

English language (UDC 811.111)

Мамаев Михаил Михайлович, Василевич Александр Петрович

Московский государственный областной университет, Московский государственный областной университет
ОБ ОДНОМ СПОСОБЕ ВЫЯВЛЕНИЯ ГЕНДЕРНО ЗНАЧИМОЙ ЛЕКСИКИ В АНГЛОЯЗЫЧНОМ ТЕКСТЕ

Аннотация

Гендерные особенности языка проявляются, прежде всего, на лексическом уровне. При этом основным исследовательским приемом является оценка употребительности: одни слова чаще встречаются в текстах, написанных женщинами, другие – в текстах мужчин. Основанная на этом подходе процедура гендерной атрибуции текста была опробована на обширном англоязычном материале (22 автора) и оказалась достаточно эффективной. Расположение авторов на «шкале маскулинности» позволяет ставить и решать ряд нетривиальных задач. В частности, выдвинута гипотеза о том, что по сравнению с авторами XIX века степень маскулинности современных авторов-мужчин в целом значительно снизилась, а авторов-женщин – повысилась.

Ключевые слова: английский художественный текст, гендерные особенности, гендерная атрибуция текста, употребительность слова.

Mamaev Mikhail Mikhailovich, Vasilevich Alexander Petrovich

Moscow Region State University, Moscow Region State University

ON THE PROCEDURE TO REVEAL GENDER-INTRINSIC LEXICS IN ENGLISH FICTION

Abstract

Gender peculiarities become apparent essentially at the lexical level. So evaluating frequency of use becomes the main research method: one set of words are more frequent in the texts written by women, other ones are more frequent in men's texts. The procedure of gender attribution based on this approach was tested on a large sample of English texts (22 authors) and proved to be rather effective. The placement of authors on the "scale of masculinity" allows to put forward and solve a number of noteworthy problems. In particular, the hypothesis was advanced that compared to the 19th century authors the level of masculinity of contemporary male authors has decreased while the level of masculinity of female authors increased.

Keywords: English fiction, gender peculiarities, gender text attribution, word frequency evaluation.

The issue of sex and gender has become very important nowadays and it is a matter of utmost interest not only scientifically but also in sociocultural sphere, sometimes even on the legislative level. Sex is regarded as a biological phenomenon; consequently men are opposed to women on the basis of purely biological characteristics, including behavioral peculiarities. As to gender, it touches upon the psychological features and in this respect contrasted notions will be masculinity and femininity (Kon 2004).

In most cases sex and gender coincide yet some examples of the opposite have been reported [1].

Since gender is a sharply defined framework of social and psychological settings, it is assumed to affect the *linguistic behavior* of an individual (Maslova 2004, p. 124).

Introducing the parameter of gender into the linguistic research has opened new prospects in the analysis of various aspects of language and speech. The term itself emerged in linguistics in the 1980s, i.e. a bit later than in other humanities – History, Psychology, and Sociology. The ideas of gender linguistics, put forward by various schools and movements, are being still moulded in a system (Mamaev 2011).

It must be stressed that the human being not only understands the meaning of a judgment uttered but realizes his or her involvement in it. Hence, the word becomes a cultural archetype and is regarded as an issue that both the outer world and the speaker have impact on (Lebedev 2008). It seems clear, then, that gender speech peculiarities are most likely to manifest themselves.

The study of gender aspects has become extremely popular among linguists lately. Up to the present efforts have been focused on specific features of men's and women's speech. For example, among feminine linguistic facts the researchers enumerate frequency of usage of euphemisms; adverbs helping to avoid the utterance categoricity (e.g. *rather, quite*); exclamatory sentences, tag questions, etc. Regarding masculine facts they reckon the pursuit for word creation; tendency to the usage of archaic, jargon and dialectal words; omissions of articles and auxiliary verbs; etc. (Antineskul 2001).

