Corpus-assisted Discourse Analysis of New Digital Genres: A Study of Online Product Descriptions in Lithuania

Tekstynais paremtas naujų skaitmeninių žanrų tyrimas: produktų aprašymai Lietuvos gamintojų internetinėse svetainėse

LINGUISTICS / KALBOTYRA

Jūratė Ruzaitė

Vytauto Didžiojo universitetas, Lithuania

http://dx.doi.org/10.5755/j01.sal.0.35.22523

Abstract

The present paper relies on a corpus linguistic approach to reveal how the new digital genre of product descriptions is constructed and shaped by Lithuanian bread manufacturers. This study primarily focuses on the Multi-Dimensional analysis of bread descriptions as compared to other genres. It investigates the dominant verbal discourses of bread promotion and the typical linguistic choices used to disseminate these discourses. The data includes product descriptions available on the websites of six major industrial bread bakers in Lithuania and consists of more than 150 product descriptions (approx. 10,000 words). The lexical choices in these texts are evaluated by referring to the Corpus of Contemporary Lithuanian (tekstynas.vdu.lt) as a general reference corpus. The results show that in Lithuanian bread descriptions two text functions dominate: descriptiveness and prepared expressiveness, which further indicates that the target texts are closest to academic prose and fiction. The analysis of collocations and word frequencies supports the prevalence of features characteristic of both academic prose and fiction. This new digital genre is used by producers to disseminate and emphasise certain values and ideologies through explicit verbal references to tradition, inheritance, and continuity.

KEYWORDS: discourse, genre, corpus analysis, text functions, bread promotion, Lithuanian.

Introduction

The current paper takes a critical approach to the new genre of online product descriptions by applying the Multi-Dimensional (MD) analytical framework (Biber 1986, 1988). This study investigates how food producers construct this new digital genre to shape very specific product identities, which are influenced to a large extent by some dominant cultural values and food ideology in general. The primary focus here lies on the text functions of bread descriptions, the dominant verbal discourses of bread promotion in Lithuania, and the typical linguistic choices used to disseminate these discourses. The linguistic choices are expected to have some distinct genre-specific features. Thus, my main aim in this paper is to identify text functions and linguistic features that pertain to a new genre that recently appeared in digital

contexts. The secondary aim of the study is to give some insight to ideological implications hidden behind the dominant linguistic features in bread descriptions.

In the recent decade, there has been an increasing number of research studies focusing on food discourse along with the general interest in healthy food, nutrition issues, and the producers' efforts at effective food promotion (for a more extensive overview, see Ruzaitė, 2013). What is important is that in the digital era, online food promotion has offered a new platform for product promotion, including new genres and new affordances to product advertising (cf. Hopearuoho & Ventola, 2009). Thanks to digitalised food promotion, marketers have started developing a new discursive tradition and are extending and/or modifying "the printbased canon or traditional means of composing" (Hull & Nelson, 2005, p. 226). The new genres are marked by increasing fragmentation and individuation, lack of stability, merged and blurred genre conventions (Kress, 2004; Kress, 2003; Kress & van Leeuwen, 1996; Kress & van Leeuwen, 2002). New genres or modified conventional genres, however, still remain an under-researched area.

In the new era of digitalised promotion, adverts have appeared even for such staples as bread, which did not exist until the end of the twentieth century. Faced with the growing competitiveness within the saturated bread market in Lithuania (for nearly 3 million inhabitants, there are four big and 241 small and medium-sized bakeries), bread producers now provide product descriptions on their official websites. As will be shown in this study, these descriptions form a distinct genre (or sub-genre). They often imitate some other conventional genres, such as academic and fictional texts, and thus disseminate very specific narratives about bread and its producers.

Since the current paper will consider the form, content, and text functions of bread descriptions mainly from the genre variation perspective, this approach is addressed in the next section.

As has already been mentioned, to identify genre-specific features of different texts on the basis of smaller units, this study applies the Multi-Dimensional approach to genre variation (MD; also referred to as the Multifeature/Multi-dimension approach). This study also uses Utka's (2004; http://tekstynas.vdu.lt/) software developed for the Lithuanian language on the basis of the MD approach.

The MD approach, which was devised and developed by Biber (1986 and 1988), has several prominent characteristics. MD, as outlined by Biber (1993, p. 332), is:

1 Corpus-based (it depends on analysis of a large quantity of naturally occurring texts);

2 Computer-based (it depends on automated analyses of linguistic features in texts);

3 Used to analyse texts, genres, text types, styles or registers, rather than individual linguistic constructions;

4 Based on variationist and comparative perspectives (it assumes that different kinds of text differ linguistically and functionally, and more than one or two text varieties should be compared);

5 Explicitly multi-dimensional (multiple parameters of variation are operative in any discourse domain);

6 Quantitative (analyses are based on frequency counts of linguistic features, and multivariate statistical techniques are applied to study the relations among linguistic features and among texts);

7 Based on both quantitative and functional approaches (the statistical results help to determine the underlying communicative functions associated with each distributional pattern);

Multi-dimensional Approach to Genre Variation

8 Based on both macroscopic and microscopic approaches (macroscopic investigations of the overall parameters of linguistic variation, which are based "on analysis of the distribution of many linguistic features across many texts and genres, are complemented by detailed analyses of particular features in particular texts" (Biber, 1993, p. 332).

The key term in the MD approach is 'dimension', which is understood as a unit consisting of a group of frequently recurring linguistic features (Biber, 1993, p. 333). Importantly, the MD framework relies on a combination of multiple dimensions that help to distinguish genres, and a single dimension can never be sufficient. Co-occurring features are associated with social, situational and cognitive functions. In addition, as Biber notes, "dimensions are continuous scales of variation rather than dichotomous distinctions (1993, p. 333). Finally, "the co-occurrence patterns underlying dimensions are identified quantitatively (by a statistical factor analysis) rather than on an a priori functional basis" (Biber, 1993, p. 333).

