Gender Assignment to English Words in the Spanish of Southern Arizona

Elise M. DuBord
Gender Assignment to English Words in the Spanish of Southern Arizona

Elise M. DuBord
The University of Arizona

Abstract
This paper examines the gender assignment of English words in the Spanish of Southern Arizona based on the categories of biological sex, phonological gender, and analogical gender. It is determined that biological sex is the greatest indicator of gender assignment, followed by phonological gender and lastly by analogical gender. There was a small (7.9%) proportion of variation in gender assignment to English words in the corpus that is attributed to a combination of words that are neither phonologically nor socially integrated into the Spanish lexicon and the linguistic insecurity of the participants.

1. Introduction
The situation of language contact in the Southwest provides a productive site for examining how lexical items from the English language are incorporated into grammatical structures of the Spanish language. This paper will examine Spanish gender assignment to English words within the Spanish dialect of Southern Arizona. Because gender is an inherent grammatical feature of all nouns in Spanish, that in order for English words to be incorporated in Spanish speech, gender assignment to English words is often necessary to adapt English lexical items into Spanish speech. This paper seeks to identify factors that influence gender assignment and examine variation in gender assignment within the speech community.

This paper will analyze gender assignment to English words in the speech of eighteen Mexican Americans from the Sonora and Arizona border region, most of which were conducted in Tucson, Arizona, based on taped interviews conducted by nine graduate students. I will first examine how gender assignment is used in Spanish words in dialects of the Southwest (García, 1998 and Chaston, 1996) as well as general theories about the assignment of gender in Spanish (Prado, 1982). Specifically related to the topic of gender assignment of loan words I will discuss several previous studies (Zamora, 1975; Barkin, 1980; Poplack, Pousada, & Sankoff, 1982; & Smead, 2000a) as well as define what criteria I used to identify English words. Using the framework proposed by Poplack, Pousada, and Sankoff (1982), I will classify types of words based on
physiological sex, phonological gender, and analogical gender.

Based on previous studies of gender assignment to loans, I expect to find that gender assignment of English words will be consistent for repetitions of the same words throughout the corpus as suggested by Poplack, Pousada, and Sankoff (1982) due to uniformity within the speech community. Based on Poplack et al. (1982), Smead (2000a), Barkin (1980), García (1998), and Chaston (1996), I expect to find that gender will be consistent with the biological sex of the noun in animate objects and for non-biologically based gender assignments there will likely be a greater correlation between gender assignment and phonological gender than analogical gender.

2. Gender in Southwest Spanish

In a study not unlike the current project, García (1998) examined gender markings in Spanish based on eleven interviews of bilingual Spanish speakers from South Texas, for whom Spanish was their home language although it may have not been their dominant language at the time of the interview. Interviewers were both native and non-native Spanish speakers and students at the University of Texas at San Antonio. The noun phrase was the unit of analysis, counting only tokens that were marked by determiners or modifiers that signaled gender. García found highly traditional gender agreement with 96% for feminine noun phrases and 94% for masculine noun phrases. This study responds to the claim that Spanish in the Southwest is going through a process of simplification that could lead to random gender assignment or the overgeneralization of one gender. García demonstrates that in fluent speakers of this dialect of Southwest Spanish, gender assignment has remained highly traditional.

Chaston (1996) also examines gender agreement in the Spanish of Texas in bilingual university students, finding that traditional gender assignment in Spanish is more probable in those that speak Spanish with their parents. The study found 96% standard agreement for masculine nouns and 89% for feminine nouns, yet standard agreement for individual participants varied from 75% to 100%. Chaston's study proposes a continuum of gender agreement based on the participants’ use of Spanish or English at home. He also suggests that words assigned non-traditional gender markers were likely learned outside of the home environment and likely to be English cognates (i.e. canal, elecciones). Also included in Chaston's study is the observation that in the 42 instances of code switching that appeared with a Spanish article, 40 were assigned masculine gender. Chaston suggests that these results could be due to a tendency to use the masculine when unsure of the assigned gender or a tendency to use the masculine with English words or new cognates. Based on García (1998) and Chaston (1996) we can conclude that gender agreement in Southwest Spanish continues to follow traditional patterns, yet individuals who use less Spanish in the home tend to employ greater non-standard gender assignment.
3. Previous Studies on Gender Assignment of Loanwords

