A computational evaluation of gender asymmetry in semantic change

Eric Schnell (eric.schnell@mail.utoronto.ca)
Cognitive Science and Mathematics Programs
University of Toronto

Yang Xu (yangxu@cs.toronto.edu)
Department of Computer Science
Cognitive Science Program
University of Toronto

Abstract
A fundamental goal in cognitive and historical linguistic research on semantic change is to characterize the regularity in how word meanings change over time. We examine a common belief that has not yet been evaluated comprehensively, which asserts that gender of a word influences its direction of semantic change. By this account, female terms like mistress should undergo pejorative change in meaning systematically more so than male terms like master. We evaluate this claim in gender-marked word pairs in English and French respectively as languages without and with grammatical gender. Our results provide supporting evidence for gender asymmetry in semantic change of English words but not French words. Our study raises questions about the generality of the claim about gender asymmetry in semantic change and provides a scalable computational framework for understanding the social roots of word meaning change.

Keywords: word meaning; semantic change; gender and language; sentiment analysis

Introduction
A theory of word meaning should explain not only how meanings are structured, but also how they vary. One dimension along which word meanings vary is time. Commonly known as semantic change or semantic shift (Bréal, 1897), a word’s meaning may change in many different ways. A fundamental goal in cognitive and historical linguistic studies on semantic change has been to characterize the regularity in semantic change (Ullmann, 1962; Sweetser, 1991; Geeraerts, 1997; Traugott & Dasher, 2001). Here we examine a common belief about gender and semantic change that has not yet been evaluated comprehensively. That is, whether gender of a word influences its direction of semantic change.

Semantic change refers to the phenomenon that meaning of a word changes over time. Extensive research from historical linguistics has classified semantic change into different types (e.g., Ullmann, 1962; Traugott & Dasher, 2001). Some prominent types include narrowing (e.g., deer: “animal”→“a ruminant animal”), broadening (e.g., Kleenex: “Kleenex tissue”→“any tissue”), amelioration (e.g., pretty: “cunning”→“attractive”), and pejoration (e.g., mistress: “housewife”→“woman having an affair”). Related work from cognitive linguistics has sought to characterize the regularity or general principles that underlie semantic change. Some representative theories have considered the cognitive and functional principles. For instance, it has been suggested that word meaning change is not arbitrary and reflects the cognitive principle of prototypicality, whereby more prototypical meanings of a word tend to be more stable over time in comparison to peripheral meanings (Geeraerts, 1997). Independently, it has been argued that functional principles such as efficient communication might have shaped the grammaticalization of word meaning (Hopper & Traugott, 2003).

Gender asymmetry in semantic change
We investigate whether social factors might influence the regularity in semantic change. In particular, we examine the role of gender or more specifically gender bias in shaping the directionality of word meaning change. Our starting point is the common belief that there exists an asymmetry in the direction of semantic change for female (or feminine) terms and male (or masculine) terms: Terms used to describe women tend to undergo pejorative semantic change, such that they become increasingly associated with a negative sentiment over time, whereas the same effect might not be observed in masculine terms (Kochman-Haladyj & Kleparski, 2010; Ng, Chan, Weatherall, & Moody, 1993).

The idea of gender asymmetry in semantic change is relevant not only from a linguistic perspective, but also for social and psychological reasons. It has been shown that language, and specifically the way in which gender is spoken about in English, reinforces bias (Bigler & Leaper, 2015). It has also been shown that gender usage in English has psychological effects on English speakers, which imposes gender stereotypes to its users. For instance, children form gender stereotypes using language as a basis (Bigler & Liben, 2006). If there exists a strong tendency in the pejoration of female terms but not male terms, then this regularity would have broad implications for the continued existence of gender bias perpetuated by language.

Figure 1 illustrates the asymmetry with the gendered word pair cow and bull in English and the equivalent pair vache and taureau in French. In early 1800s, both words were roughly on par in terms of their sentiment inferred from natural language use with cow having a slightly more positive sentiment. Over the course of 200 years, both cow and bull became increasingly more negative in meaning, but cow experienced a substantially larger decrease in sentiment than bull making it now the more negative sentiment of the two. Although gender asymmetry in semantic change has been confirmed in these anecdotal cases, to our knowledge it has not been com-
prehensively evaluated in English or other languages. One challenge for a scaled evaluation is that it has been difficult to quantify the sentiment of a word’s meaning historically. This is especially the case when dealing with implicit or connotational meaning, where a word’s definition may not be fully reflective of its sentiment at a certain time.

