

Sebastian Dunat¹

University of Bielsko-Biala

ORCID ID: 0000-0001-9188-0987

e-mail: sdunat@ath.bielsko.pl

Modality and Text Structure in American Senators' Tweets: A Corpus Study of Gendered Lexical Items

ABSTRAKT

Modalność i struktura tekstu w tweetach amerykańskich senatorów: badania korpusowe jednostek leksykalnych charakterystycznych dla danej płci

W wielu przypadkach procedury, zachowania oraz koncepcje jakichkolwiek konkretnych relacji w obrębie domeny językowej zależą od zdolności danych słów do ich zaprezentowania. Ludzie natomiast polegają na swoich umiejętnościach leksykalnych, by zaprezentować się w sieci. Celem tego badania jest wskazanie i zdefiniowanie różnic między językiem mężczyzn i kobiet u amerykańskich senatorów na Twitterze. Analiza danych wykorzystuje metody językoznawstwa korpusowego i komputerowego w celu znalezienia odpowiedzi na postawione pytania. Wybrany do badania materiał, z września 2018 r., zawiera 240 tweetów 24 amerykańskich senatorów, jednakowej liczby dla obu płci. Tekst opisuje zróżnicowanie danych korpusowych i ilościowe badania statystyczne wykonane za pomocą programu R, w celu sprawdzenia poziomu istotności prezentowanych hipotez.

SŁOWA KLUCZOWE: płeć, językoznawstwo korpusowe, językoznawstwo komputerowe, jednostki leksykalne, modalność, tweety, amerykański senat

Introduction

Procedures, behaviors, conceptions of any particular relationship within the linguistic domain depend on the words' ability to represent them, whereas people depend on their lexical abilities to represent themselves in the web. The strategies used to extract the author's involvement and blur the source of judgment, often manipulated to show tentativeness or formality, are based on the concepts of modality, politeness and hedging². Corpus linguistics creates, analyses, and

¹ Data złożenia tekstu do Redakcji „MIS”: 07.12.2018 r.; data zatwierdzenia tekstu do druku: 27.06.2019.

² K. Hyland, *Writing without conviction? Hedging in scientific research articles*. „Applied Linguistics”, 1996, No. 17,

answers the assumed questions on the basis of researching a large amount of data. This kind of linguistic research can provide information about the specificity of words usage, its changes under the influence of time and many other variables³. The purpose of this study is to determine and define the differences between the selected examples of American senators' (men and women) language on Twitter. The research material selected for the study comprises 240 tweets of 24 American senators; of both sexes equally. Ten Twitter posts for each female and male senator were obtained in September 2018; two posts selected each week. If this assumption could not have been met, first available tweets posted in September were acquired. The study describes the corpus data variation of modality markers, hedges; text structure: tenses use, parts of speech and demonstratives distribution. Quantitative statistical surveys, using the R program to check the level of the significance of the presented hypotheses, were conducted. The researcher would like to underline that this synchronic study was inspired by the work of M. Vuković⁴.

Methodology

Politeness strategies seem an important factor for Twitter posts research. The researcher would like consider the notion of *face* developed by Brown & Levinson (after Goffman's research). As authors define: „face is something that is emotionally invested, and that can be lost, maintained, or enhanced, and must be constantly attended to in interaction”⁵. It has strong connotations with English folk, corresponds to notions of embarrassment and humiliation, mainly with the *losing face* term. Maintaining face in interaction is based on mutual vulnerability because defending one's own face can threaten others. Beside the face context cultural differences, Brown and Levinson assume that „the mutual knowledge of members' public self-image or face, and the social necessity to orient oneself to it in interaction, are universal”⁶. As they establish, the negative face is the want to commence one's own actions freely. And the positive face is the want of every human to have desirable influence, during interaction, on some other people⁷.

The semantic aspects of modality connect with the expressions of tentativeness, commitment and detachment or personal attitudes. The related literature focuses on discourse analysis, corpus studies, applied linguistics, pragmatics, semantic linguistics and philosophy of logic⁸. The basic semantic nature of modality,

p. 433.

³ J. Sunderland, *Language and Gender: An advanced resource book*. Routledge: London 2006, p. 56.

⁴ M. Vuković, *Weak Epistemic modality in the UK Parliamentary Discourse*, „Journal of Interdisciplinary Approaches to Text”, 2014, No.1, p. 121-137.

⁵ P. Brown, S. C. Levinson, *Politeness: some universals in language usage*. CUP: Cambridge 1987, p. 61.

⁶ *Ibidem*.

⁷ P. Brown, S. C. Levinson, *Politeness: some universals in language usage...* op. cit., p. 62.

⁸ K. Hyland, *Hedging in Scientific Research Articles*. John Benjamins Publishing Company: Amsterdam 1998, p. 43.

proposed by Lyons, is defined as: „opinion and attitude of the speaker”⁹. Nonetheless, Lyons underlines the importance of the categories of judging the truth of statements, concerned equally with the author’s standpoint; while researching aspects of modality. Further important aspect of modality refers to its role in the potential fulfillment of an event, and not the event itself¹⁰.

Modality can be divided into two types: epistemic and deontic. „Epistemic modality expresses the speaker’s opinion or belief concerning the truth of what is said, while deontic modality indicates his or her observations about the necessity or obligation to perform particular actions”¹¹. To elucidate, epistemic modality is used to highlight confidence in the truth of propositional information. Lack of knowledge or confidence in personal judgments can use hedging as one of the aspects of epistemic modality. Hedging bears two important properties, i.e. non-factivity and subjectivity, as specified by Hyland¹². The first one carries the truth of an utterance in the speaker’s view. The second, attributed to the writer, concerns information about one’s intentions. These strategies are often used to cover the source of judgment or extract the involvement of an author. Thus, the properties can be manipulated to show tentativeness or formality.

Furthermore, one of the exponents of epistemic modality are modal verbs. There is also syntactic hedging visible through: tense, questions, tags, source attributions, passive voice, clauses of conditions and concession; all these have confirmed importance in literature¹³.

Epistemic modality is affected by the interpersonal and interactional factors (the notion of *epistemic sense*). According to Halliday, modality represents the „interpersonal semantic system of language,”¹⁴ which characterizes the personal involvement of the author in the statements and intercedes the speaker-audience relations. As Halliday defines:

„Modality is a form of participation by the speaker in the speech event. Through modality, the speaker associates with the thesis an indication of its status and validity in his own judgment; he intrudes, and takes up a position”¹⁵.