In an early study of gender assignment to loanword, Zamora (1975) surveyed Puerto Rican bilinguals on the gender assignment of both commonly used English loans and words that are not integrated loans in the Puerto Rican dialect. Despite the fact that Zamora’s method was very artificial in its elicitation (participants were asked to use words in a sentence and then explain why they chose a particular gender), he concluded that the semantic translation of the loan determined gender more frequently than the phonological shape. Zamora’s study is problematic because it does not account for gender assignment where semantic gender and phonological gender overlap. Zamora also suggests that gender assignment varies according to education as a reflection of whether loans are acquired orally or through written text because of its effect on the phonological shape of the loan word (i.e. freezer vs. frisa).

Barkin’s (1980) study of gender assignment in loanwords is another early study of English loanwords in the Spanish language. She uses Haugen’s distinction of unassimilated, partly assimilated, and wholly assimilated to classify loanwords. She proposes that with loanwords in Spanish, gender becomes an “optional category to be used or eliminated at will, for no apparent reason” (106). Barkin seems to express disapproval of the use of loanwords because it indicates a lack of control over language usage. She explains that people with more monolingual contacts use fewer loanwords and the words they do use are more wholly assimilated. She also asserts that those speakers are more “conscious of gender as a category” (107). Barkin’s explanation of gender assignment of loan words seems to leave assignment to chance in unassimilated words and suggests that gender becomes an optional category as speakers increase the quantity of loanword and as their level of phonological integration decreases.

Prado (1982) explains through his non-empirical analysis of gender assignment to loanwords in Spanish that the masculine is almost always assigned to words that originate from non-Romance languages (i.e. English). The few exceptions he cites are loanwords that have both masculine and feminine forms (i.e. la troca and el troque). Prado explains that because the feminine gender serves as the marked form in Spanish (and he argues in all other Romance languages), the masculine is the unmarked or more neutral form.

In a comprehensive study of gender assignment to English loanwords, Poplack et al. (1982) examine gender assignment of English loans in both Puerto Rican Spanish (in New York) and Montreal French. They examine Puerto Rican parents with their children to test for generational differences based on 300 hours of taped speech of 16 informants. The Montreal data was based on a computerized corpus of 120 speakers. The study was limited to single word loans, a distinction that is perhaps overly limited in its definition of loanwords. Their results demonstrate that the phonology of the loanword is more significant in determining gender in Puerto Rican Spanish than in Montreal French, which reflects monolingual language patterns. In both instances assignment is generally uniform across the speech community. The
authors explain that this factor is more influential than bilingual ability as they had originally set out to determine using different generations as a representation of different stages of language attrition. Contrary to Barkin (1980), Poplack et al. found that Spanish syntax rules determined the expressions of gender assignment rather than the phonological integration of the loan. They also conclude that there is no clear evidence for the masculine as the unmarked gender of loans, citing the lack of feminine phonological endings such as -a in English loans.

Smead (2000a) conducts a study of loanwords in two dictionaries of Spanish in the Southwest that distinguishes between a general Chicano Spanish dialect and the dialect of New Mexico and Southern Colorado. His very thorough literature review discusses in depth several previous studies on gender assignment in Spanish as well as gender assignment to loan words. An obvious deficiency of his study is the applicability of his findings based on data from dictionaries rather than actual spoken language. His findings are similar to Poplack et al. (1982), yet he adds the distinction of terminal morphemes versus terminal phonemes as well as the distinction between synonymic and hyperonymic gender. He concludes that in addition to loans with biological sex, morphological and phonological composition overwhelmingly determine gender assignment.

