VYTAUTAS MAGNUS UNIVERSITY
INSTITUTE OF THE LITHUANIAN LANGUAGE

ANDRIUS UTKA

STATISTICAL IDENTIFICATION OF TEXT FUNCTIONS

Summary of the Doctoral Dissertation
Humanities, Philology (04 H)

Kaunas, 2004

The dissertation has been prepared at Vytautas Magnus in 1999-2004.

The right for the joint doctoral studies was accorded to the Institute of the Lithuanian Language and Vytautas Magnus University on July 15, 2003 according to the decree of the Government of the Republic of Lithuania No. 926.

Research supervisor:

Assoc. Prof. Habil. Dr. **Rūta Marcinkevičienė**
Vytautas Magnus University, Humanities, Philology – 04 H

Defense council

Chair:

Prof. Habil. Dr. **Simas Karaliūnas**
Vytautas Magnus University, Humanities, Philology – 04 H

Members:

Prof. Habil. Dr. **Aloyzas Gudavičius**
Šiauliai University, Humanities, Philology – 04 H

Assoc. Prof. Dr. **Laima Grumadienė**
Institute of the Lithuanian Language, Humanities, Philology – 04 H

Assoc. Prof. Dr. **Meilutė Ramonienė**
Vilnius University, Humanities, Philology – 04 H

Assoc. Prof. Dr. **Ineta Savickienė**
Vytautas Magnus University, Humanities, Philology – 04 H

Opponents:

Prof. Habil. Dr. **Kazimieras Župerka**
Šiauliai University, Humanities, Philology – 04 H

Habil. Dr. **Aurelija Usonienė**
Vilnius University, Humanities, Philology – 04 H

The public defense of the dissertation will be held at 1 p.m. on November 16, 2004 in the Small Hall at the central building of Vytautas Magnus University.
Address: Daukanto 28, LT–44246, Kaunas, Lithuania.
Phone: +370 7 323599, fax: +370 7 203858.

The summary of the doctoral dissertation was sent out on October 16, 2004.

Those interested may acquaint themselves with the doctoral dissertation in the M. Mažvydas National Library and the Library of the Institute of the Lithuanian language in Vilnius, as well as in Vytautas Magnus University Library in Kaunas.

VYTAUTO DIDŽIOJO UNIVERSITETAS
LIETUVIŲ KALBOS INSTITUTAS

Andrius Utka

STATISTINIS TEKSTŲ FUNKCIJŲ NUSTATYMAS

Daktaro disertacijos santrauka
Humanitariniai mokslai, filologija, 04 H

Kaunas
2004

Disertacija rengta 1999–2004 metais Vytauto Didžiojo universitete.

Doktorantūros teisė suteikta Lietuvių kalbos institutui kartu su Vytauto Didžiojo universitetu Lietuvos Respublikos Vyriausybės 2003 m. liepos 15 d. nutarimu Nr. 926.

Mokslinis vadovas:

doc. habil. dr. **Rūta Marcinkevičienė**
Vytauto Didžiojo universitetas, humanitariniai mokslai, filologija – 04H

Disertacijos gynimo taryba

Pirmininkas:

prof. habil. dr. **Simas Karaliūnas**
Vytauto Didžiojo universitetas, humanitariniai mokslai, filologija – 04 H

Nariai:

prof. habil.dr. **Aloyzas Gudavičius**
Šiaulių universitetas, humanitariniai mokslai, filologija – 04 H

doc. dr. **Laima Grumadienė**
Lietuvių kalbos institutas, humanitariniai mokslai, filologija – 04 H

doc. dr. **Meilutė Ramonienė**
Vilniaus universitetas, humanitariniai mokslai, filologija – 04 H

doc. dr. **Ineta Savickienė**
Vytauto Didžiojo universitetas, humanitariniai mokslai, filologija – 04 H

Oponentai:

prof. habil. dr. **Kazys Župerka**
Šiaulių universitetas, humanitariniai mokslai, filologija – 04 H

habil. dr. **Aurelija Usonienė**
Vilniaus universitetas, humanitariniai mokslai, filologija – 04 H

Disertacija bus ginama viešame Filologijos mokslo krypties tarybos posėdyje 2004 m. lapkričio mėn. 16 d. 13 val. Vytauto Didžiojo universiteto Mažojoje salėje.
Adresas: Daukanto 28, LT–44246, Kaunas, Lietuva.
Tel.: +370 7 323599, faksas: +370 7 203858.

Disertacijos santrauka išsiuntinėta 2004 m. spalio mėn. 16 d.
Disertaciją galima peržiūrėti Nacionalinėje M. Mažvydo bei Lietuvių kalbos instituto bibliotekose Vilniuje bei Vytauto Didžiojo universiteto bibliotekoje Kaune.

1. Introduction

The object, aim, and objectives of the study

The ability to organise, systemise, and classify the existing information quickly and efficiently is among the most important issues nowadays.

Even though there have emerged many new ways to disseminate information, texts remain amongst the key information careers. Therefore the **object** of this study is a collection of texts and variation across texts.

The main **aim** of this research is to identify text functions based on the distribution of Lithuanian common words and word-forms; and to test the possibility of categorizing texts according to their functional properties.

The study needs to solve five main **objectives**:

1. To overview and evaluate existing text typologies for text classification;
2. To select an appropriate *a priori* classification of texts that would allow creating an experimental Lithuanian corpus, which would include a wide variety of texts;
3. To identify formal features, which are connected to text functions;
4. Applying factor analysis, to identify main text functions in the Lithuanian language;
5. To create a method, which would allow evaluating in what extent a given text is prototypical of the identified text functions.

Novelty of the research

The novelty of this work lies in the fact that text functions are identified using frequency distributions of common words and word-forms in texts, whereas previously in the Lithuanian linguistics functional styles have been analysed only in terms of distributions of content words or stylistically marked vocabulary.

Another innovative issue is the application of a corpus and factor analysis for identifying text functions in Lithuanian texts. Moreover, there is no previous research that has analysed text functions in Lithuanian based on such a large 10-million-word corpus (463 texts).

Research material and methodology

This research is empirical and corpus-driven. In order to achieve the aim of the research, two corpora have been compiled: the *Small Corpus of the Lithuanian Language (Mažasis lietuvių kalbos tekstynas)* comprising 25 millions of running words and the *Experimental Corpus (Eksperimentinis tekstynas)* comprising 10 millions of running words. Both the corpora are based on the *Corpus of the Contemporary Lithuanian Language (CCLL)*[1], which is being continuously compiled at the Centre of Computational Linguistics at Vytautas Magnus University.

The methodology of this work is based on empirical language evidence, which is analysed quantitatively (by factor analysis) and qualitatively (by interpretation of quantitative results). The following computer programs have been used for this work: special PERL scripts, which are used to segment sentences, to count linguistic features etc.; package of statistical program *SPPS for Windows,* which is used in all phases of factor analysis; linguistic program *WordSmith Tools,* which is used for counting linguistic features and concordance analysis; and *Microsoft Excel,* which is used for creating graphs as well as for sorting and counting numerical data.

The structure of the work

The dissertation consists of six chapters: the object and aims of this research are discussed in "Introduction"; the second chapter, "The problem of text classification", focuses on theoretical issues of text classification; third chapter, "Research Methodology for Analysing Lithuanian Text Functions", describes all practical phases of research; the fourth chapter, "Application of factor analysis for analysing text functions", deals with the application of factor analysis to the research data; the fifth chapter, "Text Functions in Lithuanian", focuses on interpretation of results of factor analysis, which involves naming and describing text functions and functional paradigms in Lithuanian texts; finally, the sixth chapter presents the main conclusions of the research.

Theses to be defended

The main theses of the dissertation are:

– distribution of common words and word-forms in texts is a significant indicator of text functions;

[1] lit. *Dabartinės lietuvių kalbos tekstynas (DLKT)*

– factor analysis is a statistical method that allows identifying groups (paradigms) of common words, word-forms and statistical features that are representative of text functions;

– it is possible to evaluate to what extent a text is prototypical of the identified text functions.

