

©Copyright 2013
Varvara Gracheva

Markers of contrast in Russian:
A corpus-based study

Varvara Gracheva

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

Master of Arts

University of Washington

2013

Committee:

Emily M. Bender

Barbara Citko

Program Authorized to Offer Degree:

Linguistics

TABLE OF CONTENTS

Chapter 1:	INTRODUCTION	2
Chapter 2:	METHODOLOGY	4
2.1	Gathering Data	4
2.2	Processing Data	6
2.3	Annotating Data	7
Chapter 3:	LITERATURE REVIEW	9
3.1	Introduction	9
3.2	Information structure in general	9
3.2.1	Definition of information structure	9
3.2.2	Components of information structure	10
3.2.2.1	Focus	12
3.2.2.2	Topic	14
3.2.2.3	Background	18
3.2.2.4	Contrast	19
3.2.2.4.1	Definition of Contrast	19
3.2.2.4.2	Contrastive and non-contrastive foci	21
3.2.2.4.3	Contrastive and non-contrastive topics	27
3.3	Information Structure in Russian	29
3.3.1	General	29
3.3.2	Word Order	30
3.3.3	Sentence/Emphatic Stress	32
3.3.4	Clefting	32
3.3.5	Clitics as Information Structure markers	33
3.3.5.1	Introduction	33
3.3.5.2	Clitics -TO and ŽE	34
3.3.5.2.1	Clitic -TO	35
3.3.5.2.2	Clitic ŽE	41
3.3.5.3	Clitic LI	47
3.4	Summary	48
Chapter 4:	ANALYSIS	49
4.1	Introduction	49
4.2	Applying Contrastiveness Tests to the Spoken Data	49
4.3	Contrastive Focus vs. Contrastive Topic	52
4.4	Summary of tests and assumptions necessary for analysis	55
4.5	CLITICS	60
4.5.1	CLITIC -TO	60
4.5.1.1	-TO as a marker of contrastive topic and contrastive focus	60
4.5.1.2	Different roles of -TO	65
4.5.1.3	-TO attaching to different parts of speech	67
4.5.1.4	-TO in interrogatives	73
4.5.1.4.1	-TO in <i>yes-no</i> questions	74
4.5.1.4.2	-TO in <i>wh</i> -questions	76
4.5.1.5	-TO in sentences without topic	81
4.5.1.6	Interaction with other lexical elements	82
4.5.1.7	-TO: SUMMARY & CONCLUSION	83
4.5.2	Clitic ŽE	85
4.5.2.1	Introduction	85
4.5.2.2	Clitic ŽE and information structure marking	86

4.5.2.3	Focus projection of <i>ŽE</i> and its position in the sentence	88
4.5.2.4	Position of contrastively focused element relative to phrasal <i>ŽE</i>	94
4.5.2.5	Clitic <i>ŽE</i> attaching to different types of words	95
4.5.2.6	Different information structure patterns marked by <i>ŽE</i>	96
4.5.2.6.1	Clitic <i>ŽE</i> attaching to deictics	96
4.5.2.6.2	Clitic <i>ŽE</i> attaching to <i>wh</i> -words	98
4.5.2.6.3	<i>Wh</i> -words + <i>ŽE</i> : Questions	99
4.5.2.6.4	<i>Wh</i> -words + <i>ŽE</i> : Declaratives	105
4.5.2.6.5	<i>Wh</i> -words + <i>ŽE</i> : Part of expression	107
4.5.2.7	<i>ŽE</i> : Summary	108
4.5.3	Clitic <i>LI</i>	109
4.5.3.1	<i>LI</i> as a Contrastive Focus Marker	109
4.5.3.2	When <i>LI</i> is not a (contrastive) focus marker	114
4.5.3.3	<i>LI</i> : Summary	115
4.5.4	Clefting	116
4.5.4.1	Clefting as signaling contrastive focus	116
4.5.4.2	Clefting: Summary	122
4.6	Summary	122
Chapter 5:	IMPLEMENTATION	124
5.1	Introduction	124
5.2	Background	124
5.2.1	The Grammar Matrix and HPSG	124
5.2.2	ICONS	126
5.3	Word order	127
5.4	<i>LI</i> implementation	128
5.5	- <i>TO</i> implementation	130
5.6	<i>ŽE</i> implementation	135
5.7	<i>It</i> -clefting implementation	135
Chapter 6:	CONCLUSION	137
6.1	Summary	137
6.2	Future Work	139
	BIBLIOGRAPHY	141
	Appendix A: Annotated sentences	146
	Appendix B: Diagnostics/analysis of information structure components	149

ACKNOWLEDGMENTS

I would first like to express my gratitude to my advisor Professor Emily Bender for introducing me to the study of information structure that finally led me to writing Master's thesis on this topic. I am deeply grateful to Professor Bender for all of her advice, ideas and constructive suggestions, for patiently reading many drafts of my thesis, as well as for always being there for all of her students, including myself. The completion of this thesis would not have been possible without her continuous guidance, encouragement, and support throughout my degree.

I would also like to thank the second reader Professor Barbara Citko for reading my thesis within very tight deadlines and providing helpful resources and suggestions on recent theory on information structure in Russian.

I appreciate the help of all of my student colleagues with whom I worked closely throughout my degree. I am grateful to Sanghoun Song for many discussions and sharing ideas on information structure, as well as for his tremendous help with implementation of information structure into Russian Grammar based on ICONS. Quantitative processing of the corpus data would not have been possible without Jeff Kahn, who helped me to identify and correct the errors in my code, as well as explained some of the main concepts in Python. Ryan Cross and Mike Goodman shared their expertise on text encoding. I am also grateful to the Grammar Matrix development group (particularly to Mike Goodman, Laurie Poulson, Antske Fokkens, and Joshua Crowgey). Lastly, I would like to acknowledge Esad Suskic for his collaboration in the grammar engineering seminar on developing Russian grammar that has been used in this thesis.

I owe a debt of gratitude to the other faculty in the program, particularly to Professor Zagana and Professor Aldridge who taught my first classes in syntax and made me even more fascinated with the study of linguistics.

I am also very grateful to Joyce Parvi for her tremendous help and guiding me through the process of application and obtaining my degree from the day I applied through the day of my graduation.

Lastly, I would like to thank my family and three native Russian speakers that have participated in the few elicitation experiments that were carried out for this thesis: Irina Cooley, Yevgeniy Grachev, and Dmitry Sharkov.

Chapter 1: INTRODUCTION

This thesis concentrates on one of the components of information structure, namely **contrast**. **Contrast** is often referred to as implying a set with limited number of members, one of which is highlighted or contrasted to other members. **Contrast**, like other components of information structure, can be marked in different ways. I am particularly interested in **contrastive clitics** in Russian, i.e. the elements that signal presence of the contrast in the sentence.

Clitics *ŽE* and *–TO* have been analyzed as adding emphasis or marking contrast in the sentence (McCoy 2001). In this thesis I further explore McCoy’s analysis of *ŽE* and *–TO* as markers of contrast. Using Russian National Corpus (Grishina 2006) available online, I explore different types of data (i.e. varying in illocutionary force, syntactic and semantic context, parts of speech to which clitics attach, etc.) to establish whether McCoy’s analysis of *ŽE* and *–TO* always applies or whether there are exceptions. Since the other goal of this thesis is to determine whether clitics *ŽE* and *–TO* can be implemented in Russian grammar within the HPSG framework, I try to examine the data with anticipation of later implementation. While understanding that there are exceptional cases, I try to account for the most frequent usage of the clitics. I later extend the analysis of clitics *ŽE* and *–TO* as marking contrastive focus to clitic *LI* and *it*-clefting, suggesting that these information structure markers traditionally analyzed as marking focus (King 1995), can be marking contrastive focus as well.

In chapter 2 I describe my methodology in gathering, processing, and annotating data necessary for analysis of information structure patterns in Russian. Chapter 3 provides a review of the literature on information structure in general (section 4.1) and information structure marking strategies in Russian, including word order, intonation, clefting, and use of clitics (section 4.2). In chapter 4 I propose my analysis of the data found in Russian National Corpus, concentrating on *–TO* (section 5.3.1), *ŽE* (section 5.3.2), and *LI* (section 5.3.3) as clitics marking contrast in the sentence, as well as *it*-cleft construction as marking contrast (section 5.3.4). In chapter 5 I start with providing a short description of the Russian Grammar (2010) that was created based on the Grammar

Matrix customization system (Bender et al. (2010) and Bender, Flickinger and Oepen (2002)) within the HPSG framework (Head-driven Phrase Structure, described in Pollard and Sag (1994), Sag et al. (2003), and Copestake et al. (2005)). I also review ICONS, i.e. constraints used in representation of information structure in MRS proposed by Song & Bender (2012) and Song (forthcoming). Then a proposal is made on how to implement earlier discussed contrast markers in LinGO Grammar Matrix based grammar. Finally, a summary, concluding remarks, and possible directions for future work are in Chapter 6.

Chapter 2: METHODOLOGY

2.1 Gathering Data

Most of the data used in this study is from the online collection of written and spoken Russian over 300 million words (Russian National Corpus). For this study, only the **spoken** data was used, which constitutes 3.9% of the entire corpus. The spoken data (referred to as “oral presentation” in the Russian corpus) consists of recordings of public speech (64.3%) and spontaneous speech (8.1%), as well as the transcripts of Russian movies (27.6%). The spoken data recordings/transcripts cover different time periods (1930 to 2007) and different geographic regions (Moscow, Saint Petersburg, Saratov, Ulianovsk, Taganrog, Ekaterinburg, etc.). In May 2013, when the data was accessed for this thesis, the spoken data in Russian National Corpus contained 3,253 documents, 1,636,928 sentences, and 10,361,479 words (Russian National Corpus). The search function in spoken corpus allows user to search the corpus by one or more words. A search for one word returns all contexts (term used in Russian National corpus for short chunks of data containing the exact form of the word). A search for more than one word returns all contexts containing these words in a sentence, but not necessarily in the same order in which it was entered in search. Each sentence in the corpus is provided with minimal context, which sometimes implies just two additional sentences, one preceding and one following the sentence. Prosodic information is sometimes included in the Russian National Corpus data. Although the non-spoken data would have been easier to process due to presence of punctuation, compared to the spoken data, which often contains large chunks of text without punctuation, there were reasons why it was the spoken data that was chosen for this thesis.

Firstly, the spoken data provides a wider range of different syntactic, morphological, and other strategies marking information structure. The reason for this is the “emotiveness” that can be present in the spoken data and is absent in the written data. Yokoyama (1986) notes that emotive sentences have sentence stress, which is an important marker of information structure in Russian. King (1995) notes that while in **non-emotive** sentences that do not have sentence stress, theme precedes the rheme, in

emotive speech the rheme (focus) precedes the theme (topic). Such variations of the word order (both emotive and non-emotive) are more visible in the spoken data.

Secondly, since a large part of this thesis deals with clitics, it was necessary to choose the data that would contain a number of clitics high enough to get a good understanding of their usage and properties. Spoken data appears to contain more clitics than written data in Russian. McCoy alludes to a study by Prokurovskaja (1977) that showed that the “the frequency of particle usage in colloquial speech has been calculated as 11.6%, compared to 1.6% of particle use in scientific genres of standard Russian” (McCoy 2001: 9-10). I observed the same pattern, i.e. higher frequency of clitics usage in spoken speech as opposed to the non-spoken speech, in the Russian National Corpus data. I isolated fiction, non-fiction, and the spoken data subsets that were comparable in size to test the frequency of clitics usage. Below is the table demonstrating the number of occurrences of three clitics *TO*, *ŽE*, and *VED’*, steadily increasing from top to bottom, i.e. from non-spoken non-fiction to fiction to the spoken data.

Table 1: Usage of clitics –*TO*, *ŽE*, *VED’* in spoken vs. non-spoken speech in the Russian National Corpus: frequency of clitic occurrence per 10,000 words

Type of data	<i>ŽE</i> occurrences	<i>VED’</i> occurrences	<i>TO</i> occurrences ¹
Non-spoken non-fiction	3.16	0.61	2.83
Non-spoken fiction	3.95	0.92	2.91
Spoken	4.16	0.99	4.38

As can be seen from the table above, the frequency of clitic usage was the lowest in the non-spoken non-fiction data, which is the written collection of official and business communication, journalism data, advertising, electronic communication, etc. The frequency of clitic usage slightly increased in the non-spoken non-fiction data, which is a collection of fiction works (adventure, drama, crime, children stories, etc.). The

¹ It should be noted that the numbers for –*TO* contain numbers for this particle written both with hyphen (–*TO*) and without the hyphen (*TO*). It did not seem necessary to isolate the numbers for –*TO* (with the hyphen), as both of them are clitics and both of them will be discussed in the section on Analysis.

frequency of clitic usage was the highest in the spoken data, which consists of public speech, spontaneous speech, and movie recordings.

2.2 Processing Data

Having access to such abundant data in the Russian National Corpus makes it possible to gain a better understanding of how clitics are used in Russian spoken data, which position they occupy in the sentence, and to which elements they usually attach.

The data used in this thesis was collected in two steps. First, I used the search function on the Russian National Corpus website, allowing me to narrow down data to only the chunks of text that contained the clitics. Then I wrote a Python program allowing me to isolate sentences containing clitics and then to process these sentences separately. Processing the sentences with clitics implied counting the frequency of clitic usage (compared to overall number of sentences), as well as identifying patterns of clitics' usage, i.e. their position in the sentences, preceding words, etc. When working with individual sentences, I was also able to retrieve their context from Russian National Corpus in order to examine their information structure patterns within their context. Examining each individual sentence within a larger context, I applied various tests and descriptions of different components of information structure in order to identify them in the sentence.

It should be noted that the data from the Russian National Corpus was noisy and hard to process. For example, finding the position of a word in the sentence could not be completely accurate due to the form of the data, which often lacked punctuation marks. Instead forward slashes “/” are used to replace all punctuation marks except for periods (.) or (...), question mark (?), and exclamation mark (!)² (Grishina 2006). Therefore, some of the very high numbers received in calculating the position of the clitic in the sentence are not accurate in terms of understanding its real position in the sentence. However, they are not surprising as they were obtained from the large chunks of data without punctuation.

² It should be noted that in some large chunks of texts these punctuation marks were missing as well, which made it impossible to identify breaks beginning and end of sentences.

It should also be noted that in addition to processing the data in terms of clitics usage, it was also challenging to analyze these sentences overall. Since the transcripts of the spoken data do not include complete (or often any) prosodic information, it was often difficult to understand the information structure patterns of the sentences, since speakers often crucially rely on prosody to convey the meaning. Additionally, due to the informal nature of spontaneous speech characterized by abrupt sentences and having very limited context provided by corpus search function, it was also challenging to understand the meaning of the utterances, or test sentences for contrastiveness, as well as to provide glosses/translations in English for this thesis.

2.3 Annotating Data

Data from Russian National Corpus was annotated using working³ information structure guidelines developed earlier for BEBO project (Gracheva & Song 2012), based on annotation guidelines developed by Dipper et al. (2007) and extended for annotation of data containing dialogues.

Since the Russian National Corpus provided context for each utterance, i.e. several utterances surrounding the utterance under consideration, context was taken into consideration during annotation. Having access to context surrounding each sentence allowed for a more accurate interpretation of the information structure patterns of the sentences. For all sentences, tests to identify contrast, focus and topic were applied in order to determine information structure components in the sentence. These tests and how they applied to the spoken data from Russian National Corpus are discussed in greater detail in Chapter 5 with Analysis. EXMARaLDA software that was developed at the Research Centre on Multilingualism at the University of Hamburg (Schmidt 2012) was used for the annotation, with files represented in XML format. Sentences have been annotated for POS (part of speech), gloss, topic, focus, contrast, and prosody. An example of a sentence from Russian National Corpus annotated in EXMARaLDA is on the next page:

³ These guidelines were not finalized. However, I am directing reader to these guidelines as the main explanation/documentation of how we annotated text for information structure.

Figure 2.3: Annotation sample: ‘It (was) Boris (who) drank vodka’

	0	1	2	3	4
TXT	Это	Борис	выпил	водку	.
[POS]	cleft	N	V	N	
[GLOSS]	it	Boris	drank	vodka	
[TOPIC]					
[FOCUS]		F			
[CONTRAST]		CF			
[PROSODY]		*			

In the sentence above *Борис* /‘Boris’ is a focused element that is also contrastive, identified as CF (contrastive focus) in the [Contrast] tier. The [Prosody] tier indicates that ‘Boris’ also appears to receive sentential stress. For more examples of annotated sentences please refer to Appendix A.

Chapter 3: LITERATURE REVIEW

3.1 Introduction

In this chapter I will provide a review of literature on information structure theory. I will start by describing the study of information structure in general in section 3.2. I will then proceed to review the components of information structure, i.e. focus, topic, background, and contrast. I will also discuss their subcategories, i.e. contrastive and non-contrastive foci, as well as contrastive and non-contrastive topics. Additionally, I will review tests to identify the above-mentioned information structure components. In section 3.3 I will review several studies on information structure in Russian. I will briefly mention recent accounts on information structure strategies used in Russian, i.e. word order, intonation, clefting, and clitics. I will finish with a more detailed review of the work published on clitics –*TO*, *ŽE*, and *LI*.

3.2 Information structure in general

3.2.1 Definition of information structure

The study of Information Structure lies at the intersection of syntax, semantics, pragmatics, and phonetics. As Lambrecht notes, there is enormous variation in the terminology and definitions that different authors use when writing about information structure. Vallduví & Vilkuna (1998) refer to information structure theory as a “terminology minefield.” The terms and definitions range from “functional sentence perspective (used by Prague School of Linguistics), to “information packaging” (Chafe 1976), “discourse pragmatics” and “informatics” (Vallduví 1990), and, finally, to “information structure” (Lambrecht 1996: 2). This thesis adopts the term **information structure** and its definition proposed by Lambrecht in (1):

- 1) INFORMATION STRUCTURE: That component of sentence grammar in which propositions as conceptual representations of states of affairs are paired with lexicogrammatical structures in accordance with the mental states of

interlocutors who use and interpret these structures as units of information in given discourse contexts.

(Lambrecht 1996: 5).

Lambrecht emphasizes the pairing of propositions with lexicogrammatical structures by concentrating on the “structural implications of discourse-pragmatic analysis,” in which only those psychological phenomena matter that have correlates in grammatical form (Lambrecht 1996: 2-3).

In his analysis of information structure and how it is expressed in the grammatical structure of the sentence, Lambrecht points out the multiplicity of different layers of grammar, i.e. syntax, semantics, pragmatics, etc., as well as their interdependency. Lambrecht emphasizes that the most promising approach to grammatical analysis presupposes seeing different layers of grammar not as “hierarchically organized independent subsystems but as interdependent forces competing with each other for the limited coding possibilities offered by the structure of the sentence” (Lambrecht 1996: 12). Lambrecht is not the only author emphasizing the interdependency between different layers of information structure. Büring also emphasizes the interdependency between prosody, syntax, and pragmatics, noting that “information structure is an aspect of syntactic representation, which interfaces with the phonological form by rules of IS realization, and receives its meaning via rules of IS interpretation” (Büring 2005: 1). As will be shown later in Chapters 4 and 5, the interaction of different layers of grammar, i.e. syntactic means (such as word order or clefting or presence of clitics) and phonological means (intonation) can be crucial in determining information structure of sentences in Russian. In Russian these layers can either work together or override one another (e.g. intonation overriding word order), creating different information structure interpretations of sentences.

3.2.2 Components of information structure

Having briefly touched on the interaction of information structure with syntactic, phonological, semantic, and pragmatic layers in the previous section, I now turn to the description of the main three components of information structure, namely **topic**, **focus**,

and **background**. The fourth component that will be examined in this thesis is **contrast**, which can either be a *contrastive focus* or a *contrastive topic*.

It should be noted that while this thesis concentrates on **topic**, **focus**, and **background** as the three components of information structure, this is not the only possible division of information structure. Several ways of subdividing information structure have been proposed. One of these approaches, described by King in her overview of existing analyses of Russian sentences, is based on a two-way distinction into **theme** and **rheme** (King 1995, Krylova & Khavronina 1988). The other approaches are based on three-way division of the sentence. One of such approaches, proposed by King, divides the sentence into **focused**, **topicalized** (referred to as ‘topic’ in this thesis), and **discourse-neutral** information (King 1995). The terms that will be used in this thesis are based on Dipper et al.’s work (2007): **focus**, **topic**, and **background**. Another three-way sentence division approach was proposed by Engdahl and Vallduví, i.e. the division of the sentence into **focus** and ground, further subdivided into **link**, and **tail** (Engdahl & Vallduví 1994). Even though King (1995) and Engdahl & Vallduví (1994) use similar three-way divisions, their terminology, as well as their reasoning behind the division of sentences into focus, background (tail), and topic (link), and their understanding of the relationship between these components of information structure are different. Engdahl & Vallduví define *focus* as the “actual update potential of a sentence,” while for King it is “new information, relative to some state in the discourse,” with “new” described as “relative to what the speaker believed the addressee knew prior to the relevant utterance” (King 1995: 70-71). While these notions do not seem to be that different from each other, King and Engdahl & Vallduví differ much more in their approach to topic. Engdahl & Vallduví describe *ground* (*link* + *tail*) in terms of its relationship to *focus*, i.e. for them *ground* is “how the information update is to be carried out” (Engdahl & Vallduví 1994: 14). Meanwhile, King describes *topic* in terms of the “shared matter of current concern” (King 1995: 67). Similar to Engdahl & Vallduví (1994), Bailyn also defines topic and focus in terms of their relationship, using the definition from Gundel and Fretheim (2001): “Topic is what the sentence is about; focus is what is predicated about the topic” (Bailyn 2011: 265). In her recent work on information structure in reference to Russian Slioussar also attempts to capture the relationship between components of information

structure by proposing a relational model of information structure. She proposes to identify topic and focus using the scales of accessibility and salience, with topic being the most accessible element (more so than focus) and focus being the most salient element (more so than topic) (Slioussar 2007). Slioussar's approach will be briefly discussed later in the sections on topic and focus accordingly.

As to the necessary presence of topic/focus in a sentence, the opinion on that varies as well. Engdahl & Vallduví argues that all sentences must contain a **focus**, while not all need to have a **background** or a **topic** (Engdahl & Vallduví 1994). Meanwhile, Büring, concentrating on the phonological form of sentences, notes that there are some sentences in English, which contain only B-accent, characteristic of topic, which might mean that there are focus-less sentences (Büring 2005: 16). Following Engdahl & Vallduví (1994), I am going to assume for this thesis that sentences always have focus, but not necessarily topic.

The means of coding information structure vary across the languages. In her research on information structure in thirty languages, Gundel finds the syntactic structure to be the “most frequent (and the only universal) means of coding the topic-comment relation across languages” (Gundel 1988, 223). The three subcategories of information structure, i.e. **topic**, **focus**, and **background**, as well as the means of coding them across languages, will be discussed in detail in the sections below.

3.2.2.1 Focus

Focus has been defined in various ways. For example, what Vallduví refers to as **focus**, has often been called *rheme*. **Focus** has also been analyzed as a syntactic part of a pragmatic **comment** (Gundel 1988). Emphasizing the syntactic nature of the focus vs. semantic nature of the comment, Gundel suggests that while “comment” refers to the speaker's accessibility of information, “focus” represents its syntactic counterpart, i.e. the part of the sentence that encodes the comment (Gundel 1988, 211).

Without going into a detailed discussion of focus' syntactic and semantic components, Dipper et al. refer to **focus** as “that part of an expression which provides the

most relevant information in a particular context as opposed to the (not so relevant) rest of information making up the background of the utterance” (Dipper et al., 2007: 170). Krifka (2007) defines focus in terms of presence of alternatives. Adopting Rooth’s (1985) definition of **focus**, he expands it to reflect the nature of these alternatives in (2) below:

- 2) A property F of an expression α is a Focus property iff F signals (a) that alternatives of (parts of) the expression α or (b) alternatives of the denotation of (parts of) α are relevant for the interpretation of α .

(Krifka 2007: 8)

Krifka’s approach to focus in terms of set of alternatives will be discussed later in the section comparing focus and contrast.

Unlike Dipper et al. (2007) who define focus in terms of the *relevancy* of information, King describes focus in terms of *newness* of information. In her discussion on the division of sentences into two parts, i.e. theme and rheme, King roughly defines rheme (focus) as “new information, relative to some state in the discourse” (King 1995: 70). Lambrecht also defines focus to be the new/most important element in the sentence (Lambrecht 1996). Slioussar defines focus in terms of *salience*, with focus as associated with new information being the most salient element on the salience scale. Since focus is often associated with newness, Slioussar does not separate these notions in her analysis (Slioussar 2007: 2007). King’s (1995) and Lambrecht’s (1996) definition of focus, i.e. in terms of its *newness* relative to the discourse, is the definition adopted in this thesis.

There are also different variations on how **focus** can be further subdivided. For example, Gundel distinguishes three types of focus: **psychological focus** (assigned according to the salience of the entity within the discourse), **semantic focus** (similar to new-information focus discussed later and answering the *wh*-question), and **contrastive focus**, all of which can overlap, but which are not equivalent (Gundel 1999).

Languages use different strategies for marking focus. They vary from syntactic marking to lexical and prosodical means. For example, one of the main strategies for marking focus is its position in the sentence. King argues that in the theme/rheme division of the sentence, rheme usually follows the theme in non-emotive speech (King 1995: 2). Additionally, focus can be expressed by the presence of lexical markers or

phonologically. As Büring notes, prominence (signifying ‘focus’) in a sentence can be characterized by pitch accent, loudness, duration, and changes in formant structure (Büring, 2005, p. 3). In English focus is usually marked with the A-accent (Jackendoff 1972).

In this thesis a two-way focus division approach will be used. Two types of focus analyzed are: **semantic** (also referred to as *new-information* or *non-contrastive* focus) focus and **contrastive** focus, distinguished by the nature of the set, from which the focus is being chosen, corresponding to Gundel’s notion of contrastive focus. Contrastive and semantic foci will be discussed in more detail in the section 4.1.2.4.2 on contrastive and non-contrastive foci, as well as in Chapter 5 describing contrast in Russian.

3.2.2.2 Topic

Different terms have been used to describe the notion of topic, including **topic** (Gundel 1988, King 1995), **theme**, and **link** (Engdahl & Vallduví 1994).⁴ Following Gundel (1988) and King (1995) the term that will be used in this thesis is **topic**.

Topic has been defined in terms of *accessibility* (Slioussar 2007) and in terms of *givenness*, i.e. as given information (King 1995: 2). Gundel has described **topic** also in terms of previous knowledge or familiarity with the entity (see (3) below):

- 3) An entity, E, can successfully serve as a topic, T, iff, both speaker and addressee have previous knowledge of or familiarity with E.

(Gundel 1988: 212).

Krifka (2007) proposes a different definition of **topic** in terms of its relation to comment (focus) with the file-card like structure to store information in (4) below:

⁴ Vallduví divides sentences into a **Focal Part** and a **Ground**, further subdividing the latter into two subcategories, **LINK** and **TAIL** (Engdahl & Vallduví 1994: 44).

- 4) The topic constituent identifies the entity or set of entities under which the information expressed in the comment constituent should be stored in the C(ommon)G(round) content.

(Krifka 2007: 41)

The notion of CG (Common Ground) used above refers to the information “mutually known to be shared and continuously modified in communication” (Krifka 2007: 4). Although the entity denoted by topic is usually described as “given” information and inferable from the context, Krifka mentions that topics can also be new entities as well, demonstrating this with the example (5):

- 5) [A good friend of mine]_{Topic} [married Britney Spears last year]_{Comment}.

According to Krifka, this sentence introduces a new entity in the discourse, uses it to denote the topic, which in Krifka’s terms equals to introducing a new file card in the Common Ground content (Krifka 2007). Krifka’s (2007) description of this type of topics in terms of newness contradicts King’s (1995) and Gundel’s (1988) defining the topic in terms of givenness. Krifka’s definitions of topic will be adopted and discussed in more detail later in this thesis.

Topic has also been defined in terms of *aboutness*, i.e. Lambrecht defines the topic to be what the proposition of the sentence is **about** (Lambrecht 1996: 118). In this thesis Lambrecht’s definition of **topic** will be used, or what the sentence is *about*, for aboutness topics. Since Lambrecht’s definition does not account for frame-setting topics (as will be shown later in this section), Krifka’s description of topic will be used as well, aiding both in accounting for frame-setting topic, as well as in differentiating between contrastive topics and contrastive foci.

One of the tests to identify a topic in the sentence is the ‘tell-me-about’ test (Choi 1999, also used in Song & Bender 2012), demonstrated in (6) below:

6) Speaker 1:

Расскажи мне про своего отца.
Tell 1SG.DAT about your father
‘Tell me about your father.’

Speaker 2:

Мой отец – инженер
My father engineer
‘_T My father] is an engineer.’

In the example above, the response of Speaker 2 is *about* their father, responding to the ‘tell-me-about’ prompt by Speaker 1. The constituent *мой отец* / ‘my father’ passes the ‘tell-me-about’ test and can be considered to be a topic in this sentence.

The topic role can be indicated by different means varying across languages, including lexical, syntactic, and phonological marking. For example, Japanese uses lexical items to mark information structure, i.e. *–wa* marker for topic (Heycock 1994). Regarding prosodic marking of topic, it is unclear whether topics are marked phonologically across all languages. Büring (1999) argues that like focus, topic is realized phonologically in the form of a specific pitch accent. However, Vallduví argues that in languages like Catalan, links (topics in this thesis) “display no particular intonational prominence” (Engdahl & Vallduví 1994: 47). In English topic is usually marked with the B-accent (Jackendoff 1972).

Topics can be contrastive and non-contrastive. Non-contrastive topics are divided into **frame-setting topics** and **aboutness topics**. Even though the topic is generally defined in terms of what the sentence is ‘about’, topic can be further split into **aboutness topic**, **frame-setting topic**, and **contrastive topic**.

Aboutness topics are the entities about which the sentence makes a predication (Dipper et al. 2007: 163). Dipper et al. note that they are usually fronted crosslinguistically and only the following can be aboutness topics: referential NPs, indefinite NPs with specific and generic interpretations, bare plurals with generic interpretations, and some finite clauses denoting concrete facts (Dipper et al. 2007: 163). An example of an aboutness topic is below:

7) (Context: ‘Today I met Lida in the park.’)

Лида работает журналисткой в газете «Ленинградская Правда».
Lida works journalist-INST in newspaper Leningradskaja Pravda.
‘Lida works as a journalist for newspaper *Pravda*.’
[_{AB-TOPIC} Lida] works as a journalist for newspaper *Pravda*.

In the example above the sentence is about woman named Lida that the speaker met in the park. The referential NP *Лиды*/'Lida' is the aboutness topic in this context.

Lambrecht's definition of topics in terms of 'aboutness' (1996) and Choi's 'tell-me-about' test (1999) capture characteristics of aboutness topics.

Frame-setting topic denotes the frame in which the main predication of the sentence is interpreted. Frame-setting topics usually describe the time or the location of the event (Dipper et al. 2007: 167). An example of a frame-setting topic is below:

8) Speaker 1:

Куда вы завтра поедете?
Where you tomorrow go
'Where 2PL.NOM you going tomorrow?'

Speaker 2:

Завтра мы поедem в Санкт-Петербург.
Tomorrow 1PL.NOM going in Saint Petersburg
'Tomorrow we are going to Saint Petersburg.'
[FS-TOPIC Tomorrow] we are going [SF to Saint Petersburg].

In the example above the temporal expression *завтра*/'tomorrow' describes the frame within which the predication *мы поедem в Санкт-Петербург*/'we are going to Saint Petersburg' should be interpreted. The temporal expression *завтра*/'tomorrow' is a frame-setting topic in the example above. It is not clear how frame-setting topic can be accounted for by Lambrecht's definition of topic in terms of 'aboutness' (1996), but it can be accounted for by Krifka's (2007) definition of topic.

Since aboutness topic and frame-setting topic are subtypes of topic, both of them should be defined in terms of what the sentence is about. However, based on the definitions and examples for aboutness and frame-setting topics above, it appears that 'aboutness' topics fit the general definition of all topics in terms of 'aboutness' better than the frame-setting topics. Therefore, it seems that either a different general definition for 'topic' is necessary (not defining topic in terms of 'aboutness') or the definition of 'frame-setting' topics should incorporate 'aboutness' as one of its characteristics. This matter deserves further consideration, but for now Dipper et al.'s (2007) definitions for aboutness and frame-setting topics will be adopted for this thesis.