4. Methods

4.1 Participant Sample

The sample in this study consisted of eighteen Mexican-American speakers of Spanish in Southern Arizona based on interviews conducted by nine graduate students (1 male and 8 female) in February and March of 2004. Each interviewer conducted interviews with two participants; 7 interviews were conducted individually and 11 were conducted with interviewees and/or others present. Each interview lasted approximately 45 minutes for individual interviews and up to 90 minutes for group interviews, with a total of approximately 25 hours of interviews. The interviews were conducted mainly in Spanish with different degrees of code-switching and English as determined by the interviewees. The relationship between interviewer and interviewees ranged from student/teacher, acquaintances, and friends. Two interviewers were native Spanish speakers and seven were non-native Spanish speakers. Each interviewer transcribed the recording of his or her own interviews. The participants (nine male and nine female) ranged in age from 18 to 56 with an average age of 33. The participants' language proficiency can be classified as: 4 Spanish monolingual, 3 Spanish dominant, 6 balanced bilingual, and 5 English dominant.

4.2 English nouns

The unit of analysis was the noun phrase that was characterized by having an English noun(s) and Spanish determiners and/or modifiers that signaled gender assignment to English words. There were 242 noun phrases identified, 68 of which were repetitions of the same noun. I therefore based my analysis on 174 tokens, counting repetitions as separate tokens only when the same nouns were assigned both male and female gender in different noun phrases (11 nouns accounting for 22 tokens). In addition, one token was assigned both male and female gender within a single noun phrase. I eliminated loan translation such as: los grados (academic grades), las aplicaciones (applications), los periodos (class periods) because
they constitute what Smead (2000b: 162) terms “calquewords” where a Spanish word takes on the meaning of an English word, displacing the original Spanish meaning. Because calquewords consistently followed traditional Spanish gender assignment in the corpus of this study, they were not considered. There is one exception of a calqueword with variable gender assignment that will be discussed in section 6. All tokens were collected from semi-spontaneous speech in interviews following the model of Poplack et al. (1982) and Blas Arroyo and Tricker (2000) rather than relying on elicited responses or written texts as other previous studies have done (see for example, Zamora, 1975; Barken, 1980; Bonfield, 1994; Sánchez, 1995; Smead, 2000a; & Callahan, 2002).

My data includes all English words that were assigned gender in the corpus regardless of whether it was a proper noun (Desert View) or not (shirt). I did not consider the phonological integration of the English words as in Barkin (1980) due to the lack of phonological information in the transcription. I also did not consider social integration that could possibly be measured through repetition of tokens due to the limited number of total tokens in the corpus.

4.3 Coding

The corpus was first divided into three categories for purposes of analysis: nouns with biological sex, proper nouns without biological sex, and the remaining nouns were categorized as general nouns. Examples of each grouping are listed below:

**Group 1.** Biological sex (country boys, tenant, Will Smith, Barney)

**Group 2.** Proper nouns (Grande Tortilla Factory, Lincoln, Ruby Road, Cactus Grill)

**Group 3.** General nouns (recess, shirt, psychology class, baggy pants)

Gender assignment of the noun was determined by Spanish determiners and/or modifiers of the English nouns. Table 1 summarizes gender assignment according to the category of noun and the total distribution of gender for all tokens.

```
<table>
<thead>
<tr>
<th>Nouns with biological sex</th>
<th>28 masculine</th>
<th>20%</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 feminine</td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td>35 total</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Proper nouns without biological sex</td>
<td>27 masculine</td>
<td>60.00%</td>
</tr>
<tr>
<td>17 feminine</td>
<td>37.83%</td>
<td></td>
</tr>
<tr>
<td>1 both</td>
<td>2.2%</td>
<td></td>
</tr>
<tr>
<td>45 total</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>General nouns</td>
<td>75 masculine</td>
<td>79.83%</td>
</tr>
<tr>
<td>19 feminine</td>
<td>20.2%</td>
<td></td>
</tr>
<tr>
<td>94 total</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Total English words</td>
<td>130 masculine</td>
<td>74.7%</td>
</tr>
<tr>
<td>43 feminine</td>
<td>24.7%</td>
<td></td>
</tr>
<tr>
<td>1 both</td>
<td>0.6%</td>
<td></td>
</tr>
<tr>
<td>174 total</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>
```

**Table 1**

Gender assignment to English nouns
5. Factors for Gender Assignment

Based on Poplack et al. (1982), three categories were identified as possible factors that determine gender assignment to English words. Biological sex, phonological gender and analogical gender were identified for each English word as masculine, feminine, or neutral.

