

# Gendered Associations of English Morphology

Jeremy M. Needle<sup>1</sup> and Janet B. Pierrehumbert<sup>2</sup>

<sup>1</sup>Department of Linguistics, Northwestern University, Evanston, IL, USA

<sup>2</sup>Department of Engineering Science, University of Oxford, Oxford, United Kingdom

Corresponding Author: Jeremy M. Needle ([jneedle@u.northwestern.edu](mailto:jneedle@u.northwestern.edu))

## Abstract

Morphological systems arise from language experience encoded in the lexicon, which includes much statistical and episodic information (see Pierrehumbert, 2006; Rácz et al., 2015). Lexical statistics have been successfully applied in theories of morphological learning and change (Bybee, 1995), but there remains much unexplained variation in speakers' morphological choices and patterns of generalization. A promising route for explanation is the role of social-indexical information in shaping morphological systems. We present a quantitative experimental study on the relationship of morphological perception to speaker gender, a highly salient aspect of the linguistic context that is known to be important in language variation and change. We show that people have significant success in associating English words with speaker gender, and that their implicit knowledge generalizes to gender associations of novel words (pseudowords) on the basis of their component morphemes. By analyzing judgments of morphological decomposition in conjunction with these indexical judgments, we also make inferences about the cognitive architecture for social-indexical effects in morphology.

**Keywords:** pseudowords, morphology, sociolinguistics, indexicality

## 1. Introduction

Morphological systems arise from experience with words as encoded in the lexicon. Both statistical and episodic information about words leave traces in mental representations (see reviews in Pierrehumbert, 2006 and 2016). Lexical statistics are known to be important in morphological learning, and learning in turn relates to change over time (Bybee, 1995; Bybee & Thompson, 1997; Komarova & Nowak, 2001; Daland, Sims & Pierrehumbert, 2007). However,

there remains much unexplained variability in how people acquire and extend morphological patterns. In particular, lexical statistics alone fail to predict why some rare patterns become much more prevalent over time (Bauer, 2001). A factor that may contribute to this variability is social-indexical information. Social-indexical effects have yielded major insights on several aspects of linguistic structure, but their interaction with derivational morphemes and compounding elements not well-studied.

Indexical associations have been documented for whole words (R. Lakoff, 1973) and for morphosyntactic patterns such as number and tense marking (Rickford & Rickford, 2000). In these domains and in others (e.g., allophonic variation), some variants become conventionally associated with different social characteristics. People can provide cues to their social identities when they choose to produce these variants (see review in Eckert, 2008). This process provides an avenue for innovations to take hold; e.g., by people imitating people they admire or identify with (Labov 2001). The extent of such associations for derivational morphemes and compounding elements is not clear. These morphemes could in principle be excellent vehicles for social-indexical information, because they encompass a large number of different forms with rather unrestricted semantics. It is possible that semantically similar affixes, such as *-ity* versus *-ness*, might be used preferentially by different groups. Some groups might use an affix where others use a compound or periphrastic (as in *roomette* versus *sleeping compartment*). Here, we present a quantitative experimental study on the relationship of speaker gender to derivational morphology and compounding patterns. Speaker gender is a highly salient aspect of the linguistic context that has played a central role in sociolinguistic theory. We show that people have significant success in associating English words with speaker gender. Their implicit knowledge generalizes to gender associations of novel words (pseudowords), such as *thrafiun* and *pelpcase*, that appear to be morphologically complex but have no established meaning. Our experimental protocol combines a morphological decomposition task with a social judgment task. By analyzing the combined results, we are also able to shed light on the cognitive architecture that is responsible for the generalization of gender associations to novel complex word forms.

### 1.1. Social-indexical information

Sociolinguistic variation arises in language when groups within a linguistic community develop different patterns of expression. Simple differences in linguistic experience can go towards explaining why people in one group may speak differently from people in another, but it does not provide the full story. Some—but not all—aspects of sociolinguistic variation enter general awareness, and are conventionally associated with specific dialects, groups of people, or with the stereotypical attributes of these groups (e.g., with attributes such as coolness, toughness, or sensitivity). When this happens, the variation has become indexical. It can be used by speakers

to convey social information concurrently with their propositional message. Indexicalization thus requires the variation not merely to exist, but also to be represented in the cognitive systems of speakers and listeners.

Social-indexical variation in the domain of phonetic variation has been intensively studied. Building on the findings of sociolinguistic fieldwork, cognitive encoding of such variation has been revealed in a variety of experimental tasks. Purnell et al. (1999) find that listeners are quite successful in identifying standard, African-American, and Chicano dialects of American English based on variation in the form of the word *hello*. Clopper & Pisoni (2004b) find that listeners are able to classify speakers into regional dialect groups. Hay et al. (2006) find that the apparent social class of the speaker influences the perception of words that are phonetically ambiguous in the context of a merger in progress. Hay & Drager (2010) show that phonetic category boundaries are impacted by subtle priming of the Australian versus New Zealand dialects. Other studies have shown that lexical encoding and memory are compromised for dialects that are low-status or non-standard, even when word recognition has not been affected (Sumner & Samuel, 2009; Clopper et al., 2016). Turning to production, German et al. (2013) describe an imitation experiment in which American English speakers learning the allophones of /t/ and /r/ of a Glaswegian English speaker generalize the target patterns to other words. They retain the ability to generalize the pattern one week later, when their knowledge of Glaswegian dialect is re-activated by hearing speech recordings that do not contain any examples of the target patterns. This behavior clearly involves a cognitive association between the Glaswegian speaker or dialect, and the allophonic pattern. Gender is one of the most salient types of social-indexical information. Gendered associations for phonetic patterns are widely documented, affecting both perception (Johnson, 2006) and production (Foulkes & Docherty, 2006). Gender is of particular interest in models of language variation and change, because women often demonstrate earlier participation in emerging sound changes, at least in English, which is the most studied language (Eckert, 1989; 2008).

The observation that men and women differ in general patterns of word use goes back to R. Lakoff (1973). Large-scale quantitative studies supporting this observation include Boulis and Ostendorf (2005), which analyzed telephone conversations, online forum postings, and web pages; and Mihalcea and Garimella (2016), which analyzed blog posts. In a historical corpus study, Nevalainen et al. (2011) report gendered associations for whole words (*ye* versus *you*), syntactic patterns (*-ing of* versus *-ing*), and also for affixes (*-th* versus *-s*). Such gendered differences may also correlate with differences in register and topic, because people tend to have social clusters based on multiple kinds of similarity. In a study of different registers, Plag et al. (1999) find that some affixes (e.g., *-ity*, *-ness*, *-ion*, *-ize*) are more productive in writing than in speech; Bucholtz (1999, 2001) in turn discusses Greco-Latinate forms as part of a constellation

of language variables used by the “nerd” community of practice at Bay City High School, a social label that reflects not only intellectual interests, but also gender and race.

For gendered social meanings to exist, gender differences in observed usage must be present. However, the presence of these usage differences is not sufficient to imply gendered social meanings. Therefore, observing gendered differences in morphemes may not mean that these morphemes are being used to carry social meanings. Indeed, Nevalainen et al. (2011) suggest that gendered differences may be explained by strong social divisions, not by gendered social meanings per se: “Women tended to lead vernacular changes, whereas men were the leaders of processes related to educated and professional written usage” (p. 4). It is important to note that indexical meaning depends on interrelated layers of context. Silverstein (2003) proposes a theory that connects chains of meaning in “indexical orders”. We can reconsider the argument of Nevalainen et al. in these terms. If the use of a specific morpheme (e.g., -s) implied that the speaker is a woman, this would be a first-order indexical token. However, if instead the use of -s implied a vernacular register, it could be the case that in some context (for example, writing letters amongst people of high social class), use of a vernacular register implied that the speaker is a woman. In this latter analysis, the gender meaning is second order: The implication of ‘woman speaker’ is indirect and mediated by the social meaning of vernacularity within the relevant context. This process is described by Ochs (1992), who argues that “few features of language directly and exclusively index gender”, and that the probing of these networks of indirect, related social meanings gives a richer and more useful understanding of gender in language. As a first step, the current study seeks evidence of gendered associations for a variety of morphemes which demonstrably vary by author gender in the source corpus. At this broad level, it is entirely possible that the gender associations of participants would derive from a variety of different paths (and from indexical tokens of different orders). The current study does not differentiate between a participant associating ‘brunette’ with a woman author that arises through any of the following four possibilities: (1) women more often use the word, (2) ‘-ette’ denotes feminine, (3) ‘-ette’ denotes diminutive (and therefore feminine), (4) ‘brunette’ describes women’s hair. These are all ways that social meaning can be mediated by the lexicon.

