

# Understanding the Effect of Gender and Stance on Opinion Expression in Debates on “Abortion”

**Esin Durmus**  
Cornell University  
ed459@cornell.edu

**Claire Cardie**  
Cornell University  
cardie@cs.cornell.edu

## Abstract

In this paper, we focus on understanding linguistic differences across groups with different self-identified gender and stance in expressing their opinions about ABORTION. We provide a new dataset consisting of users’ gender, stance on ABORTION, as well as the debates in ABORTION drawn from *debate.org*. We use the gender and stance information to identify significant linguistic differences across individuals with different gender and stance. We show the importance of considering the stance information along with the gender, since we observe significant linguistic differences across individuals with different stance even within the same gender group.

## 1 Introduction

Understanding the differences in writing style and content across different demographic groups has been an important research focus (Argamon et al., 2003; Chong and Druckman, 2007). For example, Schler et al. (2006) has studied how the writing style and the content of blogs varies with the age and the gender of the blogger. We hypothesize that to study the actual linguistic differences across different gender groups’ opinion expression on a topic, it is necessary to control for the differences that could potentially be correlated with the individuals’ stances on the topic rather than their gender.

In this paper, we analyze linguistic factors that are significantly different across individuals with different gender and stance. We limit our discussions to the topic of ABORTION in order to control for the differences in language that are correlated with the topic. We include the text of debates where a FEMALE and a MALE with opposing stances on ABORTION discusses some aspect

of this topic. We observe significant differences in language usage across FEMALE and MALE. Moreover, we see significant differences even within a gender when the individuals’ stances on the topic is different. This suggests that the stance information should be taken into account while studying the impact of differences in demographics on language use. The main contributions of our study are three-fold:

- To the best of our knowledge, this is the first work which computationally studies the effect of gender in opinion expression accounting for the individuals’ stance.
- We provide a dataset of debates on ABORTION and user information of the debaters participated in these debates.
- We investigate which linguistic features are important for discriminating between groups with different gender and stance in expressing opinions about ABORTION.

## 2 Related Work

There has been a tremendous amount of work on understanding differences in writing styles between gender groups. Argamon et al. (2003) has found that females use many more pronouns whereas males use noun specifiers more frequently in British National Corpus covering a range of genres. Litvinova et al. (2017) has looked at the differences in the frequencies of some parts of speech (POS) between different genders in Russian written texts. Morphological features have shown to be important to discriminate between genders in many European languages (Mikros, 2013; Bortolato, 2016). Schofield and Mehr (2016) has studied significant linguistic and structural features of dialogue to differentiate genders in conversation and analyzed how these

effects relate to existing literature on gender in film.

Mohammad and Yang (2011) has shown that there are marked differences across genders in how they use emotion words in work-place email. They found that women use many words from the joy-sadness axis, whereas men prefer terms from the fear-trust axis. Thelwall et al. (2010) has looked at the comments on MySpace and found that females are more likely to give and receive more positive comments than are males, but there is no difference for negative comments.

Although previous work investigates discriminative features to distinguish between different gender groups, the effect of the stance on the linguistic differences across these groups has not been very well studied. It has been shown that people with different stance talk about a particular topic in a different way (Chong and Druckman, 2007; Diermeier et al., 2012; McCaffrey and Keys, 2000). Given that the linguistic differences may also be correlated with the stance on a topic, to better understand the actual effect of gender vs stance, we propose a controlled setting where we can account for differences in stance on the topic. This allows us to study the effect of gender and the stance simultaneously.

### 3 Data Description

For this work, we present a dataset of 1639 debates on ABORTION from October of 2007 until November of 2017 drawn from *debate.org*<sup>1</sup>. The dataset includes information for 1545 users that participated in debates on ABORTION. 265 of these users have identified their gender as FEMALE and 648 of them have identified their gender as MALE. Some examples for the debate titles on ABORTION are as follows: “ABORTION IS A CHOICE, NOT A RIGHT.”, “ABORTION IS GENERALLY IMMORAL.”, “ABORTION IS MURDER AND SHOULD BE ILLEGAL.”. From 1639 debates on ABORTION, we limit our study to the debates in which one debater identifies their gender as FEMALE and the other one identifies their gender as MALE. We preferred to include only the debates where a debater interacts with someone having opposing gender since it controls for the linguistic differences that can be correlated with interacting with people from the same vs.