From numerous published results one may infer that gender peculiarities emerge, mostly, **on the lexical level**. Ordinarily, words or phrases are considered to be feminine if they appear substantially more frequently in the texts of female writers. Consequently, one should pay utmost attention to the *usage frequency* of language units.

The reverse approach may be of interest, too: basing on the lexical statistics it is possible to study the opposite task: is it possible to specify gender attribution of the text, i.e. to predict whether its author is a man or a woman. That would make the main task of our study: we undertake to find, in any random text, a group

of lexical elements which would allow its gender attribution with a high degree of reliability.

Although it was ascertained that the usage of slang and Latin terms are characteristic of the male written speech, we should bear in mind that these phenomena are not frequent in any text. The same can be said about feminine texts which should presumably include many euphemisms and tag questions. The required solution would be turning to **function words**.

First of all, function words by their nature are very frequently used and are met in any type of texts. It is no less important that they are limited in number is limited and, thus, the process of analysis wouldn't be too laborious.

It is worth noticing that gender attribution of function words hasn't been practically studied. We can cite only one attempt to dwell with the problem. It was done on English material: (Argamon, Koppel 2006; Koppel, Argamon, Shimoni 2001).

The authors are trying to solve the problem of gender attribution by dividing function words into 'mainly male' or 'mainly female' words [2]. The algorithm was introduced which is to identify the gender of the text's author with high probability. It has been tested on the large amount of text corpora and defined gender attribution with at least 80% success rate. The result seems good enough, however there still remains some opportunity to further perfect the procedure of the algorithm application.

According to the algorithm, the articles (*a, the*) as well as demonstrative pronouns (*that, these*) are considered masculine indicators, while a group of pronouns (*I, you, she, her, their, myself, yourself, herself*) indicate the belonging to the text written by a female author. The gist of algorithm is the evaluation of frequency of every word from the list. **The frequency data are supplemented with the system of coefficients, which regulates the contribution of every word to the final result.** For example, the coefficient, or the 'weight' of preposition *with* is 52, while pronoun *who* 'weighs' only 19, article *the* – 7, etc. Thus, if *with* is found in the text 4 times, then its total contribution will be 208 (4 x 52); if pronoun *who* isn't present in the text at all, it will have zero contribution; and if article *the* appears 69 times, then its contribution will be rather substantial (69 x 7) – 483.

The total sum of weights for masculine words (*who, the, as, etc.*) and feminine ones (*with, if, not, etc.*) are counted and the results are summed up. If the total sum of feminine words turns out to be more than that of masculine ones then the text is attributed as feminine.

In spite of the authors' claim that the algorithm functions rather efficiently, their suggested set of function words gives rise to some doubts. Particularly, it is not clear why the authors include forms of the verbs *to be* and *to say* in the list. The usage of *to* is also causes scrutiny: it is clear that there is serious difference if *to* functions as a

particle (with the infinitive) or a preposition (the calculations would not take this difference into consideration at all).

With a glance to the above, we made up our minds to test the algorithm on the new selection of texts. There were selected 11 novels of British authors of the 19th–20th centuries (cf.: G.K. Chesterton ‘Father Brown Stories’ / J. Austen ‘Pride and Prejudice’, etc.) and 11 modern novels of the 20th–21st centuries (cf.: P. Ness ‘A Monster Calls’ / M. Blackman ‘Noughts And Crosses’, etc.). Taking into consideration the character of the material (function words), the corpus seems quite representative. We only confined

oneself to the analysis of three fragments from every novel – in the beginning, in the middle and at the end accordingly. The amount of every fragment was nearly the same – 1500 words approximately. The list of novels is given in the end of the article.

The processing of the texts was conducted with the help of the Gender Genie algorithm was applied to each text [3]. The algorithm allows every Internet user process a text, automatically highlighting the words relevant for the authors’ gender attribution.

The example of the initial matrix is presented in Table 1.

