

National Research University Higher School of Economics

as a manuscript

Anna Khomenko

**LINGUISTIC MODELING AS A TECHNIQUE IN AUTHORSHIP
ATTRIBUTION**

Dissertation Summary
for the purpose of obtaining
academic degree Doctor of Philosophy in Philology and Linguistics

Academic Supervisor:
Tatyana Vladimirovna Romanova
Doctor of Science in Philology and Linguistics

Moscow 2021

General characteristics of the study

The research is dedicated to the solution of text attribution problems on the basis of model linguistics postulates. The thesis research involves the solution of an identification «closed class» (Juola 2008) task in a pair-wise comparison of written texts.

The author of any text is a language personality. A language personality phenomenon is studied in terms of different approaches: cognitive linguistics [Karaulov 2010; Romanova 2011], linguistics [Tarnopol'skij, Kozhushko 2019], psycholinguistics [Sedov 1999; Beljanin 2000], linguoculturology [Vorkachev 2001, Vorob'ev 2011], sociolinguistics and anthropology [Vinogradov 1961; Shuy 2005; Vul 2007], forensic linguistics [McMenamin 2002; Galyashina 2003; Coulthard 2004; Ionova, Ogorelkov 2020]. All the approaches dealing with a language personality admit that writer's individual style is a material form of a language personality. The phenomena of a language personality, idiolect [Litvinova 2019] and individual style are complicated and versatile, and therefore, they are to be modeled. Modeling usually includes the models of speech activity, namely models of analysis [Apresjan 1966: 99–107]. The attribution procedure is nothing but the model of a linguistic research «imitating the research procedures which help a linguist to find this or that language phenomenon» [Apresjan 1966: 99]. This is the set of models which is mainly used in the thesis research.

Since [Campbell 1867] and [Lutoslawski 1897] in the West and [Morozov 1915] in Russia, attribution linguistics has always been going through two parallel paths: stylometry and qualitative text analysis. Quantitative methods are now the most widespread [Korobov 2015; Murauer, Tschuggnall, Specht 2018; Muttentaler, Lucas, Amann, 2019; Litvinova, Sboev, Panicheva 2018; Custódio, Paraboni 2018; Gomzin, Laguta, Stroyev 2018; Panicheva, Mirzagitova, Ice 2018; Bachchu, Morgia, Louisiana, 2019], while the qualifying ones are mainly used in judicial author investigations [Rubcova, Ermolaeva, Bezrukova et al. 2007; Abramkina 2019; Kim 2019; Saakov 2018] as a tribute to tradition [Vul 2007; Galyashina, 2003] and in connection with the law [Federal Law of May 31, 2001 N 73-FZ; Order of

December 27, 2012 N 237]. Many of the quantitative approaches are productive and show high level results, but they consider the individual style to be a series of linguistic probabilities, not a product of individual's speech ability and competence. Thus, using only quantitative approaches based on the collection of traditional stylometric features, even in a large number of them [Bhargava, Mehndiratta, Asawa], it is impossible to create a complete model of the author's individual style that adequately reflects an author's language personality. Psycholinguistic, sociological and cognitive approaches to an individual style certainly help to make the model of an author's language personality more complete. There has been a successful at-tempt to use the integration of approaches above (quantitative and qualitative) and vector text representation in the research by [Pimonova, Durandin, Malafeev]. From our point of view, the idea of integration is quite relevant.

The methodology is implemented as follows: 1) automatic extraction of the parameters describing an individual style in terms of pragmatics, thesaurus and lexicon of the author from the text; 2) search for traditional stylometric text data; 3) determination of the «weight» of each parameter; 4) creation of the mathematic models of the texts being compared; 5) comparison of mathematic models in order to determine their similarity/difference (expert analysis of statistical data).

The formalization of the level structure of a language personality in the research is based on the ideas of [Karaulov 2010].

The relevance of the research is as follows: modern society pays much attention to the authentication of written materials. Thus, attribution methodologies are required in philological expertise of famous writers texts (F. Dostoyevsky, M. Sholokhov), forensic authorship examination, analysis of the Internet content with culpable information, solution of scientific tasks. Each of the scientific and practical spheres mentioned above requires complete and detailed methods of attribution producing objective results.

The object of study is the models of linguistic research, i.e. the models of text attribution, quantitative, qualitative and integrative methods and methodology of text attribution. **The subject** of the research is the text attribution parameters

enabling to create a complete and adequate model of the author's individual style as an explicator of a writer's language personality. In the individual style research, formalized analysis models are used, i.e. «an end number of rules to analyze an endless number of sentences» [Apresjan 1966: 107], which come from the principles of semantic syntax [Paducheva 1974], grammar of constructions [Linguistics of constructions (Lingvistika konstruktsiy) 2010] and Russian grammar [Russian grammar (Russkaja grammatika) URL: <http://rusgram.narod.ru/index.html>].

The aim of the thesis research is to develop an integrative text attribution model which combines quantitative and qualitative approaches to text analysis, and a software prototype on its basis.

The hypothesis of the research is the following: by means of a formalized rule set, it is possible to create an integrative attribution model, which is complete enough for solving identification «closed class» problem in a pair-wise comparison of written texts, imitating the original in detail and objective at the same time. This model allows a researcher to examine texts of different lengths and genres.

To achieve the aim of the research, **the following tasks** were set:

- to form a theoretical basis of the research using the analysis in theoretical linguistics, linguistics of models and applied linguistics (including computer and forensic ones);
- to formulate a working definition of the model, to design its architecture; to apply them and create automated software on text attribution;
- to work on a problem of the most frequently used quantitative and qualitative methods of individual style analysis as a representative of the author's language personality; to adapt them for application in the automated software prototype on text attribution;
- to develop a set of formalized rules in order to define linguistic characteristics enabling to identify the author of a written text;
- to set up a text collection of different genres as a material to verify the results of the integrative attribution methodology work;

- to create an authentic effective attribution model aimed at solving an identification task of text attribution;
- to test the model on a collection of texts of different genres; to define the validity of the conclusions following the analysis of the texts through the developed integrative model;
- to work out a strategy to improve the effectiveness of the methodology, if required.

The theoretical novelty of the research includes the development of a linguistic model architecture, mostly multipurpose for the attribution of texts of any length and genre and suitable to be formalized by modern computer instruments. The author also contributed to the problem stated by creating linguistic construction analysis models (formalized search rules) and other identification components of a language personality automatically.

The practical novelty of the research lies in the fact that the functionality of the created attribution model is wider than solving an identification problem of text attribution for closed class with pair-wise comparison. The resource can be used for solving diagnostic attributional problems (gender, age, etc. designation), and working under writers, journalists, etc. language personality description by forensic experts, philologists and cultural critics. Anyway, the model of a language personality will meet the principles of completeness, simplicity, adequacy, technically accurate and objective description of the original, it will be explanatory, communicative and interpretable.

