UiT The Arctic University of Norway

Faculty of Humanities, Social Sciences and Education

Picking apart Russian particles

An empirical study on the meaning and use of ***že*** and *ved'*

James David McDonald

Master's thesis in Russian language, RUS-3920, May 2021

Acknowledgements

First, I would like to give a huge thank you to my supervisor Laura Janda for all the time she has given me throughout this thesis. Laura has been there to motivate me, challenge me and encourage me, and I am so grateful for all her help and guidance. *Děkuji moc, Lauro!*

I would also like to thank the members of the CLEAR research group for always being so enthusiastic about my thesis, and always being willing to answer questions related to my topic. I want to particularly thank Svetlana Sokolova, Elena Bjørgve, Anna Endresen, Daria Kosheleva and Elmira Zhamaletdinova, for allowing me to pick your brains on Russian particles, and always cheering me on. I would also like to thank Aleksandra Bogoiavlenskaia for her native speaker intuition and assisting me with the annotation of my data. So many people have offered comments and guided me in my work, for which I am very grateful.

I would like to thank my mam, Lorraine, for always knowing what to say when I doubted myself. I couldn't have gotten through this thesis without the support of my best friend Charlotte, my Excel guru, who was always there to offer her IT skills. Finally, a huge thank you to my partner Håkon for his kindness, patience and encouragement throughout this thesis. You have always been there to motivate me when I needed it most.

Tromsø, 6th May 2021
James David McDonald

Transliterations

All examples in this thesis will be transliterated using the International Scholarly System. The names listed in the acknowledgements are transliterated according to the way the individual writes their name.

Table of contents

1 Introduction

1.1 Introducing my thesis

This thesis explores the meaning and usage of the Russian particle *že*, as well as the synonymy of *že* and the Russian particle *ved'*. According to the Cambridge English Dictionary, a particle is "a word or a part of a word that has a grammatical purpose but often has little or no meaning" (Cambridge English Dictionary, 2021). As a learner of Russian, I have often struggled to understand and effectively use Russian particles. It is for this reason that I chose this topic for this thesis: to better understand the meaning and usage of *že* and *ved'*.

In Russian it is grammatically correct to say *konečno* "of course", as well as *konečno že* "of course". In this case, what function does *že* have, and what meaning does it carry? Does *že* in *konečno že* have the same function as *že* in the constructions *v to že vremja* "at the same time", *k tomu že* "in addition to" or *srazu že* "immediately", for example? These examples demonstrate the complexity in translating *že* into English, and this motivated me to explore whether it is possible to clarify the meaning(s) of *že*.

In Endresen et al.'s 2016 investigation of Russian particles it is claimed that *že* can be replaced with *ved'* in the function of an adverbial conjunction (Endresen et al., 2016, p.123). This led me to wonder, whether it is possible to replace *že* with *ved'* in other cases, such as *konečno ved', v to ved' vremja, k tomu ved' or srazu ved'*?

My thesis consists of four main sections: a theoretical discussion and three empirical investigations. In the remainder of this chapter I will highlight the main points of the theoretical discussion, as well as present the research questions and findings of the three investigations.

In chapter two I present how *že* has been dealt with in the literature. Here I focus on how *že* has previously been classified by Padučeva (1988, as cited in Parrott, 1997), Parrott (1997), McCoy (2003a, 2003b) and Hagstrom and McCoy (2002). I then discuss the nebulous term "particle", highlighting Zwicky's proposal to eradicate this term (1985). Finally, I examine

Endresen et al.'s 2016 investigation of retagging Russian particles in the Russian National Corpus, where I concentrate on their findings for *že*.

In chapter three I use cognitive linguistic methods to investigate the meaning of *že* and how it is translated to English. I use selected texts from the RuN parallel corpus and identify submeaning categories for the different meanings and uses of *že*. Based on the English translations given, I seek to generalise how *že* can be translated to English. I then propose a radial category for *že* and present how this network interacts. Using the same method I carry out the same investigation on *ved'*. Based on my two proposals for radial categories I show how three of my proposed submeanings for *že* and *ved'* overlap, which supports the idea of a level of synonymy between *že* and *ved'*.

What factors may influence the replaceability of *že* with *ved'*? This is the main focus of chapter four. Using a new database from the Russian National Corpus (RNC) I investigate factors such as the meaning of *že* (based on my own findings from chapter three), as well as the way *že* is tagged in the RNC and the part of speech to the left of *že*. My hypothesis is that the meaning of *že* plays the most important role when deciding whether *že* can be replaced with *ved'*. Using statistical methods such as Chi-square, Cramér's V, Fisher test, Logistic Regression Model and a cTree analysis, I show that the part of speech to the left of *že* may in fact be a more influential factor than the meaning of *že* in the replaceability of *že* with *ved'*. I argue, however, that the meaning of *že* does still play a role, and I therefore do not discard it completely as an influential factor.

