PHONOLOGY OF GENDER IN ENGLISH AND FRENCH GIVEN NAMES

Lisa Sullivan[1] & Yoonjung Kang[1,2]

[1]University of Toronto, [2]University of Toronto Scarborough
lisa.sullivan@mail.utoronto.ca, kang@utsc.utoronto.ca

ABSTRACT

Traditionally, the relationship between form and meaning in language has been considered to be arbitrary. However, researchers have begun to propose that sound symbolism is more pervasive than previously assumed; for example, the phonology of given names may encode information about gender. Corpus analyses of English and French given names confirm this, finding that there are phonological patterns correlated with gender in given names. If these patterns are sound symbolic, speakers should be able to make use of them to assign gender to unfamiliar names. A name gendering experiment had participants rate the gender of nonce names which were male- or female-biased according to one of six of the phonological factors found in the corpus analyses. Female-biased names were given more female ratings than male-biased names. This suggests speakers have productive knowledge of the correlations between gender and phonological patterns and that the patterns may be sound symbolic.

Keywords: Sound symbolism, English, French, names

1. INTRODUCTION

Traditionally, the relationship between form (sound) and meaning in language has been considered to be arbitrary [14, 5]. Exceptions to this rule, such as onomatopoeia and other forms of sound symbolism, exist; however, these were considered inconsequential. More recently, researchers have begun to propose that sound symbolism is more pervasive than previously assumed [9, 16]; for example, the phonology of given names may encode information about gender. Yet, phonologically-informed investigations are limited and those that exist focus primarily on English [3, 4, 7, 16, 17, 18, 22], though recent work has looked at Japanese [15], French [18], Urdu [8] and Cantonese [21].

Corpus analyses of English given names [4, 7, 16, 17, 18, 22] have shown that there are phonological patterns which are correlated with gender. Similar patterns have been found in French [18], Urdu [8] and Cantonese [21]. Some of the patterns from English and French can be found in Table 1.

Table 1: A selection of phonological patterns correlated with gender in French (FR) and English (EN) given names [18]

Factor	Lang	Pattern
Number of Syllables	EN, FR	Female names have more syllables than male names
Final Syllable	EN, FR	Female names end in an open syllable more often
Back Vowels	EN, FR	Male names have more back vowels
Sonorant Consonants	EN, FR	Female names have more (/l/,/m/,/n/)
Stress Placement	EN	Female names have more non-initial stress
Nasal Vowels	FR	Male names have more nasal vowels

Previous studies showed that some of the sound-gender connections are productively extended to novel or nonce names; for example, in the creation of nonce names, as Sutton [19] found in a corpus analysis of American super hero names, or in determining the gender of an unfamiliar name. Experimental work on English by Cassidy et al. [3] and MacAuley et al. [7] has shown effects of stress pattern, number of syllables, final sound type and presence of a final syllabic nasal (more common in male names [7]) in assigning gender to nonce names. Furthermore, Cassidy et al. [3] also found that phonologically typical real names were more likely to be assigned the "correct" gender than phonologically atypical names.

While these studies suggest that phonological patterns are available for speakers to use, they only look at patterns native to English and, thus, do not address the potentially cross-linguistic nature of some of these patterns or the degree of language specificity at which they are encoded.

In the current study, native English speakers provide ratings for how male or female they think nonce names are. While addressing the issue of the availability of the phonology-gender patterns for speaker use, it expands upon previous studies by looking at more phonological factors (see Table 1), including factors that are common to French and English, a factor found in English but not French and a factor found in French but not English. Additionally, speakers will hear nonce names spoken by both French and English speakers. These changes are designed to address the question of the language-specificity of English speakers' knowledge of sound-gender correlation. If these patterns are encoded at language-nonspecific level, speakers should be able to make use of them regardless of speaker language and extend them to non-native patterns.

2. METHOD

2.1. Participants

27 monolingual North American English speakers over 18 years of age with normal speech, hearing and vision completed the online experiment. Participants were recruited online and received a $5 Amazon.ca gift card as compensation. 9 participants were excluded from the analysis due to technical errors (2), failure to complete the experiment properly (1), failure to meet the language requirements (2) or failure to discriminate nasal and non-nasal vowels (4).

2.2. Stimuli

40 target stimuli were created which consisted of 20 minimal pairs of nonce names whose members differed from each other in terms of one of six target factors: number of syllables (4), final syllable type (4), presence of a back vowel (4), presence of a sonorant consonant (4), stress placement (2) and presence of a nasal vowel (2). These factors cover a range of types of phonological phenomena including word length, weight, vowel quality and consonant type. The language specific factors (stress placement, presence of a nasal vowel) were selected because they were not phonologically possible in the other language. See Table 1 for a description of how these factors pattern according to gender.