For English, Biber (1988) identifies five major dimensions of variation: (1) Informational vs. Involved, (2) Narrative vs. Non-Narrative, (3) Elaborated vs. Situated Reference, (4) Overt Expression of Persuasion, and (5) Abstract vs. Non-Abstract Style. These dimensions are based on the co-occurrence distributions of 67 linguistic features across 481 spoken and written texts.

Utka's software was produced following the MD analytical framework to study Lithuanian texts and is grounded on automated factor analysis of a large variety of functionally related linguistic features which allow for analysing variation across genres (see Table 1). Linguistic features vary in their loading, or weight, which indicates the degree to which one can generalize from a certain factor to a linguistic feature (Biber, 1988). Following Biber (1993, p. 335), "features with higher loadings are … better representatives of the dimension underlying a factor".

Text factors can have positive loading if there is a positive correlation, and negative loading in case of negative correlation between the factor and register. The higher the value is, the more representative the feature is of a factor/dimension or register/genre. This way, the model takes into account both frequent and infrequent features in texts.

Table 1

Text functions and correlating features in Utka's (2004) model

Function	Dominant features	Most typical genres	Least typical genres
Spontaneous expressiveness	Spoken language features, spontaneity; deictic words; expressive language	Conversation	Official documents
Narrativeness	3rd person pronouns; spatial prepositions	Fiction	Subject-specific discussions
Prepared expressiveness	Particles; rich and expressive vocabulary	Fiction	Official documents
Persuasiveness	Argumentative style; impact on the hearer/reader	Official public discussions; spoken interaction	Official documents; fiction
Directiveness	Modals; present tense verbs; enumerations	Official and administrative texts; subject-specific discussions; academic prose	Fiction, memoirs, and conversations. Press is neutral.
Descriptivity	Long sentences; highly varied vocabulary; impersonal sentence structures	Academic prose (followed by prepared speeches)	Subject-specific discussions. Press is neutral.

The software can be used to process texts and automatically identify formal features, which are connected to text functions. By applying factor analysis, the software can automatically identify the main text functions in a Lithuanian text, their prominence in the text, and indicate which genre it is closest to (academic prose; planned speeches; press; biographical literature; official documents; fiction; conversations; or professional discussions) (Utka, 2004). Differently from other approaches, Utka's model determines text functions on the basis of frequency distributions of common words and word-forms in texts, and not distributions of content words or stylistically marked vocabulary (Utka, 2004).

Even though the software is handy to use and has a lot of potential for cross-generic analysis, it has not been applied in other research than Utka (2004) either to study conventional or new digital genres. To the best of my knowledge, the current paper is the first attempt to apply the model to compare genre-specific features of new digital and old conventional genres.

Further Implications of the MD Analysis

Though the primary aim of this research is to study genre variation, it also attempts to reveal some ideological implications of the results by pinpointing some strategies of bread manufacturers used to promote their products. Recently, a number of corpus-based discourse studies (e.g., Baker, 2005; 2006; Baker & Levon, 2015; Adolphs, 2006; Charteris-Black, 2007, Bednarek, 2010; Taylor, 2014; Al-Hejin, 2015; to mention but a few) have shown that such key aspects in corpus linguistics as frequency lists, key words, concordances, and patterns of some lexical items can be effectively used as evidence of discursively construed ideologies with regard to some sensitive or controversial groups, issues, or phenomena. A corpus analysis, based on representative amounts of data, can reveal explicit recurrent attitudes and 'uncover hidden patterns of language' (Baker, 2006, p. 19). Therefore, frequency counts and feature distributions that appear in the MDA analysis may also be indicative of some ideological aspects.

The term 'ideology' is a very broad concept used in a variety of disciplines and approaches, which, though often related to political stances, can also refer to "a set of beliefs held by an individual or group of people" (Adolphs, 2006, p. 84; for an overview of six main ways of how ideology can be defined, see Eagleton, 1991, pp. 28-30). For the purposes of this study, Eagleton's (1991) and Adolphs' (2006) definition of ideology has been adopted and is thus perceived "as a relatively neutral term that relates to beliefs and attitudes" (Adolphs, 2006, p. 84). Hence, similarly to Adolph's research on the issue of "genetics", in this paper I also focus on attitudes and beliefs in relation to the product identity shaped and framed by commercial producers on their official websites.

Ideology can be described by such notions as subjectivity, falsification, sharedness, and an identity constituent, but these are neither inherent features of ideology nor do they make up a finite list. Ideology, which is often communicated as universal truth, can sometimes be falsifying and negative, but is not inherently like that (Eagleton, 1991, p. 27; van Dijk, 1998, 2006). It is "a special form of social cognition shared by social groups" (van Dijk, 2006, p. 12) and as such is related to people's distinct worldviews and identities.

A corpus-based (or corpus-assisted) approach has been applied to study a variety of ideology-related issues, including stances in school books on physical and human geography (Stubbs, 1996), the speeches of Chris Patten (Flowerdew, 1997), the discourse of the British New Labour party (Fairclough, 2000), the image of gay men in public discourses (Baker, 2005), genetics in public discourses (Adolphs, 2006), character identities in fictional TV series (Bednarek, 2010), and Euroscepticism in Britain (Teubert, 2000). As can be seen even from this short list, the topic

range in such studies is vast and covers political issues, academic discursive practices, some scholarly and socially relevant issues, such as genetics, to mention just a few.

Discursive practices and their ideological implications in food discourse have been investigated perhaps most extensively by Cook and his project team (Cook et al., 2002, 2004a, 2004b), who studied GM food promotion. Online bread descriptions in Lithuania, as compared to the UK, have been analysed by Ruzaitė (2013). Her comparative study, which integrates multimodal discourse analysis and corpus linguistics, has shown that though bread promotion in these two countries incorporates some global aspects of food advertising (e.g. highlighting technological novelties in bread production), it clearly reflects some long-standing cultural myths about bread, especially in Lithuania, where there is a strong emphasis on tradition, inheritance, and continuity. The latter study, however, did not study text functions of bread descriptions, their generic features, or word frequency lists. It focused more on the multimodal aspects of these texts, their communicative functions, and took a more integrative approach to text analysis by trying to identify the dominant discourses, or narratives, about bread. The current study, though, will focus on the recurrent linguistic features that constitute the genre of bread descriptions and will further relate those features not only to the pertaining text functions, but also some ideological implications.