Contrastive topics will be discussed as subtypes of aboutness and frame-setting topics in more detail in the section 4.1.2.4.3 on contrastive and non-contrastive topics later.

3.2.2.3 Background

The third component of information structure, i.e. **background**, also has a variety of terms used to describe it. Some of the terms include *background* (Büring 2005), *ground* (Engdahl & Vallduví 1994: 44) or *discourse neutral material* (King 1995: 2). One of the definitions for **background** comes from Büring, stating that everything that is not focus is background. An example of a background in the sentence is below:

- 9) **Speaker 1:**
Кто играет на пианино?
who plays on piano
'Who is playing on piano?'

Speaker 2:
Иришка играет.
Irishka plays
'Irishka is playing on piano.'
[_F Irishka] is playing.

In the sentence above *узграем*/'is playing' is the background, or the discourse-neutral information in the sentence.

Like Büring, Vallduví divides sentences into similar segments, i.e. a Focal Part and a Ground, which is further divided into two subcategories, LINK (topic in this thesis) and TAIL (background in this thesis) (Engdahl & Vallduví 1994: 44).

In this thesis I am going to assume that **background** is everything but topic and focus.

3.2.2.4 Contrast

3.2.2.4.1 Definition of Contrast

As with the previous components of information structure, there are different terms and definitions used for contrast in the linguistic literature. The most common ones are *contrast*, *kontrast* or even *focus*.

Because the terms **focus** and **contrast** have been used interchangeably sometimes due to some of the characteristics they share (such as membership in a set with alternative members or “highlighting” important information), in this thesis I am going to rely on several basic characteristics of the contrast that help to differentiate it from focus:

10) Characteristics of contrast

- highlighting
- dominant contrast
- membership in a set
- limited set of candidates
- explicit mentioning of alternatives

(Molnár 2002: 101)

These five characteristics of contrast summarized by Molnár are useful in differentiating **contrastive focus** (also referred to as **identificational focus**) from **non-contrastive focus** (also referred to as **semantic focus** or **new-information focus**, or **information-focus**). Molnár argues for treating contrast as an independent category of information structure, “superimposed on topic and focus” (Molnár 2002: 112).

As to the usage of the term **kontrast**, it deserves a special attention too, particularly due to the adoption of this term in McCoy’s dissertation (2001), which is the main recent work published on Russian contrastive clitics, used in this thesis and discussed in depth in later sections. The definition of the **kontrast** was established in Vallduví and Vilkuna (1998):

- 11) The basic idea behind the notion of *kontrast* is the following: if an expression **a** is *kontrastive*, a **membership set** $M=\{ \dots ,a, \dots \}$ is generated and becomes available to semantic computation as some sort of quantificational domain. We are roughly adopting the basic semantic import of “focus” in alternative

semantics (Rooth, 1985, 1992): a set of alternatives for the focused constituent is generated as an additional denotation. *M* is a set of objects matching *a* in semantic type. Furthermore, the members of *M* have to be “comparable” to *a* (Krifka, 1991-1992: 19). The limitation to comparable objects is meant to capture ontological and contextual restrictions.

(Vallduví and Vilkuna, 1998: 83-84)

In her study on the contrastive clitics in Russian, McCoy (2001) adopts Vallduví and Vilkuna’s (1998) term **kontrast** to describe the information marking functions of some of the Russian clitics. Vallduví and Vilkuna briefly explain their choice of the term *kontrast* by trying to show its similarities with Rooth’s focus and by differentiating from the term *contrast* as the former does not cover “all instances of what has been dubbed contrast in semantics, syntax, and phonology” (Vallduví and Vilkuna 1998: 81). Further Vallduví and Vilkuna note that the term *kontrast* is fairly underspecified, subsuming identificational foci, exhaustiveness foci, contrastive foci, contrastive topics, and interrogative *wh*-words all under the cover term *kontrast* (Vallduví and Vilkuna 1998: 83-84). Noting that it would be not very productive to compare the two very different notions of *kontrast* and *contrast*, McCoy, however, identifies two main similarities between **kontrast** and **contrast** in terms of: (1) the membership, i.e. the possible number of the members in the set, and (2) the nature of the relationship between the members (McCoy 2001). The membership set in both **kontrast** and **contrast** implies a presence of at least two members that share the majority of “semantic and grammatical features” (McCoy 2001: 30). Without going into further details and making a finer differentiation between terms *contrast* and *kontrast*, in this thesis the term **contrast** will be used, with its five characteristics as described earlier in Molnár (2002).

Both topic and focus can be contrastive. In her summary of the treatment of contrast in literature, Gryllia lists Rizzi (1997), Belletti (2004), Beninca and Poletto (2004), and Kiss (1998) as differentiating between **new-information focus** and **contrastive focus**, as well as between regular **topics** and **contrastive topics**.

Contrastiveness tests have been used in order to identify whether a specific information structure component, i.e. topic or focus, is contrastive or non-contrastive. Tests adopted by Gryllia for identifying contrastive focus in Greek (Gryllia 2008) and McCoy’s tests

for identifying contrastiveness in Russian (2001) will be discussed later in the following sections, i.e. section 3.1.2.4.2 on contrastive and non-contrastive foci and section 3.1.2.4.3 on contrastive and non-contrastive topics, as well as relevant sections in the analysis of Russian clitics in chapter 4.

3.2.2.4.2 Contrastive and non-contrastive foci

There have been several approaches to subdivisions within focus. As mentioned earlier, Gundel divides focus into psychological, semantic, and contrastive focus. Following Cutler and Fodor (1979), Gundel refers to **semantic focus** as “a part of the sentence that answers the relevant wh-question (implicit or explicit) in the particular context in which sentence is used.” (Gundel 1999: 295). On the other hand, **psychological focus** for Gundel implies that “the attention of both speech participants can be assumed to be focused on it because of its salience at a given point in the discourse” (Gundel 1999: 294). And, finally, **contrastive focus** means for Gundel that the focus is being contrasted, explicitly or implicitly, with something else (Gundel 1999: 296).

King (1995) makes a distinction between **contrastive focus** and **new-information** (a.k.a. non-contrastive or semantic) **focus**. King suggests that **contrastive focus** implies a set of members, out of which a focus is chosen in contrast to the other members. Unlike **contrastive focus**, **new-information** (a.k.a. non-contrastive or semantic) focus is not chosen in contrast to other members of the set and no information is provided about other members (King 1995). Dipper et al. (2007) make the same distinction between **new-information focus** and **contrastive focus**, using newness and contrastiveness as the factors to distinguish between them. Dipper et al. define new-information focus as providing new information and/or information that carries forward the discourse, and contrastive focus as an element that is contrasted with a similar (in Dipper et al.’s terms, “semantically and/or syntactically parallel constituent”) element in a particular discourse (Dipper et al. 2007: 171-172). So, to summarize, so far contrastive focus differs from non-contrastive focus in terms of the set of members (one of the members is contrasted to other(s) in contrastive focus, while there is no information about

other members of the set in non-contrastive focus), which is very similar to the general definition of the contrast.

It will be shown in later sections, with support from data from Russian National Corpus (Grishina 2006), that semantic focus and contrastive focus can be separate from each other or they can overlap. In the former case a focused element carries forward a discourse by bringing new information or highlights an element that is contrasted with a parallel element in the discourse. In the latter case, when semantic focus overlaps with contrastive focus, the focused element both carries forward the discourse and highlights an element contrasted to a parallel element in the discourse.

Dipper et al.'s approach does not contradict King's or Lambrecht's defining ALL types of focus in terms of the *newness* of information to the discourse participant, because they qualify that new-information focus and contrastive focus are not mutually exclusive (Dipper et al. 2007: 171-172). For example, for Dipper et al. a new-information (solicited or unsolicited)⁵ focus can also be a contrastive focus. However, since focus is always defined in terms of the newness of the information, perhaps a finer distinction should be made. For this thesis I am going to make the following assumptions about semantic/non-contrastive focus and contrastive focus, based on the following examples:

12) Semantic or non-contrastive focus:

Question:

Куда ездил дедушка Егорушка?
Where traveled grandpa Egorushka
'Where did grandpa Egorushka travel?'

⁵ Below are tags used by Dipper et al. for focus subcategories:

nf	new-information focus
nf-sol	solicited new-information focus
nf-unsol	unsolicited new-information focus
cf	contrastive focus
cf-repl	replacement
cf-sel	selection
cf-part	partiality
cf-impl	implication
cf-ver	truth value (verum)

(Dipper et al. 2007: 171)

Answer:

Дедушка Егорушка ездил [_{SF} во Францию, Англию, Египет и Японию].
Grandpa Egorushka travel [_{SF} to France England Egypt and Japan].
'Grandpa Egorushka traveled to France, England, Egypt, and Japan.'

In the example above constituent [to France, England, Egypt, and Japan] is treated as semantic focus for several reasons. Firstly, this phrase denotes a set of alternative members, which is characteristic of all kinds of **focus**. Secondly, the membership set denoted by this phrase is unlimited, i.e. it represents all locations in the universe to which *дедушка Егорушка*/'grandpa Egorushka' could have possibly traveled, which suggests that it is **semantic focus**. Thirdly, since the members of the set (countries) enumerated above are not really contrasted to the other members of the set, this constituent is not treated as contrastive focus.

Below is an example of a sentence with contrastive focus:

13) Contrastive focus:

Question:

Женюшка или Ириша купила себе новую шляпку?
Ženjushka or Irisha bought self-DAT new hat
'Did Ženjushka or Irisha buy a new hat for themselves?'

Answer:

[_{CF} Женюшка] купила себе новую шляпку.
[_{CF} Ženjushka] bought self-DAT new hat
'Ženjushka bought a new hat for herself.'

In the example above with contrastive focus, as in the previous example, there is a set of alternative members implied, i.e. {Ženjushka, Irisha}. Similar to the previous example, there is an element of newness of the information to the person asking the question, as they do not know whether it was Ženjushka or Irisha who bought the hat. However, the membership set denoted in this utterance differs from the previous example in two ways. Firstly, the number of members in the set is limited, i.e. it consists of two members in this example. Secondly, the answer in this example is predictable to some extent, i.e. it is either going to be 'Ženjushka' or 'Irisha.'

Based on the sentences above, the following general assumptions will be made for this thesis, combining the approaches of Gundel (1999), King (1995), Lambrecht (1996), and Dipper et al. (2007):

14) **Focus** is the new information in the sentence and can be divided into two main subcategories:

- (a). **Semantic focus** (a.k.a. **new-information focus** or **non-contrastive focus**) (SF) is a focus with a set of alternative members. The number of members in the set is unlimited and the information as to which member from the set will be chosen, is unknown/unpredictable to the speaker. The term ‘semantic focus’ is chosen for two reasons: (1) the term ‘new-information focus’ seems to be misleading because all foci denote new information to some extent in the sentence; and (2) the term ‘non-contrastive focus’ also cannot be used due to the cases when this focus overlaps with contrastive focus, which would mistakenly result in elements labeled ‘non-contrastive contrastive focus’.
- (b). **Contrastive focus** (CF) is also a focus with a set of alternative members. However, the number of members is limited and known to the speaker, therefore it is possible to predict to some extent the answer, which is constricted to be one of the known members of the set. One of the members of the set is contrasted to another member.

As will be shown in Chapter 5 based on data examined in the Russian National Corpus, sometimes **semantic focus** can overlap with **contrastive focus**. It is not clear whether a separate subcategory (SF CF) should be created for this type of sentences or whether it should be identified as [CF [SF]] as will be seen in future sections. More research is necessary to understand the relationship between

semantic focus and **contrastive focus**. However, for the purposes of this thesis, a new category will be created:

- (c). **Semantic/contrastive focus** (SF CF) (or [CF [SF]]) in some cases with Russian data discussed in other sections) will be used to refer to the elements in which semantic focus overlaps with new-information focus. The set of members is limited to some extent, however the answers are still unpredictable, as will be seen in chapter 4 based on Russian data.

Gryllia (2008) uses several tests to identify contrastiveness of the focus in sentences, adopted from Rizzi (1997) and Kiss (1998):

15) ***Wh*-question/*contrastive answer test**

A contrastive answer is incompatible with an ordinary *wh*-question.

(test adopted from Rizzi (1997), Gryllia (2008: 40))

16) Example of *wh*-question test⁶:

Wh-question

Кто ездит на Форде?
who drives in Ford
'Who drives a Ford?'

Answers

- (1). *[_{CF} Женя] ездит на Форде.
Zhenya drives in Ford
'Zhenya drives a Ford.'
- (2). [_{SF} Женя] ездит на Форде.
Zhenya drives in Ford
'Zhenya drives a Ford.'

⁶ This is my own example.

17) Correction test

A contrastive focus can be used to answer a *yes-no* question, correcting part of the predicate information of the question (test adopted by Gryllia from Kiss (1998))

18) Example of correction test in Italian from Gryllia (2008):

Question

- a. L' ha rotto Giorgio, il vaso?
it has broken Giorgio the vase
'Has Giorgio broken the vase?'

Answer

- b. [Maria]_{C-Foc} ha rotto il vaso.
Maria has broken the vase
'It is Maria who has broken the vase.'

(Gryllia 2008: 31-32)

19) Choice test

When answering an alternative question, one alternate is contrasted to the other.
(test adopted by Gryllia from Kiss (1998))

20) Example of correction test in Italian from Gryllia (2008):

Question

- a. Chi di voi due ha rotto il vaso?
which of you two has broken the vase
'Which of you two has broken the vase?'

Answer

- b. [Maria]_{C-Foc} ha rotto il vaso.
Maria has broken the vase
'It is Maria who has broken the vase.'

(Gryllia 2008: 32)

The tests listed above to identify contrastiveness of the focus will be discussed with examples from spoken data in more detail in chapter 4.

3.2.2.4.3 Contrastive and non-contrastive topics

As has been mentioned earlier, topics have been broken into **contrastive** and **non-contrastive**.

The description that Krifka (2007) provides for contrastive topic implies an existing set of alternatives because of the contrast present in the topic (which Krifka refers to as *focus within the topic*):

- 21) “Contrastive topics are topics with a rising accent... They arguably do not constitute an information-packaging category in their own right, but represent a combination of topic and focus, as indicated in the example, in the following sense: They consist of an aboutness topic that contains a focus, which is doing what focus always does, namely indicating an alternative.”

(Krifka 2007)

For Krifka a contrastive topic is an aboutness topic, which is what the sentence is about, that contains a focus, which in its turn implies a set of alternatives. Krifka’s definition of a contrastive topic is very useful in terms of differentiating a contrastive topic from a non-contrastive topic, i.e. contrastive topic implies a membership set with alternatives, one of which is contrasted to other members. For this thesis I will adopt Krifka’s definition, at least the part that will be crucial in processing spoken data in Chapter 5, i.e. contrastive topic is a topic denoting a set with alternative members. Krifka’s definition of contrastive topic as a topic that contains a focus, however, does not eliminate the difficulty in differentiating contrastive topic from contrastive focus that will be discussed in the next section.

An example (in Italian) of contrastive topic is provided by Gryllia (2008):

- 22) *Context*: a farm producing a set of goods that are known to the people involved in the conversation.

- a. [La frutta]_{C-Top} la regaliamo, [la verdure]_{C-Top} la vendiamo.
the fruit it give the vegetables it sell
‘We give the fruit for free, while we sell the vegetables’

Gryllia suggests that in this context [la frutta] and [la verdure] are contrastive topics, demonstrating it with two tests to identify contrastive topics in the sentence. First test is substitution test adopted by Gryllia from Benincà and Poletto (2004):

23) Substitution test for contrastive topics

If two terms are interpreted with a ‘List interpretation’, then they can be substituted with ‘the former’ and ‘the latter’.

(Gryllia 2008: 33)

Gryllia shows how substitution test can be used to identify contrastive topics with the following example in Italian:

24) Applying substitution test (substituting [la frutta] for [la prima] (‘*the former*’) and [la verdure] for [la seconda] (‘*latter*’):

b. La prima la regaliamo, la seconda la vendiamo.
the first it give the second it sell
‘We give the former for free, we sell the latter.’

(Gryllia 2008: 33)

Another test that can be used to identify contrastive topics is implicit sub-question test:

25) Implicit sub-question test

- (i) When a *wh*-question can be split into sub-questions and the answer is organized per sub-question, then, there is a contrastive topic in the answer.
- (ii) When a question can be interpreted as containing more than one implicit sub-question, and the answer addresses only one of these sub-questions, rather than the general question, then, this answer contains a contrastive topic.

(Gryllia 2008: 37)

Gryllia demonstrates how implicit sub-question test works by using Buring’s example:

26) Question

What did the pop stars wear?

Answer1

#The female pop stars wore [caftans]_F

Answer2

The [female]_{CT} pop stars wore [caftans]_F.

(Gryllia 2008: 38)

These tests and their applicability to Russian data, as well as contrastive topics in general will be discussed in more detail in the section on Russian contrast, in particular in the part dealing with the information structure marking properties of the Russian clitic *–TO*. The tests to differentiate contrastive focus from contrastive topic will be discussed in chapter 5 with Analysis.

3.3 Information Structure in Russian

3.3.1 General

As King notes, a three-way division of sentences appears to be the most plausible analysis of information structure in Russian. Under this analysis, **topics** always precede **discourse-neutral information**, while **foci** can appear in different positions relative to the non-focused items, depending on the intonation and the sentence stress (King 1995: 76).

Examining Russian data, King makes several observations about topic marking in Russian language. She observes that topics usually appear in initial position; topics tend to be definite and are often pronominal; only constituents can be a topic (King 1995: 79). In Russian, while multiple topics are allowed, it is also possible to have a sentence without a topic (King 1995: 79), which goes in line with Enghdal and Vallduví (1994) and Buring's analysis that not all sentences always have a topic. In addition to the regular topics, there are also pro-dropped subjects, which appear frequently in the data and are always topics (King 1995: 69).

King divides focus in Russian into three types: contrastive focus, new-information focus, and presentational focus. The latter can be subsumed under the new-information focus, according to King (1995). Bailyn (2011) also divides focus into two categories: information focus and contrastive (or identificational) focus. Among the main strategies for marking focus in Russian are **word order**, **sentence/emphatic stress**, **clefting**, **li-questions**, and **morphological association with focus** (King 1995). Each of these

strategies will be discussed in detail below, in regards to them marking all information structure constituents, i.e. **foci**, **topics**, and **contrast**.

3.3.2 Word Order

One of the main strategies of marking information structure in Russian is **word order**. Word order in Russian has been described in great depth in recent Bailyn's (2011) work and it is also the primary interest of King's dissertation (King 1995). Although Russian is traditionally referred to as a language with free word order, King argues in her dissertation that Russian lies somewhere in between highly configurational languages like English and French, in which rigid word order encodes grammatical functions and the non-configurational languages like Warlpiri with extremely free word order (King 1995: 1). King suggests that although the surface word order is free in Russian, there is an "unmarked word order", and the so-called "free" surface word order is predicted from discourse factors (King 1995: 1-2). Dividing sentence into rheme (focus) and theme (topic + discourse-neutral material), King proposes that Russian is a language with an underlying VSO order, which is "obscured by the movement of constituents to receive discourse function interpretations" (King 1995: 4). The movement of the constituents marking information structure results in that in the sentences with neutral intonation, **topics** precede discourse-neutral constituents (**background**), which in their turn precede **foci** (King 1995: 1). Similar to King, Bailyn (2011) also assumes the following order of the components in Russian:

27) *IS Ordering Rule*: Topic > Discourse Neutral Material (DNM)) > Focus

Bailyn notes, however, that there is "no single syntactic method of encoding Topic/Focus structure with word alone" because of the availability of intonational means that can achieve the same ends (Bailyn 2011: 267).

As to the position of the topic, King notes that NPs appearing after the verb cannot be interpreted as a topic in Russian (King 1995: 106). King's examples supporting this statement are sentences with word order *object-verb-subject* and *object-subject-verb* below:

- 28) [Èto plat'e] šila [Inna].
 this dress sewed Inna
 'This dress-TOP, Inna sewed.'

(King 1995: 106)

- 29) [Staruyu lodku] [my] prodali.
 old boat we sold
 'We-TOP sold the old boat-TOP'

(King 1995: 107)

The examples above demonstrate the usual position of the **topic** in the sentence in Russian, i.e. sentence-initial position. If there are two topics in the sentence, then the new topic introduced in the sentence precedes the topic that has been introduced in an earlier sentence (King 1995).

Except for sentences with neutral intonation, where focus is placed after background, **focus** can occupy different positions in Russian. Position of focus in the sentence depends on the type of focus (contrastive vs. new-information focus), as well as the presence/absence of other means encoding information structure in the sentence. Semantic (or new-information focus, in King's terminology) focus usually occupies sentence-final position in Russian:

- 30) Čitaet knigu [otec].
 reads book father
 'Father-FOC is reading a book.'

(King 1995: 80)

As shown above, semantic focus occupies sentence-final position.

On the other hand, similar to Italian (Rizzi 1997), **contrastive focus** in Russian is usually fronted (Neeleman and Titov, 2009), or, as King notes, it appears "immediately preverbally" (King 1995: 81). In his review of focus position in Russian Bailyn (2011) mentions that the kind of focus that appears on the left edge of the sentence is still disputed, i.e. it can be occupied by contrastive focus or semantic focus.

However, in sentences with contrastive focus, too, intonation or presence of the contrastive clitics can override the word order in terms of information structure interpretation, as will be later shown in the sections describing clitics.

3.3.3 Sentence/Emphatic Stress

Intonation can override the word order in terms of signaling the focus in Russian language. In the example below, if only the word order was responsible for focus marking, then the second NP would have been the focused element. However, due to the presence of emphatic stress on the second element, the verb *vypil*, this verb receives focus marking instead of the clause-final NP:

- 31) Boris [VYPIL] vodku.
 Boris drank vodka
 Boris drank-FOC the vodka

(King 1995: 80)

In addition to its ability to override the word order, emphatic stress can also help distinguish between **contrastive focus** and **non-contrastive focus**. King notes that in Russian like in Hungarian contrastive focus is marked by emphatic stress (King 1995: 74). In her examples she shows that in a sentence with both non-contrastive and contrastive foci, the contrastive focus is going to be emphatically stressed, while the non-contrastive focus is not (King 1995: 74). Neeleman and Titov also discuss the different intonational contours characteristic of different types of focus. They note that the **new information focus** in Russian has been usually regarded as marked by **IK1** (a falling tone), whereas **contrastive focus** is marked by **IK2** (higher in tone and more intense than IK1) (Neeleman and Titov 2009: 515). Bailyn refers to contrastive focus, or left focus as ‘intonational focus’, or as “accompanied by Intonation Pattern IK-2” (Bailyn 2011: 278).

3.3.4 Clefting

Another strategy for marking focus in Russian language is *it*-clefting, exemplified in Gundel’s example (adapted by King):

32) Eto [Boris] vypil vodu.
 It Boris drank vodka
 It is Boris-FOC (who) drank the vodka.

(King 1995: 80)

Gundel discusses *it*-clefts (spelled as “jeto” or “eto”-clefts) in Russian and Japanese, in which focus (“comment” in Gundel’s work) precedes the topic in the following structures:

33) [(it) (BE) NP] [NP]

(Gundel 1988: 226)

In this example⁷ of focus-topic order in *it*-cleft constructions, the first NP is focused, while the second NP is a topic, thus overriding the default word order for focus marking. Bailyn notes that the item the item focused by *it*-cleft “immediately follows *eto* and receives primary stress” (Bailyn 2011: 282).

3.3.5 Clitics as Information Structure markers

3.3.5.1 Introduction

The main information strategy marking focus that this study is going to look at is clitics. Clitics marking information structure in Russian have been referred to as “morphological association with focus” (King 1995: 80) for clitic *ŽE* or, as in the case of a clitic *LI*, as an “interrogative complementizer” (King 1995: 139). Clitics can also be described as lexical items marking information structure in the sentence.

⁷ Below are two more examples in Russian and French demonstrating the focus marking by clefting:

Eto	Ivana	ja	videl
It/that	Ivan-ACC	I	saw
'It (was) Ivan I saw.'			

C'est	pour	ca	qu'il	a	gagne	le	prix
It's	for	that	that he	won	the	prize	
'It's for that that he won the prize'							

(Gundel 1988: 226)

In this study three Russian contrast-marking clitics⁸ will be analyzed in detail: *ŽE* (also frequently spelled *ZHE* in English and sometimes used as abbreviated version *Ž* or *ZH*), *-TO*, and *LI*. The first two clitics *-TO* and *ŽE* have been analyzed as marking contrast, while the latter clitic *LI* has been analyzed as marking focus in Russian. Used in sentences with different illocutionary force (*-TO* and *ŽE* are used in interrogatives and declaratives, while *LI* is used in interrogatives), these clitics can override the word order and play an important role in marking information structure in Russian. These three clitics represent just a tip of an iceberg of the multitude of clitics and their usage in Russian, as well as their role in the information structure marking, with each of them easily being a possible subject for one (or many) dissertations.

In the following sections I will first review literature on clitics *-TO* and *ŽE* and then review literature on clitic *LI*.

3.3.5.2 Clitics *-TO* and *ŽE*

The major recent work on clitics *-TO* and *ŽE* that has been completed, at least to my knowledge, is by McCoy in her extensive dissertation research proposing a unifying analysis of these two clitics (or in McCoy's terms, '*particles*') (McCoy 2001). McCoy analyzed clitics *-TO* and *ŽE* (as well as clitic *VED'*, which is outside the scope of this paper) as "unambiguous markers of contrast" (McCoy 2001, 2003: 319), connecting before seemingly unconnected functions and characteristics of these clitics.

McCoy adopts Vallduví and Vilkuna's (1998) term *kontrast* as the best way to describe the functions of these clitics, discussed earlier and repeated here for convenience:

⁸ Following Zwicky's differentiation of clitics vs. particles vs. independent words, in this thesis *-TO*, *ŽE*, and *VED'* are going to be referred to as clitics (Zwicky 1985), and not as particles as they have been referred to by McCoy (McCoy 2001). The reasons for referring them as *clitics* are based on Zwicky's tests. *-TO*, *ŽE*, and *VED'* are neither independent words, nor affixes. They are not capable of receiving stress, and they are not prosodically independent. Their ordering (following a focused element) and distribution (restricted to combining with the first phonological word of a clause or first focused word of a clause) are also indicative of clitics. Following Zwicky, they also cannot not be termed discourse markers, which, in Zwicky's terms, are independent words. Thus, *-TO*, *ŽE*, and *VED'* will be referred to as clitics for the purposes of this paper.

34) The basic idea behind the notion of *kontrast* is the following: if an expression *a* is *kontrastive*, a **membership set** $M=\{...,a,...\}$ is generated and becomes available to semantic computation as some sort of quantificational domain.

(Vallduví and Vilkkuna, 1998: 83)

Applying Vallduví and Vilkkuna's (1998) notion of *kontrast*, McCoy claims that clitics *-TO* and *ŽE* are lexemes whose primary function is to signal *kontrast*. The definition that she provides for **kontrastive markers**, or **k-markers**, is the following:

35) A *kontrastive marker*, or *k-marker*, is a linguistic expression – syntactic, morphological, prosodic, or lexical – that signals *kontrast*; or, in other words, causes the hearer/reader to evoke a membership set which the *kontrastively* marked element belongs to, and treat this element as a member of the set.

(McCoy 2001: 30)

McCoy's analysis is going to be used as the basis for this thesis, analyzing clitics as marking *kontrast* by implying presence of a **membership set(s)** with at least two members, one of which is the *kontrastively* marked element.

Adopting more analysis from Parrott and Yokoyama, McCoy develops an extensive unifying analysis of clitics *-TO* and *ŽE*, supporting it both with written and spoken data. Her spoken data comes from Ekaterina Protassova's corpus of the CHILDES database (MacWhinney 2000), which consists of the recordings of day-to-day conversations between Protassova's daughter, Varja, and her family.

3.3.5.2.1 Clitic *-TO*

Research on *-TO* that has been completed prior to McCoy (2001) describes the multiple functions of this clitic, which has been analyzed as (1) marker of *kontrast*; (2) marker of emphasis, (3) theme/topic marker or marker of the information known to the hearer, (4) marker of unexpectedness in addressing a topic, (5) implying plurality, and (6) having an

evaluative-expressive meaning (McCoy 2001: 210-212). McCoy suggests that her analysis of *–TO* as a contrastive marker (will be summarized below), explains all of its functions above as the natural consequences of its contrastive nature.

McCoy treats clitic *–TO* as marking **a set of sets of related propositions**, which is generated by introducing alternatives to a contrastive element within the *link* (topic) and a contrastive element within the *rheme* (focus). The referents marked by this clitic are known to the hearer, but are not currently activated in the discourse. McCoy uses the following example to illustrate the set of sets of related propositions marked by *–TO*:

- 36) **Speaker A:** a. Včera u babuški byl den' roždenija.
'Yesterday was grandma's birthday.'
- b. Ona ždala pozdravenij ot svoix vnukov.
'She was waiting for her grandchildren's greetings.'
- c. Nataša prislala otkrytku. Oleg privez podarok.
'Nataša sent a card. Oleg brought a present.'
- Speaker B:** d. Nu a Polina kak?
'And what about Polina?'
- Speaker A:** e. PoLIna-TO pozvoNila.
Polina-TO pozvonila
'And as for Polina, she CALLED.'

(McCoy 2001: 82)

McCoy argues that *–TO* in the example above and in similar examples signals to the hearer that the proposition '*Polina called*' is a part of a membership M set that has the following members:

37) M (=grandchildren greeting grandma on her birthday)

[*'Nataša sent a card. Oleg brought a present. And Polina(-TO) called.'*]

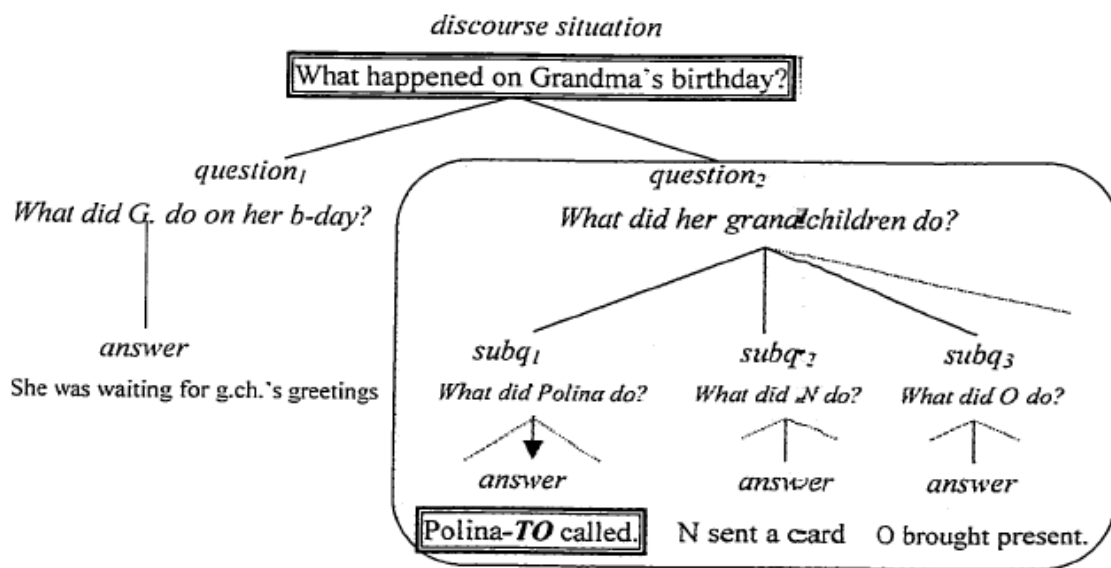
{ {Nataša sent a card; *Nataša brought a present*; *Nataša called*; ...} ;...};⁹
{*Oleg sent a card*; Oleg brought a present; *Oleg called*; ...};
{*Polina sent a card*; *Polina brought a present*; Polina called; ... }

(McCoy 2001: 83)

⁹ The punctuation and use of brackets appears as it was used in McCoy (2001). However, it seems that this curly bracket and preceding periods should be placed in the very end of the membership set.

