Zooming in on Gender Differences in Social Media

Aparna Garimella and **Rada Mihalcea**
University of Michigan
{gaparna,mihalcea}@umich.edu

Abstract

Men are from Mars and women are from Venus - or so the genre of relationship literature would have us believe. But there is some truth in this idea, and researchers in fields as diverse as psychology, sociology, and linguistics have explored ways to better understand the differences between genders. In this paper, we take another look at the problem of gender discrimination and attempt to move beyond the typical surface-level text classification approach, by (1) identifying semantic and psycholinguistic word classes that reflect systematic differences between men and women and (2) finding differences between genders in the ways they use the same words. We describe several experiments and report results on a large collection of blogs authored by men and women.

1 Introduction

Previous work on understanding gender differences has mainly focused on the authorship detection facet, trying to identify the gender of the author of a certain writing, be that a blog (Mukherjee and Liu, 2010), a tweet (Burger et al., 2011), or other works of fiction or non-fiction (Koppel et al., 2002). In this paper, we depart from this earlier research and attempt to move beyond the surface level of word occurrences and counts. We instead use semantic analysis to identify broad semantic classes that are specific to each gender, and also find differences that exist between genders in how they use certain concepts.

Specifically, the paper addresses the following two main questions. First, can we identify broad semantic and psycholinguistic classes that are predominantly used by men and women? We use linguistic ethnography in conjunction with three different resources and determine gender saliency scores associated with predefined word classes, which we can use to better understand the groups of words that differentiate men's and women's language use.

Second, can we distinguish between shades of word meanings, as used by the two genders? Do men and women use the word "car" in a similar way, or are there differences between the use of this word in their day-to-day life? We answer this question by using a word sense disambiguation framework, where each gender is regarded as a "sense," and we detect the gender corresponding to a given occurrence of a word. Using a large dataset of over 350 words, we show that gender-based word disambiguation is possible, and that there are indeed differences between the ways certain words are used by men and women.

2 Related Work

Field work in social and gender psychology has had much to say about the differences between men and women. The masculine is stereotyped as detached, rational, and aggressive, and the feminine as nurturing, gentle, and tactful (Doyle, 1985). While some stereotypes are unfounded, sociolinguists do affirm that some communication styles are gendered. It has been found for instance that men and women differ on private versus public speaking, on "report talk" versus "rapport talk"–these and other facets of relational dialectics are gendered and constitute so-called "GenderLens" (Tannen, 1991).

Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media, pages 1–10, Osaka, Japan, December 12 2016.

One of the earliest studies concerned with the language differences between men and women is due to (Lakoff, 1972), who found several characteristics of women language, including words such as "lovely" and "adorable", or phrases such as "it seems to be" or "would you mind." There is also a large body of work on the connection between language and gender in the field of sociology (e.g., (Eckert and McConnell-Ginet, 2003)), which we do not address here due to the lack of space.

In computational linguistics, several studies addressed the role of gendered language and the "gender gap" in the blogosphere (Kennedy et al., 2005; Koppel et al., 2002; Schler et al., 2006; Mukherjee and Liu, 2010); the significance of gender differences in self-disclosure strategy in teenage blogs (Huffaker and Calvert, 2005); and the validity of author gender predictions based largely on function words (e.g. pronouns, determiners) (Herring and Paolillo, 2004). Several previous studies made use of LIWC (Pennebaker and Francis, 1999) categories to investigate gender differences in writing styles and content in blogs (Nowson and Oberlander, 2006; Schler et al., 2006). Work has also been done on Twitter data, where tweets are used to predict several profile features, including gender (Rao et al., 2010; Burger et al., 2011). In (Peersman et al., 2004), age and gender prediction is performed on short messages from social networking sites. The focus in these previous studies has been primarily on investigating the use of automatic classification to distinguish between men and women writings, and also on finding words that are specific to each gender by performing statistical analysis on large amounts of data.

Other related work includes recently published research by (Nguyen et al., 2014), who showed how a person's gender identity can be constructed by using various linguistic aspects of male and female speech in language. (Gianfortoni et al., 2011) used pattern-based feature creation approach in combination with word classes to classify author's gender from blog posts. Also of interest is the work by (Prabhakaran et al., 2014), who use topic segments to predict the behavioral patterns of political leaders in election campaigns. Our work is to some extent related to that research, as we also seek to understand and model behaviors from text, however we do this for men and women rather than political figures.