Data

The present study is based primarily on the Corpus of Online Bread Descriptions (COBreaD), which includes product descriptions that were available on the websites of 10 major industrial bread bakers in Lithuania in 2014. The database contains more than 230 product descriptions (or almost 12,000 words; for more detailed statistical information, see Table 2). The database was limited to the descriptions of bread and bakery products. Cakes and pastry products were excluded, since the latter products are promoted through rather different discourses. It also needs to be noted that COBreaD represents products of industrial bakers, not bakers who would produce hand-made products, which will be an important fact when interpreting the research results.

Table 2

Statistical parameters of the data

Parameter	Result
Total number of tokens	11,865
Number of sentences	1,087
Average sentence length	10.92
Average word length	6.42
Number of types	2,356
TTR	19.86
STTR (per 500 words)	45.3

The data in COBreaD was processed using AntConc 3.2.4w. Text functions in the online bread descriptions were identified by employing the software developed by Utka (2004) following Biber's model (http://tekstynas.vdu.lt/).

To highlight what is typical of bread discourse but is less characteristic of general Lithuanian or some specific genres, the Corpus of Contemporary Lithuanian Language (CCLL; 140,921,288 words; available at tekstynas.vdu.lt) was used as a reference corpus. The corpus represents both written and spoken language. As can be seen in Table 1, it mainly comprises written texts, which include fiction, academic texts, administrative texts, publicist texts, and a segment of spoken language.

Discourse type	Texts included	Number of words
Fiction (FICT)	Novels; short stories; drama; tales	15,765,554
Academic texts (ACAD)	Scientific & educational texts; research papers; popular scientific texts	19,322,341
Administrative texts (ADMIN)	State (and church) documents; shorthand reports of LT Parliament sessions	13,625,715
Publicist texts (PUBL)	Memoirs, chronicles, essays; general, popular, & scientific periodicals	86,497,837
Spoken texts (SPOK)	Public and private communication	447,396
TOTAL:		**140,921,288**

Table 3
Composition of CCLL

Where relevant, word frequencies obtained from COBreaD are compared to the frequencies of target words in all the five sections in CCLL on the basis of relative frequency per 1000 words.

In the current paper, corpus analytical techniques will be used, first of all, to identify text functions in the target texts, which will further show which genre(s) bread descriptions are closest to. Second, frequency lists, concordance outputs, and collocates of some lexical items will be studied at the micro level to further detail out attitudes and beliefs about bread. Where relevant, wordlist information is compared to the reference corpus.

As the results of automated factor analysis indicate, neither of the six functions discussed in the previous section is positively representative of bread descriptions: all the values of the indicative features are negative (see Figure 1). The least negative results, though, are those for prepared expressiveness and descriptiveness, which can thus be said to be the most characteristic functions of bread descriptions. Therefore, they will be discussed here in greater detail, but it is also important to outline at least shortly what is especially non-characteristic of this genre.

Text Functions of Online Bread Descriptions
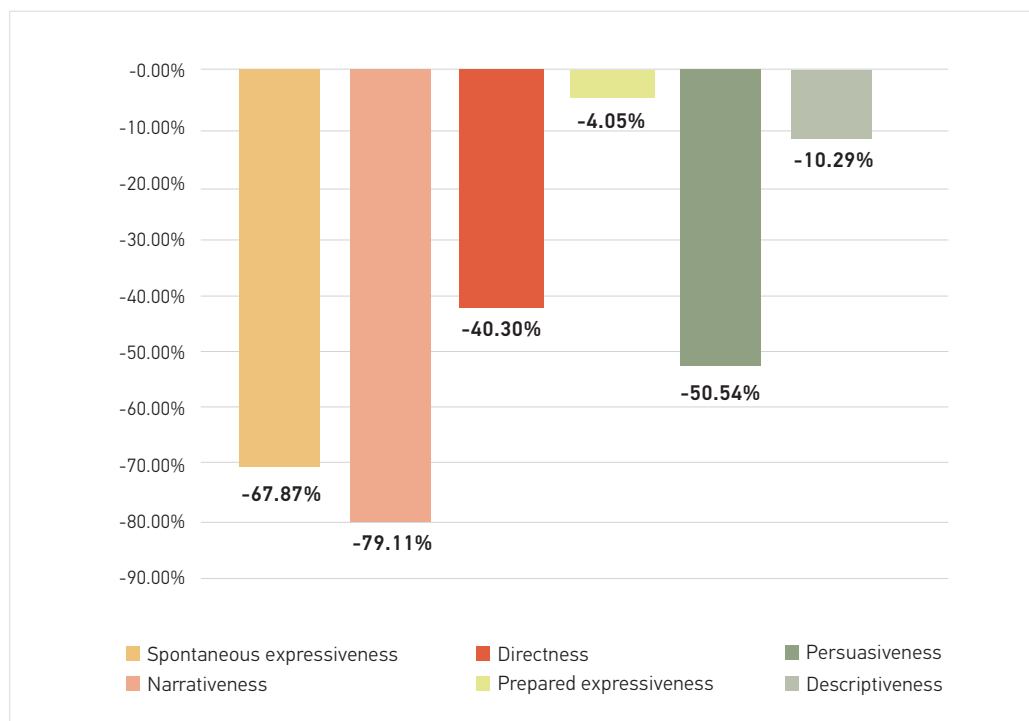
Fig. 1
Text functions of on-line bread descriptions

The function that is least represented in the current data is that of narrativeness (−79.11%), which means that in bread descriptions such features as 3rd person pronouns and spatial prepositions are scarce. As could be seen in Table 1, they are mainly encountered in fiction (in story-telling) but are not typical of subject-specific discussions. It is natural that in product promotion narrativeness is low, and they resemble more of subject-specific discussions.