The underlined propositions above are the propositions that are true. Italicized propositions are the propositions that can either be true or false. According to McCoy, the membership set marked by *-TO* is a set of set of propositions, which is “generated by introducing alternatives to the kontrastive (element within the) link and the kontrastive (element within the) rheme” (McCoy 2001: 87). McCoy uses the following diagram to represent the discourse tree for the previous example:

38) Discourse tree for set of sets of propositions marked by *-TO* in (37)

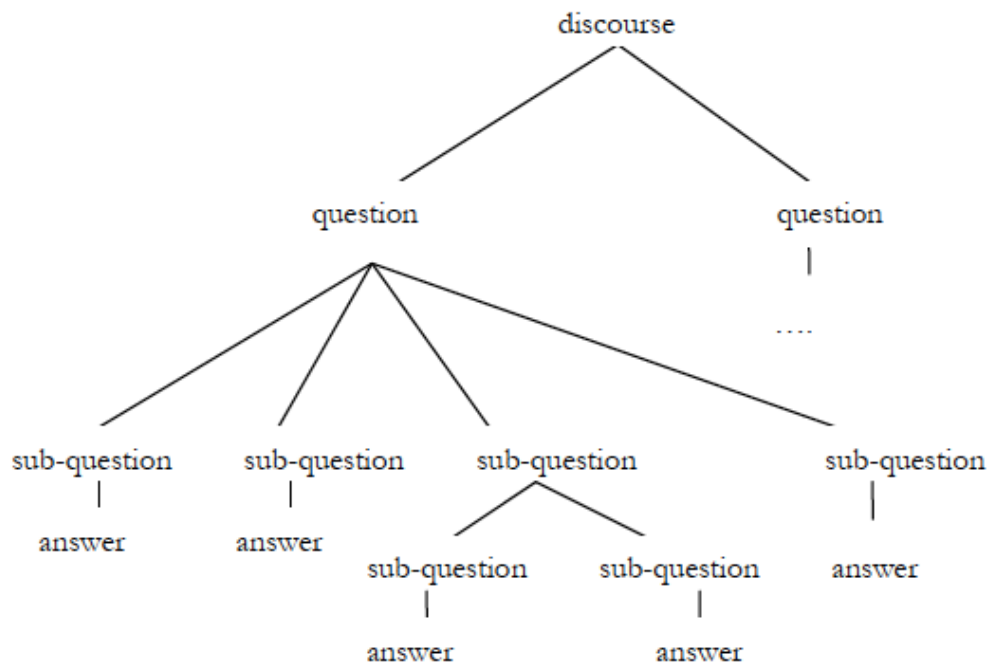


(McCoy 2001: 86)

McCoy analyzes utterance ‘*Polina-TO called*’ as evoking the set M corresponding to the area within the rectangle above. McCoy argues that “proposition ‘*Polina called*’ partially answers question 2 and makes salient other sub-questions salient (the answers to which can be implied or explicitly articulated)” (McCoy 2001: 86).

McCoy’s diagram is very similar to Buring’s (2003) hierarchical model of discourse structure used by Gryllia (2008) (see next page):

39) Buring's (2003) hierarchical model of discourse structure



Buring's diagram of discourse structure also consists of two questions, which further consist of sub-questions. McCoy (2001) breaks down sentences containing *-TO* in the same way, using a term "a set of sets of propositions", interpreting a larger set as represented by question and smaller sets of propositions as represented by sub-questions on a diagram above.

McCoy claims that clitic *-TO* encliticizes to contrastive element, which is usually the topic but it can also attach to focus when "conditions of evoking a set of sets of propositions is satisfied otherwise" (McCoy 2001: 118). On the next page is an example of *-TO* attaching to an element within a contrastive focus:

40) Context: *Varja is arranging her dog's ears so as to make the dog "beautiful"*

Mother: Xochesh' ushi kak u Pljuti?
'Do you want your ears like Pljuti's [toy dog's]?'

Varja: Tak, tak.
'This way, this way.'

- Mother: A? U devochek kogda takie ushki, eto nekrasivo, eto tol'ko u sobachek krasivo.¹⁰
Ah? No, it's not pretty. When girls have such ears, it's not pretty, it's only dogs' ears that are pretty.
- Varja: Sobaka? A tak sdelat', vot tak vot? Torchashie. A tak sdelat', vot tak? Vot tak vot sdelat', sdelat'. Krasivo? Krasivo? A tak krasivo, vot tak? Tak? Vot tak?
'Dog? And to do it this way, like this, this? Sticking out. And to do it this way, like this? Like this, like this, like this. Like this, like this do, do. Pretty? Pretty? And [it is] pretty this way, like this? This? Like this?'
- Mother: Da sobachkam, voobshche govorja,
Da [+K/-Rh sobachkam], voobshe govorja,¹¹
PART dogs-DAT generally speaking
'Well, dogs, generally speaking,
- vse eto ne ochen'-TO nuzhno.
vse eto [+K/+Rh ne ochen'-TO nuzhno].
all this NEG very-much-TO is-needed
don't need all this very much(-TO).

(McCoy 2001: 167)

McCoy states that above elements *sobachkam*/'dogs-DAT' and *ne ochen'-TO nuzhno*/'not needed very much' are contrastive topic and contrastive focus accordingly. While McCoy does not provide an explanation why the former is treated as a topic and the latter is treated as a focus, she briefly mentions that she treats *ochen'-TO nuzhno*/'not needed very much' as contrastive focus marking a set of sets of prepositions. McCoy notes that *-TO* attaching to *ne ochen'-TO nuzhno*/'not needed very much' can evoke a set of sets of prepositions because *sobachkam*/'dogs-DAT' has already been contrasted to *devochek*/'girls-GEN' in the preceding utterance by mother.

McCoy uses the following example from CHILDES database to demonstrate how *-TO* marks a set of sets of prepositions:

¹⁰ I have changed McCoy's format to save the space. Instead of using a separate line for each utterance by the same speaker I combined several utterances by the same speaker together. Also, gloss has been placed together with the utterances in Russian.

¹¹ McCoy uses Vallduví and Vilks's features in her work:

[K:+, Rh:+] – contrastive focus (also identificational focus and exhaustiveness focus)

[K:+, Rh:-] – contrastive topics

[K:-, Rh:+] – regular, non-contrastive rhemes, or foci

[K:-, Rh:-] – regular, non-contrastive themes, or topics (McCoy 2003: 320)

- 41) [Varja and her mother are looking at a picture of a dog who put her paw on a bear. They first discuss the dog, then start talking about the bear. Varja gets distracted by taking a scoop into her hand. Mother says:] (CHILDES, seance 2)

U tebjā-*TO* sovok, a čto u medvedja v lape?

At you-*TO* scoop but what at bear in paw

"YOU(-*TO*) have a SCOOP, but what does the BEAR have in his paw?"

(McCoy 2003: 322)

McCoy argues that in the example above both the **topic** (*link* in McCoy's work) and the **focus** (*rheme* in McCoy's work) are contrastive. It should be noted that this sentence might not be the best example to show the contrast-marking properties of the clitic *-TO*, due to the presence of another contrastive conjunction *a*/'but' in the same sentence. Interestingly enough, this sentence fails *wh*-test for contrastiveness of focus used in Gryllia (2008), but passes implicit sub-question test for contrastiveness of topic used in Gryllia (2008), discussed later in more detail in Chapter 5, in the section on analysis of contrastive focus vs. contrastive topic.

Below is another example from McCoy's dissertation that contains only one contrastive marker *-TO*, but follows the same information structure scheme as the previous example:

- 42) [+K/Link Tišina]-*TO* [+K/Rheme kakaja]!

Quietness-*TO* what

"How quiet it is!" or

"As for the state of quietness(-*TO*), how quiet it is!"

(McCoy 2003: 324)

As in the previous example, the element preceding clitic *-to*, noun *tišina*/'quietness' in this case, is a **contrastive topic**.

McCoy uses 'pronoun-doubling' test to identify presence of contrast in the sentence. McCoy claims that based on her research 'pronoun-doubling' is "generated only in sentences which contain kontrastive element" (McCoy 2001: 108). Using examples below, McCoy claims that pronoun-doubling is ungrammatical in this sentence until clitic *-TO* is added:

- 43) *Polina ona včera pozvonila.

Polina she yesterday called

'As for Polina in particular, she called yesterday.'

Polina-TO ona včera pozvonila.
Polina-TO she yesterday called
'As for Polina in particular, she called yesterday.'

(McCoy 2001: 111)

McCoy claims that the fact that sentences with pronoun-doubling are grammatical only upon addition of clitic *-TO*, confirms her analysis of clitic *-TO* as contrastive.

3.3.5.2.2 Clitic *ŽE*

Clitic *ŽE* has been previously analyzed as (1) a thematic/organizational/textual clitic; (2) an affective/modal clitic, (3) a contrast/contrastive focus marker, (4) a marker of emphasis, (5) a marker of (re-)activated information, and (6) a marker of reference point in the activated 'domain of reference' (McCoy 2001: 229). Prior to McCoy's work these different functions of *ŽE* have been viewed as not necessarily connected. However, as McCoy notes, there have been a few attempts, the most prominent of which are by Bitextin and Parrott, to try to find the underlying principles governing the use of this clitic. (McCoy 2001: 227). According to McCoy, Bitextin makes a distinction between *textual ŽE* (used to signal identity between two elements, when conditions of contrast or disjunction are not optimized, such as when these elements are separated by large chunks of text or if they occur in different syntactic constructions) and *modal ŽE* (usually associated with repetition and used in the absence of successful speech act conditions). Bitextin also notes the different scopes of clitic *ŽE*, i.e. scope over a term vs. scope over an entire proposition. McCoy, however, bases her analysis of *ŽE* more on Parrott's (1997) work, which describes in detail discourse functions of *ŽE*, as well as its position in the utterance as dependent on speaker's assessment of the discourse situation and of the hearer's knowledge set (McCoy 2001: 225). Similar to Bitextin and others, Parrott makes a distinction between two types of *ŽE*: thematic *ŽE* (similar in its functions to the conjunction 'a/'and or but') and sentential, a.k.a. modal *ŽE* (Parrott compares it to another contrastive clitic *VED*). McCoy notes that both Bitextin (1994) and Parrott (1997) "stop short of a truly unifying analysis that would account for a full range of its

(*ŽE*'s) contextual meanings" (McCoy 2001: 227), citing their treatment of *textual/thematic* and *modal/sentential* subtypes of *ŽE* as having separate meanings as one of the shortcomings of their not truly unifying analysis.

Following Parrott (1990,1997), McCoy mentions aggressiveness of clitic *ŽE*, comparing it to 'verbal attack' on the addressee (McCoy 2001:214). McCoy builds her work around Parrott's analysis of clitic with *ŽE*, integrating it within other frameworks and accounting for the type of sets that *ŽE* marks. While Parrott works mainly with the Transactional Discourse Model by Yokoyama (1986), McCoy integrates her analysis within other frameworks as well, i.e. Vallduví and Vilks's theory of contrast (1998), Vallduví's theory of information packaging (1992), Gundel et al.'s (1993) theory of cognitive statuses of referents in discourse, and Büring's (2000) theory of discourse structure (McCoy 2001: 226). While Parrott concentrates on the cognitive status of information marked by clitic *ŽE*, McCoy also addresses the type of sets that this clitic marks, thus accounting for a greater range of functions of the clitic *ŽE*.

McCoy proposes that clitic *ŽE* denotes a **set of propositions** that differ from each other in the value of one term or more. The members of the set of propositions marked by *ŽE* are **mutually exclusive**, per McCoy (2001). It should be noted, however, that this analysis of *ŽE* marking a set of mutually exclusive propositions should mainly refer to the proposition-level *ŽE*. Contrastive foci that can be marked by *ŽE* include verum focus¹² through verbal inflectional properties, such as tense or aspect, *wh*-words, deictics, expressions like *such*, etc. (McCoy 2001).

McCoy pays special attention to the relationship between the scope of *ŽE* and its position within the clause. Clitic *ŽE* can take different scopes (i.e. mark contrastive focus on a constituent or mark contrastive focus on entire proposition) in terms of what it marks as contrastive in the utterance. McCoy refers to two types of clitic *ŽE* taking different-level scopes as:

Phrasal *ŽE*: refers to the clitic when it contrastively marks terms, or, syntactically, phrases (McCoy 2001: 234). Below is an example of phrasal *ŽE* from CHILDES database, used by McCoy:

¹² McCoy defines verum focus as focus expressing "polarity, tense, aspect or some other semantic category associated with inflection" (McCoy 2002: 6)

44) In this example Varja and a grandparent are looking at the picture of a sparrow. Varja believes that this is a cat, while her grandparent tells her that this is a sparrow:

Grandparent:

Chto ty. Eto na stene visit kartinka, gde on narisovan,
What you. This on wall is-hanging picture, where he is-painted
'What's wrong with you. This is a PICTURE hanging on the wall, where
he is painted,

[+K etot zhe samyj vorobej].
this *že* samyi sparrow
the (*že*) VERY SAME SPARROW.'

(McCoy 2001: 233)

McCoy briefly suggests that **phrasal** *ŽE* can be analyzed as **proposition-level** *ŽE* (McCoy 2001: 240). Nevertheless, she proposes two different types for them. McCoy claims that **phrasal** *ŽE* is located in the second position within the contrastive phrase, encliticizing to the contrastively marked element (McCoy 2001: 261), as shown in the example (44) with contrastive phrase [+K *etot zhe samyj vorobej*/'very same sparrow']. Another example of a **phrasal** *ŽE* marking narrow contrastive focus over the term (phrase) is found in King's work:

45) On uedet [segodnia že].
He will leave today že
'He will leave today-FOC.'

(King 1995: 80)

In this sentence clitic *ŽE* attaches to the **contrastively focused** element on the right periphery of the sentence *сегодня*/'today'. *ŽE* occupies the 2nd position within the contrastive phrase [*segodnia že*/'today že'], as predicted by McCoy's analysis of the placement of phrasal *ŽE* (2001).

Proposition-level *ŽE* (also referred to as *clausal* *ŽE* (McCoy 2001: 118) or *sentential* *ŽE* by Parrott (1997)) refers to the clitic when it contrastively marks entire propositions or utterances. Below is an example of **proposition-level** *ŽE* from CHILDES database:

46) In this example Varja's mother is telling her that Varja's father will soon come home and bring ice cream.

Context [Varja is misbehaving: mother is changing the conversation to a more pleasant topic]

Mother:

Predstavljajsh' sebe: papa pridet i prineset morozhenoe.
Imagine to-yourself: daddy will-come and will-bring ice-cream
'Can you imagine: daddy will come and bring ice cream.'

[+K Eto zhe prekrasno].
This *že* beautiful!
'This (že) is GREAT!'

(McCoy 2001: 235)

In the example above McCoy analyzes *ŽE* located in the second position as having a scope over an entire proposition *Eto zhe prekrasno*/'This is GREAT'.

McCoy relies on Parrott's (1997) diagram for placement of **proposition-level** *ŽE* (sentential *ŽE* in Parrott's terminology) in the sentence:

47) # * *že* | | *že* | SS | *že* | | * *že* | | *že* #
(McCoy 2001: 263)

(where SS = sententially-stressed material; boxes are phonological words;
pound key (#) = boundary of utterance)

According to Parrott, **proposition-level** *ŽE* cannot be sentence-initial, nor can it encliticize to non-sentence-final element that follows the sententially stressed element SS.¹³ It is allowed in all other positions, however. McCoy extends Parrot's definition of **proposition-level** *ŽE* as "gravitating towards the center of contrast", which is the sententially stressed material. McCoy proposes the following rule for placement of **proposition-level** *ŽE*:

¹³ For written data, King notes that presence of *ŽE* indicates a contrastive focus in written Russian language, when sentence stress that is usually used for indicating contrastive focus is not available (King 1995: 81).

48) Proposition *že* encliticizes either to the clause boundaries (the initial or the final phonological word) or to the kontrastively marked element within the clause.

(McCoy 2001: 264)

To summarize, McCoy (2001) suggests that both **phrasal** and **proposition-level** *ŽE* encliticize to contrastively marked elements (or in the case of **proposition-level** *ŽE* it can also attach to clause boundaries) (McCoy 2001: 261-262). The important outcome of McCoy's treatment of clitic *ŽE* as attaching to contrastively marked elements for the analysis and implementation of this clitic discussed later in chapters 4 and 5 is that both **phrasal** and **proposition-level** *ŽE* can occupy a 2nd position within the clause/sentence.

In order to test contrastiveness of the element marked by *ŽE* McCoy attempts to use 'pronoun-doubling' test, as she did with clitic *-TO*. However, due to the absence of naturally occurring data with pronoun doubling co-occurring with clitic *ŽE*, McCoy constructs her own examples that she claims are acceptable to the speakers (except for the last one which is "slightly less acceptable" (McCoy: 2001: 114):

49) Polina ona VČERA *že* pozvonila.
Polina she yesterday že called
'But as for Polina, she called YESTERDAY.'

Polina ona včera POZVONILA *že*.
Polina she yesterday called že
'But as for Polina, yesterday she CALLED.'

??Polina *že* ona včera pozvonila.
Polina že she yesterday called
'But as for Polina, she called yesterday.'

?/??Polina ona *že* včera pozvonila.

(McCoy 2001: 114)

Based on these examples and their acceptability to native speakers, McCoy concludes that even though the combination of *ŽE* and pronoun doubling needs to be studied in more detail, it is sufficient to say that clitic "facilitates pronoun doubling" (McCoy 2001: 114), which, as McCoy believes, in its turn appears only in sentences containing contrastive elements.

Another central example of the contrast marked by *ŽE* used by McCoy is from Protassova's corpus of the CHILDES database (MacWhinney 2000). In this example child Varja sees the fly and insists that her mother kills it:

- 50) Varja: Ona muxa, muxa.
 'It's a fly, a fly.'
- Mother: Muxa, muxa, da.
 'A fly, a fly, yes.'
- Varja: Ubit', ubit' ee!
 'Kill, kill it!'
- Mother: Ona **zhe** uzhe ubita
 'It (**že**) is already killed.'

Below is just the sentence containing *ŽE*:

Ona zhe [+K uzhe ubita].
she **že** already killed(participle)
“(But) it (**že**) is already killed.”

(McCoy 2003: 326)

McCoy argues that in the example above, the contrast set consists of two mutually exclusive members: presupposition of the hearer (Varja) that the fly is alive and presupposition of the speaker (Mother) that the fly is already killed. If one presupposition (*fly is alive*) is true, another presupposition (*fly is dead*) must be false. McCoy defines *ŽE* as ‘encliticizing’ to contrastive elements. It should be noted that in terms of syntactic attachment ‘encliticizing’ can be interpreted as attaching to its host on the left. In the example above *ŽE* attaches to its host on the right and can be interpreted as ‘procliticizing’ to the contrastive element *uzhe ubita* ‘she že already killed’. To avoid the confusion, I am going to use the term ‘cliticize’ to describe *ŽE* as it does not specify the position the position of *ŽE* relative to the element it modifies, accounting for both encliticizing and procliticizing *ŽE*.

Scopal properties of *ŽE* (i.e. its ability to mark narrow contrastive focus vs. wide contrastive focus), position of *ŽE* in the sentence, as well as more tests identifying contrast in sentences containing *ŽE* will be discussed in more detail in Chapter 4.

3.3.5.3 Clitic *LI*

Clitic *LI* has been analyzed as an interrogative complementizer (both for matrix and embedded questions) that assigns a focus feature to the preceding constituent (King 1995). Clitic *LI* appears in Russian, as well as several other Slavic languages, i.e. Bosnian/Croatian/Serbian, Bulgarian, and Macedonian (Schwabe 2004, King 1995). Insertion of *LI* is one of the two main strategies to form a *yes-no* matrix question in Russian, with intonation being the second strategy. While this clitic is optional in main clause questions, it is obligatory in embedded questions (King 1995, Schwabe 2004). *LI* usually occupies the second position in the sentence, appearing after the first phonological word (King 1995).

Clitic *LI* can override the word order in terms of determining the information structure pattern of the sentence. According to King (1995), when *LI* attaches to a non-verbal constituent, it marks it as focused:

- 51) [Knigu] li ona čitaet?
Book Q she read
'Is it the book-FOC that she is reading?'

(King 1995: 80)

In the example above *LI* marks preceding noun *knigu*/'book' as focused. If *LI* attaches to a verb, the entire clause can be focused, resulting in a 'simple' yes-no question (King 1995: 153):

- 52) Živet li on zdes'?'
Live Q he here
'Does he live here?'

(King 1995: 139)

King suggests that in the matrix *yes-no* question (52) with a neutral intonation the whole clause is questioned and the verb is not focused. However, if there is an intonational stress on any of the constituents, this intonationally stressed constituent will be the focus of the sentence (King 1995: 151-152), even if *LI* does not attach to this constituent. King demonstrates this using the examples of embedded *LI*-questions, but this analysis would apply to matrix clauses as well.

Clitic *LI* and its role in marking information structure, in particular its ability to signal presence of contrastive focus in the sentence will be examined in more detail in section 4.5.3 with a more detailed analysis of *LI*.

3.4 Summary

In the preceding sections I have reviewed the literature on general theory of information structure, as well as the tests to identify components of information structure, and several accounts of information structure marking strategies in Russian, including clitics –*TO*, *ŽE*, and *LI*. Since it has already been shown that the information structure notions and tests described earlier can be contradictory, it remains a question to answer to what extent they can be used to identify information structure components in sentences when applied to naturally occurring language data.

Chapter 4: ANALYSIS

4.1 Introduction

This chapter presents an analysis of the examples I extracted from the Russian National Corpus. In section 4.2 I will briefly discuss the reasons why it is difficult or impossible to apply contrastiveness tests used by Gryllia (2008) to the spoken data containing clitics from the Russian National Corpus (Grishina 2006). In section 4.3 I will examine the differences between contrastive topic and contrastive foci and formulate diagnostics which can be used to differentiate between them. In section 4.4 I will list the tools necessary for the analysis of the spoken data, which includes new working definitions of information structure components, as well as summary of tests applicable to each component. I will also propose a working information-structure hierarchy that will be used in this thesis. In section 4.5 I will examine in detail Russian spoken data containing *–TO* and *ŽE*, applying definitions and tests formulated in section 4.4 in order to identify different information structure patterns in the sentences. Finally, I will attempt to extend the analysis of *–TO* and *ŽE* as markers of contrast to some cases of *LI* and *it*-clefts, applying the same contrastiveness tests and definitions as I have used for *–TO* and *ŽE*.

4.2 Applying Contrastiveness Tests to the Spoken Data

One possible way to identify contrastive elements in the sentence is to apply contrastiveness tests to the elements under consideration. This approach was attempted to analyze the spoken data in the Russian National Corpus used in this thesis, particularly in the cases when contrastiveness of the elements was not easily identified just based on the surrounding context of the sentence. The data that was most challenging in terms of identifying contrastive focus and topic included *wh*-questions containing clitics, as well as data in which contrastive topic was hard to distinguish from contrastive focus. These cases will be discussed in more detail in the following sections.

The contrastiveness tests used by Gryllia (2008) were not always applicable to Russian data containing clitics or produced conflicting results. The main characteristics of the data that was tested in this thesis that made testing challenging were:

I. Illocutionary force: questions vs. exclamatives vs. declaratives

Contrastiveness tests used in Gryllia (2008) are best suited to **declaratives**. Data that contains clitics –*TO* and *ŽE* is often interrogatives or exclamatives. Data that contains clitic *LI* is always interrogatives. Applying contrastiveness tests (Gryllia 2008) without ‘reversing’ them (in the rare cases when it was possible) was not possible for the questions and was often challenging for exclamatives. More details about applying contrastive tests to questions will be discussed in the sections below.

II. Presence of the clitic in the sentence

Presence of (contrastive) clitic(s) in the sentence affected the information structure pattern in such a way that contrastiveness tests used in Gryllia (2008) were challenging to apply as well:

a) Number of contrastive elements in the sentence

Often sentences with clitic –*TO* contain two contrastive elements, topic and focus. Most of the contrastiveness tests used in Gryllia (2008), except for the implicit sub-question test, are oriented towards sentences with one contrastive element.

b) Contrastive focus projection of the clitic

The contrastiveness tests used in Gryllia (2008) seem to be suited better for identifying contrast of a **constituent** and not of an entire **proposition**. When clitic marks the whole proposition as the focus, it is more challenging to apply contrastiveness tests, which will be discussed more in the section on clitic *ŽE*.

c). Contrastive focus vs. contrastive topic

The contrastiveness tests used in Gryllia (2008) are not always conclusive in differentiating between contrastive focus and contrastive topic. As will be shown in the section 5.4.1.1, the *wh*-question and implicit sub-question tests used by Gryllia (2008) are not sufficient to make a distinction between contrastive focus and contrastive topic in sentences with *-TO*.

d). Informality/Aggressiveness of clitics

As will be shown in the section describing clitic *ŽE*, most of the tests cannot really be applied (or they pass with marginal results) to the sentences containing this clitic. The reason is due to the fact that this clitic adds an element of aggressiveness to the sentence. The aggressiveness of this clitic makes sentences containing this otherwise frequently used clitic seem unacceptable or marginally acceptable in formal tests for identifying information structure components.

Clitic *-TO* is not aggressive, therefore it was easier to apply contrastiveness tests to the data containing this clitic. However, this clitic is used very informally and it is considered to be emphatic, which made some of the data containing this clitic marginal as well.

As to clitic *LI*, it was sometimes challenging to apply contrastiveness tests to data containing this clitic as well. Clitic *LI* is a second way to form questions, with another way to ask questions in Russian being forming them with intonation. As it was found in the process of analyzing and testing the spoken data for contrastiveness in this thesis, forming questions with *LI* is a less preferred way to ask questions, making questions containing *LI* more challenging to apply contrastiveness tests.

More detailed analysis of each clitic, as well as supporting data from Russian National Corpus are below.

4.3 Contrastive Focus vs. Contrastive Topic

After the discussion of **contrastive focus** and **contrastive topic** in the previous chapter, it should be noted that it is quite challenging to distinguish between these two information structure components. Sometimes **contrastive topics** are not even considered to be a separate information structure category, but are considered to be topics that contain a focus (Krifka 2007), which makes it even harder to distinguish between these two categories.

McCoy (2001) briefly mentions that the main difference between **contrastive topics** and **contrastive foci** is the information status of their referents and moves on to discuss Vallduví and Vilkuna's analysis (1998). Vallduví and Vilkuna (1998) claim that contrastive topics cannot be contrastive foci because they are "thematic" by definition. They provide two examples of **contrastive topics** (or, in Vallduví and Vilkuna's terminology, *kontrastive themes* or *thematic contrasts*):

53) **The first 100 meters** she ran [_R in a record TIME].

54) **Beer** I [_R LIKE].

(Vallduví and Vilkuna 1998: 87)

In the examples above elements in **bold** are contrastive topics, while elements in [square brackets] are semantic foci (rhemes in Vallduví and Vilkuna's terminology). Vallduví and Vilkuna (1998) note that contrastive topics (a.k.a. thematic contrasts) in English have corresponding syntactic configuration, topicalization, and/or intonational marking.

While Vallduví and Vilkuna's examples above are very useful illustrations of contrastive topics, it is still somewhat challenging to differentiate between contrastive foci and contrastive topics without some tests or other reliable diagnostics tools. Topics and foci can be identified as contrastive as sharing characteristics of contrast, described by Molnár's (2002). However, a major challenge is to further differentiate (contrastive) topics from (contrastive) foci, which is partially achievable by examining the overarching differences in the notions of **focus** and **topic**, such as 'aboutness' or 'newness' of the information.

To identify a (contrastive) topic and differentiate it from (contrastive) focus the 'tell-me-about' test (Choi 1999) could be used, as in the example below (it should be

noted that [R](heme) abbreviation used by Vallduví and Vilkkuna is replaced by [FOCUS] abbreviation in my thesis]:

55) Speaker 1:

Yesterday she ran in a marathon.

Speaker 2:

Tell me about her first 100 meters.

Speaker 1:

The first 100 meters she ran [FOCUS in a record time].

In the example above constituent '*the first 100 meters*' answers 'tell-me-about' question, thus passing Choi's 'tell-me-about' test for **topic**. Additionally, it also fits Krifka's definition of topic as an entity under which the information is expressed in the focus should be stored in the Common Ground (Krifka 2007). Furthermore, several characteristics of contrast show that this constituent is also contrastive. A set of alternatives implied in this dialogue, i.e. { '*the first 100 meters*', '*the second 100 meters*', '*the last 100 meters*', etc.) is limited. Since it is common knowledge that the marathon distance is 42,195 meters, the members in this set are bound to be within this range. According to Choi's 'tell-me-about' test, as well as Vallduví and Vilkkuna's definition of contrast and Molnár's characteristics of contrast discussed earlier, constituent '*the first 100 meters*' can be considered a **contrastive aboutness topic**.

Another test to confirm the results above is Gryllia's (2008) implicit sub-question test:

56) Speaker 1:

Yesterday she ran in a marathon.

Speaker 2:

How quickly did she run?

Speaker 1:

(a). **The first 100 meters** she ran [FOCUS in a record time].

(b). **The last 100 meters** she ran [FOCUS very slowly].

In example (56) above Speaker 2 asks a general *wh*-question that can be interpreted as containing two (or more) sub-questions, i.e. 'how quickly did she run the first 100

meters?’ and ‘how quickly did she run the last 100 meters?’ In their answer Speaker 1 answers these two sub-questions in (a) and (b) accordingly.

It is not possible to apply the substitution test used by Gryllia (2008) that identifies contrastive topic to confirm the results above. The substitution test applies only to sentences with more than one topic under consideration, while sentence above contains only one topic. Substituting the topics in question for ‘*the former*’ and ‘*the latter*’ necessary for the substitution test used by Gryllia (2008) is therefore not possible.

In addition to **contrastive aboutness topics** shown in the previous example, there are also **contrastive frame-setting topics** that should also be distinguished from **contrastive foci**. **Contrastive frame-setting topics** can also be identified using Gryllia’s implicit sub-question test, as in the example below:

57) Question:

Куда поедет спортивная команда?
Kuda poedet sportivnaya komanda?
Where go-3SG.FUT sport-SG.FEM team?
‘Where is the sports team going to go?’

Answer 1:

Сегодня они поедут в Петербург.
Segodnja oni poedut v Peterburg.
Today they go-3PL.FUT to Petersburg.
‘Today they are going to Saint Petersburg.’

Answer 2:

Завтра они поедут в Москву.
Zavtra oni poedut v Moskvu.
Tomorrow they go-PL.FUT to Moscow.
‘Tomorrow they will go to Moscow.’

Although Choi’s ‘tell-me-about’ test used for aboutness topic above cannot be applied to this example, adverb *сегодня* /‘today’ can still be identified as a topic based on Krifka’s definition (2007). Within the context of the utterance above, two implicit sub-questions can be formed to result in answers (1) and (2), confirming that the adverbs *сегодня* /‘today’ and *завтра* /‘tomorrow’ are contrastive topics. Furthermore, *сегодня* /‘today’ and *завтра* /‘tomorrow’ are also identified as frame-setting (contrastive) topics within the context above because they have characteristics of frame-setting topic

described by Dipper et al. (2007). As with the previous example, the substitution test cannot be applied because there is only one topic in the sentence tested. Additionally, due to the nature of the frame-setting topics (often formed with adverbs), the test would not be applicable to this sentence, as it is challenging to substitute words such as ‘today’ and ‘tomorrow’ with ‘former’ and ‘latter,’ which is the requirement of the substitution test.

Based on the tests above, it appears that aboutness (contrastive topics) can be identified using Choi’s (1999) ‘tell-me-about’ test to differentiate them from (contrastive) foci. However, Choi’s test does not apply to differentiate frame-setting (contrastive) topics from foci. Gryllia’s (2008) implicit sub-question test can be used to identify contrast in both aboutness and frame-setting topics, but her substitution test does not apply unless sentence contains two contrastive topics in question.