<sup>1</sup>The dataset will be made publicly available at <http://www.cs.cornell.edu/esindurmus/>.

different gender group.

**Opinions on Abortion.** Each user profile on *debate.org* includes the user’s opinions on the most controversial debate topics<sup>2</sup>. As ABORTION is one of these controversial topics, each user shares their stance on ABORTION<sup>3</sup>: either PRO (in favor), CON (against), N/O (no opinion), N/S (not saying) or UND (undecided). To study how the stance along with the gender of the debaters affect the language, from the debates including debaters with opposing genders, we keep the debates where each debater identifies their stance on ABORTION either as PRO or CON and two debaters’ self-identified stance on ABORTION is opposing (e.g. if DEBATER 1’s stance on ABORTION is PRO then DEBATER 2’s stance is CON).

This controlled setting allows us to study the differences in linguistic features across people with different self-identified gender and stance when they are interacting with someone from the opposing gender and stance.

In the following sections, we will use the following abbreviations: (PRO-FEMALE) for the debaters who self-identified their gender as FEMALE and their stance on ABORTION as PRO, (CON-FEMALE) for the debaters who self-identified their gender as FEMALE and their stance on ABORTION as CON. Similarly, for MALE, we use the abbreviations PRO-MALE and CON-MALE.

### 4 Divergence of language across different groups

To understand how people with different genders (FEMALE and MALE) and stance (PRO and CON) talk about ABORTION, first, we applied the method proposed by Monroe et al. (2009). This method estimates the divergence between the two groups’ language by modeling word-usage as multinomial distributions smoothed with a uniform Dirichlet prior. The divergence between these two distributions is measured by log-odds ratio. Table 1 shows the most discriminating words for each side of every combinations of FEMALE/MALE genders and PRO/CON sides<sup>4</sup>.

We observe that people who are PRO vs CON use different terminologies to refer to the same concept. For example, PRO- uses the word “fe-

<sup>2</sup><http://www.debate.org/big-issues/>

<sup>3</sup>The website asks the users if ABORTION should be legal or not.

<sup>4</sup>We performed stemming and lemmatization on the text.

|            |  |
|------------|--|
| PRO-FEMALE | fetus, right, woman, sex, body, care, medical, emotion, really, alive.             |
| CON-FEMALE | baby, kill, cause, murder say, moral, wrong, abortion, human, womb.                |
| PRO-MALE   | will, women, birth, pain, force, pregnancy, good, bad, society, reason.            |
| CON-MALE   | human, unborn, kill, innocent, life, baby, development, crime, wrong, alive.       |
| PRO-FEMALE | sex, people, woman, adopt, care, choice, pregnancy, emotion, really, child, legal. |
| PRO-MALE   | moral, source, argument, will, human, conclusion, show, given, claim, logic.       |
| CON-FEMALE | baby, I’m, woman, don’t, health, want, mother, birth, think.                       |
| CON-MALE   | human, argument, fetus, right, unborn, crime, case, moral, definition, claim.      |

Table 1: The most discriminating words for FEMALE/MALE genders and PRO/CON sides (e.g. first two rows shows the words which are important to discriminate between the word-usage distribution of PRO-FEMALE and CON-FEMALE. The most discriminating words for PRO-FEMALE includes “fetus”, “right”, and “woman” while the most discriminating words for CON-FEMALE includes “baby”, “kill”, and “babies”).

“tus” while CON- emphasizes words like “baby” and “unborn”. This distinction is mainly because PRO- differentiates between a “fetus” and “baby” as they claim that a “fetus” is not alive and does not have rights unlike a “baby”. However, CON- mainly argues that life starts at conception. This suggests that the differences in terminology can be indicative of the differences in stance. Also, the discriminative words suggests that different gender groups have different motivations and justifications for their stance. We see that the words used by PRO-MALE focuses more on to the ethical, moral aspects of abortion, and how society perceives ABORTION, providing examples from other sources. However, PRO-FEMALE emphasizes more on women’s rights and choices. Moreover, while CON-MALE emphasizing more on the morality and human rights to justify their arguments, CON-FEMALE emphasizes more on women, women’s health and provides more personal information and experiences. This initial analysis suggests that there may be significant differences in language usage of different groups and the aspects these groups focus on. In the next section, we conduct a more fine-grained analysis for certain linguistic features to explore which linguistic features are important to discriminate between these groups.

## 5 Analysis

In this section, we analyze the important features to discriminate the opinions from people with different genders and stance. We describe the linguistic features for which we observe a significant difference for at least one the four groups.