All in all, modality is seen as a resource which indicate the discourse status of speakers, the truth of their utterances; and it articulates interactional meanings beside the epistemic ones. Modality and its interpersonal functions indicate the

⁹ J. Lyons, *Semantics. Volume 1*. CUP: Cambridge 1977. As in: K. Hyland, *Hedging...*, p. 44.

¹⁰ A. Klinge, *The English modal auxiliaries: from lexical semantics to utterance interpretation*. „Journal of Linguistics”, 1993, No. 29(2), p. 315-357. As in: K. Hyland, *Hedging...*, p. 44.

¹¹ K. Hyland, *Hedging...*, op. cit., p. 44.

¹² *Ibidem*.

¹³ K. Hyland, *Hedging...*, op. cit., p. 45.

¹⁴ M. A. K. Halliday, *An introduction to functional grammar*. Edward Arnold, London 1994. As in: K. Hyland, *Hedging...*, p. 47.

¹⁵ M. A. K. Halliday, *Functional diversity in language as seen from a consideration of modality and mood in English*. „Foundations of Language”, 1970, No. 6, p. 335.

speaker's involvement in a social context¹⁶.

One of the important interpersonal aspects of modality discussed in literature is *mitigation*. It aims at the perlocutionary effect; as Fraser and Nolen state:

„Mitigation in spoken discourse include strategies such as indirectness, providing justifications and impersonalizing on one hand, and linguistic forms such as parenthetical verbs, disclaimers and tag questions on the other”¹⁷.

Speakers choose courses of action which minimize face-loss and work for their benefit, and regard the assessment of social distance more frequently. Thus, the notion of mitigation is a form of politeness which works as a strategy to avoid conflict. It not only reproduces the probability degree of a statement but also mirrors the relationship between the speaker/writer and audience¹⁸.

Hedging devices are strongly connected to epistemic modality which indicates commitment, willingness, unwillingness or the truth of the author's statements. They are commonly expressed with the usage of epistemic adjectives, epistemic adverbs as well as lexical and auxiliary verbs¹⁹. Common forms of hedging devices used in daily speech can be, e.g., *probably, sort of, perhaps*. All in all, hedges can reduce the utterance force²⁰.

Hedges can be used, either singly or in combinations. The value of statements can be reduced with the use of different strategies which may limit the confidence applied by the speaker in correspondence to, for example, drawing attention to the limitations of given results of experiment, or the conditions in which the data was obtained, or commenting on the possible or alternative interpretations, findings or difficulties encountered²¹. Although lexical items play a great role in such circumstances, Hyland concludes hedging is limited not only to lexical expressions. Qualification of writers' certainty and doubts in relation to the author's state of knowledge is also dependent on the use of if-clauses; along with contrastive markers. The passive voice and tense used can also be of similar importance in hedge statements. However, the use of these two poses difficulties when determining their effect and the motivation of the speaker²².

¹⁶ *Ibidem*.

¹⁷ B. Fraser, W. Nolen. *The association of deference with linguistic form*. „International Journal of the Sociology of Language”, 1981, Issue 27, p. 93-110. As in: K. Hyland, *Hedging...*, p. 48.

¹⁸ *Ibidem*.

¹⁹ K. Hyland, *Hedging...*, op. cit., p. 5.

²⁰ J. Holmes, *Gendered Talk at Work: Constructing Social Identity Through Workplace Interaction*. Blackwell Publishing, New York 2006, p. 39.

²¹ K. Hyland, *Hedging...*, op. cit., p. 6.

²² *Ibidem*.

Present study

The purpose of this study is to determine and define the differences between the selected examples of the language of female and male American senators on Twitter²³. Ten Twitter posts for each female and male senator were obtained in September 2018; two posts selected each week. If this assumption could not have been met, first available tweets posted in September were acquired. The study describes the corpus data variation of modality markers, hedges; text structure: tenses use, parts of speech and demonstratives distribution. Quantitative statistical surveys were conducted using the R program to check the level of significance of the presented hypotheses. The data of the corpus provide:

- gender,
- key tokens, key types, all tokens, all types,
- normalized key types and normalized key tokens,
- political party and time of service,
- modality markers, hedging markers,
- tense markers,
- parts of speech distribution,
- demonstratives distribution.

Quantitative statistical surveys were carried out to confirm the following theses:

- The difference in the quantitative use of the studied modal verbs between women and men will be evident, with a significant advantage of some tokens used by one group.
- There will be a significant quantitative difference in the distribution of parts of speech in sentences between the groups.
- Some particular demonstratives will have more frequent use in one of the groups.
- Distribution of some particular tense will be greater for one of the studied groups. Moreover, it will be supported by more frequent (quantitative) use in one of the research groups (the Democrats, the Republicans).

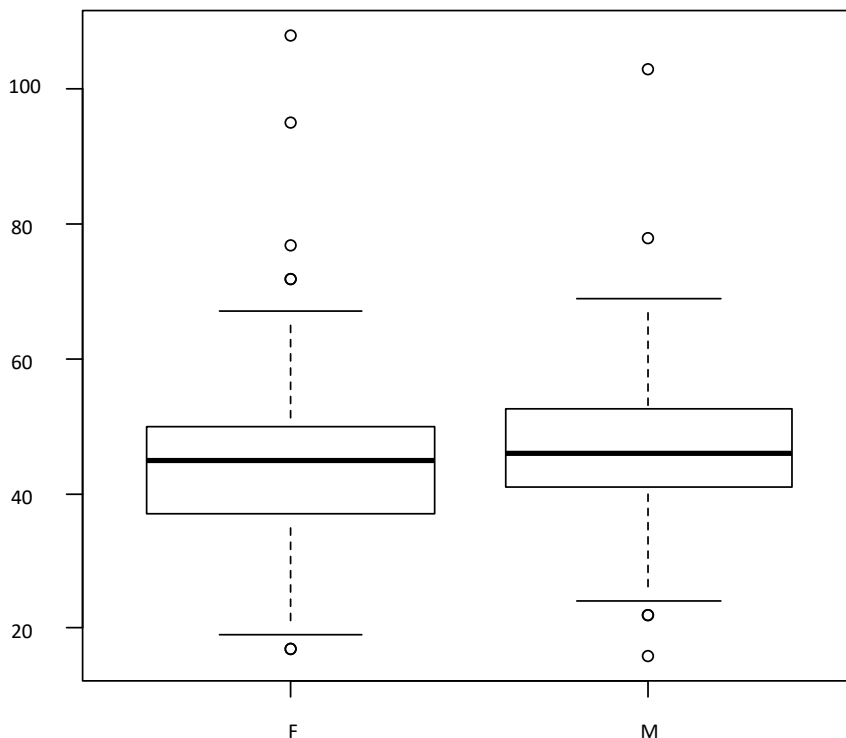
Research

The present research copied Twitter posts into Excel sheets, and saved them as .txt files for further computational approach. Then, the corpus was divided into the following sections: senators who were elected before 2010 and after 2010, senators of either the Republican or Democratic Party and, most importantly, women and men. Next, using computational techniques and programs (Prot Ant,

²³ Twitter Inc. <http://twitter.com> (01.10.2018).

