young (18-35), middle (35-50) and old (+50).

To test our hypotheses a mixed effects model, using speaker as a random effect and the independent variables described above as fixed effects, was developed using the lmer function (Bates, Maechler & Bolker, 2011) in R (R Development Core Team, 2007). The role of the fixed and random effects were first tested using the complete data set, which included both uses of *haber*, i.e. as an auxiliary and as a presentational. Auxiliaries and presentationals were then tested separately to include

the independent variables that are particular to each type. Finally, the tokens from the two geographic varieties included in the study were also tested separately.

3. Results

Our results indicate that *haiga* is less frequent than *haya* overall. The frequencies can be seen in Table 3, below.

	Haber with velar insert	Haber without velar insert
Total	**154 (36.4%)**	**269 (63.6)**
Mexico City Corpus	40 (22.6%)	137 (77.4%)
Monterrey Corpus	102 (44.7%)	126 (55.3%)
Habla Culta	12 (66.7%)	6 (33.3%)

Table 3. The frequencies of *haiga* and *haya* in the complete data set and separated by corpora.

The results of the mixed-effects model on the complete data set using speaker as a random effect show that education level is the most significant predictor of the choice between *haya* or *haiga.* The odds of less educated speakers using *haiga* are significantly higher than those of speakers with a higher education level, while the behavior of speakers in the middle education group does not differ significantly from that of speakers in the high education group. Furthermore, an interaction between Education and Type of use was found to be significant, such that presentationals are more likely to have the velar insert in low education groups than in middle or high education groups. The results can be seen in Table 4. The interaction is depicted in Figure 1.

	Estimate	*SE*	*z*Value	*p*-value
(Intercept)	-4.98	1.02	-4.88	<0.01
Education level (reference level is High ed.)				
Low education	**5.59**	**1.17**	**4.76**	**<0.01**
Middle education	0.30	1.88	0.16	0.87
Interaction (Education & Type of use)				
High education and Presentational	0.19	0.95	0.20	0.84
Middle education and Presentational	1.29	1.66	0.77	0.44
Low education and Presentational	**1.80**	**0.64**	**2.82**	**<0.01**

Table 4. Factors contributing to the insertion of /g/ in stem-final position in auxiliary and presentational uses of *haber.*

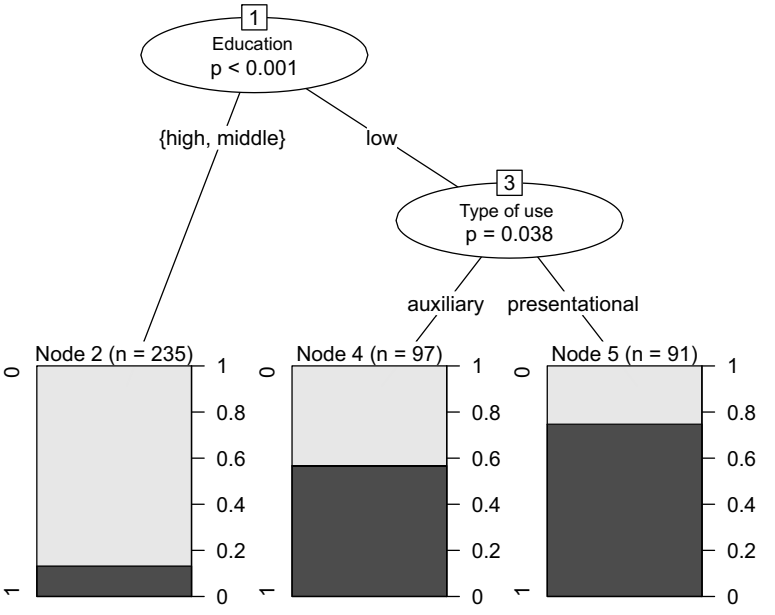

Figure 1. Conditional inference tree showing the interaction between Education and Type of use.

Conditional inference trees evaluate whether an independent variable is a useful predictor of the two possible responses of the dependent variable and divides each of the predictors into subsets that are further evaluated, considering the effect of subsequent factors. In Figure 1 the inference tree first considers whether splitting the data into two educational groups corresponds with more use of *haiga* or *haya*. The results indicate that low education corresponds with higher rates of *haiga*. Then, the importance of type of use is considered for the two subsets into which education was divided. The inference tree shows that, while the division between auxiliary and presentational is significant in the low education subset, it is not significant for the high and middle education group. In the low education subset presentational use corresponds to higher rates of *haiga*.