The results of the research were applied in teaching of the following subjects of Fundamental and applied linguistics bachelor educational program, Political Linguistics magistracy educational program, Applied Linguistics and Text Analytics magistracy educational program in the National Research University Higher School of Economics, Nizhny Novgorod: functional and cognitive models in linguistics, linguistic expertise, computer instruments of linguistic research.

The material for the research is the texts of different genres and lengths:

1) A collection of fiction texts including 10 texts by S. Dovlatov and V. Astafiev. The average length of the texts is 20,000 words.

2) A collection of modern Internet fiction texts (fan fiction stories) on «Kniga fanfikov» (<https://ficbook.net/>), divided into authors, including the texts by 3 female authors, 4 male authors; 187 texts in total; the average length of the texts is 1,500-40,000 words.

3) A collection of the texts of an Internet journal «The Village» (<https://www.the-village.ru/>), divided into authors, including the texts by 3 female authors, the texts by 3 male authors; 600 texts in total; the average length of the texts is 500-1,500 words.

4) A collection of entertaining texts on «YaPlakal» portal (<https://www.yaplakal.com/>), divided into authors, including the texts by 3 female authors, the texts by 3 male authors, 424 texts in total. The average length of the texts is 50-100 words.

5) A collection of Russian business e-correspondence, divided into authors, including the texts by 2 female authors, the texts by 2 male authors, 236 texts in total (45-49 letters by one author). The average length of the texts is 50-500 words.

The results of the research were presented in:

2021: 1) International science and practice seminar «Linguistic Diagnostics: Personality Research Methods» (Moscow, The Pushkin State Russian Language Institute); 2) 27th International conference on computer linguistics and intellectual technologies «Dialog 2021» (Moscow).

2020: X International congress on cognitive linguistics «Cognitive-Discourse Paradigm in Linguistics and Related Sciences: Contemporary Problems and Methods of Research» (Yekaterinburg, Ural State Pedagogical University); 2) International scientific conference «Russian in the Contemporary Scientific and Educational Space» (Moscow, The Peoples' Friendship University of Russia); 3) VI International science and practice conference «Language. Law. Society» (Penza, Penza state university); 4) XLIX International scientific philological conference in

the memory of Liudmila Verbitskaya (1936-2019) (Saint-Petersburg, Saint-Petersburg State University).

2019: 1) IX International congress on cognitive linguistics «Integrative Processes in Cognitive Linguistics» (Nizhny Novgorod, Higher School of Economics); 2) Contemporary Theoretical Linguistics and Problems of Forensic Examination (Moscow, The Pushkin State Russian Language Institute); 3) Artificial Intelligence and Natural Language (AINL) (Tartu, University of Tartu); 4) Linguo-Politic Personology: Discourse Turn (Yekaterinburg, Ural State Pedagogical University, The Russian Presidential Academy of National Economy and Public Administration under the President of the Russian Federation, Ural Institute of Management); 5) Language Personality and Effective Communication in the Contemporary World (Minsk, Belarusian State University).

2018: 1) V International science and practice conference «Language. Law. Society» (Penza, Penza State University); 2) Mass Media Politic Communication: Methods and Ways of Linguistic Analysis and Linguistic Expertise (Yekaterinburg, Ural State Pedagogical University);

2014: Problems of a Language World View in Synchrony and Diachrony (Nizhny Novgorod, Minin Nizhny Novgorod State Pedagogical University);

2013: International science and practice conference «Language. Law. Society» (Penza, Penza State University); 2) Artificial Intelligence and Natural Language (AINL) (Saint-Petersburg).

The theses to be put forth during the defense:

1. Modeling of the author's language personality and modeling of language structures (on the basis of the principles of semantic syntax, grammar of constructions, model schemes) used in the text is an effective instrument of text attribution;

2. Combination of qualitative and quantitative methods of attribution analysis allows building quite a complete, imitating the original in detail and, at the same time, objective attribution model.

3. To evaluate from the theoretical point of view the quality of any model,

including the model of attribution or model of an individual style as a representation of a language personality, there are relevant criteria:

- completeness of a model;
- simplicity of a model;
- precision of a model;
- economy of a model
- adequateness of a model;
- unity in its division;
- wholeness of a model;
- structural properties of a model;
- explanatoriness;
- heuristicity of a model (as a special case of explanatoriness);
- communicativeness of a model (in terms of language);
- deductiveness of a model;
- interpretability of a model;
- mathematicity, precision, unambiguity of a model;
- level of model formalization;
- level of technically precise reflection of a modeling object;
- level of real-life reflection of a modeling object;
- level of model subjectivity;
- level of importance of modal characteristics (level of abstraction (idealization) of a model);
- level of efficiency;
- level of functional and practical direction of a model;
- «hypothesis power»;
- esthetic properties of a model (optional).

4. These criteria make it possible to study the validity of the model both in terms of general characteristics, and according to the principle of the model type.

5. Evaluation of models through the classification proposed is relevant not only at the stage of choosing the model but also for examination of an existing and

tested model in order to describe the effectiveness and ineffectiveness of the model components.

6. The semi-automatic text attribution model developed during the research may be successful in solving the tasks of authorship identification for the texts of different genres and lengths.

The structure of the thesis. The thesis consists of Introduction, four chapters, Conclusion, Reference List of 259 items including 74 in English. The results of the research are contained in 20 charts, 17 figures and 14 supplements.

Contents of the Research

Introduction includes the aim and tasks of the research, the object and subject, as well as the relevance, novelty, theoretical and practical relevance; besides, there are the main theses to be put forth during the defense and the information on approbation of the research.

Chapter 1 states the problem of linguistic modeling on the basis of theoretical consideration [Bloomfield 1926], [El'msliev 2005], [Losev 2004], [Apresjan 1966], [Shtoff 1966], [Revzin 1977], [Baranov 2001], [Medvedeva 2010], [Pavlovskaya 2010], [Belousov 2010], [Bryushinkin 2009] etc. It also describes the criteria to determine the type of a linguistic model (Chart 1) and the classification of model properties to evaluate the model effectiveness from the theoretical point of view (Chart 2).