2. The problem of text classification

In Chapter 2 the related research on text categorization is reviewed and several existing criteria of text classification are discussed, such as *text, style, functional style, register, genre,* and *text type.*

It is observed that all studies of automatic text classification (text categorization) classify texts either by *topic*, *genre*, or *text type*. The present research can be viewed as belonging to the latter type of studies. The observation is made, that in terms of overall methodology this work is related to the studies by D. Biber, while in terms of usage of common word distributions it is related to the works of J. F. Burrows, H. Craig, T. Tabata, and R. Sigley.

While reviewing basic terminology for classifying texts, the conclusion is made that *genre* is the best criterion for categorising texts *a priori*, in order to create a representative experimental corpus for determining functional properties in texts. The following hierarchy of categories is accepted as appropriate for classifying texts in the experimental corpus: mode of discourse (written and spoken), super-genre (e.g. academic prose, fiction), genre (e.g. novel, short story, article), and sub-genre (e.g. financial reportage, history article).

3. The methodology of determining functional qualities in Lithuanian texts

This chapter deals with four topics: *a priori* text classification, assumptions underlying identification of text functions, identifying text functional features in a corpus, and calculation of frequency distributions of these features in the experimental corpus.

The first section of this chapter is aimed at determining a proper *a priori* text classification, which will allow compiling a representative experimental corpus. The existing text classification of CCLL is considered as a possible *a priori* text classification for the experimental corpus. It is observed that this classification does not

meet requirements of the research, as some categories in CCLL are too broad for an adequate representation of genre variation. Therefore, the decision has been taken to adopt a text classification used by Douglas Biber in his work "Variation across speech and writing"[1], since it is viewed by many researchers as being representative in terms of genre variation.

Table 1. Classification of the experimental corpus

	MODE OF DISCOURSE Super-genre Genre Sub-genre		
I	WRITTEN TEXTS	5	Press
1	Academic prose		Interviews
	Articles and books		Popular lore
	Finance		Letters to the editor
	Philology		Religion articles
	Philosophy		Reportages
	Natural sciences		Financial
	History		Computers
	Medicine		Criminal
	Political science		Cultural
	Social and behavioural science		Political
	Educology		Sports
	Technology and engineering		Skills and hobbies
	Law		Short news
2	Biographical literature	II	SPOKEN TEXTS
	Biographies	6	Professional discussions
	Memoirs		Radio discussions
3	Fiction		Parliament debates
	Novellas		Theatre discussions
	Humour stories	7	Planned speeches
	Short stories		Ceremonial speeches
	Fairy tales		Political speeches
	Novels		Lectures
4	Official documents	8	Conversations
	EU directives		
	Enterprise regulations	Total:	2 types of discourse
	Laws		8 super-genres
	Minutes		29 genres
	Agreements		17 sub-genres
	Court documents		
	Governmental resolutions		

The modified Lithuanian classification is created consisting of written and spoken texts, that are divided into 8 super-genres, 29 genres, and 17 sub-genres (see Table 1).

[1] Biber, D. (1988) *Variation across Speach and Writing.* Cambridge: Cambridge University Press.

The second section of this chapter observes that common words and word-forms have some special properties that distinguish them from rarer words. The following special properties are characteristic to common words and word-forms:

- the mathematical relationship *Zipf Law*[1], which is valid for 98% of vocabulary, does not hold for 2% of common words and word-forms that are at the top of frequency list;
- common words and word-forms make up a big proportion of any text, which is associated with their very frequent usage. For example, three most frequent words in English (*the, and, of*) as well as in Lithuanian (*ir, į, kad*) make up 5% of any corpus.
- as a rule, common words and word-forms are shorter. The higher a word or a word-form is on a frequency list, the shorter it is likely to be, and vice versa. This property on the one hand is associated with effectiveness of language usage, and on the other with the fact that shorter words carry less information.

Based on the above observations, the assumption is made that very frequent words and word-forms make up a structure in a text, while rarer words fill the structure with content. Therefore, it is reasonable to assume that distribution of common words and word-forms may be a good indicator of text functions.

A small experiment has been conducted to exemplify how common words and word-forms correlate with specific textual functions. Frequency lists from the *Small Lithuanian corpus,* fictional text, and two texts of academic prose are compared. It is shown that, for example, very frequent usage of 3rd person pronouns may point out to the narrative function of a text, while frequent usage of a certain group of verbs may be an indication that the descriptive function is dominant in a text.

Based on these observations, it is assumed that frequency distribution of common words and word-forms are not chaotic or accidental, as they can directly point to certain text functions.

The third section of this chapter deals with the practical issue of identifying a list of functional features, which consists of lexical (common words and word-forms) and statistical features (average sentence length, average word length, and type/token ratio).

[1] Zipf proved that a mathematical relation exists between the frequency of a word and the number of different words occurring with the same frequency. G. K. Zipf (1935) *The Psychobiology of Language.* New York: Houghton Mifflin.

The *Small Lithuanian corpus* is used as the basis of identifying the final list of functional features. Nouns (except *metų, metu* and *nr*) and single letter abbreviations (e.g. *a, d*) have been removed from the list of most frequent word-forms, as the former are seen as being more topic-oriented and not function-oriented, and the latter as being too ambiguous. The final list of 102 functional features, which will be used for further analysis, is presented in the Table 2 below.

Table 2. Statistical (1–3) and lexical (4–102) features

No.	Feature	No.	Feature	No.	Feature	No.	Feature	No.	Feature	No.	Feature
1	type/token r.	18	savo	35	bei	52	dabar	69	galima	86	tą
2	avg. w. l.	19	jis	36	tačiau	53	be	70	bus	87	tas
3	avg. s. l.	20	taip	37	ji	54	nes	71	mano	88	tarp
4	ir	21	nuo	38	to	55	net	72	jog	89	visi
5	į	22	apie	39	labai	56	ką	73	man	90	ją
6	kad	23	jo	40	prie	57	mes	74	tiek	91	kuris
7	iš	24	jau	41	čia	58	būtų	75	kurie	92	kiek
8	su	25	kai	42	būti	59	kur	76	jei	93	tu
9	buvo	26	dar	43	gali	60	ant	77	reikia	94	pats
10	o	27	dėl	44	arba	61	nėra	78	pagal	95	kitų
11	yra	28	jų	45	pat	62	jį	79	gal	96	tam
12	kaip	29	aš	46	jie	63	todėl	80	daugiau	97	vėl
13	tai	30	po	47	turi	64	prieš	81	vienas	98	tada
14	ar	31	už	48	mūsų	65	jeigu	82	tuo	99	juos
15	tik	32	per	49	metų	66	daug	83	jam	100	ten
16	ne	33	jos	50	iki	67	vis	84	nr	101	kurios
17	bet	34	kas	51	nors	68	nei	85	metu	102	vieną

The fourth section of this chapter presents the actual calculations of frequencies of 102 features across 463 texts in the experimental corpus. All the texts are different in length, therefore, to make the calculated frequencies comparable, the absolute frequencies need to be normalised. All word and word-form frequencies are normalised to a text length of 1000 words, type/token ratio to a text length of 500 words. Average word and sentence lengths are not normalised as they are not dependent on length of text. The final result of these calculations is a matrix, which includes normalised frequencies for all 102 functional features across 463 texts. This matrix is initial data that will be analysed using factor analysis.

4. Application of factor analysis for identifying text functions

This chapter deals with theoretical and practical issues of applying factor analysis to the frequency distribution of functional features. The main steps of factor analysis are the following:

- testing suitability of initial data for using factor analysis;
- determining factors and factor rotation;
- identifying factorial structure;
- calculating factor scores;
- interpreting factors.

The first section presents the general definition and rationale of factor analysis as well as its applicability to linguistic data. This method is commonly used to reduce the number of multiple variables into several factors without loosing essential information. It is noted, that factor analysis requires taking a number of intermediate decisions, which may influence final results and, thus, these decisions need to be justified.

The second section deals with application of suitability tests to initial data. Two suitability tests have been used: *Barttlet's test of Sphericity* and *Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy*. Both the tests justify the usage of factor analysis to the initial data.

In the third section, a number of technical solutions are presented. The method of *principal axis factoring (common factor analysis)* is chosen for determining factors. Analysis of scree plot of eigenvalues allows determining a number of significant factors. It is shown, that the most optimal solution for the given data is a solution of seven factors. In other words, the frequency distribution of 102 functional features will be reduced into seven underlying factors.

