LEXICAL FREQUENCY AND VOICED LABIODENTAL-BILABIAL VARIATION IN NEW MEXICAN SPANISH

RENA TORRES CACUYALLOS and FERNANDA FERREIRA
University of New Mexico

ABSTRACT. In this paper we look at word frequency effects on variation between voiced labiodentals and voiced bilabials in New Mexican Spanish. We find that the occurrence of labiodentals is significantly higher in words of high frequency than those of low frequency. In high frequency words, neither orthography nor English cognate status cause-with the variation, which supports the view that voiced labiodentals represent the retention of an old dialect feature. An internal origin for the labiodental is further supported by their distribution on the social boundaries of age, proficiency, and formal Spanish instruction. At the same time, the pattern of variation in low frequency words suggests that a contrast induced change is underway, where bilabials are favored when the English cognate of a Spanish word is a bilabial, because induced sound change thus seems to pattern like analogical change. In support of the Frequency-Implementation Hypothesis (B. Phillips 1999 and more generally a usage based model of phonological representation (Bybee 2001).)

1. KINDS OF \( \text{v} \) IN SPANISH. A widely accepted generalization about the distribution of voiced bilabial stops and fricatives in Spanish is shown in (1) (after D’Intremont, del Teso & Weston 1995:274).

(1) In syllable onset position, voiced bilabials are realized as stops after a pause and after a nasal, and as fricatives in all other cases. Examples: [blarnz] “let’s go”, [ðmb] “both”, but [hafit] “there was”

In short, (b) and (v) are said to be allophones in complementary distribution of a phonemes /b/ and /β/ orthographically b or v. The difference between b and v in modern Spanish is supposed to be orthographic, not phonetic, much less phonemic (cf. Buckwalter & Brown 1965:48, Banfi & Schweppre 1994:59). Labiodental (v) is said to occur when /f/ is voiced, as in Afghani” (Bartov 1969:37, note 22), or when \( \text{b} \) or \( \text{v} \)

\* A todos que participaron en el estudio: gracias.
follows its aspiration, as in the derivatory 'to wreck' or the vacular 'the cows' (Lope Blanca 1988:160, cf. Salvador 1987). Nevertheless, voiced labiodentals that cannot be attributed to assimilation or another articulatorily motivated process do occur in Spanish. A useful point of departure is provided by Lope Blanca’s classification of non-assimilation [v] into the three kinds shown in 2 (adapted from Lope Blance 1988:160-1).

2. Classes of [v]
   a. Archaic [v], as in varieties of Judeo-Spanish
   b. Language contact [v], as with bilinguals who speak a language that has [v] phoneme, such as Catalan or English
   c. Hypercorrect [v], as in certain reading styles

The clearest cases of archaic [v] are the labiodentals of Judeo-Spanish varieties, which are considered conservative or archaic (Zamora Vicente 1970:354). It is generally assumed that Old Spanish had labiodentals, especially in southern peninsular dialects (A. Alou 1955, D. Alou 1962, Penn 1976, but see Arzúa 1994:471f.). These old voiced labiodentals continued into the 16th century in Andalucía and were carried over to early Latin American Spanish (Lapresa 1981:370, Lloyd 1993:519-20). Today, an etymological distinction between /v/ and /w/ is maintained in Portuguese and in certain Catalan dialects (Parkinson 1988:139, Wheeler 1988:170), though apparently not in any modern variety of Spanish (Remírez 1982:129).

On the other hand, the labiodentals in varieties of Spanish spoken in the United States Southwest have been found to be a function of language contact. Both R. Phillips (1982) in East Los Angeles and Timm (1976) among California college students added a relation between frequency of labiodentals and degree of bilingualism, or English dominance (cf. Post 1934). Mezz (1980) in Tucson, Arizona also observed a slight tendency toward a correspondence between pronunciation and spelling, although she found that [v] was the most frequent variant overall, for both orthographic [v] and [w]. All scholars cited concurred that the least likely phonetic environment for a labiodental [v] is post-canal, where the occurrence of a bilabial stop is categorical for all intents. In any other environment a fricative variant is most likely, but among fricatives, labiodentals are most frequent.