Chart 1. Criteria to determine the model type

Model typification criteria		
№	Criterion (name)	Description (definition)
1.	Models of speech activity	imitate language processes and phenomena
<i>1.1.</i>	<i>Non-semantic</i>	imitate grammatical ability
<i>1.2.</i>	<i>Semantic</i>	imitate the ability to understand and produce meaningful sentences
2.	Analytical	contain an end number of rules to analyze an endless number of sentences
3.	Synthesizing	contain an end number of rules to produce an endless number of sentences
<i>3.1.</i>	<i>Peculiarly synthesizing</i>	from the meaning to the form
<i>3.2.</i>	<i>Generating</i>	from the basic form to the diversity of non-basic forms: include the apparatus to produce an endless number of proper sentences in a definite language and prescribe some structure characteristics to each of them
4.	Research models	imitate researcher's actions on language phenomena finding
<i>4.1.</i>	<i>Decoding models</i>	input information is only a text in its natural environment
<i>4.2.</i>	<i>Experimental models</i>	input information is the text and some rules to modify text information. The text belongs to an artificial environment and

		undergoes different experiments (modifications) conducted by an expert-researcher
4.2.1	<i>Peculiarly experimental</i>	use the text in «input» + a number of proper phrases in a language
4.2.2.	<i>Descriptive</i>	use the text in «input» + a number of proper phrases in a language + a number of semantic invariants
5.	Metamodels	imitate theoretical and experimental evaluation of the finished models of speech activity or a linguistic research (describe linguistic theories)
6.	Models-calculations	the system of permissions
7.	Models-algorithms	the system of commands
8.	Probabilistic	use the relativity theory and mathematical statistics to make decisions on the model (forecasts the behavior of objects)
9.	Structural	deterministic
10.	Mixed	statistics+determinism (mostly wide spread to solve practical tasks)
11.	Componential	include the set of features of an object written in that or those signature
12.	Homogeneous	include the description/analysis of one significant part of a phenomena
13.	Diachronic	analyze the language in development
14.	Synchronic	analyze the language of a certain period
15.	Theoretical	designed to solve theoretical problems of linguistics as such (more often, they are metamodels)
16.	Applied	designed to solve practical tasks of a customer
17.	Graphical	are a visual image of the modeling object; a graphical image is used to build the structure of an object
18.	Symbol	are a set of symbols describing the original object
19.	Computer	представляют собой набор символов определённой формализованной сигнатуры, алфавит, организованный в форме исчисления или алгоритма и реализованный на компьютере are a set of symbols with a definite formalized structure, alphabet, arranged as a calculation or an algorithm and implemented through a computer

Chart 2. Criteria for the linguistic evaluation of the model

Criteria to evaluate the features of models			
№	Criterion (name)	Description (definition)	Evaluation scale
1.	Completeness of a model	ability to reflect all the necessary information	low / moderate / high level
2.	Simplicity of a model	use of relatively a small number of means (signatures, rules) to achieve the goal of the research	low / moderate / high level
3.	Precision of a model	possibility to complete the operations with a formal apparatus supplied by the model	presence / absence
4.	Economy of a model	rational use of energy and time when the model is applied	low / moderate / high level
5.	Adequateness of a model	a property of a maximum similarity to the original object	low / moderate / high level
6.	Unity in its division	the model is always designed to be divided into parts inside the whole (a model procession always consists of subcollections)	possibility / absence of a possibility to divide into subcollections
7.	Wholeness of a model	a model procession is an undividable collection	presence / absence of wholeness (arises from the presence / absence of connection among the

			subcollections of procession elements which builds an undividable structure in its unity)
8.	Structural properties of a model	transfer of the substrate structure of the modeling object to another substrate	a) presence / absence b) successful / unsuccessful choice of the «receiving» substrate c) successful / unsuccessful arrangement of a model structure
9.	Explanatoriness	«explanatory power» of the model; ability to provide information on the reasons for the factors observed and to predict the new ones	presence / absence
10.	Heuristicity of a model (as a special case of explanatoriness)	ability of a model to search for new knowledge on the object	low / moderate / high level
11.	Communicativeness of a model (in terms of language)	any linguistic model is not based on a set of abstract statistic patterns, formulas, functions and numbers, but on the language as a communication instrument	presence / absence
12.	Deductiveness of a model	presence/absence of empiric study of language facts as a basis for modeling; modeling «from the bottom»: the use of means and methods of classical linguistics to observe/examine language facts	a) presence / absence b) low / moderate / high level of operation by peculiarly language, linguistic methods of analysis as a basis for modeling
13.	Interpretability of a model	interpretation of the model is a possibility to use objects of some subject area instead of objects (symbols) of the model	a) presence / absence b) simplicity / complexity of use (substitution)
14.	Mathematicity, precision, unambiguity of a model	correlates with the level of mathematical model formalization	a) complete, whole / incomplete apparatus of model formalization b) successful / unsuccessful operation of this apparatus as a basis of machine completion of the model
15.	Level of model formalization	is a structure described by a conventional signature of the language, or a structure described by means of mathematic, number, formula apparatus	descriptive / mathematical level
16.	Level of technically precise reflection of a modeling object	successful / unsuccessful way to formalize the model, choice of a signature; successful / unsuccessful machine completion (if any)	low / moderate / high level
17.	Level of real-life reflection of a modeling object	determines how large and full the structure of the model reflects the original object	low / moderate / high level
18.	Level of model subjectivity	presence / absence of personal evaluations and judgments of a researcher in the model structure	low / moderate / high level
19.	Level of importance of modal characteristics (level of abstraction (idealization) of a model)	successful / unsuccessful neglecting, elimination of the language facts with no meaning to	low / moderate / high level

		complete a definite task in the model	
20.	Level of efficiency	evaluation of working capacity of the model when solving the task stated	low / moderate / high level
21.	Level of functional and practical direction of a model	compliance of the model with its intended use. The aim of the model creation can be peculiarly linguistic, practical, mathematical etc.	a) complies / does not comply with the intended use b) low / moderate / high level of compliance
22.	«Hypothesis power»	determines whether the model is based on a hypothesis	presence / absence
23.	Esthetic properties of a model (optional)	harmony of the model structure arrangement	presence / absence

Chapter 2 analyzes Russian and foreign methods and methodologies, algorithms and software packages to solve attribution tasks, as well as attribution parameters of different levels used in different approaches. The author evaluates the working capacity of the components listed, describes their advantages and disadvantages.

Chapter 3 is dedicated to the most effective components and attribution parameters of the methods and software packages described in Chapter 2. Besides, it presents their testing results on authentic heterogeneous text materials (official texts, business e-correspondence, short text messages, fiction texts etc.) in order to find the best combination of parameters to solve an identification task of attribution linguistics.

Theoretical literature analysis on the topic raised and empirical research of 10 text blocks with the total size of 166,000 words enabled making a list of parameters of a language personality which are certainly important components of the author's individual style, a material explicator of a language personality of a writer, and, at the same time, can be extracted from the text automatically with a minimal preprocessing.