Ant Conc, Tag Ant, R), the corpus data were searched and elucidated according to the research theses. The following tables and graphs show distinctions between the sections with regard to all the variables used for the purpose of this research.

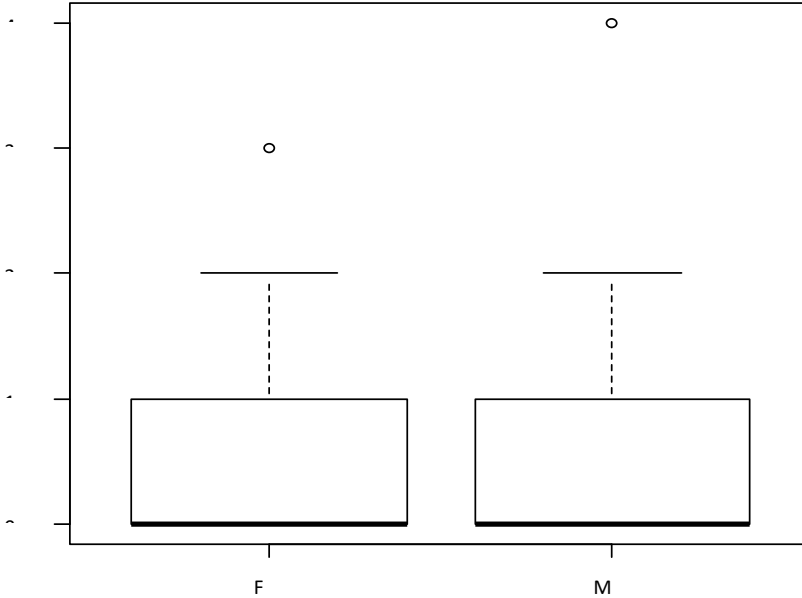
Picture 1. All tokens according to gender



Source: Author's own materials.

The first picture presents the overall word count per text for female and male members of the American Senate. As it is illustrated, the medium number of words for women and men hit 43 and 44 words, respectively. The lowest number of words used by males is lower (18 words) than that used by females (19 words). The overall range of words used by men is narrower, i.e. 25-70 words; in used by women, it is 19-68 words per text. The highest peak in women's texts hit 105 words, while in men's texts it is 102. However, as illustrated by the first graph, there are more instances of longer text written by women than by men; 5 and 2, respectively.

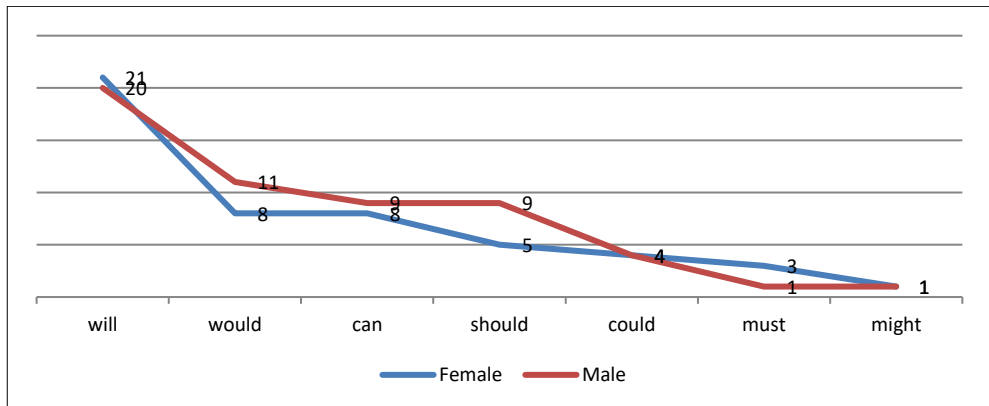
Picture 2. Key tokens according to gender



Source: Author's own materials.

The second picture illustrates the key tokens use per text. As it is visible, the medium range for both sexes is equal. Although, there is a difference in the maximum number of key tokens used in the text; 4 in male Senators' texts, and 3 in their counter group.

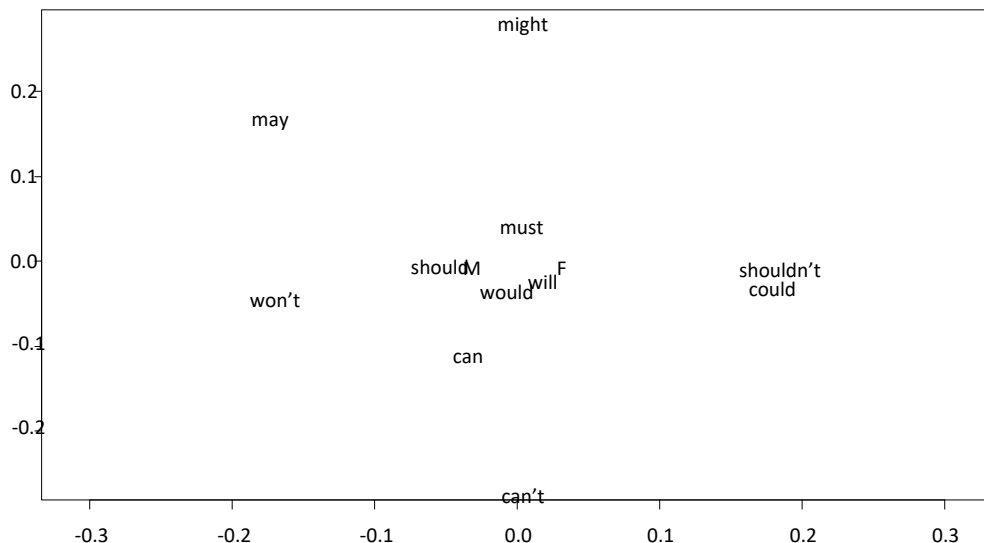
Graph 1. Modality and gender



Source: Author's own materials.

As illustrated by the first graph, there are some minor differences in usage of *should* and *would* in favor of men, and *will* and *must* in favor of women. All in all, men tend to use modal verbs more frequently than women, 55 and 50 instances, respectively.