The data was then divided by geographic variety and a mixed-effects model was developed again for each group. Results for the Mexico City speakers indicate that Education is the only significant predictor in the choice between *haiga* and *haya.* The odds of having the velar insert increase in the low education group. These results can be seen in Table 5.

	Estimate	*SE*	*z*Value	*p*-value
(Intercept)	-5.81	1.69	-3.43	<0.01
Education level (reference level is High ed.)				
Low education	**6.53**	**1.89**	**3.46**	**<0.01**
Middle education	0.56	2.61	0.22	0.83

Table 5. Factors contributing to the insertion of /g/ in stem-final position in auxiliary and presentational uses of *haber* in the Mexico City data.

Results for the Monterrey speakers resemble those from the complete data set, in that Education and the interaction between Education and Type of use are the only significant predictors in the choice between *haiga* and *haya*. The odds of finding *haiga* are significantly higher in the low education group than in the high education group. Even though the difference was not found to be significant, the middle education group patterns with the low education one, as opposed to what was reported for the complete data set and the results for Mexico City. Like the complete data set, the interaction between Education and Type of use tells us that presentationals are more likely to have the velar insert in the

low education group than in the high or middle education groups. These results are reported in Table 6.

	Estimate	*SE*	*z*Value	*p*-value
(Intercept)	-5.76	1.67	-3.45	<0.01
Education level (reference level is High ed.)				
Low education	**6.69**	**1.88**	**3.56**	**<0.01**
Middle education	5.06	3.04	1.67	0.10
Interaction (Education & Type of use)				
High education and Presentational	1.47	1.53	0.96	0.34
Middle education and Presentational	1.20	2.96	0.41	0.69
Low education and Presentational	**2.09**	**0.88**	**2.38**	**0.02**

Table 6. Factors contributing to the insertion of /g/ in stem-final position in auxiliary and presentational uses of *haber* in Monterrey Spanish.

4. Discussion

4.1. Education level

Our initial hypothesis that education level would influence the choice of *haiga* vs. *haya* was confirmed by the results above. This is not surprising, given what previous research has said regarding the distribution of the two variants across social classes. Santa Ana & Parodi (1998) propose an analysis of the Spanish spoken in Michoacan, Mexico by speech communities. The investigators divide the speakers into one of four "fields" (or speech communities) according to the expansion of their speech community and their awareness of stigmatized, regional and standard variants. This model is compatible with that of Milroy (1980), such that the members of lower fields belong to tighter, more local social networks, while members of higher fields belong to looser, more expanded social networks. Furthermore, members of higher fields are more aware of the stigmatized, regional and standard forms and can control their use of them. On the other hand, members of lower fields are less aware of these variants or not aware at all, and, thus, use stigmatized forms more freely. Santa Ana & Parodi observe that the stigmatized aspects of Spanish that they studied contain lexical, phonological and morphological elements from old American Spanish that was spoken in the Americas during the 16th century, and they cite *haiga* as one of these stigmatized forms. They note that now, these forms are labeled as rural or uneducated speech, and are censured in school. Because these forms are censured in school, we could expect speakers with higher education levels to be more aware of the stigma associated with *haiga*, and therefore more likely to avoid it. Less educated speakers would be less aware of this stigma, and thus more willing to use *haiga.* As noted by Milroy (1980), tighter social networks act as a norm reinforcement mechanism for nonstandard variants. Therefore, lower educated speakers would be more likely to have more use of *haiga* because it marks solidarity within their social network. This would motivate their lack of avoidance of the stigmatized form, and helps to explain the higher frequencies of *haiga* in the speech of our low education group.

4.2. Geographic Variety

Because the interaction between Education and Type of use was only observable in the complete data set and in the Monterrey data, but not in the Mexico City data, the role that geographic variety might play in the variation should be explored. We should emphasize that Geographic variety was not a significant predictor of velar insertion when treated as an independent variable in the analysis of the complete data set, indicating that living in Monterrey or Mexico City does not significantly influence the use of *haiga* vs. *haya*. One explanation for the observed results with regards to the interaction

between Education and Type of use is that there are more low education speakers in Monterrey than in Mexico City. The distribution of the speakers by city and education level is shown in Table 7.