Citing Labov (2001), Nevalainen et al. (2011) also suggest that abstract features like morphemes (in contrast to whole words and phonetic features) may be unlikely to be strongly associated with social meanings. Two recent experiments are, however, not entirely consistent with this skeptical view. Using the Asch “social pressure” paradigm, Beckner et al. (2015) show that in a past tense formation task, people are influenced by other people but not by humanoid robots, indicating that social judgment acts a filter in morphological processing. Using an artificial language paradigm, Rácz et al. (2017) investigate the learnability of interlocutor gender as determinant of variability in the form of the diminutive affix, finding that this contextual

condition is as learnable as a phonological condition. These two studies imply social factors in cognition for morphology, putting us one step further towards uncovering social-indexical meanings for morphological patterns. The extent to which they do so is the main concern of this study. First, we use corpus statistics to identify differences between men and women in the usage of words and morphemes. Then, we carry out a gender identification experiment using male-biased, female-biased, and gender-neutral forms. In a novel protocol, the identification task is combined with an explicit morphological decomposition task. The results have important consequences for the influence of social information on word formation and change in the lexicon.

### *1.2. Structure of the mental lexicon*

This investigation into the relationship between social-indexical information and morphemes takes place in the context of active debate over the nature of the mental lexicon and morphological systems which derive from it. If our goal is to determine at what levels and to what units indexical information may attach, then competing ideas about the lexicon impose different constraints. Under multiple-route models (e.g., in Hay & Baayen, 2005), morphemes, simple words, and complex words are specified as entries in the lexicon. Lexical entries for complex words may be accessed either directly or through the morphemes that comprise them. Phonotactic cues, frequency relationships, and semantic transparency all affect which route is more likely to succeed first, and the strength of the morphological boundary in a complex word is a gradient function of the access history. In fully analogical models, both simple and complex words are stored in the mental lexicon, and novel complex words are generated or parsed on-demand based on similarities amongst known words (Daelemans et al., 2010; Dawdy-Hesterberg et al., 2014; Rácz et al., 2015). Words and morphemes may also be undifferentiated, as in the NDL (Naïve Discriminative Learner) model of Baayen, Hendrix, and Ramscar (2013). In an NDL model, the concepts for affixes and roots have the same status, and letter sequences (e.g., trigrams) are linked directly to these concepts. This means that the meaning of the phrase *a British provincial city* is encoded as the concepts {A, BRITAIN, ISH, PROVINCE, IAL, CITY}, and these concepts are statistically associated with all the trigrams in the phrase. In the NDL, complex words are epiphenomenal results of patterns of association between phonological material and categories of meaning. In the Item-and-Process approach (Haspelmath & Sims, 2013), morphologically complex words are created by rules that add or modify simpler word forms. This is the standard approach in generative phonology, receiving a statistical implementation in the MGL (Minimum Generalization Learner) developed by Albright and Hayes (2003).

All these models would need to be augmented in some manner to support social-indexical associations. In interpreting our results, we will discuss simple model extensions, in which

anything that appears in the ontology for a model is a potential host for a social-indexical association. For example, both morphemes and words in the multiple-route model might potentially host associations. In the MGL model, both stems and rules might be associated with social factors. It exceeds the scope of the present paper to consider more complex extensions that might potentially obtain social-indexical effects indirectly.

### *1.3. Morphological decomposition*

In classical linguistic theory, the morpheme is the minimal unit of association between form and meaning, and complex words can be decomposed into two or more morphemes. A confluence of findings, reviewed in Hay & Baayen (2005), indicate that the classic theory is oversimplified, and that the decomposability of complex words is variable and gradient. Hay & Baayen (2001) address the observation that the type frequency of a morpheme is a surprisingly poor predictor of its productivity, showing that the prediction can be improved by assuming that complex words that are more frequent than their stems (such as *stairs* and *government*) are accessed as wholes, and therefore do not contribute to the effective type frequency for the suffixes they exhibit. Hay (2002) shows that English suffixes are generally ordered with more decomposable suffixes outside of less decomposable ones. Hay et al. (2004) show that participants use statistical word-boundary parsing in order to make wellformedness judgments of pseudowords. The judgement is based on the best available parse. People respond as if an internal word boundary is present in pseudowords that contain consonantal sequences that are unattested or rare within monomorphemic words. However, it does not automatically follow from such results that morphological decomposition plays a role in social-indexical processing of speech. The extent to which social associations also accrue for their morphological components is not known. It is also not known whether social associations of known words generalize to novel words, and still less whether any such generalization occurs through overall similarities in word form, or through more structured morphological parsing.

The current study is a step toward untangling these questions. It answers the call of both Pierrehumbert (2006) and Foulkes & Docherty (2006) to improve on traditional statistical models of language by developing ways to account for social effects. The cognitive model needs to be extended to explain dialect, intra- and inter-speaker variability, social interpretation, and the interaction of these factors with other cognitive factors such as word form and denotational semantics. The study considers gender association effects of whole words and morphemes, for simple real words, complex real words, and complex pseudowords. To evaluate the gender associations of morphemes, it focuses on a set of derivational suffixes and compounding elements that differ (according to a corpus study) in their rates of use by men versus women. Indexicality is evaluated by asking participants to decide whether word forms are more likely to have been

produced by a man or a woman. Participants also give an explicit decomposition for each word (or respond that no decomposition is needed), alongside the gender association response. Our analyses consider gender responses in conjunction with both the accuracy of morphological decomposition, and the objectively available support for morphological decomposition.

## 2. Methods

### 2.1 Corpus statistics

We selected the British National Corpus to survey gender bias for words and suffixes. It included material from a variety of different genres for which the gender of the author can be determined. For this study, we used the written portion of the British National Corpus, and included only those documents that could be attributed to men or women authors. The British National Corpus written subset contained 3,141 documents, for a total of 87,953,932 words; after filtering for author gender, there were 378 documents from women (13,451,416 words) and 844 documents from men (28,659,100 words). We note that the corpus had more material written by men authors than by women authors. This may have improved the statistical estimates for man-biased forms. In addition, information currently available in the British National Corpus limited us to considering gender in terms of a man–woman binary. In this corpus, as in everyday life, author gender is correlated with the topic of discussion. More than half of the ‘imaginative’ content domain was written by women, making their relative representation over twice that of men. However, men were overrepresented in the other 9 content domains, especially ‘natural science’ (2300%) and ‘commerce’ (500%). In this study, we lacked the information to tease apart these variables.

Following Mihalcea and Garimella (2016), we calculated the gender bias of each word as the ratio of use frequency by women versus men authors. This is expressed below as a log ratio. Negative values mean that the word is man-biased; women use the word less than men. Positive values mean that the word is woman-biased. The results broadly replicated Mihalcea and Garimella (2016) in finding that a large number of words display little gender bias, but a certain number are used much more by one gender than by the other. These provided targets for the experimental stimuli. Morpheme gender bias values were calculated as the ratio of grouped usage frequencies for complex words sharing the final morpheme as determined from CELEX decompositions; e.g., the calculation for *-land* includes *grassland*, *dreamland*, and so on. For compounds, the value was determined solely from appearances of the compounding element as the second element in a compound word, because some frequent compounding elements have diverged semantically from their meanings as isolated words. It may not be surprising that gender

bias was present for a variety of compounding elements, and it also proved to be present for a variety of suffixes.

## 2.2. Presentation and stimuli

The study used a new online experimental paradigm in which participants were shown a series of words and pseudowords, one at a time. Each word was presented with a user interface to allow a single marker to be placed between the letters of the word, indicating a decomposition boundary; and accompanied by a pair of named face images. For each item, the participant responds to two tasks: (a) “Split the word into two meaningful parts, if possible.” and (b) “Which author most likely used this word?”. The participant indicates a single position to split the item by clicking between the letters displayed to move the decomposition marker. To give the gender response, they click directly on the face of either the man or the woman shown above the item; see Figure 1 for images of example trials.



Figure 1. Example experiment trials. Gendered faces and decomposition responses are shown: The left panel shows a decomposition placed and the left face selected; the right panel shows a ‘no decomposition’ response, with a face not yet selected.

This paradigm was used to gather explicit morphological decomposition responses for simple and complex items, as well as the implicit gender associations for each item. The participant may complete these two tasks in either order, prior to clicking the ‘Next’ button to move to the next trial. The explicit decomposition paradigm was previously validated using a baseline experiment (216 participants, 288 items each), in which participants gave morphological decomposition responses in addition to Likert ratings for item familiarity. The items in the baseline experiment were the same as those in the current study. All real words were rated as highly familiar, and all pseudowords were rated as unfamiliar. In the baseline experiment, the average accuracy for decomposition responses was 96% for simple real words, 88% for complex real words, and 65% for complex pseudowords (taking the “correct” decomposition to be the one assumed in constructing the stimuli). Accuracy is similar for the current experiment: 96% for simple real words, 86% for complex real words, and 65% for complex pseudowords (see Section 3.1 for full analysis of parsing).

