Gender Differences in the Length of Words and Sentences on the Corpus of Congressional Speeches

Dragana Božić Lenard
Josip Juraj Strossmayer University of Osijek
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek

Abstract: Gender differences in various subfields and registers have been of an interest to sociolinguists since the 1960s. This paper aimed to examine gender differences in the categories of word count, words per sentence and long words in political speeches. The study was conducted on the corpus of the congressional speeches from the most current 113th United States Congress. The corpus consisted of 672 speeches given by the female and 2,983 by the male politicians. The speeches transcripts were downloaded from the official speech repository Thomas. The analysis was two-fold. First, the transcripts were analyzed with the text analysis software Linguistic Inquiry and Word Count which counted overall words and words per sentence and calculated the degree to which the politicians used long words. The computational analysis was followed by a statistical analysis done with the software SPSS which was used to carry out the Mann-Whitney, Kruskal-Wallis, independent sample t-test and one-way ANOVA. The results showed that the male politicians occupied the floor longer but the female politicians used longer words in their speeches, while there were no statistically significant differences in the length of sentences.

Key words: congressional speeches, gender differences, LIWC, long words, SPSS, word count, words per sentence

1. Introduction

Language is one of the most frequently used means to express one’s thoughts. Guided by thoughts, people use different expressions. Two people might speak about the same thing and use completely different expressions, which might reveal their (c)overt feelings about things from the real world.

Politics has been defined as the struggle for power and as such, it has been naturally dominated by men whose personality characteristics (directness, assertiveness, proficiency) are prototypical for it (Huddy and Terkildsen, 1993; Kahn 1996). Hence, men’s domination and women’s under-representation in politics are rather expected. Furthermore, when running for offices, women usually hold offices related to education, the elderly, social and health care services, i.e. offices that deal with issues not so rigorously related to masculinity. Female politicians are associated with solidarity issues, whereas male politicians with business, military and economy (Huddy and Terkildsen, 1993; Alexander and Andersen, 1993; Leeper, 1991). Nevertheless, women are frequently depoliticized and womanized by media (Bengoechea, 2011). Therefore, women entering the world which has traditionally been claimed by men might provoke certain behavioral and linguistic changes.

2. Theoretical background

Gender studies in language are divided in four main stages. Built upon Jespersen (1922) marking men’s language as a standard and women’s as inferior, Robin Lakoff (1975) proposed the deficit model which claimed that the social pressure guided women to “talk like ladies” making their speech powerless in comparison to men’s. At the same time, Zimmerman and West (1975) suggested the dominance approach believing that men realized their more powerful social position in floor apportionment and interrupting others while speaking. The first two stages were extensively criticized for being too male-centered so researchers started reassessing women’s language by searching for its strengths leading to an introduction of the cultural difference approach first proposed by Maltz and Borker (1982). This approach suggested that miscommunication stemmed from different communicational rules used by members of different communities. Precisely, miscommunication between men and women might result from them being raised in different peer groups. The approach was popularized by Deborah Tannen (1986; 1990) who regarded men’s and women’s language as equally valid and believed that any differences were the result of differences in a socialization process. Finally, the fourth stage in gender studies in language are a group of anti-essentialist approaches such as ethnomethodology (Garfinkel, 1967), discursive psychology (Potter and Wetherall, 1987), social constructionism (Shotter and Gergen, 1994), conversation analysis (Sacks, 1992). They
operated under the assumption that gender was not an inherited trait but something one “does”.

Since the 1980s, researchers started to look into more specific fields. Given the previous findings, researchers were interested in examining gender differences in public contexts and presented women’s language as facilitative and cooperative as opposed to argumentative and competitive men’s language (Coates, 1989, Holmes, 1992). Additionally, research (Edelsky, 1981; James and Clarke, 1992; Karpowitz and Mendelberg, 2014) suggested that when speaking in public contexts, men occupied the floor longer than women, thus being more assertive. This led to the idea that interactional norms were the male norms (Gal, 1991) and started the notion of gendered spaces (Freed, 1996). However, later research (Webster, 1990; McElhinny, 1998) pointed to women adopting masculine strategies when they speak in naturally men’s fields – political speeches being one of the most representative examples.

