

Text analysis of Maxpark and LiveJournal Russia: How is the evaluation of modern femininity  
and masculinity discussed in Russian blogs

by

Elena Dergacheva

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Arts

Humanities Computing - Modern Languages and Cultural Studies

University of Alberta

©Elena Dergacheva, 2014

## ABSTRACT

This thesis focuses on text analysis of Russian speaking blogs with the goal to examine critically discussions of femininity and masculinity in modern Russian society. The data corpora are comprised of blogs and comment entries from websites Maxpark.com and LiveJournal Russia. The central question of this research asks if the degree of patriarchal values remains higher in the evaluation of femininity concepts in comparison to masculinity. A combination of quantitative and qualitative approaches is used in data analysis to detect key topics of discussion within the research question. The quantitative modes of inquiry in this research are such methods of text analysis as keyword extraction, collocations, feature extraction (sentiment analysis) and text visualization. The qualitative part includes classic literary and linguistic analysis together with intuition.

## ACKNOWLEDGEMENT

I would sincerely like to thank Dr. Geoffrey Rockwell and Dr. Peter A. Rolland for their academic guidance, practical advice and fruitful discussions. Special thanks are due to Mayya Sharipova, a Ph.D. candidate in Computer Science from University of Saskatchewan, who contributed to the technological part of my research by building the programming code for text analysis of the Russian language blogs. Also I am very grateful to Dmitry Paranyushkin, a programmer from Nodus Labs, whose open source text visualization program I used in this research.

## TABLE OF CONTENTS

INTRODUCTION	1
CHAPTER I. LITERATURE REVIEW	4
Olga Zdravomislova’s research on woman discourse in media	6
Concept of gender in Russian folk culture	8
Changes in the gender culture from XVIII to XXth century	10
Ambiguity of modern womanhood and manhood	11
Blogs as a special communication and information environment	12
Blog as a Unique Genre and Social Space	13
Blogs Are A Studied Minuet Between Blogger And Audience	15
Blogging activity in Russia	16
From text to corpus and from instance to system	18
Technology as a trigger from the concept of parole to langue in text studies	21
Aesthetics of data	24
Algorithmic criticism with humanistic inquiry	27
CHAPTER II. METHODOLOGY	29
Hypothesis: “There is more traditionalism in the evaluation of femininity in modern Russian women than in the evaluation of masculinity in modern Russian men.”	31
Definition of traditionalism in traditional Russian gender culture	33
Statistics of websites	34
Scenario for research: Corpus – driven approach with hypothesis and categorization at the beginning and analysis with generalization at the end	41
Help of Text Analysis Software	43
Technological formula for research: Tokenizing+Annotating +Calculating+Extracting with the use of NLTK for Python and TreeTagger Russian version	44
Process of creating the lists of negative and positive personality traits for sentiment analysis	45
Principles of TreeTagger program in processing the data corpora	46
Keyword Method and its Collocations	47
Text visualization based on node-edge structure with the use of Texttexture	48
CHAPTER III. ANALYSIS	50
Overview of the “blog” development throughout the history	50
Statistics of RuNet	52
Russia is the most active user of LiveJournal	53
LiveJournal as a point of Russian collectivism and resistance at the same time	53
Maxpark – a mirror of social and political life of the country	55
Programming Code for Sentiment Analysis of Masculinity and Femininity	56
Process of Feature Extraction in the Corpora	57
Masculinity and Femininity in Russia: From Past To The Present Day	58
Calculation of Keywords/Collocations and its Frequency for Further Analysis	61
Discussion of beauty on the physical level in femininity as well as masculinity	70
Male needs given more priority than female ones	72
Discourse of female strength focused on career but not family	73

Main female need is to be needed. Adjective “needed”, words with root “need”	73
Female beauty as a double-edged weapon	74
Women hunting after men in the animal farm	79
Women more concerned about family issues compared to men	80
Negative attitude towards both modern Russian feminists and patriarchal Russian women	83
Mix of patriarchal and liberal in consideration of female vs. male way of doing	85
Discourse of responsibility has more traditionalism in the corpus of blogs	87
CONCLUSION	92
Traditionalism in the discussion of masculinity and femininity	96
What is normality within gender discussions?	97
Mixed values of traditionalism and liberalism	98
BIBLIOGRAPHY	102

## LIST OF TABLES

Table 1. Positive adjectives by frequency	63
Table 2. Negative adjectives by frequency	63
Table 3. General adjectives by frequency in user comments	68

## LIST OF FIGURES

Figure 1. Differences between the terms text and corpus	21
Figure 2. The three rises of the British novel	26
Figure 3. Corpus – based approach	41
Figure 4. Corpus – driven approach	42
Figure 5. Keyword “beautiful”	70
Figure 6. Keyword "needed"	72
Figure 7. Keyword "needed" in the comments	73
Figure 8. Keyword "strong"	75
Figure 9. Keyword "beautiful"	77
Figure 10. Keyword "good"	79
Figure 11. Keyword "normal"	81
Figure 12. Keyword "normal" and “it is normal”	82
Figure 13. Keyword "female"	84
Figure 14. Keyword "male"	86
Figure 15. Keyword "must"	88
Figure 16. Keyword "must" with "woman" and "man" collocations	89

## INTRODUCTION

Recent developments in internet technologies have created new ways of information exchange and communication globally. Blogs represent one of these new ways in today's technologically advanced world. So far a lot of research work has been done to analyze the phenomenon of blogs from the informational and communicational standpoint. However, there has been little discussion about blogs as representations of public opinion or sentiment regarding any particular cultural question. Although the number of papers analyzing public sentiment on commercial products in blogs is quite numerous, a critical blindspot still exists on cultural, as opposed to commercial, topics. This thesis will examine selected blogs of the Russian speaking internet through such methods of text analysis as keyword extraction, collocations, feature extraction (sentiment analysis) and text visualization. The central research question is the following: How is the evaluation of modern femininity and masculinity discussed in the Russian blogs? The hypothesis that will be tested is as follows: Traditionalism in the evaluation of modern femininity remains higher than in masculinity. During the text analysis of corpora the goal will also be to identify the key themes within the discussion of masculinity and femininity.

This paper engages with the data corpora assembled from the Russian blog websites LiveJournal Russia and Maxpark. The special value of this research is the combination of technological quantitative methods with qualitative elements of linguistics and classic text analysis. The reader should bear in mind that the study is based on the blogs taken within a certain limited time span and search criteria. The time period covers two years back from the cutoff date December 31, 2013. The data were found and gathered from the blogs containing such basic key words as “woman”, “man”, “girl”, “guy”, “husband”, “wife” and others associated with femininity and masculinity topics.



Computer - assisted text analysis is a relatively new area for Russian linguists and interdisciplinary scholars. There are some potential problems that the scholar can face doing research in a relatively new field. One of these challenges is the limited choice of technological programs for specific retrieval of data in the Russian language available as open source. To overcome this technological obstacle, a code was written in Python with the support of NLTK library and TreeTagger program for the Russian language. Thanks to this code it was possible to do text analysis in the data corpora. This research relies on the technology in such methods as keyword extraction, collocations, feature extraction (sentiment analysis) and text visualization. There is also a certain degree of manual work applied together with intuition.

The idea to explore the modern understanding of femininity and masculinity occurred to me as a result of my undergraduate research. At that time I was analyzing the image of the female Catherine Earnshaw in the *Wuthering Heights* by Emily Bronte. Then I decided to elaborate more on the question of gender, but in a comparative analysis. After a review of the relevant literature on Victorianism and gender discourse I found that femininity and masculinity were closely examined by generations of critics across a whole spectrum of classic literature, old movies and pictures. However, far too little attention has been paid to the analysis of contemporary women and men, especially in comparison to the concepts of traditional culture which is defined and discussed further in this paper. This was the case for both English and Russian textual sources. Most contemporary academic literature on gender discourse describes either men or women from very specific points of view, for instance women within the framework of Western feminism (Offen1988), or Catholicism (Weaver 1995). The same situation is true in terms of men, for instance it is either a general theory about masculinity (Brod and Kaufman 1994) or a particular issue like fatherhood for example within masculinity (Marsiglio 1995).

Although there is a certain amount of academic work investigating gender questions in the comparative paradigm of traditionalism and contemporary life, this amount is very small. Moreover most literature covers the question from a female perspective, e.g. Bryant (2003), Soto (1983), Malcolm (1992). As one can see the question of both masculinity and femininity within both traditional and contemporary vision was not thoroughly researched in English as well as Russian academic literature. In addition, my experience of working with blogs and social media led me to the idea that public blogs could be a better source of data about modern people in comparison with modern fiction or even magazines. Online blogs give a larger scope of outreach with representation of a large number of different people. The results of my research will contribute first of all to the field of text analysis, especially in the Russian language. Secondly, my research could be useful for social and cultural studies too, as it provides insight into many issues of gender relations in contemporary Russia.

The overall structure of the present thesis takes the form of five chapters, including this Introduction chapter. The second chapter is devoted to a description of literature and other materials that formed the theoretical and practical ground for my research. A historical overview of femininity and masculinity concepts in Russian culture is also provided. Additionally, the second chapter gives a review of blogging activity in Russia followed by a brief description of text analysis' formation as part of corpus linguistics. The third chapter deals with the explanation of the chosen methods, definitions and categorization. In the fourth chapter I describe the quantitative and qualitative findings of my text analysis. The key topics and findings are also discussed individually in the fourth chapter. Generalization together with the critique of findings is presented in the Conclusion.

## CHAPTER I

### LITERATURE REVIEW

This chapter describes the literature on key issues and methods in text analysis. Additionally the concept of the blog is discussed from different perspectives – in the form of a literary genre as well as a social and information environment. To better describe the research question a brief overview of literature on the concepts of femininity and masculinity in the Russian blogs is addressed as well.

The literature overview presented in this chapter shows that text analysis has an interdisciplinary character and has passed through significant years of development to be accepted as a mature reliable practice. It is also worth noting that the corpus and text analysis issues are discussed generally in this chapter as fundamental knowledge. The good point about text analysis is that despite its strong dependence on technology the framework of text analysis allows some intuition and freedom of actions within it. Keeping this point in mind, it is easier for researchers to choose their own path to examination of the problem. For instance, in the case of my research I have certain categories like “blog”, “femininity,” “masculinity” as well as specific type of corpora. Due to flexibility of text analysis, I can connect the general theory with specifics of my research question and corpora. The mixed approach to the analysis of my corpora became possible only thanks to this flexibility of text analysis.

The scope of my research paper is very large and encompasses technological, linguistic, cultural, and even social academic literature. It is impossible to be proficient in all of these spheres at the same time, but it is possible to analyze the research question after working with basic materials in the sphere of text analysis and gender studies. This chapter of my paper is devoted to an overview of the basic literature sources underlying my research. The analysis of

literature in this chapter is divided into three parts. First of all, I describe materials that shape the hypothesis of my research. In this first part of the literature review the work by the following scholars is presented – Olga Zdravomislova and Natalya Kigay (2002), D. Baranov and O. Baranova (2005), and Olga Gurova (2005), along with material from various online sources such as digital magazines and online Encyclopedia of Philosophy produced by the Russian Academy of Sciences. The second part of the chapter outlines the research work conducted in the analysis of blog. Additionally, the second part includes the theoretical foundation for the text analysis based on books and articles by the following scholars – Susan Herring (2004), Bonnie Nardi (2004), Kevin Crowston & Marie Williams (2000); Andrew Dillon and Barbara Gushrowski (2000), Elena Goroshko (2009) and Eugene Gorny (2006). The third section of this chapter focuses on literature about methodology informing my research from the point of text analysis. The authors here are Robert Johnson & Larry Christensen (2008), Elena Tognini-Bonelli (2001), John Sinclair (1991), John Firth (1935), Michael Stubbs (1996); Anatol Stefanowitsch and Stefan Gries (2003), Franco Moretti (2009), Stephen Ramsay (2011) and Andrey Zhalevich (2006).

To research femininity and masculinity in a broader sense of traditional culture, I explored the definition of traditionalism in English as well as Russian academic sources. According to New Philosophical Encyclopedia by Stepin (2001) “Traditionalism is orientation of the individual, group and public consciousness towards the past, which is usually opposed to values of the present” (online). Oxford dictionary gives the following definition: “The theory that all moral and religious truth comes from divine revelation passed on by tradition, human reason being incapable of attaining it.”

Also traditionalism is characterized as consciousness oriented towards the past and opposed to the values of present (online Encyclopedia <http://iph.ras.ru>). The traditions, customs

and skills developed from the past serve as foundational knowledge in the traditional society. In contrast to liberal culture that promotes gender equality in all spheres of life, there is a distinct division of gender roles within the traditional culture in any society. Traditionalism emphasizes gender differences. One of the most known traditional symbols that pictures gender differences is Chinese “yin and yang” – female and male beings moving in one circle.

### **Olga Zdravomislova’s research on woman discourse in media**

For the basis of my research I took the findings from the work by Russian sociologists Olga Zdravomislova and Natalya Kigay, who analyzed female discourse in mass media. The research was undertaken together with the Association of Female Journalists of Russia in 2002. The goal of this research was to analyze the concepts of modern Russian gender discourse with the focus on women’s topics. Professional journalists, both male and female ones, were interviewed in focus groups regarding female representation in modern journalism, as well as the roles of a woman and a man in the modern Russian society. The novelty of this research was in the effort to understand how the journalists represent women in their professional work and unconsciously in their personal conversation. All previous research done in this area used as sources only newspapers and journals, but not the people who wrote them. As a result scholars found that the most subjects in the female discourse characterized modern Russian women traditionally – orientation on family, health, sphere of emotions. Moreover, in the consciousness of most journalists in their professional writing as well as personal conversation, there was a definite division between female and male worlds. But the crucial finding of the research, which mostly aroused my interest, was the mixed presentation of female images. The perception of the modern Russian woman was a mix of classical Russian tradition with elements of female emancipation in Russia early XXI century.

The researchers noticed that traditionalism in the evaluation of a modern Russian woman is much higher than it is of a modern Russian man. The number of responses characterizing woman from traditional point of view equaled 6 responses in one focus group, 7 and 13 responses in two other focus groups. This number was bigger in comparison with the number of responses characterizing a man from traditional point of view, only 12 in one focus group (Zdravomislova and Kigay 36). Despite many objective reasons and numbers (for instance statistics of employment) that show a modern Russian woman from a feminist point of view there remain many stereotypes concerning the evaluation of a woman. I found this point interesting and decided to analyze it in a slightly different environment. I chose the nonprofessional blogosphere and focused on female as well as male representations in the gender discourse.

Another feature that makes Zdravomislova and Kigay's report important is its endeavor to analyze the area not widely explored in the social and cultural studies of modern Russia. Although research about the traditional Russian woman is quite extensive, there is not too much done in terms of the image of modern Russian woman and modern Russian gender discourse in general. Additionally, I came to the conclusion that there are very few researchers focusing on the male image in the Russian culture – both traditional and modern. This lack of research together with interesting findings from Zdravomislova and Kigay's work directed me to the importance of exploring modern femininity together with masculinity. Although Zdravomislova and Kigay's research analyzes the gender discourse looking at a woman in the context of her relationship and connection with the man in all spheres of life, the research is more focused on a woman and refers to a male figure only as a background in description or comparison of a female situation.

## **Concept of gender in Russian folk culture**

For a more thorough analysis of modern Russian femininity and masculinity I had to explore deeper its traditional and modern form in the academic literature. “Male and Female in traditional Russian Culture. Illustrated Encyclopedia” by Baranov D.A, Baranova O.G and Ziminina, T. (2005) (henceforth: Encyclopedia 2005) served as a good guide into traditional understandings of the matter. The Encyclopedia gives definitions for a variety of concepts outlining the material, cultural and social life of Russia in the patriarchal epoch. The lifestyle, responsibilities, and environment of an ordinary Russian person is described starting from early childhood and ending up in very old age.

The Encyclopedia describes many concepts of gender and folk culture that existed in the patriarchal times of Russian history (XVI – late XIX century). The authors use documentary photography from the Archive of Russian Ethnographic Museum to illustrate the gender culture of those times. There are some definitions from the vocabulary list of the Encyclopedia that were particularly important for my research. Those are the definitions of baba (a mature woman), a woman (женщина), a girl (девушка), a wife (жена), a guy (парень), a husband (муж), a man (мужчина), as well as beauty (красота) and family (семья). It is worth noting a difference in meaning between the words baba and woman. Although originally baba had a positive meaning and referred to a married woman (Encyclopedia 21), this word later acquired a more negative meaning. In the contemporary Russian language baba means uneducated and scandalous woman. After researching the Encyclopedia, I came to the conclusion that definitions of female gender outweigh to some degree male ones in the vocabulary list. Probably this difference can be attributed to the fact that female significance in traditional Russian family culture is very high. And as the Encyclopedia mostly describes the everyday family life of those times, it details many concepts in the female as well as male world of those times.

The Encyclopedia was really helpful in analyzing almost all topics of discussion in my corpora. For instance, in the analysis of a key word “beautiful” I compared the ideas and concepts written by bloggers with the ones described in the Encyclopedia. According to the Encyclopedia, female beauty was considered as both physical and mental. The good temper of a girl and her active lifestyle were the features that were valued on the top of physical ones (Encyclopedia 281). The authors of the Encyclopedia mention “Female beauty is an important factor in the choice of marriage partner but at the same time there were no general standards, everyone chose ‘beautiful’ to his/her taste” (Encyclopedia 281).

Some definitions given by the Encyclopedia refer to ages long past and have almost no connection with the present moment. This is especially true when it discusses the symbolic part of the traditional Russian culture that played a very important role in that epoch. For instance, the section about beauty has a detailed description of attributes of hair style, clothing and other elements that a beautiful woman was supposed to have in her wardrobe: “most frequently the hair braid of a woman was associated with the symbol of beauty and femininity in the folk understanding” (282). This sort of description sounds outdated in terms of modern Russian life. Nowadays one can see all kinds of women’s hair styles starting from bold and ending with long loose hair, although the braid in girls’ hair style is still admired. The presence of such dated details in the Encyclopedia helps my research in understanding of changes in female depictions and expectations.

In my research I stress the different understandings of beauty in women and men. The Encyclopedia helped me to understand the basic distinction between female and male beauty in the traditional Russian culture. The appreciation of female beauty was mostly based on the external features (although the inner qualities were also seriously considered). But in the description of a woman priority was given more to her external side. Male beauty, however, was



described with the focus on his mental and social qualities – decent, generous, honest, etc., “Female beauty is first of all embodiment of fertility potential in all of its aspects, the same like strength is representation and symbol of male beauty” (282, translated by me E.D.) As I noticed after working with the Encyclopedia as well as other materials, there are not many resources that would describe a man and masculinity. Even the Russian speaking materials about traditional culture have little on this topic, although the traditional Russian culture of the *Domostroy* (*The Household Book*) assigns a leading role to the man inside as well as outside the house life. There are not too many materials describing or analyzing a man or masculinity in the English language sources as well. There is, however, a great deal of analysis of woman and femininity in the English as well as Russian language literature on the question.

### **Changes in the gender culture from XVIII to XXth century**

The lifestyle of ordinary Russian women and men in the nineteenth century was not very different from the times described in the Encyclopedia. The change happened with the formation of the Soviet Union. Both women and men acquire social and cultural features different from the patriarchal epoch. In the Soviet period starting 1930s and almost up to the collapse of the Union consideration of femininity and masculinity was closely connected with ideology and propaganda. Olga Gurova (2005) in her work *People and Things in Soviet Culture* (*Люди и вещи в советской культуре*) offers an overview of people’s mind transformation from austerity in everything to the new cult of consumerism. Talking about a woman and her beauty in the beginning of the twentieth century, Gurova mentions “woman – worker” (женщина работник). A woman – worker is a very common and collective image of a woman in those times. The lifestyle of a woman – worker completely rejects any notion of “following the fashion” and really taking care of one’s appearance. She was supposed to be a super strong and hardworking woman whose life should mostly be spent at work and not somewhere else with family. The concept of

female beauty was primarily based on the idea of rationality. Gurova points out that to “strive for the beauty of body was considered mostly as non – Soviet, bourgeois feature as it was sort of window – dressing and put the visible part prior to the real being” (11, my translation).

Gurova does not tell a lot about/the image of a man in those times and any notion of masculine beauty. The only conclusion regarding the man one can see is that realities of those times did not favour development of male fashion as well. The shortage of goods and clothing items was unbelievably severe and produced a very strange effect on perception of a male appearance. “So everybody was given the coupons and from that moment almost all men were in the same shoes, half of the men were in the same hats.” (Gurova 43)

In the second half of the twentieth century, the economic situation of the Soviet Union improved. This brought new mottos into the life of society, for example “everything for a person, everything for the good of a person” (Gurova 17). But still this new prosperity did not change the perception of a woman and man. They both were supposed to work hard, be simple, healthy and happy.