In speech, an analysis of the most frequently used words by males and females in telephone conversations was presented in (Boulis and Ostendorf, 2005), who found that swear words are more often used by males (bullshit, sucks, damn), whereas family-relation terms are more often used by females (children, marriage, boyfriend).

One exception from the general theme of previous work on surface-level gender classification is the work by (Sarawgi et al., 2011), where topic bias is explicitly avoided, with the goal of identifying stylistic differences between men and women writings. The authors use blogs addressing predefined topics (e.g., education, travel) and scientific publications, and show that differences can be found even when the data sources are controlled for topic. In our research, we zoom in even deeper, and try to identify semantic and psycholinguistic word classes that characterize gender differences, and also find the distinctive ways in which men and women use certain words.

2.1 Data

We use a large corpus of blogposts annotated for gender, which we collected from the *Blogspot* (http://www.blogspot.com) community (Liu and Mihalcea, 2007). We chose to use *Blogspot* as opposed to other blog communities such as *LiveJournal* because it has richer blogger profile annotations including gender, age, location, occupation, and others. The kind of writing found in a weblog is ideally suited to what we wish to discover, since weblogs often give an intimate account of personal everyday life, and personal viewpoint unto current events. More than just language and syntax, weblogs contain ample evidence of experiences and perceptions, which we attempt to uncover using corpus-based modeling and semantic analysis.

Starting with the names of approximately 300,000 blogs that were updated with a new entry during the time when the crawling was performed, [1] we collected the profile page of the blog owners (bloggers) and the corresponding profile features. We discarded all the blogs maintained by more than one blogger (collective blogs), and we also discarded the blogs corresponding to bloggers who chose not to include gender information in their profile. Finally, we parsed the entries from the remaining set of blogs, and

[1] The blogs were crawled in summer 2006.

retained only the blogposts written in English and having a length within a 200–4,000 character limit. Interestingly, although a large fraction of the blogs listed on *Blogspot* are spam, the constraints that a blogger have a profile and that the size of a blogpost be within certain limits removed almost all the spam – to the point that a random hand-check of 100 blogposts revealed clean spam-free data.

The post-processing and profile-based filters left us with a total of 160,000 blog entries annotated for gender, which after balancing between male and female authors, left us with the final set of 75,000 male blog entries and 75,000 female blog entries. It is to be noted that the blog data is not balanced across different genres, as we expect any existing genre-imbalance to convey some information about the interests of different genders. Table 1 shows two sample entries written by a male and a female writer. Table 2 shows three unigrams, bigrams, and trigrams with high probability in these blogs.

Male-authored blogpost
No word back from the Georges Island people on possible use of their power so I'm going to proceed with the QRP plans. Even though the QRP stuff is smaller than the 100 watt outfit, there will still be a significant amount of stuff I'll need to wrestle on to the island. I'll bring the Pelican 1510 case outfitted with the Elecraft K 2.
Female-authored blogpost
You could probably tell that I literally enjoy dressing up in costumes and crap. I just don't have the resources nor the skills to make a good costume. But I'm a resource for outlandish ideas. I remember shocking my host dad when I told him that I enjoy dressing up like that.

Table 1: Male and female authored blogposts

	Female	Male
Unigrams	knitting	microsoft
	hubby	democrats
	yarn	poker
Bigrams	my husband	my wife
	love him	of Israel
	so excited	prime minister
Trigrams	I love him	my wife and
	so much fun	of the United
	I miss my	the Bush administration

Table 2: Unigrams/bigrams/trigrams with high probability in men/women language

3 Gender Dominant Semantic and Psycholinguistic Word Classes

Can we move beyond the word-based discrimination of men and women writings, and find semantic patterns in word usage? Given a set of semantic and psycholinguistic word classes, we calculate a score associated with each word class, and consequently identify in a principled manner the word classes that are salient in each gender.

3.1 Calculating Word Class Saliency

We calculate the saliency of a word class using the distribution of occurrences for words belonging to the class inside the men and women writings. Given a class of words $C = \{W_1, W_2, ..., W_N\}$, we define the class coverage in the female corpus F as the percentage of words from F belonging to the class C:

$$Coverage_F(C) = \frac{\sum\limits_{W_i \in C} Frequency_F(W_i)}{Size_F}$$

where $Frequency_F(W_i)$ represents the total number of occurrences of word W_i inside the corpus F, and $Size_F$ represents the total size (in words) of the corpus F.