Chapter five aims to investigate the replaceability of *že* with *ved'* further. First, I present the development and findings of a pilot experiment, which led to the main questionnaire of this chapter. I carried out a questionnaire on the addition of *že* or/and *ved'* to a specific clause in a sentence. Participants were given a sentence where *že* had been removed, and were asked two questions: whether they could add *že* to the highlighted clause in the sentence, and whether they could add *ved'* to the highlighted clause in the sentence. The sentences in the questionnaire are based on combinations of the most frequent parts of speech to the left of *že* and the most frequent meanings of *že* from my findings in chapter four. I hypothesise that when the meaning of *že* overlaps with the same meaning of *ved'*, participants agree that both

že and *ved'* can be added to the clause. My findings, however, show that this does not appear to be the case. Generally the participants of the questionnaire did not favour adding *ved'* to the highlighted clause, but levels of synonymy in the data can be seen. A cTree analysis also shows that the meaning of *že* does not play as an important role as the part of speech to the left of *že*. I conclude that this investigation does not give any definitive results, but does indicate that there is a relationship between *že* and *ved'*.

Concluding this thesis I reiterate how these three empirical investigations shed light on a notoriously difficult topic of Russian grammar for learners of Russian. This thesis demonstrates that cognitive linguistic and statistical methods can be used to gain a better understanding of the meaning of Russian particles such as *že* and *ved'*, in addition to the relationship between these particles. I show that it is possible to categorise the different submeanings of *že* and *ved'*, which may help learners of Russian in understanding and using these two lexemes. I also demonstrate that statistical methods are useful tools in linguistics to investigate the relationship between lexemes such as *že* and *ved'*. I hope this thesis brings valuable observations that can be used as a springboard to further investigate Russian particles.

1.2 Research questions

To summarise, the research questions of this thesis are:

- What is the meaning and usage of *že* and *ved',* and how do they relate?
- What factors may influence the replaceability of *že* with *ved'*?
- What is the relationship between the factors that may influence the replaceability of *že* with *ved'*?

1.3 TROLLing repository

All of the data I have collected for this thesis, as well as the Rscripts made in chapter four and five, have been made open access at the TROLLing repository. To access my data, use the following link: https://doi.org/10.18710/G9S9OW.

2 *Že in the literature*

This thesis is motivated by a personal endeavour to try to understand the Russian particle *že*. I began by investigating some of the literature already available on the topic to gain a better insight. Padučeva (1988, as cited in Parrott, 1997) made observations of the different ways in which *že* is used. In her doctoral dissertation, Parrott (1997) challenges Padučeva's classification of *že,* and further investigates the usage *že* as a discourse particle. McCoy (2003a, 2003b) and Hagstrom and McCoy (2002) move away from a descriptive analysis of *že,* as presented by Parrott, and use the theory of "kontrast" (Vallduví and Vilkuna, 1998) to link different meanings of *že*, *ved'* and *-to*. In the first section of this chapter, I will present the different classifications and functions of *že* referencing Padučeva (1988, as cited in Parrott, 1997), Parrott (1997), McCoy (2003a, 2003b) and Hagstrom and McCoy (2002), illustrated with examples.

In all the aforementioned literature, *že* is categorised as a particle and a clitic. Zwicky (1985) discusses the properties of words that are referred to as clitics and particles, and ultimately suggests eliminating "particle" as a part of speech. I will provide a summary of this and follow with a disccusion on an investigation by Endresen et al. (2016), who use corpus data to further support Zwicky's stance on particles by creating a new tagging scheme. Using the current tagging scheme in the Russian National Corpus (henceforth RNC), Endresen et al. develop a more complex tagging scheme eradicating the use of "particle" and show how their complex tagging scheme is just as effective. This investigation will be presented with particular focus on Endresen et al.'s findings on *že*.. I will summarise this chapter by explaining how this theoretical overview and discussion paved the way for this thesis.

The main topic of this thesis is the Russian particle *že*. In this thesis, however, I will also investigate the synonymity between *že* and *ved',* and therefore, *ved'* will be briefly mentioned in the current chapter, where appropriate.

All reference to Padučeva is made through Parrott's doctoral dissertation as I was unable to obtain a copy of Padučeva's original work, which is written in French. Figure 2.1 is taken from Parrott's (1997, p.17) doctoral dissertation and shows the different ways in which

Padučeva (1988, as cited in Parrott, 1997) classifies *že* on the left. The column "Proposed classification" shows Parrott's own reclassification of the usages of *že*, to better suit her view on the distribution of *že*. I will explain these usages with examples.

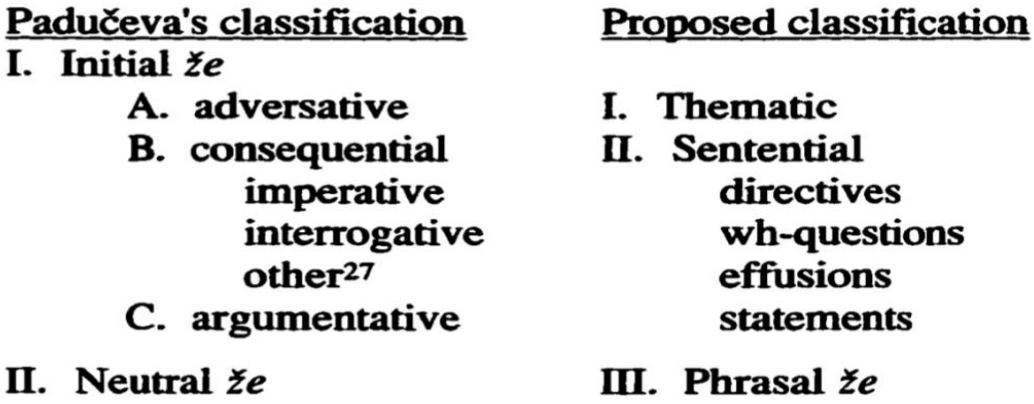

Figure 2.1: A summary of Padučeva and Parrott's classification of the usages of že, taken from Parrott (1997, p. 17)