## 2.3. Stimuli

### 2.2.1. Gendered faces

Faces for each ‘author’ were created using public domain images of 6 women and 6 men. This experiment used only faces appearing to be white adults between 25 and 40 years old (see Figure 2). While the faces sufficiently convey the intended gender cue, appropriate names are included to support the experimental narrative that stimuli should be associated with the authors. Each of the 12 images was assigned a name based on the most popular names by gender in the United States since 1917 (Social Security Administration, 2016). We consider these names to have stable gender associations and familiarity for participants of varying ages. Each name was among the 10 most popular names for the 100 year interval, and all names were in the top 200 most popular for Americans born in the 1980s (which corresponded to the face age range). None have ambiguous gender. Pairings of names and images were the same in all trials. All different-gender pairs of men and women were used to make 36 distinct pairings, and were presented in 2 orders (man–woman, woman–man) for a total of 72 face pair orderings.

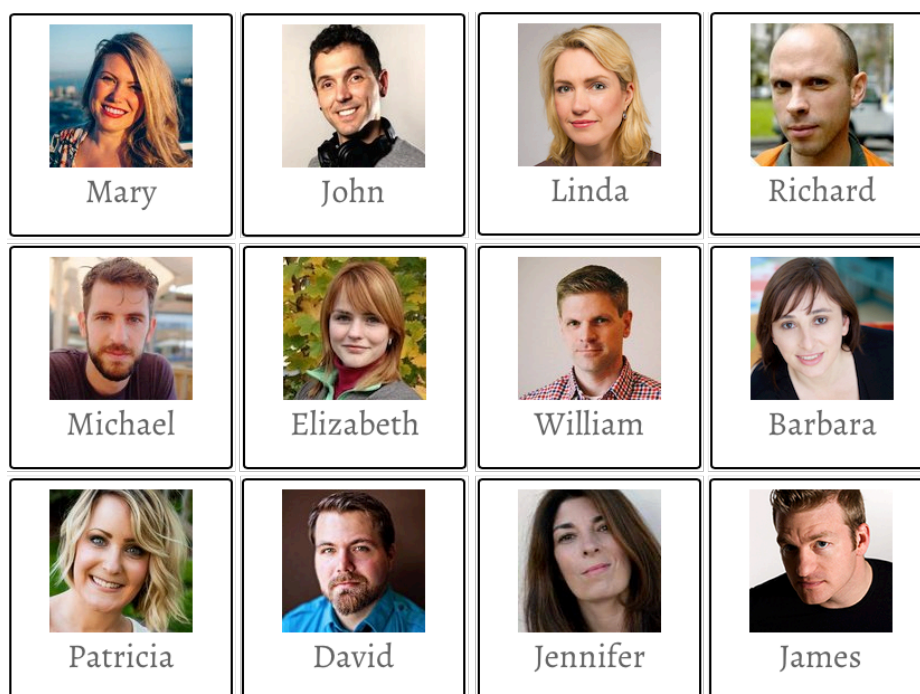


Figure 2. Name–picture pairings used in the experiment.

### 2.2.2. Items and script design

Item stimuli consisted of simple real words, complex real words, and complex pseudowords. Each real word had a whole-word gender bias value. In each complex word, the second morpheme had a morpheme gender bias value. Target morphemes included both compounding elements and suffixes. The complex pseudowords, designed to be comparable to the real complex words, consisted of a pseudo-stem and a real morpheme ending. The stems for these pseudowords were drawn from amongst the 8400 pseudowords that were generated for the norming study presented in Needle, Pierrehumbert, and Hay (under review). These varied in length and have statistical wordlikeness scores as determined by smoothed phonotactic and orthotactic scores. The stems selected for the present study all had above-median scores. In addition, stems with low ratings (regardless of score) were excluded; thus, selected stems were all of good phonotactic quality. Three additional criteria were imposed. The length distribution fell in the middle of that for real stems in the study. Stems were selected to have a phonotactically legal transition to the suffix, defined as having a digram probability within the range for the complex real words. Combinations with unanticipated word embeddings were eliminated by hand. For example, *egaussage* was not used as an example of a word with the suffix *-age* because it contains the words *gauss* and *sage*. The complex real words use different morphemes from the pseudowords, and their stems are always able to stand alone (e.g., *grass* in *grassland*).

The experiment had 288 items: 108 complex real words, 108 complex pseudowords, and 72 simple real words. Simple real words were balanced by whole-word gender bias: 24 woman-biased, 24 neutral, and 24 man-biased. Pseudowords were balanced by morpheme gender bias, with three examples each of 36 morphemes: 12 woman-biased, 12 neutral, and 12 man-biased. Complex real items were balanced for both whole-word gender bias and morpheme gender bias: 12 woman-biased, 12 neutral, and 12 man-biased morphemes; within each morpheme, there was one woman-biased, one neutral, and one man-biased whole-word example. It proved impossible to balance the gender bias values perfectly for whole words or for morphemes: For whole words, the mean values were 1.2 for woman-biased, -0.06 for neutral, and -1.5 for man-biased; for morphemes, the mean values were 0.31 for woman-biased, -0.26 for neutral, and -0.71 for man-biased. Summary statistics on characteristics of the items are provided in Table 2. During item selection, frequent morphemes and words were preferred. The morphemes used include both suffixes and compounding endings. For suffix-type morphemes, 24 are consonant-initial and 24 were vowel-initial. For compound-type morphemes, all 24 were consonant-initial. It was not possible to find 24 vowel-initial words that both occur frequently in compounds and exhibit strong gender bias. The morphemes varied in productivity, and both morphemes and whole words varied in length and frequency. For the Complex Real condition, it was necessary to reach farther down the word frequency scale than for the Simple Real condition to obtain enough items with strong gender biases (see Table 2). For examples of experiment items, see Table 1.

Table 1. Example stimuli by category, including compounds and suffixations. Gender bias for whole words and for morphemes are marked: Bias toward women, men, and neutral are marked by ‘W’, ‘M’, ‘N’.

<b>Simple Real</b> <i>72 items</i>	<b>Complex Real</b> <i>108 items, 36 real endings</i> -ism M                      -light W		<b>Complex Pseudo</b> <i>108 items, 36 real endings</i> -ium M                      -case W	
porcelain W	sexism W	firelight W	balnium	snoshcase
tennis M	modernism M	searchlight M	vodepium	clumcase
straight N	heroism N	daylight N	thrafiium	pelpcase

Table 2. Summary statistics for real word stimuli: Whole-word gender bias and frequency.

	<b>Gender Bias</b>				<b>Log Frequency</b>			
	<b>Min</b>	<b>Max</b>	<b>Median</b>	<b>SD</b>	<b>Min</b>	<b>Max</b>	<b>Median</b>	<b>SD</b>
All real words	-4.3	2.4	-0.03	1.25	0	8.8	4.4	1.65
Simple reals	-2.1	2.3	-0.1	0.99	3.9	8.1	5.0	1.01
Complex reals	-4.3	2.4	-0.03	1.0	0	8.8	3.6	1.74

#### 2.4. Participants

The study collected data from 216 participants via Amazon Mechanical Turk (111 women, 101 men; four participants declined to provide gender information). All participants reported being English speakers currently residing in the United States. Reported birth years range from 1948 to 1996 (three participants declined to answer). All participants completed the experiment between 2017-5-1 and 2017-6-1. Participants were paid \$3 for completing the task, which took up to 30 minutes. Six participants were excluded for insufficient decomposition performance on simple and complex real words combined ( $d' < 1$ ).

### 3. Results

We first evaluate gender associations of whole-word and morpheme gender bias for simple real words, complex real words, and complex pseudowords (Regression Models 1, 2, and 3). Then, we consider the role of explicit morphological decomposition: Does explicit parse accuracy necessarily imply morpheme awareness, or could performance instead be explained by phonotactic or orthotactic cues (Model 4)? Do participants need to correctly parse morphemes to be influenced by the gender associations of those morphemes (Model 5)? Finally, if explicit parsing is not required, does it nonetheless improve gender response accuracy?

Before we begin with these questions, a note about participant demographics is appropriate. When considering questions of social meaning in linguistics, it is important to consider not only who is speaking, but who is listening. Many social meanings are sensitive to social context, and vary from person to person. In addition to information about age and gender, participants reported their highest level of formal education from the following list: “Some High School or equivalent”, “High School graduate or equivalent”, “Associate's degree”, “Bachelor's degree”, “Master's degree (MA, MS, MFA, or equivalent)”, “Doctoral degree (Ph.D, LL.D, or MD)”. Our initial analyses of response effects related to participant age, gender, and education level did not yield significant results, and are not reported here. A future study may undertake a more detailed analysis, or the collection of a larger and more nuanced set of participant data.