It seems that the most reliable diagnostics to differentiate between (contrastive) topics and (contrastive) foci include Krifka’s (2007) definition of topic applicable to both aboutness and frame-setting topics, Choi’s (1999) ‘tell-me-about’ test and Lambrecht’s (1996) definition of topic in terms of ‘aboutness’ to identify (contrastive) aboutness-topics, and Gryllia’s (2008) implicit sub-question test to identify both contrastive aboutness and contrastive frame-setting topics. However, as it will be shown in the sections with analysis of Russian spoken data, the results of diagnostics listed above are not always conclusive when applied to spoken data, especially without full knowledge of the context of the sentence or when applied to interrogatives.

4.4 Summary of tests and assumptions necessary for analysis

In this section I will summarize the tools such as notions and tests to identify information structure components described in earlier sections, as well as propose some new working definitions that will be employed in this chapter to analyze data. I will be identifying information structure components in spoken data based on the following:

58) Topic will be identified using Krifka’s definition below:

The topic constituent identifies the entity or set of entities under which the

information expressed in the comment constituent should be stored in the CG¹⁴ content.

(Krifka 2007: 41)

Topics are divided into aboutness and frame-setting topics, which in their turn are divided into contrastive and non-contrastive subcategories. Instead of defining contrastive aboutness and frame-setting topics based on Krifka's (2007) definition of contrastive topics as aboutness topics that contain a focus, they are instead defined in different terms to keep notions of 'topic' and 'focus' separate from each other, but at the same time to retain Krifka's understanding of contrastive topics as implying a set of alternatives.

- **Aboutness topics** are the entities about which the sentence makes a predication (Dipper et al. 2007: 163).
 - A **non-contrastive aboutness topic** is an aboutness topic without characteristics of contrast defined later. It can be identified using 'tell-me-about' test (Choi 1999).
 - A **contrastive aboutness topic** is an aboutness topic that implies a limited membership set of alternatives in which one member is contrasted/highlighted to another syntactically or semantically comparable member. It can be identified using Gryllia's (2008) implicit sub-question test:
 - (i) When a *wh*-question can be split into sub-questions and the answer is organized per sub-question, then, there is a contrastive topic in the answer.
 - (ii) When a question can be interpreted as containing more than one implicit sub-question, and the answer addresses only one of these sub-questions, rather than the general question, then, this answer contains a contrastive topic.

(Gryllia 2008: 37)

¹⁴ As has been mentioned earlier and repeated here for convenience, the notion of CG (Common Ground) used above refers to the information "mutually known to be shared and continuously modified in communication" (Krifka 2007: 4).

- An additional test for contrastive aboutness topic (although it requires presence of two topics in question) is the substitution test (adopted from Benincà and Poletto (2004)):

If two terms are interpreted with a ‘list interpretation’, then they can be substituted with ‘the former’ and ‘the latter’.

(Gryllia 2008: 33)

- **Frame-setting topics** denote the frame in which the main predication of the sentence is interpreted. Frame-setting topics usually describe the time or the location of the event (Dipper et al. 2007: 167).
 - A **non-contrastive frame-setting topic** is a frame-setting topic without characteristics of contrast defined later.
 - A **contrastive frame-setting topic** is a frame-setting topic that implies a limited membership set of alternatives in which one member is contrasted/highlighted to another syntactically or semantically comparable member. It can be identified using Gryllia’s (2008) implicit sub-question test.

59) Focus refers to **new** information in the sentence, based on definitions provided by King (1995), Lambrecht (1996), and Dipper et al. (2007). Focus is divided into two main subcategories:

- (a). **Semantic focus** (a.k.a. **new-information focus** or **non-contrastive focus**) is a focus with a set of alternative members. The number of members in the set is unlimited and the information as to which member from the set will be chosen is unknown/unpredictable to the speaker. Test that will be used to identify semantic focus is the following:

***Wh*-question**

A focused element answers *wh*-question.

(Lambrecht, 1996; Gundel, 1999)

(b). **Contrastive focus** is also a focus with a set of alternative members.

However, the number of members is limited and known to the speaker and the hearer, therefore it is possible to predict to some extent the answer, which is restricted to be one of the known members of the set. Tests that will be used to identify **contrastive focus** used by Gryllia (2008, 31-32):

(1). “***Wh*-question/*contrastive answer test**

A contrastive answer is incompatible with an ordinary *wh*-question.”

(test adopted by Gryllia from Rizzi (1997))

(2). “**Correction test**

A contrastive focus can be used to answer a yes-no question, correcting part of the predicate¹⁵ information of the question.”

(test adopted by Gryllia from Kiss (1998))

(3). “**Choice test**

When answering an alternative question, one alternate¹⁶ is contrasted to the other.”

(test adopted by Gryllia from Kiss (1998))

60) Contrast is defined combining Molnár’s characteristics of contrast (2002) and Vilkuna’s definition of kontrast (1998) and Dipper et al.’s definition of contrast (2007):

¹⁵ Term “predicate” was not defined either by Gryllia (2008) or by Kiss (1998), but based on their usage of this term it is assumed in this thesis that it refers to a “presupposition.”

¹⁶ Gryllia uses the term “alternate” (not “alternative” in the second part of this definition)

Expression **a** is contrastive if membership set $M=\{..., a, ...\}$ is generated and is available to both the speaker (and the hearer). Set **M** has a limited number of members that are syntactically and/or semantically comparable. Member **a** is contrasted/highlighted in relation to the other members.

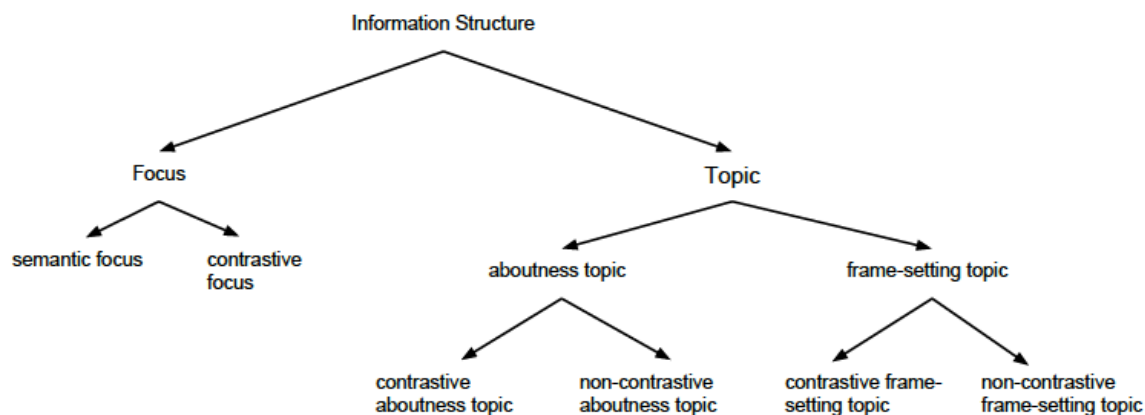
61) **Contrast marker** is defined following McCoy (2001), Molnár's characteristics of contrast (2002) and Vilkuna's definition of kontrast (1998) and Dipper et al.'s definition of contrast (2007)

A **contrast marker** is a linguistic expression – syntactic, morphological, prosodic, or lexical that signals contrast. The element marked by contrastive marker belongs to a limited membership set with syntactically and/or semantically comparable members and in which this element is contrasted/highlighted in relation to the other member(s).

Information Structure hierarchy

This thesis provisionally organizes the information structure subcategories in the following way:

Figure 4.4: Information structure hierarchy used in this thesis



It should be noted that the hierarchy above is work in progress and needs further research and analysis of data to be confirmed. A potential addition to this hierarchy is a subcategory for elements which are simultaneously contrastive and semantic foci. However, this requires further investigation in terms of understanding the differences in focus projection marked by these two foci and will not be addressed in this thesis. Also, it should be noted that the tests and working definitions adopted and summarized in this section are not always conclusive and can produce ambiguous results, as will be shown in the following sections.

4.5 CLITICS

In the following sections I will analyze spoken data containing the clitics *–TO* and *ŽE* in terms of the information structure patterns these clitics mark. For both clitics, I will start with applying several tests to identify information structure patterns in the example sentences containing these clitics used by King (1995) and McCoy (2001), as well as sentences from the Russian National Corpus. Using data containing clitics from the Russian National Corpus, I will identify the words to which these clitics most frequently attach. Additionally, I will examine information structure patterns marked by these clitics in interrogatives, both in *yes-no* and *wh*-questions from the Russian National Corpus. I will also examine information structure patterns in the topic-less sentences containing *–TO*, as well as briefly look at the interaction of *–TO* with *ŽE* and other contrastive elements. For *ŽE* I will also attempt to establish the correlation (if any) between its contrastive focus projection(s) and its position in the sentence.

4.5.1 CLITIC *–TO*

4.5.1.1 *–TO* as a marker of contrastive topic and contrastive focus

McCoy's central example denoting a set of sets of propositions, which has been described earlier, is repeated below:

- 62) [Varja and her mother are looking at a picture of a dog who put her paw on a bear. They first discuss the dog, then start talking about the bear. Varja gets distracted by taking a scoop into her hand. Mother says:] (CHILDES, seance 2)

U tebjā-*TO* sovok, a čto u medvedja v lape?

At you-*TO* scoop but what at bear in paw

"YOU(-*TO*) have a SCOOP, but what does the BEAR have in his paw?"

(McCoy 2003: 322)

McCoy argues that in the example above *tebjā*/'you' is **contrastive topic** and *sovok*/'scoop' is **contrastive focus**. In this sentence only the first clause is going to be analyzed. Below I will examine both *tebjā*/'you' and *sovok*/'scoop' for contrastiveness of topic and contrastiveness of focus accordingly using contrastiveness tests from Gryllia (2008). Below the *wh*-question test for **contrastiveness of focus** (Gryllia 2008) is applied to noun *sovok*/'scoop':

- 63) (a) Что у тебя?
What at 2.SG.GEN?
'What do you have?'

(b) *U tebjā-*TO* sovok
At 2SG.GEN-*TO* scoop
'YOU(-*TO*) have a SCOOP'

The noun *sovok*/'scoop' in the answer (b) is incompatible with the *wh*-question in (a), thus passing the *wh*-question test for **contrastive focus**. Based on the *wh*-question test used in Gryllia (2008), as well as King's (1995) and Lambrecht's (1996) definitions of focus in terms of newness of the information, *sovok*/'scoop' is a **contrastively focused** element. However, as it has been noted earlier in the general discussion of tests and how they can be applied to data containing clitics, sentence (b) above, as well as many following sentences with the clitic -*TO* (and other clitics examined in this paper) is only marginally acceptable in the context of testing, due to the emphatic, informal, and sometimes even aggressive nature of the clitics it contains. The sentences like (b) often become acceptable if the clitic is omitted (provided that the information structure role of the element marked by the clitic is preserved by some other means, e.g. prosodical means). The sentences that are marginally acceptable with the clitic and acceptable

without the clitic, will be identified by question mark (?) for marginal acceptability and (*clitic*) to differentiate them from the other sentences. An example will be used in the later in the section.

While noun *sovok*/'scoop' passes the *wh*-question test confirming McCoy's analysis of it as **contrastive focus**, pronoun *tebja*/'you' to which *-TO* attaches is more problematic in terms of its identification as **contrastive focus** or **contrastive topic**. On one hand, pronoun *tebja*/'you' passes the *wh*-question test identifying **contrastive focus** used by Gryllia (2008):

- 64) (a) У кого совок?
At whom scoop?
'Who has the scoop?'

(b) *U tebja-TO sovok
At 2SG.GEN-TO scoop
'YOU(-TO) have a SCOOP'

Based solely on the *wh*-question test used in Gryllia (2008), since answer (b) is an ungrammatical response to question (a), pronoun *tebja*/'you' is a **contrastively focused** element. This either means that McCoy's (2001) analysis of pronoun *tebja*/'you' as a contrastive topic is incorrect or that *wh*-question test used by Gryllia (2008) cannot be applied to sentence above. Presence of clitic *-TO*, which is often referred to as "emphatic" in literature, is a possible culprit in failing of *wh*-question test, which might not be a sufficient condition for identifying contrastive focus-hood in sentences containing *-TO* or other emphatic elements.

On the other hand, pronoun *tebja*/'you' passes the implicit sub-question test used by Gryllia (2008) for identifying **contrastive topic**:

65) Question:

What do they (*i.e. referents in the discourse*) have?

Answers:

- (1). (?)You(-TO) (*Varja*) have a scoop.
- (2). The bear has the dog.

Since sentence (1) above is one of the possible answers to the implicit sub-question, this sentence passes the implicit sub-question test. As mentioned earlier in the section, this sentence has a question mark (?) and clitic (*-TO*) in parentheses. This means that this

sentence is marginally acceptable with the clitic, but acceptable without the clitic (provided that *tebja*/'you' retains its contrastive focus meaning through prosodical and syntactic means, such as sentence-initial position in the sentence). Based solely on the implicit sub-question test, pronoun '*тебя*/'you' appears to be a **contrastive topic**. This either confirms McCoy's analysis of this element as a contrastive topic, or demonstrates the inconclusiveness of both this test and the earlier *wh*-question test.

Identifying pronoun *tebja*/'you' as a **contrastive topic** (instead of **focus**) is further supported by applying the implicit sub-question test above in the 'wrong way', i.e. while pronoun *tebja*/'you' passes implicit sub-question tests as a **contrastive topic** (as seen above), it fails it as a **focus**:

66) Question:

?Who has these things?

(implying '*things under discussion*', i.e. *the scoop and the dog*)

Answers:

(1). *You-TO (Varja) have a scoop.

(2). *The bear has the dog.

Firstly, it is very difficult to form a question corresponding to answers (1) and (2). Secondly, even with this question being marginally acceptable, it is not going to result in the answers (1) and (2). Both the impossibility of forming a question and its incompatibility with answers suggest that the sentence under consideration, i.e. 'You-TO (Varja) have a scoop' fails the implicit sub-question test for identifying *tebja*/'you' as a **focus** (and *sovok*/'scoop' as a **contrastive topic**).

In order to come to more conclusive results than *wh*-question and implicit sub-question tests permit, the 'tell-me-about' test (Choi 1999) could be applied to determine whether first pronoun *tebja*/'you' is a (contrastive) topic or a (contrastive) focus. Whether pronoun *tebja*/'you' can be identified as a (contrastive) topic could be based on 'tell-me-about' test:

67) Question:

Tell me about myself.

Answer:

*You-TO (Varja) have a scoop.

Based on the fact that answer above is ungrammatical in the context of preceding question, pronoun *tebja*/'you' fails the 'tell-me-about' test, suggesting that it is not a **topic** (or not a contrastive topic in the example above).

Based on the results of applying the *wh*-question and implicit sub-question tests (Gryllia 2008), 'tell-me-about' test (Choi 1999), it appears that these tests are quite inconclusive in terms of differentiating **contrastive focus** from **contrastive topic**. Instead, Krifka's definition of topic (2007) repeated below is going to be used to describe (contrastive) topic:

68) Topic

The topic constituent identifies the entity or set of entities under which the information expressed in the comment constituent should be stored in the CG content.

(Krifka 2007: 41)

Examples of topic provided by Krifka are:

- 69) a. [Aristotle Onassis]_{Topic} [married Jacqueline Kennedy]_{Comment}.
b. [Jacqueline Kennedy]_{Topic} [married Aristotle Onassis]_{Comment}.

(Krifka 2007: 42)

In the example (a) above topic [Aristotle Onassis] identifies the entity, under which the information expressed in the comment [married Jacqueline Kennedy] is stored. If the same analysis is applied to Russian sentence below, the result is very similar:

- 70) [У тебя]_{Topic-TO} [совок]_{Comment}.
[At you]_{Topic-TO} [a scoop]_{Comment}.

In the sentence above, topic [you] identifies the entity (grandchild *Varja* in this context), under which the information expressed in the comment ('has [a scoop]' in this context), is stored. This suggests that *tebja*/'you' is a **topic**. Additionally, Molnár's (2002) characteristics of contrast, describe pronoun *tebja*/'you' as **contrastive**:

To summarize the results above, implicit sub-question test (Gryllia 2008) shows that pronoun *tebja*/'you' is **contrastive topic**, while 'tell-me-about' test (Choi 1999) shows that it is **not a topic** and *wh*-question test (Gryllia 2008) shows that it is a **contrastive focus**. The inconclusivity of the currently existing tests to identify information structure components in the sentence suggests that further research is necessary in order to establish more reliable tests to identify contrastive topics and

contrastive foci in Russian spoken speech. However, for the purposes of this thesis, I am going to rely on the results of the implicit sub-question test (Gryllia 2008), Krifka's definition of **topic** (2007) and Molnár's characteristics of **contrast** (2002), identifying pronoun *tebja*/'you' is a **contrastive topic**.

4.5.1.2 Different roles of *-TO*

As mentioned earlier, clitic *-TO* appears in different contexts and plays very different roles. This clitic can be spelled with or without hyphen, and its position in the sentence varies. Since this thesis is mostly concerned with the clitic *-TO*¹⁷ spelled with the hyphen, data from Russian National Corpus¹⁸ below is calculated only for this spelling:

Table 2: Words to which clitic *-TO* attaches most frequently

Words to which <i>-TO</i> attaches	Number of sentences	% from total data
<i>wh</i> -words	474	70.85%
nouns	52	7.77%
verbs	38	5.68%
adverbs	27	5.70%
pronouns	25	5.27%
deictics	14	2.09%
personal names	6	0.90%
adjectives	3	7.89%
other ¹⁹	30	4.48%

Based on the data summarized in table 2 above, most often clitic *-TO* (in 70.85% of the total 569 sentences) attaches to *wh*-words, forming indefinite pronouns and not marking contrast. When marking contrast (in the less than 29% remaining), *-TO* most often attaches to nouns, adverbs, pronouns, and verbs. More rarely it attaches to deictics, personal names, and adjectives.

¹⁷ I did not include *TO* (spelled without the hyphen) in order to exclude all cases when it forms disjunctions or when it is a deictic, as it seemed not as relevant to this thesis.

¹⁸ Data consisted of 669 sentences with *-TO*.

¹⁹ The "other" category consists of words that were ambiguous in terms of its identification as part of speech or that were erroneously transcribed to the point of not being understandable.

In the following paragraphs I will briefly mention several cases of *-TO* which are not going to be discussed later in the thesis. These usages include cases when *-TO* does not mark contrast at all (when it forms an indefinite pronoun) and when it marks contrast, but in different types of constructions than that are discussed for this thesis, i.e. conditional contrastive propositions and disjunctions.

As mentioned earlier, clitic *-TO* most frequently attaches to *wh*-words, forming indefinite pronouns and not marking contrast. For example, if *-TO* attaches to *wh*-words, indefinite pronouns are formed: *как*/'how' (*как-то*/'somehow'), *что*/'what' (*что-то*/'something'), *какой*/'which' (*какой-то*/'some'), *где*/'where' (*где-то*/'somewhere'), *кто*/'who' (*кто-то*/'someone'), *почему*/'why' (*почему-то*/'for some reason'), etc. In the example below *-TO* attaches to the word *как*/'how':

- 71) Вообще как-то страшно. Человек работал и вдруг хоп и нет ничего.
Overall how-TO terrifying. Person working and suddenly whoops and no nothing
'All in all, it is somewhat terrifying. A person was working, and suddenly, whoops, and there is nothing.'

(Russian National Corpus)

In the example above *-TO* attaches to *wh*-word and does not mark it for contrastiveness.

There are also two cases in which *TO* (spelled without the hyphen) marks contrast, but which are not central for this thesis. First case involves *-TO* forming a disjunction (*то (ли) ... то (ли) / to (li) ... to (li) ... / 'either ... or'*):

- 72) Он то хорошо учится, то плохо.
He TO well studies TO badly
'Now he studies well, now he studies badly.'

The adverbs *хорошо*/'well' and *badly*/'плохо' above are contrasted to each other. It should also be noted that similar disjunctions are also formed with another clitic discussed in this thesis, clitic *LI*. The presence of *LI* in these sentences adds an element of doubt/questionability, which is to be expected due to its main function as an interrogative complementizer. Sentence below is an example of a disjunction formed with *TO LI*:

- 73) Он то ли хорошо учится, то ли плохо.
He TO LI well studies TO LI badly
'He is either studying well or badly'
(it is not clear to the speaker, which one it is)

As in the previous example, the adverbs *хорошо*/'well' and *badly*/'плохо' are contrasted to each other, but with an element of questionability added by clitic *LI*. A further investigation of disjunctions formed by *TO* and *TO LI* is outside the scope of this thesis, but would be beneficial to better understand the role both clitics play in marking information structure, as well as their interaction.

Another case of *TO* (spelled without the hyphen) marking a contrast but not discussed further in this thesis, is a **conditional contrastive proposition** (formed with clitics [*a to by .../otherwise would*']), as in the example below:

- 74) Я рада, что мы не пошли в магазин. А то бы я ничего не успела сделать.
 I glad that we not went to store. A TO BY I nothing not have_time do
 'I am glad that we haven't gone to the store. Otherwise I would have not had enough time to do anything.'

In the example above *TO* is a part of a contrastive propositional expression *A TO BY* and signals contrast. Presupposition '*I had the time to do something*' based on the proposition '*because I did not go to the store*' is contrasted with the presupposition '*I would not have been able to do anything*' in the case '*had I gone to the store.*'

Although outside the scope of this thesis, usages of clitic *TO* as contrastive marker including the cases discussed in this section need further investigation in order to gain a better understanding of how this clitic marks contrast.

4.5.1.3 –*TO* attaching to different parts of speech

In this section I will briefly examine data in which –*TO* attaches to different parts of speech, i.e. *nouns*, *verbs*, and *adverbs*. I will try to determine whether the part of speech to which –*TO* attaches is an important factor in defining its information structure role.²⁰ Using examples below, I will show that the variations in the information structure patterns in sentences containing –*TO* are dependent on the context, rather than on the part

²⁰ I attempted to examine the differences between information structure patterns in sentences in which –*TO* attaches to different parts of speech to establish whether these information structure patterns are different for verbs. In such case this clitic would behave similar to clitic *LI*, which has scope ambiguity resulting in two possible information structure patterns when *LI* attaches to verbs.

of speech to which *–TO* attaches. In the paragraphs below I will examine sentences in which *–TO* attaches to nouns, verbs, and adverbs.

In example (62) earlier in the chapter and examples later in the chapter clitic *–TO* attached to a pronoun or noun marks topic and focus for contrast, as predicted by McCoy (2001): [CT]-TO..[CF]. Example (75) is an example of the sentence in which *–TO* attaches to noun and does not have information structure pattern predicted by McCoy (2001). Below is an excerpt of a dialogue, which is my own example recorded from day-to-day conversation. In the example below *–TO* does not mark topic or focus for contrast:

75) Speaker 1:

Митя, у нас опять из мясного есть ничего.
 Mitja at 1PL.GEN again from meat.GEN eat nothing
 'Mitja, again we don't have anything meat(y) to eat.'

Speaker 2:

Как ничего?
 How nothing
 'How (can it be) nothing?'

Speaker 1:

Сосиски-то мы вчера доели.
 [SF Sausages-TO] 1PL.NOM yesterday [SF ate.PL]
 'We ate sausages yesterday.'

In the sentence with *–TO* above Speaker 1 is pointing to Speaker 2 the fact that they are out of meat products at home, because they have eaten the sausages, which were the last meat product at their home. Speaker 1 chooses *сосиски*/'sausages' from a set of alternatives belonging to meat products, which indicates that this is a **focus**. Since the entity denoted by *сосиски*/'sausages' is not contrasted to any other comparable entities in this utterance, noun *сосиски*/'sausages' is not contrastive. It can be concluded that the element (*сосиски*/'sausages' in this sentence) to which *–TO* attaches is non-contrastive **semantic focus**. Given examples such as (75) it is impossible to state that *–TO* is an unambiguous marker of contrast, but it can also mark semantic focus. Additional information, such as the state of the Common Ground, is necessary to disambiguate between these possibilities.

Similarly to the examples above with *–TO* attaching to nouns, sentences with *–TO* attaching to verbs can also have different information structure patterns. Below I briefly examine two types of sentences with *–TO* attaching to the verb: in (76) *–TO* marks only focus for contrast, while in (77) *–TO* does not mark contrast at all.

Below is an example of a declarative sentence in which *–TO* attaching to a verb marks only focus for contrast. In this excerpt of a recording of a family conversation a female speaker is anticipating separation with the father of her daughter. She believes that as a result he will not be visiting their daughter very often, which is not what her child needs. The speaker believes that her child needs an everyday contact with her father. However, as the speaker says, there is nothing she can do about it at this point:²¹

76) Speaker 1:

Да? А чё я Светке-то скажу?
 Yes? But what 1SG.NOM Svetka-DAT-TO say-1SG.FUT?
 ‘Yes? And as to Svetka-TO, what will I tell her?’

Будет / конечно / к ней приехать / когда время будет.
 be.3SG.FUT / of course / to 3SG.DAT come-INF / when time be.3SG.FUT
 ‘(He) will, of course, visit her when there is time.’

То есть / я так понимаю / не зачастит...
 Therefore / 1SG.NOM so understand.SG.PRS / no frequent-3SG.FUT.
 ‘So, as I understand it, he will not frequent (daughter).’

И ведь сделать-то уже ничего не поделаешь.
 And ved’ [T to_do-TO] already [CF nothing not do]
 ‘And as to doing something at this point, you cannot do anything.’
 (or ‘And there is nothing you can do at this point.’)

Но разве «этого» ребёнку надо?
 But really this-MASC.GEN child-DAT necessary?
 ‘But is this really what a child needs?’

Ей отец нужен каждый день..
 3SG.DAT father-SG.MASC.NOM necessary every day
 ‘She needs a father every day.’

(Russian National Corpus)

²¹ It should be noted that the contrastive clitic *–TO* in the first utterance will not be discussed for this example.

Below the sentence with the clitic is repeated again:

И ведь сделать-то уже ничего не поделаешь.
And ved' [_T to_do-TO] already [_{CF} nothing not do]
'And as to doing something at this point, you cannot do anything.'
(or 'And there is nothing you can do at this point.')

In the sentence above *сделать-то*/'to do-TO' is analyzed as a **non-contrastive topic** according to Krifka's definition of topic (2007), due to its lacking contrastive characteristics described by Molnár (2002), passing of the 'tell-me-about' test (to the extent that this test can be applied to identify verbal topics) to identify topic (Choi 1999), and the failure to form an implicit sub-question test to identify it as contrastive topic (Gryllia 2008). It is not analyzed as focus because there is no set of alternatives implied. Meanwhile, constituent *ничего не поделаешь*/'nothing you can do' is analyzed as **contrastive focus**, based on the failure of the *wh*-question test. The constituent '*ничего не поделаешь*/'*you cannot do anything*' is considered to be contrastive because it is a member of a set with two comparable members, one of which is contrasted to another. First member in this set represents presupposition that the speaker '*could do something about father of her child not leaving*' and another one represents presupposition '*she cannot do anything at this point, as it is too late.*'

If in the example above –*TO* marks only focus for contrast, in *wh*-question example below it does not mark for contrast either topic or focus. This is an excerpt of a recording of a day-to-day conversation:

77) Male speaker 1:

Ну а потом плюнул. Иди боком домой / думаю.
Well and then spat Go sides home / think-1SG.PRS
'Well and then I didn't care. Go home, I thought (to myself).'

[Смеются] [Laughing]

Разозлился и ушел.
Got_angry-MASC and left-MASC
'I got angry and left.'

Male speaker 2:

А ты где был-то? В подъезде?
And 2SG.NOM where be-TO? In entrance?
'And where were you? Near the entrance?'

Male speaker 2:

Ну. да! В подъезде у нее...
Well yes! In entrance at her
'Well, yes! Near the entrance to her place...'

Below only the sentence containing *–TO* is repeated:

А ты где был-то?
And 2SG.NOM where be-TO?
And 2SG.NOM [_{SF} where] [_T be-TO]?
'As to your being in terms of location, where were you?'

Following the same diagnostics used to identify contrast, topic, and focus in the sentence as were used for previous sentences, in this example verb *был*/'be' to which *–TO* attaches is a **non-contrastive topic**, while the *wh*-word is **semantic focus**, with the following information structure pattern:

78) [_{SF} *wh*-word] [_T be]-TO?

There are cases²², as shown above, in which verb '*be*' is an **aboutness non-contrastive topic** and *wh*-word is **semantic focus**, further confirming that *–TO* cannot be considered an unambiguous marker of contrast in sentences in which it attaches to the verb. Whether *–TO* is an contrast marker or marker of other components of information structure can be resolved in context.

Lastly, some of the cases with *–TO* attaching to **adverbs** support McCoy's analysis of this clitic as a contrastive marker. Per McCoy, clitic *–TO* usually attaches to

²² In some cases 'be-TO' was a contrastive topic:

Speaker 1:

Был?
Were?
'Were you?'

Speaker 2:

Быть-то я / конечно / был.
Be-TO I / of course / was
'As to being, of course, I was.'
[_{CT} be]-TO [_{CF} was]

the topic and signals that both the topic and the focus are contrastive. As mentioned earlier, not only can **aboutness topics** (as in the McCoy's examples above) be marked as contrastive by *–TO*, but **frame-setting topics** preceding clitic *–TO* can be contrastive as well. Below is an excerpt of a recording of a dialogue between a male and a female speaker in which *–TO* attaches to frame-setting topic and marks for contrast both topic and focus:

79) Female speaker:

Я завтра поеду / надо помидорки воткнуть /
 1SG.NOM tomorrow go-1SG.FUT / necessary tomatoes plant
 'I will go tomorrow. I need to plant tomatoes /

остатки... только бы мороза не было /
 remainders... only COND frost no be-SG.NEUT.PST
 'the remainders... I only hope there is no frost'

да ну / вроде бы на плюсовую выйдет...
 well no / seemingly COND at plus-SG.FEM.ACC come_out
 'well no / it seems that it will be plus (temperature)...

Male speaker:

Все равно погода такая вчера была...
 At_any_rate weather such-SG.FEM.NOM yesterday be-SG.FEM.PST...
 'At any rate, it was such a weather yesterday...

Female speaker:

Да... вчера-то холодно было. Я вчера смотрела этот... Евровидение...
 Yes... yesterday-TO cold was. I yesterday watched this... Eurovision...
 'Yes... Yesterday-TO it was cold. I watched yesterday this... Eurovision...'
 (Russian National Corpus)

Below is the sentence containing *–TO*:

Да... вчера-то холодно было.
 Yes... [CT yesterday]-TO cold was
 'Yes... Yesterday-TO it was cold.'

Applying sub-implicit question test (Gryllia 2008) to this sentence within the context provided above yields the following information structure pattern:

- 80) What is the weather?
 1. [CT Tomorrow] [CF it will be plus (temperature)].
 2. [CT Yesterday] [CF it was cold].

In the sentence (2) above adverb *вчера* ‘yesterday’ preceding *–TO* is a **contrastive frame-setting topic** (contrasted to *завтра* ‘tomorrow’ in previous sentence), while the rest of the sentence ‘*it was cold*’ is a **contrastive focus** (contrasted to ‘*it will be plus*’ in previous sentence). Although most of the sentences with adverbs seem to support McCoy’s analysis of *–TO* as marking both topic and focus for contrast, more analysis of data with *–TO* attaching to adverbs is necessary, as the dataset extracted from Russian National Corpus contained too few sentences with *–TO* attaching to adverbs to make any definite conclusions.

Based on the very limited data from Russian National Corpus discussed in this section, it appears that the part of speech to which *–TO* attaches does not influence the information structure pattern. The declarative sentence (75) in which *–TO* attaches to noun and the interrogative sentence (77) in which *–TO* attaches to verb were used to support this claim. More examples with *–TO* attaching to different parts of speech will be examined in the sections below.

4.5.1.4 *–TO* in interrogatives

In this section I will look at occurrences of *–TO* in interrogatives in order to determine the role that *–TO* plays in marking information structure in interrogative sentences. Since *–TO* appears in *yes-no* questions and *wh*-questions, I will examine information structure patterns in both of these types of questions.

In questions clitic *–TO* may appear in the position that is unusual for topics, i.e. in the middle or in the end of the sentences. This is different from its usual position in non-questions, in which clitic *–TO* often attaches to the sentence-initial topic element. In the examples of questions below *–TO* does not necessarily mark information structure pattern as predicted by McCoy (2001). In both *yes-no* questions and *wh*-questions *–TO* does not necessarily mark contrast in the sentence. When it does, it either marks both topic and focus for contrast, or just focus.