According to Yu. Karaulov, identification parameters are found at all the three levels of a language personality:

«1) verbal-semantic – the lexicon of a personality in a broad sense including, among others, the fund of their grammar knowledge;

2) linguo-cognitive – the thesaurus of a personality in which the «image of the world» or the system of knowledge about the world are captured;

3) motivational – the level of activity and communication which reflects the pragmatics of a personality: the system of their aims, motives, prescriptions and intentionalities» [Karaulov 2010: 53].

A language personality is understood as the result of its formation in a definite social environment: autobiographical, sociolinguistic and forensic linguistics approaches [Vinogradov 1961; Coulthard 2004; Shuy 2005; Vul 2007].

Chapter 4 describes the software prototype created on the basis of combination of identification parameters and linguistic modeling as a part of a semi-automated integrative model of text attribution. It also contains the results of authentic attribution model testing on the materials of different genres.

To provide computer-assisted extraction, all the formal rules were programmed and integrated into a linguistic resource «XoPOM» (KhoRom): <http://khorom-attribution.ru/#/>¹.

The user module has the following functions: an input requires two texts, A and B; the user can pre-set the genre of the text. This option is subject to the variation of the rules of search for linguistic structures in different discourses. The user can create a model not only on the basis of pre-programmed parameters, but also can choose those which are considered the most relevant for a definite pair of texts. These functions set the software developed apart from other, for example, based on machine learning [Radbil, Markina 2019; Pimonova, Durandin, Malafeev 2020] where all the parameters are pre-set not by a user, but by a developer. Therefore, a real algorithm is not fully automatic, and the final decision is up to the user.

The user may not only check where there are all the implementing parameters in the text included in frequency calculation, but also may exclude those which they consider to be interfering and then re-calculate the data for the final models.

As a result of empirical research, such parameters as parts of speech (a number of content words, correlation of different parts of speech – an index of readability, coefficient of subjectivity etc. in accordance with Golovin's concepts [Golovin

¹ KhoRom is developed by a research team led by A. Khomenko. Yu. Baranova is the technical manager of the project.

1970], the Gunning fog index, the Flesch–Kincaid readability tests with a coefficient for Russian [Solnyshkins, Guryanov, Gafiyatova, Varlamova 2018: 679]), average lengths of words, presence/absence of complex words of semi-concatenation; modal particles, interjections, presence/absence of a modal postfix «-to», preferable intensifiers were programmed for search at the verbal-semantic level. There are 10 standard algorithms and 32 authentic, unique formalized rules in total to extract linguistic structures. The formalized search for the units of this level is conducted in terms of morphologic abstract of the text, i.e. through tags of a part of speech on each word and tags of all grammar categories which are typical of this part of speech.

A verbal-semantic level (level of an idiolect with reference to the concept of [Kristall, Dejvi 1980]) , [Fedotova 2013], [Burov 2017], [Litvinova 2019]) is easily formalized, as it as such possesses «more formal» language features which are a-priori considered stable, though the issue of the stability is not being intentionally researched» [Litvinova 2019: 2].

The description of the thesaurus of a personality: to represent the fragment of this level, there are such parameters chosen as key lexemes, the most frequently used word trigrams and bigrams, explicators of axiological text dominants of «us-them» dichotomy.

This is the most difficult level to formalize. One may automatically create a material explication of the author's thesaurus [Bessmertniy, Nugumanova 2012], but to determine how the lexemes in the thesaurus «are built in an ordered, quite strict hierarchical system to some (indirect) extent reflecting the structure of the world» [Karaulov 2010: 52], to find explicators of peculiarly concepts and categories, the most important for author's world view, axiological dominants defining the philosophy of a writer, is extremely difficult. This level is represented by the least number of parameters (three standard algorithms and one authentic rule for linguistic information extraction in total), peculiarly in terms of attempts not only to formalize some components of a language personality in order to represent it computationally but also turn the final model into an interpretative one.

The pragmaticon of a language personality is formalized due to the following

parameters: parenthetical words and constructions explicating subjective modality; purpose, emphatic constructions, constructions with comparative conjunctions representing the level of the author's competences in written speech and their communicative strategy and tactics; syntax fusions describing, among others, what functional-stylistic texts the author prefers; sentences with detached appositions; complex syntax constructions; comparative subordinate clauses, verbal mononuclear sentences explicating the functional type of the narration; presence/absence and types of address as a contact-forming element. On the whole, there are 11 constructions and 107 authentic «custom» rules for extraction of information from the text.

As a result of the algorithm operation, the value of Pearson correlation coefficient (the coefficient of determination should be evaluated), linear regression, Student's t-test for models of the two text being compared are put out, as well as the values of metrics of each parameter for the two texts, metrics proving or disproving the H_0 hypothesis for that the author of the two texts might be² one person.

It is important to note that this block is not the final step in the methodology. The text statistics also needs interpreting. The software does not submit the results as output knowledge as «The author of the two texts is one person / The authors of the two texts are different people» intentionally, because the developed model assumes that the expert makes the final decision on text attribution based on, among others, statistic data, scoring charts with the results of the research (Chart 3) and their expert experience.

Chart 3. An example of a scoring chart to evaluate the results of the attribution model operation

² Probabilistic nature of the output means that it is the researcher who makes a final decision on the authorship in each separate case in accordance with the methodology developed.

Discourse type	Pearson correlation coefficient	linear regression determination coefficient	Student's t-test (p-value)	The author of the compared texts is probably ³ one person	The authors of the compared texts are probably different people	Commentary
Prose fiction	not below 0,97; usually: 1,00	not below 0,94; usually: 1,00	not below 0,91; rarely above: 0,93 Total: about 0,90	+	-	values of all metrics approximately from 0,90 to 1,00
Prose fiction	1,00	1,00	not below 0,84	+	-	There are situations in the analysis of prose fiction when the Student's t-test (p-value), the most important metrics for large texts, is quite low. In this case, the value of other metrics should equal 1 in order to be able to admit that the author of the text compared is one person
Prose fiction	may reach 0,97	may reach 0,94; usually about 0,84	about 0,50	-	+	If the Student's t-test (p-value) for prose fiction is low (about 0,50), then high values of other metrics, if any, may not be taken into consideration

The results of the model usage were evaluated from two points of view: on the one hand, the models of language personalities received were considered in terms of the theoretical evaluation scale for the models from Chart 2. On the other hand, the models received were also evaluated in terms of the identification task solution. These evaluations served to prove the hypothesis that with the fixed set of formalized

³ Probabilistic nature of the output means that it is the researcher who makes a final decision on the authorship in each separate case in accordance with the methodology developed.

rules one can create an integrative attributive model which is complete enough, imitating the original in detail and, at the same time, objective. This model allows the researcher to solve the identification task of attribution linguistics on the texts of different lengths and genres successfully.