The next least representative function is spontaneous expressiveness (−67.87%), which is typical of conversation. This finding suggests a low frequency of spoken language features, spontaneity, deictic words, and expressive language in bread descriptions, which is not surprising either, since expressiveness in written texts is carefully planned.

Persuasiveness, which correlates with argumentative style and impact on the hearer/reader, is not typical either (−50.54%). It is representative of official public discussions and spoken interaction in general and thus is not characteristic of written bread descriptions which are, regarding this function, closer to fiction.

Directiveness correlates with modals, present tense verbs, and enumerations and thus might be expected in promotional texts, but in bread descriptions it is not typical (−40.3%). This function is typical of official and administrative texts, subject-specific discussions, and academic prose. It is least characteristic of fiction, which suggests again that bread descriptions are similar to fiction.

The most representative function in bread descriptions is **prepared, or non-spontaneous, expressiveness** (−4.05%). In general, it is characteristic of texts, which contain a high frequency of particles, and rich and expressive vocabulary. It is most typical of fictional texts, expressive written language in general, and is least typical of official documents. There are 16 positive and 3 negative features that correlate with this function; Table 4 provides items that are indicative of prepared expressiveness in COBread.

Table 4

Features correlating with prepared expressiveness (freq. per 1,000)

Positive features		dar ('yet')	1.01
tik ('only')	7	kiek ('how many')	0.17
o ('and')	1.85	todėl ('so')	2.44
TTR	45.3	ir ('and')	41.47
ne ('no')	1.6	bet ('but')	1.52
net ('even')	1.1	reikia ('need')	0.17
jei ('if')	0.25		
nei ('neither')	0.84	Negative features	
be ('without')	8.26	dėl ('for')	0.84
kas ('what/who')	0.51	mes ('we')	0.17
gal ('maybe')	0.08	buvo ('was/were')	0.59

The second most typical function is that of **descriptiveness** (−10.29%). It is characteristic of texts which contain long sentences (the average sentence length in COBreaD is 10.92 words), rich vocabulary (TTR = 45.3), and impersonal constructions. This function is typical of academic prose texts, and least typical of professional discussions. As can be seen in Table 5, there are 9 positive and 7 negative features correlating with this function, but not all of them are encountered in bread descriptions.

| Positive features		Negative features	
tarp ('between')	0	nr ('No.')	0.08
avg. s. l.	10.92	dėl ('for, due to')	0.84
galima ('possible')	0.17	už ('for, by, over')	0.51
tačiau ('however')	0.42	mano ('my')	0
tuo ('with that')	0.34	man ('for me')	0
jų ('theirs')	0.42	aš ('I')	0
TTR	45.3	gal ('maybe')	0.08
buvo ('was/were')	0.59		
savo ('my, his, her, its')	1.85		

Table 5
Features correlating with descriptiveness (freq. per 1,000)

The category that does not occur in the current texts is the preposition *tarp* ('between'), which in fact is a positive feature and should appear in texts where descriptiveness is prominent. The 1st person pronoun (*aš* 'I') and its different forms (*mano* 'my', and *man* 'for me') are negative features, and, as expected in this model of function identification, they do not appear at all.

Bread descriptions are lexically highly varied, as the TTR (defined as the relationship between the number of types and the number of tokens) indicates. The more types there are in comparison to the number of tokens, the more varied is the vocabulary, i.e. there is greater lexical variety and less repetition.

To sum up, in terms of text functions, bread descriptions are closest to two genres: fiction (mainly in terms of prepared expressiveness) and academic prose (mainly in terms of descriptiveness). Further, the paper discusses the lexical choices and argumentative strategies at the microlevel by taking into account dominant lexical/semantic categories, some keywords, their frequency, and collocations.

In order to uncover the identity of bread products shaped by the producers, first the lemma DUONA ('bread') was analysed. The collocates of the lemma were identified with AntConc and further manually categorised into semantic sets represented in Figure 2. The frequency counts here and elsewhere include the frequency of tokens, not types. Where the strength of collocates is assessed, the measure of T-score is used (as calculated with AntConc). A T-score is statistically significant if it equals 2 or is higher.
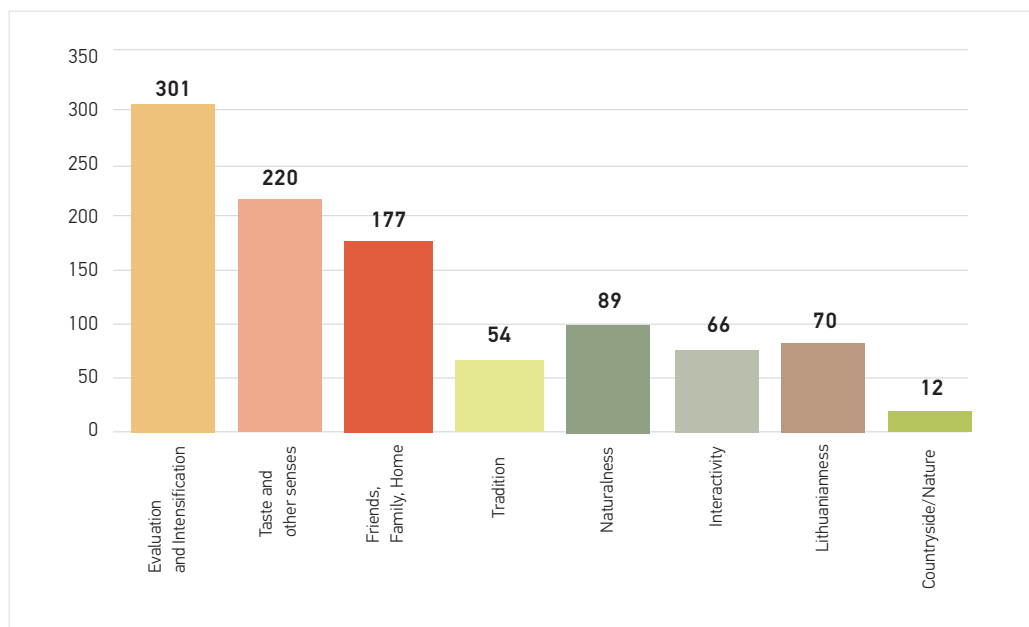
Sets of Collocates of the Lemma DUONA ('bread')

Fig. 2
Sets of collocates of the lemma DUONA ('bread')

The largest set of collocates includes units used for evaluation and intensification (301 to-kens). This finding relates well with the text function of prepared expressiveness so charac-teristic of bread descriptions.