4.5.1.4.1 -*TO* in *yes-no* questions

In this section I will examine two types of information structure patterns in the *yes-no* questions found in the spoken data. In the first sentence type -*TO* marks for contrast both focus and topic (2001). In the second sentence type -*TO* marks only focus for contrast.

- ...[CT]-*TO* [CF] ...?
- ...[CF] [T] -*TO* ...?

The first type of *yes-no* questions above supports McCoy's (2001) analysis of -*TO*. Below is an example of this type of information structure pattern, found in an excerpt of a dialogue recorded in a health resort. Based on the limited context provided in the Russian National Corpus, speaker 1 is a non-smoker that does not smoke because he recently had a heart attack. Speaker 2 is a smoker asking whether speaker 1 has an ashtray at his place:

81) Speaker 1:

Да.
yes
'Yes'

Speaker 2:

У тебя есть какая-нить пепельница?
at 2SG.GEN is some ashtray.NOM
'Is there an ashtray here (at your place)?'

Speaker 1:

Есть.
is
'(There) is.'

Speaker 2:

А ты-*то* не куришь?
and 2SG.NOM-*TO* not smoke-2SG.PRS
'And (as to) you, you don't smoke?'

Speaker 1:

Да нет. Ну я после инфаркта...
well not. well 1SG.NOM after heart_attack
'No. Well, I am after a heart attack.'

Speaker 2:

A-a!

a-a

‘Ah, (okey)!’

Below is the *yes-no* question from example above with clitic *-TO* by Speaker 2:

A ты-то не куришь?

And [CT 2SG.NOM]-*TO* [CF not smoke-2SG.PRS]?

‘And (as to) you, you don’t smoke?’

In the example above there are two members of the set marked by *-TO*: speaker 1 (non-smoker) denoted by *ты* ‘you’ and speaker 2 (smoker) who is speaking at that moment. Speaker 2 is choosing one of the members of the set *ты* ‘you’ contrasting it to himself. Pronoun *ты* ‘you’ is a **contrastive topic** in this example, per Krifka’s (2007) definition of topic and Molnár’s (2002) characteristics of contrast. It is difficult to apply ‘tell-me-about’ test (Choi 1999) and implicit sub-question test (Gryllia 2008) to an interrogative addressing 2nd person.

A different example of a *yes-no* question containing clitic *-TO*, with **contrastive focus** and **non-contrastive topic** is found in this short excerpt from a transcript of a movie *Brief Encounters*:

82) Speaker 1:

Ой / он же уже закипел.

Oh / 3SG.MASC.NOM *ŽE* already boiled.

‘Oh, it has already boiled.’

Speaker 2:

Зиночка / давайте / я вам погадаю.

Zinochka/ let_me / 1SG.NOM 2SG.DAT read_fortune-SG.FUT

‘Zinochka, let me read your fortune.’

Speaker 3:

Ой...

Oh...

‘Oh...’

Speaker 2:

Вы правильно чай-то заварили?

You correctly tea-TO brew

‘Did you correctly brew the tea?’

Speaker 3:

Правильно.
Correctly.
'Correctly'

(Russian National Corpus, *Brief Encounters*)

Below the sentence with –*TO* is repeated:

Вы правильно чай-то заварили?
you [CF correctly] [T чай]-*TO* brew?
'Did you correctly brew the tea?'
...[CF] [T]-*TO* ...?

In the sentence above *правильно*/'correctly' is the **contrastive focus**, while *чай*/'tea' is **non-contrastive topic**.

To summarize, two types of *yes-no* questions have been covered in this section with the following information structure patterns, with the first one supporting McCoy's analysis (2001) and the second two types showing information structure patterns slightly different from the ones predicted by McCoy:

- ...[CT]-*TO* [CF] ...?
- ...[CF] [T]-*TO* ...?

In *yes-no* questions (2) and (3) above clitic –*TO* marks a **contrastive focus** in sentences but attaches to a **non-contrastive topic**. However, in all three cases there is another contrastive element in the sentence, namely contrastive focus. So, even though –*TO* does not necessarily mark the contrast on the element to which it attaches, it nevertheless appears to signal the presence of a contrastive element in the sentence.

4.5.1.4.2 –*TO* in *wh*-questions

Clitic –*TO* often appears in the *wh*-questions. Depending on the context, clitic –*TO* in *wh*-questions marks different information structure patterns. In the section below I will briefly examine *wh*-questions with three different information structure patterns:

- (1). [_{SF} *wh*-word] [_{CF} verb]-*TO*?
- (2). [_{SF} *wh*-word] [_T verb]-*TO*?
- (3). [_{CF} *wh*-word] [_{T/(BG?)} verb]-*TO* ?

While in the *wh*-questions of type (1) and (3) above –*TO* signals presence of contrast, there are also *wh*-questions of type (2) containing clitic –*TO* in which there is no contrast.

Wh-question of type (1) above is examined below, demonstrating the following information structure pattern: [_{SF} *wh*-word] [_{CF} verb]-*TO*?. This information structure pattern is different from the one examined by McCoy in her dissertation (2001). This is an excerpt of a recording of a phone conversation between 70 and 71-year old female speakers:

83) Speaker 1 (female, 70 years old):

То дождь / то снег... А от воды грязь... Вот такие дела.
 TO²³ rain / TO snow.... And from water mud... Here such things...
 ‘Now rain / now snow... And water causes mud... That’s how things are.’

Speaker 2 (female, 71 years old):

Что делаешь-то?
 What doing-*TO*
 ‘What are you doing?’

Speaker 1 (female, 70 years old):

Да вот смотрела телевизор / думаю дай-ка я тебе позвоню / давно не
 ‘Well here watching television / think let-KA I you call / long no
 слышала твоего голоса. Да мне скучно было. Думаю / возьму-ка трубку
 heard your voice. But me boring was. Think / take-KA receiver
 / да позвоню.
 / and call

‘Well, I was watching television / and here I think let me call you / I haven’t heard your voice for a long time. And I was bored. So I think / let me take a receiver and call.’

(Russian National Corpus)

In the example above –*TO* marks a contrast, but in a different way than in the previous examples, i.e. in this sentence **semantic focus** overlaps with **contrastive focus**. On one hand, speaker 2 is just asking a **semantic focus** question, i.e. she is genuinely interested

²³ It should be noted that the *TO* forming disjunction in the first sentence is not analyzed.

in finding out WHAT the other speaker is doing. Speaker 1 responds to Speaker 2 with details about her activities before she called Speaker 2. On the other hand, there is an element of **contrast**, i.e. when Speaker 2 asks '*What are you doing-TO?*' she is trying to switch the conversation from discussing weather to Speaker 1 telling her about her own day. In this sense, Speaker 2 is contrasting two different topics for conversation, i.e. weather vs. daily life of Speaker 1, encouraging her to speak about the latter. To summarize, in the sentence above there is a **contrastive focus** and a **semantic focus**:

84) Что делаешь-то?
 [CF [SF what] doing]-TO?²⁴
 'What are you doing?'

As with some previous examples, it should be noted again that if the same question is asked without *-TO* (but preserving the intonation of the question with *-TO*), it will be interpreted in the same way, i.e. as speaker's attempt to change the subject of conversation. However, if the same question is asked without *-TO* and without preserving the same intonation, then this question will not have the same interpretation, i.e. it will be an open-ended question without attempt to change the subject implied. This suggests that presence of *-TO* in the sentence requires a certain intonation that might receive a contrastive focus interpretation, possibly **IK2**, which is higher in tone and more intense than **IK1**, which is used for new-information focus (Neeleman and Titov 2009: 515). More research on the connection between clitic *-TO* and the prosody of element to which it attaches is necessary to make more definite conclusions.

The example below shows that it is possible to have yet another pattern of information structure in *wh*-questions containing *-TO*. This pattern is shown below in the excerpt of a recording of a telephone conversation between two female speakers:

²⁴ It should be noted that in this sentence an entire Speaker 2's proposition '*what are you doing-TO?*' has been treated as contrasted to the preceding speaker 1's utterance about '*weather*.' However, treating only the verb '*doing-TO*' as contrastive is possible as well, if Speaker 1's utterance addressing Speaker 2 is interpreted in the following way:

You are '*talking about the weather*', but what are you '*doing*'?

More examples of both types of information structure patterns (i.e. wide contrastive focus and narrow contrastive focus in *wh*-questions containing *-TO*) should be looked in more detail in the future.

85) Speaker 1

Да ничего не идет... Хоккей.
But nothing not is_on... Hockey
'There is nothing on (TV)... Hockey.'

Speaker 2

Кто играет-то?
Who playing-TO
'Who is playing?'

Speaker 1

Россия с Германией / по-моему.
Russia with Germany / in_my_opinion
'Russia with Germany, I think.'

(Russian National Corpus)

The sentence that contains clitic *-TO* is repeated below:

Кто играет-то?
[_{SF} Who] [_T playing]-*TO*?
'Who is playing?'

Using the same diagnostics for identifying information structure components that were applied earlier, this sentence is analyzed as containing a **non-contrastive topic** *играет*/'playing' and a **semantic focus** *кто*/'who'. Both topic and focus are non-contrastive in this sentence, i.e. clitic *-TO* in this sentence does NOT mark contrast. As in the previous examples, the non-contrastive interpretation of both elements in the sentence can remain the same if the question is asked without *-TO*.

And, finally, in a *wh*-question very similar to the one above, but uttered within a different context, clitic *-TO* actually marks contrast in focus. There is a **non-contrastive topic** (*делаешь*/'doing') and a **contrastive focus** (*что*/'what'):

86) Да что ж ты **делаешь-то** / а? Я тебе как показала?
But what ž 2SG doing-TO / huh? 1SG 2SG.DAT how showed-FEM
'What (on earth) are you doing, huh? How did I show you (what to do?)'

(Russian National Corpus)

In this utterance the speaker is not asking for any new information, which explains why *'что*/'what' is not analyzed as semantic focus in this sentence. Instead, speaker is

unhappy with the actions of the hearer. This sentence has the following information marking pattern:

Да что ж ты делаешь-то / а?
But [CF *wh*-word] ž you [T/(BG?) doing]-*TO*, huh?

or, in brief:

[CF *wh*-word] ž [T/(BG?) verb]-*TO* ?

By asking ‘*What (on earth) are you doing?*’ speaker is implying that the hearer should have been doing something else instead of what (s)he is doing at the moment. The *wh*-word *что* / ‘what’ is analyzed as **contrastive focus** because it is a member of a set in which ‘*what(ever) the hearer is doing*’ is contrasted to ‘*what(ever) the hearer should have been doing instead.*’ It should be noted though that another contrastive clitic, i.e. *ž*, is present in the same sentence. It might bear partial responsibility for marking contrastive on the focused *wh*-word, which might have not been contrastive focus otherwise. More research about interaction of the contrastive clitics is necessary to come to any definite conclusions, but for purposes of this thesis *что* / ‘what’ is analyzed as **contrastive focus**. Meanwhile, verb *делаешь* / ‘doing’ is analyzed as **non-contrastive topic** because it fits Krifka’s definition of topic (2007) and does not possess contrastive characteristics described by Molnár (2002).

To conclude, three types of *wh*-questions with –*TO* attaching to the verb in the sentence-final position have been examined in this section:

- [SF *wh*-word] [CF verb]-*TO*?
- [SF *wh*-word] [T verb]-*TO*?
- [CF *wh*-word] ž [T/(BG?) verb]-*TO* ?

As has been shown above, the differences in the information structure patterns above depend mainly on the context of the sentence (semantic or syntactic). This presents a challenge in terms of predicting the information structure pattern of the sentence with –*TO* without looking at the surrounding context of the sentence.

4.5.1.5 -TO in sentences without topic

In the sentences where topic is absent, clitic *-TO* cannot mark topic for contrast, but it can mark the focused element for contrast. For example, in the sentence below, in which the contrastive marker *-TO* appears twice in two topic-less clauses, it attaches to and marks for contrast the focused elements in both clauses. This sentence is a recorded excerpt from a dialogue between a 45-year old male and a sociologist on social/political topics:

87) Speaker 1:

Дайте людям денег / а инструмент-то дайте / чтобы сделать-то.
Give people money/ but tool-TO give / so_that make-TO
'Give money to the people. But give (them) the tool in order to make (something).'

CLAUSE 1 [CF noun]-TO [SF verb] / CLAUSE 2 [SF verb]-TO.

Speaker 1:

А у вас пока не будет этого инструмента / вы и руками там ничего /
And at you while no be this tool / you and hands there nothing
'And until you have this tool, you (will be able to make) nothing with (your) hands'

только из глины можете что-то.
only from clay can something
'(you) might only be able to do something with the clay'

(Russian National Corpus)

In the sentence above *-TO* appears twice in two different clauses: the *matrix clause* and the *embedded clause*. In the first clause clitic *-TO* attaches to **contrastive focus** *инструмент*/'tools' followed by **semantic focus** *дайте*/'give':

а инструмент-то дайте
but [CF tool]-TO [SF give]

I am treating *инструмент*/'tools' as **contrastive focus** and not as contrastive topic based on the limited context (there is no indication that it should be treated as a topic according to Krifka's (2007) definition of topic) and the fact that it fails implicit sub-question test (Gryllia 2008), i.e. a sub-question cannot be formed to accommodate clause above as one of its answers. The reason I am treating *инструмент*/'tools' as contrastive is because it is contrasted to *денег*/'money' in the previous utterance. Verb *дайте*/'give' is treated as **semantic focus**, as it represents new information in the sentence, but there is no contrast implied.

In the second clause, in the absence of topic, clitic *–TO* attaches to **non-contrastive semantic focus** *сделать*/'make':

чтобы сделать-то.
so_that [_{SF} verb]-*TO*
'to make-TO'

As in the preceding example, in this sentence verb '*сделать/make*' is treated as non-contrastive semantic focus is because it represents new information in the sentence, but there is no contrast implied.

To conclude, it appears that in the sentences in which topic is absent *–TO* can mark for contrast only focus. More analysis of spoken data would be beneficial to confirm this hypothesis.

4.5.1.6 Interaction with other lexical elements

The contrastive clitic *–TO* can be used with other clitics in the same sentence, including the contrastive clitics *ŽE* and *VED'* and contrastive conjunction *a/'but'*. An example with three contrastive clitics and a contrastive conjunction co-occurring in the same sentence is below:

88) А то ведь ей не давали же прижиться.
But *TO* *VED'* her not give *ŽE* establish.
'But they were not letting her to establish herself.'

(Russian National Corpus)

Although the occurrence of four contrastive elements in one sentence as above is quite rare, sentences containing two contrastive clitics are more common. The contribution of each clitic to the sentence containing more than one contrastive clitic is outside the scope of this thesis, but deserves further consideration. Whether the presence of other contrastive clitics strengthens contrastive properties of the element in question remains an interesting problem to solve. Looking at the contrastive scope of each clitic, their interaction, as well as their positions in the sentence within the sentences containing more than one contrastive clitic could aid in understanding the input and the scope of each individual clitic.

4.5.1.7 -TO: SUMMARY & CONCLUSION

As has been noted above, (-)TO has many functions, including being a demonstrative pronoun, forming disjunctions, forming indefinite pronouns, and being a part of conditional expressions. The clitic -TO does not always mark contrast in the sentences, as has been shown with examples in which it appears in interrogatives.

McCoy (2001) argues that in the cases when it marks contrast, the clitic -TO marks **a set of sets** of propositions, usually attaching to the topic with the following information structure pattern:

89) [CT]-TO ... [CF]...

Several issues have been addressed in the preceding sections describing the difficulties with applying contrastiveness tests and theoretical concepts of information structure to McCoy's analysis, as well as with applying McCoy's analysis to Russian spoken data.

Using McCoy's example of -TO as a contrastive marker it has been shown that one of the challenges of applying McCoy's analysis involves identifying the element to which -TO attaches as a contrastive topic by applying tests information structure components reviewed in section 4.4 earlier in this thesis:

90) [CT(?)]-TO ... [CF] ...

While the implicit sub-question test used by Krifka (2007) and by Gryllia (2008) identify the element to which -TO attaches as **contrastive topic**, *wh*-question test used by Gryllia (2008) identifies it as **contrastive focus**. Additionally, the 'tell-me-about' test used by Choi (1999) suggests that it is **not a topic**. Due to the inconclusiveness of these tests, Krifka's definition of topic (2007) and Molnár's (2002) description of characteristics of contrast were also adopted in order to identify (contrastive) topic in the sentences, in addition to the tests. Applied together, the contrastiveness tests, the definition of topic and the characteristics of contrast provided support for McCoy's analysis of -TO in sentences similar to the example (62), in which the following information structure components were identified:

91) [CT]-TO ... [CF] ...

However, it has been noted that more reliable diagnostics are necessary to identify information structure components in spoken speech with clitics than the tests discussed in section 4.2.

It has also been shown in the preceding sections that *-TO* can mark both **aboutness topics** and **frame-setting topics** for contrast. However, more research is necessary in order to demonstrate that this is the most accurate analysis of the information structure pattern in sentences of this type.

Another challenge was applying McCoy's analysis to spoken data, particularly to sentences with the verb '*быть/to be*' and the interrogatives. It was found that in these sentences different information structure patterns were possible, some of which did not support McCoy's analysis of this clitic.

While some sentences with the verb '*быть/to be*' have information structure patterns of the type: [CT] *-TO* [CF], confirming McCoy's analysis, there were two other types of examples that did not confirm McCoy's analysis. One type of sentences with *-TO* attaching to verb '*be*' contra McCoy's analysis was represented by the declaratives, with *-TO* marking only contrastive focus:

92) [T *be*] *-TO* [CF]

In addition to sentences with *-TO* attaching to verbs interrogatives represent another problem for McCoy's (2001) analysis of information structure pattern of sentences with *-TO* as [CT] *-TO* [CF].

The following information structure patterns have been found in the *yes-no* questions, with the former type supporting McCoy's analysis (2001) and the latter type not predicted by McCoy's analysis (2001):

- ...[CT] *-TO* [CF] ...?
- ...[CF] [T] *-TO* ...?

In *wh*-questions there have been found the following information structure patterns, contra McCoy (2001):

- [SF *wh*-word] [CF verb] *-TO*?
- [SF *wh*-word] [T verb] *-TO*?
- [CF *wh*-word] [T/(BG?) verb] *-TO* ?

It was concluded that in both *yes-no* and *wh*-questions examined in this section *–TO* does not necessarily information structure pattern as predicted by McCoy (2001). *–TO* does not always mark contrast in interrogatives. When it marks the contrast, it does not necessarily mark both topic and focus for contrast.

Based on the examples of *yes-no* and *wh*-questions contradicting McCoy's analysis, it was concluded that the factors contributing to different information structure patterns in sentences with *–TO* depend on the illocutionary force of the sentence and the context (semantic or syntactic, such as presence of another information structure marking clitic).

In section 4.5.1.5 sentences without topic have been examined, showing that the only elements that *–TO* can mark for contrastiveness in such sentences are foci.

Lastly, the interaction of *–TO* with other lexical elements has been briefly examined, without a detailed investigation and any definite conclusions. It was suggested that interaction of *–TO* with other clitics represents a promising area for further research that would allow for a better understanding of the contribution of each clitic to the utterances.

4.5.2 Clitic *ŽE*

4.5.2.1 Introduction

In the following sections I will examine the spoken data containing *ŽE*. I will briefly mention the most frequent position of *ŽE* in the sentence, as well as the words to which *ŽE* most frequently attaches. I will also discuss information structure marking properties of *ŽE* in spoken data²⁵ in more detail.

²⁵ As has been mentioned earlier in chapter 3, clitic *ŽE* has often been considered to be aggressive or implying a verbal attack on the hearer, which explains why this clitic is so most frequently used in the spoken data in the Russian National Corpus, which is the most informal type of data in the corpus. The type of data that was next to spoken speech in terms of this clitic usage frequency was non-spoken non-fiction data, which contains many lectures and interviews, characterized by strong presence of authoritative utterances, which also explain the high usage frequency of *ŽE*.

I will start with applying several tests to identify information structure patterns in the example sentences from King (1995) and from the Russian National Corpus in section 4.5.2.2. In section 4.5.2.3 I will show that it is impossible to account for the contrastive focus projection of *ŽE* just based on its position in the sentence, as well as to account for the position of contrastively focused element relative to phrasal *ŽE* without knowledge of context or intonation. In section 4.5.2.6 I will look at different information structure patterns marked by *ŽE* when it attaches to deictics and *wh*-words. I will examine in detail cases when *ŽE* attaches to *wh*-words in interrogatives, declaratives, and within the established expressions.

4.5.2.2 Clitic *ŽE* and information structure marking

In this section I will look at two examples, one from King (1995) and another one from the Russian National Corpus and try to determine whether tests to identify information structure patterns can be easily applied to sentences containing *ŽE*.

Below is an example used by King in which she claims clitic *ŽE* marks preceding word as **contrastive focus**:

- 93) On uedet [segodnia že].
 He will leave today že
 ‘He will leave today-FOC.’

(King 1995: 80)

King argues that clitic *ŽE* in the example above makes preceding element, *segodnia*/‘today’ in this sentence, more emphatic. Following King (1995), in this thesis sentence above is analyzed as having the following information structure pattern:

- 94) ... [CF today] *ŽE*.

There are several reasons why above constituent *segodnia*/‘today’ is not analyzed as a contrastive topic in this thesis. Firstly, it fails implicit sub-question test to identify contrastive topic used by Gyllia (2008). Secondly, even though it could be failing this

test due to other reasons, such as presence of an emphatic/aggressive clitic *ŽE*, it also does not fit Krifka’s definition of topic (2007). As a result, *segodnia*/‘today’ is not analyzed as topic in this sentence. Instead, constituent *segodnia*/‘today’ is analyzed as **contrastive focus** because it fits the definition of contrast discussed earlier, i.e. it represents a set of alternatives with semantically and syntactically parallel members (‘today’, ‘tomorrow’, etc.) one of which, i.e. ‘today’ is contrasted to other members.

Even though the constituent *segodnia*/‘today’ fits definition of contrastive focus used in this thesis, contrastiveness tests used by Gryllia (2008) cannot be used to confirm that *segodnia*/‘today’ is indeed a contrastive focus. Constituent *segodnia*/‘today’ passes the *wh*-question, correction, and choice contrastiveness tests used by Gryllia (2008), but only with very marginally acceptable answers. As in the examples with clitic *–TO*, the answers to the questions required by these tests are only marginally acceptable if they contain clitic *ŽE*, but become acceptable without *ŽE* (for the details about diagnostics used for this sentence please refer to Appendix B, example I).

It should be noted that a sentence very similar to the sentence above (93) can have a very different information structure marking that can be inferred only on the basis of context. In the example below sentence-initial element *zavtra*/‘tomorrow’ preceding *ŽE* is a **non-contrastive topic**. This is an excerpt of recorded conversation between father and daughter, with father asking her about her weekend plans:

95) Father:

А завтра во сколько дома будешь?
And tomorrow at when home be
‘And when are you going to be at home tomorrow?’

Daughter:

Ой / папуль / не знаю...
Oj / Dad / not know...
‘Oh, Dad. I don’t know...’

Завтра **же** у Эльмирика день рождения.
Tomorrow *ŽE* at Elmirik day birthday
‘Tomorrow Elmirik has birthday.’

(Russian National Corpus)

[TOPIC ...] *ŽE* [CF ...].

The reason that constituent ‘*завтра/tomorrow*’ above is treated as **non-contrastive frame-setting topic** because there is no contrast implied, it fits the definition of topic and it fails the implicit sub-question test used in Gryllia (2008).

Based on just two examples above, the spoken data containing *ŽE* behaves similar to data containing *–TO* in the conditions of testing for information structure patterns, i.e. not all tests can be easily applied to this data. In the sections below I will look at more examples of sentences containing *ŽE* and their information structure patterns, as well as apply tests to more utterances.

4.5.2.3 Focus projection of *ŽE* and its position in the sentence

As has been mentioned earlier, McCoy (2001) describes *ŽE* as cliticizing (‘encliticizing’ in McCoy’s terms) to the contrastive element or cliticizing to clause boundaries. Instead of looking at *ŽE* in terms of its relation to the contrastive element or clause boundaries, I will examine phrasal *ŽE* (marking word/constituent for contrast) and proposition-level *ŽE* (marking an entire proposition for contrast) in terms of its position in the sentence.

Seeing whether the element marked contrastively by *ŽE* is narrow contrastive focus or wide contrastive focus depending on the position of *ŽE* in the sentence could inform how it should be implemented in the grammar in chapter 5. I am going to test whether *ŽE* can be interpreted as **proposition-level** when it appears in the second position in the sentence and as **phrasal** when it appears in other positions in the sentence. Using examples from the spoken data, I will demonstrate that both phrasal and proposition level *ŽE* share the same positions in the sentence.

Based on the data found in the Russian National Corpus (more details in the table 3 below), the clitic *ŽE* is most frequently placed in 2nd position after the first phonological word in the sentence. High frequency of appearance of this clitic in the second position and its steady decrease towards the end of the sentence are evident in the data from Russian National Corpus. Below is the summary of data that I calculated based on the 6,653 sentences containing the clitic *ŽE* in Russian National Corpus:

Table 3: Position of *ŽE* in the sentence based on Russian National Corpus data

Position in the sentence	% of total data ²⁶
2 nd	15
3 rd	11
4 th	5
5 th	4
7 th	3
6 th	3
Other	59

It should be noted that the data gathered from Russian National Corpus is noisy and the high cumulative number of ‘other positions’ occupied by clitic (not presented in the table above) is due to the noisiness of the data. Some of the positions in the sentence occupied by clitic *ŽE* are positions higher than 1,000, due to the lack of punctuation in large chunks of the spoken data. Such numbers are clearly not indicative of the real position of this clitic in the sentence. Additionally, the number for the “other positions” currently includes the numbers from the embedded sentences, which are not delimited by commas in the spoken corpus and therefore are challenging to process. Although outside the scope of this thesis, it would be beneficial to study the less noisy spoken data in more detail in the future to get more accurate numbers and patterns for the positioning of *ŽE* in the sentence.

Based on the data from Russian National Corpus, it is impossible to identify whether *ŽE* marks a phrase or an entire proposition as contrastively focused merely based on its position in the sentence. As will be shown below, both **proposition-level** *ŽE* and **phrasal** *ŽE* can appear in the 2nd position in the sentence, which makes it difficult to differentiate between them and to identify the elements that they mark contrastively. In the paragraphs below I will examine the ambiguity of contrastive focus projection of clitic *ŽE* when it is placed in the 2nd position in the sentence. I will look at two sentences demonstrating focus projection ambiguity of the clitic *ŽE* following a verb, as well as at

²⁶ This column represents the percentage of sentences containing *ŽE* in the position indicated in the left column, e.g. 15% of 6,653 sentences that were processed for this task contained *ŽE* in the second position. The row “Other” represents other (mostly spurious) data consisting of large chunks of text without sentence breaks.

one sentence in which *ŽE* follows a pronoun. In this sentence **phrasal *ŽE*** occupies second position in the sentence that can also be occupied by **proposition-level *ŽE***.

Example (96) below contains *ŽE*, which is placed in the 2nd position after the first phonological word *есць*/'is'. This sentence is a part of the discussion about collective memory/knowledge about the World War II:

96) Speaker 1:

Всё уже давно забылось.
All-NEUT.NOM already long_time_ago forgotten-NEUT
'Everything has been forgotten long time ago.'

Speaker 2:

Нет / кто знает.
No / who knows
'No... who knows...'

А есть же страны и не знают.
But are *že* countries-PL and not know-PL.PRS
'But there are countries that do not (even) know.'
[_{CF} But there are (*ŽE*) countries that do not (even) know.]

Вот просто показывают по телевидению опросы / они и не знают.
Here just show-PL on television-DAT polls / they and not know
'So they simply show some polls on TV, and they don't know.'

А о молодёжи-то вообще нечего говорить /
And about youth-TO overall nothing say
'And there is nothing to say about youth.'
(meaning "And it's even worse when it concerns youth")

они даже и не знают / что война такая была.
they even and not know that war.SG.FEM such.SG.FEM was.SG.FEM
'They even don't know that there was such a war.'

(Russian National Corpus, Fund *Obschestvennoe mnenie*)

Speaker 1 states that the war has been forgotten. Speaker 1 objects that there exist entire countries that do not know about the war, i.e. people in these countries have never even known about World War II to begin with. Therefore, speaker 1's presupposition, i.e. that

*‘people have known about World War II, but they have forgotten about it’²⁷ is contrasted against speaker 2’s presupposition *‘there are countries in which people have never even known about World War II’.**

In the example above **proposition-level** *ŽE* occupies 2nd position in the sentence after the first phonological word.

In the previous section I have used King’s (1995) example (93) of a sentence in which **phrasal** *ŽE* occupies non-second position in the sentence. However, **phrasal** *ŽE* can occupy the second position in the sentence, which can also be occupied by **proposition-level** *ŽE* as has been shown in previous example (96). Below is an example of a sentence in which **phrasal** *ŽE* is placed after the first phonological word in the sentence. This is an excerpt of a dialogue in a documentary film:

97) Speaker 1 (male):

И городу приятно и мне приятно.
And city-DAT pleasant and 1SG.DAT pleasant
‘It is pleasant both to me and to the city.’

[Посмеивается]
[Laughing]

Speaker 2 (female):

Львов достаточно часто делаем.
L’vov rather often do-1PL.PRS
‘We often do (film?) (city of) L’vov.’

Speaker 3 (male):

Я думаю / что вот крестьяне вывозили вот свои...
1SG.NOM think.SG / that well peasants take_out-PL.PST well their.PL.ACC
‘I think that... well... peasants were taking out their ...’

Торговать **же** они в Москву ездили и куда-то в... эту... Рязань / Москва.
Trade-INF **ŽE** 3PL.NOM to Moscow went and somewhere to... this... Ryazan’ / Moscow
‘As to trading, they went to Moscow and somewhere to this... Ryazan’... Moscow.’

²⁷ It is also possible that this construction has the following (odd) information structure pattern, in which *ŽE* marks contrastive focus in the next clause:

А есть **же** страны и не знают.
But are **že** countries and [CF not know]

И... б-бывали закупки у то... у того же Мамонтова в Абрамцево.
 And...w-were purchases at th... at that *ŽE* Mamontov-SG.GEN in Abramtsevo
 ‘And there were purchases from the same Mamontov in Abramtsevo.’

(Russian National Corpus, documentary film *Letters from the Province*)

The sentence with *ŽE* with corresponding information structure pattern is repeated below:

Торговать же они в Москву ездили...
 [CT Trade-INF] *ŽE* 3PL.NOM to Moscow went.PL.PST...
 ‘As to trading, they went to Moscow....’

The verb *торговать* /‘trade’ is treated as **contrastive topic** because it passes both ‘tell-me-about’ test used by Choi to identify topic (1999) and it passes the implicit sub-question test used by Gryllia (2008) (for more details see Appendix B, example II).

Similar scopal ambiguity of *ŽE* when it occupies second position in the sentence are found in the utterances in which *ŽE* attaches to non-verbs. Example (98) below contains **phrasal** *ŽE* in the second position in the sentence. This is an excerpt from a recording in a school, of a teacher scolding students for throwing tangerines and toilet paper in the toilet bowls. This example demonstrates placement of **phrasal** *ŽE* in 2nd position in the sentence:

98) Speaker:

Дело в том /
 Thing in that /
 ‘The thing is

что кто-то запихнул мандаринов и туалетной бумаги в унитазы /
 that someone shove-PST tangerine-PL and toilet paper in toilet_bowls /
 that someone shoved tangerines and toilet paper in the toilet bowls

следовательно / туалет засорился.
 therefore / toilet clogged_up
 Therefore, the toilet clogged up.

...

Весь четвертый и третий этаж в воде!
 Entire fourth and third floor in water
 Entire fourth and third floors are in water!

Вы же убираться будете / а не учителя.
 You *ŽE* clean will-2PL.FUT / but not teachers
 [CF You] *ŽE* [SF will be cleaning] /
 You will be the ones cleaning it, and not the teachers.’