20 non-word CVCV sequences consisting of sounds found in both French and English were selected as the base of each minimal pair. Each sequence was modified by one of the phonological factors to create an alternate that was either male or female biased. Female-biased alternants were created for number of syllables (CVCV → CV[və]CV), consonant type (CV[s,z,b,f]V → CV[l]C) and stress placement (ˈCVCV → CVˈCV). Male-biased alternants were created for final syllable type (CVCV → CVCV[v]), presence of a back vowel (C[–back V]CV → C[+back V]CV) and presence of a nasal vowel (CVCV → CṼCV). In each pair, the unmodified alternant was considered to be biased in the opposite direction of the modified alternant (e.g. for number of syllables, CVCV is male biased and CV[və]CV is female biased). /v/ and /ə/ were selected for insertion because they are neutral in terms of gender bias.

The target stimuli were divided into two groups, one to represent English names and the other, French names. Target stimuli were divided equally across language and phonological factor, meaning four pairs were created for each factor common to both French and English and two pairs for the language-specific factors. 8 additional CVCV sequences were selected to serve as training items. These were divided into two groups, such that there were 4 training items for each language.

2.2.1. Stimulus Recording and Preparation

The stimuli were recorded by two phonetically trained female linguists who were native speakers of Canadian English and Canadian French, respectively. Stimuli were recorded using a Sound Devices 722 digital audio recorder and a DPA 4011 unidirectional cardioid microphone. The speakers were presented with the stimuli corresponding to their native language in IPA. They were asked to read each stimulus item clearly three times into the microphone.

The stimuli were prepared using Praat [2]. One of the three recordings were extracted from the audio file for each word. For each extracted word, approximately 500ms of silence was added before and after the word and the intensity of the word was adjusted to 70db. Each of these sounds was then saved as a wav file and converted to an mp3 file.

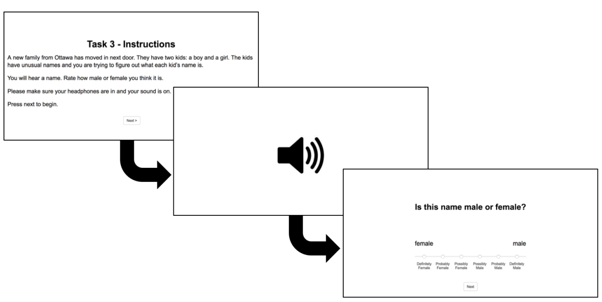
2.3. Procedure

The experiment was delivered to participants online using a script developed using jsPsych [6] and scripts developed by Hyoung Seok Kwon and Na-Young Ryu. Participants began by completing a language background questionnaire to assess their knowledge of and exposure to French, English and other languages. This was followed up by the main task, which was presented to the participant in two blocks (one for each language), each of which was followed by an AX discrimination task corresponding to the language-specific factor for that block (nasal vowel discrimination for French, stress placement discrimination for English). The order of the two blocks was counterbalanced across participants.

Each name gendering task block (Fig. 1) began with the instructions *A new family from Ottawa has moved in next door. They have two kids: a boy and a girl. The kids have unusual names and you are trying to figure out what each kid's name is. You will hear a name. Rate how male or female you think it is. Please make sure your headphones are in and your sound is on. Press next to begin.*

Each stimulus item was presented aurally. The participant heard the target name while an audio symbol was displayed on the screen. Once the sound was finished playing, they were presented with the question: *Is this name male or female?* and a 6-point Likert-type scale ranging from "definitely female" to "definitely male". Participants rated how male or female they thought the name was and clicked "Next" to proceed to the next trial. The first four items the participant heard were training items which were presented in a randomized order by participant. Following this, the participant heard the target items, which were also randomized by participant.

Figure 1: Name gendering task procedure

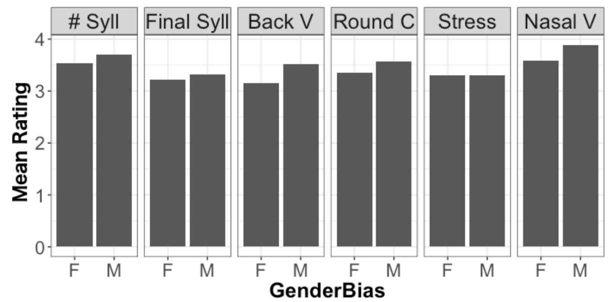


3. RESULTS

Ratings were converted to a 6-point scale where lower numbers (1-3) represented more female names and higher numbers (4-6) represented more male names. Overall, Participants rated female biased names (*M* = 3.34, *SD* = 1.17) as more female than male biased names (*M* = 3.54, *SD* = 1.21). This pattern holds for both French names (female: *M*=3.45, *SD*=1.12; male: *M*=3.72, *SD*=1.14) and English names (female: *M*=3.23, *SD*=1.21; male: *M*=3.36, *SD*=1.25), although French names were rated as more male than English names overall.