According to Parrott, Padučeva bases her classification on two main types of "prosodically different variants of the particle" (Parrott, 1997, p.11). Padučeva identifies two main categories of *že*: "initial *že*" and "neutral *že*".

Parrott does not consider Padučeva's "neutral *že*" in her study but reclassifies it as "phrasal *že*" as it contributes to "word- or phrase-level semantics" (Parrott, 1997, p.13). Parrott mentions the two distinct uses of Padučeva's "neutral *že*", shown in examples (1) & (2), taken from Padučeva (1988, as cited in Parrott, 1997, p.12):

(1) **Neutral *že*: identity.** *Za stolom sidela ta že ženščina.*
"At the table sat the same woman"

(2) **Neutral *že*: promptitude.** *On prišël v pervyj že večer*
"He came the very first evening"

Padučeva's classification "initial *že*" describes instances where *že* is positioned near the beginning of a clause. "Initial *že*" is further categorised by three different uses: adversative, consequential and argumentative.

Parrott refers to Padučeva's adversative usage as "thematic *že*" "since its host is always the theme (broadly speaking)" (Parrott, 1997, p.14). Furthermore, "thematic *že*" is used where the scope is limited to the host. Example (3) shows this usage, taken from Padučeva (1988, as cited in Parrott, 1997, p.12):

(3) **Initial *že*: adversative.** *Moeju budet ved Ljudmila, Ruslan že v grobu obrečen*
"Ludmila will be mine forever, but as for Ruslan, he is doomed to the grave"

Padučeva's classification of the consequential usage of *že* and argumentative usage of *že* are reclassified by Parrott as "sentential *že*", because "its scope is the entire proposition" (Parrott, 1997, p.14). Padučeva's consequential usage of *že* is further divided into submeanings. Examples (4-7), taken from Padučeva (1988, as cited in Parrott, 1997, p.12), show Padučeva's consequential and argumentative usage of *že:*

(4) **Initial *že*: consequential: imperative.** *Tak podi že, popljaši*
"So go ahead, dance"

(5) **Initial *že*: consequential: interrogative.** *Čto že mne v takom slučae delat'?*
"But what am I to do in such a case?"

(6) **Initial *že*: consequential: other.** *Zdorovo že ty uvleksja (esli ničego ne slyšal)!*
"You must have really been carried away (if you didn't hear anything)!"

(7) **Initial *že*: argumentative.** *On že genij*
"But he's a genius"

Parrott justifies the need for a reclassification and renaming of Padučeva's classification of the use of *že* because "the names that she [Padučeva] gives the different usages are at times misleading and easily confused" (Parrott, 1997, p.16). Whilst Parrott claims that this recategorisation presents less confusion in understanding the usages of *že*, I find both

descriptions inadequate to understand and define the meaning of *že*. Both Padučeva and Parrott address different contexts in which *že* is used (interrogative contexts, argumentative contexts), but their descriptions are not satisfactory for interpreting all examples of *že*.

McCoy further investigates the meaning and function of *že*, and presents a series of articles aimed at unifying the meaning and structure of Russian particles, with focus on *-to, že* and *ved'* (Hagstrom and McCoy, 2002, McCoy 2003a, 2003b).. Whilst McCoy focuses on the cognitive status of information marked by *že*, rather than the translation of *že*, she briefly mentions some of the ways in which *že* is rendered in English (2003a, p.126). Three main contexts of *že* that McCoy focuses are shown in examples (8-10), taken from McCoy (2003a, p.124):

I. **A declarative.** This occurrence of *že* can be translated using the contrastive "but" or similar prosodic means. This occurrence of *že* is similar to Padučeva's argumentative usage of "initial *že*", or Parrott's statement usage of "sentential *že*".

(8) *My že s toboj sami videli, čto slomalos'.*
"(But) you and I together saw that it got broken"

II. ***Že* in a rhetorical yes/no question**. Here, *že* is best translated into English as a tag question. This usage is not specified in Padučeva's classification.

(9) *Èto že ne slomalos'?*
"This isn't really broken, is it?"

III. **A wh-question.** This type of question is also rhetorical and can be translated with a wh-question such as "why in the world…?". This occurrence of *že* is similar to Padučeva's consequential interrogative usage of "initial *že*", or Parrott's wh-question usage of "sentential *že*".

(10) *Kak že ne slomalos'?*
"Why in the world isn't it broken?"

Summarising the meaning of *že* McCoy claims that

> "utterances with *že* address some contradiction that the speaker believes the addressee holds on to and are aimed at solving this contradiction by "correcting" the addressee" (McCoy 2003a, p.125).