The effects of word and morpheme bias on gender responses were analyzed using logistic mixed-effects regression with the function ‘glmer’ implemented in R package ‘lme4’ (Bates, et al., 2015) in R (R Core Team, 2014). For all regression models reported here, each continuous measure was centered in the models: Whole-word gender bias, morpheme gender bias, and log word frequency. Final models were the result of a consistent pruning procedure: For each model, analysis began by including all relevant fixed effects and their interactions, as well as slopes and intercepts for each random effect. The random effects for each item were nested under the morphemes, because the stimulus design included exactly three items for each morpheme; items cannot be independent of their morphemes. None of the models converged with random slopes included, so the first step of pruning in each case was to remove random slopes (leaving random intercepts only). Insignificant terms were removed from the models one by one, with higher-order (interaction) terms removed first. Details of model pruning are described for each model, below.

It was necessary to split the gender response analysis into 3 models: Simple real words had whole-word bias values only (Model 1), complex real words had both whole-word and morpheme bias values (Model 2), and pseudowords had morpheme bias values only (Model 3). For whole-word and morpheme gender bias values, woman-biased values were greater than 0, and man-biased values were less than 0. The response variable for Models 1, 2, and 3 was gender

response, in terms of log-odds. Summaries for Models 1, 2, and 3 are given on Table 3. In figures showing effects from these models, log-odds estimates are transformed and shown in terms of probability of woman responses chosen, from never (0) to always (1); shaded regions around the effect estimates indicate pointwise 95% confidence intervals for normal distributions.

The model for simple real items (Model 1) contained a fixed effect for whole-word gender bias, and random intercepts for each participant, item, and face image. During pruning, the interaction of word gender bias with word frequency was removed first, and then the main effect of word frequency was removed from the final model. There was a significant effect of whole-word gender bias ( $\beta = 0.42$ ,  $SE = 0.076$ ,  $z = 5.5$ ,  $p < 0.0001$ ). Participants were more likely to choose woman responses as the item became more woman-biased (see Figure 3a). For complex real items (Model 2), the model contained fixed effects for whole-word gender bias, morpheme gender bias, and log word frequency; interaction terms for word gender bias with morpheme gender bias, and for word gender bias with log word frequency; and random intercepts for each participant, morpheme, item, and face. During pruning, the three-way interaction between word gender bias, morpheme gender bias, and log word frequency was removed first; then, the two-way interaction between morpheme gender bias and log word frequency was removed. Word frequency was taken from the COBUILD corpus via CELEX. There was a significant positive effect of word gender bias ( $\beta = 0.31$ ,  $SE = 0.041$ ,  $z = 7.5$ ,  $p < 0.001$ ). Model 2 showed the same pattern for whole words as Model 1 (see Figure 3b): Responses for real words reflected their gendered statistics, with people more likely to choose woman responses as the complex real items became more woman-biased. There was also a significant positive effect of word frequency ( $\beta = 0.092$ ,  $SE = 0.031$ ,  $z = 3.0$ ,  $p = 0.0027$ ), meaning that higher-frequency words were associated more with women. The main effect of morpheme gender bias was not significant ( $\beta = 0.031$ ,  $SE = 0.12$ ,  $z = 0.26$ ,  $p = 0.79$ ) (see Figure 6b).

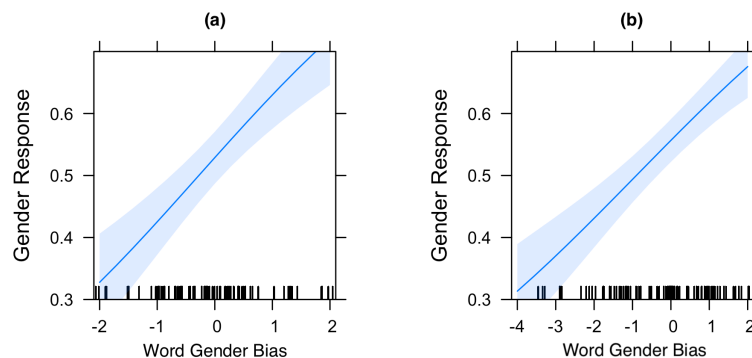


Figure 3. Effect of whole-word gender bias on gender response for simple (Model 1: a) and for complex (Model 2: b) real words. Shaded regions indicate pointwise 95% confidence intervals for normal distributions.

There were two significant interactions affecting word gender bias. The interaction of word gender bias with word frequency is significant ( $\beta = 0.086$ ,  $SE = 0.027$ ,  $z = 3.2$ ,  $p = 0.0016$ ), such that the effect of word gender bias on gender response is weaker as frequency decreases (see Figure 4). Experience with a word is needed for a gender association effect to obtain, and more experience supports better learning of the association. The interaction of word gender bias with morpheme gender bias was also significant ( $\beta = 0.20$ ,  $SE = 0.080$ ,  $z = 2.5$ ,  $p = 0.013$ ): The influence of word gender bias increases as morphemes are more woman-biased (see Figure 5). That is, amongst words containing more woman-biased morphemes, the man-biased whole words were judged to be more man-biased, and the woman-biased words were judged to be more woman-biased. We view this interaction with considerable caution, because it does not arise naturally in any current model of the mental lexicon. Insofar as the effect proves to be reliable, we speculate that it might arise indirectly from the correlation of gender bias with register and topic in the experimental stimuli. Overall, the man-biased morphemes are more typical of formal prose and the woman-biased morphemes are more typical of colloquial language. The gender association of a whole word might be more salient—and thus easier to learn—in conversational contexts than in formal prose.

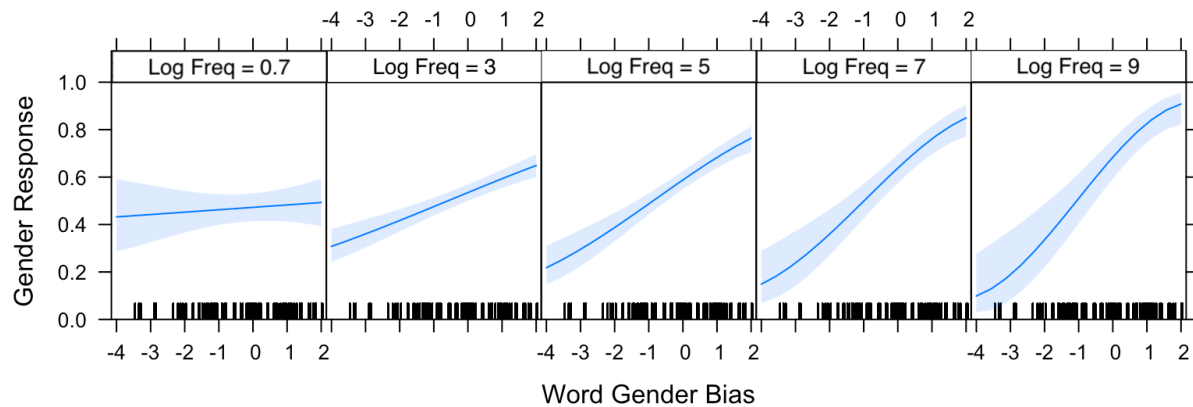


Figure 4. Interaction of word gender bias with word frequency for gender response (Model 2). Shaded regions indicate pointwise 95% confidence intervals for normal distributions.

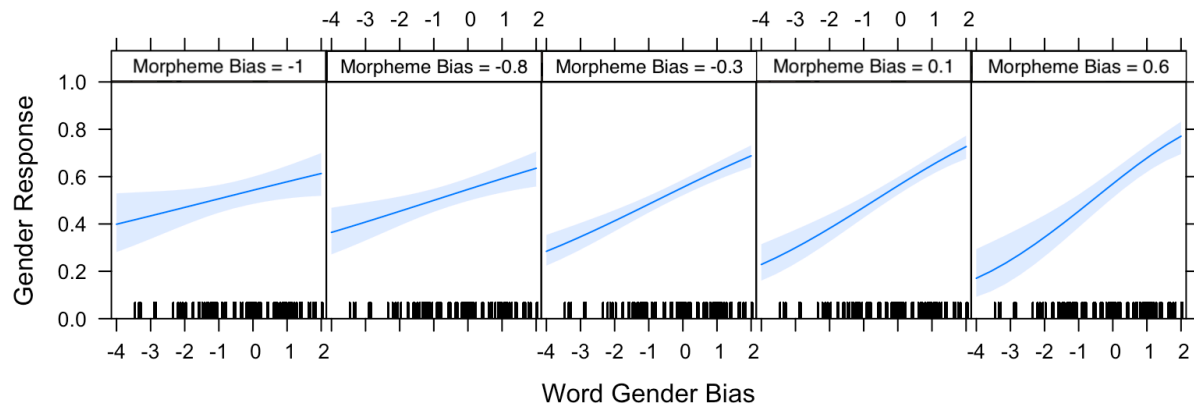


Figure 5. Interaction of morpheme gender bias with word gender bias for gender response (Model 2). Shaded regions indicate pointwise 95% confidence intervals for normal distributions.