### **Ambiguity of modern womanhood and manhood**

The literature describing and analyzing contemporary Russian women and men in social, cultural, and other aspects is not that extensive. Although it is possible to find some interesting academic and popular books as well as articles, the largest part of the modern femininity and masculinity overview for my research was based on the materials from online magazines - <http://www.rusarticles.org>, RIA News <http://ria.ru/>, [woman.ru](http://www.woman.ru), <http://www.cosmo.ru/> and other online sources. Most of the articles promote the image of a strong, successful, beautiful, career – oriented and independent woman. It is important to note that such image of a woman is widely applauded in many online magazines with positive feedback. But at the same time, a negative attitude towards the more liberal image of a woman is also present. In an article about female

beauty from RusArticles, for example, the author does not just describe this image, but also gives his evaluation of the modern femininity advanced in modern mass media (RusArticles 2012). In this author's view, the culture of extreme thinness and unachievable beauty is rooted in the economy's wish for selling more and more beauty and diet products. Although many resources depict the modern Russian woman as a winner in her social and personal life, as one who has a leading position at work and is very happy at home, statistics slightly contradict this representation. An article *Home violence in Russia* (RIA News 2013) reports that every forty minutes there is one woman dying from domestic violence. In addition to that, reported 36 000 women suffer from battery caused by their husbands. As one can see, the real picture of the modern Russian woman is not that happy. In addition, the statistics shows that the percentage of women who have a management position in modern Russia is very low (NewsRu article 2012). Summarizing all this information, one can conclude that the image of the modern Russian woman combines extremely contradictory descriptions – strong and independent, but at the same time obedient, abused and unhappy.

As to the discussion of a modern man and his masculinity features, the situation is even more complicated. Research conducted by Saint Petersburg State University shows the absence of any definite publicly accepted model of masculinity (Ushakova 2009). Moreover, the perception of the modern man is full of gender stereotypes and contradictory sets of moral values. Valentina Ushakova writes about modern Russian masculinity as a mix of myths coming from early twentieth century and supplemented with realities of the fast shift from socialism to capitalism in current Russia (2009). Masculinity is considered as a capability to “conquer, win, and keep.” And if a Russian man, for instance a politician or a top manager, is not rude, does not swear or threaten, he can be characterized “behaving like a woman” (Sillaste 2013).

## **Blogs as a special communication and information environment**

Before text analysis of the corpora, I explored the essence of the blog phenomenon in general and its development in relation to RuNet. Several definitions of the term “blog” were given throughout internet history. Initially a “blog” was looked at as a tool (early 90s), then as a service (late 90s) and gradually it turned into the process or practice. Wikipedia attributes the first appearance of the short form “blog” to Peter Merholz who mentioned it in the sidebar of his blog Peterme.com in 1999. Later on, the term was referred to as both noun and verb. Pyra Labs' Blogger product contributed a lot to the widespread use of the term. In that definition “blog” was a means or a tool for information sharing. In the late 90s blogs were evolving into the form of an online diary with great focus on the personal life of users. With the creation of services like LiveJournal, the blog gained the power of action. In addition to that the blogs became part of the online environment for journaling practice and publishing in personal and professional life. After acquiring features of the publishing environment with equal online access to everyone, the blog now represents the possibility to publish and comment on a particular topic of interest, and thus the blogs grew into community. Comparing different definitions by different scholars, I came to the conclusion that the blog is a variable online environment that can be named or characterized depending on the practices and practitioners participating in the process.

## **Blog as a Unique Genre and Social Space**

In my research I consider the blog as a social space and information environment. This understanding of the blog is based on a number of research findings that suggest the uniqueness of particular blogs as a genre and social space. The number of scholars analyzing the blog phenomenon is fairly large both in English and Russian academic literature. In my research I refer to the work done by such scholars as Kevin Crowston and Marie Williams (2000); Bonnie

A. Nardi, Diane J. Schiano, Michelle Gumbrecht (2004); Susan C. Herring (2004); Danah Boyd; Elena Bazhenova and Irina Ivanova (2012); Yana Bardashevich (2011); Eugene Gorny (2006); Bruce Etling and Karina Alexanyan (2010).

Susan Herring (2004) in her work *Bridging the Gap: A Genre Analysis of Weblogs* advocates the idea of blogs as an individualistic and intimate form of self-expression rather than a general occurrence “oriented towards external events” (1). In her research she explores if blogs are an emergent or a reproduced genre. After conducting a content analysis of 203 randomly selected weblogs, Herring found that blogs represent an emergent genre that encompasses offline print materials like magazines or journals e.g. as well as online media. In addition to this development Herring mentions that her analysis:

“Suggests that blogs are neither unique nor reproduced entirely from offline genres, but rather constitute a hybrid genre that draws from multiple sources, including other Internet genres. Specifically, it forms a de facto bridge between multimedia HTML documents and text-based computer-mediated communication.”(2)

Herring’s great expectations for this continuing and fast blog development and its high popularity reflect today’s situation – blogs have become built into our everyday life. Back in 2004 her research states that “intermediate” characteristics of blogs make them attractive to users” (Herring 11). As well, blogs give an opportunity of communication and public self – expression. But the most attractive part of the blog is that one can control and customize this interaction activity. Herring predicts growth in blogs’ popularity as well as its very diverse uses in the future. Blogs have presently become a very widespread form of communication, debate, advertising, and even learning. According to Herring, the most popular type of blogs is the personal online journal.

A number of scholars have searched for the explanation of such great popularity driving blogs and blogging today. Rebecca Blood (2002) suggest that one of the main reasons is grounded on the “promise of the web... that everyone could publish, that a thousand voices could flourish, communicate, connect.”

### **“Blogs Are A Studied Minuet Between Blogger And Audience”**

The blog was also viewed as a web-based form of communication driven by certain motivations and as a form of social activity (Nardi et al. 2004, henceforth: Nardi). A research team carried out audiotaped ethnographic interviews with bloggers and then conducted a text analysis of these blog posts. Some findings of this paper mirror Herring’s concept about the blog as a human, interpersonal phenomenon rather than only a technological oriented development. Nardi’s paper looks at blogging as a social activity, and she analyzes the motivations and relationships behind it. What drives people to blog?

In the view of Nardi the blog represents a “frequently updated webpage with a series of archived posts.” To some extent Nardi’s research has many common ideas with Herring. Nardi also finds blogs especially valuable from the perspective of human factors. In her opinion, the blog gives users an opportunity of self-representation and communication. In terms of the blog as a genre, Nardi is not assured that blogs have passed enough developmental stages to be referred to as a genre. But at the same time Nardi’s opinion about the mixed or hybrid nature of blogs is somewhat close to Herring. “It is also too early to say whether blogging is a “genre”. While the uniform format of blogs suggests genre conventions, the extraordinarily diverse content of blog posts would seem to burst the bounds of a single genre” (Nardi 10).

Nardi’s paper demonstrates that blogging can really be considered as a social activity due to a number of aspects that form the experience of bloggers and influence their behavior in this process. Researchers came up with the list of objectives that motivated human activity:

1. Update others on activities and whereabouts
2. Express opinions to influence others
3. Seek others' opinions and feedback
4. "Think by writing"
5. Release emotional tension

Additionally, Nardi notes that blogging as a social activity gives birth to other forms of communication – instant messaging, chatting, e- mailing, for example. Also, the scholar found a strong connection and dependence between the bloggers and their audience:

"We learned that blogs create the audience, but the audience also creates the blog. This linkage happened in a number of ways: friends urging friends to blog, readers letting bloggers know they were waiting for posts, bloggers crafting posts with their audience in mind, and bloggers continuing discussions with readers in other media outside the blog." (Nardi et al.3)

Nardi stresses the "consciousness of audience", its attention and feedback as "central to the blogging experience" (4). But at the same time, the key finding of the research suggests "limited interactivity of the blogging experience" (10). The conversation between bloggers and their audience happens, but it is dominated by the blogger. The situation is similar to the radio station where the audience listens to the deejay without actively participating in the discussion. The key problem with this conclusion is that scholars do not consider this relatively passive reaction from the side of audience as a drawback. Moreover, Nardi does not seek the ways to engage the audience as the main focus of her paper is on the bloggers rather than the audience.

### **Blogging activity in Russia**

There are a number of other researchers who have attempted to explore the nature of blogs. In the view of Crowston and Williams (2000), the personal home pages have an "emergent" (rather than a "reproduced") web genre. Dillon and Gushrowski (2000) find the web-

based genre unique in its own way. Goroshko describes the Internet as a communicative and informative environment that gives rise to new genres inherent only in this given environment (117). Almost all the scholars come to the conclusion that Web blogs have their own unique features and represent a new unique genre.

As one can see there is a certain degree of consensus regarding the blogging phenomenon in the English academic literature. In the Russian speaking world, the research on blogs is also quite extensive and shares many concepts with academic representatives of other countries. In many cases the idea of the blog is explored as a form of communication and individual self-expression (Bardashevich 46). Similar to many other papers mentioned above, the Russian scholars also draw attention to the ambiguity and heterogeneity of blogs as a genre and as social experience as well (Bazhenova and Ivanova 130). Data from the research by the Berkman Center at Harvard University (Etling et al. 2010) demonstrate that blogging activity can also serve as a trigger for mobilization of people for some actions (19).

After familiarizing myself with major literature available on the topic of the blog I came to conclusion that Herring's perspective on blogs' nature is one of the broadest. The vision of blogs suggested by Herring fits my research very well. At a closer look, one can notice that blogs on many websites (including my corpora) have a "hybrid" nature that "draws from multiple sources, including offline and other internet genres" (Herring et al. 2). In my examination of the blogs I also consider them as part of human social activity suggested by Nardi. Collectively, all the research papers allow me to make an assumption that blogging is a human online experience that can be a continuation of reality, a type of human conversation that takes place in real life but at a different pace, with a bigger choice of people, and expressed in different form.



## **Literature on text analysis**

In terms of text analysis there is a large and growing body of literature that investigates the methodology and application of blogs. In my research I refer to many scholars and in particular Tognini-Bonelli (2001), Johnson & Christensen (2008), Leech (2005), Danielsson (2003), McEnery and Wilson (1996), Louw (1993) and Sinclair (1991 and 2004), Firth (1935), Stubbs (1996), Stefanowitsch and Gries (2003), Moretti (2009) and Ramsay (2011). The ideas and concepts about text analysis brought up by all these scholars seriously contributed to my research.

## **From text to corpus and from instance to system**

The book *Text and Corpus Analysis* by Michael Stubbs was especially important. In his book Stubbs describes the history of text analysis with a review of its major concepts. Supported by three case studies, the book serves a good guide into computer – assisted studies of language and culture theoretically as well as practically. The theoretical part of Stubbs' book can be challenging for understanding in some places. Stubbs provides a very broad historical perspective on text and corpus analysis comparing many scholars of different times. For instance in his discussion of key principles of corpus linguistics Stubbs describes development of corpus linguistics within different stages - from Firth to Sinclair through Halliday:

Work by Halliday and Sinclair has covered the thirty years from the 1960s to the 1990s. This is precisely period of the rise and heyday of Chomskyan linguistics. Firth (1957) published one of his most famous papers *A synopsis of linguistic theory, 1930-1955*, in the same year that Chomsky (1957) published *Syntactic Structures*". (1996)

Stubbs raises a range of questions that have been in debate among corpus linguists:

- the relations of text to corpus and of instance to system;
- the concepts of sample, representativeness and typicality;

- the probabilistic nature of many textual patterns;
- the meaning of patterns of frequency and distribution;
- the function of syntactic units in texts;
- the mediation of knowledge in language;
- the dissemination of texts from authors to audience (20).

Generally the work by Stubbs represents a very rich summary and analysis of key theoretical concepts in text analysis. Stubbs points out the British linguistic school that set out the tradition for corpus studies represented by Firth (1935), Halliday and Sinclair (Stubbs 1996).

“They have made major descriptive and theoretical contributions to linguistics, and have insisted on the integrity of the linguistic methods” (Stubbs 25). Unlike more recent American and British work, the earlier British school did not consider linguistics as purely autonomous from others fields of science. Stubbs mentions that Firth looked at his work as “essentially sociological” (25), while Halliday connected his research with education linguistics. The main principle advanced by Firth is the contextual theory of meaning: ‘the complete meaning of a word is always contextual, and no study of meaning apart from a complete context can be taken seriously’ (Firth 37).

The major contribution in the corpus linguistics made by Firth was his “contextual theory of meaning”. Firth stressed the concept of form and meaning as inseparable parts in the study of meaning. This view was questioned by Sinclair who considered the meaning prior to the form: “Meaning affects the structure and this is...the principal observation of corpus linguistics in the last decade” (Sinclair 6-7).

In his approach to the text Sinclair is not too much concerned with the cultural theme, while Firth as well as Chomsky argue that “language in use transmits the culture” (Stubbs 43). Despite some differences in approaching the texts, the scholars of those times share the idea of language

as a real and social experience. In addition to that, most academic work of those times is grounded on belief in reliability of language study only when the language is researched in use. To research language in use means to explore meaning of a certain word in connection to the whole sentence it was written or pronounced in, the whole paragraph, book and context in general “Language should be studied in actual, attested, authentic instances of use, not as intuitive, invented, isolated sentences” (Stubbs 28). In the opinion of Halliday (1990) the language can program the experience: “clause by clause, texts code experience, construe it in certain way and synthesize the world view” (Stubbs 60).

In the continuation of his historical overview of text analysis, Stubbs moves to the era of computer appearance in text linguistics which is late 80s and early 90s. Computer technology helped linguists to analyze the data previously unobservable. Stubbs notes that technology brought up a new important concept in the linguistic theory which is the corpus. Using the example of data – driven linguistics, Stubbs demonstrates how computationally supported analysis of language opens new ways of “considering the relation between data and theory” (p. 46). New methodologies to read and interpret the text were given to the researchers, for example, key word in context (KWIC), concordances, collocations, frequency words and so on. On the example of data from Williams’ *the Country and the City* as well as Said’s *Orientalism* Stubbs demonstrates the effectiveness of keyword and collocation methodology that helps to study the “cultural connotations of words.” So called “cultural keywords” in the form of fixed and semi – fixed expressions (collocations, catch phrases, clichés and idioms) encode cultural information (Stubbs 169) and form the knowledge of something (Oriental atmosphere, Oriental despotism, Oriental sensuality...). Collocation is a quantitative technique in corpus analysis that shows the “associations and connotations they (words) have and therefore the assumptions which they

embody” (Stubbs 172). In this sense technology played a huge role in grasping the incredibly large amount of data and providing more reliability to the findings.

**Technology as a trigger from the concept of *parole* to *langue* in text studies**

The work by Tognini-Bonelli (2001) is really helpful in terms of practical use of text analysis. If Stubbs’ book provides a very solid theoretical ground for the research in text analysis, the work *Corpus Linguistics at Work* by Tognini-Bonelli (2001) focuses more on real practical illustrations of work with corpus data. The significance of the book is in its fresh view on the data with the input of new technologies. The use of technology provides greater “awareness of language variety” (Tognini-Bonelli 6). In her book Tognini-Bonelli outlines key issues related to the corpus work. First of all, the author offers a definition of corpus which is collection of authentic data for example texts “gathered according to explicit design criteria, with a specific purpose in mind” (Tognini-Bonelli 2). Characterizing the corpus, Tognini-Bonelli outlines key differences between the terms *text* and *corpus*.

A TEXT	A CORPUS
read whole	read fragmented
read horizontally	read vertically
read for content	read for formal patterning
read as a unique event	read for repeated events
read as an individual act	read as a sample of social practice
of will instance of parole	gives insights into langue
coherent communicative event	not a coherent comm. event

Figure 1. Differences between the terms *text* and *corpus* (Tognini-Bonelli 3)

In the framework of a corpus one can read the data vertically and consider it from the *langue* point of view: “Using a Saussurian terminology one can conclude by saying that text is an instance of *parole* while patterns shown up by corpus evidence yield insights into *langue*” (Tognini-Bonelli 3). Thanks to technology, researchers now are capable of viewing a statistically large amount of data to make representative generalizations out of it. There are some important points that text analysts should consider while building the corpus “the authenticity of the texts

included in the corpus, the representativeness of language included in the corpus and the sampling criteria used in the selection of texts” (Tognini-Bonelli 54). In the further discussion of the corpus issues, Tognini-Bonelli describes a tension between theory and practice in the example of corpus – based and corpus – driven approaches. In her opinion, the corpus – based scenario places limits on linguists as it is restricted by theory at the very beginning, while a corpus – driven approach allows a researcher more space for the analysis as it does not start with the hypothesis like in the corpus – based scenario. “In the corpus – driven approach, the path of research goes from observation to hypothesis then to generalization and only after to unification in theoretical statement”. (Tognini-Bonelli 85)

To convince readers of the richness of the corpus – driven approach, Tognini-Bonelli presents an illustration from the analysis of the English words “faced” and “facing” in the *Birmingham English* corpus and the *Economist* and *Wall Street Journal* corpus. As a verb “face” has two meanings: 1) related to a particular position and direction or 2) confronting problematic areas. After observing the collocations of face, Tognini-Bonelli came to the conclusion that the physical meaning of facing is more prevalent in the Birmingham corpus. The most frequent collocations of “face” were such words as “forward”, “palms”, “stood”, “sat”, “windows” and so on (Tognini-Bonelli 95). As to the use of “facing” in the *Wall Street Journal* corpus, it mostly had an abstract meaning (fear, serious, problems). This observation led to the conclusion that the *Birmingham English* corpus can be characterized as a more general one, while the *Economist* and *Wall Street Journal* are more specialized ones.

As an approach with open questions without a particular categorization in the beginning, the corpus – driven model provides more opportunities and ways to analyze the text. Tognini-Bonelli points out the importance of new technologies in such large-scale text analysis. This example demonstrates how technologies hastened the shift from the concept of *parole* to the

concept of *langue*. If we look at the word “facing” in any single concordance line, we can see an individual instance of language in use. This can be referred to as a parole. But when “many such instances are gathered together and sorted... on the vertical axis” (Tognini-Bonelli 98) the research effort builds up into generalization or *langue*. Before the advent of computer corpora, the patterns of *langue* were not as observable as they are nowadays.

Tognini-Bonelli describes this drawback as “an indication of the state of technology and not the state of the language” (98). Summarizing the discussion of corpus – driven approach the author also refers to Sinclair, Firth, as well as Halliday who considered connection between form and meaning as a methodological formality, “In this respect Sinclair, like Firth and Halliday, is adamant in claiming that the distinction between form and meaning is only a methodological convenience and this leads him to posit formal observations as criteria for analyzing meaning”. (Tognini-Bonelli 99). The key point suggested by Tognini-Bonelli is that connection represents a much stronger fusion, “form and meaning are seen as two aspects of the same phenomenon – language in action – and ultimately to analyze the former according to the objective criteria will yield insights into the latter” (99).

When assembling corpora for my research, I kept in mind all the points mentioned above. Although the text in my corpora is very unstructured and represents a mix of different blogs combined together, it is still authentic. The blogs were written by real people within certain websites and certain groups on the website. To make sure my corpora provide good evidence and representativeness of the results I gathered as many blogs as it was possible to do manually. In total 235 blogs were gathered along with 220 comments. In terms of my approach, none of the two approaches separately seemed effective in my research due to the limitations each of them has. That is why I combine both approaches in a one mixed approach - the hypothesis and categorization were set at the early stage of my research, but analysis with generalization was

undertaken in the second phase of the research. This mixed approach allows more time for analyzing corpora before advancing a hypothesis and testing it.

There is a large volume of published studies in the area of text analysis, especially in the English language literature. The Russian language literature has fewer materials on this subject. Even though text analysis is used by the Russian corpus linguists, their methods are still very classical and modeled more on literary analysis rather than text analysis. Russian corpus linguists accept technological innovations very reluctantly, probably due to strong devotion to tradition that has been generally developed in Russian society through centuries.

### **Aesthetics of data**

To make the results of my research clearer in its visuals and content, I applied the technique of text visualization. Alan Liu in his paper *The State of the Digital Humanities: A Report and a Critique* (2012) examines data aesthetics. In his opinion data aesthetics is one of the crucial aspects about information, text, and design that seems to remain underdeveloped in today's understanding and view of the page:

“One has only to view any typical data visualization from the text-oriented side of the digital humanities (concordance-like lists, line or bar graphs, cliché'd tag clouds, etc.) to recognize the near-total imaginative poverty of the field in crafting an aesthetics of data.

In great part, this is due to the conceptual separation in the field between the formatting of ‘data’ and the encoding of ‘metadata’” (27).

A possible solution to this problem could be a greater concentration not only on metadata creation but on the “look-and-feel” of data as well. Liu states that those two aspects together will provide for better clarity and beauty of the information.