In contrast, labiodentals in varieties of Mexican Spanish are a clear case of hypercorrect [v]. Mexican pedagogic [v], as Lope Blanche (1988) refers to the phenomenon, occurs in formal or emphatic speech, especially in reading style. These labiodentals appear not only in environments assigned to the bilabial fricative, but as often in post-canal position, as in the word convence 'to convince', or in absolute initial position. Notice that pedantic [v] is not constrained by the same phonetic environment as the Southwest labiodentals described above, which were found to be in variation with bilabial fricatives rather than stops. This difference in distribution suggests that these Southwest Spanish labiodentals do not belong to the class of hypercorrect [v]. Furthermore, an account of hypercorrection based on orthographic representation would be incongruent with the fact that Southwest Spanish varieties are overwhelmingly oral.

This brings us back to the two most plausible candidates for New Mexican Spanish labiodentals: archaic [v] and language-contact [v].

2. Predictions and Methods. Language contact phenomena are usually evaluated by showing some kind of correlation between the linguistic variant of interest and extralinguistic factors related to parameters of bilingualism. For example, Silva-Corvalán (1994) uses generational distance from immigrant monolingualism in describing broad tendencies of change in verbal categories in the Spanish spoken in Los Angeles. Generational groupings of this kind, however, would be inappropriate for New Mexican Spanish, which has been spoken in the region in relative isolation for over three hundred years, although there is increasing contact with immigrant varieties. The Spanish spoken in Northern New Mexico and southern Colorado, with features dating to the late 16th century, differs noticeably from the more modern Spanish varieties spoken in other parts of the Southwest (Lipski 1994:281, cf. Caufield 1981:80-1, Espinosa 1930). For the present study, then, we operationalized degree of contact with English by considering the independent variables of age, Spanish proficiency and use, and formal instruction in Spanish (cf. Bernal-Enríquez 2000). If voiced labiodentals represent a contact-induced change, we would expect a higher rate of occurrence among younger and less proficient speakers. Later, in Table 2, we present a summary of the distribution of speakers by these social factors.

Another way to test the hypothesis that labiodentals in New Mexican Spanish are the result of contact with English is to look at their occurrence in cognates, that is, Spanish words with an English translation-counterpart similar in orthographic representation and phonetic shape. If labiodentals occur more frequently in cognates than in non-cognates, we have support for the hypothesis that there is an active process of transfer from English to Spanish in the pronunciation of [v] among bilinguals. Thus, in addition to orthographic representations ([v] and [b]), we considered the cognate status of the word and whether the English cognate has a labiodental [v] or a bilabial [b] phoneme.

Finally, we considered the token frequency of the word in which the variable occurs as a possible conditioning factor. A growing body of usage-based functionalist work demonstrates that linguistic structure is emergent from patterns of language use (e.g. Givón 1979, Hopper 1987). In a usage-based model of phonology, Bybee (1995, 2001) argues that the basic unit of mental storage is the word, defined as a processing unit based on categorizations of actual tokens. In this model, words and frequent constructions or phrases are the domain of application for sound change.

As in morphosyntactic variation and change (Bybee & Thompson 2000), Bybee (2001) shows that token frequency—the frequency of occurrence in running text—
has two very different effects in phonology. On the one hand, high frequency words undergo articulatorily motivated changes more rapidly than low frequency words. This type of change is the most common and involves minimization or reduction. On the other hand, high frequency items are more resistant to non-articulatorily motivated change, that is, analogical change based on the analysis of other forms. This is an interference effect, as frequency of use results in a high level of lexical strength and resistance to general changes occurring elsewhere in the language. For example, while weper and other English past tense forms tend to regularize (weepsed), high frequency words such as kingkiss do not (Bybee 1985, 1995). Of particular interest for our purpose is the high frequency lexical diffusion pattern, which can provide a way to determine the mechanisms of change (Bybee 2001). B. Phillips (1984, 2001) formulates for Frequency-Implementation Hypothesis, given in 3.