	Low	Middle	High
DF	67	46	82
Monterrey	121	13	94

Chi-squared = 32.4, df = 2, p<.01
Table 7. The distribution of speakers by geographic variety and education level.

We postulate that, if the distribution of speakers by education level were more even across geographic variety, the same interaction would have been found in Mexico City. Evidence in support of this hypothesis can be found in the comparison of the results from the complete data set to those from Monterrey. Recall that the effect of Type of use within low education speakers is stronger in the complete data set ($p<0.01$) than in the Monterrey data alone ($p=0.02$). This suggests that it is not only the Monterrey data that is yielding this effect in the complete data set. If that were the case we would expect a higher *p-value* for the interaction in the complete data set than in the results from Monterrey alone. The addition of the Mexico City data helps strengthen the effect, thus suggesting that Mexico City may also show this trend. Had there been more low education speakers in the Mexico City corpus, this effect may also have been observed in the results from this corpus alone. However, further research is necessary to confirm any effect of Geographic variety.

4.3. Type of use

The results of the statistical analysis also show that the forms with the velar insert are favored in presentational uses of the verb *haber* in the Spanish spoken in Monterrey and in the complete data set within the low education group. One might think that the higher rate of velar insertion in presentational tokens could be attributed to the fact that only third person singular (and occasionally plural) is used in this type of construction. However, neither person nor phonological form was selected as significant predictors in the best fit model. It could also be plausible that frequency is contributing to this result; however, the distribution of presentational and auxiliary uses of *haber* throughout our data is quite even, at 202 and 221 tokens respectively.

The differential effect of construction type could be attributed to the semantic differences that exist between the two uses. As an auxiliary, *haber* is a grammatical word that only carries inflectional information. In its presentational use *haber* is a lexical verb that is used to introduce mainly new referents into the discourse. It is possible that the variation between *haya* and *haiga* responds to a process of morphologization of the semantic difference between the two types of use, due to a desire on the part of the speaker to further distinguish two semantically different items. Previous research on velar insertion addresses the possibility of the velar having an indexical or morphological function. Elson (1988:394), for example, claims that the velar was independent from the verb stem. According to the author, the function of the velar was to oppose derived forms to basic forms, which he assumes to be the realizations occurring in the infinitive. He explains that "The distribution of *g* versus its absence in the present is identical to that of derived forms of the verbal stem versus the basic form […] with CVC-e/i- representing the basic stem of a second or third conjugation verb, and CVC-/CVC-a- representing the derived stems". Because the presence of the velar correlates with the derived forms, it is possible that it is a functional index of them. Lenfest (1993) and Fondow (2010) also stress the independence of the velar insert in the verbal paradigm. They maintain that /g/ is a type of morphological marker that helps speakers set apart a particular class of verbs and whose function was also to increase the stability of the verbal paradigm. Lenfest (1993:641) defends this analysis saying that "The idea that /g/ is a 1st pers. sing. pres. ind. marker is supported not only by the regrammaticalized /k/ of the inchoative infix but also by etymological alternations such as *digo dices; hago haces* and by a number of gratuitous irregularities in the 1st pers. sing. pres. ind. caused by

metathesis (*quepo cabes*); reduction (*sé sabes*); and agglutination (*doy das*), etc.". While the literature posits a variety of reasons that led to the extension of /g/ in the Spanish verb paradigm, we find that in these analyses the potential of /g/ to acquire morphological load is emphasized. We propose that this indexical function is further exploited in the case of *haiga/haya* in order to differentiate two semantic and functional properties.

Based on these observations, we hypothesize that, in the case of Monterrey, the speakers are using the presence or absence of the velar to mark semantic differences. Monterrey had the most low education speakers. Since *haiga* is more present overall in the low education group, it is plausible that the velar insert is more accessible to them for use in marking this semantic difference than it is for high and middle education speakers, who are more likely to be aware of the stigma associated with this form, and thus less likely to make use of it despite its differentiating potential. This could explain why this factor is significant in Monterrey, where lower education speakers are making use of the velar insert for presentationals.