Picture 3. Multiple correspondence analysis of the first modal verb usage



Source: Author’s own materials.

Multiple correspondence analysis of the first modal verb used in the texts (picture 3) shows the female language correspondences with *should* and *would* modal verbs. The opposite group, that is the male language, correspondence is *must* and *will* modals. Moreover, what is visible is that the negative usage of modal verbs by both groups is equally low.

Table 1. Significance test for modality

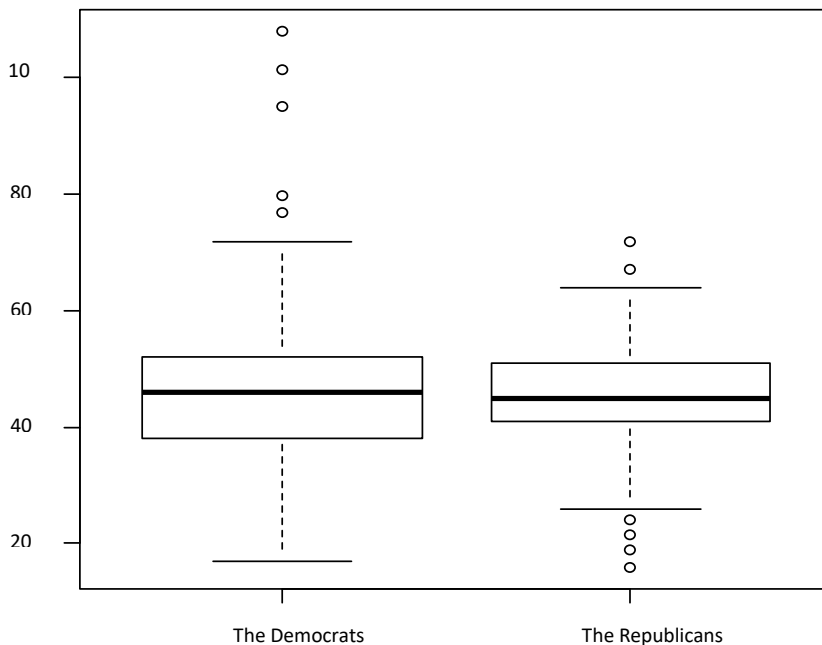
Modality	Female	Male	Pearson’s chi-squared test
will	21	20	X-squared = 3.3709, df = 7, p-value = 0.8487
would	8	11	
can	8	9	
should	5	9	
could	4	4	

must	3	1	
might	1	1	
perhaps	0	1	
Modality 2	Female	Male	Pearson's chi-squared test
will	21	20	X-squared = 1.1682, df = 3, p-value = 0.7606
would	8	11	
can	8	9	
should	5	9	

Source: Author's own materials.

The first table illustrates the score of Pearson's chi-squared test for significance in modality with regard to gender. There is no significant correspondence between the use of particular modal verbs and gender. The tests scored the p-value of 0.84 and 0.76 in men and women's texts, respectively.

Picture 4. All tokens according to the party



Source: Author's own materials.

The fourth picture presents all tokens' division in texts dependent on the party. The medium length of a text for both parties hit 43 and 44 words; for the Democrats and Republicans, respectively. The Democratic party has a wider medium range (38-55) than the Republican party (42-51). There are significantly more instances of shorter texts published in the Democratic party, and the range of words used by its members varies between 18 and 73.

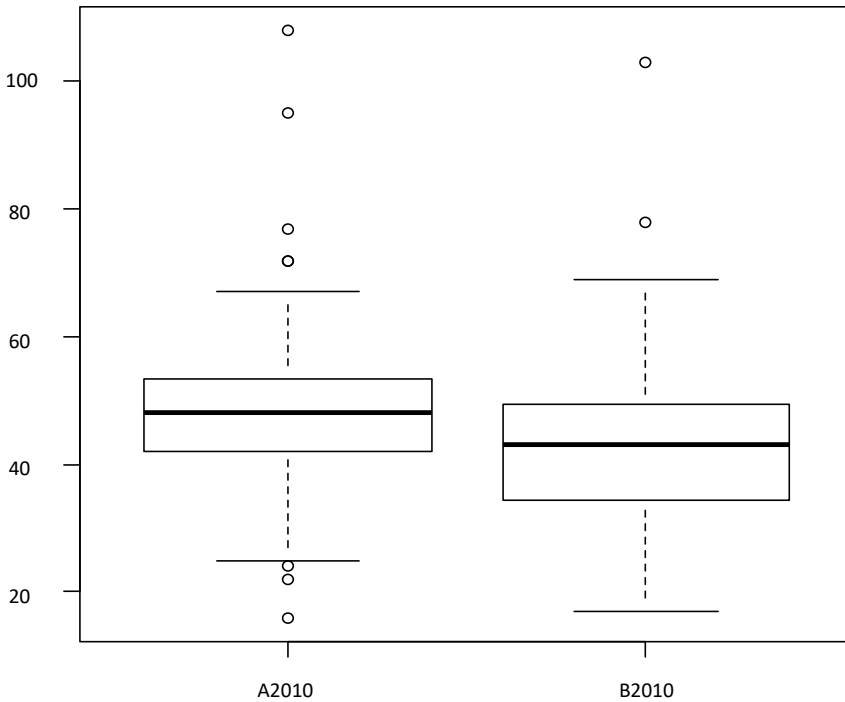
Table 2. Modal verbs distribution according to the party

The Democrats				The Republicans			
1st modal	Female	Male	Total	1st modal	Female	Male	Total
can	5	7	12				
could	2	-	2	could	1	-	1
may	-	1	1				
might	-	1	1	might	1	-	1
must	2	1	3	must	-	1	1
should	4	6	10	should	-	2	2
will	6	7	13	will	12	8	20
would	5	6	11	would	2	2	4
(none)	46	31	77	(none)	34	47	81
Total	70	60	130	Total	50	60	110
2nd modal	Female	Male	Total	2nd modal	Female	Male	Total
can	2	2	4	must	1	1	2
will	1	2	3	will	1	-	1
would	1	-	1	would	-	1	1
(none)	66	56	122	(none)	48	58	106
Total	70	60	130	Total	50	60	110

Source: Author's own materials.

Table 2 presents the party division between the selected Twitter users. As is visible, there is a tendency to use *will* in female Republicans texts. Next, an even more visible difference concerns the usage of *can* in the Democratic party, which does not occur on the Republican side. The Democrats use *would* more frequently than the opposite group. *Should* is also more frequent in the Democratic party members' texts. In conclusion, the Democrats tend to use modal verbs more frequently; 53 cases in total.