Similarly, we define the class C coverage for the male corpus M:

Resource	Class	Score	Sample words
		Female	
LIWC	GROOM	1.74	cleaner, washer, perfume, shaved, shampoo, cleansing, soap, shower, toothpaste
LIWC	SLEEP	1.65	tiresome, sleeping, dazed, sleeps, insomnia, napping, siesta, nightmare, dreams
LIWC	I	1.52	me, myself, my, mine, I
LIWC	FAMILY	1.51	auntie, mommy, nephews, parents, daughter, motherhood, grandma, wives, cousin
LIWC	EATING	1.46	fat, dinner, tasting, drunken, fed, breakfast, cookie, eats, tasted, skinny, cookbook
WA	DISGUST	1.59	sickening, revolting, horror, sick, offensive, obscene, nauseous, wicked, offensive
WA	FEAR	1.23	suspense, creep, dismay, fright, terrible, terror, afraid, scare, alarmed, panicked
Roget	SEWING	3.46	mending, stitching, knitter, mend, tailor, suture, embroidery, seamstress, needle
Roget	PURPLENESS	1.87	purple, mauve, magenta, lilac, lavender, orchid, violet, mauve, mulberry, purply
Roget	SWEETNESS	1.80	syrup, honey, sugar, bakery, nectar, sweet, frost, sugary, dessert, glaze, nut
Roget	BROWNNESS	1.45	coffee, biscuit, walnut, rust, berry, brown, brunette, cinnamon, mahogany, caramel
Roget	CHASTITY	1.38	shame, elegant, decent, virtue, virgin, delicate, faithfulness, platonic, purity, spotless
		Male	
LIWC	RELIG	1.47	bless, satanism, angel, communion, spirit, lord, faithful, immortal, theology, prayers
LIWC	METAPH	1.43	suicide, meditation, cemetary, temples, drained, immortalized, mercy, mourning
LIWC	SPORTS	1.41	running, jogged, pool, basketball, swimming, exercise, fitness, teams, aerobic
LIWC	TV	1.39	show, ad, comedies, tv, actors, drama, soaps, video, theaters, commercials, films
LIWC	JOB	1.30	credentials, department, financials, desktop, manage, employ, work, career
Roget	OPONENT	1.88	finalist, rival, enemy, competitor, foe, opposite, defendant, player, dissident
Roget	THEOLOGY	1.88	creed, scholastic, religious, secularism, theology, religion, divine, faith, dogma
Roget	UNIFORMITY	1.88	evenness, constancy, persistence, accordance, steadiness, firmness, stability
Roget	ENGINEERING	1.60	automotive, process, industrial, manufacture, measure, construction, technician
Roget	INFLUENCE	1.60	power, force, weak, weakness, inflexible, ineffective, charisma, charm, wimpy

Table 3: Sample dominant word classes in male and female blogs.

$$Coverage_M(C) = \frac{\sum\limits_{W_i \in C} Frequency_M(W_i)}{Size_M}$$

The *dominance score* of the class C in the female corpus F is then defined as the ratio between the coverage of the class in the corpus F with respect to the coverage of the same class in the male corpus M:

$$Dominance_F(C) = \frac{Coverage_F(C)}{Coverage_M(C)} \tag{1}$$

A dominance score close to 1 indicates a similar distribution of the words in the class C in both the female and the male corpora. Instead, a score significantly higher than 1 indicates a class that is dominant in the female corpus, and thus likely to be a characteristic of the texts in this corpus. In a similar way, we define the $Dominance_M(C)$ score as the ratio between $Coverage_M(C)$ and $Coverage_F(C)$, where a score significantly higher than 1 indicates a class that is salient in the male corpus.

3.2 Word Classes

We use classes of words as defined in three large lexical resources: Roget's Thesaurus, Linguistic Inquiry and Word Count, and the six main emotions from WordNet Affect. For each lexical resource, we only keep the words and their corresponding class. Note that some resources include the lemmatised form of the words (e.g., Roget), while others include an inflected form (e.g., LIWC); we keep the words as they originally appear in each resource. Any other information such as morphological or semantic annotations is removed for consistency purposes, since not all the resources have such annotations available.

Roget. Roget is a thesaurus of the English language, with words and phrases grouped into hierarchical classes. A word class usually includes synonyms, as well as other words that are semantically related. Classes are typically divided into sections, subsections, heads and paragraphs, allowing for various granularities of the semantic relations used in a word class. We only use one of the broader groupings, namely the heads. The most recent version of Roget (1987) includes about 100,000 words grouped into nearly 1,000 head classes.