Some main categories in this set include intensifiers, superlatives, and overgeneralisations. The most frequent intensifier is *tik* ('only'; freq. 43 tokens; T-score = 355.126), as in example (1).

1. ...o gamybos procese naudojamos **tik** natūralios žaliavos.
   '...and in the process of production **only** natural ingredients are used.'

The most frequent superlative is *aukščiausios* ('highest'; freq. 7 tokens; T-score = 186.747), which is typically used to refer to the outstanding quality of ingredients, as in example (2).

2. ...gamybai naudojami **tik aukščiausios** kokybės ingredientai.
   '...for its production **only** the ingredients of the **highest** quality are used.'

Overgeneralizations are typically realised with universal pronouns, such as *vis** (derivatives of 'all'; freq. 8 tokens). As in example (3), such unduly exaggerated claims are presented as universal truths, which clash with objectivity and are indicative of ideological implications.

3. Mūsų **visų** pamėgti batonai – naujose pakuotėse.
   'The baguettes loved by **all** of us – in new packages.'

The second largest set of collocates includes sensual and tactile imagery, namely words re-ferring to different senses: taste, smell, touching, vision (Table 6), and hearing (much less common and thus not represented in the table below). Here and elsewhere, the table indicates how often the collocate appears to the left (Freq. (L)) and to the right (Freq. (R)) of the node.

Table 6

References to senses

Rank	Freq.	Freq. (L)	Freq. (R)	T-score	Collocate	EN
Taste						
54	16	5	11	359.027	pagardinta	'enriched with'
69	13	5	8	319.762	skonį	'taste' (sg. Acc.)
93	10	5	5	305.597	sodraus	'rich' (sg. Gen.)
767	1	0	1	0.91595	sodresnį	'richer' (sg. Acc. comp.)
119	8	3	5	263.528	skonis	'taste' (sg. Nom.)
176	6	2	4	229.509	gardi	'delicious' (fem.sg.Nom.)
238	4	2	2	195.798	skoninių	'of taste'
240	4	2	2	195.798	skaniausių	'tastiest' (pl. Gen.)
328	3	0	3	127.107	skoniu	'taste' (Instr.)
342	3	0	3	156.222	paskaninta	'enriched with'
Aroma						
80	12	8	4	328.213	kvapni	'fragrant'
208	5	3	2	208.572	kvepianti	'fragrant'
222	5	1	4	184.141	aromato	'aroma' (sg. Gen.)
223	5	4	1	212.331	aromatinga	'aromatic'
Other senses						
40	21	6	15	429.830	minkšta	'soft' (fem.)
61	14	5	9	351.703	švelnaus	'delicate' (Gen.)
63	1	1	0	0.83191	spalvą	'colour' (sg. Acc.)

Again, this category can be directly related to the high dominance of prepared expressiveness observed in the analysis of text functions.

One of the most dominant discourses of bread is the discourse of bread as a family product typically associated with home (cf. Ruzaitė, 2015). This is clearly reflected in the frequency of references to family and especially home in the current study (see Table 7).

Rank	Freq.	Freq. (L)	Freq. (R)	T-Score	Collocate
151	6	3	3	234.655	šeimai
55	16	12	4	386.342	namiškių
87	11	7	4	322.793	naminės
122	8	7	1	273.928	naminė
504	2	2	0	129.535	naminę
953	1	0	1	0.87393	namų
954	1	0	1	0.57977	namus
955	1	0	1	0.95798	namuose
956	1	0	1	0.91595	namiškius
957	1	1	0	0.87393	namie
958	1	0	1	0.95798	namais
Total freq.: 49					

Table 7

Frequency of references to home and family (*nam* / šeim**)

The importance of tradition in relation to bread is manifested in the frequency of collocates, which are derivatives of *tradicija* ('tradition'; freq. 44 tokens) and derivatives of *senas* ('old'; freq. 9 tokens), as in examples (4)–(5).

4. Ruošiama **tradiciniu** lietuvišku būdu, nenaudojant konservantų...
   'Prepared in the **traditional** Lithuanian way without any preservatives...'

5. Iš močiučių perėmėme **senąsias** duonos kepimo technologijas, todėl „**Senolių**" duona...
   'From our grandmothers we inherited the **old** bread baking technologies; therefore, "**Ancestors'**" bread...'

A related category of collocates includes references to Lithuania (70 tokens) and countryside (12 tokens), which suggests that bread is often represented by the producers as an iconic national product produced by following rural traditions and authentic technologies:

6. ...kuri gimtųjų namų pasiilgusiems primins tikros **lietuviškos** duonos skonį...
   '...which to those missing home will remind of the taste of the real Lithuanian bread...'

7. ...ilgai išlieka šviežia, lyg tik ką ištraukta iš karštos **kaimiškos** krosnies.
   '...stays long fresh, as if just taken out of the hot **rustic** stove.'

Here a clear contradiction can be seen between the nature of producers represented in the data (industrial manufacturers) and their claims about ancient technologies, which are possible only if producers are non-industrial.

When frequent items in the wordlist were compared to the results in the general reference corpus CCLL, it appeared that some frequent items in bread descriptions only very infrequently occur or do not occur at all in other genres. As Table 8 demonstrates, none of the special terms frequently used in bread descriptions is used in spoken discourse, which is not a surprising finding, since these terms form the special jargon of bread descriptions and have features of academic jargon used for special purposes. The meaning and usage of these words are very specific, and semantically they may not be transparent to laypeople.