Picture 5. All tokens according to time of service



Source: Author's own materials.

Picture five visualises all tokens' spectrum in the groups of the senators elected before and after 2010. As is visible, there is a higher use of words in texts of the senators elected after 2010; 49 words medium, and 42 words for the group elected before 2010. The wider range of words used occurs in the group elected before (18-70 words); the group A2010's range is between 25-69 words. There are more instances of longer texts published in the after-2010 group; four examples. There are also fewer instances of shorter texts published in this group.

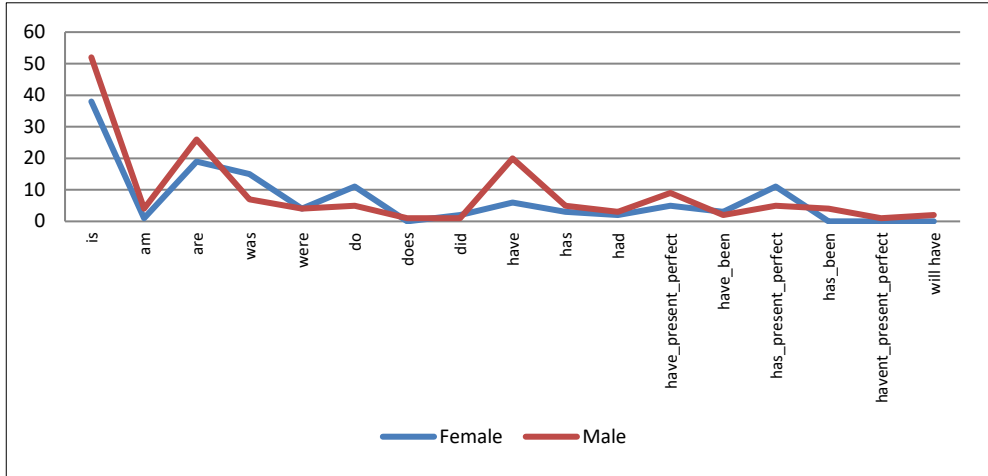
Table 3. Modal verbs usage according to time of service

1st modal	Female	Male	Total	2nd modal	Female	Male	Total
A2010	22	21	43	A2010	22	21	43
can	2	2	4	can	-	2	2
can't	1	1	2	cannot	1	-	1
could	1	-	1	can't	1	-	1
may	-	1	1	must	-	1	1
might	1	-	1	will	-	1	1
must	1	2	3	would	1	-	1
should	3	4	7	(none)	19	17	36
shouldn't	1	-	1				
will	10	9	19				
would	2	2	4				
B2010	19	22	41	B2010	19	22	41
can	2	4	6	can	1	-	1
could	2	-	2	cannot	-	1	1
might	-	1	1	must	1	-	1
must	1	-	1	will	2	1	3
should	1	4	5	would	-	1	1
will	8	6	14	(none)	15	19	34
won't	-	1	1				
would	5	6	11				
Total	41	43	84	Total	41	43	84

Source: Author's own materials.

As the third table presents, the senators elected after 2010 tend to use *should* and *will* modal verbs more often. On the other hand, those elected earlier seem to use *would* more frequently. There are only a few examples of negation, which are: *can't* (4 uses in the group of senators elected after 2010, and 1 elected before 2010) and *cannot* (1 use in each research group). Additionally, one extracted text uses *won't* in the group elected before 2010.

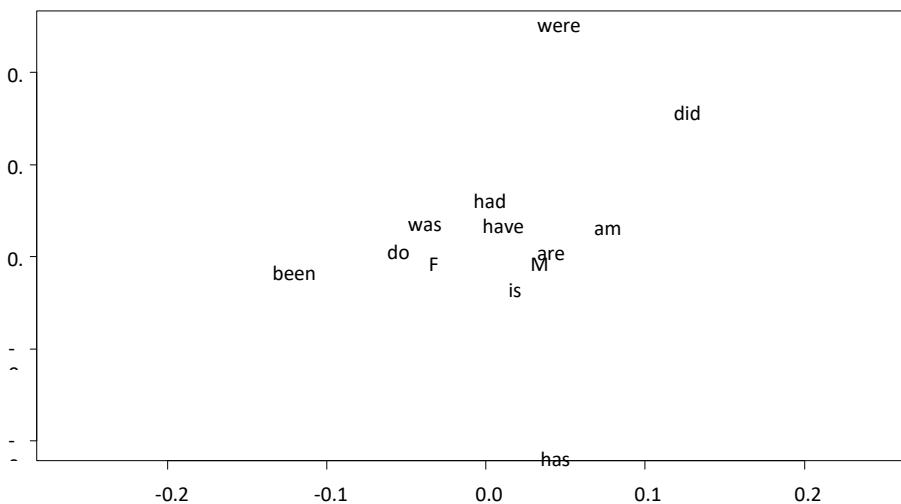
Graph 2. Tense markers for gender



Source: Author's own materials.

As illustrated by the second graph, men tend to use present tense markers (*am*, *is*, *are*) more frequently than women; 52 and 38, respectively, in the case of *is*. The same situation occurs for the verb *have* in present tense; 20 instances in males and 6 in females' group. Women, in comparison, seem to use past tense markers more frequently (*was* and *has* in present perfect). What is interesting, both groups rarely use the *am* tense marker.

Picture 6. Multiple correspondence analysis of the first tense marker of gender



Source: Author's own materials.

Picture six presents a multiple correspondence analysis of the first tense marker with regard to gender. As can be seen, the female language corresponds to the past tense markers *was* and *been* as well as to the present marker *do*. The male language corresponds to the markers of present tense (*i.e. is, am* and *are*). The usage frequency of the *have* and *had* markers is very similar in both groups.

Table 4. Tense markers for gender

Tense marker	Female	Male	Pearson's chi-squared test
is	38	52	X-squared = 28.013, df = 17, p-value = 0.04478
am	1	4	
are	19	26	
was	15	7	
were	4	4	
do	11	5	
does	0	1	
did	2	1	
have	6	20	
has	3	5	
had	2	3	
have_present_perfect	5	9	
have_been	3	2	
has_present_perfect	11	5	
has_been	0	4	
havent_present_perfect	0	1	
willhave	0	2	
past_simple+have	0	1	
Tense marker	Female	Male	
have	6	20	X-squared = 14.371, df = 5, p-value = 0.01342
is	38	52	
are	19	26	
was	15	7	
have_present_perfect	5	9	
has_present_perfect	11	5	

Source: Author's own materials.