Table 1 – The initial matrix of the calculation results (fragment)

	Feminine words					Masculine words			
	Word	Was used, times	Weight	Total contribution		Word	Was used, times	Weight	Total contribution
Charles Dickens	her	27	20	540	the	209	6	1254	
	me	26	20	520	a	116	10	1160	
	your	12	40	480	what	15	35	525	
	
	with	43	1	43	many	4	6	24	
	myself	2	4	8	more	7	2	14	
	where	3	2	6	below	1	8	8	
	Total					5456			
Ann Brontë	her	61	20	1220	the	170	6	1020	
	me	65	18	1170	as	32	30	960	
	not	65	8	520	a	86	10	860	
	should	8	50	400	what	23	35	805	
	
	myself	17	4	68	more	7	2	14	
	with	51	1	51	these	1	8	8	
where	3	2	6	many	1	6	6		
	Total					4939			

According to Table 1, the sum of masculine words prevails in Ch. Dickens’ text (5456 > 4179), consequently, this text is written by a male author. On the contrary, in A. Brontë’s text feminine words prevail (5400 > 4939), so the text is obviously written by a female. As we see, the program has worked well in both cases.

However, having analyzed all the 22 texts, the results didn’t turn out that successful. Suffice it to say that the program has correctly attributed only 6 out of 11 texts by male-writers, i.e. made mistakes in nearly a half of the cases. Keeping to the hypothesis that basing on a limited number of function words we can make gender attribution of the text correctly enough, we have to concede that either some words from the given list aren’t appropriate, or the weight system needs correction, or both are true.

Our next step had two main aims:

- 1) making sure that the initial division of words into masculine and feminine was accurate and suggesting corrections, if necessary;
- 2) improving the weight system of the words used.

We summarized the initial matrix data of all the 22 authors. In case the algorithm attributes a word to the feminine group, than its total usage frequency in female-writers’ texts should substantially exceed that in male-writers’ texts, and vice versa.

The analysis shows that the confirmative data have been received for 10 words out of 15 (e.g. *me* was found 360 times in female-writers’ texts and only 180 times in male-writers’ texts; *your* 138 : 74 and so on). For 4 words no preferences have been spotted (e.g. *with* – 360 : 360, *when* – 160 : 150), and in one case there is even an opposite result (*was* – 860 : 1033).

Still more disappointing was the result for the masculine words. The confirmation was acquired for only 6 words (e.g. *the* – 3136 : 2330 or *as* – 486 : 293). In the rest 10 cases data were nearly the same for males and females (*these* – 36 : 35, *below* – 7 : 5, *many* – 24 : 20) or even with the prevalence for females (*to* – 1083 : 1240, *are* – 91 : 132).

In consideration of these data we have taken into account only the cases with the significant difference in the usage frequency of words in male- and female-writers’ texts (10 feminine and 7 male words). The rest of the words from the initial list were eliminated.

Our next task was to correct the weight system of the 17 words.

The way of reasoning in applying weights to words by the algorithm has been quite dim to us. Here are a few examples (Table 2)

Table 2 – The analysis of the lexical units’ weights suggested by the algorithm

	Word	Found in male- authors’ texts	Found in female- authors’ texts	Word weight
1	should	32	48	50
	around	43	49	10
	myself	37	51	4
2	we	102	169	45
	what	136	190	35
	if	118	146	28
3	a	1347	1111	10
	the	3136	2330	6
	to	1083	1240	2

The groups (1–3) contain the words of nearly the same usage frequency. As one can see, the weights in every group differ significantly, so the necessity to somehow correct the weight system seems obvious.

In working out the new weight system we took two main factors into consideration. The first one is the range of frequency difference. For example, the words in Group 2 have more differentiating force than the words in Group 1, and consequently they should be given more weight.

On the other hand, some words (especially articles) have considerably higher frequency, and applying much weight to them will cause their taking the substantial part of the general result, minimizing all other words’ role. In setting a proper weight, we

thought it reasonable to ensure that a word’s contribution to the whole sum shouldn’t exceed 25%. So, we applied the weights of 4 and 6 to the articles *a* and *the* respectively, while *if* got the weight of 28, and *myself* – 50.