The model for complex pseudowords (Model 3) contained a fixed effect for morpheme gender bias, and random intercepts for each participant, morpheme, item, and face image. All interactions were pruned from the final model for lack of significance. Complex pseudowords were significantly more likely to be associated with women faces when the morpheme group was more woman-biased ( $\beta = 0.22$ ,  $SE = 0.06$ ,  $z = 3.5$ ,  $p < 0.001$ ). Figure 6 compares the morpheme gender bias effect for complex pseudowords versus complex real words.

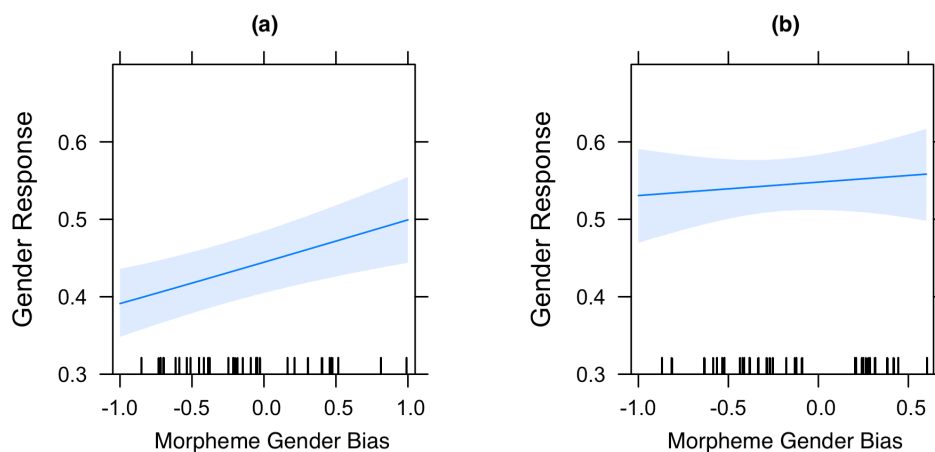


Figure 6. Effect of morpheme gender bias on gender response for complex pseudowords (Model 3: a) and for complex real words (Model 2: b). Shaded regions indicate pointwise 95% confidence intervals for normal distributions.

Table 3. Regression model summaries for Models 1, 2, and 3.

<b>Model 1:</b> <b>gender_response ~ word_gender + (1 workerId) + (1 item) + (1 face_1) + (1 face_2)</b>				
	Estimate	Std. Error	z-value	p(>  z )
<i>word_gender</i>	0.42	0.076	5.5	< 0.001
<b>Model 2:</b> <b>gender_response ~ word_gender + morph_gender + log_freq + word_gender:morph_gender +</b>				
<i>word_gender</i>	0.31	0.041	7.5	< 0.001
<i>morph_gender</i>	0.031	0.12	0.26	0.79
<i>log_freq</i>	0.092	0.031	3.0	0.0027
<i>word_gender:morph_gender</i>	0.20	0.080	2.5	0.013
<i>word_gender:log_freq</i>	0.086	0.027	3.2	0.0016
<b>Model 3:</b> <b>gender_response ~ morph_gender + (1 workerId) + (1 morph) + (1 morph:item) + (1 face_1)</b>				
<i>morph_gender</i>	0.22	0.063	3.5	< 0.001

### 3.1. Decomposition accuracy

We now turn to the accuracy of decomposition responses, as a prerequisite to considering the role played by explicit morphological decomposition in the gender associations. A decomposition was judged accurate for complex items if it exactly parses the intended morpheme, and for simple items if the response was ‘no decomposition’. As in the baseline experiment used to validate the paradigm, decomposition accuracy rates were above chance for simple real words, complex real words, and complex pseudowords. Performance on complex words also exceeded the level predicted by phonotactic boundary statistics alone (described below).

Using logistic mixed-effects regression, we evaluated the contribution of phonological information to decomposition accuracy; specifically, we tested the possibility that participants are parsing items based on the statistical cues to the presence of a word boundary, without any perception of morphemes. For both phonemic and orthographic bigrams at the location of the expected decomposition, we compared the likelihood of a boundary being present versus absent (taking the difference of the log likelihoods) (cf. Daland & Pierrehumbert, 2011). The boundary

likelihood ratio was derived from orthographic and phonemic bigram statistics in the 10931 CELEX monomorphemes, a list made as discussed in Hay et al. (2004) by hand-checking the lexical entries in the CELEX lexicon (Baayen et al., 1995). Monomorphemes were used so that the bigram statistics accurately reflect words without internal boundaries. Boundary likelihood ratio was defined as the probability that a boundary is present, divided by the probability that a boundary is not present. To estimate the probabilities for bigrams with boundaries, we make the simplifying assumption that words can combine freely.

For morphological decomposition accuracy, Model 4 contained fixed effects for orthographic boundary likelihood ratio, phonemic boundary likelihood ratio, and for lexicality (real word or pseudoword); and random intercepts for each participant, morpheme, and item (see Table 4). In Figure 7, log-odds estimates were transformed and shown in terms of probability of expected decomposition responses given, where correct is 1 and incorrect is 0. During pruning, the three-way interaction between orthographic boundary likelihood ratio, phonemic boundary likelihood ratio, and lexicality was removed first; then, each two-way interaction was removed. Model 4 included response data for complex real words and complex pseudowords only; 7% of pseudoword data and 8% of complex real word data were excluded because boundary ratio statistics were not available for the expected boundary. In Model 4, there was a significant positive effect of lexicality: Participants parse pseudowords less accurately than complex real words overall ( $\beta = 2.1$ ,  $SE = 0.34$ ,  $z = 6.2$ ,  $p < 0.0001$ ) (see Figure 7c). This lexicality effect may mean that participants gained a boost from recognizing two morphemes (the stem and the affix) instead of only the affix; or that they had explicit morphological knowledge of the familiar real words. The strength of the orthographic cue was significantly associated with decomposition accuracy ( $\beta = 0.41$ ,  $SE = 0.093$ ,  $z = 4.4$ ,  $p < 0.0001$ ) (see Figure 7a). Participants parsed complex stimuli more accurately when the expected boundary was orthographically likely. The effect of phonemic boundary cue was not significant ( $\beta = 0.11$ ,  $SE = 0.083$ ,  $z = 1.3$ ,  $p = 0.21$ ) (see Figure 7b).

The usefulness of boundary likelihood is limited: The orthographic boundary cue led to a correct parse in only 37% of complex items (42% for complex real words, 32% for complex pseudowords), but participants gave correct parses for 88% of the complex real words and 65% of the complex pseudowords. This discrepancy means that participants were making significant use of other information for morphological decomposition, such as recognition of morphemes per se.

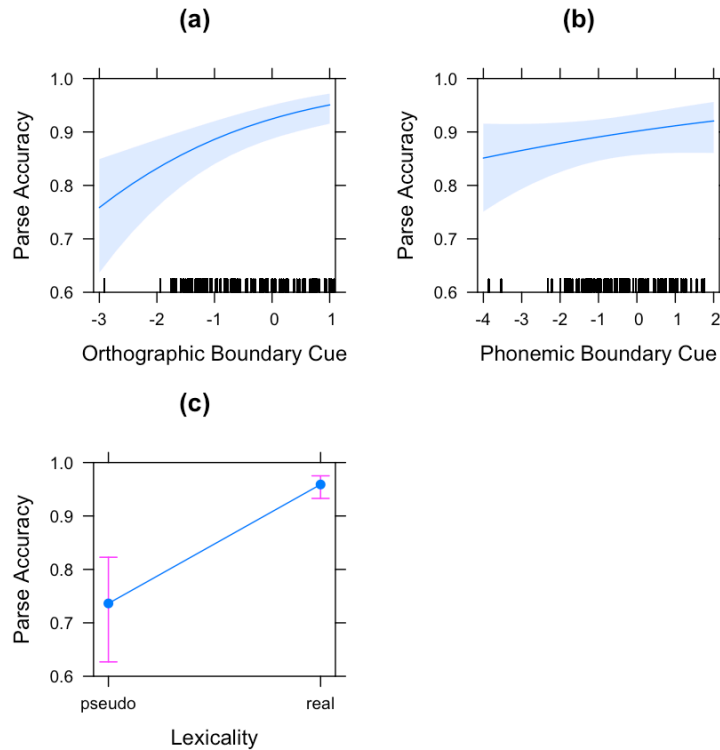


Figure 7. Effects on decomposition accuracy (Model 4) of orthographic boundary cue (a), phonemic boundary cue b), and lexicity (c). Shaded regions indicate pointwise 95% confidence intervals for normal distributions.

Table 4. Regression model summary for Model 4.