3. Methodology

The aim of this paper was to examine if the speeches of the male and female politicians who served at the 113th US Congress differed in terms of the word count, words per sentence and long (more than six letter) words. 113th Congress was chosen because it is the most currently completed one and it had the highest number of female politicians. It was composed of 103 female and 405 male politicians. The House of Representatives and Senate consisted of Democrats and Republicans as shown in Figure 1. These two variables (chamber seats and party affiliation) were also taken into consideration when studying gender differences.

The transcripts of all the speeches made in the 113th Congress were downloaded from the official repository Thomas available at http://thomas.loc.gov/home/thomas.php. The corpus was composed of 672 speeches made by the female and 2,983 speeches by the male politicians. Two types of analyses were done in this research. First, the computational analysis which counted words and calculated the degree to which the politicians used long words was done with the software Linguistic Inquiry and Word Count. The raw results were uploaded in the software for the statistical analysis SPSS. Given the distribution of data, the nonparametric Mann-Whitney and Kruskal-Wallis tests were used with the variables of word count and words per sentences, while the parametric independent sample t-test and one-way ANOVA were used with the six-letter words variable.

4. Results and Discussion

The results of the mentioned variables will be presented in their respective subsections accompanied by the similar analytical results of other researchers.

4.1. Word count

A great deal of previous research into gender differences in language has focused on the issue of verbosity. Marjorie Swacker (1976) investigated gender differences in asking questions and providing answers at academic conferences. Her research found that women contributed only 27.4% with questions asking. Questions asked by women were twice as short as men's questions. The differences in both the structure and length of women’s and men’s questions were in men introducing the opening question with a statement, asking more than one question and responding to a speaker’s answer with additional questions or comments. Additionally, when invited to ask questions, almost exclusively men were first to ask questions; they asked more questions and their questions were longer. The findings led Swacker to conclude that women were less comfortable than men in speaking before a large group of people in a public meeting. Two years later, Westbrook Eakins and Eakins (1978) tape recorded seven university faculty meetings. Their findings supported Swacker’s – with only one exception, men spoke more frequently and their speeches were longer than women’s. The authors were also interested in turn-takings. Their study recorded that women’s turns lasted from 3 to 10 seconds while men’s lasted from 10.66 to 17.07 seconds. According to the research findings, Westbrook Eakins and
Eakins concluded that women were reluctant to speak in a public event attended by a larger group of people.

In her book, Dale Spender (1980) suggested that people intuitively believed that women should be seen and not heard. From her point of view, when talking equally, women were perceived as the ones who talked more. She believed that happened because of women speaking in various situations (home, social situations, on a phone), which men could not understand. Similar findings were reported by Sadker and Sadker in 1985. Teachers were shown a videotape of a classroom discussion and asked to conclude who spoke more. They believed that girls talked more when in fact boys talked three times more than girls.

In a ground-breaking paper *Who’s Got the Floor?* Edelsky (1981) distinguished between two types of floors: singly developed where one speaker speaks at a time and collaboratively developed which is open to all participants simultaneously. To paraphrase, a collaboratively developed floor, also known as a polyphonic floor (Chafe, 1995), included overlapping speech and co-construction of utterances. Edelsky’s research showed that men talked more and took longer turns in a singly developed floor. In comparison, turn length and frequency differences were naturalized and women were more actively engaged in speech in a collaboratively developed floor. These research findings were supported by numerous studies (Falk, 1980; Chafe, 1995; Coates, 1996; Coates, 1997; Coates and Jordan, 1997) conducted in English-speaking communities in Australia, Britain and North America.

Other women’s personal experiences of their husbands being talkative at work and life of parties and simultaneously being mute at home inspired the linguist Deborah Tannen to introduce the concepts of report talk and rapport talk in 1990. She believed that men and women used language for different purposes. Women see language as a way of establishing connections and negotiating relationships – a rapport. Men, on the other hand, use language to maintain independence and negotiate status in a hierarchical order – a report. These differences result from different styles of upbringing, expectations and talking to boys and girls. However, in spite of being raised differently, men do not intentionally prevent women from speaking in public settings. Rather, they see women as equals and implicitly invite them, as well as other men, to fight for the floor.