McCoy summarises previous proposals in the literature for describing the function of *že*. She claims, however, that these descriptive approaches only deal with one specific aspect of the particle, and are therefore limited (McCoy 2003a, p.126). Building on this, McCoy attempts to unify the meaning of *že* by using Vallduví and Vilkuna's theory of "kontrast" (Vallduví and Vilkuna, 1998). McCoy's interpretation is that the notion of kontrast describes an instance where a set of alternatives are generated (McCoy 2003b, p.1). Using *že* as a marker of kontrast, McCoy proposes the following kontrast set for *že*:

M={X,X'}, where X=¬X'
(X is true if and only if X' is false)

(McCoy 2003a, p.127)

This model is demonstrated in bold in the following example of *že* as a declarative:

(11) (A conversation between Varja and her mother about a fly on the windowsill)
Varja: *Ubit', ubit', ee*
Kill, kill it!
Mother: ***Ona (že) uže ubita (že)***
It (*že*) is already killed (*že*)

(Hagstrom and McCoy, 2002, p.4)

In this case *že* is a marker of kontrast because *že* contains a set of alternatives that have mutually exclusive propositions: that the fly is killed, and that the fly is not killed. McCoy (2003b) also notes the cognitive status of information that this use of *že* marks. In example (11) the presupposition of Varja's utterance is that the fly is alive. *Že* as a marker of kontrast indicates that the mother claims that Varja should know that the fly is no longer alive. By using *že* in this sentence, the mother is informing Varja that the proposition "the fly is not

killed" is no longer possible. Hagstrom and McCoy (2002) and McCoy (2003a, 2003b) further demonstrate *že* as a marker of kontrast by presenting how this framework works for *že* as a yes/no question and a wh-question.

McCoy (2003b) does not only focus on *že* as a marker of kontrast, but also the Russian particles *-to* and *ved'* (due to the scope of this thesis, *-to* will not be discussed). Whilst McCoy notes that *že* and *ved'* are similar in that they mark a set of propositions that differ from each other (2003b), she also states that *že* and *ved'* are different in two ways. First, unlike *že,* the set of alternatives for *ved'* is restricted to propositions where an allegation or proclamation is made based on an opinion (McCoy 2003b, p.11). Secondly, the members of the set for *ved'* are supplementary, unlike for *že* where they are mutually exclusive (McCoy 2003b).

McCoy presents her articles with the use of corpus data. However, corpora have since developed vastly and grown significantly in size. To my knowledge, the most recent studies focusing on *že* were carried out by McCoy (2003a, 2003b), where the focus was on the function of the particle *že*. A search in the literature found another study carried out by McCoy-Rusanova (2008) focusing on the combinations of *-to, ved'* and *že*, however I was unable to access this article. Other studies include Endresen et al. (2016), where the current tagging system of nine Russian particles (two of which are and *že* and *ved'*) in the Russian National Corpus (henceforth RNC) is challenged by proposing a new tagging system, and McCoy-Rusanova (2017), where the function of multiple discourse particles (*-to, že* and *ved'* combined) is the focal point.

In their research Parrott and McCoy categorise *že* as a particle in terms of part of speech. As a second-language learner, I find particles in Russian difficult to understand, use and translate. In grammars *že* is predominantly classified as a 'particle' (Dunn & Khairov, 2009, p. 215), a term used to describe words that do not fall into well-established categories such as nouns, verbs and adjectives. Other words in Russian that are categorised as particles, according to Dunn & Khairov, include *vot, by, li, ved',* and *-to*. Deriving from the Greek *ἐγκλιτικός enklitikós* "leaning", *že* is further described as a clitic, meaning that it is phonologically dependent upon the preceding word. As a result, *že* can never be the initial word in a sentence. This is shown in examples (12-13). Example (12) is grammatical because *že* is not sentence

initial, it ‘leans on’ the pronoun *on*, whereas example (13) is not grammatical as *že* does not have anything to lean on.

(12) *On že nikogda ne govoril ob ètom*

He *že* never not spoke about this

He never spoke about this

(13) **Že on nikogda ne govoril ob ètom*

The part of speech categorisation of *že* as a particle has been a debated topic for many years. In his 1985 article on clitics and particles, Zwicky raises the issue of distinguishing between clitics and independent words and suggests a series of tests to resolve this distinction. Zwicky states that these tests “point to characteristic symptoms of a linguistic state of affairs, not to invariant concomitants of it” (Zwicky, 1985, p. 285). There are numerous tests, categorised as phonological tests, an accentual test and syntactic tests. Zwicky demonstrates a grammatical hierarchy of units: affix > clitic > word > phrase > clause. He argues that there is no reason to add another unit, namely particle, as “languages contain no ‘particles’, but only words belonging to syntactic categories, clitics, and [...] affixes” (Zwicky, 1985, p. 294).

Zwicky discusses the properties of words that are labelled as particles. Although he primarily focuses on English, Zwicky points out that many different collections of words are assigned to the category particle in other languages, e.g. honorifics, case-markers and markers of emphasis, to name but a few. These different collections of words further highlight the lack of clarity with particle as a part of speech. Furthermore, particles are “…words left over when all the others have been assigned to syntactic categories” (Zwicky, 1985, p. 292).