The extraction of seven factors is followed by factor rotation, which helps to interpret the factors. The oblique *Promax* rotation is chosen for the current data. Oblique rotation methods are recommended for describing textual data, as they permit slight correlations among factors. This choice is justified by the fact that linguistic phenomena as a rule correlate.

The factorial structure is presented in the fourth section. The factorial structure consists of seven groups of features, which significantly load on the factors. The features are sorted by their weight. The bigger is the weight, the more important the

feature is to the factor. Only features that have weights larger than |±0,3| are considered as significant to a factor.

In the fifth section, it is presented how factor scores are calculated for each text. The calculation of factor scores is used for confirming and interpreting the established factorial structure. Seven factor scores are calculated for each text in the experimental corpus, so that the texts can be sorted according to their prototipicality in respect to the seven factors.

5. Text functions in Lithuanian

The first section of this chapter discusses basic principles of factor interpretation. Statistical factor analysis, which is performed by a computer, allows determining factors, but the meaning of these factors needs to be interpreted and explained by a researcher. A particular group of functional features loads on each factor. As a rule, the first factor is the strongest with the biggest number of significant features, while the last is the weakest with the least number of significant features. Functional features of a factor that have positive weights have a tendency to occur in certain texts more frequently, while negative less frequently. In other words, positive features correlate positively with a factor, and negative features – negatively.

It is reasonable to assume, that significant functional features of a factor share some common functions in texts.

1st Factor: Spontaneous Expressiveness

32 features load on the first factor (Table 3). 30 features have positive weights and 2 features have negative weights. This factor is the strongest and the most reliable factor of all seven factors. Therefore, its interpretation is given the biggest attention in the dissertation. *Average word length* and *type/token ratio* are negative features of the factor, which point out to shorter words and repetitive vocabulary.

Microanalysis of usage of the above enumerated features demonstrates that there are certain functions shared by all of them. It is shown that these features are commonly used to express:

– spontaneity;

– dependence on situation (reference to location and time);

– expressiveness.

Table 3. Significant features of the first factor

1st FACTOR	
Positive features	***Weight***
1. **tai** (it)[1]	**1,04**
2. **jau** (already)	**0,83**
3. **čia** (here)	**0,82**
4. **ten** (there)	**0,80**
5. **tas** (that)	**0,70**
6. **ką** (what)	**0,69**
7. **tą** (that)	**0,67**
8. **tiek** (so much)	**0,62**
9. **taip** (yes)	**0,61**
10. **mes** (we)	**0,60**
11. **abai** (very)	**0,59**
12. **bet** (but)	**0,58**
13. **dabar** (now)	**0,58**
14. **aš** (I)	**0,53**
15. **kaip** (as)	**0,48**
16. **kur** (where)	**0,46**
17. **nors** (although)	**0,46**
18. **vis** (over)	**0,45**
19. **tada** (then)	**0,39**
20. **daug** (many)	**0,39**
21. **nes** (because)	**0,39**
22. **o** (and)	**0,38**
23. **kai** (as)	**0,37**
24. **kiek** (how many)	**0,36**
25. **(to** (į)	**0,32)**[2]
26. **(buvo** (was)	**0,32)**
27. **(gal** (maybe)	**0,31)**
28. **(visi** (all)	**0,31)**
29. **(jeigu** (if)	**0,31)**
30. **(dar** (yet)	**0,30)**
Negative features	***Weight***
31. avg.w.l.	**-0,64**
32. type/token ratio	**-0,43**

The main functions of this factor are designated as **functions of spontaneous expressiveness**. The group of features that expresses these functions is referred to as the **paradigm of spontaneous expressiveness** or the **first paradigm.** The following definition characterizes this function:

> *The function of spontaneous expressiveness, which is expressed by the first paradigm, is characteristic to texts, which have not been prepared in advance, have features of spoken language, are closely associated with the context of situation and highly emphatic. This function is most typical of conversations.*

Factor scores can be calculated for individual texts as well as for larger groups of texts. Only those features exceeding 0,35 are used in the calculation of factor scores. The factor scores for eight super-genres in respect to the first paradigm are given in Figure 1 below.

[1] The presented translation of common words is not exaustive, as it is just inteded to help non-Lithuanian speakers.

[2] Features with weights less that 0,35 are in parentheses. Although they may be used when interpreting factors, they are not used in calculating factor scores.

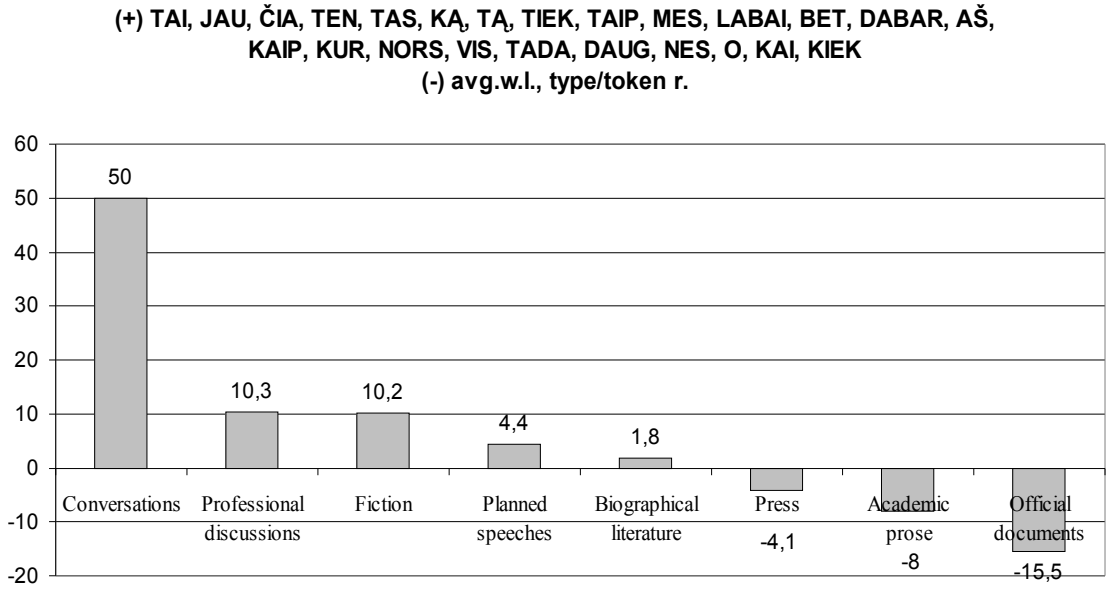

Figure 1. Factor scores of eight super-genres in respect to the paradigm of spontaneous expressiveness

Figure 1 illustrates that the most prototypical super-genre in respect to the paradigm of spontaneous expressiveness is *conversations,* while the least prototypical is *official documents.*

2nd Factor: Narrativeness

13 features load on the second factor (Table 4). Even 7 features are different cases of pronouns the 3rd person pronouns *jis* (*he*) and *ji* (*she*). It shows that the functions of this factor are associated with frequent usage of the 3rd person pronouns.

Table 4. Significant features of the second factor

2nd FACTOR			
Positive features	***Weight***		
1. **jis** (he)	**0,83**	7. **jį** (him)	**0,58**
2. **ji** (she)	**0,68**	8. **ant** (on)	**0,50**
3. **jos** (her)	**0,66**	9. **į** (to)	**0,49**
4. **jam** (him)	**0,65**	10. **prie** (near, by)	**0,46**
5. **ją** (her)	**0,64**	11. **savo** (my, his, her, its)	**0,45**
6. **jo** (his)	**0,63**	12. **tu** (you)	**0,40**
		13. **vėl** (again)	**0,35**

It is generally known that the frequent usage of 3rd person pronouns is characteristic to narrations. Prepositions *ant, į, prie* are also frequently used in narrations.

The main function of this factor is designated as the **function of narrativeness**. The group of features that expresses this function is referred to as **the paradigm of**

narrativeness or **the second paradigm.** The following definition characterizes this function:

> *The function of narrativeness, which is expressed by the second paradigm, is characteristic to texts where story telling is dominating. This function is most typical of fictional texts.*

The factor scores for eight super-genres in respect to the second paradigm are given in Figure 2 below.