Recent advances in distributed semantic representations of word meaning have provided support that broad gender biases are reflected in natural language use (Bolukbasi, Chang, Zou, Saligrama, & Kalai, 2016; Caliskan, Bryson, & Narayanan, 2017). These findings suggest that gender biases can be captured and measured by computational means in large textual corpora (e.g., digitized books and news). Using similar techniques, we quantify the degree to which gender bias may influence the directionality of semantic change. We draw on computational methods to measure the sentiment of words automatically at scale (Turney & Littman, 2003). Furthermore, by using historical text corpora, we track the directionality of semantic change in words over a long temporal span (Cook & Stevenson, 2010) by measuring whether a word undergoes amelioration, pejoration, or neither (see Figure 1).

**Scope and hypothesis**

Socio-linguistic research has found that language use can be biased, such that people can associate words implicitly with different groups following group stereotypes (Banaji & Hardin, 1996). One such example is that words relating to science are associated more often with men, whereas words relating to art are associated more often with women, reflecting gender stereotypes (Nosek, Banaji, & Greenwald, 2002). Psychological work has also shown that as a word becomes more associated with a particular gender, it tends to adopt the stereotypes of that gender (Ng et al., 1993). This suggests that the change in sentiment of a gendered word could be the result of a change in that word’s gender alignment rather than the direct result of a word’s fixed gender. To prevent the change in a word’s gender alignment from influencing our evaluation, we focus on analyzing words that belong to a fixed gender pair. This is to say that we only consider pairs of words which have the same meaning, but have overt forms that distinguish the two gender classes. In English, although gender marking in word forms is not obligatory, there exists a relatively large set of noun pairs that do make a gender distinction (e.g., *actress* vs. *actor*).

To evaluate the generality of the proposal on gender asymmetry in semantic change, we also consider an analysis in French which has an explicit grammatical gender system. For instance, every noun and adjective in French is assigned to a gender (e.g., *sorcière* and *sorcier* correspond to the feminine and masculine forms of the concept “witch” respectively).

If the hypothesis about gender asymmetry holds generally, we should expect that the sentiment of female terms in both English and French to become more negative over history. We do not make any predictions on the directionality of sentiment change for male terms, although we do predict that the change in sentiment of male terms should be overall more positive than that of female terms. In particular, even if male terms decrease in sentiment over time, they should do so to a lesser degree than the female counterpart terms.

**Computational methodology**

To quantify the sentiment of a word, we use the Sentiment Orientation – Pointwise Mutual Information (SO-PMI) formulation, which is an established method in computational linguistics described in Turney and Littman (2003). We use Positive Pointwise Mutual Information (PPMI), a vari-

---

Figure 1: An illustration of asymmetry in historical sentiment change of gendered words in English (left) and French (right). Solid arrows indicate temporal changes in sentiment between consecutive decades during a period of 200 years (1800–2000). The dots indicate the starting and ending positions of a word in sentiment space.
ant of Pointwise Mutual Information (PMI), that represents the meaning of a word as a non-negative vector based on its co-occurrence with context words in text. The formula for calculating sentiment orientation then follows the method in Turney and Littman (2003):

$$\text{SO} - \text{PPMI}(\text{word}) = \sum_{\text{pwords} \in \text{Pwords}} \text{PPMI}(\text{word}, \text{pword}) - \sum_{\text{nwords} \in \text{Nwords}} \text{PPMI}(\text{word}, \text{nword})$$

(1)

Here \(\text{Pwords}\) represents the set of positive seed words (i.e., words with canonically positive sentiment such as \textit{good}), \(\text{Nwords}\) represents the set of negative seed words (i.e., words with canonically negative sentiment such as \textit{bad}), and \(\text{PPMI}(\text{word}, \text{word})\) equals the Positive Pointwise Mutual Information between two words, calculated identically to Hamilton, Leskovec, and Jurafsky (2016) which includes historical data for our analysis (described later).