(3) Changes that ensue analysis—whether syntactic, morphological, or phonological—during their implementation affect the least frequent words first; others affect the most frequent words first. This is because frequent words are processed as automatized units, while less frequent words are more likely to be analyzed into constituent morphemes and phonemes.

In the case of the New Mexican Spanish-bilingual variation studied here, the changes does not seem to have as articulatory motivation, whether the direction of change is from monolingual to bilingual licit or vice versa. It would not be an assimilationary, since both varieties occur in intervocalic position. Nor does it appear to be recessive. While these features could be viewed as a weakening with respect to the English reflexes that werejawsoreb and lower lip is lost, they are typologically more part, at least in Romance (e.g. Portuguese, Catalan, French, Italian have [v] rather than [b]). The change, then, should pattern like non-articulatorily motivated changes, that is, it should affect high frequency words more than high frequency items. Thus, we can formulate the two sets of predictions about high frequency words in New Mexican Spanish given in 4.

(4) a. If the change is from standard Spanish bilinguals to contact-induced Spanish, we would expect more monolinguals in low-frequency words.

b. If the change involves loss of (archaic) Spanish bilinguals, we would expect more monolinguals in high frequency words.

The prediction in 4a is based on our proposal that change due to language contact is different in lexical diffusion pattern than regular sound changes resulting from articulatorily reduction. Contact-induced change differs through the lexicon like changes requiring minimal articulatory changes or targets. Low frequency words are assumed to be retained for the reason that the structure of bilingual speakers are more likely to analyze these words at some level, as being similar or different from those in the other language.
We obtained data by asking speakers to translate English words or phrases, the technique used by Mrz (1980) and R. Phillips (1974). This was a way to avoid having speakers read words, as in Timm (1976). Although we would have preferred to use recordings of sociolinguistic interviews, we found it difficult to distinguish a labiodental from a bilabial fricative acoustically. Therefore, we watched speakers' mouths and used visual criteria to decide which variant had been produced. For nearly all interviews, two colleagues evaluated the variant. We counted only cases where both raters' evaluations coincided.

A total of 599 tokens from 18 speakers were coded. The average occurrence of labiodentals was 61%. Bilabial fricatives occurred in 30% and bilabial stops in 9% of all tokens. These tokens were submitted to variable rule analysis using VARRULS (Piotrowski 1988), a type of multivariate analysis that considers factors simultaneously and selects those that contribute a statistically significant effect to the choice of variants. While individual speaker differences were found to be significant for all words, cognate status and orthographic representation turned out to be significant conditioning factors in low frequency words only. We first discuss social variables and then return to frequency effects.

3. SPEAKER DIFFERENCES AND EXTRANLICITARY FACTORS. Table 1 presents labiodental frequencies and VARRUL weights by speaker. Individual frequencies of \( v \) range from 31% to 94% and weights range from .10 to .92. In this study VARRUL weights above .5 are interpreted as favoring the occurrence of the labiodental variant, while figures below .5 reflect the opposite.

Table 2 shows the distribution of speakers and frequencies of \( v \) by groupings of age, proficiency in Spanish, level of Spanish use, and formal Spanish instruction. We are cautious about drawing firm conclusions from these results because the sample is small and biased toward speakers who prefer to use English. With one exception, all speakers indicated that they speak English as well as, or better than, Spanish. Nevertheless, certain interesting patterns emerge with respect to the role played by formal instruction and contact with other varieties of Spanish.

Let us look first at the average occurrence of \( v \) by age grouping. There are ten speakers in the 18-25 group, five in the 40-55 group, and three in the over 65 years old group. The \( v \) frequencies are 57% for the younger age group, 69% for the middle age group, and 65% for the older age group. The only statistically significant difference found was that between the younger and the middle age group.