Zwicky raises the idea that particles are acategorial, whereby they do not belong to any syntactic category but are rather introduced by syntactic rules. For example, the English word ‘*only*’ can appear before a determiner and noun phrase, a verb and noun and prepositional phrase, or before a preposition and a noun phrase. Zwicky, however, rejects the notion of particles being acategorial, because “every word (in every language) belongs to one of the syntactic categories provided by (universal) grammatical theory.” (Zwicky, 1985, p. 294). By rejecting the idea of acategorial words, Zwicky assumes that there exists “an elaborated

theory of syntactic categories" (Zwicky, 1985, p. 294) consisting of subcategories within already established categories.

Zwicky's recommendation to eliminate the term "particle" as well as his assumption of the subcategorisation of syntactic categories motivate of Endresen et al.'s 2016 study of particles in the Russian language. Supporting Zwicky's stance on the topic, Endresen et al. use advancements in corpus linguistics to further support their findings. An interesting observation from their research shows that in the RNC 'particle' as a part of speech is widely used as a tag to categorise words, "…accounting for approximately 4.5% of all words in a corpus …" (Endresen et al., 2016, p.104). Despite this, Endresen et al. analyse a sample of the RNC to propose a reclassification of nine Russian words commonly called particles, one of which is *že*.

In their article Endresen et al. raise both theoretical and practical problems related to particles. Firstly, from a theoretical perspective "particle" is commonly used as a part-of-speech category and Endresen et al. look at what a part of speech is and highlight the different ways to classify parts of speech. These include formal characteristics (observing morphological classes), a distributional approach (that prepositions appear before nouns) and a semantic approach (nouns signify entities) (Endresen et al., 2016, p.105).

Endresen et al. suggest that linguists combine strategies when identifying parts of speech (Endresen et al., 2016, p.105). Furthermore, they state that Croft proposes a "conceptual space for parts of speech" (2001, cited in Endresen et al., 2016, p.105), and highlight that the details of this conceptual space differ from language to language. By labelling it as a 'conceptual space', Croft allows for the possibility that different categories (nouns, verbs, adjective) overlap. When it comes to defining categories such as 'verb' Endresen et al. highlight that part-of-speech categories contain prototypical members and non-prototypical members. An example of a prototypical characteristic of a verb in Russian includes the transitive construction, and a non-prototypical member is a participle. As Endresen et al. demonstrate, non-prototypical members can overlap with other categories. This is shown by their example of *vydajuščijsja*, which is categorised as both a participle and as an adjective (Endresen et al., 2016, p.106). Zwicky (1985) claims that a plethora of types of words have been categorised as

particles, which makes it impossible to determine a prototype that has common properties with every type of particle. This further supports Endresen et al.'s attempt at reclassifying particles.

Secondly, Endresen et al. present practical problems related to particles. As previously stated, particles are well attested in the Russian language, according to corpus data. The RNC is regarded as a reliable source of data, representing both spoken and written Russian of a wide variety of genres. It is possible to obtain grammatical information about each entry in the corpus (known as a token), including the token's part of speech. The laborious task of tagging each token is carried out automatically by trained computer programs, which according to Endresen et al. is not always successful in Russian. Like in the example of *vydajuščijsja* particles can be ambiguous in meaning, overlapping with other parts of speech such as adverbs, conjunctions, and interjections (Endresen et al., 2016, p. 110). Endresen et al. present an example from the RNC where the lexeme *ved'* has been tagged in two different ways (2016, p.112). In example (14) *ved'* is tagged as a particle and as a conjunction in example (15), despite the fact that the word *ved'* serves functions that are both syntactically and semantically identical.

(14) *Ved' vy ne znaete, možet, on na vas takoe nagovoril…*
But you don't know, maybe he has made up a story about you…
(15) *Ved' vy ne znaete goroda…*
But you don't know the city…

This ambiguity and inconsistency can cause problems in identifying Russian particles, in addition to the lack of consistency amongst scholars as to how many particles there are in Russian (Endresen et al., 2016, p.108). The lack of clarity this topic further supports the need for Endresen et al.'s work.

In the following section I will present Endresen et al.'s two experiments, focusing mainly on the findings most relevant to this thesis, namely the findings for *že.*

Endresen et al. found that in the RNC many of the lexemes tagged as particles were tagged as two, three and in some cases four categories for parts of speech. For the experiment Endresen et al. chose nine particles of high frequency that were tagged in the RNC as belonging to two categories. In addition to "particle" these other categories were: 'adverb', 'conjunction' or 'predicative'. *Že* was the second most frequent particle with 21,350 entries and was tagged in the RNC as both a particle and as a conjunction (Endresen et al., 2016, p.113).