The fourth table visualises the use of tense markers with regard to gender. As we can observe, Pearson's chi-squared test for significance equals the p-value of 0.04, which is of significance for the presented material. The simplified, second part of table four emphasises an even stronger value of the presented female and male language correspondences with reference to a given tense use. The overall p-value for the corrected data score is 0.01, which proves the quantitative difference between the female and male American senators' language use.

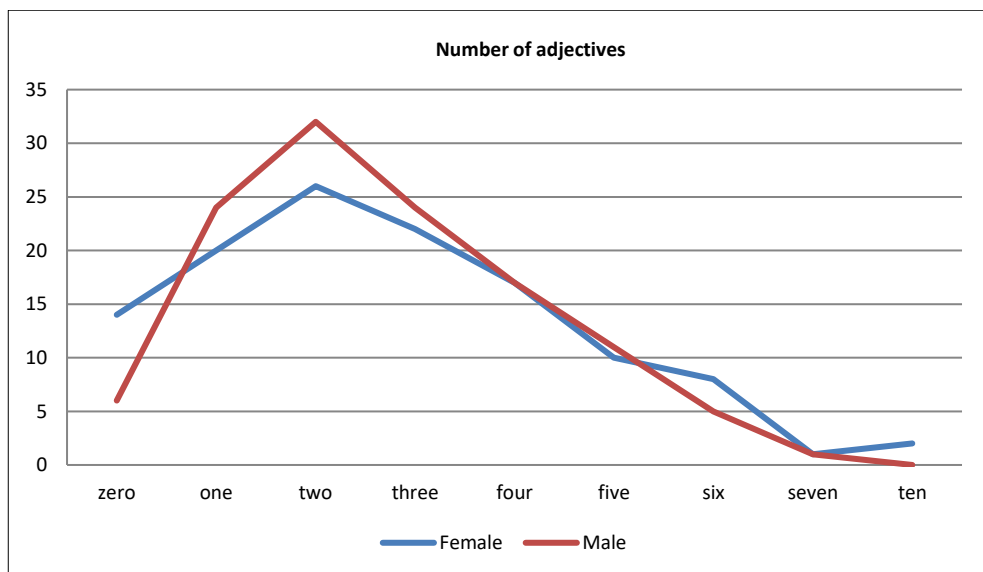
Table 5. Tense markers for particular parties

Tense marker	The Democrats	The Republicans	Pearson's chi-squared test
is	52	38	X-squared = 21.557, df = 17, p-value = 0.2024
am	1	4	
are	24	21	
was	10	12	
were	5	3	
do	12	4	
does	1	0	
did	2	1	
have	11	15	
has	3	4	
had	3	2	
have_present_perfect	3	12	
have_been	2	3	
has_present_perfect	9	8	
has_been	1	3	
modal+have	1	4	
havent_present_perfect	0	1	
hasnt_present_perfect	0	1	
Tense marker	The Democrats	The Republicans	
have	11	15	X-squared = 8.5936, df = 5, p-value = 0.1264
is	52	38	
are	24	21	
was	10	12	
have_present_perfect	3	12	
has_present_perfect	9	8	

Source: Author's own materials.

Table five shows the tense markers distribution in groups of both parties members. As we can see, there is a tendency to use present perfect in the Republican group more frequently. The Democrat party members use present tense, mainly, *is* and *do*, more frequently than the Republican ones. Pearson’s chi-squared test for significance unveils that there is no significant correspondence between the party and the tense markers in both groups, since the p-value for all tense markers equals 0.20. The second test for significance checks equal tense markers. In the simplified part of table four, the p-value of this data scores 0.12; which merits further research but, for the sake of conciseness, is not pursued in this text.

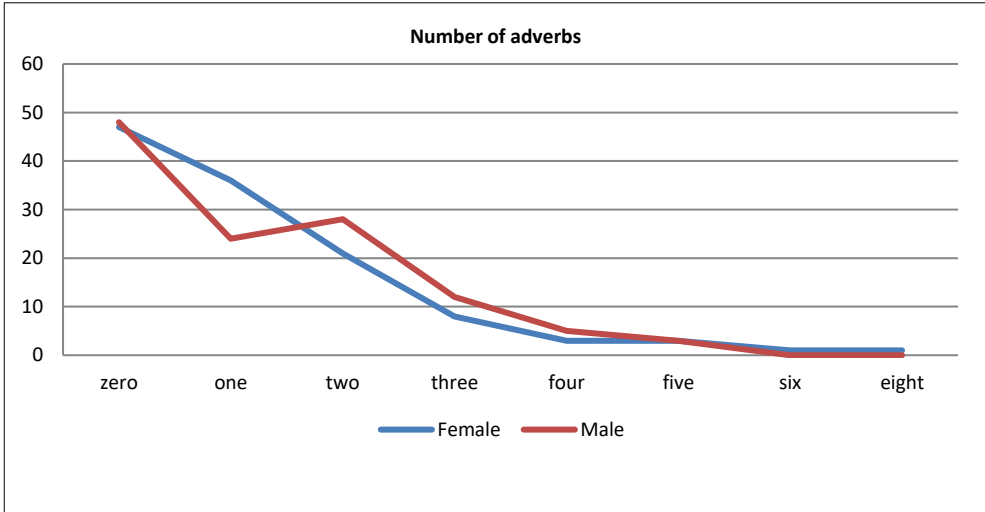
Graph 3. Number of adjectives used in texts



Source: Author’s own materials.

The third graph presents the number of adjectives used in the selected texts. As we can see, the distribution of adjectives in tweets authored by males varies between one and four and is similar in women’s counter group. However, it should be noted that there is a peak in the number of women’s texts which have five to seven adjectives (see graph 3).