Testing and approbation of the algorithm developed were conducted on the text collections described in **the material of the research**. The following **results** were achieved:

- 1) A collection of fiction texts: accuracy, precision and recall equal 100%, F-score - 1⁴;
- 2) A collection of modern Internet fiction texts (Kniga fanfikov): accuracy – 83%, precision – 67%, recall – 100%, F-score – 0,8;
- 3) A collection of online journalistic texts (The Village): accuracy, precision and recall are equal to 100%, F-score – 1;
- 4) A collection of entertaining texts (YaPlakal): accuracy – 40%, precision – 0, recall – 0, F-score – 0;
- 5) A collection of Russian business e-correspondence: accuracy – 83%, precision – 67%, recall – 100%, F-score – 0,8.

The analysis led to the following **conclusions**:

- 1) Student t-statistics is the most informative for the analysis of fiction texts (both by famous authors and fan fiction);
- 2) Stylometry pool is not informative for modern fan fiction as, according to the experimental data, the values of stylometry parameters are close for all the texts studied;
- 3) The values of correlation coefficients should be equal to 1 in order to define the author of a journalistic text and prove that the H_0 hypothesis is true. The necessity of such a high level is related to the length and specificity of the text material. It should be admitted that t-statistics, the most informative index for fiction discourse, is far less relevant for journalistic discourse. As for gender of the author,

⁴ The values of metrics are indicated here and elsewhere due to the interpretation of statistic data through guidelines and scoring charts created for the analysis.

one should note that «female» journalistic texts are more similar to «female» ones, as well as «male» – to «male» ones; the most significant correlation differences are observed in individual styles of language personalities of different genders;

4) As for short text messages (business e-correspondence, commentaries on the Internet), it is necessary to make a representative selection from the massive of texts not exceeding 500 words. The 100-word limit deduced by S.M. Vul and still relevant for forensic authorship in Russia [Rubtsova, Yermolayeva, Bezrukova and others 2007] should be enlarged for the proposed method because of mathematical statistics usage in parameterized model. To improve the algorithm operation on the material given, additional parameters are now being worked out in order to build the models of an individual style as a representation of a language personality of a writer, they are related to so-called «digital handwriting»:

- Graphic liturative;
- Graphic hybridization;
- Use of archaic affixes for purpose;
- Use of elements of the text in capital letters;
- Emoticons and other graphic symbols expressing the emotionality of speech.

5) The texts of different genres can also be validly studied with this integrative methodology (one can, for example, compare the text of an electronic message with a journalistic article): accuracy – 83%, precision – 67%, recall – 100%, F-score – 0,8.

The output data of an automated algorithm are not the most valuable in the use of the methodology, but the models of an individual style as a representation of language personalities of writers created by its means are. These models are clear and simple, they are easily interpreted by experts, on the one hand, and imitating the original object properly, on the other hand.

The results of the attribution model operation can be compared with the results of other model operation based on machine learning and neuron networks [Khomenko, Baranova, Romanov, Zadvornov 2021].

Therefore, the algorithm operation based on machine learning depends on the length of text data. The attribution model proposed in the research is less sensitive to the text length as the difference is neglected by a relative nature of the frequencies used and the proper parameterization of the analysis model for each text pair separately.

Nevertheless, it should be noted that the comparison of the model proposed in the thesis with other algorithms driven by machine learning and neuron networks may be almost irrelevant as the model has absolutely different basis in comparison with fully automated systems: this model is always to be interpreted by the researcher, and the final decision on text attribution is still made not by algorithms, but by the researcher.

Conclusion contains the final outcomes of the research, recommendations on improvement of the model architecture which attribution analysis is based on according to the methods proposed.

Supplements include empirical material.

The main contents of the thesis are supported by the following references: publications indexed in Scopus:

1) Romanova T. V., Khomenko A. Yu. Funkcionirovanie elementov semanticheskogo polya social'naya znachimost' v russkom i anglijskom yazykah po dannym slovarnyh i korpusnyh istochnikov // Vestnik Sankt-Peterburgskogo universiteta. Yazyk i literatura. 2020. T. 17. № 1. S. 49-73;

publications included in the list of high-level journals prepared at the Higher School of Economics:

2) Khomenko A. YU., Korsakov K. Sudebnyj PR v Rossii: ponyatie i ego sut', yavlenie i ego social'no-gosudarstvennaya znachimost' // Vestnik Voronezhskogo gosudarstvennogo universiteta. Seriya: Pravo. 2019. № 4 . S. 284-300.

3) Khomenko A. YU. Lingvisticheskoe modelirovanie kak instrument vyyavleniya iskazhenij rechevyh navykov avtora pis'mennogo rechevogo proizvedeniya. Opyt prakticheskogo issledovaniya // Voprosy psiholingvistiki. 2018. № 2 (36). S. 209-226;

in other publications:

4) Khomenko A., Baranova Yu., Romanov A., Zadvornov K. Linguistic Modeling as a Basis for Creating Authorship Attribution Software [Lingvisticheskoe modelirovanie kak osnova dlya sozdaniya elektronnoho atribucionnogo resursa]. Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2021” Moscow, June 16–19, 2021, URL: <http://www.dialog-21.ru/media/5315/khomenkoaplusetal048.pdf>.

4) Khomenko A. Yu., Ben'kovich E. R., Gajnutdinova D. I., Gasanova L. R., Kostina A. A., Mazunina Z. O., Nikolaeva A. S., Pimonova E. V. Avtomaticheskaya obrabotka teksta i lingvisticheskoe modelirovanie kak sposoby resheniya problem atribucionnoj lingvistiki // Politicheskaya lingvistika. 2020. № 3(81). S. 215-224.

5) Romanova T. V., Khomenko A. Yu. Lingvisticheskoe modelirovanie kak instrument atribucii teksta // V kn.: Russkij yazyk v sovremennom nauchnom i obrazovatel'nom prostranstve. Izdatel'stvo RUDN, 2020. S. 250-252.

6) Khomenko A. Yu. Vozmozhnost' identifikacii lica po raznorodnomu materialu - ustnoj i pis'mennoj rechi – v usloviyah odnoj ekspertnoj zadachi // V kn.: Lingvopoliticheskaya personologiya: diskursivnyj povorot : materialy Mezhdunar. nauch. konf. (Ekaterinburg, 29—30 noyab. 2019 g.). [b.i.], 2019. S. 214-218.

7) Khomenko A. Yu., Zinina A. M. Global'nyj anglijskij i internet-slang. Vliyanie na yazykovuyu lichnost' sovremennoj molodezhi (na materiale russkogo i kitajskogo yazykov) // V kn.: YAzykovaya lichnost' i effektivnaya kommunikaciya v sovremennom polikul'turnom mire : materialy V Mezhdunar. nauch.-prakt. konf., posvyashch. 20-letiyu osnovaniya kaf. teorii i praktiki perevoda fak. sociokul'tur. kommunikacij Belorus. gos. un-ta, Minsk, 24–25 okt. 2019 g. Mn. : Belorusskij gosudarstvennyj universitet, 2019. S. 152-158.