5.1 Biological sex

Biological sex was coded as either masculine (M) or feminine (F) according to the sex of the person or animal represented by the noun as exemplified in the following list.

Batman – M  
sophomores – M  
Miss Ross – F  
baby – F (referring to a female)

Both proper and general nouns were included in this category of analysis. Thirty-five tokens were considered in the analysis of biological sex. As Table 2 demonstrates, masculine gender assignment is highly standardized for biologically masculine nouns (96.2%). The one token that was assigned feminine gender despite masculine biological gender, was first assigned feminine gender and then later assigned masculine gender by the same speaker (la Ashton Kutcher, el Ashton Kutcher). There were significantly fewer total tokens of biologically feminine nouns with 75% receiving feminine gender assignment. The two exceptions in this instance were el baby and el Bachelorette. According to the context of the token, baby referred to a female, yet standard gender assignment to bebé in Spanish is masculine, regardless of gender, although la bebé is used in several dialects to refer to a female baby. The inclusion of el Bachelorette (the name of a television program) in the category of biological sex is debatable because it could be categorized as having analogical gender based on an analogical translation based on show (el programa), rather than bachelorette (la soltera). This token was included based on the most transparent syntactical meaning of bachelorette. Because of the ambiguity of the source of gender assignment in the example of el Bachelorette and constraints of the analogical translation of el baby, it is likely that these two tokens are not representative of feminine gender assignment according to biological sex.

<table>
<thead>
<tr>
<th>Biologically masculine</th>
<th>Masculine assignment</th>
<th>Feminine assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>26/27</td>
<td>1/27</td>
</tr>
<tr>
<td></td>
<td>96.2% masc.</td>
<td>3.8% fem.</td>
</tr>
<tr>
<td>Biologically feminine</td>
<td>28</td>
<td>6/8</td>
</tr>
<tr>
<td></td>
<td>25% masc.</td>
<td>75% fem.</td>
</tr>
</tbody>
</table>

Table 2

Gender assignment according to biological sex

5.2 Phonological and analogical gender

All tokens were also coded according to the phonological and analogical gender of the English
word. This included proper nouns and general nouns, while eliminating nouns from the category of biological sex.

5.2.1 Phonological gender

Phonological gender was determined according to the phonological shape of the word ending based on Smead (2000a) and Teschner and Alatorre's (1984) classifications of Spanish gender based on word endings. Words ending orthographically or phonetically in -l, -o, -n, -e, -r, and -s were coded as masculine (M). Words ending in -a, -d, -ion, -is, were coded as feminine (F). Words ending in –z, which are typically divided between masculine (40%) and feminine gender (60%) in Spanish (Teschner, 1983: 255), were not encountered in the corpus of English words. Words that did not correspond with the ending mentioned above were coded as phonetically neutral (N). The following list demonstrates examples of phonological coding of English words from the corpus.

dishwasher –M
parquiadero – M
Lincoln – M
beads – F
troca – F
Pima – F
project – N
truck – N
Spanish – N

Table 3 shows the distribution of masculine and feminine gender assignment according to the phonological shape of the English word. English words were first analyzed separately in the category of proper nouns (group 2) and general nouns (group 3) and then together for all proper and general nouns without biological sex. For proper nouns there was a higher percentage of feminine assignment across phonologically masculine, feminine and neutral words as compared to general nouns, but these finding are based on significantly lower numbers of tokens. When the two groups are analyzed together, phonologically masculine words are assigned masculine gender in 82.3% of tokens, phonologically feminine words are assigned feminine gender in 38.9 of tokens, and phonologically neutral words are assigned 70.7% masculine and 29.3% feminine. In all categories masculine gender assignment prevails in the following order according to greatest percentage of masculine assignment first: phonologically masculine, phonologically neutral, and phonologically feminine.
Analogical gender was coded based on Spanish translation of the English word as masculine (M), feminine (F), or neutral (N) when a semantic equivalent could not be identified. For proper nouns without biological sex, the analogical gender was coded according to the corresponding general noun (i.e. *Panda* was coded as analogically masculine because it refers to a restaurant, *restaurante* is masculine). A native Spanish speaker from northern Mexico confirmed the translation of the English words to assure that the analogical gender was determined by the word that most closely corresponded with the semantic translation of the English word. The following list provides several English nouns with their semantic translations and resulting analogical gender.