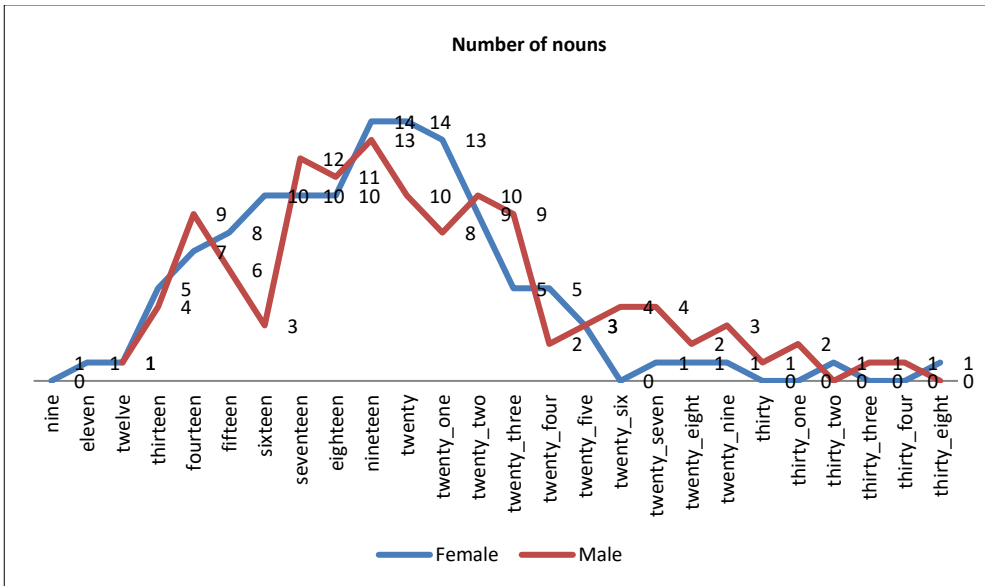
Graph 4. Number of adverbs used in text



Source: Author's own materials.

Graph number four illustrates the number of adverbs used throughout the corpus. As is shown, men use more adverbs (between two to four), while women use, mostly, one adverb per text.

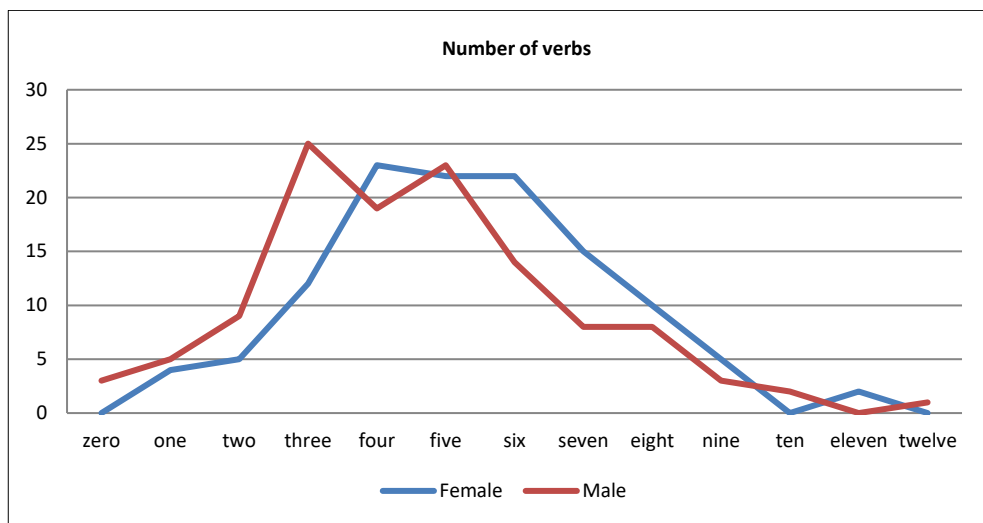
Graph 5. Number of nouns used in texts



Source: Author's own materials.

The fifth graph shows the number of nouns used in the corpus. There are three peaks in the number of nouns used in the texts of men, and three in the texts of women. The ranges for men are: 13-14, 17-19, 22-24; and for women: 13-16, 19-21, 24-25. All in all, women tend to use more nouns in their texts than men.

Graph 6. Number of verbs used in texts



Source: Author’s own materials.

Graph number six presents the data concerning the use of verbs throughout the texts in the corpus. As illustrated above, women use four to nine verbs in their texts, while men use one up to five verbs per text. There are two peaks in the texts of both men and women when it comes to their use of verbs. These are, for males, three and five and, for females, four and six.

Table 6. Parts of speech number and chi-squared tests

Number of adjectives	Female	Male	Pearson’s chi-squared test
zero	14	6	X-squared = 7.0112, df = 8, p-value = 0.5354
one	20	24	
two	26	32	
three	22	24	
four	17	17	
five	10	11	
six	8	5	

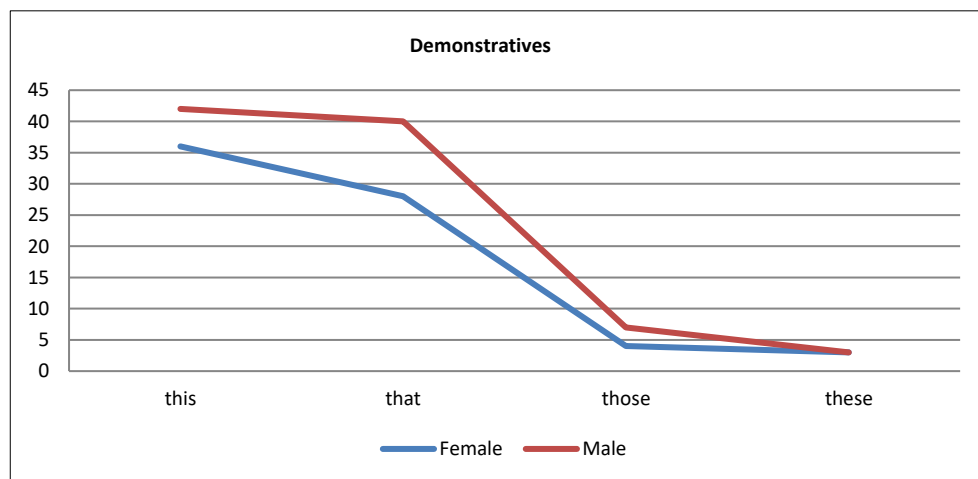
seven	1	1		
ten	2	0		
Number of adverbs	Female	Male	Pearson's chi-squared test	
zero	47	48	X-squared = 6.7105, df = 7, p-value = 0.4596	
one	36	24		
two	21	28		
three	8	12		
four	3	5		
five	3	3		
six	1	0		
eight	1	0		
Number of nouns	Female	Male		Pearson's chi-squared test
thirteen	5	4		X-squared = 7.2533, df = 10, p-value = 0.7013
fourteen	7	9		
fifteen	8	6		
sixteen	10	3		
seventeen	10	12		
eighteen	10	11		
nineteen	14	13		
twenty	14	10		
twenty_one	13	8		
twenty_two	9	10		
twenty_three	5	9		
Number of verbs	Female	Male	Pearson's chi-squared test	
zero	0	3	X-squared = 18.855, df = 12, p-value = 0.09208	
one	4	5		
two	5	9		
three	12	25		
four	23	19		
five	22	23		
six	22	14		
seven	15	8		
eight	10	8		
nine	5	3		

ten	0	2	Pearson's chi-squared test X-squared = 10.204, df = 6, p-value = 0.1163
eleven	2	0	
twelve	0	1	
Number of verbs 2	Female	Male	
two	5	9	
three	12	25	
four	23	19	
five	22	23	
six	22	14	
seven	15	8	
eight	10	8	

Source: Author's own materials.