8) Khomenko A. Yu. Lingvisticheskaya atribucionnaya ekspertiza v otechestvennoj i zarubezhnoj shkolah. perspektivy razvitiya avtorovedcheskih metodik v Rossii // V kn.: Mezhdunarodnaya nauchnaya konferenciya «Sovremennaya teoreticheskaya lingvistika i problemy sudebnoj ekspertizy». M. : Gosudarstvennyj institut russkogo yazyka im. A.S. Pushkina, 2019. S. 536-550.

9) Khomenko A. Yu. Lingvisticheskoe atribucionnoe issledovanie korotkih pis'mennyh tekstov: kachestvennye i kolichestvennye metody // Politicheskaya lingvistika. 2019. № 2 (74). S. 177-187.

10) Khomenko A. Yu. Modelirovanie kognitivnyh struktur na osnove metodik avtorovedcheskogo analiza // V kn.: Kognitivnye issledovaniya yazyka. Vyp. XXXVII: Integrativnye processy v kognitivnoj lingvistike: materialy mezhdunarodnogo kongressa po kognitivnoj lingvistike / Otv. red.: T. V. Romanova. T. XXXVII: Integrativnye processy v kognitivnoj lingvistike: materialy mezhdunarodnogo kongressa po kognitivnoj lingvistike. Dekom, 2019. S. 1069-1074.

11) Khomenko A. Yu. Atribuciya tekstov malogo ob'yoma. Statisticheskie zakonomernosti // V kn.: YAzyk. Pravo. Obshchestvo: sb. st. V Mezhdunar. nauch.-prakt. konf. Penza : PGU, 2018. S. 123-127.

12) Khomenko A. Yu. Opredelenie lingvisticheskikh priznakov podgotovlennoj i nepodgotovlennoj rechi na materiale ustanovlennogo doslovnogo sodержaniya videozapisi pri proizvodstve kompleksnoj sudebno-lingvo-fonoskopicheskoy ekspertizy // V kn.: Sudebno-psihologicheskaya ekspertiza i kompleksnye sudebnye issledovaniya videozapisej: sb. st. Mezhdunar. nauch.-prakt. konf. (g. Moskva, 16 marta 2017 g.). RGUP, 2017. S. 192-198.

13) Khomenko A. Yu., Romanova T. V. Algoritm avtomaticheskoy atribucii pis'mennogo teksta v lingvokriminalistike // V kn.: Sovremennye problemy v oblasti ekonomiki, menedzhmenta, biznes-informatiki, yurisprudencii i social'no-gumanitarnyh nauk: materialy XII-oj nauchno-prakticheskoy konferencii studentov i преподаvatelej NF NIU VSHE / Sost.: E. A. Aslanyan. N. Novgorod : Nacional'nyj issledovatel'skij universitet Vysshaya shkola ekonomiki v Nizhnem Novgorode, 2014. S. 220-223.

14) Khomenko A. Yu. Algoritm dlya avtomaticheskoy identifikacii avtora pis'mennogo rechevogo proizvedeniya v sudebnom avtorovedenii // YUrislingvistika. 2014. № 3. S. 83-93.

15) Khomenko A. Yu. Analiz yazykovoj lichnosti avtora teksta s primeneniem

metodov matematicheskoy statistiki kak sposob ustanovleniya avtorstva teksta. Problemy yazykovoy kartiny mira v sinhronii i diahronii // V kn.: Problemy yazykovoy kartiny mira v sinhronii i diahronii. Sbornik statej po material Vserossijskoj nauchnoj konferencii molodyh uchyonyh Vyp. 12. N. Novgorod : Federal'noe gosudarstvennoe byudzhethoe obrazovatel'noe uchrezhdenie vysshego professional'nogo obrazovaniya "Nizhegorodskij gosudarstvennyj pedagogicheskij universitet imeni Koz'my Minina", 2014. S. 293-297.

16) Khomenko A. Yu. K voprosu ob issledovanii pis'mennogo rechevogo proizvedeniya v ramkah avtorovedcheskoj ekspertizy na predmet ego original'nosti // Politicheskaya lingvistika. 2014. № 4. S. 306-312.

17) Khomenko A. Yu. Aprobaciya metodov matematicheskoy statistiki pri atribucii teksta v ramkah sudebnogo avtorovedeniya // V kn.: Yazyk. Pravo. Obshchestvo: sb.st.Vseros.nauch.-prakt.konf. Penza : Izdatel'stvo PGU, 2013. Gl. 52. S. 356-369.

The dissertation research was supported by the Russian Foundation for Basic Research, project number 19-31-27001"Postgraduates".

Bibliography

Abramkina E. E. Protokol doprosa kak ob»ekt avtorovedcheskoj identifikacionnoj ekspertizy: zadachi, problemy, metodika analiza // Sbornik materialov Mezhdunarodnoj nauchnaya konferenciya «Sovremennaya teoreticheskaya lingvistika i problemy sudebnoj ekspertizy», g. Moskva, 1-2 oktyabrya 2019 g. – M.: Gosudarstvennyj institut russkogo yazyka im. A.S. Pushkina, 2019. Pp. 492-506.

Apresjan Ju.D. Idei i metody sovremennoj strukturnoj lingvistiki. M., 1966. 305 p.

Bacciu A., Morgia M. La, Mei A., Nemmi E. N., Neri V., Stefa J. CrossDomain Authorship Attribution Combining Instance-Based and Profile-Based Features // Notebook for PAN at CLEF 2019, 2019. – URL: http://ceur-ws.org/Vol2380/paper_220.pdf (accessed: 05.07.2020).

Baranov A.N. Vvedenie v prikladnuju lingvistiku. M.: Jeditorial URSS, 2001.360 p.

Beljanin V.P. Osnovy psiholingvisticheskoy diagnos-tiki: modeli mira v literature.