A number of studies have been done to provide theoretical and practical ways to connect technology and text as well as technology and criticism. In particular these are questions the

area of digital humanities is concerned with. As my research also deals with these questions of text analysis and data visualization it is worth noting two scholars who brought serious influence on digital humanities. This is Franco Moretti with his work *Graphs, Maps, Trees: Abstract models for a literary history* and Stephen Ramsay with *Reading Machines: Toward an algorithmic criticism*.

Franco Moretti in his book *Graphs, Maps, Trees: Abstract models for a literary history* strongly advocates of distant reading as a practice that enables readers to see the whole picture out of small pieces of literary works:

“A canon of two hundred novels ...a field this large cannot be understood by stitching together separate bits of knowledge about individual cases, because it is not a sum of individual cases: it is a collective system, that should be grasped as such, as a whole-and the graphs that follow are one way to begin doing this.” (Moretti 4)

On the example of graphs Moretti demonstrates development of such genre as the British novel from 1700 until 1840. The graph below shows growth in the yearly output of new novel titles going through the period of steady growth and reaching its peak in 1840.



FIGURE 2: *The three rises of the British novel*

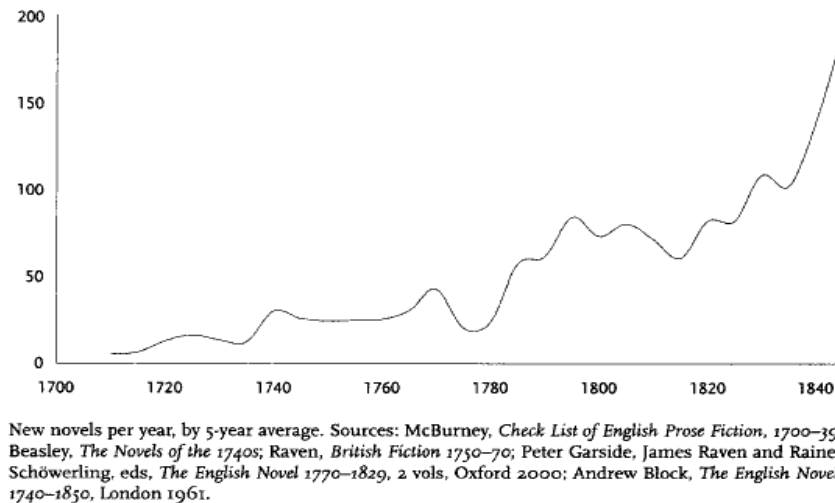


Figure 2.

With the help of charts, maps, and time lines Moretti looks at different literary genres, for instance novel or gothic, as a large total system. The use of graphs makes it possible to track even gender ratio of novel authors. For instance reading separate literary pieces of that period could give an idea about gender of some of the writers. But in order to see a more systematic outline of authors Moretti turns to the form of graph:

“Clearly, a mid-Victorian 'invasion' presupposes a reversal of the gender shift of the 1740s. And, in fact, this is what the historical record shows: if between 1750 and 1780, as a result of the initial shift, men publish indeed twice as many novels as women, in the late 1780s a second shift reverses the gender ratio, as one can see in Garside's breakdown for a slightly later period”. (Moretti 27)

Moretti continues his discussion of distant reading on the example of mapping literary genres. The visualization of Mary Mitford's *Our Village* turns the text into “a little solar system, with the village at the centre of the pattern, and two roughly concentric rings around it” (Moretti 37). Each part of this “solar system” has meaning and open for analysis and interpretation. One

ring tells about personal relationships of heroes while another ring connects the collective events of the village. This spatial representation of the text enables the analysis of narrative space. In the visualization of *Our Village* the “narrative space is not linear, it is circular” (Moretti 37). Moretti explains the circular form of visualization in the case of *Our Village* by a “perceptual system in which the village is still largely self-sufficient, and can therefore feel at the centre of 'its own' space” (38). One can easily find connection between form and meaning in all visualizations by Moretti.

### **Algorithmic criticism with humanistic inquiry**

Stephen Ramsay to some extent can be positioned as antagonist to Moretti. The work of the latter is centered more on technology while Ramsay starts with text and goes to technology. Although his method is called “algorithmic criticism” Ramsay’s approach is dictated by humanistic inquiry first of all, “The digital revolution has not penetrated the core activity of literary studies,..., it remains mostly concerned with the interpretative analysis of written cultural artifacts” (Ramsay 2). Ramsay does not find surprising the fact that literary studies did not fraternize so easily with technology right from the very beginning. With reference to Roberto Busa, an Italian Jesuit priest who was the pioneer of computer usage for literary analysis of Thomas Aquinas’ works, Ramsay mentions that “Busa’s own revolution was firmly rooted in the philological traditions to which modern criticism was largely a reaction” (Ramsay 2).

Ramsay gives practical example of “criticism derived from algorithmic manipulation of text” (2). Calculation of term frequency in Virginia Woolf’s *The Waves* shows that criticism can easily speak the language of mathematics and statistically add value to the analysis of texts. Running a certain algorithm one can find the frequency of terms that speakers use. This frequency in its turn helps to detect importance of a particular question for a certain hero of the book. For instance Ramsay finds the term “accent” especially distinctive to Louis, one of the

heroes in the book who is Australian while others are English. Other heroes do not mention the word “accent” so frequently neither regarding their own accent nor Louis’. It becomes obvious that Louis is concerned with the way he talks while others are not (Ramsay 12). For better explanation of “algorithmic criticism” Ramsay points out that its mechanisms do not seek to “solve Woolf” or any other literary piece. What he says is “we are trying to ensure that discussion of *The Waves* continues” (15). Ramsay also highlights the need for computational tools in literary studies that has not been addressed for a long time but can be solved now:

“Optical character – recognition software reveals the fragility of the grapheme.

Tokenization forces us to confront the fact that the notion of the word is neither unambiguous nor satisfactorily definable for all circumstances. Rather than hindering the process of critical engagement, this relentless exactitude produces a critical self – consciousness that is difficult to achieve otherwise”. (34)

## CHAPTER II

### METHODOLOGY

This research is focused on the analysis of femininity and masculinity as represented and discussed on Russian social websites. The blog articles from Maxpark and LiveJournal Russia were assembled in two data corpora for the subsequent text analysis. The goal of study is to test the following hypothesis advanced by the Russian sociologists Olga Zdravomislova and Natalya Kigay in their research of the representation of women in media: “There is more traditionalism in the evaluation of femininity in modern Russian women than in the evaluation of masculinity in modern Russian men” (36).

Such text analysis methods as word frequency, collocations and feature extraction (sentiment analysis) are used in my research. For better generalization of key words and concepts in the text I use the text visualization method as well. The visualization is done with the help of free program Texttexture <http://texttexture.com/>. My research is based on a mixed approach (Johnson & Christensen 2008) combining both quantitative and qualitative methods. In other words, I chose a bottom up approach allowing multiple forms of data to be collected and looked at from multiple perspectives. I use the data to prove or test a particular hypothesis. Moreover, my analysis of the data can lead me to the conclusion that disproves the hypothesis. The result of this analysis should represent statistical findings with in-depth qualitative analysis and finding of overall themes in the discussion of femininity and masculinity.

The data corpora for this analysis were collected from the blog articles and comments on websites Maxpark and LiveJournal Russia. There were certain criteria in building the corpus. No particular age category was put into the corpus criteria as it is really hard to determine the age of users online. Most users in Russian social media do not mention their age or make it invisible to those who are not friends with the user. As it is impossible for a researcher to

become friends with all the users, the option is to find texts with topic orientation as the main criteria. The data set was assembled manually by making a search of text posts from Maxpark and LiveJournal by using the specific set of search parameters. A time span of two years was determined for the blog coverage. The key component of search was topic or interest orientation – modern woman, modern man, femininity, masculinity, female/male relationship.

By saying a “blog” I mean a blog article without relevant comments to it in the discussion below the original post. The comments were collected in a separate text file. All the blogs and comments were marked up with an <entry> tag and gathered in a plain text file. In the corpus of comments each comment entry includes all the comments under the blog posts. For example a certain blog had 115 users commenting on it, this means the entry gathers all 115 comments in one entry. Additionally, the text file of the data corpus contains links to the web page the blogs were taken from. There is also a separate Word document with all the sentences containing the keyword and its collocations. In total there are 235 blog entries in the corpus of blogs and 220 comment entries in the corpus of comments. The reason why the number of comment entries is less than the number of blogs is because some of the blogs did not have any comments. The time span for all the entries is starting December 2011 until December 2013.

In the initial phase of the research, all the blog articles for only the last month or two were taken into the corpus. The compilation of blog articles was done manually by searching for articles using such key words and tags as ‘family’, ‘relationship’, ‘gender’, ‘woman’, ‘femininity’, ‘mother’, ‘wife’, ‘husband’, ‘male’, ‘masculinity’ and other words relevant to the main topic. The fact that all blog articles for the indicated time span were included in the corpus should provide good representativeness of the research question as it will touch upon many different points of view. The comments to all the blog posts were also included. This approach helped to look at the research question from different perspectives.

Such text analysis methods as word frequency and collocations identify key topics of discussion in the data corpus, serving as the quantitative part of the research (tokenizing, counting, etc.). The feature extraction method describes the content and summarizes it. Feature extraction stands for the qualitative part. This way both qualitative and quantitative methods are used in the research.

The portion of manual work in my study is substantial, including data collection as well as partial feature extraction. It is important to note that due to the Russian language of the data, I also had to work manually to understand the gender of words. The Russian language has female, male and neuter genders. The programming code that I describe later in this chapter counts the adjectives and particular words of all the genders together. Although it was difficult to determine the age of users, it was quite easy to determine the gender of users just by reading their nicknames. Cases of misleading information in the names of users happen very often, but still most users create nicknames that correspond to their gender.

**Hypothesis: “There is more traditionalism in the evaluation of femininity in modern Russian women than in the evaluation of masculinity in modern Russian men.”**

The Russian sociologists Zdravomislova and Kigay investigated discourse on women in the Russian media written by professional journalists. The project consisted of researching a number of written Russian media as well as interviewing a number of Russian journalists in focus groups. As a result, Zdravomislova and Kigay came to a very interesting conclusion - traditionalism in the evaluation of femininity is higher than in the evaluation of masculinity. She found that the image of a modern Russian woman combines two controversial ideas. First, the vision of a modern Russian woman is based on traditional principles of a fundamental difference between a woman and man. Secondly, the representation of a modern woman is strongly influenced by the idea of feminism and emancipation. The researcher cites a fact about high

employment activity among women in all social classes which was not the case even fifty years ago. This means Russian women work to provide for themselves and become more independent. Most of the interviews conducted by Zdravomislova and Kigay show much subjectivity and very strict requirements for women on all levels of life. For instance the survey conducted as part of Zdravomislova and Kigay's research showed the rating of character qualities necessary for women and men. According to this rating the woman first of all needs to be self-confident, with good organizational and leadership skills as well as independent and ambitious. But at the same time the list of qualities for the woman included classical Russian notions such as conservatism, emotionality and openness (8). Generally speaking the survey results portrayed the ideal woman as an "independent and smart woman who knows how to achieve her goals and manage difficult situations" (Zdravomislova and Kigay 9). As regards to the man the following qualities were prioritized: self-confidence, firmness, presence of mind, ability to convince and persistence. This leads to the conclusion that the man is portrayed with bigger focus on activity and some degree of aggression. The researchers draw attention to the presence of classical Russian notions about masculinity and femininity together with more modern and liberal representation in the descriptions of both modern Russian women and men (9).

In many cases interviewees of the survey openly told about the "deficiency of women", their inferiority in many areas. Many respondents referred to a woman as a human being "that cannot exist without man, the man and the woman are both human beings but the man is more full-fledged as he can be a politician, an economist, male authored press is more interesting" (Zdravomislova and Kigay 38). Other association words for "woman" that respondents mentioned were "synonym for love" and "home". At the same time many respondents expected women to have the same responsibilities as men – to earn good money, to succeed in a career

(Zdravomislova and Kigay 52). Summarizing the results of interviews Zdravomislova and Kigay mention “ideology with two layers” (48) whose external side claims for “gender equality, political correctness and social tolerance while the inner side represents pseudo-patriarchal ‘collusion’ with a mix of hard pragmatism and romanticism” (48).

### **Definition of traditionalism in traditional Russian gender culture**

Talking about the idea of traditionalism it is important to mention that notions of traditionalism and conservatism are very close, although conservatism does not deny the evolutionary development of society. Depending on nationality, there can be different features of gender traditionalism. In the Russian culture traditionally a woman and man had the following descriptions (Encyclopedia 2005):

**Woman:** In Russian traditional society, marriage was considered as the main purpose of life of any girl born to this world. It was inevitability. For centuries Russian people developed the idealized image of a woman as a “kind” wife. In the system of traditional life, to be a mother was the crucial role of a woman, and motherhood is the main purpose of a woman (Encyclopedia 339).

**Man:** A man had the leading position in the village society. He was the head of family, the master and main proprietor. His responsibilities included solving economic problems, paying taxes and responsibility for the wellbeing of the family. Supremacy of the man is a characteristic feature not only in Russia but in many other cultures as well. It is explained by many ideological and economic circumstances. In the case of Russian culture an important role was played by the right to own land. A man had a right to own the land. (Encyclopedia 372)



## **Maxpark and LiveJournal – two diverse blogospheres in one corpus**

Unlike many other social websites where user communication happens mostly by the exchange of short messages and statuses, Maxpark represents a platform for discussion on a variety of topics. The idea of Maxpark is intelligent virtual communication that can benefit and unite website users in real life. Every user of the website – Maxparker – gets a chance to create his own independent information environment making news, creating blogs, surveys, exchanging data and socializing with other Maxparkers. Other features that make Maxpark special and interesting for research are the possibility of discussion, as well as socializing and meeting popular people. The texts presented on Maxpark vary in their structure, topic orientation, types of author who write these blog articles, and other characteristics. My research focuses on electronic texts (blog articles) written by ordinary people. How can one distinguish the blogs by ordinary people from those by professional journalists?

In the case of a professional journalist, Maxpark provides a link to the original source - a website where the journalist is publishing (it is a rule of the website). If there is no link at the end of blog article and the user's description of profession has no sign of professional journalism, then, this blog article is written by an ordinary user and can be included into the corpora. The reason for focusing on ordinary people is because the goal of this study is to analyze the online discussion of the ordinary public, how ordinary people talk about the research question: the representations of and perceived notions regarding the modern Russian woman and traditionalism.

The domain for the text selection definitely represents widespread popular, rather than academic culture, as it seeks out the voices of the grassroots. One more reason that makes Maxpark a good social website for analysis of public discussion is that Maxpark collects blogs

from other professional and non-professional blog sources. When the blog article is taken from another online resource, there is a reference to that resource.

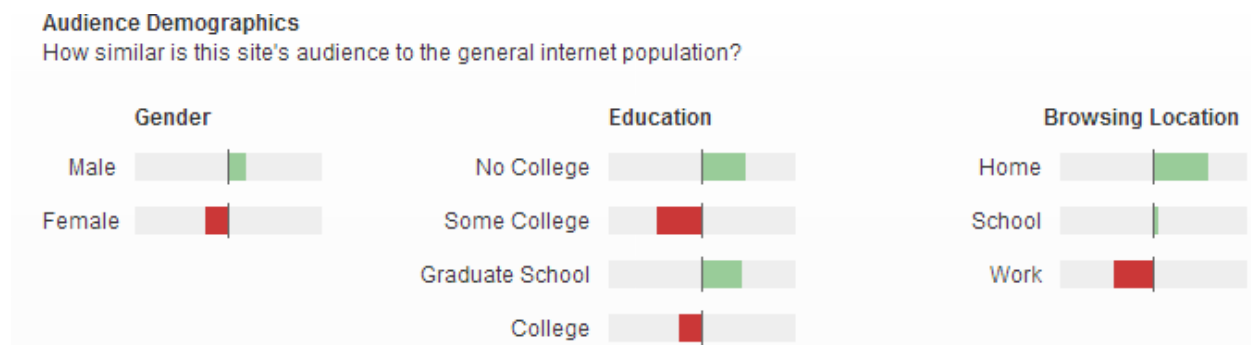
Other features of Maxpark include a good collection of data, heterogeneity of text styles, and a diversity of opinions. I came to this conclusion after carrying out a pilot study for the purpose of a general overview of the website, by testing whether the website contains enough text materials about the research question and what kind of texts are available on this website in general. At first sight the website itself and the blogs look unstructured and scattered. It is hard to see the logic or idea behind each blog post. But despite such ambiguity, a second closer look reveals that biggest part of the front page is devoted to the blog articles with the highest rating of discussion. Moreover, those blog articles in the rating are gathered from different interest groups automatically, regardless of topic orientation and position of the group in the overall website rating.

If a particular blog article has been actively discussed for quite a long time, it automatically shows up on the front page. But in case a user wants to see only fresh blog articles even though they are not “hot” ones, then there is a section entitled “New.” Browsing the front page, one can see multitude of topics being discussed. Titles of the blog articles on the front page signify a diversity of topics and opinions that can be found on the website: *Should kids help to bring up the younger brothers and sisters?; Russia does not need immigrants at all; Workers dying on “Kolskaya” (jack – up rig) were forbidden to set the signal SOS; About the Russian village; How to integrate immigrants?*, etc. Exploration of the website audience was also part of this pilot study.

### **Statistics of websites**

I used the Web information program <http://www.alexacom> to find the general statistics about users. Alexa is the website that supplies accurate metrics of site traffic, audience and

benchmarking. Although very detailed information about any website through Alexa is available for paid subscribers only, there are some free options as well. During the time period of this research (end of 2013 – early 2014), Alexa Traffic Rank places Maxpark in the 600<sup>th</sup> place on the Russian internet. This is an estimate of Maxpark’s popularity in Russia. The way Alexa calculates this ranking is through a combination of average daily visitors and page views on the website from users of that specific country over the past month. The ranking works in descending order with #1 as the highest combination of visitors and page views. Alexa showed the basic audience demographics for Maxpark ; (<http://www.alexa.com/siteinfo/maxpark.com>).



Graph 1. Demographics of Maxpark

All the indicators are shown as relative to the general internet population. The graphs show that males are overrepresented while females are underrepresented. People without college education are over – represented while those with some college education are under – represented, but at the same time the audience with graduate school education is over-represented in relation to those who went to college. One more factor that can help to allow a deeper insight into the audience of Maxpark is the browsing location. Based on the fact that the population reading Maxpark from home is overrepresented and those reading from school have almost zero presence, one can conclude that Maxpark readers are quite mature. For comparison, Facebook

statistics on Alexa show school as a major browsing location, which corresponds to the fact that the biggest Facebook audience consists of young people. (Emarketer 2012)

It is important to keep in mind the level of data confidence that Alexa provides for any website. Maxpark was given medium confidence in all demographic categories. To ensure certainty of the statistics of the website, there was a need for additional checking. After a relatively long period of surfing on the website it became clear that around 60% of users do not mention their age, whereas the other 40% can be split into two categories – people of 35 – 50 years old or those stating 100 or 120 as their age (jokers). Summarizing these statistics, it is possible to say that the majority of Maxpark readers are people of middle age with higher education. This general description of the age category is not enough as my research is focused on a specific discourse on Maxpark, thus it is reasonable to explore the audience that participates in this discourse.

The pilot study of Maxpark was a preliminary step before building the specialized data corpus proved that the website contains interesting data from the linguistic and analytical points of view. Linguistically, the data is interesting as it combines different language styles, for instance, formal and informal, literate or slang. If we look at the blog articles and the comments we can find different uses of grammatical patterns, stylistic features, and syntactical frequency. Analytically, the data can be interesting as the website covers a mature audience with a certain level of education. The fact that the website audience does not visit Maxpark from work but finds time to discuss and write blogs from home means that these people are hardworking, broad – minded, and not indifferent. Moreover, the density of texts already produces the impression of blog articles that are grounded on some ideas. One might say that the length of the blog entries does not necessarily stand for a higher quality than any other usual blog of a shorter size. Comparing Maxpark to other Russian social websites (VK.com, <http://www.odnoklassniki.ru/>)

during the pilot study, however, I observed that Maxpark texts have not only greater density, but also more meaningful blog posts and line of comments.

Moreover, Maxpark offers a larger coverage of audience and accordingly opinions. The main mechanisms of conversations and discussions on Maxpark are text oriented. There are almost no pictures used in blog posts comparing to other social websites like Live Journal, VK.com, etc. The average number of comments to each blog post on Maxpark is at least 200 comments, while other social websites may have the same intense discussions but they are not relevant to my specific research question. The pilot study showed that Maxpark contains a large percentage of blog posts on the topic of relationships, gender questions, and many other subjects within the research question. After comparing Maxpark to other social websites, I came to the conclusion that Maxpark offers a solid ground for creating large corpora for analysis of femininity and masculinity.