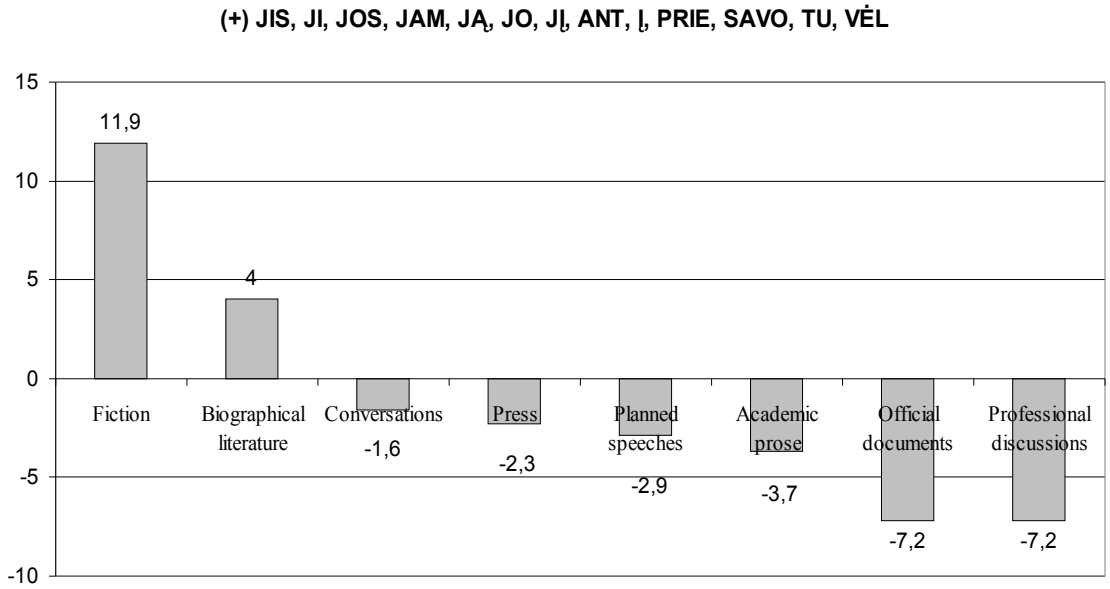

Figure 2. Factor scores of eight super-genres in respect to the paradigm of narrativeness

Figure 2 shows that the most prototypical super-genre in respect to the paradigm of narrativeness is *fiction,* while the least prototypical is *professional discussions.*

3rd Factor: Directiveness

15 features load on the third factor (see Table 5 below). 13 features have positive weights, and 2 features – negative. The fact of appearance of 7 verbs (*būti, gali, turi, galima, reikia, yra, buvo*) among significant features has proved to be important to interpretation.

Table 5. Significant features of the third factor

3rd FACTOR	
Positive features	***Weight***
1. **būti** (be)	**0,70**
2. **arba** (or)	**0,69**
3. **gali** (can)	**0,68**
4. **turi** (have)	**0,67**
5. **ar** (or)	**0,65**
6. **tam** (to that)	**0,52**
7. **jei** (if)	**0,47**
8. **pagal** (according)	**0,42**
9. **jeigu** (if)	**0,42**
10. **(kurios** (which)	**0,32)**
11. **(galima** (possible)	**0,31)**
12. **(reikia** (need)	**0,31)**
13. **(yra** (is)	**0,30)**
Negative features	***Weight***
14. **buvo** (was)	**-0,48**
15. (type/token ratio	**-0,31)**

Closer analysis has shown that most of these features are used in official documents to express authority. Concordance analysis of these features has shown that word combinations *gali būti* (*can be*) and *turi būti* (*have to be*) are the most important when identifying the official language.

The main function of this factor is designated as **the function of directiveness**. The group of features that expresses this function is referred to as **the paradigm of directiveness** or **the third paradigm**. The following definition describes this function:

> *The function of directiveness, which is expressed by the third paradigm, is characteristic to texts, where certain modal and present tense verbs are used more often, as well as certain alternative and conditional constructions. This function is typical of official documents.*

The factor scores for eight super-genres in respect to the third paradigm are given in Figure 3 below.

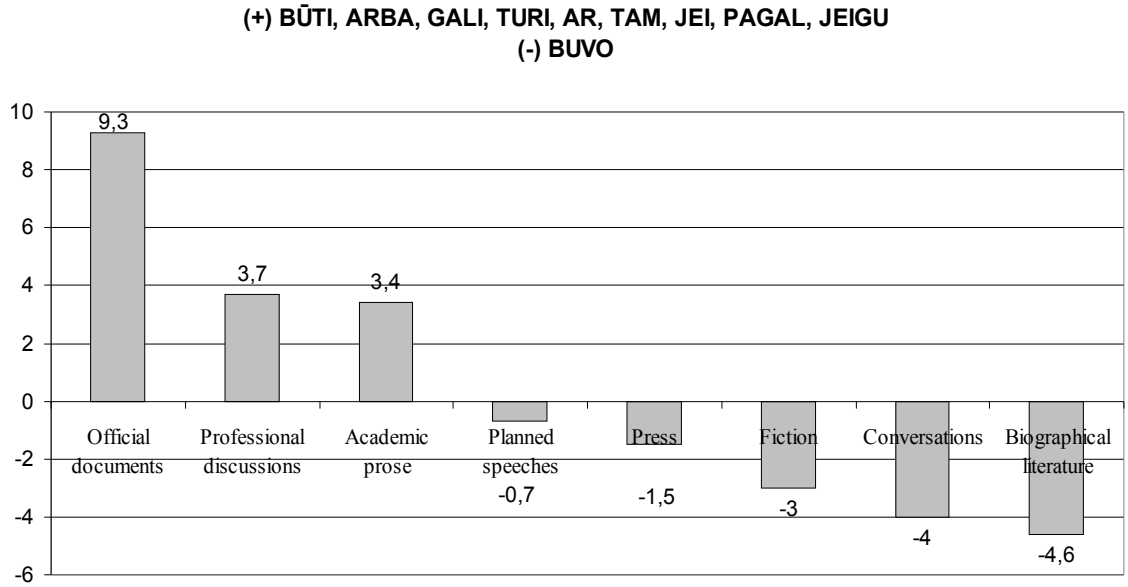

Figure 3. Factor scores of eight super-genres in respect to the paradigm of directiveness

Figure 3 illustrates that the most prototypical super-genre in respect to the paradigm of directiveness is *official documents,* while the least prototypical is *biographical literature.*

4th Factor: Prepared Expressiveness

19 features load on the fourth factor (Table 6), most of them are particles. 16 of them have positive weights, and 3 – negative.

Table 6. Significant features of the fourth factor

4th FACTOR		10.**gal** (maybe)	**0,38**
Positive features	***Weight***	11.**dar** (yet)	**0,37**
1. **tik** (only)	**0,66**	12.**kiek** (how many)	**0,36**
2. **o** (and)	**0,58**	13.**(todėl** (so)	**0,33)**
3. type/token ratio	**0,58**	14.**(ir** (and)	**0,31)**
4. **ne** (no)	**0,53**	15.**(bet** (but)	**0,31)**
5. **net** (even)	**0,49**	16.**(reikia** (need)	**0,30)**
6. **jei** (if)	**0,43**	***Negative features***	***Weight***
7. **nei** (nor)	**0,41**	17.**dėl** (for)	**-0,37**
8. **be** (without)	**0,41**	18.**mes** (we)	**-0,35**
9. **kas** (what)	**0,39**	19.**(buvo** (was, were**)**	**-0,31)**

It is shown in this section that the presence of particles and rich vocabulary[1] are characteristic to written expressive language.

[1] The higher weight of the type/token ratio indicates richer vocabulary in texts.

The main function of this factor is designated as **the function of prepared expressiveness**. The group of features that expresses this function is referred to **as the paradigm of prepared expressiveness** or **the fourth paradigm**.

> *The function of prepared expressiveness, which is expressed by the fourth paradigm, is characteristic to texts, which have rich and expressive vocabulary. This function is typical of fictional texts.*

The factor scores for eight super-genres in respect to the fourth paradigm are given in Figure 4 below.