Furthermore, studying gender differences in the workplace, Kendall and Tannen (1997) found that men talked more often, their speeches were longer and they interrupted a person speaking, while women were more interrupted even by a person subordinate to them. Similar findings of men talking more in formal and women in informal settings were reported by James and Drakich (1993) with women’s speech often being trivialized and labeled as gossiping (Weatherall, 2002). Starting from the hypothesis that women speak less than men in formal settings, Power and Berardone (1998) carried out an analysis of first speeches in the Australian parliament. Their study showed that there was no statistically significant difference in the amount of men’s and women’s speech; however, women spoke about a wider range of topics significantly more (p<.01) than men.

According to the previous research findings, we expected that the men’s and women’s speech in the US Congress would differ in terms of a word count. The word count and words per sentence categories, unlike others, are not expressed in the percentage form, i.e. the software provides the exact number of used words. The total number of word count in our corpus was 2,615,264 – 2,203,595 words spoken by the men and 411,669 by the women. Even though a conclusion can be drawn from these raw numbers, we did the Mann-Whitney test and found that there was a significant difference in the men’s (M = 205.59) and the women’s (M = 176.20) word count (U = 12719, Z = - 2.239, p = .025, two-tailed).

Once we detected a statistically significant difference between the men and women as groups, we were interested in more subtle differences such as differences between and within groups based on the party affiliation, chamber and education level. We then conducted Kruskal-Wallis and the post hoc tests. The results showed that the men Democrats (M = 216.80) spoke more than the women Democrats (M = 168.60) with a statistically significant difference (p = .017). Furthermore, the post hoc tests revealed that the women Representatives (M = 145.23) spoke less than the women Senators (M = 303.15) as well as the men Senators (M = 318.71) with p = .000 in both cases. The men Representatives (M = 165.26) spoke less than the men Senators (M = 318.71) and the women Senators (M = 303.15) with p = .000 in both cases. There was no statistically significant difference (p = .112) in the word count between the men and the women regarding the level of education.

Statistical evidence showed that the Senators, regardless of gender, spoke more than the Representatives which may have happened because of the time limitation in the House of Representatives. Providing statistical evidence that the men spoke more in a public setting, our research contributed to a growing body of research reporting that men speak more than women in various types of public settings thus trying to establish themselves in a hierarchical order.
The previous studies on gender differences in the length of sentences were contradictory. In 1979, Poole interviewed 96 sixteen-year-olds who were divided in equal social class and gender groups. The verbatim transcripts of the undertaken interviews showed that girls used longer sentences than boys. One of the most productive researchers in the field is Anthony Mulac who, together with Lundell, reported the same results in 1986. They assessed oral descriptions of landscape photographs provided by 40 subjects who were sixth grade students, university freshmen and sophomore students, teaching assistants and older residents from the adjacent town. They audiotaped and later on transcribed the descriptions attributing gender codes to each subject. Compared to Poole’s study which was equalized in terms of the education level and social class, demographic characteristics (background, education level, and race) of their subjects varied substantially. They also recorded women using longer sentences than men in oral descriptions.

Furthermore, Mulac and his colleagues (1986) took one-minute speech transcripts by 30 university students and asked 11 trained coders to analyze them linguistically for 35 language features chosen as potential discriminators of speakers’ gender. The results of the discriminant analysis showed 100% accuracy of gender prediction based on a combination of 20 linguistic features. One of the discriminant features were longer sentences used by female speakers. The same feature was recorded in Kerstin Thelander’s 1986 study on parliamentary language in Sweden (cited in Romaine, 1999). Additionally, Thelander invited her participants to describe each other’s language styles. Men’s speech was described as abstract, authoritative, impersonal and pompous, whereas women’s speech was described as soft, simple, spontaneous, clear and sensitive. Interestingly, both men and women ascribed negative features to men’s speech.