For our analysis in this section, we combined

all the text utterances of PRO-FEMALE, CON-FEMALE, PRO-MALE, and CON-MALE from the debates. Then, we split combined text for each group to sentences. Our final dataset includes 2716 sentences for PRO-FEMALE, 2215 sentences for CON-FEMALE, 3010 sentences for PRO-MALE and 2444 for CON-MALE. All our analysis in this section is done at the sentence-level.

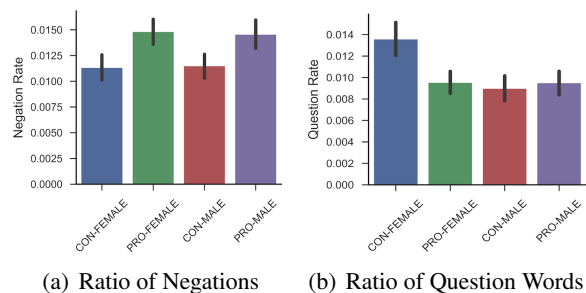


Figure 1: Ratio of negations and questions to number of tokens. PRO-MALE and PRO-FEMALE uses significantly more negation marks than CON-MALE and CON-FEMALE. CON-FEMALE uses significantly more question words than any other group.

**Sentence length.** Previous work has found that when mean sentence length is calculated, women come out as the wordier gender in writing (Weitz, 1976; Mulac and Lundell, 1994). However, in our experiments, we find that number of tokens of the text of CON-MALE is significantly higher than PRO-MALE and CON-FEMALE ( $p < 0.05$ )<sup>5</sup>. Similarly, number of characters of the text of CON-MALE is significantly higher than

<sup>5</sup>All reported  $p$  values are computed with Welch’s t-test.

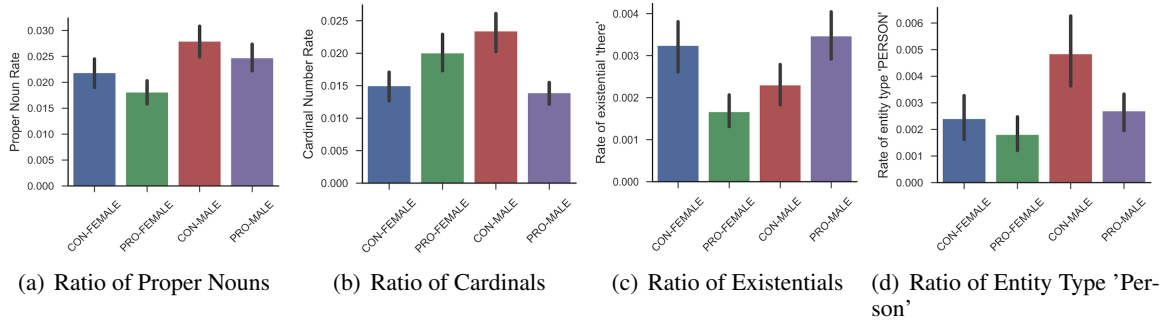


Figure 2: Ratio of Proper Nouns, Cardinals, Existential “there” and PERSON named entity.

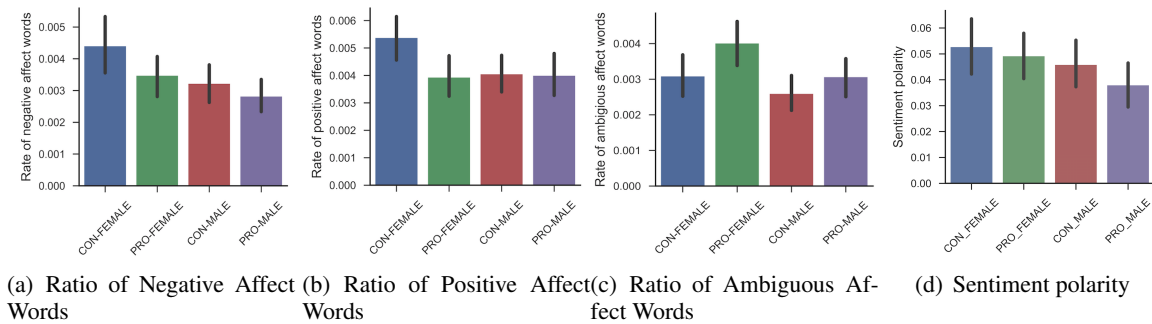


Figure 3: Ratio of negative, positive and ambiguous affect words, and sentiment polarity.