For greater transparency, representativeness, and accuracy in the research, the data corpus must be large enough. The basic assumption in finding the texts on Maxpark was a search engine within the website. The keywords ‘woman,’ ‘mother,’ ‘gender,’ ‘girlfriend,’ ‘wife,’ ‘housewife,’ ‘widow,’ ‘femininity,’ ‘womanhood,’ ‘masculinity,’ ‘husband,’ ‘man,’ ‘boyfriend,’ ‘guy’ formed the parameters in the search. This strategy of key words made it possible to find some blog articles, but not too many of them. The blog articles found by a manual search with keywords made up the corpus. In total 45371 blog articles were found this way. But not all of them met the criteria of the data corpus for this study. Some of the blog articles were written by professional journalists or taken from other websites where professional journalists publish their works. For example this blog <http://maxpark.com/community/4057/content/2927579> was reposted by the blogger who took from another website [www.1prime.ru](http://www.1prime.ru)

Most data were found by manual search through the most popular interest groups devoted to the topic of the research question. The search by key words revealed that the following interest groups contain the biggest number of the words “woman” and “man”: the link to the interest groups can be found at

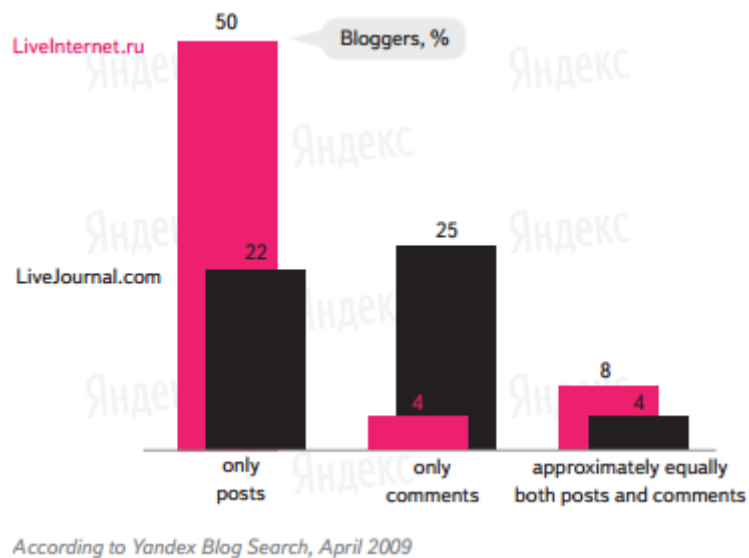
<http://maxpark.com/communities/search?search=%D0%B6%D0%B5%D0%BD%D1%89%D0%B8%D0%BD%D0%B0>

In total the search resulted into 115 interest groups found. The number of participants in these groups varies from 50 to 5000 people in the group. (All the blog articles from the groups were incorporated into the corpora). The most discussed blog articles from all 115 interest groups formed the corpora. The search was done manually by copy pasting all the blog articles from Maxpark into Notepad. Those 115 communities on Maxpark fall under the website category Relationship, Society, and General Interests.

Building the corpus with all the interest groups adds representativeness and reliability to the data. For instance, some of the groups have the name “We Russian women,” but the groups with the name “We Russian men” are also included. This means that the corpus does not have a one sided nature. As it comes from different interest groups or, as users say, communities inside the website, the corpus encompasses different perspectives and views on the research question. The data corpora assembled from Maxpark and LiveJournal Russia blogs are gathered in two separate text files – one for blogs posts, another for user comments.

LiveJournal is another website whose blogs became part of my data corpus. According to LiveJournal statistics of December 5, 2012, the following countries have the highest blog activity in LiveJournal: 1) USA, 2) Russia, 3) Great Britain, 4) Canada, 5) Ukraine. In general, bloggers from Russia and Russian-speaking countries make about a million blog entries every day - about

300,000 posts and about 700,000 comments (Yandex Blog Search 2009). Communities are updated more often than blogs: new posts in communities appear 1.3 times more often and comments 2.1 times more often than in blogs. Regarding the proportion of people who write and comment, the LiveJournal has an approximately equal number of both.



Graph 2.

The popularity of LiveJournal makes it a good website for data collection, ensuring a reliable representativeness.

During the actual process of building the data corpus, I familiarized myself on the kind of texts my corpus will have in terms of genre. Browsing manually through the websites during “construction” of data corpus, I researched the description of the interest groups and the titles of the articles taken into the corpus to be more prepared for the analysis from the linguistic standpoint.

**Scenario for research: Corpus – driven approach with hypothesis and categorization at the beginning and analysis with generalization at the end**

It was a little bit confusing which path to choose for the actual analysis of the corpora. Two main approaches widely accepted in corpus linguistics are the corpus – based approach and the corpus – driven approach; both methods fit my research very well. For instance, the scheme of the corpus – based approach allows for detailed and deep exploration of the corpus, but it is relatively time consuming as it requires an annotated corpus as well as categorization analysis at the beginning of the research (see the Graph 4).

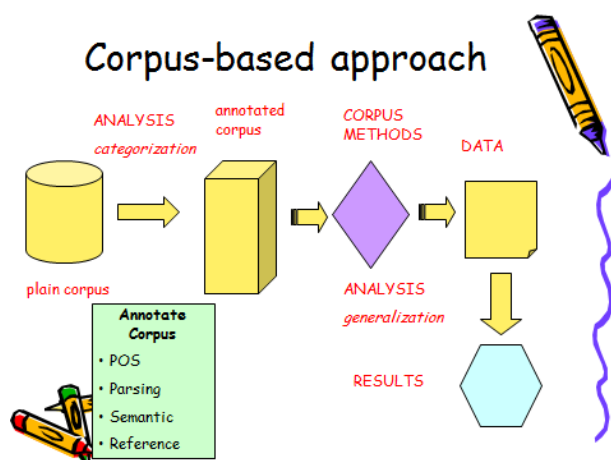


Figure 3. Corpus – based approach. Source: [www.lexically.net](http://www.lexically.net)

The corpus – driven approach makes it possible to work with a plain corpus, applying only corpus methods to the data analysis with generalization and categorization at the end. The corpus – driven approach allows more freedom in the analysis and it is not so much attached to the annotated corpus.

As Tognini-Bonelli (2001) explains, the main difference between those two approaches is that the corpus – based approach is much too dependent on some pre-existing theory. Graph 4 clearly demonstrates analysis categorization as a starting point in the research. The pre-existing hypotheses make use of corpora to prove assumptions or theoretical positions, but they very rarely challenge these positions or find new ones. In Tognini-Bonelli’s view there is less freedom and variability of language within the corpus – based model as it gives priority to the pre –



existing theoretical statement that “insulates, standardizes and reduces” this variability. The corpus – driven model is more flexible, since it uses the corpus not merely as a validation of appropriate ideas.

Tognini-Bonelli describes the corpus – driven approach as the one that builds up the theory step by step in the presence of the evidence, the observation of certain patterns leads to a hypothesis, which in turn leads to the generalization in terms of rules of usage and finally finds unification in a theoretical statement.

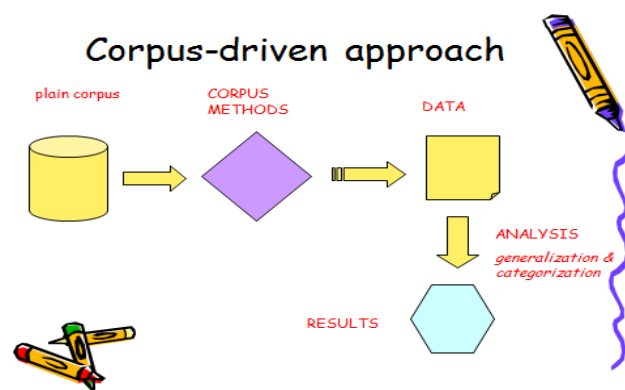


Figure 4. Corpus – driven approach. Source: [www.lexically.net](http://www.lexically.net)

My research takes the corpus - driven model as its base in the methodological path to the corpus analysis. The corpus – driven model allows for the research use of mixed methods. Even though I have a hypothesis which is the element of corpus – based approach, I would like to explore the question openly with freedom and variability of language. Due to the interdisciplinary nature of my research, the mixed approach to the corpus analysis is the best option. I undertake both hypothesis and categorization at the early stage of my research, but am using the data for generalization and analysis at the end.

According to Graph 4, the corpus – driven model does not necessarily include annotation, but my research will nonetheless include a slightly annotated corpus (as a part of corpus – based

approach). This is achieved with the help of TreeTagger automatic tagging and annotation library for the Russian language, available at <http://corpus.leeds.ac.uk/mocky/>.

Leech (2005) defines corpus analysis as an annotation process. In his understanding they are indispensable and self-supplementing parts of the text analysis. Word frequency, collocation and feature extraction serve as effective methods in this research. As long as the research seeks not to prove underlying assumptions but to find and analyze an open question, one shouldn't just rely on automation. Talking about methodology in text analysis, Danielsson (2003) says that "the distribution of words in texts is far more complex than a mathematical formula can perceive" (114). That's why logic and intuition should be used in any kind of analysis dealing with the text. A more in depth interpretation of the findings within my data corpus is possible when both qualitative and quantitative parts of analysis take place throughout the research process. I would like to explain the importance of this combination referring to McEnery and Wilson (1996) who say that quantitative analysis "enables one to separate the wheat from the chaff" (63). And on the other hand qualitative analysis that gives freedom for interpretation makes it possible for researcher to see the differences between the wheat and the chaff.

### **Help of Text Analysis Software**

To speed up the analysis process, I turned to the text analysis software. The types of free text analysis software for the Russian language are not so numerous. To analyze these Russian texts I programmed custom code with the help of a programmer from the University of Saskatchewan – Mayya Sharipova. The code was written in Python with the support of the NLTK library and TreeTagger program for the Russian language. As a result of this code (please see the attached text file), one can see the calculations of all parameters of the text necessary for the analysis.

## **Technological formula for research: Tokenizing+Annotating+Calculating+Extracting with the use of NLTK for Python and TreeTagger Russian version**

The programming code uses the NLTK library in Python program with built-in annotation tool - TreeTagger. This annotation tool annotates text with part-of-speech and lemma information for the Russian language. The goal was to write the code that could process the corpora of the Russian text and provide statistical information about it in English – key words, collocations and word frequency. Two Python processing files were included in the program. The code is attached. With the help of code the following text analysis was conducted with the corpus:

1) Tokenizing the text corpus into separate words - tokens. To do this I first eliminated the punctuation, and then used the NLTK word tokenizer to split text sentences into separate words-tokens. NLTK is a Natural Language Toolkit for Python, which allows a researcher to process human language data (<http://nltk.org/>).

2) Annotating tokens with morphological information. Morphological information includes information about parts of speech, gender, plural/singular form, case, and lemma of a token (a standard dictionary form of the word). For this data, I used TreeTagger software (<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>), and developed an annotated Russian corpora for reading the data (<http://corpus.leeds.ac.uk/mocky/>). This way, each word of my corpus was read by the program and given its own characteristics. For example, the word “woman” has a symbol Ncfpiy. Each letter of the symbol 'Ncfpiy' has a meaning. After looking this symbol up in the TreeTagger Russian description, one can find the following definition: The word “woman” is a noun, used in the instrumental case, feminine, plural number.

3) Calculating various types of statistical information about the corpus. The statistical information includes the following: number of all words in the corpus, number of different words, number of adjectives, the most frequent collocations, the most frequent adjectives, and all collocations that occur at least three times in the text corpus. For this we wrote a python script, making use of another library of NLTK called Collocations.

### **Process of creating the lists of negative and positive personality traits for sentiment analysis**

Finally, to be able to summarize and generalize the ideas presented in the data, we built a program for feature extraction. Code written in Python counts the number of positive and negative adjectives of both genders – feminine and masculine. Before building the program, I created two lists of adjectives – one for negative character features and another for positive features. The basis for these two lists was the website by popular Russian psychologist Andrey Zhalevich ([www.zhalevich.com](http://www.zhalevich.com)), who suggests the positive character features of a person (any gender) should generally have in order to be happy and successful and what kind of features a person should alter or improve to avoid misfortunes in his or her life.

Since Zhalevich's lists are pretty neutral, one cannot exactly say that (or how) these features can be attached specifically to women or men. The lists offered by Zhalevich are very long and cover almost all aspects of human character – 333 positive adjectives and 649 negative adjectives. But as my research question examines the traditional character of both women and men, I had to adjust the lists based on the traditional values of Russian culture the nature of which I discuss in the Analysis chapter. I had to switch some of the adjectives in Zhalevich lists because they simply did not fit into the traditional image of either the Russian woman or man. For instance, Zhalevich put “sympathetic” (сострадающий, сочувствующий) into the negative list

but I moved it to the positive list for women as “sympathetic” is present in many traditional notions about femininity. In general, however, Zhalevich’s lists are very realistic and reasonable for a sentiment analysis. First of all, the code that counts the adjectives helps to identify positive and negative associations used in the discourse. Secondly, it becomes easier to find and analyze features or topics of discussion in the data corpus. By the end of my research it was possible to formulate the rubric of features in the discussion of masculinity and femininity.

### **Principles of TreeTagger program in processing the data corpora**

During the procedure of adjective collection (positive and negative), to detect feminine and masculine adjectives in the corpus, I needed to normalize these adjectives in the corpus as well as bring them to their lemma form in order to compare them with adjectives from dictionaries (positive and negative). To find adjectives and normalize them, I used TreeTagger. The program works as follows:

- 1) The program takes the corpus, removes punctuation symbols, and breaks it into the tokens, or separate words, using the NLTK word tokenizer. The program creates a file tokens.txt – a list of separate tokens. This file is necessary for further application with TreeTagger.
- 2) TreeTagger tags every token with a part of speech tag, and other word forms including gender.
- 3) From the tagged tokens the program extracts adjectives and their lemmas. The program compares these adjectives with adjectives from dictionaries of positive and negative adjectives, and counts how many positive and negative adjectives were encountered. It then writes the result into the result file.
- 4) The result file displays the positive and negative adjectives by frequency.

Although this program works effectively, I had to double check manually the collocations of all the adjectives to make sure the program read the gender correctly.

All these quantitative, qualitative and lexical parameters of the text help to research the main question –representation of femininity and masculinity in the Runet. Numerical statistics (word frequency) show the importance of a word within the corpus. The words used the most frequently, or “key words,” help to build the overall context around a particular topic.

Collocations in their turn provide a more reliable overview of the meaning because they indicate not just single word-forms but phrases.

### **Keyword Method and its Collocations**

The nature of collocations reflects the idea of a semantic prosody in linguistics. The concept of semantic prosody (SP) was first introduced by Bill Louw in 1993 and widely explored by John Sinclair, John Rupert Firth, Michael Stubbs, Alan Partington, as well as Susan Hunston and Gill Francis. Semantic prosody is an important aspect in corpus linguistics that creates room for discussion, analysis, and interpretation of the text. The first definition that Louw offered for the phenomenon of semantic prosody was a “consistent aura of meaning with which a form is imbued by its collocates” (1993). In other words, Louw emphasizes that form and meaning cannot act inseparably, but are bound up with each other.

Semantic prosody can also be referred to as discourse prosody. The reason why semantic prosody or collocation has its own important niche in text analysis is because it helps to dig deeper into the keywords. With the help of technology it is relatively easy to identify the keywords. The analysis of the broader meanings and contexts and the ability to understand positive or negative associations around a particular word, however, are strongly enabled by semantic prosody.

The word expressed through frequent occurrences with particular collocations helps to understand the context and discourse within any given text. The most popular practical example of semantic prosody is exploration of the English verb “cause,” which is mostly used in the negative context (cause disease, loss, etc.), although cases of its positive meanings also occur, but are very rare e.g. “cause happiness” (Stefanowitsch and Gries 209-243). The equivalent meaning of “cause” can be found in the Russian language, e.g. “причинять боль” (cause pain) and, more rarely, the use of “причинять радость” (cause joy). Similar examples of verbs with negative prosody used in positive meaning in Russian can be “наносить добро” (inflict something good), “подвергать ласке” (expose to care).

The code I designed for this research makes it possible to see the context and understand the meaning systematically. Moreover, the combination of technology and human logic used in analyzing the text opens wider horizons. It becomes possible to see both pragmatic and semantic meanings. Semantic meaning mostly addresses the definition of the word and its lexical meaning in general, while pragmatic meaning examines the word in a particular situation and context within a given text.

### **Text visualization based on node-edge structure with the use of Texttexture**

Numerous studies show that text visualization helps for the representation of abstract data to reinforce human cognition (Moretti 2005, Liu 2012, Ramsay 2011). In my thesis, text visualization has the same function – to make understanding of the abstract data easier. Additionally, even the simple visualization provided by Texttexture significantly contributes to the generalization of categories or topics in the research area. After extracting the key words and its collocations from my corpora, I then ran the excerpts through the Texttexture program.

The Texttexture program was created in 2011 by Nodus Labs based on the open-source free Gephi software and Sigma.js toolkit. In addition to the relatively simple handling of data, the

program is effective for my research as it is free and handles Russian language. The goal of the program is to provide an easy-to-use online tool for text network visualization, fast non-linear reading, and comparative analysis of textual data. The methodology behind Texttexture is quite simple. The scanning of data includes three steps:

- 1) Removal of all the most frequently used “stopwords”, such as “are”, “is”, “the”, “a”, etc.
- 2) Turning of the word into its morpheme (e.g. “took” becomes “take”, “plates” becomes “plate”)
- 3) The resulting sequence is then scanned so that every word is encoded as a node and their co-occurrence is encoded as the connection between them.

The connection of nodes happens not only because of proximity in the text but also according to the paragraph and sentence structures. Designers call this process a Landscape reading model. As a result, the node-edge structure is encoded into a graph format, then processed by the Gephi server-side Java toolkit, which calculates the basic metrics and applies community detection algorithms for the graph. The size of the nodes is then ranged according to their betweenness centrality. The nodes are then colored according to the community they belong to. The nodes are considered to belong to the same community if they are more densely connected together than with the rest of the network. The modularity iterative algorithm embedded in Gephi locates these communities.

As the final step, the graph is visualized using Sigma.js library by Alexis Jacomy (2009), which also applies Force Atlas layout, pushing the nodes that belong to the same community together and pushing those communities apart in order to provide a more visually readable image of the graph (<http://texttexture.com/>). It is important to note that snapshots of the data visualized by Texttexture are shown as static in my paper. But originally the visualization works in the dynamics when processing the data on the website.



## CHAPTER III

### ANALYSIS

The following chapter of this paper provides description of the *blog* phenomenon in the World Wide Web and within RuNet with further text analysis of Russian blogs. To address the technological process of text analysis the details of programming code are explored together with the discussion how it works in Russian language texts. For better understanding of the results of text analysis the chapter moves on to examine in greater detail the comparative historical overview of femininity and masculinity notions in the traditional Russian culture. Overall this chapter reports the results of text analysis with subsequent data visualization made by Texttexture online program.

#### **Overview of the “blog” development throughout the history**

Before the actual text analysis of selected blogs, I should contextualize blogs in general and their development in relation to RuNet. Several definitions of the term “blog” appear throughout the history of the internet. Comparing different definitions by other scholars, I concluded that blog forms an information environment that can be named or characterized depending on practices and practitioners participating in the process. The term “blog” was intended initially as a tool (early 90s), then as a service (late 90s), and gradually the term turned into the process or practice. Blood (2002) attributes the first mention of the short form “blog” to Peter Merholz, who mentioned this neologism in the sidebar of his blog Peterme.com in 1999. Later on, the term was referred to as both noun and verb. Pyra Labs' Blogger product (1999) contributed a lot to widespread use of the term. In that definition, “blog” was a means or a tool of information sharing. In late 90s blogs evolved into the form of an online diary with great focus on the personal life of users. With creation of services like LiveJournal, the blog gained the power of action, turning into the online environment for journaling practice and self-publishing in the area

of personal and professional life. The blog now acquired features of a publishing environment with equal online access for everyone and the possibility to publish and comment on a particular topic of interest. From this point, the blog grew into community.

As I already mentioned in the Literature Review chapter there are a number of scholars researching phenomenon of blog as a genre and social activity. After looking through many definitions of the blogs advanced by such scholars as Herring (2004), Nardi (2004), Blood (2002) and others, I came to the following three conclusions: 1) blogs represent an environment with their own participants (bloggers), 2) this environment is productive as it produces information and reaction to this information in the form of comments, 3) the blogs perform significant communicative functions. All these features together can turn any prospective blog into a social space and information environment. As the literature review indicates, there is much research analyzing blogs as social spaces and information environments, but little research engaging the notion of blogs as language texts. Therefore, in my research I want to perform a deeper analysis to explore a certain component of blogs which is text. Starting from the assumption that the blog is text, I will analyze blogs using the text analysis methods discussed earlier. Later, I will connect the results of my text analysis with existing social and informative definitions of the term “blog”.