- M.: Trivola, 2000. 247 p.
- Belousov K.I. Model'naja lingvistika i problemy modelirovanija jazykovoj real'nosti. Model'naja lingvistika i problemy modelirovanija jazykovoj real'nosti // Vestnik Orenburgskogo gosudarstven-nogo universiteta. 2010. № 11 (117). Pp. 94-97.
- Bessmertnyj I.A., Nugumanova A.B. Metod avto-matcheskogo postroenija tezausov na osnove statisticheskoj obrabotki tekstov na estestvennom jazyke // Izvestija Tomskogo politehnicheskogo uni-versiteta. 2012. № 5. Pp. 125-130.
- Bhargava M., Mehndiratta P., Asawa K. Stylometric Analysis for Authorship Attribution on Twitter Author's // International Conference on Big Data Analytics, 2013. – URL: <https://www.researchgate.net/publication/299669552> (accessed: 03.05.2021).
- Bloomfield L. A set of postulates for the science of language // Language, 1926. №2 (3). Pp. 153–164.
- Bryushinkin V.N. Kognitivnyj podhod k argumentacii // RACIO.ru. 2009. № 2. Pp. 2-22.
- Burov A.A. K voprosu ob idiostile sovremennogo oratora kak jazykovoj lichnosti // Zapiski Gornogo instituta. T.160. CHast' 1. 2017. Pp.8-9.
- Campbell L. The Sophistries and Polilicus of Plato / L. Campbell. – Oxford : Clarendon, 1867. – 170 p.
- Coulthard M. Author Identification, Idiolect, and Linguistic Uniqueness // Applied Linguistics. 2004. № 24 (4). Pp. 431-447.
- Custódio J. E., Paraboni I. EACH-USP Ensemble Cross-domain Authorship Attribution // Notebook for PAN at CLEF 2018, 2018. – URL: http://ceur-ws.org/Vol-2125/paper_76.pdf (accessed: 05.07.2020).
- El'mslev L. Prolegomeny k teorii yazyka / L. El'mslev ; per. s angl. [V. A. Zvegincev i dr.]. - Moskva : URSS, 2005. 243 p., Translated from.: Hjelmslev, Louis Prolegomena to a theory of language.
- Federal Law of May 31, 2001 N 73-FL «On State Forensic Science Activities in the

- Russian Federa-tion» [Federal'nyj zakon ot 31 maya 2001 g. N 73-FZ «O gosudarstvennoj sudebno-ekspertnoj deyatel'nosti v Rossijskoj Federacii»]. — Russian newspaper [Rossiyskaya Gazeta]. — N 256 of De-cember 31, 2001.
- Fedotova M. A. K voprosu o razgranichenii ponyatij idios-til' i idiolekt yazykovoj lichnosti // Zapiski romano-germanskoj filologii, 2013. №1. Pp. 220-226.
- Galyashina E.I. Osnovy sudebnogo rechevedeniya: monografiya. M.: STJeNSI, 2003. 236 p.
- Golovin B. N. YAzyk i statistika. — M.: Prosveshchenie, 1970.
- Gomzin, A., Laguta, A., StroeV, V., Turdakov, D. Detection of author's educational level and age based on comments analysis // Paper presented at Dialogue 2018, Moscow, 30 May–2 June 2018. — URL: http://www.dialog-21.ru/media/4279/gomzin_turdakov.pdf (2018) (accessed: 05.07.2020).
- Ionova S.V., Ogorel'kov I.V. Rechevaya diagnostika lichnosti po gendernomu priznaku v avtorovedenii: kvantitativnyj podhod // Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Jazykoznanie. 2020. T. 19. № 1. Pp. 115-127.
- Ivchenko G.I., Medvedev YU.I. Matematicheskaya statistika: Uchebnoe posobie. — M.: Vysshaya shkola, 1984. 600 p.
- Karasik V.I. YAzykovoj krug: lichnost', koncepty, diskurs / V. I. Karasik; Nauch.-issled. lab. «Aksiol. Lingvistika». - M.: GNOZIS, 2004 (GUP Smol. obl. tip. im. V.I. Smirnova). 389 p.
- Karaulov Ju.N. Russkij jazyk i jazykovaya lichnost'. M.: LKI, 2010. 264 p.
- Kim L. G. Holisticheskie principy provedeniya identifikacionnoj avtorovedcheskoj ekspertizy raznozhanrovyyh tekstov // Mezhdunarodnaya nauchnaya konferenciya «Sovremennaya teoreticheskaya lingvistika i problemy sudebnoj ekspertizy». M.: Gosudarstvennyj institut russkogo yazyka im. A.S. Pushkina, 2019. Pp. 507-516.
- Korobov M.: Morphological analyzer and generator for Russian and Ukrainian languages / Khachay, M.Y., Konstantinova, N., Panchenko, A., Ignatov, D.I.,

- Labunets, V.G. (eds.) // AIST 2015. CCIS, vol. 542, 2015. Pp. 320–332.
- Kristall D., Dejvi D. Stilisticheskij analiz // Novoe v zarubezhnoj lingvistike, 1980. №9. Pp. 148–165.
- Lingvistika konstrukcij / Otv. red. E. V. Rahilina. — M.: «Izdatel'skij centr «Azbukovnik», 2010. — 584 p.
- Litvinova T. A. Idiolekt kak ob»ekt korpusnoj idiolektologii: k stanovleniyu novogo lingvisticheskogo napravleniya. // Uchenye zapiski Novgorodskogo gosudarstvennogo universiteta imeni YAroslava Mudrogo. No 7 (25). Pp. 1–5.
- Litvinova T. A., Sboev A. G., Panicheva P. V. (2018), Profiling the Age of Russian Bloggers // Proceedings of the 7th International Conference, AINL 2018, St. Petersburg, 2018. Pp. 167–177.
- Losev A.F. Vvedenie v obshhuyu teoriju jazykovykh modelej. M.: Editorial URSS, 2004. 296 p.
- Lutoslawski W. The origin and growth of Plato's logic. London. Longmans, Green and Co. 1897.
- McMenamin G.R. Forensic Linguistics: advances in forensic stylistics. / G.R. McMenamin. 2002. – 331 p.
- Medvedeva T.N. Formal'nye modeli v lingvistike : Uchebnoe posobie / T.N. Medvedeva. – Saratov: Nauchnaya kniga, 2010. – 56 p.
- Morozov N. A. Lingvisticheskie spektry: Sredstvo dlya otlicheniya plagiatov ot istinnykh proizvedenij togo ili drugogo izvestnogo avtora : Stilemetricheskij etyud / N. A. Morozov // Izvestiya Otdela russkogo yazyka i slovesnosti Imperatorskoj Akademii nauk. 1915. T. 20, kn. 4. P. 93–127.
- Murauer B., Tschuggnall M., Specht G. Dynamic Parameter Search for Cross-Domain Authorship Attribution // Notebook for PAN at CLEF 2018, 2018. – URL: http://ceur-ws.org/Vol-2125/paper_84.pdf (accessed: 05.07.2020).
- Muttenthaler L., Lucas G., Amann J. Authorship Attribution in Fan-Fictional Texts given variable length Character and Word N-Grams // Notebook for PAN at CLEF 2019, 2019. – URL: http://ceur-ws.org/Vol-2380/paper_49.pdf

(accessed: 05.07.2020).