Linguistic Inquiry and Word Count (LIWC). LIWC was developed as a resource for psycholinguistic analysis (Pennebaker and Francis, 1999; Pennebaker and King, 1999). The 2001 version of LIWC includes about 2,200 words and word stems grouped into about 70 broad categories relevant to psychological processes (e.g., emotion, cognition). The LIWC lexicon has been validated by showing significant correlation between human ratings of a large number of written texts and the rating obtained through LIWC-based analyses of the same texts.

WordNet Affect (WA). WA (Strapparava and Valitutti, 2004) is a resource that was created starting with WordNet (Miller et al., 1993), by annotating synsets with several emotions. It uses several resources for affective information, including the emotion classification of Ortony (Ortony et al., 1987). We build an affective lexicon by extracting the words corresponding to the six basic emotions defined by (Ortony et al., 1987), namely anger, disgust, fear, joy, sadness, and surprise.

3.3 Gender Dominant Word Classes

Applying the word class saliency metric on the blog dataset using the three resources described before results in a score associated with each class. The following word classes were found to be dominant in either the female corpus or the male corpus, with a score that is away from the neutral score of 1 by a margin larger or equal to 0.20.

Roget. Female: SEWING (3.46), PURPLENESS (1.87), SWEETNESS (1.8), BROWNNESS (1.45), ORANGENESS (1.45), CHASTITY (1.38), TOUCH (1.38), ASCETICISM (1.37), FASTING (1.37), SPELL CHARM (1.37), SEMILIQUIDITY (1.35), PREDICTION (1.34), ENVY (1.34), BLUENESS (1.31), PULPINESS (1.31), SOURNESS (1.31), RAIN (1.29), GREENNESS (1.29), SENSATIONS OF TOUCH (1.29), ROUGHNESS (1.29), RECESSION (1.27), FORESIGHT (1.27), EVILDOER (1.26), TEXTURE (1.25), REFRIGERATION (1.24), REDNESS (1.23), SELFISHNESS (1.23), VIRTUE (1.23), INSOLENCE (1.22), RESINS GUMS (1.22), COURTESY (1.22), UNORTHODOXY (1.22), ONENESS (1.22), UNINTELLIGIBILITY (1.21), MATHEMATICS (1.2), CLOTHING MATERIALS (1.2), SECRETION (1.2), OVERESTIMATION (1.2) Male: THEOLOGY (1.88), OPPONENT (1.88), UNIFORMITY (1.88), UNSANCTITY (1.75), ENGINEERING (1.60), INFLUENCE (1.60), MISSILERY (1.60), PROHIBITION (1.58), QUADRUPLICATION (1.58), INSIPIDNESS (1.56), PHRASE (1.51), IDOLATRY (1.51), PRECEPT (1.49), ELECTRONICS (1.49), MISTEACHING (1.49), RELIGIONS CULTS SECTS (1.43), BODY OF LAND (1.43), PUBLIC SPIRIT (1.43), MECHANICS (1.43), ILLEGALITY (1.41), ETHICS (1.41), PREJUDGMENT (1.40), THIEF (1.39), LAND (1.34), UNITED NATIONS INTERNATIONAL ORGANIZATIONS (1.34), INORGANIC MATTER (1.34), PRECURSOR (1.34), FUEL (1.34), EARTH SCIENCE (1.33), WISE PERSON (1.33), AVIATOR (1.33), ARCHITECTURE DESIGN (1.31), MERCHANDISE (1.31), TRIBUNAL (1.30), DISCORD (1.30), TREATISE (1.28), ROCK (1.28), REVOLUTION (1.28), FOUR (1.28), REGION (1.26), TEACHER (1.26), NONRELIGIOUSNESS (1.26), FICTION (1.25), COUNTRY (1.25), LETTER (1.25)

LIWC. Female: GROOM (1.74), SLEEP (1.65), I (1.52), FAMILY (1.51), NONFL (1.48), EATING (1.46), SELF (1.44), POSFEEL (1.36), HOME (1.36), FEEL (1.34), FRIENDS (1.33), PHYSICAL (1.33), SEXUAL (1.31), PRONOUN (1.29), ASSENT (1.27), BODY (1.23), SIMILES (1.22) Male: RELIG (1.47), METAPH (1.43), SPORTS (1.41), TV (1.39), JOB (1.30).

WA: Female: DISGUST (1.25), FEAR (1.23)

Table 3 shows several salient word classes along with sample words belonging to these classes.

A few interesting observations can be made. First, there are indeed word classes, both semantic and psycholinguistic, which are dominant in one gender. While previous work has mainly focused on identifying individual words that have high frequencies in either men's or women's writings, our method allows us to identify patterns over these differences in the form of linguistically justified word classes.