If we look at the blog as a text under analytical consideration, then it would be reasonable to consider its genres. The phenomenon of blogs or, speaking even more generally, the World Wide Web is considered new genres by many scholars. Susan C. Herring (2004) in her genre analysis of weblogs refers to various observations about the genre nature of the blog. The main question for Herring in identifying blog genre is “whether blogs are an emergent or a reproduced genre” (2). In the view of Crowston and Williams, the personal home pages have an "emergent" (rather than a "reproduced") web genre (10). Dillon and Gushrowski (2000) characterize blogs as unique genre because it is the first web – based text. Goroshko describes the internet as a

communicative and informative environment that gives rise to new genres inherent only in this given environment (Goroshko 117). Almost all the scholars reported that Web blogs have their own unique features and represent a new unique genre. The specific feature that Herring finds in the weblog genre is its “hybrid” nature that “draws from multiple sources, including offline and other internet genres” (2).

The numerous weblogs in the World Wide Web can be unique and common at the same time. They are unique in their diversity of text styles and languages as a means of expression for different ideas; the different people writing with diverse motivations also add to weblogs’ unique status. Yet they are common because of their technological means of text representation and structure. Even though the blog websites can be built and designed in one or another web platform or programming language, the weblogs still have their own different features. Keeping in mind this observation, in addition to the fact that the blogosphere in Russia and generally throughout the WWW is always a constantly evolving and relatively new area, I try to avoid generalizations about web blogs as a genre at this point of my research. Before proceeding to examine masculinity and femininity in the Russian blogs, it is important to discuss basic statistics about the websites Maxpark and LiveJournal Russia where the data corpora were assembled.

### **Statistics of RuNet**

The Russian internet (RuNet) has reached its highest peaks of usage only in recent years. It is hard to calculate exact statistics about RuNet because the term RuNet covers not only Russian speaking websites designed in Russia, but also Russian speaking websites in the former Soviet Union region. Despite such ambiguity, different Russian and international monitoring agencies conduct research on RuNet usage. According to *Internet World Stats*, the internet growth index in Russia made up 2.1% (3.1 million people) in 2000. By 2007 this index reached 803.2%, which was 19.5% (28 million people). According to the Public Opinion Foundation, the

semi – annual readership of RuNet at the age of 18 and older made up 33% or 37.5 million people in 2007. The recent data by TNS Russia (2013) indicate that the number of RuNet readers reached 76.5 million people or 53.5% of the general Russian population. It is important to note the geographical locations of the readers. Most are located in major cities. For instance, in Moscow and Saint – Petersburg almost 75% of population turn to the internet every day. As to the cities with the population less than 100.000 people, the penetration is only 50%. In 2013 the number of mobile internet users grew to 45%. The statistics mentioned above show that RuNet is experiencing its major growth in the beginning of the twenty-first century. Another important aspect here is the age of RuNet users. The research presented in *Russian Internet Forum* together with *Internet and Business* indicates that users in the age of 18 – 24 make up to 96.6% of RuNet.

### **Russia is the most active user of LiveJournal**

Although my research focuses on two particular websites, Maxpark and LiveJournal, the statistics above help fit my analysis of discussion on those websites within a general context of the internet situation in RuNet. LiveJournal enjoys great popularity among Russian users. The information monitoring company Alexa scores Russia as the most active user of LiveJournal with the most popular browsing location – with schools and females overrepresented. In contrast to LiveJournal, the statistics of Maxpark indicate that its readers are quite mature. Alexa summarized that males are overrepresented and the average age of most users is over 35 on Maxpark. These facts reveal the diversity of people blogging on these two websites. This diversity can also be found in blog styles, text genres, and many other aspects that I will analyze later in this chapter.

### **LiveJournal as a point of Russian collectivism and resistance at the same time**

The websites Maxpark and LiveJournal Russia differ from each other in many ways – in their structure, topic orientation, audience, visual design, visitor rating and many other aspects.

Although LiveJournal Russia (LJR) is part of international LiveJournal network, there are certain specific features in LJR. The difference is in peculiarities of national mentality and cultural traditions, as well as in economic development in the Russian speaking region of the CIS. Eugene Gorny (2006) was one of the first scholars who researched LJR. Although his work dates back to the early 2000s, there are certain findings in Gorny's work that still contribute to the description of LJR currently. Gorny observes that "LiveJournal Russia was first populated by users who had authority and could influence others to adopt the innovation" (89). This observation reflects the fact of relatively slow penetration of LiveJournal in RuNet. After engaging a little bit with LiveJournal and realizing all its advantages of community – building, the Russian public appreciated LJR. Gorny states that it is probably the Russian "tendency towards collectivism" that made LiveJournal interesting for the country, "LiveJournal has meshed well with the Russian cultural identity and social circumstances, resulting in the special value of informal networks" (89). The opportunity to belong to the community and feeling of collectivism which is so inherent in the Russian culture turned LiveJournal into one of the biggest website preferences of Russians since its launch and until the moment of the present research (2013 – 2014).

Together with this positive collaboration opportunity, LJR nonetheless brought some difficulties for the Russian users or, as Gorny calls it, "a slight tinge of coercion" (89). To be part of the LJR community, one has to register first by sharing your personal information with the website and other users. It was a question for many Russian people if it is safe to share your info in early 90s. As absolutely new phenomena, LiveJournal and RuNet in general took time to gain a certain degree of trust from the people. Even after sharing some minimal portion of personal information, people were reluctant to actively use LJR as one becomes tied to the certain web place, interface, website and functions. Happily, the desire to communicate prevailed over all other fears and LJR became more and more popular. Gorny holds the view that Russians saw the

potential in LJR to transform virtual community into real life, and exactly this feature made the website so popular. Gorny refers to the off line events and projects that were discussed and initiated in the LJR first and communally realized in real life. One of those events is the Current Music Festival conducted in Moscow yearly since 2000 (Gorny 90). Nowadays one can find more examples of such mobilization from online into real life projects. In everyday life it is especially popular starting from meeting arrangements and up to more global social level in the form of fundraising events.

### **Maxpark – a mirror of social and political life of the country**

Maxpark is the Russian website or weblog that offers almost the same collaboration and information sharing opportunities as LJR. The main difference is that Maxpark positions the blogging primarily as debate activity or even a duel. The website offers to mark your connections as friends or opponents. Many popular experts, journalists, actors share their ideas and participate in debates on different topics. If LJR is mostly focused on the grassroots and young people in the process of getting an education Maxpark's audience is more mature, representing mostly people with high education. The difference is also in topics being discussed. The most popular areas of interest on LJR are current events and news, music, cultural life and travelling (Link to the list of most popular posts <http://www.livejournal.com/interests.bml?view=popular&mode=text>).

Exploring live stream of the blogs on LJR I noticed some trends in topics and blogging style. LJR posts are very rich in multimedia, sparkling with lot of pictures taken by users everywhere. Most posts have a descriptive character telling about the event a blogger visited or describing someone's mood by the end of the day. The titles of the posts are usually very short e.g. *Last day in London; Pluschenko Out*. The posts on Maxpark bear different style and topic choices. The most popular topics for discussion on the website are in politics, sociology, and culture. Maxpark posts have more scientific and analytical character

<http://maxpark.com/communities>. The core two parts of website include People and Groups (Communities). The groups are split into the most active, new, and popular categories. Rating of the most active groups almost every day includes the following communities: 1) Politics, 2) Moment of Truth, 3) Interesting News, 4) Back in the USSR, 5) Alternative friends: from word to action, 6) Maxpark Newsfeed, 7) Russia and the world, 8) South Russia, 9) We are Russian, 10) Haven for optimists etc. Titles of the posts are longer than on Maxpark and in most cases they express a point of view, or contain the appropriate idea or fact rather than describing an idea. For instance *Lace panties became prohibited in the countries of the Customs Union*.

### **Programming Code for Sentiment Analysis of Masculinity and Femininity**

The technological process of text analysis is outlined in the following section. The programming code that was described in the previous Methodology chapter is used to achieve the following goals: 1) get statistics about keywords and its collocations, 2) apply the statistics for the analysis of discussions, in particularly sentiment analysis. In my case the key words in context are “woman”, “man”, “wife”, “husband”, “male”, “female” and those adjectives collocated. Sentiment analysis identifies a positive or negative opinion expressed about masculinity and femininity in the texts. Sentiment analysis will be applied to blogs as well as comments gathered in two separate corpora. Comments represent special value in the research of any internet discussions as they represent the voice of the grassroots. Someone might say that bloggers also stand for the ordinary public, but bloggers are very often referred to as professional amateurs in writing. Although in most cases they do not have any special training, they still have certain skills to write an article that fits a particular website. One single blog article can contain a thousand of comments that represent many public voices. Keeping this in mind, I decided to research the discussions in both blogs and comments.

## **Process of Feature Extraction in the Corpora**

First, the extraction of negative/positive adjectives was done to examine sentiment analysis on the image of modern Russian woman and man in the corpora. Moreover, my goal was to connect this sentiment analysis with the idea of traditionalism. Then the same operation was done with the corpus of comments. Before running the data corpus I built the lists of positive and negative adjectives characterizing the image of traditional Russian woman. As the research also deals with men, I had to adjust the adjectives so that they reflect realities of Russian culture. The original lists of adjectives were taken from the website by popular Russian psychologist Andrey Zhalevich. As a psychologist of the new school Zhalevich proposes personality traits characterizing a new Russian person, not traditional one. He describes those traits as not gendered and applicable to any person who wants to achieve success and happiness. Due to the nature of research, I had to make those adjective lists gendered and more relevant to the traditional image of Russian woman and man.

As I already mentioned in the previous chapter, the adjective “sympathetic” (сочувственный, сочувствующий) usually characterizes a person and especially a woman in a positive way. That is why I switched this adjective from negative to positive one. Another trait is “timidity” (робость in Russian) which also has positive meaning for a woman in traditional Russian culture, but was put into negative list by Zhalevich. After these light adjustments the lists of adjectives fitted into description of traditional Russian woman and man. The data corpora were run through the programming code to detect sentiments regarding traditional Russian woman and man. Then the results of sentiment analysis were used to explore the main topics of discussions and degree of traditionalism in the representations of masculinity and femininity.



So far this chapter has focused on the phenomenon of blog and technological side of the research. The following section will discuss the concepts of femininity and masculinity in the traditional Russian culture with subsequent overview of quantitative results of the text analysis of corpora.

### **Masculinity and Femininity in Russia: From Past To The Present Day**

Russian sociologists Olga Zdravomislova and Natalya Kigay (2002) conducted the research on woman discourse in the Russian media written by professional journalists. She came to very interesting conclusions. Traditionalism in the evaluation of femininity is higher than in the evaluation of masculinity. She found that the image of a modern Russian woman combines two controversial ideas. First, the vision of a modern Russian woman is based on traditional classical principles of fundamental difference between woman and man. Secondly, the representation of a modern woman is strongly influenced by the idea of feminism and emancipation. The researcher cites a fact about high employment activity among women in all social classes (Zdravomislova and Kigay 2002) In this chapter I will try to question this hypothesis, whether the texts from popular Russian blog websites prove this notion or not.

Before the actual discourse analysis of the corpora I would like to describe shortly the evolution of understanding what masculinity and femininity are throughout the Russian history. The traditional image of Russian woman is seriously and very broadly reflected in classic Russian literature (XIX century). Many Russian writers tried to create what they thought were very realistic images that accumulate the personality traits inherent in the Russian culture, especially in women. The greatest writers like Pushkin, Ostrovsky, Turgenev, Nekrasov and Tolstoy showed a number of interesting and true images of Russian women. In the literary pieces by Fyodor Dostoyevsky we also come across different images of Russian women. Most of them fall into the general theme of all Dostoyevsky's literature that narrates about difficult unhappy

life of the poor, honest and kind people of Russia in the XIXth century. All these classical writers together revealed the richness of moral and intellectual virtues of women living in those times, their pureness, mind, a heart full of love, patience, striving for freedom and fight for justice. Those are the features typical for the traditional image of Russian woman in classic Russian literature.

As time has passed the cultural and social life of the country also underwent some changes. In the middle of XX century the industrial mood of the Soviet Union was on the rise. This led to a new type of a woman that was working hard in the factory rejecting “bourgeois consumptive approach to life as well as commercial dominance of sexuality in female beauty” (Gurova 2005). The image of woman in Soviet culture was based on such notions as health, strong will, naturalness and devotion to the Party. The woman becomes a comrade or a citizen who considers public interests superior to her personal ones (Shlapentokh 1989). The women of Soviet times are not so subtle emotionally or mentally as those described in the early XIX century.

Coming closer to the present days the image of a woman acquires new traits and becomes a more mixed type. There is no exact definition or description of a modern Russian woman that would be widely accepted by most of the people. This question is very often a point of debates. Some social and cultural experts refer to the glamorous and careless girl, kind of ‘social butterfly’ as the most common image of a modern Russian woman (Gurova 2008). But this assumption also finds its opponents who claim that “the image of a skinny glamorous girl” was imposed on modern Russian culture and will not last long (Female beauty 2012). Sociologists connect the image of modern Russian woman with the difficult situation in terms of domestic violence. According to statistics 36 000 Russian women suffer battery daily from their husbands (RIA News 2013) This fact adds to the traditional image of Russian woman from a poor village family

who runs through brick walls to support the family and to provide everything for family without getting any appreciation in return:

A run-away horse she will master,  
Walk straight to a hut that's aflame  
("Коня на скаку остановит,  
в горящую избу войдёт". Nekrasov 1974  
translated by Soskice, J)

To explore the traditional image of Russian man I turned again to the classic Russian literature as well as Encyclopedia of Male and Female in Traditional Russian Culture (Мужики и бабы: Мужское и женское в традиционной русской культуре, Иллюстрированная Энциклопедия) written by Baranov, D. and Baranova, O. in 2005. The Encyclopedia describes the man using such words as husband and "muzhik". Muzhik, an authentic Russian word, means ordinary Russian man who has strong will as well as very hardworking, plain and sometimes rough personality. The image of muzhik reflects the views of masculinity in traditional Russian culture. A man was the one who is married and actively participates in the life of village and home. He was supposed to be the master of home, the one who has the last word in the family, "responsible for family and farm before people and God" (Encyclopedia 372). This image especially corresponds with the times of serfdom and up to the moment when it was abolished in 1861.

Later in XVIII - XIX centuries we see the image of Russian man described in classic Russian literature. He comes from a rich aristocratic family of a good name. He is a nobleman, sometimes social party goer like Evgeniy Onegin by Pushkin, or an officer of the Russian army like Pechorin by Mikhail Lermontov. But these figures do not represent the vast majority of people of that epoch as these heroes come from the upper class of society.

Soviet culture teemed with the plain guys who were as easy as old Tilly, very positive, in most cases working in the factory and living in family. Soviet propaganda paid no attention to their appearance, their clothes were shabby, their families were supposed to be poor but friendly and happy.

In general, there are fewer resources describing Russian men from the point of traditional Russian culture. For instance even the term “muzhik” does not have as much coverage as the term “woman” or “baba” does in the Encyclopedia. The resources analyzing Russian women outweigh those depicting Russian men. That’s why it is hard to characterize any men. But speaking about the image of modern Russian man, I would like to refer to the most explicit examples from movies. In early 90s the typical hero was a New Russian, a bandit with lot of money that he earned through criminal activity. Towards the 2000s the male image took a more pleasant form – now this can be a businessman or a manager stressed out at work. The presence of family in the life of modern man is no longer paid as much attention as it was in other historical periods.

### **Calculation of Keywords/Collocations and its Frequency for Further Analysis**

The next part of this chapter demonstrates the results of text analysis together with sentiment analysis of corpora. After running the corpora through the programming code the statistics of the most frequent negative and positive adjectives was revealed in a corpus of author blogs and a corpus of user comments. It is also important to note that the programming code did not reveal the list of gendered adjectives ready for analysis. To make sure the adjectives belong to a particular male or female gender the analysis of the collocated words was done manually.

There are certain theoretical aspects of text analysis process in this research that need to be addressed before moving further to the discussion of actual results. In the view of J. R. Firth (1935) it is not only the words that should be looked at during the data analysis but their

distribution within the text, as he says “formal scatter” (Stubbs 53). Firth proposes a contextual theory of meaning. To grasp the meaning of the negative/positive adjectives within the overall context of my data corpus I applied the collocation method. This approach helped to explore not only the sentiment or attitude of the blogger towards the key word but also an idea and general context.

Thanks to this statistics it was possible to characterize specific sentiment expressed about a woman and a man in two corpora. The total number of blogs was equal to 235. Statistics in the corpus of author blogs is the following: the number of positive adjectives was 394 while negative adjectives made up 163. It is clear that use of positive adjectives is almost twice higher than negative ones. To be more specific program showed the following statistics:

Beautiful красивый	49	Responsible ответственный	7
Needed нужный	31	Natural естественный	6
Strong сильный	18	Fully fledged полноценный	6
Emotional эмоциональный	17	Concrete конкретный	5
Confident уверенный	13	Independent независимый	5
Kind добрый	12	Constant постоянный	5

Simple простой	11	Timid Робкий	5
Lively живой	10	True истинный	4
Direct прямой	9	Pleasant приятный	4
Bright яркий	9	Useful полезный	4
Light светлый	8	Soft мягкий	4
Wise мудрый	8	Grateful благодарный	4
Cordial сердечный	7	Faithful верный	4
Worthy достойный	7	self-reliant уверенный в себе	4
Correct Корректный	7	Warm Теплый	4

Table 1. Positive adjectives by frequency

Extreme крайний	23	Closed закрытый	2
Material	9	Indifferent	2

материальный		безразличный	
Cold холодный	8	Unrequited безответный	2
Drunken пьяный	6	Outdated устаревший	2
Evil злой	4	Boring скучный	2
Dull тупой	4	Tough жесткий	2
Passive пассивный	4	Negative негативный	2
Annoying раздражающий	3	Vicious порочный	2
Naïve наивный	3	Rough грубый	2
Wild дикий	3	Bad плохой	2
Forcible насильственный	3	Merciless безжалостный	2
Sly хитрый	3	Defective дефективный	2
Dependent зависящий	3	Weird странный	2

Stupid тупой	3	Unhappy несчастный	2
Criminal криминальный	2	Cruel жестокий	2

Table 2. Negative adjectives by frequency

In both lists there was an adjective with the highest score and break away. It was positive adjective “beautiful” (49) and negative “extreme” (23).

The first 27 positive adjectives that got the highest frequency reflect the features typical to traditional image of Russian woman in particular. Although the first adjective “beautiful” (красивый) has descriptive meaning of physical appearance, almost all the rest describe the inner world of a woman e.g. “confident”, “strong”, “cordial”, “kind” etc. In sum all these 27 adjectives correspond with the traditional portrait of a woman – “needed”, “bright”, “wise”, “timid”, “grateful”. It is very interesting how “timid” and “independent” reside in the same list of positive adjectives. This reminds of complexity of the female image in Russian culture. As we see from the list those demands required her to be “strong”, “independent”, “correct”, “confident”, “responsible” and have other qualities that one could attach to a man as well. But at the same time she cannot lose traits specific to femininity, such as “beautiful”, “obedient”, “gentle”.

As to the negative adjectives found, we see that they also suggest traits of character relevant to the inner world of the person. Despite the fact that most adjectives in the lists were purposely chosen to describe the character (inner world of the person), the words “beautiful,” “tidy” are included in the list and scored high frequency in the blogs. As to the word “unattractive” it was included in the negative list. This choice is explained by the purpose to analyze the attitude toward physical features of a woman and a man. How important and how frequent this word is mentioned. As it shows the statistics of “unattractive,”



(непривлекательный) this word has a very low frequency in both corpora. This leads to the conclusion that being “unattractive” is not the problem in case of both genders. Although the concept of physical beauty was regarded as a virtue for a woman in the traditional Russian culture, its absence was not openly criticized. The situation with a man is different, traditional culture pays less attention to the physical attractiveness of a man but also does not openly criticize lack of this attractiveness. Another explanation to the little use of the word “unattractive” in the blog articles can be different interpretation of the concept “beauty” in different cultures.

In terms of Russian traditional culture female beauty was very often referred to the combination of physical beauty with mental power, with sun or light. Actually this combination can be found in many other traditional ethnic cultures. For example Aphrodite and Venus are described as “fiery” with “lightsome smile” in Greek myths (Telenkova online article) In terms of male beauty, its most popular symbol in Russian ancient epic folk stories is *bogatyr* (from *baghadhara* in Sanskrit “somebody that possesses happiness and luck”) which refers to a physically very strong male hero with divine powers who protected the Russian land, for example Dobrynya Nikitich (Gusterin 2012). We will turn to the discussion of male beauty later in this paper.

The negative adjectives with the highest frequency could be good characterization of “b\*tch” as they refer to very practical and cynical woman (cold, material) with bad habits (drunken, criminal) and narrow mind. “Extreme” became the adjective with the highest frequency for both men and women. What can the use of “extreme”, “extremely”, “absolutely” tell about a woman?