- **mall** (*el centro comercial*) M
- **cell phone** (*el celular*) M
- **Safeway** (name of supermarket – *el supermercado*) M
- **shirt** (*la camisa*) F
- **high school** (*la preparatoria*) F
- **Salpointe** (name of school – *la escuela*) F
- **spring break** (no equivalent term) N

As suggested by Rodríguez (1986), gender assignment to acronyms in Spanish tend to be determined by the principle name in the phrase, and acronyms from other languages tend to be determined by the gender of the equivalent translation of the principle name in the foreign language.
(English in our study). The one token that is an acronym in our corpus *la ASDB*, whose complete name is *Arizona School of the Deaf and the Blind*, follows Rodríguez’s rule for gender assignment using *School* as the principle word whose analogical equivalent is feminine in Spanish (*la escuela*).

In Table 4 the distribution of masculine and feminine assignment according to analogical gender is provided for general nouns and proper nouns and their gender. The distribution between masculine and feminine assignment is remarkably similar across all categories when comparing analogical gender with the distribution of phonological gender from the previous section.

<table>
<thead>
<tr>
<th></th>
<th>Masculine assignment</th>
<th>Feminine assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proper nouns</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analogically masculine</td>
<td>16/22 72.7% masc.</td>
<td>6/22 27.3% fem.</td>
</tr>
<tr>
<td>Analogically feminine</td>
<td>11/22 50.0% masc.</td>
<td>11/22 50.0% fem.</td>
</tr>
<tr>
<td><strong>General English nouns</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analogically masculine</td>
<td>49/57 86.0% masc.</td>
<td>8/57 14.0% fem.</td>
</tr>
<tr>
<td>Analogically feminine</td>
<td>22/33 66.7% masc.</td>
<td>11/33 33.3% fem.</td>
</tr>
<tr>
<td>Analogically neutral</td>
<td>4/4 100% masc.</td>
<td>0/4 0% fem.</td>
</tr>
<tr>
<td><strong>Proper and General</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analogically masculine</td>
<td>65/79 82.3% masc.</td>
<td>14/79 17.7% fem.</td>
</tr>
<tr>
<td>Analogically feminine</td>
<td>33/55 60.0% masc.</td>
<td>22/55 40.0% fem.</td>
</tr>
<tr>
<td>Analogically neutral</td>
<td>4/4 100% masc.</td>
<td>0/4 0% fem.</td>
</tr>
</tbody>
</table>

Table 4

Gender assignment according to analogical gender

Due to the minimal distinction between the distribution of gender assignment according to phonological and analogical gender it is necessary to compare the two factors together.

**5.2.3 Phonological gender vs. analogical gender**

In order to determine if phonological or analogical gender was a greater indicator of masculine or feminine gender assignment for English nouns (having first removed the category of biological sex), nouns were grouped in six categories: 1) masculine phonologically and analogically, 2) masculine phonologically and feminine analogically, 3) feminine phonologically and analogically, 4) feminine phonologically and masculine analogically, 5) neutral phonologically and masculine analogically, and 6) neutral phonologically and feminine analogically. Words that were analogically neutral were removed from the sample because there were only four tokens and all were assigned masculine gender. Words in the six categories were identified as either being assigned masculine or feminine gender (and both in one case). Table 5 reports the distribution between masculine and
feminine gender assignment for each of the six categories. An example from the corpus is shown in italics for both masculine and feminine gender assignment in each category.