PRO-MALE ( $p < 0.05$ ). We do not observe any significant differences in sentence length between other groups.

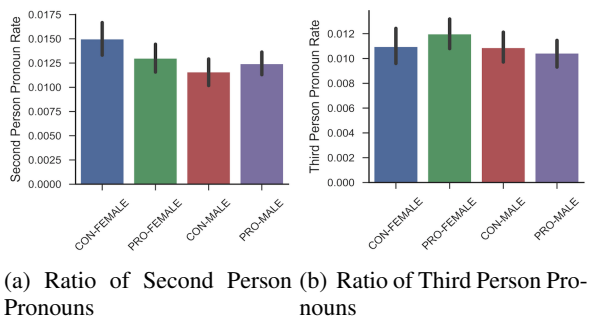


Figure 4: Ratio of number of personal pronouns to number of tokens.

**Personal pronouns.** Gleser et al. (1959) and Mulac et al. (1986) have shown that women frequently are the higher users of the personal pronouns when the entire category of personal pronouns is considered. Despite this finding, we do not observe any significant differences across different groups when we look at the total number of personal pronouns. We also do not observe any significant difference across these groups in first-

person pronoun usage. Figure 4 shows the ratio of number of second-person pronouns and third-person pronouns to number of tokens for each group. We find that PRO-FEMALE uses significantly more<sup>6</sup> third person pronouns than PRO-MALE. Moreover, CON-FEMALE uses significantly ( $p < 0.05$ ) more second person pronouns than any other group.

**Negation.** Mulac et al. (2000) has shown that male uses more negation words than female in a setting where 36 female and 50 male managers giving professional criticism in a role play. In our experiments, we did not observe significant differences across individuals with different genders; however, we observed significant differences across individuals with different stance. Figure 1(a) shows the ratio of negation marks to number of tokens. PRO-MALE and PRO-FEMALE uses significantly more ( $p < 0.001$ ) negation marks than both CON-MALE and CON-FEMALE.

**Question words.** We look at the ratio of total number of question words<sup>7</sup> to number of tokens. As we see in Figure 1(b), CON-FEMALE uses

<sup>6</sup>Our comparisons are made after we normalize the feature values with number of tokens in the sentences.

<sup>7</sup>includes “why”, “when”, “how”, “what”, “who”, “whose”, “whom”, “where”, “whose”, “whether”.

significantly more question words than any other group ( $p < 0.0001$ ). We observe that the question words are used not only to ask questions but also to form adjective clauses. These clauses are used to provide more specific information.

**POS tag types.** In Figure 2, we see the ratio of Proper Nouns 2(a), Cardinals 2(b), and Existential “there” 2(c) to number of tokens for each group. We observe that CON-MALE uses significantly more ( $p < 0.001$ ) Proper Nouns than both PRO-FEMALE and CON-FEMALE. Moreover, PRO-MALE uses significantly more ( $p < 0.001$ ) Proper Nouns than PRO-FEMALE. We find that users use Proper Nouns generally to cite relevant sources.

CON-MALE uses significantly more Cardinals than any other group<sup>8</sup>. PRO-FEMALE uses significantly more Cardinals than CON-FEMALE and PRO-MALE ( $p < 0.05$ ). We observe that Cardinals are used to refer to a source or something happened in the past (e.g. to provide a date).

We also find that CON-FEMALE and PRO-MALE uses significantly more existential “there” than any other group ( $p < 0.001$ ) and CON-MALE uses significantly more existential “there” than PRO-FEMALE ( $p < 0.05$ ).

**Named Entity Mentions.** For each entity type such as PERSON, LOCATION, ORGANIZATION, we look at the ratio of words belonging to these classes to number of tokens. As 2(d) shows, CON-MALE has significantly more mentions for entity type PERSON than any other group. We observe that CON-MALE uses PERSON entity type while citing other people’s ideas.