*The number of actual tokens is about 25% less than the possible number (18 speakers x 48 words = 864), partly because of non-responses and partly because of indiscriminate variants. We would like to thank Robin Fettes, Antonio G. Sempere, and Devra Jenkins for help in coding the variants.

**The difference between the younger and middle age group in Table 2 is significant at \( p < .05 \) but not at \( p < .01 \), which is probably a more acceptable level of significance given the small sample size.
the low frequency, non-cognate, orthographic b-cell. Surprisingly, all speakers who pronounced this word did so with a bilateral fricative. In contrast, all other words in this cell had more than half their tokens with a labiodental. Caballa ‘onion’ and cahorro ‘sheep’ averaged 51% [v], while mid-high frequency words meeting the same criteria nuebe ‘cloud’, caballo ‘horse’, and aborno ‘open-Adj.’ averaged 71% [v]. We explain this unexpected behavior of rubio by its status as a learned word in New Mexican Spanish, that is, a word learned in Spanish class. Indeed, most speakers translated blond as gisoro, the preferred lexical variant in the region. Another factor that seems to disfavor labiodentals is contact with other varieties of Spanish. The lowest individual [v] frequency is 31%, for speakers m., j. and d (see Table 1). While m. and j. have studied Spanish at the university level, we cannot attribute speaker d’s low [v] frequency to formal Spanish instruction. This speaker has far fewer labiodentals than his grandmother, speaker f, whose [v] average is 61%. A combination of circumstances might have contributed to d’s low [v] frequency: that he grew up in the urban center of Albuquerque, where immigrant Spanish varieties are more prominent than in other parts of northern New Mexico; that he was in contact with other varieties of Spanish in the Army; and that he mostly speaks Spanish with Mexican coworkers. Taken together, these biographical facts point to early and intense contact with non-local varieties. The case of speaker e, who has the lowest next [v] frequency at 33%, points in the same direction. This speaker told us that she mainly speaks Spanish with her Mexican cousins (on her father’s side—her mother is New Mexican).

In summary, several extralinguistic factors seem to interact in the occurrence of labiodentals. No significant differences based on age or proficiency groupings were found. On the other hand, speakers who have received formal instruction in Spanish present a lower frequency of labiodentals. These results suggest that early or intense contact with non-native varieties of Spanish, whether standard varieties in school or immigrant varieties among peers, reduces the occurrence of labiodentals.

4. DIFFERENCES BETWEEN HIGH AND LOW FREQUENCY WORDS. A comparison of average [v] frequencies verifies that there is a lexical frequency effect on labiodental bilabial variation, as predicted by the theory of gradual lexical diffusion (cf. Bybee 2001). As shown in Table 3, high frequency words overall average 73% labiodentals, while low frequency words average 40% labiodentals or have taken classes at the high-school level only, with a combined [v] frequency of 66%, α2 (1, N = 467) = 29.06, p < .000. The higher proportion of labiodentals in entrenched high frequency items supports the hypothesis that at least some voiced labiodentals are archaïc [v], or retentions of a dialect feature of New Mexican Spanish.

For the VARPURL analysis, three independent variables or factor groups were considered: individual speaker differences (discussed in Section 3), word frequency, and orthographic representation of cognates and non-cognates. For cognates, two possibilities were distinguished: cases in which orthographic represen-
LEXICAL FREQUENCY AND VOiced LABIODENTAL

analyses, for high frequency words apart, and for low and mid-level frequency words apart. The results for high frequency words are summarized in Table 4. The only factor group selected was that of individual speakers differences; cognate status—orthography was not. Chi-square tests confirmed that the difference in the proportion of labiodentals between words with an orthographic b and words with an orthographic h is not significant. This is in stark contrast with labiodentals in Mexico and among Spanish-language instructors in the U.S., which overwhelmingly correspond to orthographic v (Lepe Blanch 1988:164-5, Stevenson 2000:142).