We take the seed words in English from the General Inquirer (Stone, Littman, Namenwirth, & Ogilvie, 1962). In French, we use Polarimots (Gala & Brun, 2012). We use General Inquirer in English because it is standard in sentiment research and was used in a prior study that explores historical semantic orientation change (Cook & Stevenson, 2010). We use Polarimots in French since it includes only single-word entries and is available to the general public. Both seed sets can include a word multiple times if it has multiple meanings. These duplicated entries can either be of the same sentiment or of opposite sentiments. We remove duplicates of the same sentiment and if a word has an entry in both the positive and negative sentiment, because these cancel out in the formula above and hence has no effect on the sentiment orientation.

In all our analyses, we quantify how the sentiment of a word changes over time. To calculate a word’s change in sentiment, we take the difference between the word’s sentiment between the final and starting periods (in our case, these correspond to decades 1990s and 1800s respectively).

**Data**

To evaluate our hypothesis, we collect data from two sources: 1) historical word statistics that allow us to quantify sentiment change of words based on their context over the past 200 years; 2) a control set of paired gendered words, such as \textit{(actress, actor)} in both English and French.

**Historical word usage statistics**

We use data from Hamilton et al. (2016) (https://nlp.stanford.edu/projects/histwords/) as our basis for investigating sentiment change in English and French words. Both the English and the French datasets are constructed from the Google Book Ngram corpus (https://books.google.com/ngrams). Hamilton et al.’s (2016) data contains multiple statistics related to word usage, particularly PPMI scores between each pair of words in the dataset. This dataset is also historical and spans from 1800 to 2000, and it is binned by decade. This allows us to study the change in a word’s sentiment over a 200-year temporal span. For our analyses of sentiment change, words must occur at least 500 times within a decade. This is the same threshold used by Hamilton et al. (2016) in their historical analysis of semantic change.

**Gendered word pairs**

We took the gendered word pairs analyzed in English from Zhao, Zhou, Li, Wang, and Chang (2018). This list is comprised of nouns and includes both singular and plural forms of words. We omitted the plural forms to ensure that we were not double counting results for a pair. This list is not intended to be a definitive list of all gendered pairs in English, but rather representative enough of the phenomenon to properly test behaviour. We translated the English pairs into French using the Larousse English-French dictionary (Larousse Dictionnaire Anglais-Français, 2020). Some words did not have direct translations, so these were removed. To ensure our dataset properly represented French gendered noun pairs, we also included the 100 most common such pairs based on the numbers of entries in books (New, Pallier, Brysbaert, & Ferrand, 2004). In addition, we also analyzed a list of French ad- djective pairs which similarly distinguish between masculine and feminine forms. For adjectives we considered the 200 most common pairs (New et al., 2004). Table 1 shows a sample of gendered word pairs in English and French that span the spectrum from occupational concepts to other human and non-human concepts.

<table>
<thead>
<tr>
<th>Occupation</th>
<th>English</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actor–Actress</td>
<td>Auteur–Actrice</td>
<td></td>
</tr>
<tr>
<td>Waiter–Waitress</td>
<td>Serveur–Serveuse</td>
<td></td>
</tr>
<tr>
<td>Policeman–Policewoman</td>
<td>Policier–Policière</td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Son–Daughter</td>
<td>Fils–Fille</td>
<td></td>
</tr>
<tr>
<td>Fatherhood–Motherhood</td>
<td>Paternité–Maternité</td>
<td></td>
</tr>
<tr>
<td>Heir–Heiress</td>
<td>Héritier–Héritière</td>
<td></td>
</tr>
<tr>
<td>Non-human</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bull–Cow</td>
<td>Taureau–Vache</td>
<td></td>
</tr>
<tr>
<td>Rooster–Hen</td>
<td>Coq–Poule</td>
<td></td>
</tr>
<tr>
<td>Stallion–Mare</td>
<td>Étalon–Jument</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Sample gendered pairs in English and French.

**Results**

We evaluate the hypothesis about gender asymmetry in semantic change in three sets of words: 1) English gendered noun pairs, 2) French gendered noun pairs, and 3) French gendered adjective pairs.