This would not be the first case of the verb *haber* grammaticalizing since, historically, it grammaticalized from a verb of possession to an auxiliary verb, used even with intransitive verbs (Penny 2002) and ultimately replaced by *tener* for possession.

4.4. Other factors

Other factors were considered that our model did not select as significant in the choice of *haiga* or *haya* in our data. Age, for example, was not chosen as part of the best fit model. In the case of a change in progress, we might expect to see older speakers using less of one variable than younger speakers (Bailey et al., 1991, Bailey 2002, Boberg 2004). Lack of an age effect in our data suggests that the phenomenon is not a change in progress. This is not surprising, since, according to Fondow (2010), occurrences of both forms have been in existence since the beginning of 14th century.

The fact that no other linguistic factors were selected as significant explains why this variation continues to be found in *haber*, but is less extensive in other Spanish verbs that previously showed synchronic variation between the forms with and without the velar. The linguistic factors that were not selected, such as person, number, animacy of the subject, overtness of the subject and the presence or absence of negation would presumably apply to all verbs, if these factors played a role in the choice of variants. Type of use, particularly the presentational use, applies uniquely to *haber,* thus explaining why this variation remained prominent in this verb, but less so in other Spanish verbs.

5. Conclusion

The results of this study indicate that education and speaker are the most important factors in the choice of *haiga* or *haya* in the corpora examined. The results for speaker as a random variable suggest that individual speakers have tendencies to use one form or the other. This is supported by the fact that mostly the external factor of education, and not other linguistic factors examined (such as person, number, negation, etc), seems to govern the choice of one variant over the other.

Type of use was only selected as a significant factor in the data from Monterrey and in the complete data set. In these data, the velar insertion is significantly more likely to occur in presentational uses than in auxiliaries within the low education group. Given that this effect is stronger in the complete data set than in the Monterrey data, we postulate that the Mexico City data must contribute to this effect in the complete data set, thus indicating that this is not a regional difference. We attribute the effect being found in Monterrey but not in Mexico City to the fact that the corpora provided many more low educated speakers in Monterrey than in Mexico City. We suspect that with more low education speakers in Mexico City in our sample, the same effect may be found there. More data would help corroborate this.

As for why *haiga* is more likely to occur in presentationals, we propose that the association of each variant to a different type of construction points to a desire on the part of the speaker to differentiate two semantic properties of *haber*. The difference between presentational and auxiliary *haber,* and the productivity of this verb, might explain why the variation between velar and standard forms still takes place in modern Spanish while the rest of the verbs that showed synchronic variation

either completely lost one of the variants or show much lower rates of velar insertion at present.

Erker & Guy (2012) propose that internal constraints on Spanish subject personal pronoun use become more apparent as frequency increases. More generally, they propose that frequency acts as a "gatekeeper" such that there exists a frequency threshold, above which significant linguistic constraints emerge, and below which they do not. Future studies on velar insertion should incorporate a linguistic analysis of its insertion in the less frequent verbs (eg: *ir, dar, ver, etc.*). Because *haber* is the most frequent of these verbs (as evidenced by Table 2), if Erker & Guy's theory holds, we would not expect to find linguistic constraints emerge for less frequent verbs that may contain the velar insert that do not emerge for *haber* variation.

Schwenter & Torres Cacoullos (2010) analyze clitic placement in Spanish, showing that when *ir a* is used to indicate motion (the lexical meaning), more enclisis is favored. When *ir a* is used as a future construction (the grammaticalized, auxiliary meaning), enclisis is disfavored, and more proclisis (the innovative form) is found. With respect to the *haiga* vs. *haya* alternation, while *haiga* is not actually an older form, it is considered to be a remnant of old American Spanish (Santa Ana & Parodi 1998). Then, a comparison can be drawn between what Schwenter & Torres Cacoullos found for clitic placement, and what we have identified in the alternation between these forms of *haber.* Auxiliary uses of *haber,* which express a more grammaticalized meaning, favor *haya*, while *haiga* is found more often with presentational uses of *haber,* the lexical, and diachronically older, meaning.

References

Bailey, Guy, Wikle, Tom, Tillery, Jan & Sand, Lori. (1991). "The Apparent Time Construct." *Language Variation and Change,* Vol. 3, pp. 241-264.