Lexical Choices in Bread Descriptions as Compared to Other Genres

Table 8

Distribution of specialist terms in COBreaD and CCLL

	COBreaD (Raw)	COBreaD (Raw) COBreaD f/1000	Fiction f/1000	ACAD books f/1000	Admin docs f/1000	Publicist discourse f/1000	SPOK f/1000	Total CCLL f/1000
konservantų 'preservatives'	90	7.59	0.0005	0.0002	0	0.002	0	0.001
riebalų 'fats'	51	4.3	0.01	0.02	0.01	0.02	0	0.02
mielės 'yeast'	51	4.3	0.0006	0.001	0.003	0.001	0	0.001
angliavandenių 'carbohydrates'	51	4.3	0.002	0.01	0.0002	0.01	0	0.01
energinė 'energy'	47	3.96	0	0.001	0	0.0003	0	0.0003
baltymų 'protein'	47	3.96	0.006	0.04	0.001	0.02	0	0.02
kcal	47	3.96	0	0.003	0	0.002	0	0.002
kj	46	3.88	0	0.003	0	0.002	0	0.001
priedų 'additives'	38	3.2	0.003	0.01	0.03	0.01	0	0.01

In none of the genres in CCLL, is the frequency of these terms as high as their frequency in bread descriptions. Their frequency is especially low in fiction and administrative texts. The other terms restricted to the area of bread production include references to the type of flour and other ingredients of bread, e.g., *pilno grūdo* ('wholemeal') and *salyklinių kviečių dribsnių* ('malted wheat flakes').

Some terms, namely the words for 'fats', 'carbohydrates', and 'additives', occur more frequently in academic and publicist discourse, which supports that these items are part of academic jargon. This also reinforces the earlier observation that, in terms of dominant text functions, bread descriptions are close to academic prose. It can also be seen in Table 8 that *priedų* ('additives') are relatively frequent in administrative documents, but here this polysemous word is used with the meaning of 'annexes', not 'additives'.

The distribution of references to taste and quality (provided in Table 9) shows again that some terms (*tinka* 'suits', *natūralų* 'natural', and *skonį* 'taste') are more typical of academic texts than any other genres. The items *skonį* ('taste') and *puikiai* ('perfectly') are most frequent in fiction.

Table 9

Distribution of references to taste and quality in COBreaD and CCL

	COBreaD (Raw)	COBreaD Per 1000	Fiction Per 1000	ACAD books Per 1000	Admin docs Per 1000	Publicist discourse Per 1000	SPOK Per 1000	Total CCLL Per 000
skonio 'of...taste'	76	6.41	0.02	0.03	0	0.02	0.02	0.02
tinka 'suits'	46	3.88	0.05	0.15	0.02	0.02	0.09	0.07
pagardinta 'enriched'	38	3.2	0.0005	0	0	0	0	0.0001
natūralų 'natural'	34	2.87	0.004	0.01	0.003	0.005	0	0.01
skonį 'taste' (Acc.)	33	2.78	0.03	0.02	0.0005	0.02	0.02	0.02
puikiai 'perfectly'	32	2.7	0.19	0.09	0.035	0.12	0.12	0.11

These findings support the dubious nature of bread descriptions: on the one hand, they resemble academic texts, and on the other hand, they are close to fiction. That is, they contain both expressive and lexically rich vocabulary and special terms resembling of academic jargon.

So far, the analysis has focused on the surface features of bread descriptions without considering broader textual context, or co-text. This section will shortly discuss how different discursive features discussed above are intertwined in bread descriptions to form the intricate ideologically loaded texture (for a more detailed analysis of dominant bread discourses, see Ruzaitė, 2012, 2013).

Manual content analysis of bread descriptions has revealed that they often resort to health discourse, which tends to be combined with the discourse of 'science' (e.g. 'useful for the human body', as in example (8)). Such descriptions suggest that behind bread production there is scientific grounding. This argumentation is achieved through special terminology (e.g. names of different types of grains) used to refer to the beneficial nutritional qualities of bread, as in example (8).

> **8.** 5 grūdų batonas – tai tamsus batonas, išsiskiriantis ypač turtinga receptūra. Šio batono sudėtyje yra ne tik kvietinių, bet ir ruginių miltų, todėl jis yra tamsus, tačiau priešingai nei ruginė duona – purus ir lengvas. Turtinga receptūra sukurta naudojant unikalų grūdų ir sėklų mišinį, kurį sudaro linai, aguonos, soros grūdai, saulėgrąžos ir sezamo sėklos. Visi šie grūdai ir sėklos yra **labai naudingi žmogaus organizmui**.

> 'The 5-grain baguette is a dark baguette with particularly rich ingredients. This baguette contains not only wheat, but also rye flour, so it is dark, but unlike rye bread, it is fluffy and light. The rich recipe was made using a unique mixture of cereals and seeds, consisting of flax, poppy seeds, millet grains, sunflower seeds and sesame seeds. All these grains and seeds are **very useful for the human body**.'

Here the description contains features of typical academic argumentation: benefits of the recipe are listed as arguments used to ground the conclusion (highlighted in bold) and to construct credibility.

Most commonly bread descriptions integrate several discourses (as in example (9)): tradition and authenticity (underlined), academic discourse and terminology (bold), naturalness (double underline), and an appeal to senses (dashed underline).

> **9. Ajerų duona**
> Ši ant ajerų kepama senovinės formos tamsi duona išsiskiria karališkai švelniu saldoku skoniu, kuris persipina su ryškiu ruginio salyklo aromatu. Ruošiama su **plikiniu** ir **natūraliu raugu**, be jokių **konservantų** ir **maisto priedų**. Ajerais kvepianti duona ilgai išlieka šviežia, lyg tik ką ištraukta iš karštos kaimiškos krosnies.