Nor are there significant differences between cognates of any kind and non-cognates. In short, neither orthography nor cognate status is a significant factor in the occurrence of labiodentals in high frequency words.

<table>
<thead>
<tr>
<th>COGNATE STATUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognate, Eng. v &amp; Span. v</td>
</tr>
<tr>
<td>Cognate, Eng. v &amp; Span. b</td>
</tr>
<tr>
<td>Non-cognate, Spanish v</td>
</tr>
<tr>
<td>Non-cognate, Spanish b</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NUMBER</th>
<th>PERCENT (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>47</td>
<td>81%</td>
</tr>
<tr>
<td>33</td>
<td>73%</td>
</tr>
<tr>
<td>75</td>
<td>72%</td>
</tr>
<tr>
<td>54</td>
<td>67%</td>
</tr>
</tbody>
</table>

ORTHOGRAPHY

- Orthographic v
- Orthographic h

<table>
<thead>
<tr>
<th>NUMBER</th>
<th>PERCENT (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>122</td>
<td>75%</td>
</tr>
<tr>
<td>87</td>
<td>69%</td>
</tr>
</tbody>
</table>

None of the differences is significant.

Table 4. Frequencies of Labiodentals in High Frequency Words
By Cognate Status and Orthography

In contrast, cognate status and orthographic representation turns out to be a significant constraining factor in mid and low frequency words, as shown in Table 5. Labiodentals are most likely to appear in cognates with an orthographic v in both languages (92) or non-cognates with a Spanish orthographic v (91). Labiodentals are highly unlikely in cognates with an orthographic h in both Spanish and English (33). However, labiodentals are about as equally likely to occur as bilabials in non-cognates with an orthographic b (53).

Table 6 confirms that the difference in the proportion of labiodentals between words with an orthographic v and words with an orthographic h is significant for low and mid frequency words. Both this result and the results for cognate status in Table 5 contrast with our findings for high frequency words, where neither orthography nor cognate status were significant factors (Table 4).

In summary, we have found that labiodentals occur more in high than low frequency words and do so independently of orthography or cognate status. At the

*Thanks are due to John Bergen for coming up with these.
TABLE 5. Voiced Labiodental-Bilabial Variation in Low and Mid Frequency Words

<table>
<thead>
<tr>
<th>Orthography</th>
<th>N</th>
<th>Percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Orthographic v</td>
<td>167</td>
<td>84%</td>
</tr>
<tr>
<td>Orthographic b</td>
<td>223</td>
<td>32%</td>
</tr>
</tbody>
</table>

\(x^2(1, N = 390) = 102.26, p = .000\)

The high frequency of labiodentals in highly entrenched, high frequency words for both orthographic \(v\) and orthographic \(b\), and independent of cognate status, points to an archaic \(v\) as in varieties of Judeo-Spanish. High frequency words such as habla ‘there was/there were’ were written with a \(v\) or a \(b\) in Old Spanish texts and colonial Mexican texts (Lope Blanck 1985-66). It is likely that these words spread to other words, regardless of etymology. This seems to have occurred in Bucharest Judeo-Spanish, where \(v\) has extended to cases of intervocalic \(b\) originating in Latin \(p\), for example escorcer [es'corcir] (Sala 1971). In our sample, the same word escorcer was pronounced with a labiodental in 70% of all tokens. Although it is likely that contact with English has favored \(v\), its origin in New Mexico appears to be the Old Spanish labiodental.

Two additional sets of facts from this area support the hypothesis that New Mexico labiodentals are a dialect feature originating in Spanish. First, labiodentals occurred in traditional New Mexican forms. All three speakers who used the archaistic form vide ‘law’, which has been transmitted in New Mexican Spanish since before English was ever heard in the region, pronounced it with a labiodental. Overall (as wide) presented a \(v\) frequency of 87%. Similarly, the form sabore ‘taste, flavor’ with the paragogic \(v\) typical of New Mexican Spanish (Hernández Cubillos 1990) was pronounced with a labiodental by the speaker who used it. This word had an overall \(v\) frequency of 71%.