The first experiment was carried out to test how well the current RNC tagging system works. 100 examples were randomly extracted for each of nine particles to create a database representative of the dataset. This database was then used to make a Hidden Markov Model (HMM), a statistical model used in part-of-speech tagging (Endresen et al., 2016, p.114). The tagging distribution of the 100 examples across the categories is given, and in the case of *že* ninety-four examples were tagged in the RNC as a particle and six tagged as a conjunction. The database was put to the test by dividing it into ten sections and then carrying out a ten-fold cross-validation, using ninety sentences as the training set and ten sentences as the test set (Endresen et al., 2016, p.114). The aim of this was to see how Endresen et al.'s own HMM tagger fared against the distribution of original tags. In the case of six of the nine lexemes, including *že*, the tagger showed worse results (-5% in accuracy for *že*). This experiment further highlights the unreliability of the tagging system in the RNC and justifies the need for improvements of this issue.

The second experiment is based on Endresen et al.'s own proposed scheme for tagging Russian particles. Using the same 100 randomly sampled sentences for each of the nine high-frequency particles in experiment 1, Endresen et al. present their own, more complex system. Whilst in experiment 1 *že* was tagged as a particle and as a conjunction in the RNC, Endresen et al. suggest reclassifying *že* with the categories 'adverbial conjunction' (13 examples), 'coordinating conjunction' (6 examples) and 'emphasiser' (81 examples). Endresen et al. claim that the most common use of *že* is as an emphasiser (2016, p.116).

In their analysis of reclassifying *že*, Endresen et al. claim that there are factors, not absolute rules, that can assist in determining whether the *že* is an adverbial conjunction, coordinating conjunction or emphasiser. These factors are comprised of whether *že* is preposed or

postposed, and syntactically optional or obligatory. Furthermore, for the two conjunction types another distinguishing factor is the replaceability of *že* with semantically equivalent conjunctions.

As *že* is a clitic it is dependent on another stressed lexeme and can be positioned both before (preposed) and after (postposed) the stressed lexeme. As previously stated *že* cannot, however, appear preposed if the stressed lexeme to which it is dependent is in initial position. In cases where *že* is postposed the part of speech is often an emphasiser or a coordinating conjunction and mostly an adverbial conjunction when *že* is preposed.

Syntactic optionality refers to instances where *že* can be removed without changing the syntax of a sentence and is true for emphasiser and adverbial conjunctions. Endresen et al. state that *že* as a coordinating conjunction "…is obligatory for creating an explicit contrast between syntactic constituents" (2016, p.116).

As a coordinating conjunction *že* can be replaced with the conjunction *a* without affecting the semantics of the sentence, although syntactic changes occur. As an adverbial conjunction *že* can be replaced with *ved'*, although the register of the utterance's politeness is altered. Table 2.1 is a visual summary of Endresen et al.'s reclassification of *že*:

	EMPHASISER	**ADVERBIAL CONJUNCTION**	**COORDINATING CONJUNCTION**
Preposed/postposed	Postposed	Preposed	Postposed
Syntactic optionality/Obligatory	Syntactic optionality	Syntactic optionality	Obligatory
Replaceability with equivalent conjunction	-	*ved'*	*a*

Table 2.1: A summary of Endresen et al.'s reclassification of že

For experiment 2 Endresen et al. used the same database from experiment 1, but instead trained the HMM tagger on their own scheme for classifying the nine particles. As there are more tags in Endresen et al.'s scheme, it should be more difficult for the HMM tagger to

perform well in experiment 2. Whereas all but one particle had two tags in experiment 1, almost all (six out of nine) particles were assigned to three, four or five distinct tags in experiment 2. The baseline for each particle in experiment 1 is the highest number of the original tags in the RNC. For example, the baseline for *že* was ninety-four, as out of the 100 example sentences randomly selected six were tagged as a conjunction, and ninety-four as a particle. The baseline for each particle in experiment 2 is the highest number of the new tags proposed by Endresen et al. For *že* the new baseline was eighty-one (coordinating conjunction: 6, adverbial conjunction: 13, emphasiser: 81) (2016, p.122).

The results of experiment 2 show that whilst there were no significant differences, in total there was a gain over the baseline. This could be interpreted that the HMM tagger was not negatively affected by the new tagging system that was much more complicated. In the case of *že* the result was actually a loss over the baseline of 5%, which was the same for experiment 1 (Endresen et al., 2016, p.129).

Endresen et al.'s experiments show that Zwicky's claim that the term 'particle' should be eliminated as a part of speech is justified. Given that the automatic tagger overcame the challenge of tagging a more complex tagging scheme, it would be beneficial to linguists and learners of Russian to replace the part of speech 'particle' with a more enriched scheme, such as the one suggested by Endresen et al..

This literature has given me the foundation for this thesis and was vital in developing my research questions. Parrott and McCoy present some of the functions of *že*, but an in-depth discussion into the meaning and translation of *že* into English is lacking. In addition, the similarities and differences between *že* and *ved'* in terms of their function are only briefly mentioned. I decided to therefore investigate the meaning of *že* and *ved'* and how they can be rendered in English, to find whether there are factors that influence the synonymy of *že* and *ved'*. Inspired by Endresen et al. I decided to carry out this thesis using empirical methods, with the aim of better understanding a notoriously difficult topic of Russian grammar.

Given the arguments presented by both Zwicky and Endresen et al., it is not beneficial to refer to *že* and other words classed as particles as "particle". I will henceforth use the neutral term "lexeme" in this thesis.