<b>Model 4:</b>				
<b>accurate ~ boundary_ortho + boundary_phono + lexicity + (1 workerId) + (1 morph) + (1 morph:item)</b>				
	Estimate	Std. Error	z-value	p(>  z )
<i>boundary_ortho</i>	0.41	0.093	4.4	< 0.001
<i>boundary_phono</i>	0.11	0.083	1.3	0.21
<i>lexicity = 'real'</i>	2.1	0.34	6.2	< 0.001

### 3.2. Gender accuracy in relation to decomposition of pseudowords

As shown in Models 2 and 3, the morpheme gender bias had a significant effect only for pseudowords. How did this effect come about? The decomposition analysis showed that people had moderate success in decomposing pseudowords, and it suggested that they were using both morphological awareness and orthotactic cues. We can now ask whether morphological parsing

influences the morpheme gender effect. Decomposition could provide a boost to the activation of embedded morphemes, or decomposition might be required for the morphemes to be activated at all. We evaluated this question with Model 5, which considered factors affecting gender response accuracy for pseudowords only (Table 5). The gender response accuracy measure reflects whether participants gave the expected gender response for each pseudoword, based on their embedded morphemes: I.e., choosing a woman's face when the morpheme gender bias was greater than 0, and choosing a man's face when the morpheme gender bias was less than 0 (neutral items are excluded). We evaluated the hypotheses that (1) gender response accuracy was higher when participants made accurate decompositions, and (2) higher for word forms that contained stronger boundary cues to the decomposition. Along with these decomposition-related factors, Model 5 included morpheme gender bias magnitude, which is the absolute value of the morpheme gender bias value (which was critical in Model 3): More extreme bias (toward either men or women) should help participants to choose the expected gender response, improving gender response accuracy. We used the absolute value in this model to place the gender bias factor on similar footing with the decomposition factors: Higher gender accuracy was hypothesized in the presence of an accurate decomposition, of stronger boundary cues, and of stronger morpheme gender bias.

Table 5. Regression model summary for Model 5.

<b>Model 5:</b>				
<b>gender_accurate ~ accurate + boundary_ortho + boundary_phono + abs(morph_gender) + (1   workerId) + (1   morph) + (1   morph:item)</b>				
	Estimate	Std. Error	z-value	p(>  z )
accurate = 'TRUE'	-0.035	0.042	-0.84	0.40
boundary_ortho	0.026	0.043	0.60	0.55
boundary_phono	0.028	0.050	0.57	0.57
<i>abs(morph_gender)</i>	<i>0.58</i>	<i>0.13</i>	<i>4.6</i>	<i>&lt; 0.001</i>

Actual parse accuracy and both boundary cues were included to cover the possible case that participants had poor awareness of morphological parsing information that nonetheless implicitly affected their gender responses. All three cues could be included in a single model because they were not excessively correlated. Fixed effects were included for orthographic boundary cue, phonemic boundary cue, morpheme gender bias magnitude (the absolute value of

morpheme gender bias), and parse accuracy (true or false); and random intercepts were included for participant, item, morpheme, and face image. During pruning, the four-way interaction was removed first, followed by all three-way interactions, and then all two-way interactions. There was no significant effect of orthographic boundary cue ( $\beta = 0.026$ ,  $SE = 0.043$ ,  $z = 0.60$ ,  $p = 0.40$ ) (Figure 8a), phonemic boundary cue ( $\beta = 0.028$ ,  $SE = 0.050$ ,  $z = 0.57$ ,  $p = 0.57$ ), or parse accuracy ( $\beta = -0.035$ ,  $SE = 0.042$ ,  $z = -0.84$ ,  $p = 0.40$ ). These results suggest that participants' gender response accuracy was not affected by whether they decomposed the items. In contrast, the magnitude of morpheme gender bias was a highly significant predictor of gender response accuracy ( $\beta = 0.57$ ,  $SE = 0.13$ ,  $z = 4.5$ ,  $p < 0.0001$ ) (Figure 8b); participants were more likely to choose the gender face that matched the expected morpheme gender as the gender bias increased.

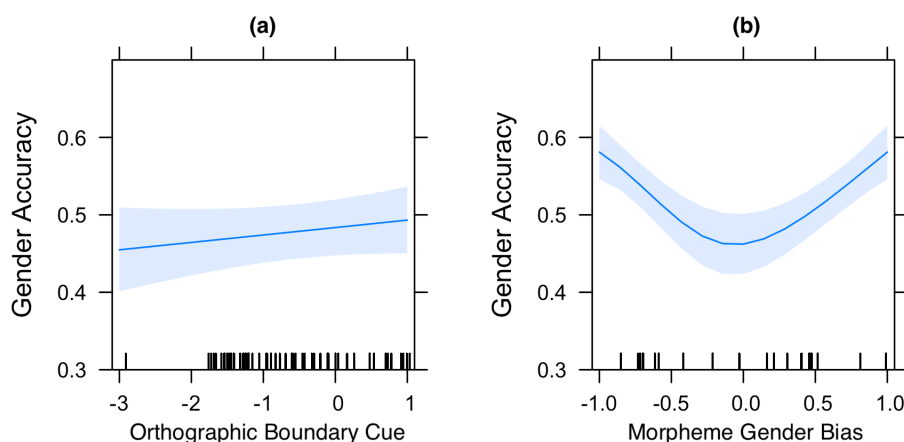


Figure 8. Gender accuracy effects (Model 5) of orthographic boundary cue (a) and morpheme gender bias (b). The morpheme gender bias effect is shown after reverting the absolute value transform used in Model 5. Shaded regions indicate pointwise 95% confidence intervals for normal distributions.

### 3.3. Summary of main results

The main results of this study are given on Table 6. Participants showed high accuracy in decomposition of simple and complex stimuli, though accuracy was lower for pseudowords than for real words. We found that speakers have social-indexical associations between words and gender, so that their gender response correlate with the gender bias measures. The effect of word gender bias on gender responses is influenced by two relevant interactions: With word frequency, and with morpheme gender bias. We also found gender associations for morphemes within pseudowords, but not for real words.

Table 6. Summary of major results.

Lexicality	Morphological Complexity	Decomposition Accuracy	Gender Effect of Word Bias	Gender Effect of Morpheme Bias
Real Word	Simple	96%	yes	—
Real Word	Complex	86%	yes	no
Pseudoword	Complex	65%	—	yes

#### 4. Discussion

Our results for whole words extend related findings such as Quina et al. (1987) and Bucholtz (1999, 2001). For both simple and complex real words, participants reliably matched the gender bias of the whole word as estimated from corpus statistics, suggesting that their intuitions are the result of gendered language experience. This gendered experience might be with the words per se, or with the concepts associated with the words; e.g., we cannot distinguish whether whole-word gender responses reflect associations of words like *sodium* and *iridium* with men (e.g., by hearing men use these words), or associations of *sodium* and *iridium* with atomic elements as a science topic, which is in turn associated with men. The interaction of word frequency with gender association for complex words supports the hypothesis that intuitions are the result of gendered language experience: Even though all of the real words in this study were rated as highly familiar in the baseline experiment, the gender association is strongest for the most frequent words, and disappears for the rarest words. This result is reminiscent of the pattern found by Clopper and Pisoni (2004a) for dialectal experience: Their listeners were better able to associate speakers with regional dialects when they had more exposure to relevant speech variation. The effect of word frequency on gender association was not significant for the simple word model, which may be explained by the different frequency ranges for simple and complex real words: At 150, the median simple word frequency is higher than 75% of complex real words (for which the median is 37). Additional work with rarer simple word stimuli might show the same disappearance of the gender association effect.

We did not see a main effect of morpheme gender bias for complex real word stimuli. Instead, we understand the significant interaction between morpheme gender bias and word gender bias from the perspective of the word gender bias main effect: The effects of word gender bias were more polarized in words containing more woman-biased morphemes. This means that

the effect of morpheme gender is not cumulative with the effect of word gender, but enhances the word gender effect. We view this interaction with extreme caution due to both the size and the nature of the effect, which is not predicted by any of the morphological theories considered. We suggested that this pattern, if it is a real one, might arise as an artifact of the different communicative situations in which the various words in our study are encountered: Formal and textual, versus casual and face to face.

For the pseudoword stimuli, where whole word knowledge does not exist, we see a main effect of morpheme gender bias on gender responses. Our results show that participants significantly associated pseudowords with gendered faces that matched the gendered corpus statistics for their component morphemes. In addition, when gender bias was stronger, participants were better able to choose the predicted gender response. The morpheme gender association effect obtains regardless either of participants' accuracy in explicitly parsing the pseudowords, or of the presence of partially useful orthotactic cues to the presence of an internal word boundary. That is, the ability to identify the morpheme itself does not change the gender association effect. These results indicate that people's gender associations with unknown words are influenced implicitly by the presence of real morphemes. They contrast with the outcomes observed for real words, in which gender associations of morphemes were not significant as a main effect.