Among the semantic word classes from Roget, many of the ones found to be dominant for women refer to sensorial concepts, e.g., PURPLENESS, GREENNESS, SWEETNESS, TOUCH, SOURNESS, TEXTURE, etc., which suggests that women have an increased sense of perception of the surrounding world. The ones that are predominant for men reflect a concern with religion, e.g., PUBLIC SPIRIT, THEOLOGY, RELIGIONS CULTS SECTS, or science and engineering, e.g., ARCHITECTURE DESIGN, AVIATOR, EARTH SCIENCE, INORGANIC MATTER, ENGINEERING.

In terms of psycholinguistic classes (LIWC), women appear to be more interested in family, e.g., FAMILY, HOME, FRIENDS and personal well being, e.g., GROOM, SLEEP, SELF, BODY, whereas men seem to be more interested in RELIGION, SPORTS, and JOB related topics.

Perhaps not surprisingly, among the WordNet Affect word classes, there are no emotions that are dominant for men. Instead, two emotions, DISGUST and FEAR, are salient for women.[2]

[2]All the other emotions had a $Dominance_F(C)$ score higher than 1 (even if below 1.20), which is probably justified by

4 Gender-based Word Disambiguation

We now turn to our second question, which is concerned with whether some words are used differently by men and women, which can be regarded as a reflection of the differences in how they see the world around them. To test our hypothesis, we use examples drawn from men's and women's writings for a large number of words, and build disambiguation models centered on these target words. We are therefore formulating our task as a word sense disambiguation problem, and attempt to automatically identify the gender of the person using a certain target word.

4.1 Target Words

The choice of target words for our experiments is driven by the phenomena we aim to analyze. Because we want to investigate the behavior of words in the language of the two genders, and verify whether the difference in word behavior comes from changes in sense or changes in wording in the context, we choose a mixture of polysemous words and monosemous words (according to WordNet 3.0 (Miller, 1995)), and also words that are frequent in the writings of both genders, as well as words that are frequently used by only one gender.

According to these criteria, for each open class (nouns, verbs, adjectives, adverbs) we select 100 words, 50 of which have multiple senses, and 50 with one sense only. Each of these two sets has a 30-10-10 distribution: 30 words that are frequent in both men and women writings, with a distribution in the two genders falling in the [40%-60%] range, and 10 words per each gender such that these words are only frequent in one gender (i.e., words that have a frequency for the dominant gender higher than 70%).

The initial set of target words consists of 400 open class words, uniformly distributed over the 4 parts of speech, uniformly distributed over multiple-sense/unique sense words, and with the frequency based sample as described above. From this initial set of words, we could not identify enough examples for 36,[3] which left us with a final set of 364 words.

4.2 Data Preprocessing

For each target word in our dataset, we collect 300 examples from each gender, for a maximum of 600 examples per target word. The average number of examples is 492 examples per target word.

All the extracted snippets are then processed: the text is tokenized and part-of-speech tagged using the Stanford tagger (Toutanova et al., 2003), and contexts that do not include the target word with the specified part-of-speech are removed. The position of the target word is also identified and recorded as an offset along with the example.

4.3 Gender Disambiguation Algorithm

The classification algorithm we use is inspired by previous work on data-driven word sense disambiguation. Specifically, we use a system that integrates both local and topical features. The *local features* include: the current word and its part-of-speech; a local context of three words to the left and right of the ambiguous word; the parts-of-speech of the surrounding words; the first noun before and after the target word; the first verb before and after the target word. The *topical features* are determined from the global context and are implemented through class-specific keywords, which are determined as a list of at most five words occurring at least three times in the contexts defining a certain word class (or epoch). The features are then integrated in a Naive Bayes classifier. The final disambiguation system is similar to several word sense disambiguation systems described in previous work (Dandala et al., 2013).

For evaluation, we calculate the average accuracy obtained through ten-fold cross-validations applied on the data collected for each word. To place results in perspective, we also calculate a simple baseline, which assigns the most frequent class by default.

---

the more emotional nature of women.

[3] A minimum of 100 total examples was required for a word to be considered in the dataset.

4.4 Results and Discussion

Table 6 summarizes the results obtained for the 364 words.[4] Overall, we find that there are indeed differences between the ways men and women use predefined target words, with an average error rate reduction of 7.64%. While improvements are obtained for all parts-of-speech, the nouns lead to the highest disambiguation results, with the largest improvement over the baseline, which interestingly aligns with previous observations from work on word sense disambiguation (Mihalcea and Edmonds, 2004; Agirre et al., 2007).