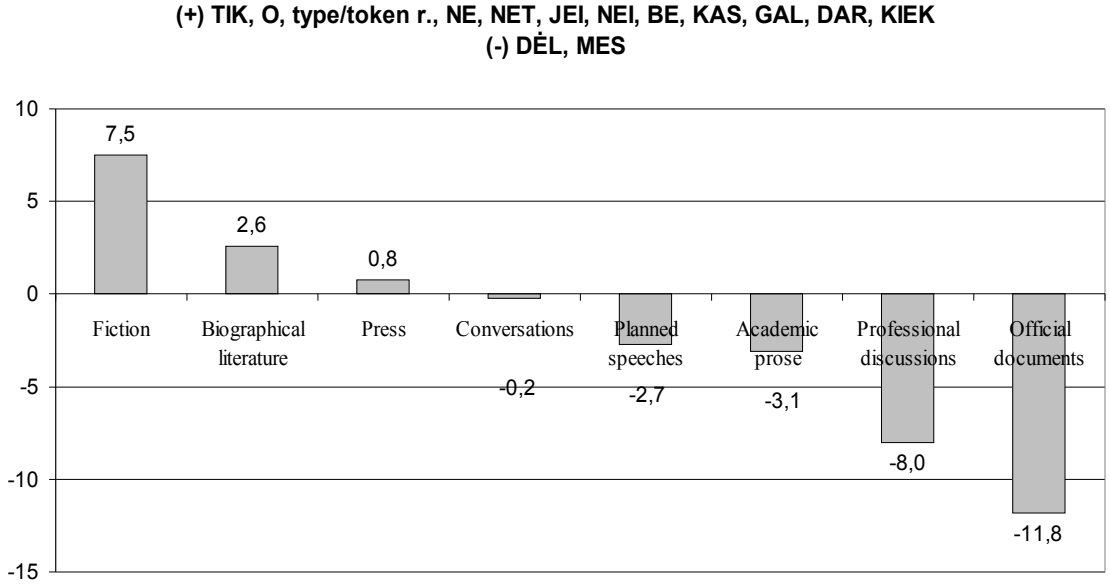

Figure 4. Factor scores of eight super-genres in respect to the paradigm of prepared expressiveness

Figure 4 shows that the most prototypical super-genre in respect to the paradigm of prepared expressiveness is *fiction,* while the least prototypical is *official documents.*

5th Factor: Persuasiveness

12 features load on the fifth factor (see Table 5 below). All the features have positive weights.

Table 7. Significant features of the fifth factor

5th FACTOR			
Positive features	***Weight***		
1. **kad** (that)	**0,69**	6. **yra** (is)	**0,44**
2. **nėra** (is not)	**0,582**	7. **mes** (we)	**0,39**
3. **būtų** (would be)	**0,565**	8. **mūsų** (ours)	**0,381**
4. **dėl** (for, due to)	**0,479**	9. **todėl** (so)	**0,367**
5. **kurie** (which)	**0,475**	10. **nes** (because)	**0,366**
		11. **(to** (him, his)	**0,337)**
		12. **(kuris** (which)	**0,314)**

It is demonstrated that most of these features can be used to persuade and influence people. For example, words *kad, dėl, todėl, nes* denote the opposition of cause and result, which is characteristic to the language of argumentation.

The main function of this factor is designated as **the function of persuasiveness**. The group of features that expresses this function is referred to as **the paradigm of persuasiveness** or **the fifth paradigm**. The following definition characterizes this function:

> *The function of persuasiveness, which is expressed by the fifth paradigm, is characteristic to texts, where the language of argumentation is used in order to influence listeners or readers. This function is typical of official debates.*

The factor scores for eight super-genres in respect to the fifth paradigm are given in Figure 5 below.

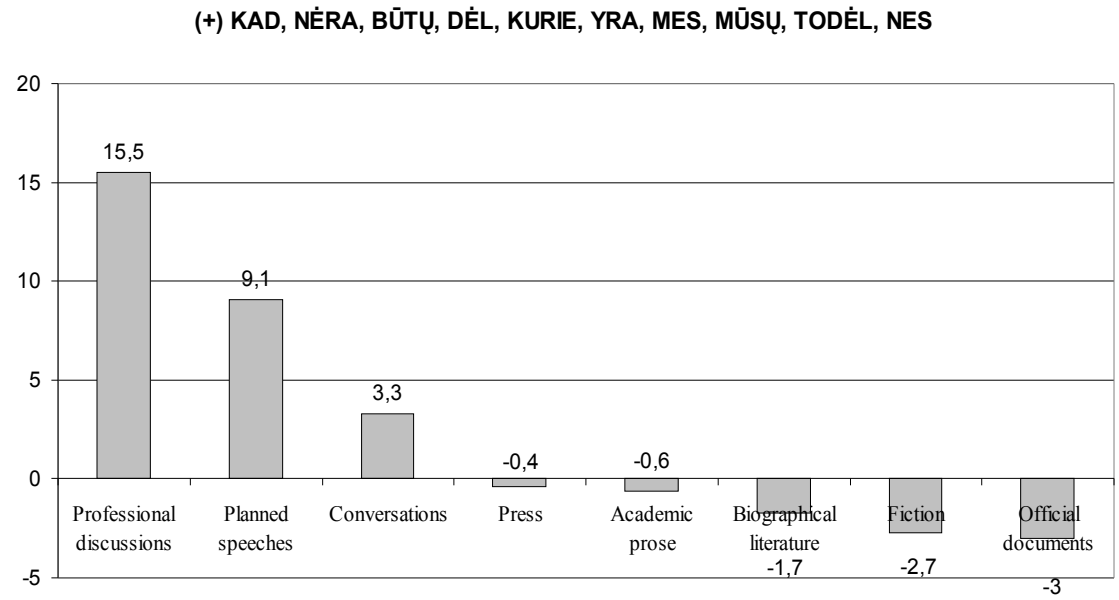

Figure 5. Factor scores of eight super-genres in respect to the paradigm of persuasiveness

Figure 5 shows that the most prototypical super-genre in respect to the paradigm of persuasiveness is *professional discussions,* while the least prototypical are *fiction* and *official documents.*

6th Factor: Descriptiveness

16 significant features load on this factor (see Table 8 below). 9 of them have positive weights, and even 7 – negative. Such a factorial structure means that the function of this factor is expressed not only by frequent usage of positive features, but also by the infrequent usage of negative features.

Table 8. Significant features of the sixth factor

6th FACTOR		9. **(savo** (my, his, her, its)	**0,30)**
Positive features	***Weight***	***Negative features***	***Weight***
1. **tarp** (between)	**0,50**	10. **nr** (No.)	**-0,54**
2. avg. s. l.	**0,45**	11. **už** (for, by, over)	**-0,44**
3. **galima** (possible)	**0,37**	12. **dėl** (for, due to)	**-0,42**
4. **tačiau** (however)	**0,36**	13. **mano** (my)	**-0,41**
5. **tuo** (with that)	**0,36**	14. **man** (my)	**-0,40**
6. **jų** (theirs)	**0,35**	15. **(aš** (I)	**-0,30)**
7. (type/token ratio	**0,32)**	16. **(gal** (maybe)	**-0,30)**
8. **(buvo** (was, were)	**0,30)**		

The analysis of the features indicates that they are prevalent in the academic discourse.

The main function of this factor is designated as **the function of descriptiveness**. The group of features that expresses this function is referred to as **the paradigm of descriptiveness** or **the sixth paradigm**. The following definition characterizes this function:

> *The function of descriptiveness, which is expressed by the sixth paradigm, is characteristic to texts, which have long sentences, rich vocabulary, and impersonal constructions. This function is typical of academic prose texts.*

The factor scores for eight super-genres in respect to the sixth paradigm are given in Figure 6 below.