A second, initially unexpected, set of facts also supports our hypothesis. Labiodentals were high in certain low and mid-level frequency words written with a \(b\), such as caballo ‘horse’, with 69% \(v\), and nube ‘cloud’, with 87% \(v\). This is the opposite of the results for ruido ‘noisy’, which was assigned to the same non-cognate, orthographic \(b\) cell and which had no labiodental tokens at all. Why might this be? We suggest that certain nouns for everyday things, such as caballo, may be of relatively higher frequency in child language than in adult language as reflected in frequency dictionaries. Words learned from grandma would be highly entrenched, especially so in cases of incomplete acquisition where speakers never acquire adult varieties and vocabulary remains limited to those early language words. One speaker told us that the only Spanish she spoke as a child were single words referring to animals and other commonplace objects. She subsequently took Spanish at the university. Her overall \(v\) frequency was 45%, which is lower than the average (speaker \(k\), Table 1), yet she used \(v\) for all nouns in this cell (nube, sabana, caballo, and caballo). The one exception was school-learned ruido (see Section 3). What about the language-contact \(v\)? In low frequency words, the variation between labiodentals and bilabials follows the pattern of English cognates. As we saw in Table 5, labiodentals are most likely to occur in cognates with an English \(b\) and bilabials are most favored in cognates with an English \(v\).

Let us take a closer look at the bilabial variants. Despite the fact that these are said not to occur in intervocalic position, as in example (1), there was a total of 55 such cases (9% of all tokens). Table 7 shows that stops are most favored in cognates with an English orthographic \(b\) (96%) and somewhat less favored in non-cognates.
nates with a Spanish orthographic b (64). There were no stop occurrences at all in cognates with an English v. Overall, 69% of these bilabial stops occurred in cognates with an English /v/, such as abuso, habitantes, atributos. These results suggest a change in progress (at least among some speakers) to the detriment of labiodentals, toward a sound-spelling correspondence under English influence. A It is important, however, that 80% of all stop tokens occurred in low frequency words and that there were no occurrences at all in high frequency words. This distribution pattern further supports the hypothesis that English-to-Spanish transfer is limited to low frequency words.

<table>
<thead>
<tr>
<th>FACTOR GROUP</th>
<th>FACTOR</th>
<th>N (%)</th>
<th>WEIGHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognate Status</td>
<td>Orthography</td>
<td>Cognate, Eng. v &amp; Spnam. v</td>
<td>133</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cognate, Eng. b &amp; Spnam. b</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cognate, Eng. v &amp; Spnam. b</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-cognate, Spanish v</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Non-cognate, Spanish b</td>
<td>134</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word Frequency</td>
<td>High</td>
<td>0%</td>
<td>Not</td>
</tr>
<tr>
<td></td>
<td>Mid</td>
<td>8%</td>
<td>Selected</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>17%</td>
<td></td>
</tr>
</tbody>
</table>

N = 544, Input probability = .03, p = .000, Application value = bilabial stop [b]

TABLE 7: Bilabial Stop - Bilabial and Labiodental Fricative Variation

5. Conclusion. We have shown a lexical frequency effect on voiced labiodental-bilabial variation in New Mexican Spanish. Based on the differences between high and low token frequency words, we may distinguish two kinds of labiodentals: archaic [v] and language-contact [v]. Labiodentals in high frequency words represent the retention of an old dialect feature, while low frequency words show a tendency to follow English patterns of distribution of /b/ and /v/. More generally, we propose that change due to language contact has a lexical diffusion pattern like analogical changes. Unlike regular sound change, contact-induced change affects low frequency items first.

References