> **'Calamus bread'**
> 'This dark bread of an old traditional shape is baked on calamus leaves and is distinct for its royal slightly sweet taste with the pronounced flavour of rye malt. The bread is made from **scalded and natural sourdough** and **does not contain any preservatives or additives**. This bread having the smell of calamus stays fresh for a long time and tastes as if it has just been taken out of the hot rustic bread oven.'

The form of bread descriptions, which is highly limited in length, predetermines the density of discourses and shapes them as succinct lexically rich texts with an interplay between a variety of discourses, which form a dense texture.

Conclusion

Shared (cultural) beliefs about food products, such as bread, are well reflected in producers' product descriptions, which form a specific product identity and disseminate it using the affordances of new digital genres. In the genre of online product descriptions, features of two genres, fiction and academic prose, are merged and fused to achieve the appellative effect. This also helps to shape the identity of the producer as a reinforcer of ancient cultural traditions and 'hand-made' technologies, and a responsible perpetrator of national heritage, even if this producer is an industrial manufacturer (cf. similar observations by Weiss, 2004, p. 48 in relation to food packaging of kosher food).

The results related to text functions have disclosed a rather paradoxical nature of bread descriptions: prepared expressiveness (a prototypical function of fiction) and descriptiveness (a prototypical function of academic texts) are contrastive at least to some extent. The former is more associated with subjectivity and emotionality, whereas the latter is associated with objectivity and neutrality. This way, bread descriptions combine features of 'reason ads' and 'tickle ads', to use Cook's (2001, p. 15) classification, where tickle ads appeal to emotion, and reason ads rely on objective argumentation (cf. Ruzaitė, 2013).

The analysis of wordlists, collocations, and concordance lines shows that Lithuanian bread discourse disseminates and emphasises certain values and ideologies through explicit verbal references to exceptional quality, health, tradition, inheritance, and continuity, which are further reinforced and elaborated on through non-linguistic semiotic choices (see Ruzaitė 2012). Similar appeals to shared values, including family values, and the importance of good impressions are typical of probiotic web advertising in Koteyko and Nerlich (2007), who also show that probiotic web advertising extensively employs the discourse of health and science. A high degree of expressiveness and exaggerations has been observed in GM food promotion by Cook and his project team (Cook et al., 2002, 2004a, and 2004b).

As the findings in the current study demonstrate, online bread descriptions use some distinct lexis, part of which is typical of academic prose and/or fiction, but some lexical items appear only in bread descriptions. This implies that they are developing as a distinct genre (or perhaps a sub-genre) with some vocabulary of restricted usage. Some dominant lexical choices (e.g. *pagardinta* 'enriched with') can be hypothesised, though, to be typical of food discourse in general, but this needs further research. An analysis of a broader variety of food products could reveal which of the aspects identified in this study pertain to food discourse in general, and whether there are any features characteristic only of bread discourse. It could also be revealing to further compare descriptions of staple products to luxury products (e.g. chocolate) to identify if, how, and to what extent the discourses shaping identities of different products vary.

References

1   Adolphs, S., 2006. Introducing Electronic Text Analysis: A Practical Guide for Language and Literary Studies. London: Routledge. https://doi.org/10.4324/9780203087701

2   Al-Hejin, B., 2015. Covering Muslim Women: Semantic Macrostructures in BBC News. In: Discourse and Communication, no 9(1), pp. 19-46. https://doi.org/10.1177/1750481314555262

3   Baker, P., 2005. Public Discourses of Gay Men. London: Routledge. https://doi.org/10.4324/9780203643532

4   Baker, P., & Levon, E., 2015. Picking the Right Cherries? A Comparison of Corpus-Based and Qualitative Analyses of News Articles about Masculinity. In: Discourse & Communication, no 9(2), pp. 1-16. https://doi.org/10.1177/1750481314568542

5   Bednarek, M., 2010. The Language of Fictional Television: Drama and Identity. London, New York: Continuum.

6   Biber, D., 1986. Spoken and Written Textual Dimensions in English: Resolving the Contradic-

tory Findings. Language, no 62, pp. 384-414. https://doi.org/10.2307/414678

7   Biber, D., 1988. Variation across Speech and Writing. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511621024

8   CCL-Corpus of Contemporary Lithuanian. www.tekstynas.vdu.lt

9   Charteris-Black, J., 2004. Corpus Approaches to Critical Metaphor Analysis. London: Palgrave Mac-Millan. https://doi.org/10.1057/9780230000612

10   Cook, G., 2001. The Discourse of Advertising. 2nd edition. London: Routledge. https://doi.org/10.4324/9780203978153

11   Cook, G., 2004. Genetically Modified Language. London: Routledge. https://doi.org/10.4324/9780203561232

12   Cook, G., Pieri, E., & Robbins, P. T., 2002. The Discourse of the GM Food Debate: How Language Choices Affect Public Trust. ESRC Research Grant. Final Report, Award number RES-000-22-0132. http://www.regard.ac.uk/.

13   Cook, G., Pieri, E., & Robbins, P. T., 2004a. The Presentation of GM Crop Research to Non-specialists: A Case Study. ESRC Research Grant. Final Report. Award number R000223725. http://www.regard.ac.uk/.

14   Cook, G., Pieri, E., & Robbins, P. T., 2004b. The Scientists Think and the Public Feels: Expert Perceptions of the Discourse of GM Food. In: Discourse and Society, no 15(4), pp. 433-449. https://doi.org/10.1177/0957926504043708

15   de Beaugrande, R., 1999. Discourse Studies and Ideology: On 'Liberalism' and 'Liberalisation' in Three Large Corpora of English'. In: Discourse Studies, no 1(3), pp. 259-95. https://doi.org/10.1177/1461445699001003001

16   Eagleton, T., 1991. Ideology: An Introduction. London, New York: Verso.

17   Fairclough, N., 2000. New Labour, New Language? London: Routledge.

18   Flowerdew, J., 1997. Competing Public Discourses in Transitional Hong Kong. In: Journal of Pragmatics, no 28, pp. 533-53. https://doi.org/10.1016/S0378-2166(97)00033-7

19   Hopearuoho, A., & and Ventola, E., 2009. Multisemiotic Marketing and Advertising. In:

The World Told and the World Shown, (eds.) Ventola, E., & Guijjaro, A. J. M., pp. 183-204, Basingstoke: Palgrave Macmillan.