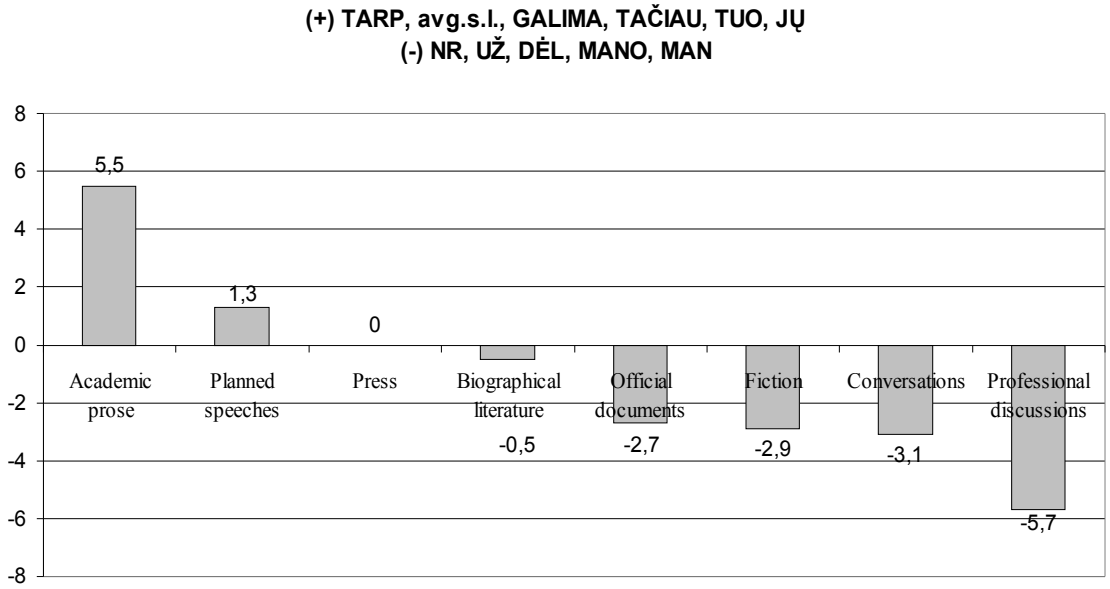

Figure 6. Factor scores of eight super-genres in respect to the paradigm of descriptiveness

Figure 6 shows that the most prototypical super-genre in respect to the paradigm of descriptiveness is *academic prose* texts, while the least prototypical is *professional discussions*.

7th Factor: Temporality

Only 9 features load on the seventh factor (Table 9). This is the weakest factor as none of the features have weight bigger that 0,5. Its suitability to classify texts is yet to be confirmed.

Table 9. Significant features of the seventh factor

7th FACTOR					
Positive features	***Weight***		5.	**(bus** (will be)	**0,34)**
1. **per** (over, during)	**0,47**		6.	**(daugiau** (more)	**0,33)**
2. **metų** (years)	**0,46**		7.	**(iki** (to, untill)	**0,31)**
3. **po** (after)	**0,36**		***Negative features***		***Weight***
4. **(prieš** (before)	**0,34)**		8.	**ir** (and)	**-0,47**
			9.	**(jos** (her)	**-0,31)**

The analysis of the features has shown that they express qualities that are more typical of press. The features *per, metų, po, prieš, bus, iki* can all be used to express temporal relations. The main function of this factor could be called as the **function temporality,** and the group of features that expresses this function **the paradigm of temporality**. However, such a small number of features does not ensure reliable classification of texts; therefore, this function may be seen as a theoretical construct not suitable for practical calculations.

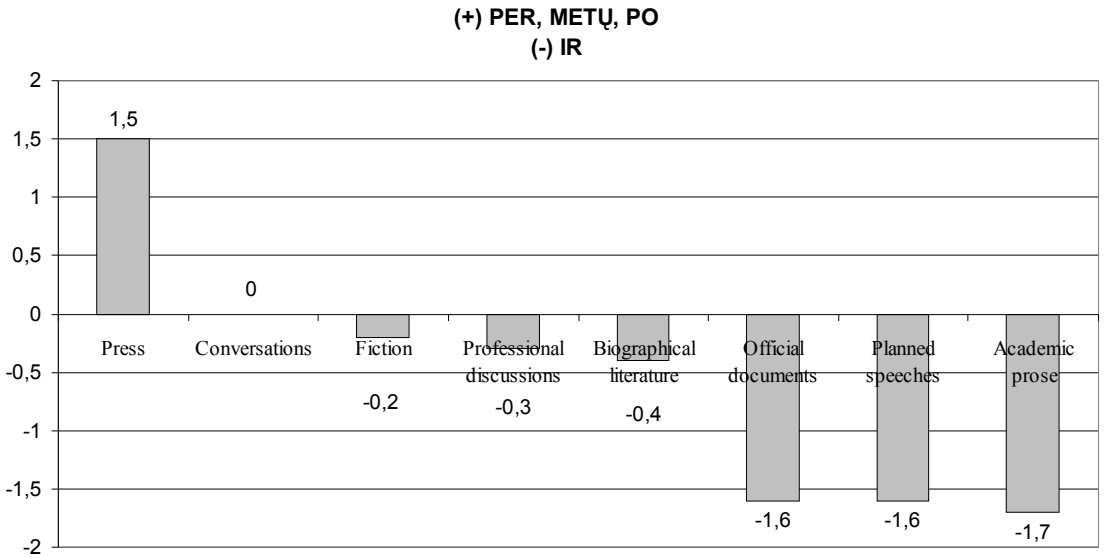

Figure 7. Factor scores of eight super-genres in respect to the paradigm of temporality

The factor scores for eight super-genres in respect to the seventh paradigm are given in Figure 7 above.

Figure 7 shows that the most prototypical super-genre in respect to the paradigm of temporality is *press,* while the least prototypical are *official documents, planned speeches,* and *academic prose.*

6. Conclusions

The following conclusions have been formulated in this dissertation:

1. The relation between language as a system and text as a unit of communication is asserted through works of corpus linguistics, which can be summarised by one of the basic postulates of corpus linguistics that "authentic language occurs as language-in-text"[1]. In this work this conception helps to differentiate two groups of notions. The first group includes *functional style* and *register,* which are associated with language as a system; and the second group includes *individual style, genre,* and *text type,* which are associated with text as a unit of communication. In this work classification according to *genre* is chosen as *a priori* text classification, while classification according to *text type* as *a posteriori* text classification.

2. Distribution of common words and word-forms in texts is not chaotic or accidental. While being the most frequent text elements, common words and word-forms are closely associated with functional text properties, therefore they are seen as significant indicators of text functions.

3. Big variety of texts can be well represented by a corpus, which has the following hierarchy of classifying categories: spoken and written texts, super-genres, genres, and sub-genres. This classification of text has been used for the compilation of the Lithuanian experimental corpus, which consists of spoken and written texts, 8 super-genres, 29 genres, and 17 sub-genres.

4. With the help of factor analysis seven factors and their significant features have been identified in the experimental corpus. The factor interpretation has demonstrated that features of each factor represent specific text functions. Each function is represented by a group of features that is referred to as a *functional paradigm.* The functions and appropriate functional paradigms have been designated as *spontaneous*

[1] Stubbs, M. (1993) British Traditions in Text Analysis: From Firth to Sinclair. In M. Baker, G. Francis, and E. Tognini-Bonelli (eds.), *Text and Technology: In Honour of John Sinclair,* 1–33. Philadelphia: John Benjamins Publishing Company, p. 11.

expressiveness, narrativeness, directiveness, prepared expressiveness, persuasiveness, descriptiveness, and *temporality.*

5. Based on results of factor analysis, each text can be evaluated how prototypical it is in respect of each paradigm. The practical text evaluation method is created, which allows a researcher to evaluate text in respect of six functional paradigms (*spontaneous expressiveness, narrativeness, directiveness, prepared expressiveness, persuasiveness, descriptiveness).*

6. The current research once more confirms that form and content as well as form and function are inseparable in texts. This unity is demonstrated by the fact that paradigms of common words and word-forms credibly indicate text functions.

STATISTINIS TEKSTŲ FUNKCIJŲ NUSTATYMAS

Reziumė

Darbo objektas, tikslas, uždaviniai. Šiuolaikinė visuomenė neretai susiduria su viena iš modernaus pasaulio rykščių – informacijos pertekliumi. Šiandienos problema yra ne informacijos stoka, bet gebėjimas orientuotis jos sraute: paieška, tvarkymas ir sisteminimas. Dabar didelė dalis informacijos pateikiama virtualioje erdvėje elektroniniu pavidalu, ir jai skleisti kuriami vis nauji informacijos perdavimo būdai. Nors atsirado daugybė naujų informacijos perdavimo būdų bei formų, kalba ir ja kuriami tekstai išlieka tarp pačių svarbiausiųjų. Todėl tekstai ir jų įvairovė yra šio darbo tyrimo **objektas**.