Order of December 27, 2012 N 237 «On approval of the List of types of forensic examinations performed in federal budgetary forensic institutions of the Ministry of Justice of Russia, and the List of expert specialties for which the right to independently conduct forensic examinations in federal budgetary judicial expert institutions of the Ministry of Justice of Russia» [Priказ ot 27 dekabrya 2012 goda N 237 «Ob utverzhdenii Perechnya rodov (vidov) sudebnyh ekspertiz, vypolnyaemyh v federal'nyh byudzhetnyh sudebno-ekspertnyh uchrezhdeniyah Minyusta Rossii, i Perechnya ekspertnyh special'nostej, po kotorym predstavlyaetsya pravo samostoyatel'nogo proizvodstva sudebnyh ekspertiz v federal'nyh byudzhetnyh sudebno-ekspertnyh uchrezhdeniyah Minyusta Rossii»] (as amended on September 13, 2018).

Paducheva E.V. O semantike sintaksisa. M.: Nauka, 1974.

Panicheva, P., Mirzagitova, A., Ledovaya, Y.: Semantic feature aggregation for gender identification in Russian Facebook. // Filchenkov, A., Pivovarova, L., Žižka, J. (eds.) AINL 2017. CCIS, 2018. Vol. 789. Pp. 3–15.

Pavlovskaya N.YU. Modelirovanie yazykovykh kategorij v svete kognitivno orientirovannoj paradigmy nauchnogo znaniya. // Modeli v sovremennoj nauke: edinstvo i mnogoobrazie: sb. nauch.tr. / pod red. S.S. Vaulinoj, V.I. Greshnyh. – Kaliningrad: Izd-vo RGU im. I.Kanta, 2010 – 472 p. Pp. 73 – 80.

Pimonova E., Durandin O., Malafeev A. Doc2vec or better interpretability? A method study for authorship attribution // Paper presented at Dialogue 2020, Moscow, June 15–20, 2020, 2020.

Radbil' T.B., Markina M.V. Veroyatnostno-statisticheskie modeli v proizvodstve avto-rovedcheskoj jekspertizy russkojazychnyh tekstov // Politicheskaja lingvistika. 2019. № 2 (74). Pp. 156-166.

Revzin I. I. Sovremennaya strukturnaya lingvistika: Problemy i metody / I.I. Revzin ; AN SSSR, In-t slavyanovedeniya i balkanistiki. - Moskva: Nauka, 1977. - 263 p.

- Rodionova S. E. Semantika intensivnosti i ee vyrazhenie v sovremennom russkom yazyke / S. E. Rodionova // Problemy funkcional'noj grammatiki. Polevye struktury. SPb.: Izdatel'stvo «Nauka», 2005. P. 150-169.
- Romanov A., Kurtukova A., Fedotova A., Shelupanov A., Goncharov V. Authorship Identification of a Russian-Language Text Using Support Vector Machine and Deep Neural Networks // Future Internet. – 2021. – Volume 13. – Issue 1. – 3. – 16 p.
- Romanov A.S., Sobolev A.A., Kurtukova A.V, Fedotova A.M., Shelupanov A.A. Determining the Age of the Author of the Text Based on Deep Neural Network Models // Information. – 2020. –Volume 11. – Issue 12. – 589. – 12 p.
- Romanova T.V. Chelovek i vremja: Jazyk. Diskurs. Jazyko-vaja lichnost'. N. Novgorod: Nizhegorodskij gosudarstvennyj lingvistich-eskij un-t im. N.A. Dobroljubova, 2011.
- Rubcova I.I., Ermolaeva E.I., Bezrukova A.I. i dr. Kompleksnaya metodika proizvodstva avtorovedcheskih ekspertiz: Metodicheskie rekomendacii. – M:EKU MVD Rossii, 2007. — 192 p.
- Russkaja grammatika: nauchnye trudy: v 2 t. URL: <http://rusgram.narod.ru/index.html> (accessed: 12.01.2021).
- Saakov T.A. K voprosu o ponyatii demograficheskikh harakteristik avtora v sudebnoj avtorovedcheskoj ekspertize. // YAzyk. Pravo. Obshchestvo: sb. st. V Mezhdunar. nauch.-prakt. konf. Penza : PGU, 2018. Pp. 99-102.
- Sedov K. F. Stanovlenie struktury ustnogo diskursa kak vyrazhenie jevoljucii jazykovoj lichnosti: dis. ... d-ra filol. nauk. Saratov, 1999.
- Shtoff V.A. Modelirovanie i filosofiya / V.A. Shtoff. – M.; L.: Nauka, 1966. 304 p.
- Shuy R. W. Creating language crimes: How law enforcement uses (and misuses) language. New York: Oxford University Press, 2005. – 194 p.
- Solnyshkins M., Guryanov I., Gafiyatova E., Varlamova E. (2018). Readability metrics: the case of Russian educational texts // Proceedings of ADVED 2018-4th International Conference on Advances in Education and Social Sciences. – Istanbul, Turkey, 2018.

- Stepanenko A.A. Gendernaja atribucija tekstov komp'juternoj kommunikacii: statisticheskij analiz ispol'zovanija mestoimenij // Vestnik Tomskogo gosudarstvennogo universiteta. 2017. № 415. Pp. 17-25.
- Tarnopol'skij O.B., Kozhushko S.P. Kul'turno-obuslovlennye sostavljajushhie i metodika formiro-vanija vtorichnoj jazykovoj lichnosti u izuchajushhih in-ostrannyj jazyk v ego vuzovskom kurse // Jazykovaja lich-nost' i jeffektivnaja kommunikacija v sovremennom polikul'turnom mire : materialy V Mezhdunar. nauch.-prakt. konf., posvjashh. 20-letiju osnovanija kaf. teorii i praktiki perevoda fak. sociokul'tur. komunikacij Belorus. gos. un-ta, Minsk, 24-25 okt. 2019 g. Minsk: BGU, 2019. Pp. 300-304.
- Vinogradov V.V. Problema avtorstva i teorija stilej. M.: Goslitizdat, 1961. 614 p.
- Vorkachev S.G. Lingvokul'turologija, jazykovaja lichnost', koncept: stanovlenie antropocentricheskoj paradigmy v jazykoznanii // Filologicheskie nauki. 2001. № 1. Pp. 64-72.
- Vorob'ev V.V. Jazykovaja lichnost' v lingvokul'tur-ologii // Tez. dokl. Jazykovaja lichnost': Lingvistika. Lingvokul'turologija. Lingvodidaktika. BashGU. Nojabr' 2011 g. Ufa: RIC BashGU, 2011. Pp. 234-237.
- Vul S.M. Sudebno-avtorovedcheskaja identif-ikacionnaja jekspertiza: metodicheskie osnovy. Xar'kov: HNIISJe, 2007. 64 p.