Female names were also rated more female for five of the six phonological factors (Fig. 2): number of syllables (female: *M*=3.54, *SD*=1.24; male: *M*=3.69, *SD*=1.12), final syllable type (female: *M*=3.22, *SD*=1.06; male: *M*=3.32, *SD*=1.32), presence of a back vowel (female: *M*=3.15, *SD*=1.18; male: *M*=3.51, *SD*=1.21), presence of a round con-

Figure 2: Mean ratings (rating range: 1-6) for each phonological factor separated by gender bias



sonant (female: *M*=3.47, *SD*=1.22; male: *M*=3.57, *SD*=1.25) and presence of a nasal vowel (female: *M*=3.58, *SD*=1.13; male: *M*=3.89, *SD*=1.01). For stress placement, female (*M* = 3.31, *SD* = 1.09) and male (*M* = 3.31, *SD* = 1.17) names had the same mean rating.

A mixed effects linear regression model was built to assess the effects of phonological gender bias, speaker language and phonological factor on participants' ratings of nonce names. The response variable was rating. The fixed effects were gender bias (male or female), phonological factor (the six phonological factors), language (French and English), and interactions between gender bias and language, and gender bias and phonological factor. Simple coding [20] was used for gender bias (female = -0.5, male = 0.5), phonological factor (6 levels) and language (English = -0.5, French = 0.5). The random factors were random intercepts for pair and participant and a random slope for gender bias by participant.

The model was run in R [13] using the lmer() function from the lme4 package [1]. There are significant effects of intercept ($\beta = 3.454$, $t = 21.970$, $p < 0.001$) and gender bias ($\beta = 0.190$, $t = 2.315$, $p = 0.0209$). There were no other significant effects. An ANOVA of the model shows no significant main effects of phonological factor ($F_{(5, 13)}=0.205$, $p=0.955$) or language($F_{(1, 13)}=0.751$, $p=0.402$) and no significant effects of the interactions between gender bias and phonological factor ($F_{(5, 676)}=0.332$, $p=0.894$) and gender bias and language ($F_{(1, 676)}=0.313$, $p=0.576$).

Model comparison by likelihood ratio test using the anova() function in R [13] was used to compare the model described above to a trimmed-down model with the same random effects, but with gender bias as the only fixed effect. The results indicate that there is no significant difference between the two models, thus the larger model does not give a significantly better fit than the trimmed-down one. The trimmed-down model can, therefore, be considered an appropriate fit for the data.

The trimmed down model has significant effects of intercept ($\beta = 3.440$, $t = 24.582$, $p < 0.001$) and gender bias ($\beta = 0.197$, $t = 2.542$, $p = 0.0113$). The significant effect of intercept indicates that ratings were, on average, more male than female. The positive coefficient for gender bias indicates that a male gender bias results in a more male rating.

4. DISCUSSION

The results indicate that gender bias, but not phonological factor or speaker language has a significant effect on participant rating of the gender of nonce names. These results pattern in the expected direction, with female-biased names receiving more female ratings than male-biased names. The results also indicate that there are no significant differences in the interaction between gender bias and phonological factor or speaker language, meaning ratings patterned the same way regardless of variance in language or factor.

Overall, these results suggest that speakers may have productive knowledge of these phonological patterns which they can make use of to assign gender to names. This confirms the findings of previous experimental work in which speakers used phonological patterns to assign gender to nonce and real English names [3, 7]. It also expands on this research, finding that speakers are able to use segmental, and not just suprasegmental, information to assign gender to nonce names.

Furthermore, the fact that the English and French language results pattern similarly, including the nasal vowel factor, which is only available in French, indicates that speakers may be able to apply the patterns to novel situations. This suggests that this sound-gender correlation may be available to speakers in a more abstract, language-nonspecific way. It is not clear, however, how speakers acquired these patterns. The fact that the French results pattern with the English ones allows for the possibility that the patterns are learned by a method other than exposure to a specific language. Rather, some of the correlations may have an iconic, universal basis [10, 11] and speakers' knowledge of these correlations may be independent of language-specific experience. However, as most of the factors are the same across languages, there is not sufficient evidence to support this hypothesis over the alternative that the patterns are learned from exposure to English or French.