Šio darbo **tikslas** yra išanalizuoti galimybę įvertinti ir klasifikuoti lietuvių kalbos elektroninius tekstus pagal jų funkcijas bei sukurti greitą ir efektyvų automatinį klasifikavimo metodą. Tekstų funkcijos bus nustatomos remiantis lengvai identifikuojamų formalių kalbinių požymių dažnumo pasiskirstymu.

Darbo tikslui pasiekti keliami penki pagrindiniai **uždaviniai**:

1. Apžvelgti ir įvertinti egzistuojančias tekstų tipologijas bei jų pritaikymo galimybes automatinei tekstų funkcijų analizei;

2. Pasirinkti tinkamą *a priori* tekstų klasifikaciją, kuri leistų sukurti kuo didesnę tekstų įvairovę apimantį eksperimentinį lietuvių kalbos tekstyną;

3. Nustatyti formalius kalbinius požymius, susijusius su tekstų funkcijomis;

4. Naudojant statistinę faktorių analizę, nustatyti pagrindines lietuvių kalbos tekstų funkcijas;

5. Sukurti metodą, leidžiantį įvertinti duotojo teksto prototipiškumą nustatytųjų funkcijų atžvilgiu.

Darbo naujumas ir aktualumas. Šio darbo naujumas yra tas, kad teksto funkcijoms nustatyti pasitelkiamas labai dažnų žodžio formų pasiskirstymas tekstuose. Lietuvių kalbotyros funkcinių stilių tyrimuose daugiau yra analizuojami stilistiškai žymėti prasminiai žodžiai, o statistiniai funkcinių stilių tyrimai remiasi daugiausia nominatyvinių kalbos dalių pasiskirstymo analize.

Naujas dalykas lietuvių kalbotyroje taip pat yra faktorių analizės metodo taikymas tekstų tipologijai tirti tokią didelę tekstų įvairovę apimančiame tekstyne (463 tekstai; 10,68 mln. žodžių).

Manoma, kad tekstų klasifikacija pagal jų funkcijas, paremta tekstų kalbinių požymių dažnumų pasiskirstymų statistine analize, yra tinkamas ir ateityje naudotinas tekstų klasifikavimo būdas. Todėl svarstytina galimybė šią klasifikaciją pritaikyti Dabartinės lietuvių kalbos tekstynui, kuriame derintųsi ši ir jau egzistuojanti apriorinė tekstų klasifikacija.

Tiriamoji medžiaga ir tyrimo metodika. Ši disertacija yra empirinis tekstynų lingvistikos darbas, kuriame tekstų tipologijos tyrimo pradinė medžiaga yra tekstynas. Darbo tikslams pasiekti naudojami 100 mln. žodžių Dabartinės lietuvių kalbos tekstyno (Marcinkevičienė 1997) pagrindu sukurti du mažesni tekstynai: 25 mln. žodžių Mažasis lietuvių kalbos tekstynas (MLKT) ir 10 mln. žodžių Eksperimentinis tekstynas (ET), subalansuotas žanriniu požiūriu. Mažuoju lietuvių kalbos tekstynu remiamasi nustatant dažniausias žodžių formas, o šių žodžių formų pasiskirstymas įvairiuose žanruose tiriamas Eksperimentiniame tekstyne (plačiau šie tekstynai ir jų sandara aptariami 3.1.2 skyriuje ir 1-ajame priede).

Šio darbo metodika paremta empiriniais kalbos duomenimis, kurie analizuojami kiekybiškai (faktorių analizės metodu) ir kokybiškai (rezultatų interpretacija). Darbe naudojamos šios kompiuterinės programos: specialiai šiam darbui PERL'o kalba sukurtos programos (naudojamos kalbiniams požymiams skaičiuoti, sakiniams segmentuoti ir kt.), statistinis programinis paketas *SPSS for Windows* (taikomas visiems faktorių analizės etapams), programinis paketas *WordSmith Tools* (taikomas kai kuriems kalbiniams požymiams skaičiuoti ir konkordansams analizuoti), *Microsoft Excel* (taikomas duomenų grafikams kurti ir duomenims rūšiuoti bei skaičiuoti).

Darbo struktūra. Darbą sudaro šeši skyriai:

Pirmasis skyrius, „Įvadas“, skirtas darbo tikslams, uždaviniams, nagrinėjimai problematikai ir kitoms įvadinėms temoms;

Antrajame skyriuje, „Tekstų klasifikavimo problema“, aptariamos teorinės tekstų klasifikavimo problemos;

Trečiajame skyriuje, „Lietuvių kalbos tekstų funkcijų tyrimo metodologija“, aprašomi visi darbo tyrimo etapai,

Ketvirtasis skyrius, „Faktorių analizės taikymas tekstų funkcijoms tirti“, skirtas faktorių analizės etapams aprašyti

Penktajame skyriuje, „Lietuvių kalbos tekstų funkcijos“, analizuojami bei interpretuojami faktorių analizės rezultatai ir įvardijamos bei aprašomos septynios tekstų funkcijos ir jų paradigmos; paskutiniame,

Šeštajame skyriuje, „Išvados“, apibendrinami darbo rezultatai ir pateikiamos pagrindinės darbo išvados.

Pagrindiniai ginamieji teiginiai:

– Labai dažnų žodžių formų pasiskirstymas tekstuose yra reikšmingas tekstų funkcijų rodiklis.

– Faktorių analizės metodu analizuojant labai dažnų žodžių formų ir statistinių požymių pasiskirstymą didelę tekstų įvairovę apimančiame tekstyne, galima patikimai nustatyti labai dažnų žodžių formų ir statistinių požymių grupes (funkcines paradigmas), kuriomis reiškiamos tam tikros tekstų funkcijos.

– Remiantis funkcinių paradigmų pasiskirstymu tekstuose, galima automatiškai įvertinti tekstų prototipiškumą nustatytųjų funkcijų atžvilgiu.

DARBO REZULTATŲ APIBENDRINIMAS

Funkcijų tyrimas tekstuose leidžia daryti tokias pagrindines išvadas:

1. Tekstynų lingvistikoje egzistuojantis ryšys tarp kalbos kaip sistemos ir teksto kaip komunikacinio vieneto, yra apibendrinamas teiginiu jog autentiška kalba yra teksto kalba. Šiame darbe ši nuostata padeda skirti dvi sampratų grupes: viena jų, kuriai priklauso *funkcinis stilius* ir *registras,* susijusi su kalba kaip sistema, o kita, kuriai priklauso *individualusis stilius, žanras* ir *teksto tipas,* susijusi su tekstu kaip komunikaciniu vienetu. Todėl šiame tyrime apriorinė tekstų klasifikacija yra sudaroma žanriniu pagrindu, o aposteriorinė – funkciniu pagrindu. Pirmuoju atveju tekstai klasifikuojami pagal išorinius, o antruoju pagal eksperimentiškai nustatomus vidinius požymius.

2. Labai dažnų žodžių formų (*ldžf*) dažnumų pasiskirstymas tekstuose nėra chaotiškas ar atsitiktinis. Būdami dažniausi struktūriniai teksto vienetai, (*ldžf*) yra tiesiogiai susiję su teksto funkcijomis, todėl kartu su kitomis formaliomis teksto ypatybėmis jie yra reikšmingi teksto funkcijų rodikliai.

3. Didelę tekstų įvairovę apimančiam tekstynui sudaryti geriausiai tiktų keturių pakopų žanrinė klasifikacija: *kalbos atmaina*, *superžanras*, *žanras* ir *požanris*. Žanrinė

klasifikacija, kuri apima 2 kalbos atmainas, 8 superžanrus, 29 žanrus ir 17 požanrių, yra taikoma eksperimentiniam tekstynui (ET) sudaryti.