Our results present us with questions about the different roles of knowledge about whole words versus word parts. For real words, whole-word knowledge affects gender responses, to the exclusion of word-part knowledge. For pseudowords, only word-part knowledge is available, and it affects gender responses. However, explicit decomposition results do not control or influence this effect, so word-part knowledge appears unrelated to decomposition responses in these tasks. In section 1.2, we summarized four different current theories about morphological representation and processing in the mental lexicon: Models related to multiple-route, general analogy, probabilistic rule application, and the NDL. In light of the results presented, we can engage more deeply with these theories and consider how readily they can be extended to encompass the socio-indexical patterns that we found.

We can consider the multiple-route approach as a more sophisticated alternative to an obligatory decomposition model. With obligatory decomposition (see Taft, 2004), people would always recognize and activate morphemes, and retrieve stored gender information about each morpheme. The pattern of current results is not consistent with the predictions of an obligatory decomposition model, under which we would have expected judgments of complex real words to reflect the gender associations of the parts (even if gender associations of the whole word also play a role in the people's judgments). We might also have expected the morpheme effect to be stronger when the whole word frequency is low (that is, too low for a whole-word gender

association to obtain). However, there was no significant effect of morpheme gender bias for complex real items, regardless of whether the whole word frequency was high or low. Under this always-decompose theory, we would further have expected that morphological decomposition would feed into gender judgments for pseudowords: The association with gender would be stronger when the gendered morpheme was identified in the parse. However, this expectation is not fulfilled.

A multiple-route approach is more readily extended to encompass our results, although some points of difficulty in doing so require additional assumptions about the dependence of the results on the tasks in the experiment. In this approach, the phonological or orthographic representations of words are recognized either as wholes, or by decomposition into constituent morphemes. Both routes lead to activating the meaning of the whole word, and the question of which route wins is subject to concerns such as the relative frequencies of the whole word and its morphemes. Both whole words and morphemes are present in the mental lexicon, with their own associated information. This information includes the gender-biased experience assumed in the current study: Both morphemes and whole words can have gendered associations. If we assume that activation of morphemes should be reflected in the gender response task, it predicts a pattern of results in which pseudowords reflect morpheme gender bias because they are decomposed and processed by parts, as no whole-word option is available. A morpheme gender effect could be observed for the highly-decomposable complex real word stimuli; specifically, those words with lower whole-word frequency and high-frequency morphemes would be decomposed and processed by parts. In contrast, for real words accessed by the whole-word route, the associations of the morphemes would not be activated. However, this prediction depends on the assumption that complex pseudowords are aggressively decomposed, and that many complex real words are not decomposed during lexical access. These assumptions are not well supported by our decomposition results. Our real words were highly decomposable, and people did decompose them (with parsing accuracy above 85%). The pseudowords were less reliably decomposed and the decomposition was not predictive of gender responses. The lack of morpheme gender influence for real words might be explained by the nature of the gender response task, which can be considered to be a slow or high-level task. This means that participants have plenty of time to activate the relevant meaning representation for complex real words, regardless of the route used, so their gender responses reflect only knowledge about that meaning representation. This view is compatible with our pseudoword gender results, though it may require a considerable disconnect between the implicit process of morphological decomposition during lexical access, and the information about morphological decomposability collected with our protocol.

General analogy models as presented in Nosofsky (1988, 1990), Daelemans et al. (2010) and Dawdy-Hesterberg et al. (2014) readily capture our results, with the proviso that analogical forces determine judgments about unknown words, but have only a weak influence on judgments for known words (as claimed in Daland et al. (2007)). The mechanism for pseudowords to have apparent gender associations under such an approach is comparable to that proposed in Johnson (2006) for social identity correlates of allophonic variation to emerge. If the gender effect for pseudowords does not come from explicit gender information known about the morphemes themselves, general analogy provides an alternative mechanism: Pseudowords inherit the implicit associations of similar real words. In this case, similarity derives in part from sharing a morpheme: The unknown word *glonitis* would be similar to *bronchitis*, *arthritis*, etc., so it would get the gender association of the overall group. This mechanism depends on whole word gender associations, which have been demonstrated previously and which this study replicated. This mechanism is reminiscent of results in Nation and Cocksey's (2009) study on semantic interference. They found semantic interference from sub-word orthographic matches (e.g., *hip* in *ship*) when the sub-word took beginning, middle, or final position in the word, or even when the sub-word involved phonological mismatches to the target (e.g., for the letter 'h' in *hip* and *ship*). The semantic associations in that experiment clearly result from overall word similarities and not from morphological decomposition.

The Albright and Hayes (2003) MDL model could capture gendered associations of morphemes by probabilistically associating gender with morphological rules, effectively capturing the results for pseudowords that were actually decomposed. For real complex words, it would be necessary to add the proviso that knowledge about the whole word takes priority over predictions from the rule system. While this proviso is not clearly stated in Albright and Hayes (2003), it is independently necessary to explain why real words have highly stable inflectional morphology even if they belong to groups of word forms whose morphology varies. For example, the past tense of *keep* is *kept* and the past tense of *beep* is *beeped*; only a novel word such as *fleep* exhibits instability (e.g., *fleeped*, *flept*). The challenge for this model would be to explain why the gendered associations for pseudowords were found to be unrelated to the decomposition judgments, or to the cues for decomposition.

The NDL model of Baayen, Hendrix, and Ramscar (2013) is very different from the other approaches presented. Under this theory, there are no lexicon representations for whole words or morphemes at the orthographic or phonological level, but only at a semantic level. Instead, phonological sequences (e.g., triphones) are probabilistically associated with meanings. Kuperman (2013) analyzes the behavior of the NDL in relation to a study of the effects of the emotional and sensory connotations of English compounds and the words comprising them. The study found that the connotations of the parts in general had no effect on the processing of the

compound words. Kuperman interprets this result as supporting the NDL, in which any knowledge associated with the whole word will supersede the more weakly activated conceptual associations with subparts of the word. This account correctly predicts that the gender responses for the simple and complex real word stimuli will reflect only the gender bias of the whole word. It follows that, for pseudowords, the fullest available form is the morpheme, so the gender response would reflect the morpheme, as in our results. In addition, the NDL explains the gender effect for pseudowords without recourse to decomposition, which means that it accords with our results showing no link between explicit decomposition and morpheme gender bias effects. An exception to the general pattern in Kuperman (2013) was the outcome for the dimension of emotional valency: Morphemes with negative emotional connotations contributed to slower reaction times for compound words. Kuperman interprets this results as indicating that selective attention can affect the activation of non-denotational meanings, and that humans are particularly vigilant for emotionally negative information, so that such morphemes capture attention away from the whole word. We suggest that an attentional explanation might be appropriate for the unexpected interaction we found for complex real words between whole-word gender bias and morpheme gender bias. If the words containing woman-biased morphemes are more casual and more likely to be used in face to face interactions, then speaker gender is less likely to be ignored. This would enhance encoding of gender for the whole word stimuli in this context. Stronger integration of gender information into the whole-word meaning could give rise to the observed effect that whole-word gender bias is relatively enhanced for words containing more strongly woman-biased morphemes.

To summarize, modifying any current model of the lexicon to capture our results involves ensuring that knowledge about whole words takes priority over a compositional analysis, to the extent that such knowledge is available: Not only is whole-word knowledge stronger than word-part knowledge, but word-part knowledge is superseded when whole-word knowledge is available. Given this proviso, which is often motivated independently by the existence of irregular morphological forms, the results are most readily captured by assuming that social-indexical effects in morphology operate through a general analogical mechanism. While morphological parsing is known to be relevant within the phonology and morpho-syntax, such structured processing may be confined to these parts of the linguistic system. These findings leave several avenues to explore: Attention should be paid to rarer real words, to more lower-level or faster experimental tasks, and to pseudowords made of only real morphemes. The interaction between word bias and morpheme bias points toward a new stimulus set that controls for word register and sociolinguistic context, with the exciting possibility that communication mode plays an important role in the encoding and association of indexical information with whole words and morphemes.

## Additional files

### File 1

A text file (gen\_assoc\_appendix\_A.txt) containing the complete list of stimuli for this study. The format is tab-delimited plain text, with columns for: Item (“item”), morpheme group (“morph”), real or pseudoword status (“lexicality”), simple or complex item (“complexity”), compound or suffixation (“structure”), morpheme gender bias category (“morph\_gender”), whole word gender bias category (“token\_gender”), and orthographic boundary ratio (“boundary\_cue\_ortho”). DOI: []

## Acknowledgments

This project was made possible through the support of a grant to Northwestern University from the John Templeton Foundation (Award ID 36617). The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. For assistance in stimuli preparation, data collection, and statistical analysis, the authors would like to thank Jennifer Hay, Chun Liang Chan, and Matt Goldrick. We thank the audience at LSA 2018 for their feedback.