3 A radial category for *že and ved'*

3.1 Introduction

This chapter deals with understanding the lexemes *že* and *ved'* as well as how they can be translated into English. The ultimate aim of this study is to facilitate linguists' understanding of a notoriously complicated area of Russian grammar. For English speakers, the lexemes *že* and *ved'* can be difficult to translate. In some instances, primarily when *že* is used to emphasise and stress a statement or opinion in Russian, the meaning in English is conveyed via intonation when speaking. This can also be the case for *ved'*, which assumes that *že* and *ved'* can be synonymous. However, as presented in the literature review in chapter two, *že* and *ved'* can differ greatly in their meaning, demonstrating that they are not always synonymous.

In this investigation I explore how *že* and *ved'* are translated to English. Based on data collected from the RuN parallel corpus of Russian texts and their English translations, I decipher submeanings for *že* and *ved'*. I also propose two radial categories: one for *že* and one for *ved'*. The goal of deciphering these radial categories is to attempt to show:

1) the ways in which *že* and *ved'* are rendered in English and
2) that these meanings, whilst they may seem very different to each other, may in fact be interconnected via a network and share common properties.

In section 3.2 I present a general introduction to understanding radial categories and prototypes. I go on to explain how I collected and analysed my data in section 3.3, with a detailed description of the proposed submeanings for my radial category of *že*, using examples from my data. I present my radial category in section 3.3.13 and explain how I built the network. In section 3.4 I present my data collection and description of submeanings for *ved'.* My proposal for a radial category for *ved'* is presented in section 3.4.11. In section 3.5 I present further analysis of my findings. I show how my radial categories for *že* and *ved'* relate in section 3.5.1. Using the data collected for both radial categories, I present how the distribution of each radial category differs by looking at two texts that appear in both the dataset for *že* and for *ved'* in sections 3.5.3 and 3.5.4. Finally, in section 3.5.5 I offer suggestions for using the findings from this investigation in language teaching.

3.2 Radial categories and prototypes

As human beings we categorise everything around us. When we come across a new entity and learn of its properties, we connect it with those entities that we are familiar with and share similar properties. These entities can be both physical objects such as a ball or a tree, as well as abstract concepts such as love and hate. Categorisation is primarily innate, and we seem to only be aware of it in instances where categorisation is problematic, for example where an entity has unfamiliar properties. This general view on categories, known as the Classical Theory, was once regarded as definitional truth (Lakoff, 1987, p.6). This perspective has since changed, with Eleonor Rosch as the pioneer in developing the "prototype theory" (Lakoff, 1987, p. 39). According to the prototype theory, the individual assigns a "prototype" for an object, one's notion of what is typical for that category.

In the 1970s Rosch investigated the prototype effects of the category "bird". Her data showed that robins and sparrows were the best examples of birds (prototype), and ostriches, emus and penguins were regarded as peripheral examples (Lakoff, G., 1987, p.44-45). To be able to state that one type of bird is a better example than another, there must be some kind of internal structure within the category of birds. This is known as the radial category which shows how the individual evaluates each type of bird. One way of visualising this is by creating a network, showing different categories and how they interconnect. Each submeaning shares features with the prototype. Some may share more features than others. In Rosch's example of birds, one characteristic of birds that the participants in her experiment could have rated highly was the ability to fly. This factor could give reason to birds such as ostriches, emus and penguins being rated as weaker examples, as they do not fly. This does not mean, however, that these birds are any less of a member of this category; they are merely different and more peripheral in the category of birds when the prototype is robin.
These three birds may also be subcategorised further, as penguins are physically different from emus and ostriches: they cannot move as quickly for example, as well as the difference in habitat. Within the internal structure of the bird category, penguins may also be related to ducks, geese and swans as all four birds have the ability to swim. This shows how some subcategories interconnect through shared properties, a necessary component to constructing a radial category.

It should also be noted that whilst the concept of categories and radial categories spans across cultures and languages, their internal structure is often different. Categorisation differs from person to person, and from language to language. Using the example of birds, a robin may not be the prototypical bird for a Norwegian speaker in Norway, where seagulls are very common. Another example of this is the colour blue: English speakers conceptualise the colour blue with "light blue" and "dark blue" categorised as types of blue; for Russian speakers, on the other hand, *goluboj* (light blue) and *sinij* (dark blue) are two separate colours, and therefore are categorised differently (Winawer et al., 2007, p.7,780).

When identifying prototypes, there are certain characteristics that are used. A radial category network represents the relationship between different subcategories that are motivated by a central subcategory (prototype). The prototype is the subcategory that is semantically most representative of the radial category network and the other non-central subcategories are motivated by the prototype. The non-central subcategories act as variants of the prototype and do not have to share properties with the prototype (Lackoff, 1987, p.379)

3.3 *Že*

3.3.1 How I collected my data

In order to examine how *že* has been translated into English in written texts, I used a parallel corpus. The RuN corpus is a parallel corpus focusing predominantly on Norwegian and Russian texts, containing approximately two million words for each language. The corpus does however also include other languages, with around 900,000 words in English. With English being well represented in this corpus I consider it a useful data source. The RuN corpus only contains texts of literary prose, and therefore this investigation is restricted to literary prose.