4. Specialiai šiam darbui sudarytame eksperimentiniame tekstyne faktorių analizės metodu buvo nustatyti septyni faktoriai ir jiems reikšmingi požymiai, kurių kokybinė interpretacija leido nustatyti, įvardinti ir aprašyti septynias tekstų funkcijas ir jų paradigmas: *spontaniško ekspresyvumo, naratyvumo, direktyvumo, nespontaniško ekspresyvumo, apeliatyvumo, deskriptyvumo ir temporatyvumo*.

4.1 ***Spontaniško ekspresyvumo*** paradigma, kuria reiškiama spontaniško ekspresyvumo funkcija, yra pati stipriausia, nes ją sudaro net 32 reikšmingi požymiai: iš jų 30 yra teigiami (*tai, jau, čia, ten, tas, ką, tą, tiek, taip, mes, labai, bet, dabar, aš, kaip, kur, nors, vis, tada, daug, nes, o, kai, kiek, to, buvo, gal, visi, jeigu, dar*) ir 2 neigiami (vidutinis žodžio ilgis, iteracijos indeksas).

Tekstai, kuriuose dominuoja spontaniško ekspresyvumo paradigma, yra iš anksto neparengti, spontaniški, turintys sakytinės kalbos bruožų, susiję su pasakojimo ar pokalbio situacija ir pabrėžtinai ekspresyvūs.

Prototipiškiausias superžanras šios paradigmos atžvilgiu yra *pokalbiai*, neprototipiškiausias – *oficialieji dokumentai*.

4.2 ***Naratyvumo paradigmą***, kuria yra reiškiama naratyvumo funkcija, sudaro 13 teigiamų požymių (*jis, ji, jos, jam, ją, jo, jį, ant, į, prie, savo, tu, vėl*).

Tekstams, kuriuose dominuoja naratyvumo paradigma, yra būdingas 3-ojo asmens įvardžių bei erdvės santykius reiškiančių prielinksnių dominavimas. Su šiais požymiais yra paprastai susijęs pasakojimas.

Prototipiškiausias superžanras šios paradigmos atžvilgiu yra *grožinė literatūra*, neprototipiškiausias – *dalykinės diskusijos*.

4.3. ***Direktyvumo paradigmą***, kuria reiškiama direktyvumo funkcija, sudaro 13 teigiamų (*būti, arba, gali, turi, ar, tam, jei, pagal, jeigu, kurios, galima, reikia, yra*) ir 2 neigiami (*buvo*, iteracijos indeksas) požymiai. Tarp reikšmingųjų požymių daug modalumą reiškiančių veiksmažodžių.

Tekstams, kuriuose dominuoja direktyvumo paradigma, būdinga modalumo raiška, esamasis laikas, alternatyvų ir sąlygų vardijimas.

Prototipiškiausias superžanras šios paradigmos atžvilgiu yra *oficialieji dokumentai*, neprototipiškiausias – *memuarinė literatūra*.

4.4 ***Nespontaniško ekspresyvumo paradigmą***, kuria reiškiama nespontaniško ekspresyvumo funkcija, sudaro 19 reikšmingų požymių: iš jų 16 teigiamų (*tik, o,* iteracijos indeksas, *ne, net, jei, nei, be, kas, gal, dar, kiek, todėl, ir, bet, reikia*) ir 3 neigiami (*dėl, mes, buvo*).

Tekstams, kuriuose dominuoja nespontaniško ekspresyvumo paradigma, būdingas ekspresyvus turtingas žodynas.

Prototipiškiausias šios paradigmos superžanras yra *grožinė literatūra*, neprototipiškiausias – *oficialieji dokumentai*.

4.5 ***Apeliatyvumo paradigmą***, kuria reiškiama apeliatyvumo funkcija, sudaro 12 teigiamų požymių (*kad, nėra, būtų, dėl, kurie, yra, mes, mūsų, todėl, nes, to, kuris*).

Tekstams, kuriuose dominuoja apeliatyvumo paradigma, būdingas argumentuotos kalbos vartojimas, stengiantis daryti poveikį klausytojui ar skaitytojui.

Prototipiškiausias šios paradigmos superžanras yra *dalykinės diskusijos*, neprototipiškiausias – *oficialiųjų dokumentų* superžanras.

4.6 ***Deskriptyvumo paradigmą***, kuria reiškiama deskriptyvumo funkcija, sudaro 16 požymių: iš jų 9 yra teigiami (*tarp*, sakinio ilgis, *galima, tačiau, tuo, jų,* iteracijos indeksas, *buvo, savo*) ir 7 neigiami (*nr, už, dėl, mano, man, aš, gal*).

Tekstams, kuriuose dominuoja deskriptyvumo paradigma, būdingi ilgi sakiniai, beasmenės konstrukcijos ir turtingesnis žodynas.

Prototipiškiausias šios paradigmos superžanras yra *akademinė proza*, neprototipiškiausias – *dalykinių diskusijų* superžanras.

4.7 ***Temporatyvumo paradigma*** yra pati silpniausia, nes ją sudaro tik 9 požymiai: iš jų 7 teigiami (*per, metų, po, prieš, bus, daugiau, iki*) ir 2 neigiami (*ir, jos*).

Tekstams, kuriuose dominuoja temporatyvumo paradigma, dominuoja laiką žyminčių žodžių gausesnis vartojamas. Nedidelis reikšmingų požymių kiekis neužtikrina patikimo tekstų įvertinimo, todėl ši paradigma laikytina nepatikimu temporatyvumo funkcijos rodikliu.

Prototipiškiausias šios paradigmos superžanras yra *spauda*, o neprototipiškiausias – *akademinė proza*.

5. Sukurta tekstų įvertinimo metodika pagal prototipiškumo laipsnį šešių paradigmų atžvilgiu (*spontaniško ekspresyvumo, naratyvumo, direktyvumo, nespontaniško ekspresyvumo, apeliatyvumo ir deskriptyvumo*). Bet koks tekstas gali būti įvertintas procentine išraiška kiek jis yra prototipiškas kiekvienos paradigmos atžvilgiu.

6. Atlikta tekstų funkcijų analizė dar kartą patvirtina tekstynų lingvistikos postuluojamą teksto formos ir funkcijos bei formos ir turinio vienovę, kuri šiame darbe išryškėja iš to, kad labai dažnų žodžių formų paradigmos gerai atspindi teksto funkcijas.

Publications on the topic of the dissertation

1. Utka, A. (1999) Ką reiškia *mintis* filosofinėje literatūroje? *Darbai ir dienos 10 (19):* 141-154.
2. Utka, A. (2000) Kalbinė įranga ir jos galimybės. *Darbai ir dienos 24:* 275-285.
3. Utka, A. (2005) Labai dažnų lietuvių kalbos žodžių ir žodžių formų ypatybės. *Lituanistica 1(61): 48-55.*

Publications on other topics

4. Danielsson, P. and A. Utka (2003) „Academic Research and Standards: a Discussion on Standards for Multilingual Language Resources“: *Corpus Linguistics* 2003, Lancaster University, March 28-31, *Workshop: Multilingual Corpora: Linguistic requirements and Technical Perspectives*, 35-42.
5. Utka, A. (2004) English-Lithuanian Phases of Translation Corpus: Compilation and Analysis. *IJCL 9:2: 195-224.*

About the author

In 1993–1999 Andrius Utka studied at Vytautas Magnus University, and in 199 he graduated with Masters' degree in English philology. Since 1998 until now he has worked at the Centre of Corpus Linguistics at Vytautas Magnus University. His duties included compiling *the Corpus of Contemporary Lithuanian Language* and parallel corpora, as well as participating in various linguistics projects. In 1999 he entered doctoral studies in philology at Vytautas Magnus University. During academic leave in 2000–2001 he studied and in 2002 graduated from the University of Birmingham with the Degree of Master of Philology in Corpus linguistics. In 2000–2003 he also worked at the Centre for Corpus Linguistics at the University of Birmingham, UK.

Andrius Utka

STATISTICAL IDENTIFICATION OF TEXT FUNCTIONS

daktaro disertacijos santrauka

Leidėjas – Vytauto Didžiojo universitetas (S.Daukanto g.28, Kaunas)
SL 1557. Užsakymo Nr. 37. Tiražas 40 egz. 2004 10 14
Nemokamai.