The pronoun *вы*/'you.PL' preceding the clitic *ŽE* in the sentence above is treated as a **contrastive focus** in this thesis because it passes (1) *wh*-question test and (2) correction test (with marginally accepted answer that can be explained by "aggressive" nature of the clitic *ŽE*) for contrastiveness of focus, and fails (3) Choi's 'tell-me-about' and (4) Gryllia's implicit sub-question tests for topic and contrastive topic respectively (Choi 1999, Gryllia 2001) (for more details please refer to Appendix B, example III). In the example above the pronoun *вы*/'you.PL' is contrasted to the noun *учителя*/'teachers'. **Phrasal** *ŽE* attaches to contrastively focused pronoun, also first phonological word *вы*/'you.PL' and occupies second position in the sentence.

Below is an example in which *ŽE* also attaches to a pronoun and occupies second position in the sentence, but this time it is a **proposition-level** *ŽE*. This is an excerpt of a recording of a conversation between a male and female speaker:

99) Female speaker:

Ну и ходи так.
Well and walk this_way
'Well you can walk around in such way' or
'Walk around as you wish.'

Male speaker:

Ну правда / че / плохой цвет лица?
Well really / what / bad color face-GEN
'Really, what, (I have) bad complexion?'

Female speaker :

Хороший / хороший. Но ты же молодой пока. А состаришься?
Good / good But 2SG *ŽE* young so far But age-2SG.FUT
'(It's) good, are good. But you are still young *ŽE*. What about when you age?'

Male speaker:

Не состарюсь. Я Горец.
Not age-1SG.FUT. I mountaineer.
'I will not age. I am from the mountains.'

(Russian National Corpus)

The sentence containing *ŽE* is repeated below for convenience:

Но ты **же** молодой пока.
 But 2SG **ŽE** young so far
 [_{CF} But 2SG **ŽE** young so far]
 ‘But you are still young (**ŽE**).’

In the example above **proposition-level** **ŽE** occupies the second position after the first phonological word and attaches to the pronoun *ты* ‘you’. The female speaker seems to be unhappy about the way that male speaker walks around (perhaps not wearing sunscreen, smoking, or some kind of behavior that is supposed to affect his complexion, as she believes). The male speaker responds to her that regardless of his behavior his complexion looks great. The female speaker objects that the reason that male speaker’s complexion looks great is because he is still young, but it is not going to look so great when he is old. The female speaker’s proposition ‘*but you are young so far*,’ implying that the young age of the hearer might be the main reason why he still has good complexion is contrasted to the male speaker’s belief that he ‘just’ has a good complexion despite his lifestyle or some actions.

To conclude, examples in this section demonstrate that the contrastive focus projection of **ŽE** when it is placed in the 2nd position in the sentence is ambiguous between **phrasal** **ŽE** and **proposition-level** **ŽE**. More research would be beneficial in terms of identifying other factors besides clitic’s position in the sentence that may determine whether **ŽE** marks a constituent or an entire proposition as contrastively focused.

4.5.2.4 Position of contrastively focused element relative to phrasal **ŽE**

Similar to difficulties in determining the contrastive focus projection of **ŽE**, it is also challenging to determine position of the contrastively focused element relative to phrasal **ŽE**. The contrastively focused element can be placed before, as in the majority of examples with phrasal **ŽE** in the previous section, or after **ŽE**, as will be shown in the example below. This is an excerpt of a dialogue between a female and male speaker about privatization, recorded by fund *Public Opinion*:

100) Speaker 1:

Прихвaтизирoвал.
Privatized
'Privatized.'

Speaker 2:

Ну / воспoльзoвался и воспoльзoвался.
Well/ used and used
'Well, he used – so he used' ("Well, he used. So what?")

Этo **же** он воспoльзoвался / a не ктo-тo.
It **ŽE** [_{CF} he] used / but not someone
'It was him that used it, but not someone (else).'

(Russian National Corpus, fund *Obschestvennoe mnenie*)

In the sentence above pronoun *он*/'he' is contrastively focused. It is a member of the set in which it is contrasted to *кто-то*/'someone (else)'. In this sentence contrastively focused element *он*/'he' is placed after clitic *ŽE*. This example is one of the many examples in Russian National Corpus demonstrating that it is impossible to always accurately predict the position of a contrastively focused element based on the position of clitic *ŽE* without the context. The contrastively focused element can be placed before or after phrasal clitic *ŽE*.

To summarize the findings from the last two sections, proposition-level *ŽE* and phrasal *ŽE* both can occupy 2nd position in the sentence, supporting treatment of this clitic as ambiguous in terms of its marking an entire contrastive focus projection. The position of the contrastive element relative to the position of phrasal *ŽE* is ambiguous as well, i.e. the element contrastively marked by phrasal *ŽE* can precede or follow *ŽE*.

4.5.2.5 Clitic *ŽE* attaching to different types of words

Below is the table summarizing the results that I calculated based on the data from the Russian National Corpus. It lists words to which clitic *ŽE* most frequently attaches, in the descending order, based on 6,653 sentences total:

Table 4: Words to which clitic *ŽE* attaches most frequently

Words to which <i>ŽE</i> attaches	Translation	% of total data
я	I	4.76
это	this	3.68
то	that	3.44
как	how	2.88
так	so	2.7
что	what	1.9
там	there	1.66
опять	once again	1.53
конечно	of course	1.39
тут	here	1.38

Based on the data from the Russian National Corpus, the words to which clitic *ŽE* attaches most frequently are: deictics, *wh*-words, and personal pronouns. These are discussed in the following sections.

4.5.2.6 Different information structure patterns marked by *ŽE*

In the following sections I will examine different information structure patterns marked by *ŽE* depending on the word to which it attaches and the illocutionary force of utterance in which it appears. Due to the limited scope of this thesis, I will examine only the following cases of *ŽE*:

- *ŽE* attaching to deictics
- *ŽE* attaching to *wh*-words:
 - in interrogatives
 - in declaratives
 - in expressions

4.5.2.6.1 Clitic *ŽE* attaching to deictics

Clitic *ŽE* is often placed after deictics *this*, *that*, *this*, etc., resulting in accordingly '*этом же/this same one*', '*том же/that same one*', '*так же/in the same way*' (note that this is different from '*также/also*'), etc. Below is an excerpt of a recording of a dialogue

between two customers in the market. They are discussing a certain product, comparing it to some other product in terms of the material from which it is made and the location where it was made:

101) Male Speaker 1:

Да там одно и тоже / что Эрих Крауз / что это.
But there one and same/ that Erich Krause / that this
'But it's all the same there. (Same as) Erich Krause. (Same as) this.'

Female Speaker 2:

Они просто надежней и прочнее.
They just more reliable and sturdier
'They are just more reliable and sturdier.'

Male Speaker 1:

Да ладно. Из одного и того же пластика. В том же Китае делается всё.
But okay. From one and that *že* plastic. In that *že* China is made everything

'No way. (Made) from the same plastic. Everything is made in China all the same.'
But okay. From [CF one and the same] *ŽE* plastic. In [CF in the same *ŽE*] China...

Female Speaker 2:

Ну понятное дело.
Well understood thing.
'Well, that is understandable.'

(Russian National Corpus)

In the example above, Speaker 2 says that a certain product is more reliable and sturdier (than the other product that is presumably worse). However, Speaker 1 does not believe Speaker 2. Speaker 1 objects to her that the product that Speaker 2 believes to be '*more reliable and sturdier and made from a better material*', is actually not any more reliable or sturdier than from the 'worse' product. Speaker 1 explains it by stating that this presumably 'better' product has been made from the same material and in the same country (China) where the 'worse' product has been made. There are two mutually exclusive presuppositions here: Speaker 2's presupposition '*product under consideration is more reliable and sturdier than the 'worse' product because it is **made from a different material and in a different place** than the 'worse' product*' and Speaker 1's presupposition that '*the product under consideration is made **from the same material and in the same place** that the 'worse' product is made, so it is just as (un)reliable and*

as (un)sturdy as the ‘worse product’. These constituents, i.e. ‘(made) from a different material’ vs. ‘(made) from the same material’, as well as ‘(made) in the same place’ vs. ‘(made) in a different place’ are mutually exclusive and are marked by phrasal *ŽE*:

102) ..._[CF deictic] *ŽE*

Example above, as well as similar examples from Russian National Corpus not included in this thesis show that McCoy’s (2001) analysis can be applied to sentences in which clitic *ŽE* attaches to deictics. In these sentences *ŽE* can mark a set of mutually exclusive **contrastive propositions**, as predicted by McCoy (2001).

4.5.2.6.2 Clitic *ŽE* attaching to *wh*-words

The contrastive clitic *ŽE* also attaches to *wh*-words, although not as frequently as *–TO*. McCoy notes in her dissertation that the combination of *wh*-words with clitic *ŽE* represents a challenge for her analysis of this clitic as a K-marker with set of mutually exclusive members. Her suggestion is to “analyze the *wh*-set marked by *ŽE* as (estimated by the speaker) to be empty of predictable, or reasonable, answers. So, by using *ŽE* with a *wh*-word, the speaker indicates that the set of possible answers is empty of reasonable answers and opens it to some unpredicted, unreasonable, unexpected answers. However, whatever is viewed by the speaker as an unpredicted, unreasonable, unexpected answer might not be so for the hearer.” (McCoy 2001: 281).

In the following sections I will briefly examine several sentences from the Russian National Corpus with different illocutionary force with *ŽE* attaching to *wh*-words. First, I will examine interrogatives (*wh*-questions in this case) in which *ŽE* marks **contrastive focus** or **contrastive focus/semantic focus**, possibly implying a set of mutually exclusive members. Next, I will look at the declaratives in which *ŽE* marks contrastive focus, but does not imply a set of mutually exclusive members. Lastly, I will look at the declaratives in which *ŽE* is a part of an expression and does not mark contrast at all.

It should be noted that examining all possibilities of the clitic *ŽE* combining with *wh*-questions is not possible for this thesis. However, given the existence of easily accessible spoken data online nowadays, it would be very beneficial to examine the different types of sentences in which *ŽE* combines with *wh*-words for a more complete analysis.

4.5.2.6.3 *Wh*-words + *ŽE*: Questions

Based on data from the Russian National Corpus, McCoy's analysis of the *wh*-words marked with *ŽE* as marking contrast and denoting members of a set with mutually exclusive members partially applies to the *wh*-questions, as will be shown in subsection 4.5.2.6.3a. However, it should be noted that sentences in which the clitic *ŽE* attaches to the *wh*-words represent a challenge for McCoy's analysis for several reasons.

Firstly, as McCoy notes herself, contrasted foci in these sentences represent a challenge to her analysis of them as a set denoting mutually exclusive members. McCoy enumerates several types of *wh*-questions containing *ŽE*, supporting them with examples. McCoy suggests that one possible way to resolve this issue is to analyze the *wh*-set marked by *ŽE* as empty of reasonable or predictable answers, while opening it up to some unreasonable and unpredicted answers (McCoy 2001: 281). McCoy leaves the matter for further consideration.

Secondly, and perhaps more importantly, it should be noted that the element marked as **contrastive focus** by *ŽE* in these sentences can also be **semantic focus**, as will be shown in subsection 4.5.2.6.3b. Although this does not contradict McCoy's proposed analysis of the *wh*-set marked by *ŽE* as a set with mutually exclusive members, it suggests the presence of an overlap of the set with mutually exclusive members with the set of not mutually exclusive members in the sentences, resulting in its turn an overlap of **contrastive focus** with **semantic focus**.

4.5.2.6.3a Questions (wh-words + *ŽE*): Contrastive focus with mutually exclusive members

Usually the clitic *ŽE* in *wh*-questions marks a set with mutually exclusive members, as predicted by McCoy. Below is an example, which is an excerpt of a recording of a talk on radio station Echo Moskvy (the discussion on the nature of terrorism):

- 103) Ведь там горе / настоящее горе / потому что жертвами этих
Ved' there grief / real grief / because victims these
'But there is grief, genuine grief, because the victims of

действий стали совершенно невинные люди. Какая **же** это месть?
actions became completely innocent people. What *ŽE* this revenge?
these actions were completely innocent people. What kind of revenge is that?

В нашем обществе такая месть не принимается / она осуждается.
In our society such revenge not accepted / it condemned.
In our society such revenge is not accepted. It is condemned.'

Translation of entire excerpt:

'But there is grief / real grief/ because completely innocent people have become the victims of these actions. What *ŽE* (kind of) revenge is that? In our society such revenge is not accepted / it is condemned.'

(Russian National Corpus)

Below the sentence containing **proposition-level** *ŽE* is repeated:

Какая же это месть?
[_{CF} what *ŽE* this revenge]?
'What kind of revenge is that?'

By asking a question '*What ŽE (kind of) revenge is that?*' above, the speaker does not expect the hearer (audience) to give a reasonable answer to his question, i.e. to describe what kind of revenge this action of terrorism is. The speaker is neither making a request for new information, neither he is expecting to hear any new information. Since there is no request for new information involved in this question, it does not contain semantic focus. In fact, this is not even a question, but rather an assertion by the speaker. Asking his question, the speaker presupposes that an act of revenge implies revenge aimed at the targets of the ones seeking revenge. In the case of act of terrorism, the targets of terrorists are usually governments or military forces. Therefore, since innocent civilians are not the

real targets of the terrorists, the terrorist act in which innocent civilians are victims cannot be considered the ‘revenge of the terrorists’. By asking this ‘question’ the speaker merely states his presupposition, or makes a proposition that this act of terrorism is not an act of revenge because innocent civilians have died and therefore this action cannot be defined as a ‘revenge’ act.

It is possible to see the example above as marking a set of mutually exclusive contrastive propositions. An (invisible) audience holds an assertion (at least as presupposed by the speaker) that *‘any action of terrorism is an act of revenge’*. The radio host is objecting/contrasting the audience’s presupposition with his own presupposition, i.e. *‘an action of terrorism cannot be really called ‘revenge’ if innocent people (instead of the real targets of the terrorists) are hurt’*.

The above analysis of the sentence implying contrasting presupposition supports McCoy’s analysis of *ŽE* marking contrast with the set of mutually exclusive members, even when it attaches to *wh*-words. The only difficulty with analyzing propositions marked by *ŽE* as mutually exclusive is that it is possible to apply the same analysis to any dialogue in which speakers disagree, as their statements/presuppositions can be seen as contrastive and mutually exclusive.

4.5.2.6.3b Questions (*wh*-words + *ŽE*): Contrastive focus (with mutually exclusive members) overlapping with semantic focus

McCoy’s analysis of *ŽE* marking contrast with the set of mutually exclusive members could also be applied to *wh*-questions containing **semantic foci**. There are some *wh*-questions in which clitic *ŽE*, in addition to marking **contrastive focus**, also marks **semantic focus**. Below is an excerpt of a recording of a small dialogue between two speakers (their gender and age are unknown):

- 104) **Speaker 1:** Ты водку пьешь? /
 2SG.NOM vodka drink /
 ‘Do you drink vodka?’

Speaker 2: Нет / я даже запах не переношу / а ты? /
No / 1SG.NOM even smell not stand / and 2SG.NOM /
'No, I cannot even stand the smell, and you?'

Speaker 1: Да! / А что же ты пьешь? /
Yes! / And what *že* 2SG.NOM drink? /
'Yes. So, what do you drink (then)?'

Speaker 2: Шампанское / вино могу / только не очень крепкое! /
Champagne / wine can / only not very strong /
'Champagne... I can drink wine, only not very strong!'

Speaker 1: А коньяк? /
And cognac?
'And cognac?'

Speaker 2: Фу / гадость какая-то!
Yuck / grossness some
'Yuck, (how) gross!'

(Russian National Corpus)

Question with *ŽE* is repeated below:

Да! / А что же ты пьешь? /
Yes! / And [CF [SF what] *ŽE* 2SG.NOM drink? /
'Yes. So, what do you drink (then)?'

The presence of the clitic *ŽE* makes the question above more informal and perhaps slightly aggressive. When trying to find out what the speaker 2 drinks, the speaker 1 makes two presuppositions/assertions: (1) based on the previous utterance of speaker 2, speaker 2 does not drink vodka and (2) the only reasonable thing to drink is vodka and what on earth can the speaker 2 drink if (s)he does not drink vodka. Thus, speaker 1 places bounds/limitations on what the speaker 2 can reply. The bounds are set by the speaker 1's presuppositions above to all alcoholic drinks *but vodka*. However, although speaker 1 already partially knows the answer to his/her question implied by presupposition '*anything but vodka*', speaker 1 still expects to receive some new information when asking the question. Speaker 1 receives this information when the speaker 2 starts enumerating different drinks (s)he consumes instead of vodka. The request for new information suggests that *что*/'what' is also a **semantic focus** in this sentence. To summarize, the question above serves two purposes: it makes a

presupposition (resulting in presence of **contrastive focus**) and asks a question (resulting in presence of **semantic focus**). As a result, there appear to be two membership sets denoted by the *wh*-word in the example above, one of which is nested within the other membership set:

{_A vodka, non-vodka drinks {_a champagne, wine, cognac, etc } }

Membership set **A** is a limited set with only two members, one of which is contrasted to another member, representing **contrastive focus** in the sentence above. One of **A**'s members, i.e. member 'non-vodka drinks' contains another membership set **a**. Set **a** is unlimited set, representing semantic-focus.

This example of a *wh*-question, in which the clitic *ŽE* attaches to the *wh*-word, suggests that in the examples such as the one above **semantic focus** (with a set of unknown members) overlaps with **contrastive focus** (with a set of mutually exclusive members):

105) [CF [_{SF} *wh*-word]] *ŽE* ...?

In the example above *ŽE* has contrastive scope over the semantically focused *wh*-word.

It should be noted that there is another smaller subset of *wh*-questions with *ŽE* attaching to the *wh*-word that looks very similar to the example above, but has a slightly different information structure pattern in an appropriate context. These are questions, the answers to which are known to the speaker, but not to the audience, with a change of subject implied. Similar to the previous example in terms of syntactic structure, in the sentence below **contrastive focus** also overlaps with **semantic focus**, but with a different focus projection. In the example below the *wh*-word is semantic focus, while *ŽE* has scope over larger proposition, in which speaker is trying to change an entire subject of conversation from '*we are calling the second person*' to '*what is our main prize for today*'. Below is an example which is an excerpt of a recording from a morning radio show on radiostation MAXIMUM:

106) Speaker (Radio Host):

Все / вперед.
That's_it / onward
'That's it, onward'

Ну и побежал мужчина / побежал / побежал.
 So and ran man / ran / ran
 ‘So and the man ran, and ran, and ran.’

Ну а мы второму звоним.
 Well and we second call
 ‘Well, (meanwhile) we are calling the second one.’

А что же у нас за главный приз на сегодня?
 And what *že* at us for main prize for today
 ‘And/but what is our main prize for today?’

Стать пассажиром первого класса в настоящем гоночном
 To become passenger first class in real race
 ‘To become a passenger of the first class in a real

автомобиле / а можно отправиться на чемпионат России по
 automobile / and is possible leave to championship Russia-GEN of
 automobile, or it is possible to leave for a championship of Russia of

шоссейно-кольцевым гонкам / то есть попасть в самое нутро гоночного
 бизнеса.
 belt-line races / that is get into very entrails racing-GEN
 business.
 the belt-line races, that is to say, to get into the very entrails of the racing
 business.’

(Russian National Corpus, Radio Maximum)

In this sentence both the **contrastive focus** and **semantic focus** are present:

А что же у нас за главный приз на сегодня?
 And [_{SF} [_{SF} *what*] *ŽE* at us for main prize for today]?
 ‘And/but what is our main prize for today?’

107) ...[_{CF} [_{SF} *wh-word*] *ŽE*]... ?

In the sentence above there is a **semantic (new-information) focus** *что* / ‘what’, since the prize is unknown to the audience. As in the earlier example (83) with clitic **–TO** with a telephone conversation between two women implying the change of the subject, this sentence also contains **contrastive focus**, as there is a change of topic (of conversation) implied. Radio host’s question implies: ‘*while we are calling another listener, let us (change the subject from the call for now and) ask what is our main prize?*’ Two subjects of conversation are contrasted: ‘*we are calling the second person*’ is contrasted to ‘*what*

is our main prize for today'. As in the previous example, there are two membership sets denoted in this utterance:

{_A discussing the call, discussing the prize {_a participating as a passenger in the race, going to a race championship,} }

Membership set **A** denotes topics for conversations and is a limited set with only two members, one of which is contrasted to another representing **contrastive focus** in the sentence above. One of **A**'s members, i.e. 'discussing the prize' contains another membership set **a**. Set **a** is unlimited set, representing **semantic-focus**.

To conclude, in *wh*-questions with the clitic *ŽE*, this clitic is usually a strong indicator of contrast, usually of contrastive focus or contrastive focus overlapping with semantic focus. Questions examined in this section seem to require an analysis different from declaratives due to the nature of the utterances (*wh*-questions, usually implying presence of semantic focus) and the presence of the clitic *ŽE* (usually implying presence of contrastive focus because of the presence of contrasting entities or entire propositions in sentences containing this clitic). There are *wh*-questions in which the speaker makes an assertion instead of asking a question, in which case they contain **contrastive focus**. There are also *wh*-questions in which speaker both asks for new information and makes an assertion, which can be contrasted to another proposition in the context, in which case utterance contains both **semantic focus** and **contrastive focus**. There are also *wh*-questions in which the speaker changes the subject (contrasting one subject of conversation to another subject of conversation), but also makes a request for new information, in which case utterance also contains both **semantic focus** and **contrastive focus**.

4.5.2.6.4 *Wh*-words + *ŽE*: Declaratives

McCoy's analysis of the *wh*-words marked with *ŽE* as contrastive and being members of a set with mutually exclusive members also partially applies to declaratives. In **declaratives** below the clitic *ŽE* usually marks contrast, but not necessarily with

mutually exclusive members. Below is an excerpt from a recording of a casual conversation between two female students:

108) **Speaker 1:**

А как твой мч [молодой человек]?
And how your bf [boyfriend]?
'And how is your boyfriend?'

Speaker 2:

Да / вот говорю ж сёдня на коньках катались.
Well / here say ž today on skates rode.
'Well, I was just saying today we skated.'

Я иногда думаю / блин / какой **же** он умный / не то что я.
I sometimes think / dang / what **že** he smart / not that what I
'I sometimes think, how smart he is, not like me.'

Или у него жизненного опыта больше?
Or at him life experience more
'Or he has more life experience?'

Speaker 1:

Ну он же старше все-таки!
Well he že older after_all
'Well, he is older, after all!'

(Russian National Corpus)

In the sentence above, Speaker 2 is comparing herself with her boyfriend, saying that he is smarter than she is. There seems to be a scale of intelligence presupposed by Speaker 2, and on this scale of intelligence Speaker 2 ranks lower than her boyfriend.

The scale presupposed by the speaker/hearer in the statements with construction [... *wh*-word *že*...] is similar to the scales described in Beaver and Clark (2008). In their analysis of **scalar additives** (such as '*even*'), Beaver and Clark propose that these words, describing the addressees' expectations "relative to a salient ordering of propositions" are **focus sensitive** (Beaver and Clark 2008: 70-71). Similar to their description of '*even*' as focus sensitive, in the example above the *wh*-word *kakaja*/'how.SG.FEM.NOM' is **focused**²⁸:

²⁸ It is also possible that an entire proposition '*How smart he is*?' is focused, resulting in the following information structure pattern: [CF *wh*-word *ŽE* ...]. More research is necessary.

109) [_{CF} *wh*-word] *ŽE* ...

However, the set marked by clitic *ŽE* in this sentence of type [... *wh*-word *že*...] does not appear to be a set of mutually exclusive members, unless we consider the possibility that the elements weighted very far from each other, such as those placed on the opposite ends of the scale, can be interpreted as mutually exclusive.

To conclude, the clitic in the statements with *ŽE* attaching to *wh*-words usually signals the presence of **contrastive focus**, but it is **not always a set of mutually exclusive members**. An exception to when the *wh*-set marked by *ŽE* should be analyzed as a set of mutually exclusive members is when these members are measured on a presupposed scale.

4.5.2.6.5 *Wh*-words + *ŽE*: Part of expression

Another case when clitic *ŽE* attaching to *wh*-words does not mark contrastiveness is when it is used as a part of the established expression, as in the example below:

- 110) Ну что же, Нарбеков. Будем готовить приказ о вашем отчислении.
 Well what *ŽE* Narbekov. Will prepare order about your dismissal
 ‘So what (do we have here), Narbekov. We will be preparing order of your dismissal.’

(Russian National Corpus)

In the example above clitic *ŽE* can be potentially analyzed as a clitic signaling contrast. In the sentence above the speaker seems to mean that given the ‘present course of events’, the reasonable outcome is to prepare order of Narbekov’s dismissal. However, the presence of *ну что же*/‘so what *ŽE*’ also suggests that the course of events could have been different, in which case it would have not been necessary to dismiss Narbekov. So, indeed, there is a contrast present in the presupposition of ‘*what the events are*’ vs. ‘*what the events could have been*’. However, this analysis of the clitic *ŽE* as contrastive in this combination *ну что ЖЕ*/‘well what *ŽE*’ is slightly far-fetched, as this analysis: (1) can be applied to many sentences that are usually not considered contrastive, but just hint

at the possibility of a different course of events, and, most importantly, (2) can be applied to the same sentence as above without *ŽE*. For the purposes of this thesis, this use of *ŽE* will not be analyzed as contrastive, but further research would be beneficial to understand whether *ŽE* marks contrast when it is a part of this type of expression.

4.5.2.7 *ŽE*: Summary

In the preceding sections I have examined spoken data from the Russian National Corpus containing *ŽE*. The data showed that *ŽE* is most frequently placed in the second position in the sentence, after the first phonological word in the sentence. The words to which clitic *ŽE* attaches most frequently are: the deictics *this/that*, *wh*-words, and personal pronouns.

I analyzed clitic *ŽE* following McCoy (2001) as having two possible contrastive focus projections: **phrasal** (narrow contrastive focus projection) and **proposition-level** (wide contrastive focus projection). I examined the spoken data in the Russian National Corpus to see whether the contrastive focus projection of *ŽE* can be determined in relationship to its position in the sentence. The data showed that it is challenging to identify the contrastive focus projection of *ŽE* based on its position in the sentence, without knowledge of sentential stress or context. Both propositional and phrasal *ŽE* can be placed in the second position in the sentence. Additionally, it is also challenging to determine the position of the contrastively focused element relative to phrasal *ŽE*. Although usually contrastively focused element is placed before phrasal *ŽE* it can also be placed after phrasal *ŽE*.

McCoy's proposal to analyze the combination of *wh*-words with *ŽE* as marking contrast and a set of mutually exclusive members often applies to the spoken data, but with some exceptions. Particularly problematic for her analysis are sentences in which *ŽE* attaches to the *wh*-word in *wh*-questions (1 below), statements with *wh*-words (2 below), and *wh*-expressions (3 below):

1. *Wh*-question (overlap of semantic focus and contrastive focus):
[SF CF *wh*-word] *ŽE* ...?
2. *Wh*-statement (contrast, but not a set of mutually exclusive members)

3. *Wh*-expression (no contrast at all)

It should be noted that most of the *wh*-questions above with the clitic *ŽE* can be considered partially questions asking for new/unknown information (thus requiring semantic focus) and partially presuppositions/assertions restating what the hearer already knows and contrasting it to some presupposition inferable from the context (explaining the presence of contrastive focus). It could be the nature of these questions combined with presence of the clitic that makes them different from the other examples. Further research of the interrogatives containing *ŽE*, both *yes-no* and *wh*-questions, is required to understand the information structure patterns in interrogatives, as well as the role that clitics play in them.

4.5.3 Clitic *LI*

4.5.3.1 *LI* as a Contrastive Focus Marker

The interrogative clitic *LI* is usually analyzed as a focus marker. However, if the analysis of the clitics *-TO* and *ŽE* completed in preceding sections is extended to clitic *LI*, it appears possible that *LI* also can mark contrastive focus in the sentence, at least in the sentences in which it attaches to non-verbal constituents. These are the sentences in which the element preceding *LI* appears to be a member of a set in which it is contrasted to other alternative members. Analyzing *LI* as marking contrastive focus in the preceding element is also supported by the fact that the element preceding *LI* occupies a sentence-initial position, which is usually a position for contrastive focus (or topic) in Russian. Another reason why *LI* could be analyzed as a contrastive marker is because it is incompatible with *wh*-questions, which is characteristic of contrastive focus. Additionally, some of the contrastiveness tests used by Gryllia (2008) can be applied to sentences with *LI*, although in a ‘reverse’ way, suggesting that questions with *LI* contain contrastive focus.

In the sentence below ‘*Bukmop/Victor*’ appears to be a member in the set, in which it is contrasted with some other member that could have possibly done the action that is discussed in this utterance:

- 111) Виктор ли это сделал?
 [_{CF} Victor] *LI* this done
 ‘Was it Victor who did this?’

In addition to the notion (Vallduví & Vilkuna 1998) and characteristics (Molnár 2002) of contrast discussed in earlier sections, the contrastiveness tests used by Gryllia (2008) also suggest that the element preceding clitic *LI* can be considered a contrastive focus. The choice test cannot be applied to the sentence above, as it is a question. As for Kiss’ correction test used in Gryllia (2008), a statement contains contrastive focus if the element under consideration can be an answer correcting a question making a wrong presupposition. Since the correction test works by asking a question of a statement, it cannot be applied to the sentence above, since this it is already a question. However, it is possible to apply Kiss’ correction test in a ‘reverse’ way. In the ‘reverse’ correction test a question would be considered to contain a contrastively focused element if it can result in an answer with a contrastive focus correcting the presupposition in the question. If an answer with contrastively focused element answers a question containing clitic *LI*, the constituent preceded by *LI* can then be considered contrastively focused element:

- 112) *Question*
 Виктор ли это сделал?
 [_{CF} Victor] *LI* this done
 ‘Was it Victor who did this?’

Answer
 Нет, это сделал Данила.
 No it did [_{CF} Danila]
 ‘No, it was Danila (who did this).’

The correction test works in the ‘reverse’ way in this situation, providing us with a correcting answer with a contrastively focused element to the *yes-no LI*-question. This example, as well as similar examples, in which *LI* attaches to non-verbal elements in *yes-no* matrix questions, suggests that *LI* might be a clitic marking contrastively focused element in the sentence. Matrix *yes-no LI*-questions, in which *LI* attaches to the verb, are characterized by ambiguity of the scope of the clitic *LI* in such sentences, as in these

cases either the verb preceding *LI* can be contrastively focused, or an entire proposition. They deserve further consideration.

To provide more support for analysis of *LI* as a marker of contrastive focus, I carried out a small experiment with three native speakers basing it on the ‘reverse’ correction test. Although inconclusive, this test serves as another indication of the difficulties with (and sometimes inapplicability of) tests aimed at identifying contrastive elements in sentences that contain clitics.

Native speakers were provided with three sentences. First two sentences were pronounced with contrastive focus intonation (element in CAPS) and third sentence with a neutral intonation:

113) Response-sentences provided to native speakers:

1. ZHURNAL ona chitaet.
MAGAZINE she reads
‘It is the magazine that she is reading.’
2. Net. V BIBLIOTEKE ona.
no IN LIBRARY she
‘No. It is in the library where she is.’
3. Ona v biblioteke.
she in library
‘She is in the library.’

Then native speakers were asked to form a question that would produce the answers above. The goal of this test was to see whether speakers would ask a *LI*-question to result in the answers containing contrasting elements, which would suggest that *LI* signals presence of contrastive elements. However, native speakers did not form *LI*-questions, but instead they asked *yes/no* questions with intonation implying the presence of contrastive focus in the first two questions and a *wh*-question (as expected) for the third sentence with a non-contrastive response. All questions were of the type described below:

114) Questions formed by native speakers to produce responses above:

1. Ona chitaet knigu?
she reads book
‘Is she reading the book?’

2. Ona poshla v magazin?
she went to store
'Did she go to the store?'
3. Gde ona?
where she
'Where is she?'

Additionally, three native speakers were provided with short dialogues containing questions with and without *LI* and were asked to rate the acceptability of these short discourses:

115) I. (a) *Question*

Она читает книгу?
3SG.FEM.NOM reads book.ACC
'Is it a book that she is reading?'

Answer

Журнал она читает.
Magazine-ACC 2SG.FEM.NOM reads
'It is a magazine that she is reading.'