**Evidence for gender asymmetry in English**

Table 2 summarizes the results and detailed statistics from our analysis of the English word pairs. These results indicate that there is a significant split between the average change in a word’s sentiment across the gender dimension \(p < 0.02\)
from a permutation test), with the female words exhibiting more negative sentiment than the male words over time. In particular, we find that sentiment of male words generally increases from the time of their first entry in the dataset to their last entry. However, the average sentiment of female words decreases during this time period (see Table 2). When we separate our data by word gender, this difference in average sentiment change shows a significant difference between the two gender groups. One thing to note about these results is the uneven dataset sizes for male and female words. Here we are concerned with all possible words from our list, which means that in the case where one word in a pair does not meet the inclusion criteria but the other one does, we still include the qualified word. For this reason, we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summarizes this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summarizes this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summarizes this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summarizes this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summarizes this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summarizes this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summarizes this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs. Table 3 summaries this paired analysis. We observe that our previous results hold similarly in this controlled analysis, where we perform the same analysis except here we only consider pairs where both words match the inclusion criteria. In this analysis we include words multiple times if they belong to multiple pairs.
Evidence against gender asymmetry in French

To further evaluate whether the hypothesis holds generally across languages, we repeat the same analysis with the French word pairs as in the case of English, except that we separate the results for nouns and adjectives. Tables 7 - 12 summarize these results. Although the French datasets are comparable in sample size as the English dataset, we observe no significant results that support the hypothesis about gender asymmetry in semantic change of French words. If anything, we find the opposite effect (with no significance at $\alpha = 0.05$) in the case of French nouns, where the male nouns appear to show less positive sentiment change on average than the female nouns.

These results provide evidence against the hypothesis, but we performed the final analysis on a more focused set. Recent work has suggested that the effect of gender bias is present most prominently on occupation terms across languages including French (Lewis & Lupyan, 2020). Based on this observation, we redid our analysis of French but only on word pairs that concern people’s occupations. This focused analysis led to results closer to the hypothesis (e.g., female occupation terms show on average a negative trend in sentiment, whereas the male occupation terms show on average a positive trend, as shown in Table 13), but we still observe no statistical significance ($p = 0.1494$ in a paired t-test) in this reduced sample of word pairs.

In French, we also examine the regression towards the mean by separating the pairs into two sets depending on if the male term or the female term is more positive initially. We find a similar effect of regression towards the mean as is found in English (see Tables 14 and 15; note we only show the results for nouns but they are similar for adjectives). Unlike in English, this effect is significant both when the male term is initially more positive and when the female term is initially more positive. This may explain why our overall findings in French are null, because the effect of regression towards the mean is strong in both directions. This effect might explain the difference in results between English and French, namely the reason for the lack of gendered role on semantic change in French.

Overall, this set of analyses suggests that the gender asymmetry observed in semantic change of English words does not hold for French, which is a grammatically gendered language.

Discussion

We have presented a scalable computational approach to evaluate the hypothesis that gender of a word influences its direction of semantic change (Kochman-Haładyj & Kleparski, 2010; Ng et al., 1993).

In English, our results support the hypothesis. There is a significant relationship between a word’s gender and the direction of its sentiment change. In fact, male words appear to increase in sentiment over time and female words ap-
<table>
<thead>
<tr>
<th>Dataset</th>
<th>n</th>
<th>Average</th>
<th>SD</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>25</td>
<td>1.3529</td>
<td>4.1479</td>
<td>0.1494</td>
</tr>
<tr>
<td>Female</td>
<td>25</td>
<td>0.0576</td>
<td>4.7116</td>
<td></td>
</tr>
</tbody>
</table>

Table 13: Average sentiment change pairs - French Occupation Nouns. p-value is calculated by a one-tailed paired t-test.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>n</th>
<th>Average</th>
<th>SD</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>54</td>
<td>2.3552</td>
<td>5.5858</td>
<td>0.0015*</td>
</tr>
<tr>
<td>Female</td>
<td>54</td>
<td>0.0355</td>
<td>4.4467</td>
<td></td>
</tr>
</tbody>
</table>

Table 14: Average sentiment change when female is more positive - French Nouns. p-value is calculated by a one-tailed paired t-test.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>n</th>
<th>Average</th>
<th>SD</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>46</td>
<td>-1.0682</td>
<td>4.2423</td>
<td>0.0066*</td>
</tr>
<tr>
<td>Female</td>
<td>46</td>
<td>1.6327</td>
<td>4.6922</td>
<td></td>
</tr>
</tbody>
</table>

Table 15: Average sentiment change when male is more positive - French Nouns. p-value is calculated by a two-tailed paired t-test.

...
Acknowledgments
We thank Angela Li for research assistance. YX is funded through an NSERC Discovery Grant, a SSHRC Insight Grant, and an Ontario ERA award.

References