Lastly, we have perfected the procedure in one more aspect. The initial algorithm only allows to attribute a text as ‘male’ or ‘female’. We’d like to go further and arrange the texts on a kind of “scale of masculinity”. This will allow not only distinguish a male text from a female one but also compare two texts of the same gender. As a result, we’ve introduced a **masculinity index M**, which may serve as an important feature of the author’s gender style. Index **M** is based directly on the masculine / feminine words’ indicator and is calculated according to the formula:

$$M_i = \frac{masc_{ai}}{av_{masc}} - \frac{fem_{ai}}{av_{fem}}$$

where M_i – index of masculinity of author (*i*);

$masc_{ai}$ – total sum of masculine words in author *i*’s text;

fem_{ai} – total sum of female words in author *i*’s text;

av_{masc} (and av_{fem}) – total sum and masculine (and feminine) words on the average in the whole group of texts.

Here comes an example of calculations for texts by G.S. Chesterton and A. Brontë. In Table 3 one can see fragments of initial data matrices.

Table 3 – Initial data matrices (fragments)

	Feminine words				Masculine words				
	Word	Frequency	Weight	Total	Word	Frequency	Weight	Total	
G.S. Chesterton	we	7	40	280	the	292	4	1168	
	be	22	10	220	a	163	6	978	
	myself	4	50	200	as	44	20	880	
	
	she	1	12	12	around	0	20	0	
	Total				1585	Total			
A. Brontë	me	65	20	1300	the	170	4	680	
	her	61	15	915	as	32	20	640	
	myself	17	50	850	a	86	6	516	
	
	be	17	10	170	around	2	20	40	
	Total				5515	Total			

Table 4 illustrates the stages of calculating the index **M**.

Table 4 – The calculations of parameters for the index **M**

	mas _{C_i}	av _{masc}	$\frac{mas_{C_i}}{av_{masc}}$	fem _i	av _{fem}	$\frac{fem_i}{av_{fem}}$	M
Chesterton	4572	3695	1,237	1585	3704	0,428	0,809
A. Brontë	2684		0,726	5515		1,489	-0,763

Table 5 – Index **M** of the whole group of authours

Male authors	mas _{C_i}	fem _i	M	Female authors	mas _{C_i}	fem _i	M
O. Wilde	6414	3348	0,83	J.K. Rowling	3904	3572	0,09
G.K. Chesterton	4572	1585	0,81	J. Cox	3268	3368	-0,02
R. Stevenson	5046	3246	0,49	M. Rosoff	3876	3991	-0,03
B. Bryson	3520	2083	0,39	Ch. Brontë	3048	3188	-0,04
A. Conan Doyle	5308	4064	0,34	M. Shelley	3148	3724	-0,15
P. Ness	3462	2396	0,29	A. Christie	2800	3503	-0,19
M. Peet	3220	2294	0,25	J. Valentine	3428	4968	-0,41
M. Burgess	2802	2095	0,19	A. Fine	4088	5836	-0,47
Ch. Dickens	3424	3509	-0,02	M. Blackman	3166	5661	-0,67
W. Thackeray	3764	4015	-0,06	J. Austen	2704	5446	-0,74
K. Brooks	3650	4088	-0,12	A. Brontë	2684	5515	-0,76

According to Table 5, index **M** attributes the texts rather accurately. Certainly, as it usually occurs, there are some borderline results when **M** approaches zero (in this category we find 3 male and 4 female authors; they are marked grey). Perhaps, further improvements of the procedure might reduce the number of such cases, however efforts in another direction seem no less promising. In fact, as we mentioned earlier, gender reflects an individual’s linguistic consciousness, and it shouldn’t inevitably coincide with his or her biological sex. Consequently, if of an author’s index **M** stands out from the rest, this should cast some doubt on his or her gender identity. To clarify the case, one may need additional

facts from related sciences – Literary Criticism, Psychology, etc. From this point of view it would be interesting, for example, to study the results of Brontë sisters whose texts turned out to be in opposite ends of the “scale of masculinity”.