After running the corpus consisting of user comments I came across the following statistics. The number of positive adjectives in the comments equaled to 299 while negative adjectives constituted 155. The first 27 adjectives in both positive and negative list of user

comment corpus differ from the first 27 in the author blog corpus. Below you can find the lists of adjectives by frequency:

Local местный	1591	Bad Плохой	109
tall высокий	1013	Beautiful красивый	109
Black черный	281	Similar подобный	101
Needed нужный	211	New новый	96
Good хороший	203	Simple простой	93
Must должный	171	personal близкий	92
Normal нормальный	168	Important важный	88
Female женский	164	Smart толковый	83
Equal равный	146	Russian русский	82
Dull глупый	143	Last последний	80

Free свободный	128	Interesting интересный	78
Different разный	126	Young молодой	76
Concordant совместимый	125	Happy счастливый	71
Male мужской	120	Complete Цельный	70
Big большой	111	Ready готовый	68

Table 3. General adjectives by frequency in user comments

If we look at the first 20-30 most frequent adjectives in general category one can see that program counted personal last names as adjectives as well. The reason for that could be the ending of last names in Russian –oy –a, -ya, (for example Krutoy, Krutaya) the same endings can belong to the adjectives. After taking the last names out of the list it is possible to see statistics about actual adjectives in the corpus of comments. The general statistics of adjectives shows that positive adjectives are used more frequently than negative ones: “needed” (нужный 211), “good” (хороший 203), “normal” (нормальный 168), “female” (женский 164), “equal” (равный 146), “free” (свободный 128). “Dull” (нудный 143) and “bad” (плохой 109) are the only adjectives with negative meaning among frequently used ones.

Regarding the negative adjectives with the highest frequency they are “cynical” (циничный 17), “wild” (дикий 13), “extreme” (крайний 12), “stupid” (глупый 9), “silly” (неумный 6). Among the positive adjectives, those with the highest frequency are “needed”

(нужный 63), “beautiful” (красивый 48), “simple” (простой 16), “confident” (уверенный 14), “independent” (независимый 9). A closer look at separate lists of positive and negative adjectives gives a better insight into sentiment expressed in the corpus of comments.

In terms of traditionalism almost all of these adjectives can be applied to the image of traditional Russian woman except for “independent”. If we speak about the image of traditional man then “beautiful” does not really fit in, because male beauty in the traditional Russian culture is sort of physical and mental strength. But to be able to analyze what exactly the commentators mean by using the adjective “beautiful” we should look at the collocations of the key words in contexts.

There are some coincidences as well as differences in the use of most frequent adjectives. The corpus of user comments includes such adjectives as “kind” (добрый), “independent” (независимый), “concrete” (конкретный) and “lively” (живой). These frequent positive adjectives are absent in the list of ones used in corpus of author blogs. This does not mean that these adjectives were not used at all. They were used but not so frequently. It is also important to keep in mind probability of error, as the programming code does the mechanical work. The machine has an artificial brain, not a human one.

The most frequent negative adjectives in the corpus of comments are also those that are not presented in the corpus of blogs. “Cynical” (циничный), “wild” (дикий), “stupid” (неумный), “superfluous” (излишний) and “naïve” (наивный) were used very often in the comments’ corpus and not so much in the blog corpus. In general, the negative adjectives in the comment section have a stronger meaning, and the nature of language is more informal comparing to the author blogs.

### **Discussion of beauty on the physical level in femininity as well as masculinity**

The rest part of this chapter is going to focus on the keywords and the most frequently used adjectives (positive/negative) in both corpora. Coming back to the collocations of the word “beautiful”, the adjective beautiful was included in the list of positive features of a woman. But as it shows the graph, “beautiful” is frequently collocated with nouns of male gender as well e.g. “beautiful IT – nerd guy”, “beautiful man”. The cases of collocation with female subjects and objects are still higher. This shows significance of this feature in a woman. In the overall statistics of corpus of blogs the adjective “beautiful” is mentioned 105 times.



Figure 5. Keyword “beautiful”

Nouns placed near beautiful do not have an abstract character. On the contrary they mean real things or people like “beautiful woman” or “beautiful house”. But the word “beauty” as a noun in most cases is used on the abstract level. If we look at the collocations of “beauty” it is obvious, that as a noun it has mental evaluative meaning rather than physical evaluative in case of “beautiful” as an adjective.

“While they are young and beautiful” (пока они молоды и красивы) – physical evaluative meaning,

“To find beauty and joy in personal life” (Найти красоту и радость в собственной жизни) – mental evaluative meaning,

“Flexible strength”, “shining beauty”, “femininity”, “radiance of happiness” (Гибкая сила, лучистая красота, женственность, сияние счастья.)

In terms of semantic prosody, “beauty” as well as “beautiful” are mostly used in the positive meaning. The physical as well as mental (abstract) beauty is praised. The cases of negative prosody also take place, “her monotonously beautiful” smile hides inner futility of her soul” (Её однообразно-красивая» улыбка словно скрывает внутреннюю пустоту души). Here the blogger is talking about boredom, monotony or uniformity that can be found in the modern image of a beautiful woman. Despite physical attractiveness, the blogger finds the traits of an empty, futile inner world that is reflected in the smile. In the example of “monotonously beautiful” we witness loss of original meaning of the word with the acquisition of a new discourse role. Elena Tognini-Bonelli describes such situation with the adjective as delexicalisation. She connects the phenomenon of delexicalisation with the “extension in size of the unit of meaning,” “a signal of a new role for the original unit of meaning” (Tognini-Bonelli 120). The case of “monotonously beautiful” partly falls under the category of delexicalization. The full delexicalization however would be if the term “beautiful” lost completely the evaluative meaning. Tognini-Bonelli in her example of the word *bel* (beautiful in Italian) demonstrated its complete delexicalization when *bel* was used in such expression as *un bel po* - a good bit or *un bel nulla* – nothing at all.

To sum up the analysis of discussion with the key word “beautiful”, it is important to notice that traditionally Russian culture did not pay attention to the physical beauty of a man. Frequency of the adjective “beautiful” with the collocation “man” demonstrates increased importance of physical beauty in appreciation of male qualities.

**Male needs given more priority than female ones**

The second adjective in the list of positive ones in the corpus of blogs is “needed” and “necessary”, in Russian «нужный» (31). I split this word into two for the English language to be able to give the proper translation. In the English meaning of “necessary” (необходимо) as the adverb is used 15 times, while in other 16 cases it is used as an adjective – “needed” (нужный), very rarely as a noun “need” (нужда).

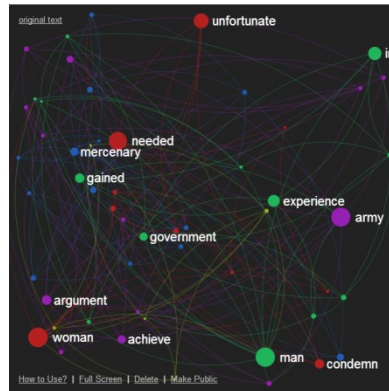


Figure 6. Keyword "needed"

In many of the sentences above “needed” as well as “necessary” function not only as an adjective but as a verb as well. The patterns of “need” as a verb and the case of its collocations help to identify whose needs are described and focused on in the text. It is obvious that mostly “need”, “needed” and “necessary” are mentioned in the discussion oriented on male topics, despite the close collocation with the word “woman”. The word is reaching its peak of use in the discussion of military topic e.g. “there is a need for a soldier”, or “a soldier is needed” (нужен солдат). The repetitive expression addressing the woman “Who needs you?” (Кому ты нужна) echoes the traditional understanding of Russian woman whose happiness was supposed to be full only in case someone (a man in particular) needs her, waits for her at home. It is more important for a woman to be needed rather than she needs someone or something. In the corpus of comments the indignant female voices follow with the emphasis on what woman needs:

“attention”, “gifts” and “compliments” (внимание, подарки, комплименты) The analysis of another keyword “must” serves good continuation of the discussion about “need”.

**Main female need is to be needed. Adjective “needed”, words with root “need”**

There are similarities in the use of adjective “needed” (нужный) in both corpora. Most of sentences with “needed” as well as “need” are presented in the discussions about both genders. The graph below shows connection of the key word “needed” with woman and man. Moreover, the words “not needed” (не нужный) or “not necessary” (не нужно) are also used quite frequently. Like in the first corpus of blogs most excerpts here discuss the type of personality that modern women and men look for, also what they need and do not as well as other aspects of relationships.

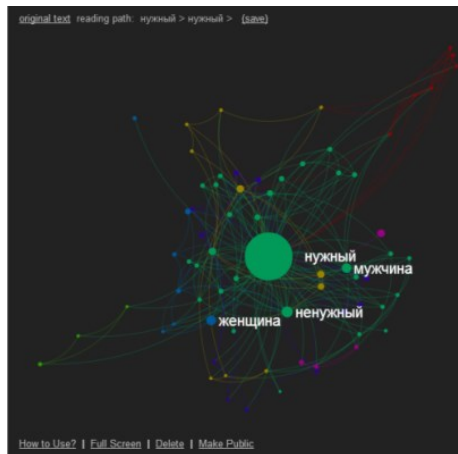


Figure 7. Keyword "needed" in the comments

The very meaning of the word “need” signifies that discussion in this part of comments’ corpus has a little bit educative character. Many comments contain the phrases like “it needs to be done” (это нужно сделать) or “it is necessary that” (это необходимо сделать). In terms of the use of “needed” as an adjective we see the same trend as in the first corpus of blogs. Mostly, this adjective is attributed to females rather than males: “she is not needed by anyone”, “no one needs



her” (она никому не нужна), “to understand if he needs you” (чтобы понять нуждается ли он в тебе).

Traditionalism is reflected almost in all the sentences with adjective “needed” in female form. Many commentators express the idea of a female happiness that can be full only if a man needs her. The interesting observation is that some female bloggers explicitly confess this idea and express regret about not feeling that they are needed e.g. “I felt that I wasn’t really needed” (я не чувствовала себя действительно нужной). Female needs are obviously underrepresented comparing to male needs. Many excerpts tell that “a man needs a woman for entertainment” (мужчине нужна женщина для развлечения), or “to feed and provide for him” (кормить и обеспечивать его).

There are also instances that narrate about modern men that need “no women at all, neither smart nor rich women” (женщины вообще не нужны, не умные, не богатые). As you see, there is almost zero traditionalism in comments about male needs. Looking back in the traditional Russian culture, a man is the head of a family, the guarantor of the continuation of his family with commitments before the state and society. At the same time, when the discussion touches upon women’s “need”, the commentators describe basic female “need” as being needed by men. Imbalance in appreciating female and male needs based on the traditional concepts is quite explicit in the corpus with comments.

### **Discourse of female strength focused on career but not family**

The next adjective in the list of positive ones is “strong” (сильный). Frequency in the overall adjective statistics is 89 in the corpus of blogs. This adjective is the third most frequent in the collocation with word woman in the corpus of blogs. Although “strong” as an adjective was mentioned only 89 times, “strong” in the other grammatical forms like adverb “strongly”

(сильно) for instance scored higher use. But according to the basic principle of the text analysis “strong” and “strongly” are two different words although they share the same root “strong”.

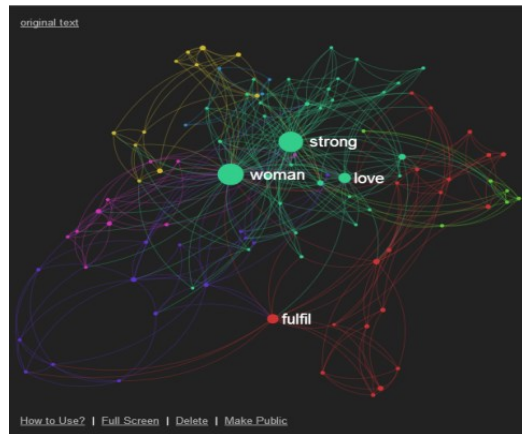


Figure 8. Keyword "strong"

Collocations of “strong”: woman, man, love, fulfill, feeling

In most sentences “strong” is mentioned in the discussions of power, responsibility, equality between woman and man as well as distribution of their social roles. In most sentences it is the mental or spiritual strength (not physical) being discussed. Moreover, this strength is attributed to women as a feature of their character. The attitude towards this feature varies depending on the group a particular blog was written in. Some bloggers say it is a good feature to have, others have doubts regarding it. Examples of “strong” as a feature of a man also exist but it has a meaning of physical strength rather than spiritual e.g. “a man strongly “forechecks” women” (и мужчина сильно прессингует женщин) or “men being stronger went hunting” (мужчины как более сильные, они уходили на охоту).

In the case of a man, the adjective “strong” is always mentioned in positive semantic prosody in both corpora, with collocates expressing approval, admiration or simply positive public opinion e.g. “women admired by strong men, he becomes strong, or tell to your man that

he is the strongest” (женщины восхищаются сильными мужчинами, он становится сильным, скажи ему, что он самый сильный). These sentences suggest an idea that a man would like to be strong and moreover a woman would like him to be strong. The strength of a man is taken with very positive attitude in both corpora while attitude towards female strength is more dubious. For instance there are bloggers that say in the corpus of comments “A girl should be strong” (девушка должна быть сильной); or “I must be strong to be able to help you” (я должна быть сильной, чтобы помочь тебе); or “since early childhood they are suggested an idea that a woman must be strong, never be ill” (с самого раннего детства им внушают идею, что женщина должна быть сильной и должна никогда не болеть).

The blogs positioning a strong woman positively outweigh those with a negative attitude towards this feature. The voices criticizing women for being much too strong and powerful are also fairly numerous. Some of the blogs and comments suggest that modern Russian women are too strong, and that they took too much over their shoulders. By being so strong women deprived men of any reason for being strong e.g. “When we start obeying we fulfill our female purpose and this way we facilitate growth of masculinities in our favorite representatives of the sterner sex” (Когда мы начинаем слушаться – мы выполняем женское предназначение и способствуем росту мужских качеств в наших любимых представителях сильного пола). In this excerpt the blogger questions the whole idea of female strength. There is a popular idiom in the Russian language that states that “female strength is in her weakness” (сила женщины в её слабости). In other words only a strong woman can let herself be weak and let a man be strong. But in the traditional Russian culture this female strength is directed to the matters of family, her husband and home.

The biggest part of texts in the blog corpus is looking at female strength in terms of social equality, career success, and public appreciation in general. In their discussion bloggers put

woman into realities of modern life in Russia where a woman is supposed to work and provide for the family, but at the same time, she is expected to fulfill the female purpose in family. The findings of the analysis of “strong” reflect Zdravomislova and Kigay’s idea about mixed representation of a woman. In the corpus of comments the adjective “strong” is not used very often, its frequency just counts 49.

### **Female beauty as a double-edged weapon**

The next positive adjective in the corpus of comments is “beautiful” (красивый 48). According to statistics the word is mostly collocated with such words as “she”, “woman”, “wife” and other nouns of female gender. The sentences with this adjective relate to the discussion of beauty as a physical and mental feature of a woman. One can see some comments where beautiful is collocated with male nouns like a “man”, “guy”, “husband” etc. But mostly this adjective is mentioned in the discussion about woman. The graph below shows red node “woman” and yellow node “man” as the collocations of “beautiful”. It means that “woman” has wider use in the collocation with beautiful.

Collocations for “beautiful”: woman, man, meet, interesting, scandal, expensive, good

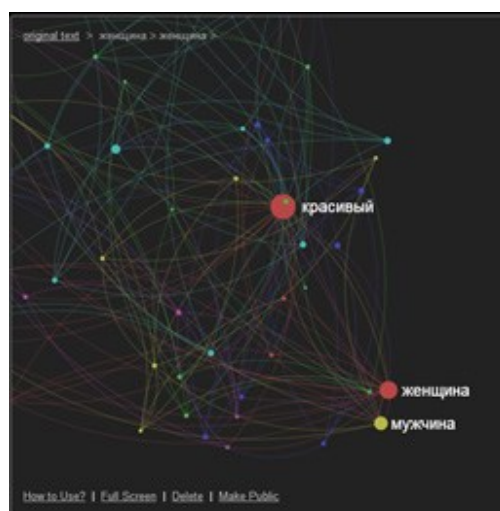


Figure 9. Keyword "beautiful"

If we look at the names of users participating in the discussion one can notice that the number of male users is quite numerous. Female users are also present in the discussion, but not in such large numbers. Another interesting moment is that when talking about being beautiful, even female users mean exactly female beauty, not male one. As it shows the statistics of collocations being “beautiful” is good feature to have for any woman: “it is interesting for women to be well dressed up and look beautiful” (женщине интересно быть хорошо одетой и выглядеть красиво); “smart man will never marry a stupid woman, not rather a one that is not beautiful” (умный мужчина никогда не женится на глупой женщине, и на некрасивой тоже); “if you want to meet smart and beautiful woman you will meet her for sure” (если вы хотите встретить умную и красивую женщину, то вы её обязательно встретите).

The comments mentioned above demonstrate positive sentiment towards beautiful woman. Many women want to be like this, many men want to meet women like this. But at the same time there is a certain element of negative attitude, especially in the discussion about pitfalls that a man can come across living with a beautiful woman. This mood shows especially from male users: “I spend time with beautiful women too, but short time” (я провожу время с красивыми женщинами, но короткое время); “when everyone cares about your beautiful girlfriend – it is hard” (Когда всем есть дело до красивой девушки-это очень сложно); “weak men are afraid of beautiful women” (слабые мужчины боятся красивых женщин); “beautiful women can cost you more money than ordinary ones but there is no big difference” (красивые женщины обходятся дороже, но большой разницы нет); “all beautiful women are happy, but they are bitches” (все красивые счастливы, но – суки); “Beautiful, well – groomed woman is especially good in making you a cuckold” (Красивая, ухоженная женщина выдается ...вместе с рогами).

These comments presented above obviously demonstrate stereotypes regarding behavior of a beautiful woman. Many of them lack any proof, they are just expression of feelings regarding a particular issue. In all the 48 excerpts with the word “beautiful” there are very few phrases that have to deal with life experience of a person, his mental outlook or reference to some scientific source. Such collocation as “in my experience” (по моему опыту) has a very low frequency – 3 times. Collocation “as statistics show” (как показывает статистика) are not mentioned at all. Summarizing the analysis of the keyword “beautiful” and its collocations, I can conclude that portion of traditionalism in evaluating female beauty is relatively large in the comments. It is expressed in the high demand for a woman to be beautiful and well – groomed.

### **Women hunting after men in the animal farm**

The adjective “good” (хороший 203) is mostly collocated with male nouns – husband, man, guy.

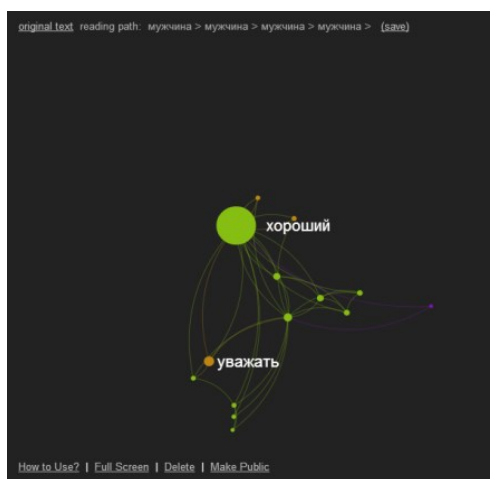


Figure 10. Keyword "good"

In the corpus of comments the adjective “good” is pronounced mostly in discussion about men, and just occasionally about women. In the conversation about relationship between wife and husband many users express themselves in so – called animal language. The reason why I call

such language type an animal language is because the words like a dog, or cattle are featured in this discussion. Even though the discussion is about a human being, a man in particular.

Based on the comments it is possible to form an image of a man with the following characteristics: an animal who has only instincts and no feelings or commitments, without brain and will: “cynical cattle living on transistors and beer; a man is milk-and-water cattle free of brains” (циничная скотина на транзисторах и пиве). When talking about good husbands the language becomes less brutal, but still with the traits of animal and marketing language: “good dogs are taken very fast while they are still puppies” (Хороших собак ещё щенками разбирают); “in good hands any man forgets about his “mentality” (в хороших руках любой мужчина забывает слово "менталитет"); “if a man is good, take him immediately” (если мужчина хороший, берите его немедленно). Other collocations of the word husband include “breadwinner” (кормилец), “cow” (скот), “protector” (защитник), “dog” (собака), “cattle” (скотина), “stallion” (жеребец), “to feed cattle well” (хорошо кормить скотину).

The general feature of all the comments is their informal language, very often profanity. The biggest part of the comments with keyword “good” shows a man in a very passive way, “he is taken” (его берут) and affected somehow by a woman. While a woman in this discussion is described as an active participant of relationship whose word is the last.