**Affect and Sentiment.** We used Wordnet-Affect (Valitutti, 2004) to find affective concepts correlated with affective words and compare the affective concepts across these different groups. We look at the words associated with negative, positive and ambiguous emotions. Figure 3 shows the ratio of negative 3(a), positive 3(b) and ambiguous 3(d) affect words to number of tokens for each group. We find that CON-FEMALE uses significantly more ( $p < 0.05$ ) negative and positive affect words than any other group and PRO-FEMALE uses significantly more ambiguous affect words any other group ( $p < 0.05$ ). We also look at the overall sentiment of the sentences and we see that both CON-FEMALE and PRO-FEMALE have significantly more positive sentiment than PRO-MALE ( $p < 0.05$ ). Although we observe that

<sup>8</sup> $p < 0.0001$  for CON-FEMALE and PRO-MALE,  $p < 0.05$  for PRO-FEMALE.

female uses more affect and sentiment words in some cases as also found by Danner et al. (2001), the usage of affect and sentiment words changes depending on the stance.

Our analysis in this section shows that there are significant differences in linguistic features between different genders and the individuals with different stance within the same gender group. For example, we see that the linguistic differences in use of negation is more correlated with stance than the gender. This finding highlights the importance of accounting for the stance information to understand actual linguistic differences between different genders.

## 6 Task : Predicting the gender and the stance.

In the previous section, we demonstrate some linguistic differences between groups with different gender and stance. This analysis highlights the importance of considering the stance while analyzing the differences in opinion expression for different gender groups. In this section, we look at whether the linguistic features we analyze in the previous section are predictive of the group of the person who utters a given piece of text. From the debates we described, we extracted all the turns and label whether they come from PRO-FEMALE, CON-FEMALE, PRO-MALE or CON-MALE. Our dataset for this task includes 451 turns<sup>9</sup>. The task is given a turn, predicting which one of these four groups the person who utters this turn belongs to.

### 6.1 Methodology

We treat this task as a multi-class (4-class) classification task and we used Logistic Regression<sup>10</sup>. We evaluate prediction accuracy, precision and recall using 5-fold cross validation approach. We pick the model parameters for each split with 3-fold cross validation on the training set. We perform ablation tests for the linguistic features analyzed in the previous section.

<sup>9</sup> 105 of them are coming from CON-MALE, 120 of them are coming from CON-FEMALE, 127 of them are coming from PRO-MALE, and 99 of them are coming from PRO-FEMALE.

<sup>10</sup>with one-vs.-rest strategy. We optimized the regularizer ( $\ell_1$  or  $\ell_2$ ) and the regularization parameter C (between  $10^{-5}$  and  $10^5$ ).



| Features                                | Accuracy     | Precision    | Recall       |
|---|--------------|--------------|--------------|
| Majority                                | 26.61        | 6.65         | 25.00        |
| Length                                  | 27.48        | 19.82        | 26.65        |
| Tf-idf                                  | 43.46        | 43.86        | 42.82        |
| Negation                                | 29.25        | 14.84        | 27.06        |
| Named entity: PERSON                    | 31.04        | 23.30        | 28.78        |
| Questions                               | 32.37        | 18.00        | 29.77        |
| Ambiguous affect words                  | 29.92        | 21.08        | 27.93        |
| Tf-idf+Questions+Ambiguous affect words | <b>44.34</b> | <b>44.01</b> | <b>43.34</b> |

Table 2: Results for the ablation tests. The best model includes *Tf-idf*, *Questions* and *Ambiguous affect words* features.

### 6.1.1 Baselines

**Majority.** Picking the majority group from the training data, as the predicted group.

**Length.** Using number of tokens in the utterance as a feature.

## 6.2 Result and Analysis

Table 2 includes the macro-average scores for the baselines and results for the ablation tests for the linguistic features performing better than the baselines. We see that *Tf-idf* features perform significantly better than baseline. We see that important *Tf-idf* features corresponds to the discriminative words described in Section 4. These features include “kill”, “murder”, “right”, “life”, “woman”, “baby”, “mother” etc. This suggest that *Tf-idf* features are helpful to capture the differences in terminology and the motivation behind an individual’s reasoning. We see that some of the linguistic features such as *Negation*, *Named Entity: PERSON*, *Questions* and *Ambiguous affect words* that we find to be discriminative in Section 5 have significantly more predictive power than majority and length baselines. The best predictive model includes *Tf-idf* features, questions, and *Ambiguous affect words*.

## 7 Conclusion

In this paper we show that accounting for the stance is important in order to effectively study the difference in linguistic features between different genders. We used both stance and gender information of the users and we analyzed differences in language across these different groups. We find that some linguistic features are more correlated with the stance of the individuals while others are more correlated with the gender. As a future direction, we would like to explore methods for rep-

resenting user information, that captures various aspects of a user such as education level, political ideology, and religious beliefs, in order study linguistic difference across different groups of people in a more controlled manner.