The sixth table presents Pearson's chi-squared test scores for the parts of speech usage in female and male texts. As is visible, there are no significant correspondences in the sections devoted to the usage of adjectives, adverbs and nouns. The chi-squared tests outcome is 0.53, 0.45 and 0.70, respectively. What is important is that the usage of verbs comes close to the assumed level of significance, the p-value equals 0.09, which merits further research in this field. Although, the simplified data table (without numbers lower than 5) hit the p-value of merely 0.11.

Graph 7. Number of demonstratives used in texts



Source: Author's own materials.

Graph number seven shows data corresponding to the use of demonstratives in the corpus. As it is visible, men tend to use more demonstratives in total. Both groups use *this* and *that* demonstratives much more frequently than their opposites *these* and *those*. Seventy-eight texts, in total, use *this* demonstrative at least once; 42 in those by male authors, and 36 in female authors. *That* demonstrative hit 68 in the whole corpus; 28 and 40 instances in females and males' texts, respectively.

Table 7. Pearson's chi-squared test for demonstratives

Demonstrative	Female	Male	Pearson's chi-squared test
this	36	42	X-squared = 0.19135, df = 1, p-value = 0.6618
that	28	40	
those	4	7	
these	3	3	

Source: Author's own materials.

Pearson's chi-squared test for significance of the use of demonstratives in both research groups shows that there is no significant correspondence when it comes to the gendered use of singular or plural demonstratives in the corpus. The p-value of the test equals 0.66 point.

Conclusions

In conclusion, there are some minor differences in the usage of *should* and *would*, in favor of male Twitter users, and in *will* and *must* in favor of women. All in all, men tend to use modal verbs more frequently than women, 55 and 50 instances, respectively. Moreover, there is an evident and noteworthy advantage of given tokens' use by each of the groups; Pearson chi-squared test for its significance equals 0.84 and 0.74 in the simplified table (see table 1). The difference in the quantitative use of the studied modal verbs is slight and cannot be taken into consideration as significant.

Women use past tenses more frequently, while men use present tenses with greater frequency. The tense markers comparison scored the p-value of 0.04, while specified present and past tense markers have the p-value equal to 0.01. The distribution of the particular tense markers, as presented in table 4, is more frequent in both groups and proved to be significant with the use of Pearson's chi-squared test. Although it was not supported by more frequent use in research groups in the party section, the chi-squared test resulted in the p-value of 0.20, and

0.12 in the second test (for the same data labelling as in the gender section). The party section tests can settle the basis for further research in this field.

There is a slight significance in the number of verbs used in the posts authored by women. Women use between 4 and 8 verbs per tweet most frequently; while men use 3-6 verbs most often. The significance of verbs usage in Twitter posts scored the p-value of 0.09. All in all, it is not proved to be significant by Pearson's chi-squared test, as assumed by the researcher in the hypothesis section.

Demonstratives corresponding to singular nouns have more frequent use in both of the research groups. Men use all demonstratives more frequently, but Person's Chi-squared result does not prove it to be significant, since the p-value equals 0.66.

Bibliography

- Brown P., Levinson S. C., *Politeness: Some Universals in Language Usage*. CUP: Cambridge, 1987.
- Fraser B., Nolen W., *The Association of Deference with Linguistic Form*. "International Journal of the Sociology of Language", 1981, Issue 27.
- Halliday M. A. K., *Functional Diversity in Language as Seen from a Consideration of Modality and Mood in English*. "Foundations of Language", 1970, No. 6.
- Halliday M. A. K., *An Introduction to Functional Grammar*. Edward Arnold: London, 1994.
- Holmes J., *Gendered Talk at Work: Constructing Social Identity Through Workplace Interaction*. Blackwell Publishing: New York, 2006.
- Hyland K., *Hedging in Scientific Research Articles*. John Benjamins Publishing Company: Amsterdam, 1998.
- Hyland K., *Writing Without Conviction? Hedging in Scientific Research Articles*. "Applied Linguistics", 1996, No.17, p. 433.
- Klinge A., *The English Modal Auxiliaries: From Lexical Semantics to Utterance Interpretation*. "Journal of Linguistics", 1993, No. 29(2), p. 315-357.
- Lyons J., *Semantics*. Volume 1. CUP: Cambridge, 1977.
- Sunderland J., *Language and Gender: An Advanced Resource Book*. Routledge: London, 2006.
- Vuković M., *Weak Epistemic Modality in the UK Parliamentary Discourse*. "Journal of Interdisciplinary Approaches to Text", 2014, No.1, p. 121-137.

Online references

- Laurence A., AntConc x32 version 3.5.7., 23 Apr. 2018. Freeware software.
<http://www.laurenceanthony.net/software/antconc>.
- Laurence A., ProtAnt x64 version 0.1., 21 Mar. 2017. Freeware software.
<http://www.laurenceanthony.net/software/protant>.
- Laurence A., TagAnt x64 version 1.2.0., 15 Sep. 2015. Freeware software.
<http://www.laurenceanthony.net/software/tagant>.
- The R Foundation for Statistical Computing, R x64 version 3.4.1., 30 Jun. 2017. Freeware software.
<http://cran.r-project.org>.
- Twitter Inc. <http://twitter.com> (01.10.2018).

**Modality and Text Structure in American Senators' Tweets:
A Corpus Study of Gendered Lexical Items**

Summary

In many cases, procedures, behaviors and conceptions of any particular relationship within the linguistic domain depend on the words' ability to represent them, whereas people depend on their lexical abilities to present themselves in the web. The purpose of this study is to determine and define the differences between the selected examples of American senators' (men and women) language on Twitter. Corpus and computational linguistics methods were used to find answers to the assumed questions. Research material selected for this study was collected in September 2018 and comprises 240 tweets of 24 American senators; of both sexes equally. The text discusses the corpus data differentiation and quantitative statistical surveys conducted in the R program to check the level of the significance of the presented hypotheses.

Keywords: gender, corpus linguistics, computational linguistics, lexical items, modality, tweets, American senators