### **Women more concerned about family issues compared to men**

Collocations of “normal”: married (замужем, женат), problem (проблема), man (мужчина), woman (женщина)

The adjective “normal” (нормальный) has a frequency of 168 which is relatively high. After researching the collocations of this adjective, I found out that discussion on both female and male topics take place. Moreover, they are mostly equally presented. Nouns of male gender

outweigh a little bit female ones. The collocations of “normal” in male gender count 17 while female gender 14.

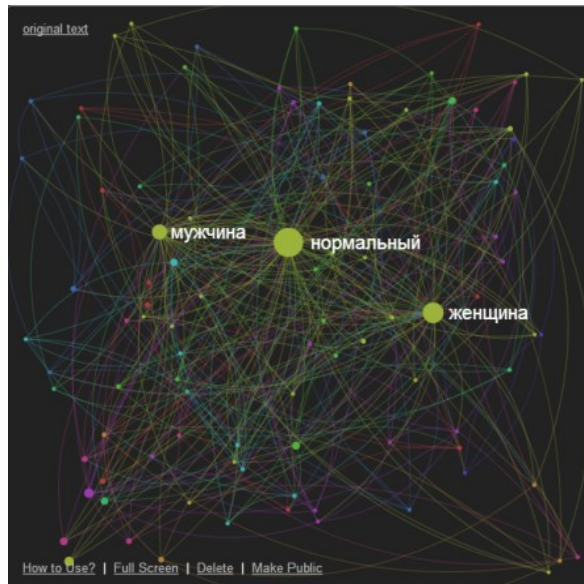


Figure 11. Keyword "normal"

The adverb “normally” (нормально) has a frequency of 83 and is not related to the adjective “normal” (нормальный). Users comment on their understanding of a modern normal woman and man. Many comments sound in the form of complaint from women about men and the opposite way. After reading the comments and collocations of “normal,” one can think that it is very hard nowadays to find a partner both for women and men. Many users say that all the “normal” men and women already have partners, and those who are still around are not a good match at all: “Alas, there are fewer and fewer normal women and more and more screwed – up women crazy about money and career” (Увы, такая жизнь настала всё меньше и меньше нормальных женщин но всё больше закомплексованных . повёрнутых на имидже деньгах и карьере фирмен); “one is a drinker, another one – drug addict, someone is a gay, and maybe just one is a normal guy but already married” (кто пьяница, кто наркоман , кто голубой, а тот кто нормальный - того уже оженили).



Regret from the side of women is more prominent. It seems that there are fewer normal men compared to normal women, as there are more complaints from the side of women rather than men. Generally speaking, comments with the key word “normal” are full of negative sentiment. Furthermore, they create an impression that there are no normal young people in the modern Russian society at all. Some of the comments pronounce this idea very explicitly: “It turns out that there are almost no normal guys and girls” (Вот и получается что нормальных парней и девушек почти нет). Then the question appears “What is normality?” To answer this question I analyzed collocations of the phrase “it is normal” (это нормально). And can the comments and blogs help to understand meaning the users put in this word? To answer this question I turned to the adjective “normal” in the corpus of blogs.

Statistics of the word “normal” in the corpus of blogs shows frequency of 93 times. This means that discourse about “normality” in the blogs has less popularity than in the corpus of comments. Additionally the program code counted the phrase “it is normal” (нормально) to track sentiment regarding what people consider to be normal. The phrase “it is normal” (нормально) is met almost in the half of those 93 occurrences. The code revealed the following:

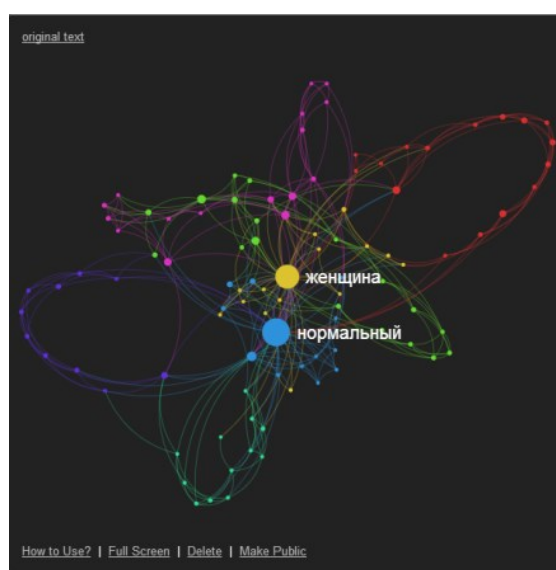


Figure 12. Keyword "normal" and "it is normal"

Collocations for the phrase it is normal: "to love and be loved"; "romantics in the beginning of love relationship"; "women to love kids more"; "women to dream about marriage".

The phrases above are the most frequently used collocations for "it is normal". The notions contained in these phrases fit very well into traditional conservative norm of Russian culture. So users in their comments about normality refer to traditionalism in the relationship between woman and man. For both females and males the "normal" person is the one who is inclined to the behavior described in the collocations of "it is normal." The sentiment analysis of the "normal" in both corpora shows that the number of young people who want to create family is relatively large. But at the same time, there is a certain number of those who do not seek traditionalism and do not put family values on top. Although both female and male users were complaining about lack of normal people for creation of family, there are definitely more female voices talking about this problem.

### **Negative attitude towards both modern Russian feminists and patriarchal Russian women**

Discussion of patriarchal and feminist nature and type of female behavior is rolling out in the collocations of the word "female" in the corpus of comments. The frequency of the word is 164.



Figure 13. Keyword "female"

Collocations of “female”: chauvinism, degradation, ass, shame

The code revealed that most collocations of this adjective have negative meaning: “female chauvinism” (женский шовинизм), “female degradation” (женская деградация), “female ass” (женский зад), “shame on women” (стыдно за женщин). Even though the conversation is about female nature, for instance female beauty, the language of users sometimes is very cynical and harsh. The frequent use of the words like “function”, “calculate”, “money” are probably mentioned by the users to stress on mercenary and material interests of modern women: “Barbie style female beauty is created and exposed as a function to sell oneself out on the sex market with the goal to earn money” (Барби'туратная женская красота создаётся и прокачивается как функция собственной востребованности на сексуальном рынке с целью заработать средства); “sex became a female currency and value that is sold out” (А вот секс стал женской валютой, ценностью, которая продается). The idea of emancipation and woman’s strive for independence can be heard in many comments: “A modern girl is a result of female emancipation, a wish to be free and independent from men” (Современная девушка - это плод

женской эмансипации, это желание быть свободной и независимой от мужчин). But at the same time it is possible to find complaints about “unhappy female fate” (несчастливая женская доля), “women taking on too much stuff on their own” (женщины тащат на себе слишком много).

The collocations of “female” give an idea of a woman with mixed personality who is free and emancipated but at the same time taking on too much and unhappy. Some of the comments suggest this idea explicitly: “What you described is not a feminist but quite a patriarchal woman who is overloaded and comforts herself with an idea of “female lot” (Вы описали не феминистку, а вполне себе патриархальную женщину, которая тянет за двоих и утешается сказочкой о женской доле).

Summarizing collocations of “female” I can conclude that there are two types of female personalities that are being discussed in the corpus of comments. The first image of a woman is oriented towards traditional values of patriarchal society and the second is a free, emancipated and independent woman. The program revealed that both images have negative sentiment expressed about it.

### **Mix of patriarchal and liberal in consideration of female vs. male way of doing**

In terms of “male” word, its frequency in the corpus of comments is 120. Although the keyword is “male,” its most frequent collocations are “female”. One can notice that discussion in this part of comments is based on the opposition of male and female. The graph below demonstrates very explicitly how both these words are equally distributed in the context.

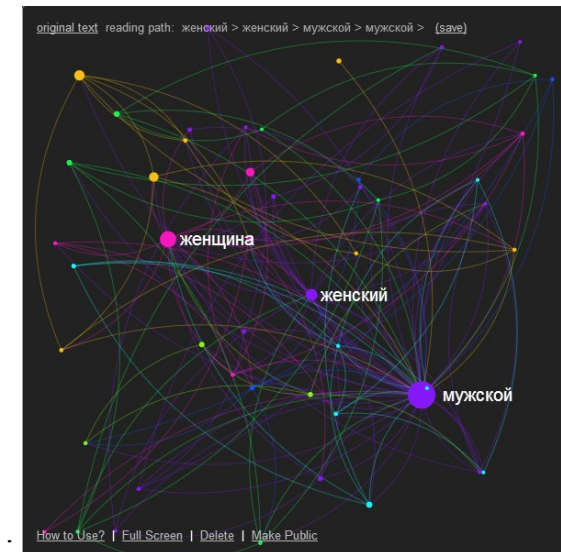


Figure 14. Keyword "male"

**Male** collocations: doesn't owe anything to anyone, wish to rule and humiliate

**Female** collocations: routine everyday work, stay home, can't think globally

The big violet node is the keyword “male”, as you can see it is connected with other keywords such as “female”, “woman”. All other words just participate in the context created by these key words. Other most frequent collocations are “chauvinism” (шовинизм), “alcoholism” (алкоголизм), “psychology” (психология), “roles” (роли), “duties” (обязанности) and “tasks” (задания).

The discussions in this part of comments sometimes have an extreme character because they deal with such vulnerable aspects of social life as human rights, family relations and alcoholic problems. In many instances the users talk about “share of power” on the global socio – political and micro every day levels of life. The vocabulary of comments in this part of corpus reflects the diverse nature of language. If you look at the smaller dots on the graph you will find adjectives and nouns that have social, political meaning as well as routine every day meaning or plain language: “society” (общество), “oath” (присяга), “poop” (какать), “bullshit” (ерунда),

“patriarchal” (патриархальный), “social values” (социальные ценности). Then users comment on how women and men fulfill their responsibilities in life attributing female features to men and male features to women: “Women consider it right to look, think and act in the same female way like Sergey Zverev in a male way” (считают себя в праве выглядеть, думать и поступать так по-женски, как Сергей Зверев по-мужски). The joke here is that Sergey Zverev is a fashion designer notorious for promoting extremely womanish fashion and behavior being a man himself. Moreover, his last name has a word for ‘animal.’ In the sentence with Sergey Zverev we see the example of negative semantic prosody with a degree of irony and joke.

This part of corpus shows that there is no certain agreement on gender roles. The very fact that users debate which way of thinking and doing is better – female or male – shows that the concept of gender equality hasn’t caught on in the Russian society. Many collocations of “female” with “routine everyday work” (рутинная и ежедневная работа) or “stay home” (оставаться дома), “can’t think globally” (не может мыслить глобально) show traditionalism in evaluation of female social role. At the same time, collocations of the word “man” for example in the case of “doesn’t owe anything to anyone” (никому ничего не должен) demonstrate lack of traditionalism and focus on more liberal values in evaluation of male social role in modern Russian society.

### **Discourse of responsibility has more traditionalism in the corpus of blogs**

Must (должен): 171 in corpus of comments, and in the corpus of blogs – 190

After running blogs with “must” in Texttexture the program revealed the following visualization:

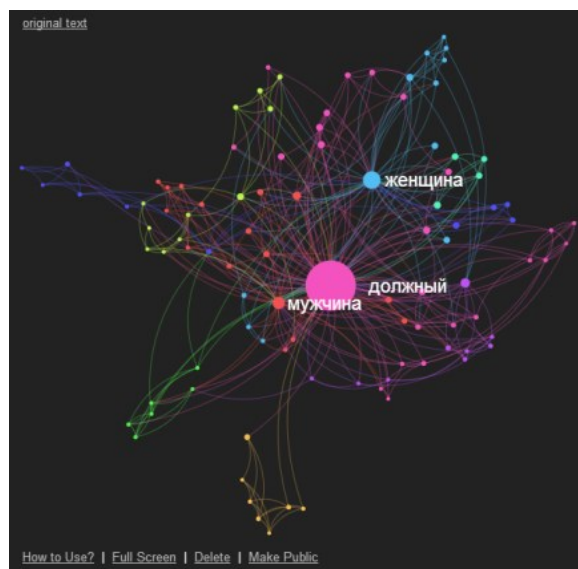


Figure 15. Keyword "must"

As you can see on the graph the verb “must” (должен) (big pink node) has more frequent collocations with the word “woman”, rather than “man”. The scarlet node “man” is smaller and has fewer connections in meaning with “must”. The colors help to interpret the connections of meaning. “Man” has more green and blue line connections. “Green” connects the node “man” with such words as “initiative” (инициатива), “show” (показывать), “do” (делать), “earn” (зарабатывать). Blue line connects the node “woman” with words “problem” (проблема), “wash” (убираться), “must” (должна), “everyday” (каждый день). The pink line builds in single connection such key words in context as “wife” (жена), “to liberate” (освободить), “to force” (заставлять), “to obey” (слушаться), “husband” (муж).

If we look at the collocations “woman must” (женщина должна) one can conclude that most of collocations reflect traditional patriarchal point of view regarding main purpose and duty in female life: “domestic goddess” (хранительница домашнего очага), “deliver babies” (рожать детей), “be very beautiful” (быть очень красивой), unquestionably obey her husband (беспрекословно подчиняться своему мужу). These descriptions fall under the category of

traditional Russian woman. Collocations of “man must” (мужчина должен) consist of the following words: “provide for family” (обеспечивать семью), “be angry and stinking” (свиреп и вонюч), “must serve in the arms” (должен служить в армии). These words also describe a man in the traditional patriarchal fashion.

In the corpus of comments the picture looks a little bit different. Both collocations “man” and “woman” are equally discussed.

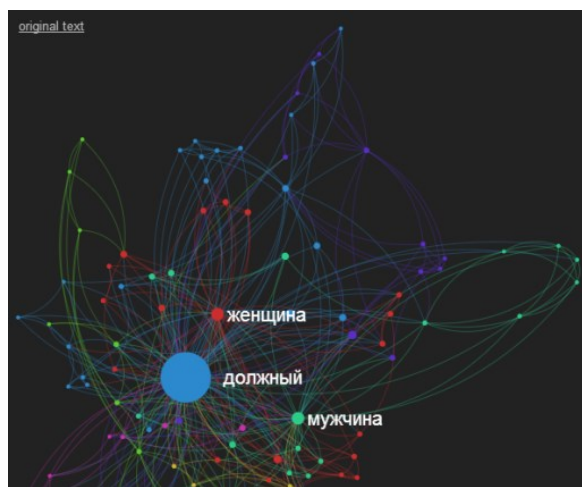


Figure 16. Keyword "must" with "woman" and "man" collocations

The graph displays the red dot “woman” the same size as the green dot “man.” The green line connects “man” with collocations “real” (настоящий), “selfish” (эгоистичный), “loser” (неудачник), “treat” (относиться), “sublime” (возвышенны, тонкий), “normal” (нормальный), “strong” (сильный). The red line of “woman” topic has the following collocations: “to be on a high level” (быть на высоком уровне), “house” (дом), “to cook” (готовить), “to clean” (убирать), “rely on own resources” (полагаться на свои собственные силы). In both corpora the discussion of “must” with “woman” and “man” is used in very informal style. The sentiment nature is mixed. Neither comments’ corpus nor blogs’ corpus has a particular sentiment expressed about a woman or a man.



If you see the full list of collocations one can understand that both positive and negative sentiments are used in the discussion about duty and responsibility in female and male cases. In addition to mixed sentiment, the collocations with “must”, “woman” and “man” as key words in context show that perception of female as well as male duties in society has changed with the shift from traditional and patriarchal to more liberal. Example of collocations of the keyword “woman” – “earn money” (зарабатывать деньги), “be strong” (быть сильной), “support her own life” (обеспечивать себя самой), “independent” (независимая); “man – does not owe anything” (ничего не должен), “not working” (не работающий), “privileged” (привилегированный), “free choice” (свободный выбор). The collocations in the corpus of comments definitely have less traditional ideas than the corpus of blog articles. The same feature for both corpora is that none of them suggest typical traditionalism in the discussion of responsibility. Both corpora combine traditional and modern tendencies.

Summarizing the results of sentiment analysis of the corpora, first I would like to mention that nature of my corpora is much unstructured. This multiplicity makes text analysis in general more complicated and interesting at the same time. Diversity of opinions expressed in the texts of corpora adds trustworthiness to the research.

Text analysis enabled me to grasp the patterns of text through the methods of key words, collocations and visualization. In addition to the mechanical statistical side I applied some linguistic analysis as well. But as I was working with corpus it was also important to see the ‘corpus stylistics’ when the units of text are examined separately as well as in the whole structure. Quantitative stylistics can “take those units of language that are so frequent that we hardly notice them, and show how powerfully they contribute to the construction of meaning” (Moretti 2009). Visualization of my data corpora definitely contributes to the meaning and its interpretation. In my research I rely on the text visualization provided by free online program

Texttexture. The possibilities of this program are relatively modest and do not largely contribute to the aesthetics of my data. But at the same time, Texttexture works very well fulfilling the basic functions: identifying key words, showing context out of it and connecting context in the network.

The next chapter (Conclusion) is going to analyze and summarize the hybrid nature of traditionalism in modern Russian women and men based on the findings of Analysis chapter.

## CONCLUSION

The findings of this research suggest that blogs represent a very complicated phenomenon expressed through the means of language and electronic text. With regard to the characteristics of blogs it is possible to make the following conclusion. Immediacy of information sharing granted to all the bloggers of any website provides equal opportunity to participate in the discussion. The active discussions on the electronic pages demonstrate that blogs have a very reproductive nature as they lead to other forms of communication. Thanks to new technologies, the users of websites have an opportunity not only to read the text but also to create their own text, in the form of comments or actual blogs. The audience acquires a more active role in the reading process. In some sense the digital environment of the blogosphere exists under very liberal rules, if we omit the moderation requirements which are different depending on a website and country of its production.

The process of text analysis showed that in their active discussions of a particular topic bloggers express the ideas, exchange opinions, and form the debate. Generally speaking readers participate in the creation of digital content being active cultural producers. This observation supports the idea that Web – based media have given the ordinary readers a chance not only to be consumers, but also producers (Gross 2009) changing the topdown tyranny of the media (Gross 67). It is possible to address bloggers as amateur writers or “producers” (Bird 2011) as their reaction to the information does not always meet certain criteria of professionalism but at the same time stands above the average level.

To connect the finding about productivity of blogger I would like to refer to the study made by Repici (2008). Repici in the analysis of Google and YouTube found out that the number of people who posted creative videos was very small, while the number of those who commented critically was even less. The vast majority of people just watched the videos, and some of them

gave incomprehensible comments that showed “no indication of any focused thought process”, but were banal, profane and remarkably uninteresting (Repici 13). Although my research was not specifically designed to evaluate “produsage” potential of bloggers, it is worth mentioning that bloggers turned out to be very creative in their expression. Diversity of the language and opinions in my corpora echo the idea advanced by Van Dijck (2009), the scholar who views the users as an invaluable data provider that helps to collect metadata on the social behavior. Mostly those are marketers who track online activity of the users to identify the appropriate preferences and tastes of people to capitalize on. My research shows that besides commercial use, one can track people’s sentiment regarding some social or cultural norms for academic purposes.

The technological challenge of my research was in the effort to get as much data as possible for better representativeness and, at the same time, to sift the grain from the chaff in this large amount of data. From the technological side it is difficult to do text analysis with texts in the Russian language. Automated processing works very well in identifying certain roots of the word. But due to its complicated grammar, semantics, and morphology the Russian language requires a lot of manual work in addition to the technological part. For instance, in the process of calculating the keyword “modern” (современный) I came across the problem of matching genders. The programming code revealed that this word was used 43 times in the male gender in the corpus of comments. But the manual search of this keyword showed that this statistics of 43 also includes little number of female gendered words. They say to err is human. The same can be said about technology. The programming code that I wrote was customized specifically for my research and basically it fulfilled all its major tasks very well. During the tokenization the code assigned a particular token to a particular word and its form using the Russian language signs of NLTK library. For example the keyword “modern”:

современной Afpfsgf      современный

According to the description here <http://corpus.leeds.ac.uk/mocky/ru-table.tab> the sign “Afpfsgf” has the following characteristics – adjective, genitive, full-art positive, feminine, singular and qualificative. But statistics given by code indicates the sign Afpfsgf for the word “modern” in both feminine and masculine forms. Probably it is the same case with other languages that have different endings to indicate gender and case, for example in German, French or Arabic. Despite some technological imperfection generally the programming code really helped and calculated a lot of words that would be impossible to do manually. Thanks to the code it was possible to get an insight into keywords and collocations. Moreover the statistics of positive and negative adjectives provided by the code assisted in making generalizations regarding the research question discussed further on in this chapter.