(b) *Question*

Книгу ли она читает?
Book-ACC *LI* 3SG.FEM.NOM reads
'Is it a book that she is reading?'

Answer

Журнал она читает.
Magazine-ACC 2SG.FEM.NOM reads
'It is a magazine that she is reading.'

II. (a) *Question*

Она пошла в магазин?
3SG.FEM.NOM went-FEM in store
'Is it the store that she went to?'

Answer

В библиотеке она.
In library-PREP 3SG.FEM.NOM
'It is in the library (where) she (is right now).'

(b) *Question*

В магазин ли она пошла?
In store *LI* 3SG.FEM.NOM went-FEM
'Is it the store that she went to?'

Answer

В библиотеке она.

In library-PREP 3SG.FEM.NOM

‘It is in the library (where) she (is right now).’

Interestingly, while rating all dialogues above as acceptable, in both contexts (I) and (II) all three speakers preferred the (a) questions without *LI*, but for the reasons that are not connected to the information structure. Questions containing *LI* were described as “long”, “too formal”, “bookish” and in one case as “unpleasant” in terms of possibly interpreted as implying some kind of a catch in the question.²⁹

As mentioned earlier, this test was not helpful in terms of identifying contrastiveness of elements in the questions with *LI*. However, this test suggests that it is possible that native speakers were just hesitant to form questions with *LI* (or other clitics as has been shown earlier) if other options were available, such as forming the same questions without these clitics but with intonation. Interestingly, if we were to assume that this test was nevertheless applicable and accurate, in this case King’s and others’ analysis of *LI* as marking focus is incorrect as well, as the 3rd question produced by native speakers and requiring an answer containing focus, was also not formed with the help of the clitic *LI*, but instead with intonation. As has been mentioned in previous sections on –*TO* and *ŽE*, more research is necessary in order to understand whether existing tests identifying information structure components can serve as an accurate measure of contrastiveness in spoken speech containing clitics in Russian.

Another example of *LI* marking contrastive focus is when it is a part of a disjunction *TO LI*, in which *LI* is a part of an expression aiding in forming a disjunction, as in the example below, which is an excerpt of a recording of a conversation between young women:

116) Speaker 1:

А у неё ремень с большой пряжкой...

And at 3SG.GEN belt with large buckle

‘And she has a belt with a big buckle...’

²⁹ Additionally, all three native speakers made comments regarding the sentence-initial position of contrastive focus in these dialogues as awkward. For them both replies would be more “normal” if the contrastively focused element was placed in the sentence-final position and highlighted as contrastive by stress.

Speaker 2:

Ты тоже ехала / да?
You also going / yes?
'You were also riding (the bus), yes?'

Speaker 1:

Не / мне кто-то рассказывал.
No 1SG.DAT someone told
'No, someone told (this) to me.'

То ли Маша / то ли Валя.
TO LI [_{CF} Masha] / TO LI [_{CF} Valja]
'Either Masha, or Valja.'

(Russian National Corpus)

In the dialogue above Speaker 1 retells Speaker 2 the story about a woman riding the bus with a big buckle on her belt. Speaker 1 does not remember who told her this story, but she remembers that it was either Masha or Valja. The above disjunction **TO LI** contains focused entity [_{CF} Masha], which is contrasted to another comparable entity [_{CF} Valja].

It should be noted, however, that as in the sentences and tests with clitics –*TO* and *ŽE*, the same question can be asked using intonation without *LI* in which case the constituent preceding it (focused in King's analysis or contrastively focused in this thesis) can still have the same interpretation if the intonation is preserved. More research on the relationship between prosody and information structure (particularly contrast) is necessary to make definite conclusions about whether clitic *LI* obligatorily marks preceding element for contrastive focus or whether it can be considered a clitic strongly signaling a presence of the clitic.

4.5.3.2 When *LI* is not a (contrastive) focus marker

There are exceptions to the analysis of *LI* as a (contrastive) focus marker. One of these exceptions is when *LI* is a part of an established expression.

If clitic *LI* is preceded by *wh*-word '*čto/what*' in the expression *čto LI*, it does not mark contrastive focus. An approximate meaning of *čto LI* in English is 'or what?', as in the example below:

- 117) Ты больной, что ли?
You sick *čto li*
'You are crazy, or what?'

(Russian National Corpus)

In the sentence above the *wh*-word *čto* preceding the clitic is not contrastively focused. Other cases in which *LI* does not necessarily mark (contrastive) focus will not be discussed in this thesis.

4.5.3.3 *LI*: Summary

To conclude, there seems to be evidence that the element preceding the clitic *LI* is not just focused, but can be also contrastively focused, according to contrastiveness tests and characteristics of contrast discussed above. More analysis of the spoken data with the clitic *LI* is necessary to further support the hypothesis that the clitic *LI* marks contrastive focus in the sentence.

Due to the limited scope of this thesis, only a few sentences have been analyzed with *LI*. Further analysis of spoken data containing *LI* is necessary to make any conclusions on whether (and when) *LI* can be regarded as marking contrastive focus instead of semantic focus. For the purposes of the current thesis and implementation of *LI* in grammar this clitic is analyzed as marking focus, i.e. underspecified between semantic focus and contrastive focus.

4.5.4 Clefting

4.5.4.1 Clefting as signaling contrastive focus

In this section I am going to extend the analysis of clitics as signaling contrastive focus to *it*-clefting in Russian. While *it*-clefting is usually analyzed as a construction involving focus (or semantic focus), it seems that it could also signal the presence of **contrastive focus** in Russian.

Below is an example from Russian National Corpus containing an *it*-cleft. This is an excerpt of a recording between man and woman on the street:

118) Female Speaker:

А ты же говорил, что они не глупые?
but 2SG.NOM že said.MASC that 3PL.NOM not stupid.PL.NOM
'But didn't you say that they are not stupid?'

Male Speaker:

Это она мне говорила, что не глупая...
it 3SG.NOM 1SG.DAT said.FEM that not stupid
'It was her who said that she is not stupid.'

(Russian National Corpus)

In the example above in second sentence speaker is contrasting '*her*' with '*himself*', in the sense that '*it was her (not him) that said that she is not stupid*'. A membership set is generated and is available to the speaker. This membership set has at least two members: '*he*' (male speaker) and '*she*' (male speaker's girlfriend), one of which is contrasted with another semantically and/or syntactically parallel constituent. This suggests that it is a contrastive element, based on the working definition defined in chapter 4.

Moreover, Rizzi's and Kiss' contrastiveness tests used by Gryllia (2008) also suggest that *it*-clefts in Russian mark **contrastive focus**. Gundel's example from her discussion on *it*-clefting in Russian and Japanese is repeated below and will be tested with contrastiveness tests used by Gryllia (2008):

- 119) Eto [Boris] vypil vodu.
It Boris drank vodka
It is Boris-FOC (who) drank the vodka.

(Gundel 1988, *The Role of Topic*: 1, as quoted in King 80)

Three tests used by Gryllia (2008) to access contrastiveness of the focus are applied to the sentence above (repeated here):

120) ***Wh*-question/*contrastive answer test**

A contrastive answer is incompatible with an ordinary *wh*-question.
(test adopted by Gryllia from Rizzi (1997)).

Question:

Kto vypil vodku?
Who drank vodka?
'Who drank vodka?'

Answer (1):

*Eto [Boris] vypil vodku.
It Boris drank vodka
'It is Boris-FOC (who) drank the vodka.'

Answer (3):

[Boris].
Boris
'Boris.'

Answer (3):

[Boris] vypil vodku.
Boris drank vodka
'Boris-FOC drank the vodka.'

For the *wh*-question above, answer (1) was never elicited from the native speakers.³⁰ This utterance appears ungrammatical in the current context, as *it*-cleft would imply that speaker presupposes that somebody else (i.e. not Boris) drank the vodka. However, this presupposition is false, hence answer (1) is ungrammatical. On the other hand, both answers (2) and (3) not containing *it*-cleft were elicited from native speakers. This suggests that *it*-cleft is incompatible with *wh*-questions, meaning that *it*-cleft passes the *wh*-question test for **contrastive focus**.

121) **Correction test**

A contrastive focus can be used to answer a yes-no question, correcting part of the predicate information of the question.

(test adopted by Gryllia from Kiss (1998))

³⁰ I carried out this small elicitation experiment. In this 4 native speakers were shown a picture of a man named Boris next to an empty bottle of vodka and were asked question 'Kto vypil vodku?/Who drank vodka?'

Applying the correction test to the *it*-cleft sentence ‘*Eto Boris vypil vodku*’ in Russian shows that this sentence can be used to answer a *yes-no* question ‘*Eto Ivan vypil vodku?*’:

Question

Eto Ivan vypil vodku?
It Ivan drank vodka?
‘(Was) it Ivan (that) drank vodka?’

Answer

(Net.) Eto [Boris] vypil vodku.
(No.) It Boris drank vodka
‘(No). It (was) Boris (that) drank vodka.’

Therefore, applying the correction test to the *it*-cleft above also suggests that *it*-cleft marks **contrastive focus**.

The constituent [Boris] also passes choice test used by Gryllia (2008):

122) **Choice test**

When answering an alternative question, one alternate is contrasted to the other.

Question:

Eto Ivan ili Boris vypil vodku?
It Ivan or Boris drank vodka?
(Was) this Ivan or Boris (that) drank vodka?

Answer:

Eto [Boris] vypil vodku.
It Boris drank vodka
‘It is Boris-FOC (who) drank the vodka.’

Applying the choice test to the *it*-cleft sentence ‘*Eto Boris vypil vodku*’ in Russian shows that this sentence can be used to answer an alternative question, in which one alternative member of the set ‘*Ivan*’ is contrasted to another alternative member of the set ‘*Boris*’.

The choice test above also suggests that *it*-cleft in Russian marks **contrastive focus**.

Interestingly, *it*-cleft can co-occur with the other contrastive clitics *TO* (without the hyphen), *ŽE* (and *VED*). Below is an example of the co-occurrence of *it*-cleft with *ŽE*:

- 123) Eto *ŽE* [Boris] vypil vodu.
 It *ŽE* Boris drank vodka
 'It is Boris-CF (who) drank the vodka.'

If the clitic *ŽE* is taken out from the sentence, the constituent [Boris] retains its contrastive meaning (this sentence was discussed earlier in the section). If the *it*-cleft is taken out of the sentence in (with clitic *ŽE* placed in the second position, as it is ungrammatical to have it in the sentence-initial position), there are several possibilities of contrastive focus projection, including narrow projection on [Boris] in (124) or wide projection with an all-contrast sentence interpretation as in (125):

- 124) [Boris] же vypil vodu.
 [Boris] *ŽE* drank vodka
 Boris-CF (who) drank the vodka.

- 125) [Boris же vypil vodu.]
 [Boris *ŽE* drank vodka]
 'Boris did drink the vodka.'

This suggests that both *it*-clefting and *ŽE* can be used interchangeably without the element losing its contrastive interpretation.

A similar analysis below is applied to a sentence from the Russian National Corpus. In this sentence *it*-cleft marks contrastive focus in spoken data from Russian National Corpus.³¹ This is a recording of two women in the train going to dacha and passing the train station Vtoraya Rechka. Speakers complain about the smell:

³¹ It-clefts are difficult to locate in the Russian National Corpus. Searching for NP following *это*/'it' returns many undesired results, including constructions of the following types:

- (1). Это Миша...
 This Misha
 'This is Misha...' (Russian National Corpus)
- (2). Это платье как-то плохо влияет на моего папу...
 This-SG.NEUT.NOM dress.SG.NEUT.NOM somehow badly influences on my Dad
 'This dress influences my Dad in some kind of bad way.' (Russian National Corpus)
- (3). А как это она так?
 And how this 3SG.FEM.NOM so
 'And how come she is this (way)?' (Russian National Corpus)
- (4). Это они едят только по праздникам вообще.
 This.NEUT.ACC 3PL.NOM eat-PL only on holidays in general
 'They eat this only during holidays in general.' (Russian National Corpus)

In most of the types above *это*/'it' (except for possibly (4)) does not mark contrastive focus.

126) Speaker 1

Вторая Речка / вонь такая // Как здесь люди живут!
Vtoraya Rechka / stench such // How here people live!
'Vtoraya Rechka... Such stench... How do people live here?!'

Speaker 2

Трубы какие-то//
pipes some
'Some (water) pipes...'

Speaker1

Какие трубы! Это сточные воды текут со всего города//
What pipes It [CF waste waters flow from entire city]
'What pipes?! It is waste waters that are flowing from the entire city'
(Russian National Corpus)

In the example above the constituent *сточные воды*/'waste waters'³² is contrasted to the proposition *трубы*/'pipes' mentioned by speaker 1. The same three tests used by Gryllia (2008) were applied to identify contrastiveness of the focus:

127) *Wh-question*/***contrastive answer test**

A contrastive answer is incompatible with an ordinary *wh*-question.
(test adopted by Gryllia from Rizzi (1997)).

Question:

Что это?
What that?
'What's that?' (referring to the subject that both speakers see)

Answer (1):

*³³Это сточные воды текут со всего города.
It waste waters flow from entire city]
'It is waste waters that are flowing from the entire city'

Answer (1) above suggests that *it*-cleft is incompatible with *wh*-questions. The element marked by *it*-cleft, i.e. *сточные воды (текут со всего города)*/'waste waters (flow here from entire city)' passes the *wh*-question test for **contrastive focus**.

³² or possibly the entire proposition *сточные воды текут со всего города*/'waste waters flow from entire city'

³³ This utterance is ungrammatical if the same intonation (highlighting 'waste waters' in comparison to 'pipes') is preserved.

128) **Correction test**

A contrastive focus can be used to answer a yes-no question, correcting part of the predicate information of the question.

(test adopted by Gryllia from Kiss (1998))

Below the correction test is applied to the *it*-cleft sentence *Это сточные воды текут со всего города*/'It is waste waters (that) flow here from entire city'. As shown below, this *it*-cleft sentence can be used to correct a statement *Это трубы*/'These are pipes':

Statement:

Это трубы.
this pipes
'(Is) it pipes?'

Answer:

Это сточные воды текут со всего города//
it waste waters flow from entire city
'It is waste waters that are flowing from the entire city'

Therefore, applying the correction test to the *it*-cleft above also suggests that *it*-cleft marks **contrastive focus**.

The proposition *сточные воды текут со всего города*/'waste waters (that) flow from entire city' also passes choice test used by Gryllia (2008):

129) **Choice test**

When answering an alternative question, one alternate is contrasted to the other.

Question:

Это трубы или сточные воды (текут со всего города)?
It pipes or waste waters flow from entire city
'Is it pipes or waste waters (that flow here from from the entire city)?'

Answer

Это сточные воды текут со всего города.
it waste waters flow from entire city
'It is waste waters that are flowing here from the entire city'

The *it*-cleft sentence *Это сточные воды текут со всего города*/'waste waters flow from entire city' answers an alternative question, in which one alternative member of the set *трубы*/'pipes' is contrasted to another alternative member of the set *сточные воды*/'waste waters'. The element marked by *it*-cleft *сточные воды (текут со всего*

зопода)/‘waste waters (flow here from entire city)’ passes choice test to identify **contrastive focus**.

It can be concluded that based on the definition and characteristics of contrast discussed earlier, the denoting of a membership set with semantically or syntactically parallel constituents, one of which is contrasted to another, as well as passing three tests for contrastiveness of focus by elements marked by *it*-cleft in both examples above, *it*-cleft in Russian signals presence of **contrastive focus** in the sentence.

4.5.4.2 Clefting: Summary

Based on the relationship between members in the set marked by *it*-cleft and application of Rizzi’s and Kiss’ contrastiveness tests to sentences with *it*-cleft in Russian it has been suggested that *it*-cleft³⁴ can be considered as an element strongly signaling presence of contrastive focus in Russian. However, more research and analysis of the spoken data is necessary to determine the plausibility of this hypothesis.

4.6 Summary

In this chapter I have discussed different diagnostics to identify information structure components in the spoken data. Using examples from the Russian National Corpus, I have showed that it is difficult to apply existing contrastiveness tests used by Gryllia (2008) to spoken data. In section 4.3 I have also tried to identify the diagnostics that could be used for differentiating between contrastive topics and contrastive foci. In section 4.4 I summarized the tests that could be applied to identify specific information

³⁴ It is possible that in addition to *it*-clefting, other types of cleft could also be analyzed as signaling presence of contrastiveness in Russian. Below is the sentence that I have constructed:

- 1) Прийти он пришел, но вот еды уже для него не осталось.
[CT To_come] 3SG.NOM [CF came] but here food already for 3SG.GEN no left
‘To come, he did come, but there was no left food for him.’

The information structure pattern in this sentence appears to be very similar to sentences with the clitic *-TO* (*Прийти-то он пришел, но вот еды уже для него не осталось* / [CT To_come-TO] 3SG.NOM [CF came] but here food already for 3SG.GEN no left), containing a **contrastive topic** (as identified by applying implicit-subquestion test used by Gryllia (2008)) and a **contrastive focus**.

structure components in the spoken data from the Russian National Corpus. I also formulated new working definitions and hierarchy of information structure components that I used for the analysis of the spoken data in later sections of this chapter. In section 4.5 I examined the spoken data containing *–TO* and *ŽE*, applying definitions and tests formulated in section 4.4. Finally, I proposed to extend the analysis of *–TO* and *ŽE* as markers of contrast to *it*-clefts and some cases of *LI*, applying the same contrastiveness tests and definitions as I have used for *–TO* and *ŽE*. While the analysis of *LI* as marking contrastive focus in some utterances is very inconclusive and needs further research, the analysis of *it*-clefting as marking contrastive focus appears more plausible, requiring more examples of the utterances from the spoken data to confirm this hypothesis.

Chapter 5: IMPLEMENTATION

5.1 Introduction

In the following sections I will briefly describe implementation of information structure marking means in Russian grammar, based on Song's ICONS analysis (Song & Bender 2012, Song forthcoming). Currently, out of four information marking means discussed in chapters 4 and 5 only two have been partially implemented, i.e. word order and clitics. Sentence/emphatic stress and *it*-clefting have not been implemented.

In the following sections I will discuss the information structure marking strategies in Russian that have already been implemented, as well as the information structure marking strategies that still remain to be implemented in the future. I will start with discussing current implementation of word order in Russian grammar. Then I will proceed to the discussion of current implementation of clitics *-TO*, *ŽE*, and *LI*. In last section of this chapter I will propose tentative implementation for *it*-clefting in Russian.

5.2 Background

5.2.1 The Grammar Matrix and HPSG

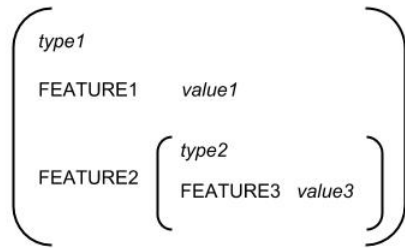
The Russian grammar that was used for implementation for this thesis was built by Esad Suskic and Varvara (Varya) Gracheva in the seminar LING567 taught by Professor Emily Bender. This small-sized grammar was built on the basis of the LinGO Grammar Matrix customization system (Bender et al. (2010) and Bender, Flickinger and Oepen (2002)). It was later extended to cover more linguistic phenomena by Esad Suskic, Sanghoun Song, and Varvara Gracheva. Some of the revisions to the grammar involving implementation of the information structure in Russian were made by Sanghoun Song and Varvara Gracheva and were based on the addition of the information structure library to the Grammar Matrix customization system by Sanghoun Song.

The Grammar Matrix provides an online customization system (Bender & Flickinger 2005, Bender et al. 2010), allowing user to describe different phenomena in

their language by providing answers in the questionnaire. The Grammar Matrix system contains libraries on word order, information structure, tense and aspect, negation, argument optionality, case, etc., reflecting the choices available to user on the in the Matrix customization questionnaire. After completion of the questionnaire the user can download a customized grammar and either use it for their purposes in its present state or expand it further if some of the linguistic phenomena are not covered in the Grammar Matrix customization system.

The Grammar Matrix is based on HPSG (Head-driven Phrase Structure), described in Pollard and Sag (1994) and Sag et al. (2003). HPSG allows modeling linguistic phenomena using **types** that have **features**, forming **typed feature structures**. The **types** are arranged into **type hierarchies** and organized by a system of **inheritance**. **Typed feature structures** are expressed through **attribute value matrices (AVMs)**. An example of an AVM is below:

Figure 5.1: Attribute Value Matrix (AVM)



(Drellishak 2009:7)

The AVM notation above is of *type1* that has two features, FEATURE1 and FEATURE2. The value of FEATURE1 is *value1*. The value of FEATURE2 is a nested feature structure of *type2*; its feature FEATURE3 has a value *value3*.

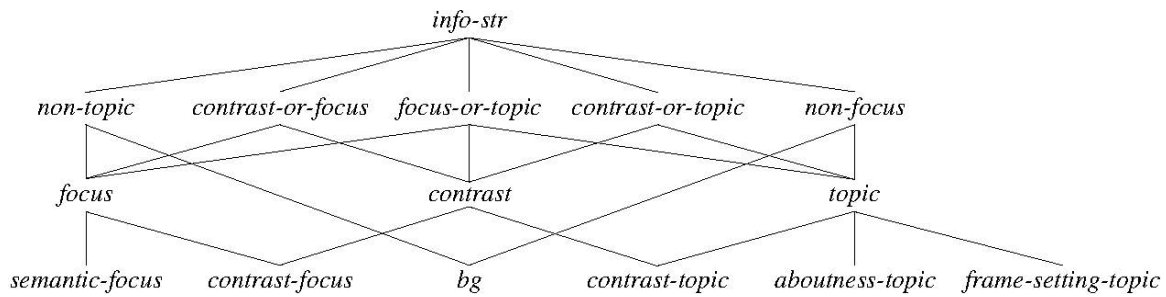
HPSG is the **constraint**-based system that works through **unification** by combining compatible typed feature structures. More details about unification operation and use of constraints as they relate to information structure implementation will be discussed in the next section discussing Song & Bender (2012) and Song (forthcoming).

5.2.2 ICONS

For the implementation of information structure in Russian, I will use Minimal Recursion Semantics (MRS)-based analysis (Copestake et al. 2005) of information structure proposed by Song & Bender (2012) and Song (forthcoming). The feature ICONS (Individual CONstraintS) adds coverage of the information structure in the MRS. The value of ICONS is a list of objects of type *info-str*. The objects of type *info-str* have the features CLAUSE and TARGET. The subtypes of *info-str* indicate the information structural role played by the TARGET with respect to the CLAUSE. The use of ICONS is aimed to account for focus projection from underspecified information structure representations, as well as for multiclausal sentences.

The use of ICONS in the Grammar Matrix within the HPSG framework allows for implementation of different information structure marking strategies in Russian and accounting for their interaction using constraints and underspecification. Use of constraints allows for coverage of the interaction of information structure marking strategies in Russian that can override each other. For example, in the absence of clitics marking information structure components in the sentence the default information structure pattern of the sentence is determined by the word order. The value of the sentence-final constituent can be constrained to *semantic-focus*. However, presence of clitic *-TO* overrides the word order in the sentence. The value of the sentence-final constituent of the construction containing *-TO* would be further constrained to *contrast-focus*. As to the underspecification, it allows for implementation of varying information structure components in the sentence as having an overarching value, e.g. if a constituent marked by a clitic can be contrastively or semantically focused, it can be underspecified as *focus*. Song (forthcoming) proposes the following information structure hierarchy (see next page):

Figure 5.2: Information structure hierarchy (Song forthcoming)



Song’s analysis of information structure will be adopted for the implementation of information structure in Russian. Song’s approach allows for placement of appropriate constraints on the values in the cases when the informational structural role of an element is clear. At the same time, it allows for underspecification of the value in the cases where there is not enough information to determine the information structural role of an element in a sentence.

5.3 Word order

As to the word order³⁵ implementation, there are two elements competing for the **sentence-initial position**: **topic** and **contrastive focus**. Since it is impossible to identify one or the other in this position without knowledge of surrounding context and intonation, sentence-initial position in the sentence remains underspecified as *info-str*, in order to also account for cases when presence of clitics overrides word order and sentence-initial position is occupied by semantic focus or background.

Usually there are two elements competing for **sentence-final position**: **semantic focus** and **contrastive focus**. **Semantic focus** usually occupies sentence-final position unless word order is overridden by clitics or by intonation in which the case sentence-final position can be occupied by **contrastive focus**. To capture the analysis that both

³⁵ It should be noted for this and other sections in implementation that intonation is not implemented. If intonation were to be taken into consideration, it could override all other means of marking information structure in Russian.

contrastive focus and **semantic focus** can appear in the sentence-final position, this position is for now underspecified as *focus*. Below is an excerpt from russian.tdl file constraining the phrase in the clause-final position to be *focus*:

```
130) clause-final-focus-phrase := binary-headed-phrase & clause-final-bound-phrase &
    [ SYNSEM [ CL-BOUND.CL-FINAL +,
      LOCAL.CAT [ MC na,
        MKG fc-only ] ],
    HEAD-DTR.SYNSEM.LOCAL.CAT.MKG non-fc,
    NON-HEAD-DTR.SYNSEM.LOCAL.CONT.ICONs <! focus !> ].
```

This rule constrains the ICONS value of the non-head daughter final element in the sentence to *focus*.

Below is an example of sentence with sentence-final position identified as semantic focus:

```
131) Спит Иван.
      Sleeps [ SF Ivan ]
      'It is Ivan that is sleeping.'
```

Currently in the grammar the value of SYNSEM.LOCAL.CONT.HOOK.ICONs-KEY of sentence-final NP 'Иван/Ivan' is underspecified as *focus*.

To summarize, sentence-initial position (usually reserved for topics or contrastive foci) has been underspecified to *info-str* and sentence-final position (reserved for semantic foci and contrastive foci) has been underspecified as *focus*.

5.4 *LI* implementation

Currently interrogative clitic *LI* is implemented in grammar as *focus*, which is underspecified between two analyses of this clitic, i.e. existing King's (1995) analysis of this clitic as **(semantic) focus** and analysis suggested in this thesis of it as **(contrastive) focus**. *LI* is a second-position clitic that attaches to sentence-initial elements constraining them to be focused.

In the lexicon.tdl file *LI*-marker is specified as a focus-marking-mod-lex with the stem spelling "ли":

132) ли-marker := focus-marking-mod-lex &
 [STEM < "ли" >].

Below is an excerpt from russian.tdl file for the focus-marking-mod-lex:

133) focus-marking-mod-lex := infostr-marking-mod-lex &
 [SYNSEM.LOCAL.CAT [MKG fc,
 HEAD.MOD < [L-PERIPH luk,
 LOCAL [CAT.HEAD +nv,
 CONT.HOOK.ICONS-KEY focus]] >]].

In the current implementation *LI* attaches as a modifier to nouns or verbs on the left-periphery of the sentence. *LI* marks elements for *focus* and constrains the ICONS-KEY value of the element it modifies to *focus*. *LI* has the following supertype in russian.tdl file, infostr-marking-mod-lex:

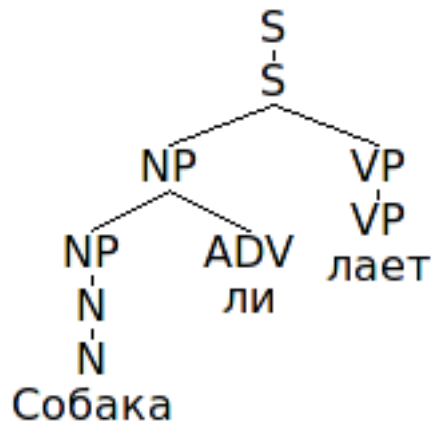
134) infostr-marking-mod-lex := no-rels-hcons-icons-lex-item &
 [SYNSEM [LIGHT +,
 LOCAL.CAT [MKG fc-only,
 HEAD adv &
 [MOD < [] >],
 VAL [SUBJ <>,
 COMPS <>,
 SPR <>,
 SPEC <>]]]].

The supertype for *LI* marks words for *focus*. The head of the phrase with *LI* is the modifier *LI* itself, which is treated as an adverb in our grammar.

As a result of this implementation, sentence (139) has the following syntactic representation (figure 6.2):

135) Собака ли лает?
 Sobaka LI laet
 Dog LI barks
 'Is it the dog that is barking?'

Figure 5.4: Собака ли лает?



The focus-marking-mod-lex *ли*/'li' is located in the second position in the sentence, attaching to the sentence-initial element *собака*/'dog'. The ICONS-KEY value of the NP *собака*/'dog' is constrained to *focus*.

If future research of data containing *LI* shows that *LI* marks constituent that it modifies as *contrast-focus* or as *semantic-focus* depending on the part of speech of the constituent to which it attaches, then *LI* could be unspecified as *focus* in order to be compatible with both analyses.

Future work in the implementation of clitic *LI* could also include the implementation of *LI* when it attaches to verbs as ambiguous between marking narrow focus on the verb and wide focus on the entire sentence.

5.5 -*TO* implementation

Although -*TO* has been referred to as “clitic” in this thesis, it is not implemented as such in the grammar. Instead, -*TO* is implemented as an affix because it attaches to words and is constrained by lexical rules. -*TO* does not displace other 2nd position clitics such as *LI*, as it is an affix and can be located in any position. Currently -*TO* marks the constituent to which it attaches as a *topic*, which turns out to be the shared characteristic of the majority of the information structure patterns summarized in the paragraphs below. The other *TO* (spelled without the hyphen) that has been briefly discussed earlier in this thesis and that

is used for forming disjunctions or is a demonstrative pronoun is not currently implemented in the grammar.

The most common information structure pattern for declaratives containing *–TO* appears to be the one described by McCoy (2001):

- [CT]–*TO* ... [CF]...

However, although McCoy’s analysis accounts for the information structure pattern in some interrogatives and sentences containing verb *бѳмѳ* ‘to be’, there are other interrogatives and sentences with *бѳмѳ* ‘to be’ that have different information structure patterns. For sentences with verb *бѳмѳ* ‘to be’ containing *–TO* the following information structure patterns have been discovered in chapter 5:

- [CT *be*]–*TO* [CF]
- [T *be*]–*TO* [CF]
- [SF *wh*-word] [T *be*]–*TO* ?

As to the interrogatives, the following information structure patterns have been discovered in chapter 5 for the *yes-no* questions and *wh*-questions containing *–TO*:

Yes-no questions:

- ...[CT]–*TO* [CF] ...?
- ...[CF] [T]–*TO* ...?

Wh-questions:

- [SF *wh*-word] [CF verb]–*TO*?
- [SF *wh*-word] [T verb]–*TO*?
- [CF *wh*-word] [T/(BG?) verb]–*TO*?

The information structure patterns in sentences with verb *бѳмѳ* ‘to be’ and the interrogatives described above share one common characteristic. In all of these patterns, except for one *wh*-question, clitic *–TO* attaches to a topic (contrastive or non-contrastive). Implementing the constituent to which *–TO* attaches as a *topic* (underspecified for contrast) allows capturing the majority of possible information structure patterns in the sentences containing *–TO*, regardless of the illocutionary force and context.

However, although it does not cover all potential information structure patterns, implementing constituent to which *–TO* attaches as a *topic* allows accommodating **majority** of the utterances containing *–TO*. Additionally, constraining the value of the

constituent to which *–TO* attaches as a *topic* specifies (at least in a limited number of cases) this information structure component, which is underspecified in all other sentences in which *–TO* is absent. Lastly, underspecifying this constituent for *contrast* covers both contrastive and non-contrastive topics, which are both possible as elements to which *–TO* attaches.

Currently *–TO* is implemented only as attaching to verbs or nouns. Below is the following excerpt from *irules.tdl*:

```
136) to-after-noun-suffix :=
      %suffix (* -to)
      to-after-noun-lex-rule.

      to-after-verb-suffix :=
      %suffix (* -to)
      to-after-verb-lex-rule.
```

The following lexical rules have been added to *russian.tdl* to constrain *–TO* attaching to nouns and to verbs:

```
137) to-after-noun-lex-rule := noun-with-TO-topic-marker-lex-rule-super &
      [ SYNSEM.LOCAL [ CAT.MKG tp,
        CONT.HOOK.ICON-KEY topic ] ].

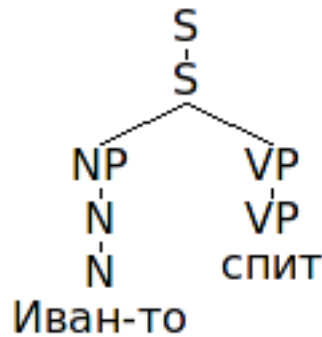
138) to-after-verb-lex-rule := verb-with-TO-topic-marker-lex-rule-super &
      [ SYNSEM.LOCAL [ CAT.MKG tp,
        CONT.HOOK.ICON-KEY topic ] ].
```

Lexical rules above, i.e. *to-after-noun-lex-rule* and *to-after-verb-lex-rule* both constrain *–TO* to be a topic marker. Additionally, they constrain the *ICON-KEY* value of the element to which they attach as *topic*.