Among the words considered, there are words that experience very large improvements over the baseline, such as *husband* (with an absolute increase over the baseline of 15.50%), *read* (13.89%) or *here* (13.66%). There are also words that experience very small improvements, such as *laugh* (1.86%), *tonight* (1.62%) or awesome (1.56%), and even a few words which are dominant in one gender, and for which the disambiguation accuracy is below the baseline, such as *shop* (-18.82%), *largely* (-11.23%) and *pink* (-7.39%).

To understand to what extent the change in frequency has an impact on gender-based word disambiguation (GD), in Table 4 we report results for words that have high frequency in both genders, or in only one gender at a time. Somehow surprisingly, the words that are used more often by one gender are harder to disambiguate. While this may be an artifact of the higher baseline, it may also suggest that the words that "belong" to a gender are used in a similar way by both genders (e.g., *cozy*), unlike words that are frequent in both genders, which get loaded with gender-specific meaning (e.g., *helpful*).

POS	No. words	Avg. no. examples	Baseline	GD
High frequency in both genders				
Noun	56	594	50.00%	56.98%*
Verb	60	451	52.53%	57.98%*
Adjective	53	590	50.98%	57.08%*
Adverb	60	560	50.39%	56.96%*
OVERALL	234	533	50.99%	57.26%*
High frequency in one gender				
Noun	41	565	50.95%	57.38%*
Verb	30	350	61.11%	58.71%
Adjective	40	344	64.14%	57.85%
Adverb	19	367	65.13%	58.13%
OVERALL	130	419	59.42%	57.94%

Table 4: Results for words that have high frequency in both genders, or in one gender at a time

The second analysis that we perform is concerned with the accuracy of polysemous words as compared to monosemous words. Comparative results are reported in Table 5. Monosemous words do not have sense changes between men and women, so being able to classify them with respect to the gender of the speaker relies exclusively on variations in their context. The fact that we obtain similar improvements for both monosemous and polysemous words is an indication that the gender differences that we observe are not due to the use of different word meanings, but rather to men and women using a certain word in different ways.

To further understand the relation between word senses and gender, we select 12 words (adjectives: *young, strong, new*; adverbs: *together, later, fast*; nouns: *party, idea, couple*; verbs: *heat, cause, understand*), randomly choose 100 examples for each of these words with equal split between male and female, and manually annotate their senses using WordNet (Miller, 1995). From these annotations, we observe that the predominant senses used by each gender are largely the same for most words. For instance, the words *party* and *heat*, shown in Figure 1 have a similar distribution over word senses. There are also a few exceptions, as illustrated for instance for the adjective *strong* in Figure 1, where the sense

[4]Disambiguation results that are significantly better than the baseline are marked with $*$ (statistical significance measured using a t-test, $p < 0.05$).

POS	No. words	Avg. no. examples	Baseline	GD
Polysemous words				
Noun	51	581	50.48%	57.44%*
Verb	50	460	54.72%	57.78%*
Adjective	50	463	56.13%	57.23%
Adverb	43	509	54.76%	57.89%*
OVERALL	194	504	53.98%	57.57%*
Monosemous words				
Noun	46	582	50.30%	56.82%*
Verb	40	363	56.23%	58.78%*
Adjective	48	445	56.58%	57.57%
Adverb	36	518	52.94%	56.46%*
OVERALL	170	478	54.03%	57.42%*

Table 5: Results for words that are polysemous or monosemous.

POS	No. words	Avg. no. examples	Baseline	GD
Noun	97	582	50.39%	57.15%*
Verb	90	417	55.39%	58.22%*
Adjective	98	454	56.35%	57.40%
Adverb	79	513	53.93%	57.24%*
OVERALL	364	492	53.98%	57.50%*

Table 6: Results for different parts-of-speech.

of (*firm, strong and sure*) is more often used by females, while the sense of (*having strength or power greater than average or expected*) is more frequently used by males. An interesting example is the word *together*, where males use more often the sense of (*assembled in one place*), while females use it with the sense of (*in each other's company*). This is in line with the observation made before using semantic classes, that women focus more on family and friends, while men talk more about groups and work.