In conclusion let us raise one more point. As it is seen from the list of works, our selection of fiction texts is balanced not only according to gender (11 male and 11 female authors), but also according to the time of writing (11 texts of 19th–20th century and 11 texts of 20th–21st century). This enables us to juxtapose the index **M** values of the two epoques (Table 6).

Table 6 – The correlation between index **M** and the epoque of writing (average data)

	Epoque	Masculine words	Feminine words	M
Males	19 th –20 th	4755	3294	0,34
	20 th –21 st	3363	2591	0,25
Females	19 th –20 th	2877	4275	-0,40
	20 th –21 st	3622	4566	-0,21
Range	19 th –20 th	0,74		
$\frac{M_{male\ authors}}{M_{fem\ authors}}$	20 th –21 st	0,45		

One can definitely infer, from the data, that gender differentiation among the 19th century writers is more distinct (the range is 0,74). Among contemporary writers we see some tendency for convergence, the difference between male and female authors being only 0,45.

No doubt, such facts call for more research involving new literary material.

Certainly, the urgent problem might be to develop the procedure for defining the index **M** for the material in Russian. If it is successful, this may help discover new possibilities – for comparative analysis in the first place. In particular, a very interesting task would be to analyze the cases when, for example, the text written by a female-author is translated by a male-translator and vice versa.

[1] It’s worth mentioning that English is among the few languages where “sex” and “gender” are differentiated on the lexical level.

[2] In this study the class of function words includes pronouns and modal verbs (30 words in total).

[3] The Gender Genie. URL: <http://bookblog.net/gender/genie.php> (date of reference: 19.02.2013).

References

1. Antineskul O.L. Gender as text-formation parameter: Special course teaching aid. – Perm, 2001.
2. Kon I.S. Sexology. Teaching aid. – M., 2004.
3. Lebedev S.A. Historical-philosophical methodological introduction in linguistic theory // Philosophy of social and studies and humanities. – M., 2008.
4. Mamaev M.M. Multiple-aspect character of studying gender factor in linguistics // Bulletin of Moscow State Regional University. Series "Linguistics", №2. – M., 2011.
5. Maslova V.A. Lingvoculturology. – M., 2004.
6. Argamon S., Koppel M., Fine J., Shimoni A.R. Gender, Genre, and Writing Style in Formal Written Texts // Interdisciplinary Journal for the Study of Discourse. 2006, vol. 23, Issue 3, pp. 321–346.
7. M. Koppel, Argamon S., Shimoni A.R. Automatically determining the gender of a text's author // Bar-Ilan University Technical Report BIU-TR-01-32. 2001.

General questions relating to both linguistics and literature. Philology (UDC 80)

Шустова Ирина Николаевна

Воронежский государственный педагогический университет

АКСИОЛОГИЧЕСКИЙ ПОТЕНЦИАЛ ПОЛИТИЧЕСКИХ ПРОЗВИЩ

Аннотация

Статья посвящена изучению аксиологических возможностей политических прозвищ. Политические прозвища очень экспрессивны. Они могут быть референтными и безреферентными. Некоторые прозвища утрачивают своё исходное значение и входят в состав оценочной политической лексики. В языке политики прозвища часто выступают не только средством оценки, но и становятся идеологическим оружием.

Ключевые слова: прозвища, аксиологические возможности, выразительные слова, средство оценки, аббревиатура, референтные и безреферентные прозвища, оценочная политическая лексика, коннотация, язык политики.