In the second study on the effects of writing, Mulac and Lundell (1994) asked 40 communication class university students to write descriptions of landscape photographs. Their results were consistent with the ones from Mulac and Lundell’s study (1986) which recorded women using longer sentences than men. Another study by Mulac et al. (2000) on the differences in language use and effects of men and women managers giving criticisms to their subordinates confirmed that men used more words overall, while women used longer sentences.

The following year Mulac et al. provided empirical support for the Maltz and Borker’s (1982) hypothesis that gender differences can be explained by gender-as-culture approach. Mulac et al. (2000) located 16 language features that had consistently indicated communicator gender among which was women’s usage of lengthy sentences. In addition, statistically significant difference (p = .002) on women using longer sentences when writing about the previous summer in e-mails and letters to male and female friends was reported in Colley et al. (2004). Reporting the same findings, Mulac (2006: 236) stated that men and women “grew up in different sociolinguistic cultural groups and have subtly different styles while accomplishing the same communication task”. In a more recent study, Mulac et al. (2013) described men’s language as reflecting higher on dynamism, while women’s reflected higher socio-intellectual status because women used intensive adverbs, hedges, dependent clauses and longer sentences.

Several studies reported contradictory results. Examining eight-minute problem-solving interactions which involved 108 university students (54 women and 54 men), Mulac (1989) discovered that regardless of their partner’s gender, men spoke in longer utterances. Interested in written speeches, Mulac et al. (1990) studied fourth-grade students’ essays. The analysis showed that boys used longer sentences than girls.

The previous research reported the difference in the mean length of sentences used by women and men in both written texts and oral speeches in a variety of communicational situations. The majority of them indicated that women used longer sentences than men. Our Mann-Whitney analysis pointed to no statistically significant difference (U = 14307.5, Z = - .640, p = .522, two-tailed) in the length of sentences, i.e. the men (M = 200.17) and the women (M = 191.77) used equally long sentences. Additionally, the post hoc Kruskal-Wallis analysis revealed no significant difference in neither within nor among groups (for gender and party p = .103, gender and chamber p = .065, gender and education level p = .241).

No statistically significant difference was recorded with respect to the length of sentences used by the female and the male politicians; hence, both the female and the male politicians were equally elaborative when giving their speeches. These results are likely to be related to the formality of a political setting. Another possible explanation is in the participants’ preparation for the speech written to be spoken rather than giving one in an impromptu way.

4.3. Six-letter words

Several studies investigating gender differences paid closer attention to the length of words. Kučera and Francis (1967) compiled a million-word corpus
of present-day English language. Their corpus consisted of 500 samples of about 2,000 words per text. The texts, selected from American publications from 1961, were chosen to include a broad range of styles and topics and were grouped into 15 genres. The computational analysis of the corpus showed that women used less frequent and longer words than men. These results were not supported by similar later research.

Biber et al. (1998) employed a factor analysis technique to study text registers. Their findings showed that male authors used long words and nouns, whereas female authors used more pronouns and present tense verbs. The results led to the conclusion that men used more complex language and their style was informational and uninvolved, while the women’s style was more involved. Furthermore, using British National Corpus texts, Koppel et al. (2002) carried on an investigation on predicting a writer’s gender simply on words usage. They used a list of 30 words reported in the previous research as having extreme usage variations and possessing distinguishing features across gender and age groups. Reporting 80% accuracy in determining a writer’s gender, men’s writing was characterized with long words. Newman and his colleagues (2008) analyzed a database of over 14,000 text files from 70 studies and 22 laboratories from the United States (63 studies), New Zealand (4 studies) and United Kingdom (3 studies). The studies were carried on over a 22-year-period (1980-2002) and contained 93% of written texts and 7% of transcribed speeches. Two-thirds of the participants were college students. Their research findings were consistent with the previous ones reporting men to use longer and more complex words than women. A recent Yu’s study (2013) on gender differences conducted on a large corpus of Congressional speeches from the 101st to the 110th Congress (1989-2008) confirmed the consistent usage of long words as a masculine style feature. The same was confirmed by Jones (2015) in an article submitted for publication in Perspectives on Politics. She studied speech given by Hillary Clinton as one of the arguably most prominent female American politicians. Analyzing 564 interviews and candidate debates (1992-2013), using LIWC Jones hypothesized that Clinton has been using long words and other masculine style features, changing her language over the years into a more masculine one.