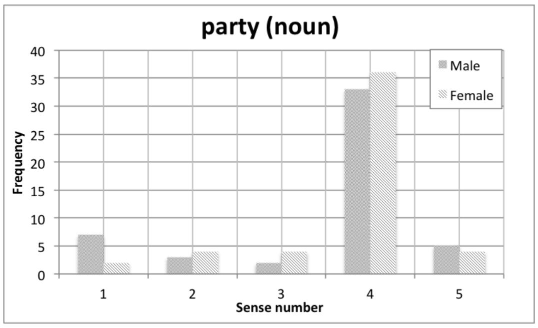
In general we find that the distribution of WordNet word senses for men and women for the 12 selected words is mostly similar. For an overall quantification, we use the Pearson and Spearman correlation metrics to calculate the correlation of word sense frequencies for the two genders, which resulted in a Pearson score of 0.94 and a Spearman score of 0.88, which reflect a high correlation. This suggests once again that the concept-centered differences that we observed between men and women are not due to distinct word meanings, but rather to different ways of using a certain word.

5 Conclusions

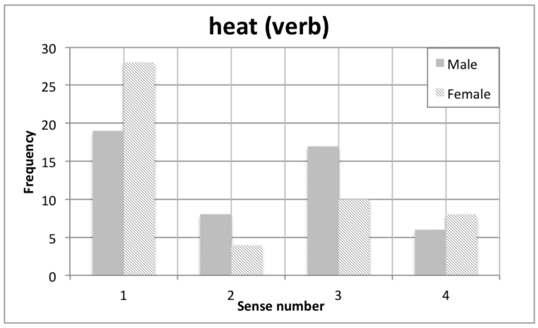
In this paper, we moved beyond the surface-level text classification approach to gender discrimination, and attempted to gain insights into the differences between men and women by using semantic methods that can point to salient word classes or differences in concept usage. We believe these distinctions at a deeper semantic level can be regarded as a reflection of the differences between the genders' perception of the world around them.

We first defined a metric for measuring the saliency of word classes, which we then used in conjunction with three semantic and psycholinguistic resources, resulting in a set of dominant word classes. With this metric, we were able to identify semantic and psycholinguistic word classes that are predominantly used by a gender, shading light on their interests and concerns.

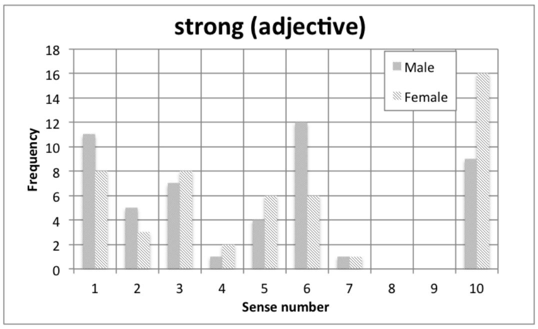
We also introduced the task of "gender-based word disambiguation," and using examples drawn from a large collection of blogposts for over 350 words, we showed that we can identify the gender of the person using a word with an accuracy significantly higher than the most frequent baseline. Additional analyses suggested that changes in frequency and context contribute to these differences, while the distribution of word senses is mainly similar.



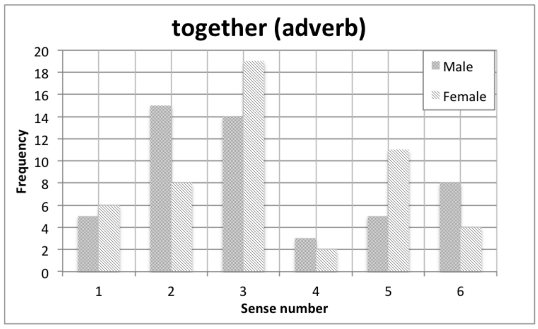
(a)



(b)



(c)



(d)

Figure 1: Distribution of WordNet senses for four words for male and female (100 examples)

In future work, we plan to extend the use of word classes to other resources, and also improve the disambiguation algorithm by including sociolinguistic and psycholinguistic features. We would also like to perform an in-depth analysis of the features that best characterize the differences in word usage between men and women.

Acknowledgments

This material is based in part upon work supported by the National Science Foundation (#1344257), the John Templeton Foundation (#48503), and the Michigan Institute for Data Science. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation, the John Templeton Foundation, or the Michigan Institute for Data Science.

References

E. Agirre, L. Marquez, and R. Wicentowski, editors. 2007. *Proceedings of the 4th International Workshop on Semantic Evaluations*, Prague, Czech Republic.

C. Boulis and M. Ostendorf. 2005. A quantitative analysis of lexical differences between genders in telephone conversations. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*, pages 435–442, Ann Arbor.

J. Burger, J. Henderson, G. Kim, and G. Zarrella. 2011. Discriminating gender on twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1301–1309.