Comparison between the two corpora (blog posts and comment posts) were made using text analysis methods – keywords, collocations as well as sentiment analysis. Additionally semantic prosody technique was applied to investigate how different word choices interact with each other in a close proximity. The corpus as a whole represents environment within which different meanings coexist in the form of separate items. To analyze these items comprehensively and make generalizations out of it, the item and environment were studied in the systematic connection:

“The computer will start, by default, by assuming that the unit of meaning is the word, defined as a string of characters between two spaces. The systematic interconnections between the item and its environment which become apparent on the vertical axis of the concordance, and the realization on the part of the linguist that the formal co – textual features surrounding a word determine its meaning and its function in a specific discourse, cast serious doubts on the idea that item and environment can be clearly separated” (Tognini-Bonelli 101).

The findings of my research are consistent with this concept of interconnection between the item of meaning and its environment. Statistics of keywords and collocations definitely bring value to my research but for better insight in the overall context, so – called environment of this meaning, solely technology is not enough. The technique of semantic prosody together with intuition played additional very important role in the research. The most obvious example of semantic prosody in action was keyword “beautiful”. The basic meaning of the word denotes physical attractiveness but the case when the same word acquires different meaning also takes place: “monotonously beautiful”, “plastic beauty”. In these examples we see how original meaning of “beauty” and “beautiful” achieves different negative coloring when connected with such words as “monotonously” and “plastic.” As it was already mentioned in the Analysis chapter here we saw example of delexicalisation.

To summarize the key findings of this research I would like to indicate the general characteristics about both corpora. In case of blog posts’ corpus it turned out to be more focused on mental features of a person. Positive adjectives with the highest frequency here are the following: “beautiful” (красивый), “needed” (нужный), “strong” (сильный), “emotional” (эмоциональный), “confident” (уверенный). As to the negative adjectives they are “extreme” (крайний), “material” (материальный), “cold” (холодный), “drunken” (пьяный), “evil” (злой). Most of these adjectives fit into traditional description of a Russian man as well as woman. Also it is important to note that the most frequently used adjective is “beautiful” and it was used in regards to female description. This means that concept of beauty has great importance in representation of femininity rather than masculinity. However one should keep in mind that Russian word “beautiful” is the same for a woman and man whilst in the English language “beautiful” refers to a woman and “handsome” to a man. This is the difficulty of research with translated corpus. In my approach to these linguistic patterns I kept in mind the difference

between unit of meaning and unit of translation. Units of meaning are defined contextually while units of translation – structurally (Tognini-Bonelli 133). In terms of stylistics the corpus of blog posts has more formal and restrained language.

The corpus of comments is stylistically and thematically different from the corpus of blog posts. The language that audience uses to express their ideas is more informal, in the style of slang. Another feature is the degree of freedom in the expression of ideas. For instance one of the users allows himself to pronounce the following comment: “Putinism and prostitution are synonyms” (Путинизм и проституция это же синонимы). The interesting point is that this comment passed somehow through moderation. The blog posts do not contain such frivolous claims like in the comment section. Blog authors are more polite in their expression. The most frequent positive adjectives in the corpus of comments are “needed”, “beautiful”, “simple”, “confident” and “independent”. The negative adjectives are “cynical”, “wild”, “extreme”, “stupid” and “dull”. The large number of comments indicate that response rate of the users is very high.

### **Traditionalism in the discussion of masculinity and femininity**

As a result of text analysis in both corpora, the following topics of discussion were identified in the corpora: physical beauty, needs, strength of character, search for a partner, family concerns, active discussion about feminists and patriarchal women, debate over distribution of rights and power between women and men, question of responsibility. Analyzing these themes it was possible to find the degree of traditionalism in masculinity and femininity.

### **Theme of beauty**

The analysis of collocations of the word “beautiful” suggests an increased interest in the physical appearance of the woman. Although the mental, inner world is discussed as well it is less prioritized. Surprisingly the importance of being handsome was also identified in the theme

of masculinity. This finding does not accord with the image of traditional Russian man described in the earlier chapters. It is interesting to note that corpus of blogs had more discussion about the value of female beauty while corpus of comments talked more about male being handsome.

### **Theme of being needed by someone**

The collocations of “need” demonstrate very traditional ideas in the relationship between a man and a woman. Male needs are mentioned more frequently in comparison to female ones. Especially in the corpus of comments, even female users discuss more male needs and admit that women very frequently put needs of men on top of their own.

### **Theme of strength**

In the discussion of mental strength there is no explicit sign of traditionalism. The image of strong woman oriented on success in career is appreciated. In regards to man, he is depicted as a passive “animal” who does not decide anything. Although traditional Russian culture assigns much strength to the man both mentally and physically, the corpus analysis did not reveal many repetitions of “strong” with male collocations. Another issue contrary to the norms of traditionalism in femininity is lack of interest in family.

### **What is normality within gender discussions?**

It was found that both women and men think of such positive notions as romance, love and relationship. Surprisingly, discussions about such pleasant spheres of life showed a high degree of negative sentiment regarding the image of both masculinity and femininity. In both corpora there were discussions about “normality” that people seek in the opposite sex. What is surprising is that analysis of collocations of “normal” and collocations of “strong” showed different perception of normality. In the discussion of female strength the image of business lady was favored while in the discussion about “normal” woman one can see the statements positioning career oriented women in a negative way “unfortunately there are less normal women



and more femmes crazy about money and career” (к сожалению всё меньше нормальных женщин и всё больше повёрнутых на имидже деньгах и карьере фирмен). Negative sentiment was also found in the discussion about “normal” man. It was expressed in the frequent use of such words as “drunkard”, “drug addict” (пьяница, наркоман). Generally speaking, career – oriented ladies as well as drunkards were not accepted as normality in the sentences with “normal”. This observation of the word “normal” leads to the conclusion that the image of business lady has not yet fitted well into public consciousness of Russian society. Drawn on this finding it is possible to conclude that results of text analysis echo the hypothesis of this research “Traditionalism in the evaluation of modern femininity remains higher than in masculinity.” But on the other hand, the results also point out traditional values in representation of masculinity, although in less degree.

### **Mixed values of traditionalism and liberalism**

The findings from the analysis of words “female” (женский) and “male” (мужской) suggest a mixed representation of patriarchal and liberal values in evaluation of femininity. As to discussion of masculinity the collocations of “male” indicate a greater degree of liberal attitude. The most frequent collocations of “male doesn’t owe anything to anyone”, “wish to rule and humiliate”, “against discrimination of a man and duty of service” indicate relatively liberal position in the consideration of responsibility from the side of a man. The collocations with “female” (routine everyday work, stay home, can’t think globally) indicate signs of patriarchal look at responsibility of a woman. The interesting observation is the combination of liberal as well as traditional norms in the discussions.

Despite numerous collocations of “female” that consider the woman from very patriarchal point of view, there were also collocations describing woman from the liberal perspective: “modern girl is a result of female emancipation”, “women should work as this way they take care

of themselves better and stir male interest easier” (Современная девушка - это плод женской эмансипации; противник мужской дискриминации и противник воинской обязанности). The collocations describing a man within traditional norms also took place – “earn money”, “protect country” (зарабатывать деньги, защищать страну). Some portion of negative sentiment was found in both discussions about masculinity and femininity – “female chauvinism”, “male discrimination”.

As one can see, there are many discussions where the man is positioned as a victim. To summarize the discussion about responsibility, I would like to mention the collocations of the keywords “man” and “woman” generally in both corpora. Woman – “earn money”, “be strong”, “support her own life”, “independent”; man – “does not owe anything to anyone”, “not working”, “privileged”, “free choice”. In general, the collocations containing traditional ideas had higher frequency in the corpus of comments rather than corpus of blogs. The analysis of discussions with the keyword “strong”, “normal”, “female/male” and “must” are consistent with the ideas advanced by Zdravomislova and Kigay about mixed representation of both genders with biased attitude and stereotype thinking about woman.

The results of text analysis of the blogs provide important insights into discussion of gender issues in the modern Russian society. As it was already identified before the language transmits culture and information. In the case of my research the language was expressed in the form of blogs. The methods of text analysis helped to identify key topics of discussion within the research question of masculinity and femininity. The results of sentiment analysis in the corpora revealed some correlation with the concept suggested by Zdravomislova and Kigay about biased attitude towards the woman and mixed representation of both genders. Especially, this correlation was obvious in the discussion of responsibility within private and public life, social realization and human rights. There was a certain degree of traditionalism found in the representation of both

femininity and masculinity in the corpora. But the most interesting result that emerges from the data analysis is the representation of femininity as well as masculinity that combines both traditional and liberal values. The analysis of collocations found a certain level of traditionalism in female discussions, but at the same time, features describing a woman from the liberal point of view are present too.

The situation in the discussion of masculinity also contained the mix of patriarchal and liberal concepts. The investigation of masculinity themes found a degree of traditionalism but in less proportion than in femininity. It is worth noting that the orientation on traditional values was higher in the evaluation of femininity than masculinity. For instance one can meet the phrase “female purpose” (женское предназначение) more often than “male purpose” (мужское предназначение). Moreover the collocations of “female” word are more frequently used in the verb form and have a meaning of action – “to do”, “to wash”, “to fulfill” etc. If you look at the words collocated with “male” they mostly have a descriptive character like “normal” or abstract meaning, for instance the verb “wish”.

The results of this research provide evidence for the concepts advanced by Zdravomislova and Kigay. It is possible to connect the result of text analysis with the social and cultural life of Russia. As they say, internet is continuation of reality, and probably these findings somehow mirror the situation in the modern Russian society. The traditions are changing, and modern women and men do not represent the same kind of people who lived even 20 – 30 years ago in Russia. The active discussion in the blogs on both websites demonstrates that people read and blog for information, pleasure and socialization. Seriousness is obviously present in the discussions. But the shade of humor and irony can also be found. Moreover it can be assumed from the variety of ideas being discussed that blogs possess a certain consolidation power. Despite many differences that bloggers can have in terms of their opinion, social class, gender or

even geographical location, the internet is the common space where they share and produce information. Concentration of so many different people on single website makes the internet a really useful tool in exploring the public opinion. My research demonstrates the potential of text analysis in particular for detecting public sentiment on certain cultural and social issues.

This research work contributes to the existing knowledge of text analysis in the Russian language field by offering the programming code that can read, tokenize and tag the Russian language. Moreover, the designed programming code shows statistics of keywords, collocations and the most frequent adjectives in the corpus. The current findings add value to the area of gender studies in the Russian internet discourse, opening many rooms for exploration in the future. Further research should therefore concentrate on the topics of femininity and masculinity separately, or together, in comparison with bigger amount of data. Additionally the research can be extended into other languages for comparative analysis of the phenomenon based on blogs in other countries. Moreover, the hybrid nature of the modern Russian femininity and masculinity can be investigated more thoroughly with the use of text visualization programs available in the Russian academic field. The visualizations of modern corpora can be compared with the texts dating several centuries back. As a result one can have a visual comparison. Another option for the future exploration could be content analysis of the blog websites. For instance, the use of smiley and different visual content uploaded by users was not in the focus of my research although it can also contribute to the analysis of research question.

## BIBLIOGRAPHY

- Baranov, Dmitriy, Olga Baranova and Tatiana Zimina. *Male and Female in traditional Russian Culture. Illustrated Encyclopedia*. Saint – Petersburg: Iskusstvo-SPB Press, 2005. Print.
- Bardashevich, Yana. *Specifics of internet blog genre*. Vyatka State University of Humanities, 2011
- Bazhenova, Elena and Irina Ivanova. *Blog as an internet genre*. Perm State National Research University, 2012
- Blood, Rebecca. *The Weblog Handbook: Practical Advice on Creating and Maintaining Your Blog*. Cambridge: Perseus Publishing, 2002. Print.
- Brod, Harry and Michael Kaufman. *Theorizing Masculinities*. London: SAGE, 1994. Print.
- Bryant, Alyssa N. "Changes in Attitudes toward Women's Roles: Predicting Gender-Role Traditionalism among College Students." *Sex Roles: A Journal of Research* 48, 3-4, (2003): 131-42. Print.
- Crowston, Kevin and Marie Williams. "Reproduced and emergent genres of communication on the World-Wide Web." *Information Society* 16 (2000): 201-215. Syracuse University School of Information Studies. Web.
- Danielsson, Pernilla. "Automatic extraction of meaningful units from corpora: A corpus-driven approach using the word stroke." *International Journal of Corpus Linguistics* 8:1 (2003): 109-127. Print.
- Dillon, Andrew and Barbara Gushrowski. "Genres and the Web - is the home page the first digital genre?" *Journal of the American Society for Information Science*, 51.2, (2000): 202-205. Print.
- Gorny, Eugene. "Russian LiveJournal. The Impact of Cultural Identity on the Development of a Virtual Community." *Control + Shift. Public and Private Usages of the Russian Internet*. Ed. Henrike Schmidt. Katy Teubener and Natalja Konradova. Norderstedt: Books on Demand, 2006. 73-90. Print.
- Goroshko, Olena. "Internet-genre and functioning of language on the Internet: an attempt to reflection." *Speech genres* 6. Saratov: Publishing Center "Science", 2009. 111-127. Print.
- Gurova, Olga. *People and Things in Soviet Culture*. Novosibirsk: Novosibirsk University Press, 2005. Print
- Gusterin, Pavel. "Our Bogatyr Power." *Russian Planet*, Web. 25 August 2012.  
[http://rusplt.ru/articles/life/life\\_782.html](http://rusplt.ru/articles/life/life_782.html)

“Emerging markets drive Facebook user growth.” *eMarketer* analysis, Web. 9 May 2013.  
<http://www.emarketer.com/newsroom/index.php/emarketer-emerging-markets-drive-facebook-user-growth/>

Etling, Bruce and Karina Alexanyan, John Kelly, Robert Faris, John Palfrey, and Urs Gasser. “Public Discourse in the Russian Blogosphere: Mapping RuNet politics and mobilization.” *Berkman Center for Internet and Society*, Harvard University, 2010.

“Female beauty: Ideas imposed from without.” *RusArticles* database. Web. 31 March 2012.  
<http://www.rusarticles.org/health-and-beauty/nutrition/2779-zhenskaya-krasota-predstavleniya-navyazannye-izvne.html>

Firth, John Rupert. “The technique of semantics.” *Transactions of the Philological Society* 34.1 (1935): 36-72. Print.

Fletcher, William. “Corpus Analysis of the World Wide Web.” *Encyclopedia of Applied Linguistic* Ed. Chapelle, Carol. United States Naval Academy, 2011. *Wiley Online Library*. 5 November 2012. Web. [http://webascorpus.org/Corpus\\_Analysis\\_of\\_the\\_World\\_Wide\\_Web.pdf](http://webascorpus.org/Corpus_Analysis_of_the_World_Wide_Web.pdf)

Jacomy, Mathieu, Tommaso Venturini, Sebastien Heymann and Mathieu Bastian. “ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software.” *PLoS ONE* 9(6). Web. 10 June 2014  
<http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0098679>

Johnson, Robert Burke and Larry Christensen. *Educational Research: Quantitative, Qualitative, and Mixed Approaches*. Thousand Oaks, CA: SAGE, 2008. Print.

Herring, Susan, Lois Ann Scheidt, Sabrina Bonus and Elijah Wright. “Bridging the Gap: A Genre Analysis of Weblogs.” *IEEE Computer Society*. Washington, 2004. Print

“Home violence in Russia.” *RIA News*. 2013. Web  
<http://ria.ru/infografika/20130129/920211298.html>

Kravchenko, Irina. “Traditionalism.” *New Philosophical Online Encyclopedia*. Ed. Institute of Philosophy of the Russian Science Academy, 2010. Web.

Leech, Geoffrey. “Adding Linguistic Annotation.” *Developing Linguistic Corpora: a Guide to Good Practice*. Ed. Martin Wynne. Oxford: Oxbow Books, 2005. 17-29. Print

Liu, Alan. “The State of the Digital Humanities: A Report and a Critique.” *Arts and Humanities in Higher Education* 11.1-2, 2012. Web.

*LiveJournal Stats*, 2011. Web <http://www.livejournal.com/stats.bml>

Louw, Bill. “Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies.” *Text and Technology*. Ed. Mona Baker, Gill Francis and Elena Tognini-Bonelli, Philadelphia/Amsterdam: John Benjamins, 1993. Print.

- Malcolm, Kari Torjesen. *Women at the crossroads : a path beyond feminism and traditionalism*. Downers Grove: InterVarsity Press, 1982. Print.
- Marsiglio, William. "Fatherhood: Contemporary theory, research, and social policy." Gainesville: SAGE, 1995. Print.
- McEnery, Tony and Andrew Wilson. *Corpus Linguistics*. Edinburgh: Edinburgh University Press, 1996.
- Moretti, Franco. *Graphs, Maps, Trees: Abstract Models for a Literary History*. London, New York: Verso, 2005.
- Moretti, Franco. "Style, inc. reflections on seven thousand titles (British novels, 1740-1850)." *Critical Inquiry* 36 (2009): 134–58. Print.
- Nardi, Bonnie, Diane J. Schiano and Michelle Gumbrecht. "Blogging as Social Activity, or, Would You Let 900 Million People Read Your Diary?" *Association for Computing Machinery*, New York (2004): 222-231. Print.
- Nekrasov, Nicolay. "Poems by Nicholas Nekrassov." Translated by Juliet Soskice. *Oxford World's Classics*. Oxford: Scholarly Resources, 1974. Print.
- Offen, Karen. "Defining Feminism: A Comparative Historical Approach." *Signs* 14.1 (1988): 119-157. Print.
- Ramsay, Stephen. *Reading Machines: Toward an Algorithmic Criticism (Topics in the Digital Humanities)*. Urbana, Chicago and Springfield: University of Illinois Press, 2011. Print.
- Repici, Michael. "Active audience and cultural dopes as media consumers and producers: a study of audience participation in Internet forums." Unpublished manuscript, University of South Florida, 2008.
- Report "Monthly audience of RuNet is more than 15 million people." *TNS Gallup Media*, 2006, Web. [http://web.archive.org/web/20070930122103re\\_/www.2sun.ru/default.asp?artID=566](http://web.archive.org/web/20070930122103re_/www.2sun.ru/default.asp?artID=566)
- Report "Russia Internet Usage and Marketing Report 2000 – 2010" *Internet World Stats*, 2010, Web. <http://www.internetworldstats.com/euro/ru.htm>
- Sillaste, Galina. *Male gender community: common factors and specifics of development*. Moscow: Publishing house Mask, 2013. Print
- Sinclair, John. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991. Print
- Sinclair, John. *Trust the Text: Language, Corpus and Discourse*. London: Routledge, 2004. Print.

Shlapentokh, Vladimir. "Public and Private Life of the Soviet People: Changing Values in the Post-Stalinist Russia." Oxford: Oxford University Press, 1989. Print.

Soto, Elaine. "Sex-role traditionalism and assertiveness in puerto rican women living in the United States." *Journal of Community Psychology* 11.4 (1983): 346–354. Print.

Stefanowitsch, A and Stefan Gries. "Collostructions: Investigating the Interaction of Words and Constructions." *International Journal of Corpus Linguistics* 8:2, (2003): 209-243. Print.

Stubbs, Michael. *Text and Corpus Analysis. Computer – assisted Studies of Language and Culture*. Oxford: Blackwell Publishers, 1996. Print.

Surveys "Internet in Russia/Russia in Internet." *Public Opinion Foundation* 19, 2007. Web. 23 March 2007. <http://bd.fom.ru/report/map/projects/internet/internet0702/int0702>

Telenkova, Irina. "Beauty standards in ancient Rome and Greece", 2013. Web. <http://ladyincity.ru/myself/articles/466/848/>

"Texttexture non-linear reading machine," Case Study by *Nodus Labs*, Web. <http://noduslabs.com/cases/texttexture-non-linear-reading-machine/>

Tognini-Bonelli, Elena. *Corpus Linguistics at Work*. Volume 6. Amsterdam/Philadelphia: John Benjamins Publishing, 2001. Print.

"Trends in the Russian Blogosphere", *Yandex's Blog Search*, 2009. Web. [http://download.yandex.ru/company/ya\\_blogosphere\\_report\\_eng.pdf](http://download.yandex.ru/company/ya_blogosphere_report_eng.pdf)

Ushakova, Valentina. "Family and marriage orientations of contemporary youth". Sociological research by Saint Petersburg State University, 2008 – 2009.

Van Dijck, José. "Users like you? Theorizing agency in user-generated content." *Media Culture Society*, 2009. Web <http://quigleyadam.files.wordpress.com/2010/12/users-like-you-dijck.pdf>

Weaver, Mary. *New Catholic Women: A Contemporary Challenge to Traditional Religious Authority*. Bloomington: Indiana University Press, 1995. Print.

Zdravomislova, Olga and Natalya Kigay. *Women's theme in media*. Moscow: Association of Journalists' Press, 2002.

Zhalevich, Andrey. "Negative personality features." Web <http://zhalevich.com/mudrost/harakter/940-puti-k-sovershenstvu.html>



Zhalevich, Andrey. "Personality personality features." Web <http://zhalevich.com/mudrost/79-harakter/941-zoloto-harakter.html>