In our study, the independent sample t-test was conducted to compare six-letters words usage by the women and men. There was a significant difference (p = .000) in the usage of six-letter words by the men (M = 23.5, SD = 3.61) and the women (M = 24.9, SD = 3.16); t (199) = -3.81. These results suggested that the women used six-letter words significantly more than the men. We then conducted the one-way ANOVA test to see if there were significant differences within and among groups. The one-way ANOVA test revealed that there was a significant difference on the six-letter words usage at the p<.05 level regarding gender and party [F (3, 391) = 4.62, p = .003]. The post hoc comparisons using the Tukey HSD test indicated that the mean score for the women Democrats (M = 24.84, SD = 2.98) was significantly different (p = .010) than the men Republicans (M = 23.33, SD = 3.83). Other groups did not differ significantly.

Additionally, we performed another one-way ANOVA test on gender and chamber groups. The test showed that there was a significant difference in the six-letter words usage at the p<.05 level regarding gender and chamber [F (3, 391) = 7.76, p = .000]. The post hoc Tukey HSD test showed that the mean score for the women Representatives (M = 24.87, SD = 3.27) was significantly different (p = .001) than the men Representatives (M = 23.11, SD = 3.76). Furthermore, the women Senators (M = 25.23, SD = 2.72) significantly differed (p = .047) from the men Representatives (M = 23.11, SD = 3.76). There was also a significant difference among a group of men. Precisely, the men Representatives (M = 23.11, SD = 3.76) significantly differed (p = .009) from the men Senators (M = 24.58, SD = 2.93). The women Representatives (M = 24.87, SD = 3.27) and the women Senators (M = 25.23, SD = 2.72) did not significantly differ on the six-letter words usage.

We were especially interested in testing the gender differences and education level. The one-way ANOVA signified a statistical difference at the p<.05 level [F (7, 387) = 3, p = .007]. The post hoc Tukey HSD test showed only one significant difference. The mean score for the women who have a graduate level of education (M = 25.14, SD = 3.15) was significantly different (p = .008) than the men with the same level of education (M = 23.46, SD = 3.55). It shows that the female politicians use more complex words than their male counterparts.

To summarize, these results are not in accord with the previous studies claiming that men use long words significantly more than women. Our results suggested that the women affiliated to the Republican party, who served in the Senate, used six-letter words the most thus employing what is traditionally considered as a masculine style feature. The women using complex and less frequent words may be explained by them being more appreciative of the setting formality. Furthermore, by using more complicated and scholarly vocabulary, the women might have tried to establish themselves as valuable contributors to political society. The women serving in the Senate,
whose serving position is more competitive than in the House of Representatives, may have wanted to prove their election integrity.

5. Conclusions

Using the corpus of congressional speeches, this research analyzed gender differences in the length of words and sentences and the floor apportionment. 672 speeches by the female and 2,983 by the male politicians given in the 113th United States Congress were included in the analysis. The computational analysis done with the software LIWC was accompanied by the statistical analysis carried out with the software SPSS which was used to perform the nonparametric Mann-Whitney and Kruskal-Wallis tests and the parametric independent sample t-test and one-way ANOVA test. The results showed that the male politicians statistically spoke more than the female politicians in Congress thus confirming the stereotypical view that men feel more comfortable when speaking in public contexts. Both the male and female Senators spoke longer than the Representatives due to the time limitation in the House of Representatives. Furthermore, the evidence demonstrated that the male and female politicians used equally long sentences, which might be due to their preparation for speeches and acknowledgment of the formal setting. The results of the six-letter words variable pointed to a statistically significant difference with the female politicians using the variable more than their male counterparts. Additionally, the female Representatives used long words more than the male Representatives, which might be due to the female politicians feeling that they needed to prove themselves more than their male colleagues, hence the higher and more scholar preparation for speeches.

6. References