B. Dandala, R. Mihalcea, and R. Bunescu, 2013. *Word Sense Disambiguation using Wikipedia*. Springer book series.

J.A. Doyle. 1985. *Sex and Gender: The Human Experience*. Wm. C. Brown Publishers.

P. Eckert and S. McConnell-Ginet. 2003. *Language and gender*. Cambridge University Press.

P. Gianfortoni, D. Adamson, and C. Rosé. 2011. Modeling of stylistic variation in social media with stretchy patterns. In *Proceedings of the First Workshop on Algorithms and Resources for Modelling of Dialects and Language Varieties*, pages 49–59. Association for Computational Linguistics.

S. Herring and J. Paolillo. 2004. Gender and genre variation in weblogs. *Journal of Sociolinguistics*.

D. Huffaker and S. L. Calvert. 2005. Gender, identity and language use in teenage blogs. *Journal of Computer-Mediated Communication*.

T. L. M. Kennedy, J. S. Robinson, and K. Trammell, 2005. *Does gender matter? Examining conversations in the blogosphere*.

M. Koppel, S. Argamon, and A. Shimoni. 2002. Automatically categorizing written texts by author gender. *Literary and Linguistic Computing*, 4(17):401–412.

R.T. Lakoff. 1972. *Language and woman's place*. Cambridge Univ Press.

H. Liu and R. Mihalcea. 2007. Of men, women, and computers: Data-driven gender modeling for improved user interfaces. In *International Conference on Weblogs and Social Media*.

R. Mihalcea and P. Edmonds, editors. 2004. *Proceedings of SENSEVAL-3, Association for Computational Linguistics Workshop*, Barcelona, Spain.

G. Miller, C. Leacock, T. Randee, and R. Bunker. 1993. A semantic concordance. In *Proceedings of the 3rd DARPA Workshop on Human Language Technology*, Plainsboro, New Jersey.

G. Miller. 1995. Wordnet: A lexical database. *Communication of the ACM*, 38(11).

A. Mukherjee and B. Liu. 2010. Improving gender classification of blog authors. In *Proceedings of the Conference on Empirical Methods in natural Language Processing*, pages 207–217.

D. Nguyen, D. Trieschnigg, A.S. Dogruöz, R. Gravel, M. Theune, T. Meder, and F. Jong. 2014. Why gender and age prediction from tweets is hard: Lessons from a crowdsourcing experiment. In *Proceedings of the 25th International Conference on Computational Linguistics*.

S. Nowson and J. Oberlander. 2006. The identity of bloggers: Openness and gender in personal weblogs. In *AAAI spring symposium: Computational approaches to analyzing weblogs*, pages 163–167.

A. Ortony, G. L. Clore, and M. A. Foss. 1987. The referential structure of the affective lexicon. *Cognitive Science*, (11).

C. Peersman, W. Daelemans, and L. Van Vaerenbergh. 2004. Predicting age and gender in online social networks. In *Proceedings of the 3rd Workshop on Search and Mining UserGenerated Contents*.

J. Pennebaker and M. Francis. 1999. Linguistic inquiry and word count: LIWC. Erlbaum Publishers.

J. Pennebaker and L. King. 1999. Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology*, (77).

V. Prabhakaran, A. Arora, and O. Rambow. 2014. Staying on topic: An indicator of power in political debates. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, October. Association for Computational Linguistics*.

D. Rao, D. Yarowsky, A. Shreevats, and M. Gupta. 2010. Classifying latent user attributes in twitter. In *Proceedings of the Second international workshop on Search and mining user-generated contents*, pages 37–44.

R. Sarawgi, K. Gajulapalli, and Y. Choi. 2011. Gender attribution: tracing stylometric evidence beyond topic and genre. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning*, pages 78–86. Association for Computational Linguistics.

J. Schler, M. Koppel, S. Argamon, and J. Pennebaker. 2006. Effects of age and gender on blogging. In *Proceedings of 2006 AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs*, pages 199–204, Stanford.

C. Strapparava and A. Valitutti. 2004. Wordnet-affect: an affective extension of wordnet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*, Lisbon.

D. Tannen. 1991. *You Just Don't Understand: Women and Men in Conversation*. London, Virago.

K. Toutanova, D. Klein, C. Manning, and Y. Singer. 2003. Feature-rich part-of-speech tagging with a cyclic dependency network. In *Proceedings of Human Language Technology Conference (HLT-NAACL 2003)*, Edmonton, Canada, May.