Table 5

General and Proper Nouns (Phonological and Analogical Gender)

Graph 1 illustrates the percentage of gender assignment for each of the six categories in the cross tabulation of phonological and analogical gender. The number of tokens in each category is listed below the name of the category.

Graph 1

Distribution of masculine and feminine gender assignment according to phonological and analogical gender
Ph=Phonological gender
An=Analogical gender
M=Masculine
F=Feminine
N=Neutral

The first category of M(Ph)/M(An) has the greatest proportion of masculine gender assignment (89.7%) as was expected. The second and third categories of M(Ph)/F(An) and F(Ph)/F(An) were closely related statistically (68.2% masculine and 31.8% feminine vs. 66.7% masculine and 33.3% feminine), although it should be noted that the second category had more than double the quantity of tokens found in the third category. This suggests that phonological shape has a lower effect on determining gender when the word is analogically feminine. The comparison that is most interesting for our analysis is that of category two M(Ph)/F(An) and category four F(Ph)/M(An) because phonological and analogical genders are oppositional. Category two demonstrates a higher distribution of masculine gender than category four, and category four has a higher distribution of feminine gender than category two, suggesting that phonological gender is a greater indicator of gender assignment when phonological and analogical gender do not coincide. When phonological shape is not applicable (phonologically neutral), gender assignment is largely determined by gender (78.1% masculine assignment for phonologically masculine words and 45.8% feminine for phonologically feminine words). We must be cautious in interpreting the implications of this data due to the low number of tokens in each category, especially for words that were phonologically feminine with only nine words in each category.

6. Variation in Gender Assignment

Poplack et al. (1982: 26) suggest that gender assignment is consistent within the speech community, with only 4% of their sample characterized by vacillation between forms. They conclude that bilingual norms established in the speech community are more influential than bilingual ability in the gender assignment to loanwords because once gender is assigned to a loanword, it is usually uniform between speakers. This data from the current analysis includes 152 total English words with gender assigned, twelve of which showed variation in gender assignment. This accounted for 7.9% of the corpus. Nouns that appeared with gender variation are listed below.

Ashton Kutcher
beads
club
drive-in
fake IDs
high school
middle school
ordenes
Pima (name of community collage)
Pin
Safeway (name of supermarket)
Southwest Supermarket

For variation in gender assignment in our corpus, one English dominant participant used both variables within his own speech for three words (Ashton Kutcher, beads, and Southwest Supermarket) and two of his gender assignments contrasted with other participants (club and fake IDs). The remaining variation was divided between bilinguals and English dominant speakers. No Spanish dominant participants varied in gender assignment to English words in their own speech or with other participants. The one surprising exception came from a monolingual Spanish speaker in his use of a calqueword (órdenes- meaning orders, pedidos or encargados in Spanish) that was included as a token because it appeared with both masculine and feminine gender assignment (las órdenes and los órdenes).

7. Conclusions

Biological sex is the most significant factor for determining gender assignment of English words. For nouns without biological sex, analogical and phonological genders have the same influence on gender assignment when analyzed separately. When phonological and analogical genders are analyzed together, phonological gender has a greater influence on gender assignment than analogical gender. These results confirm the findings of both Poplack et al. (1982) and Smead (2000a) who demonstrated that after biological sex is removed, phonological shape is a greater determiner of gender assignment than analogical shape in Spanish gender assignment to English words.

Variation in gender assignment of English words appears to characterize a small proportion of English words. Variation is generally found in unestablished loans and proper nouns with a few exceptions (high school and club). Variation appears to be greater for English dominant and bilingual individuals who use Spanish less frequently than other participants in the study. This variation could be a possible indicator of linguistic insecurity or language attrition.

For future research, the present study would benefit from a larger corpus of data and greater specification of phonological integration of English words. This could lead to a more comprehensive analysis of the integration of English words into the Spanish lexicon. A larger corpus would allow also for more repetition of tokens that would lead to broader analysis of variation in gender assignment.

Notes:
1 I thank Cindy Ducar for the use of her categorization of the participants’ language proficiency.
Works Cited