Bailey, Guy. (2002). *Real and Apparent Time*. In J.K. Chambers, Peter Trudgill, & Natalie Schilling-Estes [ed], *The Handbook of Language Variation and Change* (pp 312-331). Oxford, England: Blackwell.

Bates, Douglas, Maechler, Martin & Bolker, Ben. (2011). *lme4: Linear mixed-effects models using s4 classes* (R package version 0.999375-39)

Boberg, Charles. (2004). Real and Apparent Time in language Change: Late Adoption of Changes in Montreal English. *American Speech,* Vol. 79, No. 3, pp. 250-269.

Butragueño, Pedro Martín & Lastra, Yolanda. (2011). *Corpus sociolingüístico de la ciudad de México*. México: El Colegio de México, manuscrito.

Cortés Rodríguez, Luis. (1994). *Tendencias actuales en el estudio del español hablado*. Almería: Universidad de Almería.

Elson, Mark J. 1988. "The Synchronic Status and the Evolution of the *g* in Spanish *vengo, salgo,* etc. Revisited." *Hispania*, Vol. 71, No. 2, pp. 392-400.

Erker, Daniel & Guy, Gregory. (2012). "The Role of Lexical Frequency in Syntactic Variability: Variable Subject Personal Pronoun Expression in Spanish." *Language*, Vol. 88, No. 3, pp. 526-557.

Fondow, Steven Richard. (2010). Spanish Velar-insertion and Analogy: A Usage-based Diachronic Analysis. PhD dissertation. The Ohio State University.

Kania, Sonia. (2011). "The Spread of the Velar Insert in Medieval Spanish Verbs." *Bulletin of Hispanic Studies* Vol. 88, No. 2, pp. 129-159.

Lenfest, Donald Edgar. (1993). "Tengo-Vengo, an Update." *Hispania,* Vol. 76, No. 3, pp. 634-644.

Lloyd, Paul M. (1987). *From Latin to Spanish*. Philadelphia: American Philosophical Society.

Lope Blanch, Juan M. (1976). *El habla popular de México: materiales para su estudio*. México: Universidad Nacional Autónoma de México (UNAM).

Luria, Max A. (1936). "The Pronunciation of siegat in the Old Spanish Glosses of Silos." *Language*, Vol. 12, No. 3, pp. 193-195.

Malkiel, Yakov. (1974). "New Problems in Romance Interfixation (I): The Velar Insert in the Present Tense (with an Excursus on *–zer/-zir* Verbs)". *Romance Philology*, 27, pp. 304-355.

Milroy, Lesley. (1980). *Language and social networks.* Oxford: Blackwell.

Santa Ana, Otto & Claudia Parodi. (1998). "Modeling the Speech Community: Configuration and Variable Types in the Mexican Spanish Setting." *Language in Society,* Vol. 27, No. 1, pp. 23-51. Cambridge University Press.

Penny, Ralph. (2002). *A history of the Spanish language.* Cambridge: Cambridge University Press.

Penny, Ralph. (2004). *Variation and change in Spanish.* Cambridge: Cambridge University Press.

R Development Core Team. (2011). R: A language and environment for statistical computing. Vienna: Author.

Rodriguez Alfano, Lidia. (2006). *El habla de Monterrey, proyecto sociolingüístico*. PRESEA.

Schwenter, Scott & Torres Cacoullos, Rena. (2010, April 23). *Pragmatic Factors in the Placement of Direct Object Clitics in Spanish.* Paper presented at The Ohio State University Congress on Hispanic and Lusophone Linguistics, Columbus, OH.

Selected Proceedings of the 6th Workshop on Spanish Sociolinguistics

edited by Ana M. Carvalho and Sara Beaudrie

Cascadilla Proceedings Project Somerville, MA 2013

Copyright information

ISBN 978-1-57473-456-0 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Johnson, Mary and Sonia Barnes. 2013. *Haya* vs. *Haiga*: An Analysis of the Variation Observed in Mexican Spanish Using a Mixed Effects Model. In *Selected Proceedings of the 6th Workshop on Spanish Sociolinguistics*, ed. Ana M. Carvalho and Sara Beaudrie, 32-40. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2854.