The behaviour of participants on the presence of a nasal vowel factor is particularly interesting as this pattern is not present in English. One possible interpretation for this is that this is a cross-linguistic pattern which is available to all speakers; however, more research into the correlation between nasal vowels and gender in additional languages with phonemic nasal vowels would be necessary to confirm this. Another is that, despite being monolingual speakers, the participants, who are primarily Canadian, have had sufficient exposure to French to have learned this pattern. Testing an unfamiliar pattern from a language which participants have no exposure to could address this issue.

A final possibility is that participants were extending their knowledge of the relationship between syllable structure and gender in English to these names. Corpus analyses of English names [17, 18] indicate that male names have more closed syllables than female names. Some abstract analyses of French phonology (such as [12]) consider nasal vowel final syllables to be closed. Thus, changing an oral vowel to a nasal vowel, as was done for the nasal vowel stimuli, resulted in an initial closed syllable, which increased the proportion of closed syllables in the name, effectively making these names more male according to a phonological factor found in English. However, English doesn't have phonemic nasal vowels, so this would involve extending the syllable structure pattern to novel items. This possibility could be further addressed by testing participants on nonce word pairs in which a nasal vowel-final syllable alternates with a consonant-final syllable. Since these would both have the same syllable type, there should be no difference in rating if participants are extending the syllable structure pattern.

5. CONCLUSION

Corpus analyses have shown that there are phonological patterns correlated with gender in given names [4, 7, 8, 16, 17, 18, 22, 21]. Experimental results suggest these are available to speakers for use in assigning gender to unfamiliar names [3, 7]. The current study finds that speakers are able, not only to use these patterns to determine gender in names in their native language, but also to extend them to names in a non-native language. Furthermore, speakers were able to make use of a non-native language pattern, either because the pattern is cross-linguistic or because they were able to extend patterns from their native language to a novel situation. Speakers' ability to extend these patterns to novel situations suggests they are available in a language-nonspecific way. Future research should further explore the nature of the availability of phonology-gender patterns for speaker use.

6. REFERENCES

[1] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48.

[2] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5(9/10), 341–345.

[3] Cassidy, K. W., Kelly, M. H., Sharoni, L. J. 1999. Inferring gender from name phonology. *Journal of Experimental Psychology* 128(3), 362–381.

[4] Cutler, A., Mcqueen, J., Robinson, K. 1990. Elizabeth and john, sound patterns of men's and women's names. *Journal of Linguistics* 26(2), 471?482.

[5] Hockett, C. 1960. The origin of speech. *Scientific American* 203(3), 88–97.

[6] de Leeuw, J. R. 2015. jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods* 47(1), 1–12.

[7] MacAuley, L., Siddiqi, D., Toivonen, I. 2018. The age of aidans: Cognitive underpinnings of a new trend in english boys' names. *Names* 1–15.

[8] Mohsin, N., Y., K. 2018. Gender phonology of urdu first names. *Summer Phonology Forum 2018*. University of Toronto.

[9] Monaghan, P., Shillcock, R. C., Christiansen, M. H., Kirby, S. 2014. How arbitrary is language? *Phil. Trans. R. Soc. B* 369.

[10] Ohala, J. J. 1984. An ethological perspective on common cross-language utilization of f0 of voice. *Phonetica* 41(1), 1–16.

[11] Ohala, J. J. 2004. Sound symbolism.

[12] Paradis, C., Prunet, J. F. 2000. Nasal vowels as two segments: Evidence from borrowings. *Language* 324–357.

[13] R Core Team, 2016. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

[14] de Saussure, F. 1916. *Course in General Linguistics*.

[15] Shinohara, K., Kawahara, S. 2013. The sound symbolic nature of japanese maid names. *Proceeding of the 13th meeting of Japanese Cognitive Linguistics Association* 183–193.

[16] Sidhu, D. M., Pexman, P. M. 2015. What's in a name? sound symbolism and gender in first names. *PLoS ONE* 10(5), 1–22.

[17] Slater, A. S., Feinman, S. 1985. Gender and the phonology of north american first names. *Sex Roles* 13, 429–440.

[18] Sullivan, L. 2018. Phonolgoy of gender in french and english given names. Master's thesis University of Toronto.

[19] Sutton, L. 2016. Aliens are just like us: Personal names in the legion of super-heroes. *Names* 64(2), 109–119.

[20] UCLA Institute for Digital Research and Education, 2011. R library contrast coding systems for categorical variables.

[21] Wong, K., Kang, Y. in prep. Phonology of gender in cantonese first names. Manuscript in preparation.

[22] Wright, S. K., Hay, J., Bent, T. 2005. Ladies first? phonology, frequency, and the naming conspiracy. *Linguistics* 43(3), 531–561.