Shustova Irina Nikolaevna

Voronezh State Pedagogical University

AXIOLOGICAL POTENTIAL OF POLITICAL NICKNAMES

Abstract

The article is devoted to the study of axiological possibilities of political nicknames. Political nicknames are very expressive. They can be personal and impersonal. Some nicknames lose their primary meaning to become a part of evaluative political lexis. In the language of politics nicknames often serve not only as means of assessment, but also as ideological weapon.

Keywords: nicknames, axiological possibilities, expressive words, means of assessment, abbreviation, personal and impersonal nicknames, evaluative political lexis, connotation, language of politics.

The language of modern politics tends to be less formal. On the contrary, it becomes more popular among politicians to speak like common people do, so that they would be recognized as close to the electorate. This tendency prevails especially in the USA where most people, figuratively speaking, got used to "judge a book by its cover" and are not prone to intellectual reasoning. During the electoral period politicians are preoccupied with their image and try hard as they can to win the popularity. Language in this case is one of the most important means of producing certain image and achieving success.

Regarding the fact that the language of a successful politician must be simple, smart and bright, speechwriters and image makers are very particular about the word choice. Specialists advise to choose such words and phrases that can become memorable, fresh and entertaining at the same time. That is why the language of politics contains a lot of expressive words and phrases endowed with axiological potential. Among the words having rich evaluative possibilities are political nicknames which meet all the above mentioned requirements.

Nickname – a name used informally instead of a person's own name, usu. a short form of the actual name or a name connected with one's character or history. Nicknames are often given at school to annoy or upset other children, and many last into adult life (Longman Dictionary, p.899).

Oxford English Dictionary gives the following definition of the word 'nickname': 'A name or appellation added to, or substituted for, the proper name of a person, place, etc., usually given in ridicule or pleasantry' (Oxford English Dictionary, 1989).

It is evident that such words as nicknames possess high axiological possibilities because they serve not only as means of assessment, but also correspond the evaluative attitude, forming the opinion concerning what is good and what is bad in the society. The negative potential of nicknames can be applied while characterizing political opponents. As for political supporters or the politician himself, nicknames with a humorous colour or positive connotation are often used. Certainly it depends on the purpose of giving a nickname.

Many American politicians especially presidents have

nicknames. It can be a short form of the personal name, for example, Ike – Eisenhower, Teddy – T. Roosevelt, Bill – W. Clinton and so on. Another form of a nickname is the abbreviation of the full name of a person, for example, JFK – J. Kennedy, GWB – George Bush – junior.

As for George Bush – junior and his father George Bush – senior, we noticed that these politicians are sometimes named as the numbers of their presidential periods – Number 41 and Number 43. For example,... Although it is currently fashionable to lampoon Number 43 for his verbal gaffes, we know that Number 41 was in a class of his own (The Daily Telegraph, Feb.14, 2004, p.6).

Anyway, nicknames make presidents closer to people. And it is their main function together with the function of assessment in the language of politics.

In course of our research we also noticed that political nicknames can be personal when there is a reference to a real person and impersonal when it is applied to a group of people or a political party or a political movement.

For example, Robbery Hillham/ Hilla the Hun are offensive nicknames given to Hillary Clinton during the period of political fight when she unsuccessfully attempted to become President of the USA. Both nicknames are personal and characterize Mrs. Clinton from the negative side, eventually contributing to the destruction of her positive political image.

Concerning impersonal nicknames, we also provide some examples. For instance, hawks and doves used to be nicknames but later became ideologically loaded political words.

Hawk – 2. a person who believes in strong action or the use of force, esp. one who supports warlike political ideas.

Dove – 2. (in politics) a person in favour of peace and compromise. (Longman Dictionary, pp.382, 605).

Other examples are donkeys and elephants, boll weevils and gypsy moths. These are nicknames of American political parties – Democratic and Republican.

Gypsy moths – those liberal and moderate Republicans in the US House of Representatives who tend to deny support to President Ronald Reagan's domestic and foreign policies. They are called gypsy moths, in contrast to boll weevils, after a leaf-eating moth found in the north,