As a result of this implementation, sentence (135) has the following syntactic representation (figure 6.3):

```
139) Иван-то спит.
      Ivan-TO sleeps
      'As to Ivan, he is sleeping.'
```


Figure 5.5: *Иван-то спит*/'Ivan-TO sleeps'



The to-after-noun-suffix *-mo*/'TO' attaches to the noun *Иван*/'Ivan'. The ICONS-KEY value of the NP *Иван*/'Ivan' is constrained to *topic*.

As to the future work in the implementation of clitic *–TO*, it could possibly go in several directions, which would include:

- (1). Extending the coverage of *–TO* attaching to other parts of speech, such as adverbs and adjectives. Currently *–TO* is constrained to attach only to nouns and verbs.
- (2). Implementing *–TO* as marking the other element in the sentence as *contrast-focus*.
- (3). It should also be noted that the current implementation approach constraining value of the constituent that *–TO* modifies to *topic* does not cover the following information structure patterns below:

- *wh*-question of type: [_{SF} *wh*-word] [_{CF} verb]-*TO*?
- topic-less sentences of type: [_{CF}]-*TO*?

The types accounting for information structure patterns above could be implemented in the future with creation of another entry for *–TO* that marks the constituent to which it attaches as *contrast-focus*.

- (4). It also appears that the information structure patterns listed earlier in this section share another common characteristic, i.e. in all of them (except for *wh*-questions and one type of *yes-no* questions) the sentence-final

constituent is a **contrastive focus**. Implementing sentence-final constituent as *contrast-focus* would account for the majority of utterances containing –*TO*, except for the *wh*-questions, as well as *yes-no* questions in which contrastively focused constituent precedes the word to which –*TO* attaches:

Yes-no questions:

- ...[_{CF}] [_T]–*TO* ...?

Wh-questions:

- [_{SF} *wh*-word] [_{CF} verb]–*TO*?
- [_{SF} *wh*-word] [_T verb]–*TO*?
- [_{CF} *wh*-word] [_{T/(BG?)} verb]–*TO*?

Based on the information structure patterns listed above, future implementation of *wh*-questions containing –*TO* could include implementing *wh*-words as *semantic focus* for all of the *wh*-questions except the ones in which *wh*-word should be implemented as *contrast-focus*:

[_{CF} *wh*-word] [_{T/(BG?)} verb]–*TO* ?

(5). Lastly, clitic *TO* (spelled without the hyphen) forming disjunctions and briefly described with examples earlier in chapter 5, also deserves further research and consideration as a marker of information structure in the sentence, potentially implemented as *contrast-focus*.

Accounting for –*TO* attaching to different parts of speech or to *wh*-words, for –*TO*’s role in topicless sentences, *yes-no* and *wh*-questions, as well as for *TO* (without the hyphen) forming disjunctions would yield a more extensive implementation of this clitic in Russian grammar.

5.6 *ŽE* implementation

The clitic *ŽE* is currently not implemented in grammar. As has been shown in chapter 5, it is challenging to predict whether it marks narrow or wide focus, i.e. if it is a phrasal vs. proposition-level *ŽE*. It is also difficult to predict, whether element contrastively marked by phrasal *ŽE* is located before or after *ŽE*.

The only prediction that can be made at this time is that when *ŽE* attaches to *wh*-words, *wh*-words can be underspecified as *focus* (accounting for cases when *wh*-words in sentences with *ŽE* are *semantic-focus* or *contrast-focus* or in some cases both). Thus, *ŽE* attaching to *wh*-words could be underspecified as *focus* or possibly implemented as possibly marking *wh*-word to which it attaches for different types of focus, i.e. *ŽE* marking *semantic-focus* or *contrast-focus* or both.

The analysis and resulting implementation of *ŽE* when it attaches to any other words deserves further research and consideration.

5.7 *It*-clefting implementation

While it has been shown in chapter 5 that *it*-clefting possibly marks **contrastive focus** on the element that follows it, it has not been yet implemented as such. An example sentence examined in chapter 5 is repeated in (140):

- 140) Eto [Boris] vypil vodu.
It Boris drank vodka
It is Boris-FOC (who) drank the vodka.
(Gundel 1988, *The Role of Topic*: 1, as quoted in King 80)

In the future this type of *it*-cleft could be implemented as marking *contrast-focus* on the constituent that is placed after it, with the rest of the sentence as background. A similar example would be a sentence in which *it*-cleft is followed by contrastive clitic *ŽE*:

- 141) Это же Борис выпил водку.
It *ŽE* [_{CF} Boris] vypil vodu.
'(But) it was BORIS (who) drank vodka.'

In the sentence above in which '*jeto*' is followed by *ŽE*, it still marks **contrastive focus**, but the element that should be implemented as *contrast-focus* is now AFTER the clitic *ŽE*.

Chapter 6: CONCLUSION

6.1 Summary

In this thesis I have looked at the information structure marking in Russian, with an emphasis of the clitics marking contrast in the sentence. I have reviewed the literature on information structure in general and information structure in Russian in particular. In my review of the literature on information structure in general I have included various tests and notions/characteristics used to identify information structure components, i.e. focus, topic, and contrast. I have also briefly discussed different information structure strategies employed in Russian, i.e. word order, intonation, *it*-clefting, and clitics marking information structure.

For my analysis, I have mostly concentrated on clitics *-TO* and *ŽE* marking contrast in Russian, extending McCoy's (2001) analysis to spoken data from the Russian National Corpus (Grishina 2006). While trying to identify the elements that these clitics mark for contrast, I found that the currently used diagnostics for identifying information structure components in the sentence are not easily applicable to spoken data. It appears that contrastiveness, topic, and focus tests often yield inconclusive results when applied to data containing clitics used in this thesis. Additionally, when testing data with obtaining grammaticality judgments from native speakers, it was found that while *-TO*, *ŽE*, and *LI* are very frequently used in informal spoken speech, native speakers are very reluctant to produce them in a test environment. This is possibly caused by several characteristics of these clitics, such as the informality of *-TO* and *ŽE*, aggressiveness of *ŽE*, and formality of *LI*, as well as the availability of forming sentences using intonation with the same meaning without the usage of *-TO*, *ŽE*, and *LI*.

As a possible solution to this problem, I applied the above-mentioned tests in conjunction with the notions/characteristics of contrast/topic/focus to identify information structure components more accurately. Even though they were not always conclusive, these tests are helpful in understanding the relationship between speech acts and information structure marking. It appears that speakers use contrastive focus to correct presuppositions and answer alternative questions, while contrastive topic is used to indicate existing alternatives, which makes it similar to contrastive focus in this regard.

However, still better diagnostics for identifying information structure components in spoken speech are preferable.

As to the different usage of clitics *-TO* and *ŽE*, as well as the different marking of information structure patterns by these clitics in different contexts, only a few possibilities were explored, due to the limited scope of this thesis. Based on the examples from the Russian National Corpus, it was confirmed that the clitics *-TO* and *ŽE* often mark the presence of contrast in the sentence. However, there are many exceptions to the generalization, at the foundation of McCoy's (2001) analysis, that these clitics mark contrast. Different information structure patterns marked by these clitics can be determined by context and by the illocutionary force of the utterances.

I have also briefly reviewed several usages of the clitic *-TO*, concentrating mainly on its usage as a contrastive marker. I tried to apply some of the existing tests to identify information structure components, as well as the theory of information structure to McCoy's analysis of clitic *-TO* and to the spoken data in the Russian National Corpus. It was found that some of the currently existing contrastiveness, topic, and focus tests are not always applicable to the spoken data because of some of the characteristics of the clitics or the illocutionary force of the utterances. When these tests are applicable, they are often inconclusive because they can predict opposite results. Additionally, I tried to apply McCoy's analysis of *-TO* as attaching to topic and marking both topic and focus for contrastiveness to data in Russian National Corpus. Close examination of the spoken data showed that in some declaratives, as well as some yes/no questions, and *wh*-questions the clitic *-TO* marks only contrastive focus, but attaches to a non-contrastive topic. Additionally, in some of the *wh*-questions *-TO* is not a marker of contrast, i.e. it does not mark for contrast either focus or topic. It was concluded that *-TO* is not an unambiguous contrast marker and that it marks different patterns of information structure in utterances with different illocutionary force. The interaction of *-TO* with other lexical elements was briefly examined, suggesting that further research on the utterances containing several clitics could shed some light on the contribution of each individual clitic to the utterance.

As to the clitic *ŽE*, I also tried to apply existing tests to identify information structure components in this sentence. Similar to the clitic *-TO*, tests to identify information structure components used in this thesis were not always applicable to data

with *ŽE* because of the aggressiveness of the clitic or illocutionary force of the utterances. Based on the data examined, *ŽE* always marks contrastive focus, sometimes overlapping with semantic focus. I attempted to analyze *ŽE* as marking different elements for contrastive focus (i.e. constituents vs. entire propositions) relative to its position in the sentence, with proposition-level *ŽE* often appearing in the 2nd position in the sentence and with phrasal *ŽE* often appearing in other positions in the sentence. However, based on the data found in the Russian National Corpus, this analysis does not apply to *ŽE*. The clitic *ŽE* in 2nd position is ambiguous in terms of marking a constituent vs. an entire proposition for focus. Additionally, the position of the contrastively focused element relative to phrasal *ŽE* is ambiguous as well, as it can be placed both before and after the clitic.

The means of marking information structure in Russian discussed in this thesis have been partially implemented in a small HPSG grammar. As has been mentioned earlier, intonation, clefting and *ŽE* have not been implemented, but other information structure marking strategies have been implemented to the extent allowed by scope of this thesis and limited analysis of data from the Russian National Corpus. For word order implementation sentence-initial position has been underspecified as *info-str*, while sentence-final position has been specified as *contrast-focus*. The clitics *LI* and *-TO* have been partially implemented, while the implementation of *ŽE* requires further research.

6.2 Future Work

Based on the investigation of the spoken data completed in this thesis, possible phenomena for future implementation of contrast in Russian include *it*-clefting (possibly implemented as *contrast-focus*), *LI*-interrogatives with constituents to which *LI* attaches possibly marked as *contrast-focus*, as well as more types of information structure patterns in utterances containing *-TO* and *ŽE*. However, more research on clitics and the spoken data containing clitics is necessary to determine how the above-mentioned information structure strategies should be implemented in grammar.

As was shown in Chapter 2, clitics are used very frequently in informal speech, which makes spoken data the most useful source of utterances containing clitics. Analysis

of the spoken data with clitics would provide a better understanding of the clitics' usage in the spoken data and their role in marking information structure. Several characteristics of clitics should be examined in more depth: different functions of each clitic; the usage of clitics in utterances with different illocutionary force (declaratives vs. interrogatives vs. exclamatives); properties of clitics that make it difficult to apply diagnostics identifying information structure components to utterances containing clitics; and, finally, properties of clitics that make it difficult to obtain native speakers' judgments on grammaticality of utterances containing clitics. It would also be helpful to examine interaction of the contrastive clitics by looking at the utterances containing more than one contrastive clitic. In addition to understanding of how usage of several contrastive clitics in the utterance affects its interpretation, this might also allow for better understanding of the scope and input of each individual clitic. Finally, it would be helpful to examine in depth interrogatives containing clitics, both *yes/no* questions and *wh*-questions. As has been shown in Chapter 5, some *wh*-questions with *ŽE* can be considered interrogatives containing presuppositions/assertions, which impacts their information structure patterns. Examining *yes-no* and *wh*-questions with clitics will hopefully result a better understanding of information structure patterns in interrogatives, as well as the role played by contrastive clitics in interrogatives.

A more detailed investigation of each clitic individually, interaction between contrastive clitics, and role that clitics play in interrogatives would hopefully inform a better understanding of contrastive clitics and their use in information structure marking. All of the above could make a contribution to a more accurate analysis and implementation of contrast and other information structure components in Russian.

BIBLIOGRAPHY

- Bailyn, J. (2011). *The syntax of Russian*. Cambridge University Press.
- Beaver, D., & Clark, B. (2008). *Sense and sensitivity: How focus determines meaning*. Blackwell Publishing.
- Belletti, A. (2004). Aspects of the low IP area. In L. Rizzi (Ed.), *The Structure of CP and IP: The Cartography of Syntactic Structures*. Oxford: Oxford University Press, 2, 52-76.
- Bender, E., Drellishak, S., Fokkens, A., Poulson, L., & Saleem, S. (2010). Grammar customization. *Research on Language and Computation*, 8(1), 23-72.
- Bender, E., Flickinger, D., & Oepen, S. (2002). [The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars](#). In John Carroll, Nelleke Oostdijk, and Richard Sutcliffe, editors, *Proceedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics*, pp. 8–14, Taipei, Taiwan.
- Bender, E., & Flickinger, D. (2005). Rapid Prototyping of Scalable Grammars: Towards Modularity in Extensions to a Language-Independent Core. In *Proceedings of the 2nd International Joint Conference on Natural Language Processing IJCNLP-05 (Posters/Demos)*.
- Benincà, P., & Poletto, C. (2004). Topic, focus, and V2: Defining the CP sublayers. In L. Rizzi (Ed.), *The Structure of CP and IP: The Cartography of Syntactic Structures*. Oxford: Oxford University Press, 2, 52-76.
- Bitextin, A. B. (1994). *Chastitsy -TO, ZHE, VED' i vvodnye konstruksii tipa KAK IZVESTNO kak sredstva ukazaniya na izvestnost' propozitsional'nogo sodержaniya predlozheniya slushajuschemu*. Unpublished doctoral dissertation, Moskovskij gosudarstvennyj universitet, Moskva.
- Bjerre, Anne (2011). Topic and focus in local subject extractions in Danish. Presented at the 18th International Conference on Head-Driven Phrase Structure Grammar, August 23, 2011.
- Büring, D. (2003). On D-Trees, Beans, and B-Accents. *Linguistics and Philosophy*, 26(5): 511-545.
- Büring, D. (2005). *Semantics, Intonation, and Information Structure*. Retrieved from <http://semanticsarchive.net/Archive/GQ0YjgxM/buring.information.structure.v2005.pdf>
- Büring, D. (1997). *The Meaning of Topic and Focus — The 59th Street Bridge Accent*. London: Routledge.
- Büring, D. (1999). *Topic*. In P. Bosch and R.v.d. Sandt (eds.), *Focus: Linguistic, Cognitive, and Computational Perspectives*. Cambridge University Press, 142-165. Argues that like focus, topic also is realized phonologically in the form of a specific pitch accent.

- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics and point of view. In C. Li (Ed.), *Subject and Topic*, 27-55. New York: Academic Press.
- Choi, H.-W. (1999). *Optimizing Structure in Context: Scrambling and Information Structure*. Stanford, CA: CSLI Publications.
- Copestake, A. (2002). Implementing Typed Feature Structure Grammars. CSLI Publications.
- Copestake, A., Flickinger, D., Pollard, C., & Sag, I. (2005). Minimal Recursion Semantics: An Introduction. *Research on Language & Computation*, 3(4), 281–332.
- Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition*, 7, 49-59.
- Dipper, S., Goetze, M., & Skopeteas, S. (2007). *Information Structure in Cross-linguistic Corpora: Annotation Guidelines for Phonology, Morphology, Syntax, Semantics and Information Structure*. Universitätsverlag Potsdam.
- Drellishak, S. (2009). [Widespread but Not Universal: Improving the Typological Coverage of the Grammar Matrix](#). PhD thesis, University of Washington.
- Engdahl, E., & Vallduví, E. (1994). Information packaging and grammar architecture: A constraint-based approach. In Elisabet Engdahl (Ed.), *Integrating information structure into constraint-based and categorical approaches* (DYANA-2 Report R.1.3.B), 39-79. ILLC, Amsterdam. Available from http://www.justinecassell.com/discourse/pdfs/engdahl_Vallduví94.pdf
- EXMARaLDA. Available from <http://www.exmaralda.org/>
- Fillmore, C. J. (1976). Pragmatics and the description of discourse. In S. Schmidt (ed.), *Pragmatik II*. Munich: Wilhelm Fink Verlag. Reprinted in Cole (Ed.), 1981. (Quoted in Lambrecht 1996).
- Ginzburg, J., & Sag, I. (2000). Interrogative investigations: The form, meaning, and use of English interrogatives. CSLI Publications.
- Gracheva, V., & Song, S. (2012). Information Structure Annotation: Adapting Potsdam SFB 632 Information Structure Guidelines to Narrative Data Containing Dialogues. Available from <http://corpus.mireene.com/ssh/publications/BeboAnnotationGuideline.pdf>
- Grishina, E. (2006). Spoken Russian in the Russian National Corpus (RNC). Proceedings of the 5th International Conference on Language Resources and Evaluation, 121-124.
- Gryllia, S. (2008). On the nature of preverbal Focus in Greek: A theoretical and experimental approach. Dissertation.
- Gundel, J. K. (1988). Universals of topic-comment structure. In M. Hammond et al, (Eds.), *Studies in Syntactic Typology*, John Benjamins, 209-239.

- Gundel, J. K. (1999). *On different kinds of focus*. In P. Bosch and R.v.d. Sandt (Eds.), *Focus: Linguistic, Cognitive, and Computational Perspectives*. Cambridge University Press, 293-305.
- Gundel, J., & Fretheim, T. (2001). Topic and Focus. In L. Horn and G. Ward (Eds.), *Handbook of Pragmatic Theory*. Oxford: Blackwell.
- Halliday, M.A.K. (1967). *Notes on Transitivity and Theme in English*, Part II. *Journal of Linguistics*, 3, 199–244.
- Heycock, C. (1994). Focus Projection in Japanese. In proceedings of *North East Linguistic Society*, vol. 24.
- Junghanns, U., & Zubatow, G. (1995). Syntax and information structure of Russian clauses. Annual Workshop on Formal Approaches to Slavic Linguistics: The Cornell Meeting 1995. Michigan Slavic Publications.
- Jackendoff, R. S. (1972). *Semantic Interpretation in Generative Grammar*. Cambridge, MA: The MIT Press.
- King, T. H. (1995). Configuring Topic and Focus in Russian. Dissertation.
- Kiss, K. E. (1998). Identificational focus versus information focus. *Language*, 74, 245-273.
- Krifka, M. (1991/2). A Compositional Semantics for Multiple Focus Constructions. In Jacobs (1991/2a), 17–53.
- Krifka, M. (2007). Basic notions of information structure. In C. Féry, G. Fanselow, and M. Krifka (Eds.), *Interdisciplinary Studies of Information Structure*, 6, 13-56. Potsdam.
- Krylova, O., & Khavronina, S. (1988). *Word order in Russian*. Moscow, USSR: Russky Yazyk Publishers.
- Lambrecht, K. (1996). *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge, NY: Cambridge University Press.
- Leipzig Glossing Rules. Available from: <http://www.eva.mpg.de/lingua/resources/glossing-rules.php>
- Levontina, I. B. (2008). Загадки частицы уж. The riddles of the Russian particle *uzh*. Труды международной конференции «Диалог 2008». Retrieved from <http://www.dialog-21.ru/dialog2008/materials/html/47.htm>
- Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, 8, 339-359.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- McCoy, S. (2001). Colloquial Russian particle –to, že, and ved’ as set-generating (“konstrastive”) markers: A unifying analysis. Dissertation. Boston University.

- McCoy, S. (2002). Semantic and discourse properties of colloquial Russian construction of the form “X-to X, a...” *Glossos*, 3. Available from: <http://seelrc.org/glossos/issues/3/mccoy.pdf>
- McCoy, S. (2003). Connecting Information Structure and Discourse Structure through ‘Kontrast’: The Case of Colloquial Russian Particles -TO, ŽE, and VED’. *Journal of Logic, Language, and Information*, 12(3), 319-335.
- McCoy, S. (2007, May 4-6). Decomposing Particles in Combination: Colloquial Russian –to, že, and ved’. FASL 16, Stony Brook University.
- Molnár, V. (2002). Contrast from a contrastive perspective. In H. Hasselgrd, S. Johansson, B. Behrens, & C. Fabricius-Hansen (Eds.), *Information Structure in a Cross-Linguistic Perspective*, 147–162. Amsterdam, Netherland: Rodopi.
- Neeleman, A., & Titov, E. (2009). Focus, contrast, and stress in Russian. *Linguistic Inquiry*, 40(3), 514–524.
- Parrott, L. (1990). Argumentative že in discourse. *Harvard Studies in Slavic Linguistics* 1: 83-104.
- Parrott, L. (1997). Discourse organization and inference: The usage of the Russian particles že and ved’. Ph.D. dissertation, Harvard University.
- Pollard, C., & Sag, I. (1994). Head-driven Phrase Structure Grammar. University of Chicago Press.
- Prokurovskaja, N. A. (1977). Neznamental’naja leksika v ruskoj razgovornoj reči (sostav i funkcii). AKD: Saratov.
- Sag, I., Wasow, T., & Bender, E. (2003). Syntactic Theory: A formal introduction. CSLI Publications.
- Rizzi, L. (1997). The fine structure of the left periphery. In I. Haegeman (Ed.), *Elements of Grammar*, pp. 281-337. Dordrecht, Kluwer.
- Rooth, M. (1985). Association with Focus. Doctoral Dissertation, University of Massachusetts, Amherst.
- Rooth, M. (1992). A Theory Of Focus Interpretation. *Natural Language Semantics*, 1, 75-116.
- Russian Grammar (2010). Developed in LING 567 class taught by Emily Bender.
- Russian National Corpus. Available from <http://ruscorpora.ru/en/index.html>
- Schmidt, T. (2012). EXMARaLDA and the FOLK tools. In Proceedings of LREC, ELRA. Available from http://www.lrec-conf.org/proceedings/lrec2012/pdf/529_Paper.pdf
- Schütze, C. T. (1994). Serbo-Croatian Second Position Clitic Placement and the Phonology-Syntax Interface. Revised version of previously published manuscript. *MIT Working Papers in Linguistics*, 21, Papers on Phonology and Morphology, pp. 373-473. Current version retrieved from <http://www.linguistics.ucla.edu/people/cschutz/sc2pclit.pdf>

- Schwabe, K. (2004). The Particle *li* and the Left Periphery of Slavic Yes/No Interrogatives. Retrieved from http://www.zas.gwz-berlin.de/fileadmin/mitarbeiter/schwabe/papers/The_particle_li_2004.pdf
- Slioussar, N. (2007). Grammar and Information Structure: A study with reference to Russian. Dissertation.
- Song, S. (forthcoming). Dissertation. University of Washington.
- Song, S., & Bender, E. (2012). Individual Constraints for Information Structure. In Stefan Müller (Ed.), *Proceedings of the 19th International Conference on Head-Driven Phrase Structure Grammar* (pp. 330-348). CSLI Publications.
- Vallduví, E. (1990). The informational component. Ph.D. dissertation: University of Pennsylvania. (Quoted in Lambrecht 1996).
- Vallduví, E., & Vilks, M. (1998). On rheme and kontrast. In P. Culicover & L. McNally (Eds.), *Syntax and Semantics*, 29, 79-108. The limits of syntax. New York: Academic Press.
- Yokoyama, O. (1986). *Discourse and word order*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Zwicky, A. (1985). Clitics and Particles. *Language*, 61(2), 283-305.

Appendix A: Annotated sentences

- (I) Eto [Boris] vypil vodu.
 It Boris drank vodka
 It is Boris-FOC (who) drank the vodka.

(King 1995: 80)

	0	1	2	3	4
TXT	Это	Борис	выпил	водку	.
[POS]	cleft	N	V	N	
[GLOSS]	it	Boris	drank	vodka	
[TOPIC]					
[FOCUS]		F			
[CONTRAST]		CF			
[PROSODY]		*			

- (II) (Context: 'Today I met Lida in the park.')
- Лида работает журналисткой в газете «Ленинградская Правда».
 Lida works journalist-INST in newspaper Leningradskaia Pravda.
 'Lida works as a journalist for newspaper *Pravda*.'
 [AB-TOPIC Lida] works as a journalist for newspaper *Pravda*.

	0	1	2	3	4	5	6	7	8	9
TXT	Лида	работает	журналисткой	в	газете	"	Ленинградская	правда	"	.
[POS]	N	V	N	P	N		Adj	N		
[GLOSS]	Lida	works	journalist-INST	in	newspaper		Leningrad	truth		
[TOPIC]	AB-T									
[FOCUS]										
[CONTRAST]										
[PROSODY]			*		*			*		

- (III) Завтра мы поедem в Санкт-Петербург.
 Tomorrow 1PL.NOM going in Saint Petersburg
 'Tomorrow we are going to Saint Petersburg.'
 [FS-TOPIC Tomorrow] we are going [SF to Saint Petersburg].

	0	1	2	3	4	5
[TXT]	Завтра	мы	поедем	в	Санкт-Петербург	.
[POS]	Adv	N	V	P	N	
[GLOSS]	tomorrow	1PL.NOM	go.1PL.FUT	to	Saint-Petersburg	
[TOPIC]	FS-T					
[FOCUS]				SF		
[CONTRAST]						
[PROSODY]					*	

- (IV). U teбja-**TO** sovok (...a что u медведja в лапe?)
 At you-**TO** scoop (...but what at bear in paw)
 "YOU(-**TO**) have a SCOOP (...but what does the BEAR have in his paw?)"

	0	1	2	3	4
[TXT]	У	тебя	-то	совок	...
[POS]	P	N	-TO	N	...
[GLOSS]	At	2SG.GEN-TO	CLIT	scoop	
[TOPIC]		AB-T			
[FOCUS]				F	
[CONTRAST]		AB-CT		CF	
[PROSODY]		*		*	

- (V). Вы **же** убирать будете / а не учителя.
 You **ŽE** clean will-2PL.FUT / but not teachers
 [CF You] **ŽE** [SF will be cleaning] /
 You will be the ones cleaning it, and not the teachers.'

	0	1	2	3	4
TXT	Вы	же	убирать	будете	...
[POS]	N	CLIT	V	V	
[GLOSS]	2PL.NOM	ŽE	clean_up-INF	will	
[TOPIC]					
[FOCUS]	F				
[CONTRAST]	CF				
[PROSODY]	*				

- (VI). Виктор **ли** это сделал?
 [CF Victor] **LI** this done
 'Was it Victor who did this?'

	0	1	2	3	4
TXT	Виктор	ли	это	сделал	?
[POS]	N	CLIT	DET	V	
[GLOSS]	Viktor	LI	this-ACC	did.MASC	
[TOPIC]					
[FOCUS]	F				
[CONTRAST]	CF				
[PROSODY]	*				

Appendix B: Diagnostics/analysis of information structure components

(I) Sentence:

On uedet [segodnia že].
He will leave today že
'He will leave today-FOC.'

(King 1995: 80)

Diagnostics:

1) Wh-question test (Gryllia 2008)

Question:

Kogda on uedet?
When 3SG.MASC.NOM leave?
When is he going to leave?

Answer 1:

*On uedet [segodnia že].
He will leave today že
'He will leave today-FOC.'

Answer 2:

On uedet segodnia.
He will leave today
'He will leave today.'

The constituent *segodnia že* is incompatible with the *wh*-question, thus passing the *wh*-question test to identify it as **contrastive focus**.

2) Correction test (Gryllia 2008):

Statement:

On uedet zavtra.
3SG.MASC.NOM leave-FUT tomorrow
'He will leave tomorrow.'

Correction 1:

?On uedet [segodnia že].
He will leave today že
'He will leave today-FOC.'

Correction 2:

On uedet segodnia.
He will leave today
'He will leave today.'

The correcting sentence 1 containing *že* is marginally acceptable, while the correcting sentence 2 without *že* (but retaining contrastive interpretation of *segodnia*) is fully grammatical. This suggests that the constituent *segodnia že* passes the correction test that identifies it as **contrastive focus**.

3) Choice test (Gryllia 2008):

Question:

On uedet zavtra ili segodnia?
3SG.MASC.NOM leave-FUT tomorrow or today?
'Will he leave tomorrow or today?'

Answer 1:

?On uedet [segodnia že].
He will leave today že
'He will leave today-FOC.'

Answer 2:

On uedet segodnia.
He will leave today
'He will leave today.'

It is not completely clear what makes the sentences above with **ŽE** marginally acceptable and the same sentences without **ŽE** acceptable. When provided with the above choices for the answers and asked to choose the most plausible answer, three native speakers were hesitant to choose the answers containing the clitic. Instead they chose the sentences without the clitic. One of the native speakers mentioned that the reason for the choice was that the sentences with **ŽE** would be 'too much'. The informality/aggressiveness of this clitic seems to be a possible culprit for the marginal applicability of the contrastiveness tests above to the sentences containing emphatic/informal clitics.

(II) **Sentence:**

Торговать же они в Москву ездили...
[CT Trade-INF] **ŽE** 3PL.NOM to Moscow went.PL.PST...
'As to trading, they went to Moscow....'

Diagnostics:

- 1) ‘Tell-me-about’ test (Choi 1999) to identify **topic** was slightly modified to be applied to this sentence, i.e. verb ‘to trade’ was replaced with noun ‘trading’:

Question: Расскажи мне про их торговлю.
‘Tell me about their trade (trading).’

Answer: Торговать же они в Москву ездили...
‘To trade they went to Moscow.’

- 2) Implicit sub-question test (Gryllia 2008) to identify **contrastive topic**:

It should be noted that for this (as for many other sentences with clitics) the answer (a) containing *ŽE* is marginally acceptable/ungrammatical. However, the same sentence without *ŽE* in (b) (if the contrastive prosody and interpretation of the verb is retained) is grammatical.

Speaker 1:

Where did they go with their goods?

Speaker 2:

1. (a) ?/*Торговать **же** они в Москву ездили...
Trade-INF *ŽE* 3PL.NOM to Moscow went
‘As to trading, they went to Moscow...’

- (b) Торговать они в Москву ездили...
Trade-INF 3PL.NOM to Moscow went
‘As to trading, they went to Moscow...’

1. И... б-бывали закупки у то... у того же Мамонтова в
Абрамцево.
And...w-were purchases at th... at that *ŽE* Mamontov-SG.GEN in
Abramtsevo
‘And there were purchases from the same Mamontov in Abramtsevo.’
же они в Москву ездили...

Both tests above, the ‘tell-me-about’ test and the implicit sub-question test suggest that within the given context *торговать* /‘trade’ is a **contrastive focus**.

(III) **Sentence:**

Вы **же** убираться будете / а не учителя.
You **ŽE** clean will-2PL.FUT / but not teachers
[_{CF} You] **ŽE** [_{SF} will be cleaning] /
You will be the ones cleaning it, and not the teachers.'

Diagnostics:

- Passes *wh*-question test (Gryllia 2008)

(Q) Кто будет убирать(ся) в туалете?
Who is going to be cleaning the toilet?

(A) *Вы **же** убираться будете / а не учителя.
You **ŽE** clean will-2PL.FUT / but not teachers

- Passes (marginally) correction test (Gryllia 2008)

Speaker 1:

Учителя будут убирать(ся) в туалете.
Teachers are going to be cleaning the toilet.

Speaker 2:

?Вы **же** убираться будете / а не учителя.
You **ŽE** clean will-2PL.FUT / but not teachers

- Fails 'tell-me-about' test (Choi 1999)

Speaker 1: Расскажите мне о себе.
Tell 1SG-DAT about self-PREP

Speaker 2: *Вы **же** убираться будете / а не учителя.
You **ŽE** clean will-2PL.FUT / but not teachers

- Fails **implicit sub-question test** (Gryllia 2001)

(Q) Вы (ученики и учителя) собираетесь что делать?
You (students and teachers) are going to be doing what?

(A) *Вы **же** убираться будете / а не учителя.
You **ŽE** clean will-2PL.FUT / but not teachers