## Competing interests

The authors have no competing interests to declare.

## References

- Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, **90**(2), 119-161. doi:10.1016/S0010-0277(03)00146-X
- Baayen, R. H., Hendrix, P., & Ramscar, M. (2013). Sidestepping the combinatorial explosion: An explanation of n-gram frequency effects based on naive discriminative learning. *Language and Speech*, **56**(3), 329-347. doi:10.1177/0023830913484896
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database (Release 2). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1-48.

- Bauer, L. (2001). *Morphological productivity*. Cambridge University Press.
- Beckner, C., Rácz, P., Hay, J., Brandstetter, J., & Bartneck, C. (2016). Participants conform to humans but not to humanoid robots in an English past tense formation task. *Journal of Language and Social Psychology*, **35**(2), 158-179. doi:10.1177/0261927X15584682
- Boulis, C., & Ostendorf, M. (2005, June). A quantitative analysis of lexical differences between genders in telephone conversations. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics* (pp. 435-442). Association for Computational Linguistics. doi:10.3115/1219840.1219894
- Bucholtz, M. (1999). "Why be normal?": Language and identity practices in a community of nerd girls. *Language in Society*, **28**(2), 203-223.
- Bucholtz, M. (2001). The whiteness of nerds: Superstandard English and racial markedness. *Journal of Linguistic Anthropology*, **11**(1), 84-100. doi:10.1525/jlin.2001.11.1.84
- Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, **10**(5), 425-455. doi:10.1080/01690969508407111
- Bybee, J., & Thompson, S. (1997). Three frequency effects in syntax. In *Annual Meeting of the Berkeley Linguistics Society* (Vol. 23, No. 1, pp. 378-388).
- The British National Corpus*, Version 3 (BNC XML Edition). 2007. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. URL: <http://www.natcorp.ox.ac.uk/>
- Clopper, C. G., & Pisoni, D. B. (2004a). Homebodies and army brats: Some effects of early linguistic experience and residential history on dialect categorization. *Language Variation and Change*, **16**(1), 31-48. doi:10.1017/S0954394504161036
- Clopper, C. G., & Pisoni, D. B. (2004b). Some acoustic cues for the perceptual categorization of American English regional dialects. *Journal of Phonetics*, **32**(1), 111-140. doi:10.1016/S0095-4470(03)00009-3
- Clopper, C. G., Tamati, T. N., & Pierrehumbert, J. B. (2016). Variation in the strength of lexical encoding across dialects. *Journal of Phonetics*, **58**, 87-103. doi:10.1016/j.wocn.2016.06.002
- Daelemans, W., Zavrel, J., Van der Sloot, K., and Van den Bosch, A. (2010). *TiMBL: Tilburg Memory Based Learner, version 6.3, Reference Guide*. ILK Research Group Technical Report Series no. 10-01.
- Daland, R., & Pierrehumbert, J. B. (2011). Learning diphone-based segmentation. *Cognitive Science*, **35**(1), 119-155. doi:10.1111/j.1551-6709.2010.01160.x
- Daland, R., Sims, A. D., & Pierrehumbert, J. (2007, June). Much ado about nothing: A social network model of Russian paradigmatic gaps. In *Annual Meeting - Association for Computational Linguistics* (Vol. 45, No. 1, p. 936).

- Dawdy-Hesterberg, L. G., & Pierrehumbert, J. B. (2014). Learnability and generalisation of Arabic broken plural nouns. *Language, Cognition and Neuroscience*, **29**(10), 1268-1282. doi:10.1080/23273798.2014.899377
- Eckert, P. (1989). The whole woman: Sex and gender differences in variation. *Language Variation and Change*, **1**(3), 245-267. doi:10.1017/S095439450000017X
- Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, **12**(4), 453-476. doi:10.1111/j.1467-9841.2008.00374.x
- Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, **34**(4), 409-438. doi:10.1016/j.wocn.2005.08.002
- German, J. S., Carlson, K., & Pierrehumbert, J. B. (2013). Reassignment of consonant allophones in rapid dialect acquisition. *Journal of Phonetics*, **41**(3), 228-248. doi:10.1016/j.wocn.2013.03.001
- Haspelmath, M., & Sims, A. (2013). *Understanding morphology* (2nd ed.). Routledge, Taylor Francis Group. London.
- Hay, J. (2002). From speech perception to morphology: Affix ordering revisited. *Language*, **78**(3), 527-555. doi:10.1353/lan.2002.0159
- Hay, J., & Baayen, H. (2001). Parsing and productivity. In *Yearbook of Morphology 2001* (pp. 203-235). Springer Netherlands.
- Hay, J. B., & Baayen, R. H. (2005). Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences*, **9**(7), 342-348. doi:10.1016/j.tics.2005.04.002
- Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, **48**(4), 865-892. doi:10.1515/ling.2010.027
- Hay, J., Pierrehumbert, J., & Beckman, M. (2004). Speech perception, well-formedness, and the statistics of the lexicon. *Papers in Laboratory Phonology VI*, 58-74.
- Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, **34**(4), 458-484. doi:10.1016/j.wocn.2005.10.001
- Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics*, **34**(4), 485-499. doi:10.1016/j.wocn.2005.08.004
- Komarova, N. L., & Nowak, M. A. (2001). The evolutionary dynamics of the lexical matrix. *Bulletin of Mathematical Biology*, **63**(3), 451-484. doi:10.1006/bulm.2000.0222
- Kuperman, V. (2013). Accentuate the positive: Semantic access in English compounds. *Frontiers in Psychology*, **4**, 203. doi:10.3389/fpsyg.2013.00203
- Labov, W. (2001). *Principles of linguistic change. Vol. 2. Social factors*. Oxford, UK: Blackwell.
- Lakoff, R. (1973). Language and woman's place. *Language in Society*, **2**(1), 45-79. doi:10.1017/S0047404500000051

- Mihalcea, R., & Garimella, A. (2016). What men say, what women hear: Finding gender-specific meaning shades. *IEEE Intelligent Systems*, **31**(4), 62-67.  
doi:10.1109/MIS.2016.71
- Nation, K., & Cocksey, J. (2009). Beginning readers activate semantics from sub-word orthography. *Cognition*, **110**(2), 273-278. doi:10.1016/j.cognition.2008.11.004
- Needle, J. M., Pierrehumbert, J. B., & Hay, J. B. (under review). Phonotactic and Morphological Effects in the Acceptability of Pseudowords.
- Nevalainen, T., Raumolin-Brunberg, H., & Mannila, H. (2011). The diffusion of language change in real time: Progressive and conservative individuals and the time depth of change. *Language Variation and Change*, **23**(1), 1-43. doi:10.1017/S0954394510000207
- Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14**(1), 54-65.  
doi:10.1037/0278-7393.14.4.700
- Nosofsky, R. M. (1990). Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology*, **34**(4), 393-418. doi:10.1016/0022-2496(90)90020-A
- Ochs, E. (1992). Indexing gender. In *Rethinking context: Language as an interactive phenomenon*, **11**, 335.
- Pierrehumbert, J. B. (2006). The next toolkit. *Journal of Phonetics*, **34**(4), 516-530.
- Pierrehumbert, J. B. (2016). Phonological representation: beyond abstract versus episodic. *Annual Review of Linguistics*, **2**, 33-52. doi:10.1146/annurev-linguistics-030514-125050
- Plag, I., Dalton-Puffer, C., & Baayen, H. (1999). Morphological productivity across speech and writing. *English Language & Linguistics*, **3**(2), 209-228.
- Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology*, **18**(1), 10-30.  
doi:10.1177/0261927X99018001002
- Quina, K., Wingard, J. A., & Bates, H. G. (1987). Language style and gender stereotypes in person perception. *Psychology of Women Quarterly*, **11**(1), 111-122. doi:10.1111/j.1471-6402.1987.tb00778.x
- R Core Team (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available from <http://www.R-project.org/>.
- Rácz, P., Hay, J. B., & Pierrehumbert, J. B. (2017). Social Salience Discriminates Learnability of Contextual Cues in an Artificial Language. *Frontiers in Psychology*, **8**.  
doi:10.3389/fpsyg.2017.00051
- Rácz, P., Pierrehumbert, J. B., Hay, J. B., & Papp, V. (2015). Morphological emergence. *The Handbook of Language Emergence*, 123-146. doi:10.1002/9781118346136.ch5

- Rickford, J. R. & Rickford, R. J. (2000) *Spoken Soul: The Story of Black English*. John Wiley and Sons.
- Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, **23**(3-4), 193-229. doi:10.1016/S0271-5309(03)00013-2
- Social Security Administration. (2016, May 5). *Top Names Over The Last 100 Years*. Retrieved from <https://www.ssa.gov/OACT/babynames/decades/century.html>.
- Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, **60**(4), 487-501. doi:10.1016/j.jml.2009.01.001
- Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology*, **57A**(4), 745-765.