20   Hull, G. A., & Nelson, M. E. 2005. Locating the Semiotic Power of Multimodality. In: Written Communication, no 22, pp. 224-261. https://doi.org/10.1177/0741088304274170

21   Kress, G., 2004. 'Reading Images: Multimodality, Representation and New Media.' http://www.knowledgepresentation.org/BuildingTheFuture/Kress2/Kress2.html. https://doi.org/10.1075/idjdd.12.2.03kre

22   Kress, G. R., 2003. Literacy in the New Media Age. London: Routledge Falmer. https://doi.org/10.4324/9780203299234

23   Kress, G. R. & van Leeuwen, T., 1996. Reading Images: The Grammar of Graphic Design. London: Routledge.

24   Kress, G. R. & Van Leeuwen, T., 2002. Multimodal Discourse: The Modes and Media of Contemporary Communication. London: Edward Arnold.

25   Ruzaitė, J., 2012. The Discourse of Food Promotion: A Case Study of Lithuanian Bread Packages. In: Multiple Perspectives in Linguistic Research on Baltic Languages, (eds.) Usonienė, A., Nau, N., & Dabašinskienė, I., pp. 104-136, Cambridge: Cambridge Scholars.

26   Ruzaitė, J., 2013. What is a Culinary Crime? A Study of Online Bread Promotion in Lithuania and the UK. In: Anglistics in Lithuania: Cross-Linguistic and Cross-Cultural Aspects of Study, (eds.) Šeškauskienė, I., & Grigaliūnienė, J., pp. 28-63, Cambridge: Cambridge Scholars.

27   Stubbs, M., 1996. Text and Corpus Analysis. Oxford: Blackwell.

28   Teubert, W., 2000. "A Province of a Federal Superstate, Ruled by an Unelected Bureaucracy": Keywords of the Euro-sceptic Discourse in Britain'. In: Attitudes Towards Europe, (eds.) Good, C., Musolff, A., Points, P., & Wittlinger, R., pp. 45-88. Abingdon: Ashgate. https://doi.org/10.4324/9781315096742-3

29   Taylor, C. (2014) Investigating the representation of migrants in the UK and Italian press: A cross-linguistic corpus-assisted discourse analysis. International Journal of Corpus

Linguistics, 19(3), 368-400. https://doi.org/10.1075/ijcl.19.3.03tay

30 Utka, A., 2004. Statistinis tekstų funkcijų nustatymas. Unpublished PhD Thesis. Kaunas: Vytautas Magnus University.

31 van Dijk, T. A., 2006. Discourse, Ideology and Context. In: Folia Linguistica, no XXXV/1-2, pp. 11-40

32 van Dijk, T. A., 1998. Ideology: A multidisciplinary approach. London: Sage Publications Ltd.

33 Weiss, E., 2004. Packaging Jewishness: Novelty and Tradition in Kosher Food Packaging. In: Design Issues, no 20(1), pp. 48-61. https://doi.org/10.1162/07479360477293376

Santrauka

Jūratė Ruzaitė. Tekstynais paremtas naujų skaitmeninių žanrų tyrimas: produktų aprašymai Lietuvos gamintojų internetinėse svetainėse

Šio darbo tikslas – remiantis tekstynų lingvistikos prieiga atskleisti, kaip Lietuvos duonos gamintojai konstruoja palyginti naujai atsiradusį skaitmeninį žanrą – gaminamų produktų aprašymus internetinėse gamintojų svetainėse. Taikant Andriaus Utkos sukurtą automatinio teksto funkcijų nustatymo programą, šiuo tyrimu siekiama parodyti, kuo ir kiek šis žanras yra panašus į penkis šios programos identifikuojamus superžanrus. Remiantis statistinėmis analizuojamų tekstų charakteristikomis, dažniniais žodžių sąrašais ir dažniausių žodžių analize kontekste, aptariami vyraujantys duonos reklamos diskursai bei gamintojų kuriama produktų tapatybė.

Tyrimui naudotus duomenis sudaro duonos gaminių aprašymai, surinkti iš šešių pagrindinių pramoninių duonos kepėjų internetinių svetainių. Duomenų bazę sudaro daugiau nei 150 produktų aprašymų arba apie 10000 žodžių. Siekiant nustatyti išskirtinius tiriamojo žanro ypatumus, duonos aprašymų leksika vertinama lyginant jos vartojimą su Dabartinės lietuvių kalbos tekstyno (tekstynas.vdu.lt) duomenimis. Gauti rezultatai rodo, kad lietuviškuose duonos aprašymuose vyrauja dvi teksto funkcijos: deskriptyvumas ir nespontaniškas ekspresyvumas, kas rodo, jog tirtieji tekstai turi daugiausiai akademinės prozos ir grožinės literatūros bruožų. Kolokacijų ir žodžių dažnumo duonos aprašymuose analizė taip pat patvirtina, kad juose dažnai vartojami akademinei prozai ir grožinei literatūrai būdingi leksiniai ypatumai. Šį naują skaitmeninį žanrą gamintojai naudoja tam tikroms vertybėms ir ideologijoms skleisti ir dažniausiai pabrėžia tokius sociokultūrinius aspektus kaip tradicijos, tautinis paveldas ir tęstinumas.

About the Author

Jūratė Ruzaitė

Professor at the Department of Foreign Language, Literary and Translation Studies at Vytautas Magnus University

Research interests

(Multimodal) discourse analysis, academic discourse, sociolinguistics, and corpus linguistics

Address

Centre of Intercultural Communication and Multilingualism, Vytautas Magnus University, K. Donelaičio g. 58, LT-44248 Kaunas, Lithuania

E-mail

jurate.ruzaite@vdu.lt