## 8 Acknowledgements

This work was supported in part by NSF grant SES-1741441 and DARPA DEFT Grant FA8750-13-2-0015. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of NSF, DARPA or the U.S. Government. We thank Faisal Ladhak, Ana Smith, Amr Sharaf and the anonymous reviewers for their helpful feedback and suggestions. We also thank the Cornell NLP group for their insightful comments.

## References

- Shlomo Argamon, Moshe Koppel, Jonathan M. Fine, and Anat Rachel Shimoni. 2003. Gender, genre, and writing style in formal written texts.
- Claudia Bortolato. 2016. Intertextual distance of function words as a tool to detect an author’s gender: A corpus-based study on contemporary italian literature. *Glottometrics*, 34:28–43.
- Dennis Chong and James N. Druckman. 2007. *Framing theory*. *Annual Review of Political Science*, 10(1):103–126.
- Deborah D Danner, David A Snowdon, and Wallace V Friesen. 2001. Positive emotions in early life and longevity: Findings from the nun study. *Journal of Personality and Social Psychology*, 80(5):804–813.
- Daniel Diermeier, Jean-Francois Godbout, Bei Yu, and Stefan Kaufmann. 2012. *Language and ideology in congress*. *British Journal of Political Science*, 42(1):3155.

- Goldine C. Gleser, Louis A. Gottschalk, and Watkins John. 1959. [The relationship of sex and intelligence to choice of words: A normative study of verbal behavior](#). *Journal of Clinical Psychology*, 15(2):182–191.
- Tatiana Litvinova, Pavel Seredin, Olga Litvinova, and Olga Zagorovskaya. 2017. [Differences in type-token ratio and part-of-speech frequencies in male and female russian written texts](#). In *Proceedings of the Workshop on Stylistic Variation*, pages 69–73. Association for Computational Linguistics.
- Dawn McCaffrey and Jennifer Keys. 2000. [Competitive framing processes in the abortion debate: Polarization-vilification, frame saving, and frame debunking](#). *The Sociological Quarterly*, 41(1):41–61.
- GK Mikros. 2013. Systematic stylometric differences in men and women authors: a corpus-based study. *Issues in quantitative linguistics 3. Dedicated to Karl-Heinz Best on the occasion of his 70th birthday*, pages 206–223.
- Saif M Mohammad and Tony Wenda Yang. 2011. Tracking sentiment in mail: How genders differ on emotional axes. In *Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis*, pages 70–79. Association for Computational Linguistics.
- Burt L. Monroe, Michael Colaresi, and Kevin M. Quinn. 2009. Fightin words: Lexical feature selection and evaluation for identifying the content of political conflict.
- Anthony Mulac and Torborg Louisa Lundell. 1994. Effects of gender-linked language differences in adults’ written discourse: Multivariate tests of language effects. *Language & Communication*, 14(3):299–309.
- Anthony Mulac, Torborg Louisa Lundell, and James J. Bradac. 1986. Male/female language differences and attributional consequences in a public speaking situation: Toward an explanation of the genderlinked language effect. *Communication Monographs*, 53(2):115–129.
- Anthony Mulac, David R Seibold, and Jennifer Lee Farris. 2000. Female and male managers and professionals criticism giving: Differences in language use and effects. *Journal of Language and Social Psychology*, 19(4):389–415.
- Jonathan Schler, Moshe Koppel, Shlomo Argamon, and James Pennebaker. 2006. Effects of age and gender on blogging. In *Computational Approaches to Analyzing Weblogs - Papers from the AAAI Spring Symposium, Technical Report*, volume SS-06-03, pages 191–197.
- Alexandra Schofield and Leo Mehr. 2016. [Gender-distinguishing features in film dialogue](#). In *Proceedings of the Fifth Workshop on Computational Linguistics for Literature*, pages 32–39. Association for Computational Linguistics.
- Mike Thelwall, David Wilkinson, and Sukhvinder Upal. 2010. [Data mining emotion in social network communication: Gender differences in myspace](#). *Journal of the American Society for Information Science and Technology*, 61(1):190–199.
- Ro Valitutti. 2004. Wordnet-affect: an affective extension of wordnet. In *In Proceedings of the 4th International Conference on Language Resources and Evaluation*, pages 1083–1086.
- Shirley Weitz. 1976. [Sex differences in nonverbal communication](#). *Sex Roles*, 2(2):175–184.