I first downloaded all examples of *že* in the Russian texts where an English translation was also given. *Že* can interchangeably be written as the single letter *ž*, but this spelling was not included in my search. The corpus gave me a total of 4,555 examples. For the purposes of this study it was necessary to reduce this number. In order to focus on the modern language, I excluded any texts that were written pre-1950. This reduced the number of examples

significantly, leaving me with 486 examples from five texts. The distribution of examples across the five texts was 130, 141, 170, 3, and 42. I decided to not include the final two texts for this dataset to keep the data sample as homogenous as possible. The total number of examples was 441 from three texts: *Piknik na l'du* by Andrej Kurkov (130 examples), *Koronacija, ili Poslednij iz romanov* by Boris Akunin (141 examples) and *Medeja i eë deti* by Ljudmila Ulickaja (170 examples).

3.3.2 Limitations

As stated, the RuN corpus only contains texts of literary prose. It would have been optimal to compare the translations of *že* across different genres in order to gain a wider perspective across the entire spectrum of texts. Another challenge was the use of translations. Whilst this study in no way doubts or criticises the ability of the translators and their work, translations, particularly in fictional literature, are subjective. It can also be argued that any findings in this project are limited solely to the three texts from which the examples were taken. A broader range of texts would have strengthened the conclusions found from this data for the Russian language as a whole. It should also be pointed out that for each example of *že* only the selected sentence, or in some cases part of sentence, were used. This sometimes made it challenging to understand and interpret sentences as more context to the sentence would have been optimal, but for the purposes of this study a deeper analysis of the example texts was not carried out.

When carrying out this investigation I had to consider how to treat constructions involving *že*. "Constructions are stored pairings of form and function, including morphemes, words, idioms, partially lexically filled and fully general linguistic patterns" (Goldberg, 2003, p.219). Example (16) demonstrates a construction with *že*, taken from the Russian Constructicon (Bast et al., 2021), a database of Russian constructions. "NP-Nom" shows that after *že* a noun phrase in the nominative case occurs.

(16) Construction: *(a) kak že* NP-Nom?
Example: *A kak že mama*?
"But what about mum?"

Constructions and their meanings are intertwined and can be difficult to separate. The goal of this thesis, and in particular this chapter, is to investigate the meaning of *že* and *ved'*, however I recognise that constructions are involved in some examples of my dataset.

3.3.3 Data analysis

In order to ensure consistency and rigour in analysing my data, I took some systematic steps. Firstly, I went through each example and annotated how *že* had been translated. At this stage I began to identify some patterns. Certain fixed constructions containing *že* appeared frequently, with the same or synonymous translations, such as in (17) and (18). Note that the example of *že* or construction with *že* as well as the translator's translation of *že* are highlighted in bold where available.

> (17) *On **tut že** poprosil sekretaršu otmenit' dal'nejšij priem i bol'še nikogo k nemu ne puskat'. (Piknik na l'du*)
> ...and [he] **immediately** told his secretary to cancel all his remaining appointments and admit no one else.

> (18) *Alik rascelovalsja s Medeej i **tut že** sunul ej kartonnuju korobku, ego obyčnoe professional'noe podnošenie ... (Medeja i eë deti*)
> He [Alik] kissed Medea three times and **immediately** pressed a cardboard box into her hands, his usual professional contribution...

However, this was not the case for every example. There were some cases where the translator seemed to ignore *že* in the translation such as in (19).

> (19) *Viktor sprjatal dollary v **tu že** sumku, gde ležal podarennyj emu pistolet i opustil sumku v pogreb. (Piknik na l'du*)
> Putting the dollars with the gift gun into the shopping bag, he dropped both into the cellar.

It can be argued that *že* is lexicalised here, with the use of the definite article in the English translation being used to refer to a specific shopping bag, namely one that has just been mentioned. In this sense the idea of it being the "same" bag is implied.

An example where *že* was not directly translated, but the *že* in the Russian could be conveyed in English through a change in intonation when read aloud, is shown in (20). "*Endlung*" is highlighted in bold in this example to demonstrate where the change of intonation could occur.

> (20) *Èndlung* ***že*** *uselsja raskladyvat' pas'jans v maloj gostinoj, potomu čto ottuda byla vidna komnata mistera Karra. (Koronacija, ili Poslednij iz Romanov*)
> **Endlung** sat down to lay out a game of patience in the small drawing room, because he could see Mr Carr 's room from there.

The next step was to attempt to categorise the translations. As shown in examples (19) and (20) this was not always straightforward. After analysing all of the examples I identified eight submeanings, which are presented and explained with examples in the rest of section 3.3: EMPHASISER, ADDITION, IDENTICAL, SIMULTANEOUSLY, CONTRADICTION, CONTRAST, CONFIRMATION, URGENCY. Figure 3.1 shows the distribution of the submeanings of *že* across the dataset. It should be pointed out that this graph represents the data where one submeaning has been assigned to the example. As will be explained in section 3.3.12 some examples are motivated by more than one submeaning. For this dataset I tagged 52 examples or 11.8% of the data as being multiply motivated. For these examples I have assigned what I consider to be the dominant submeaning for the multiply motivated examples.

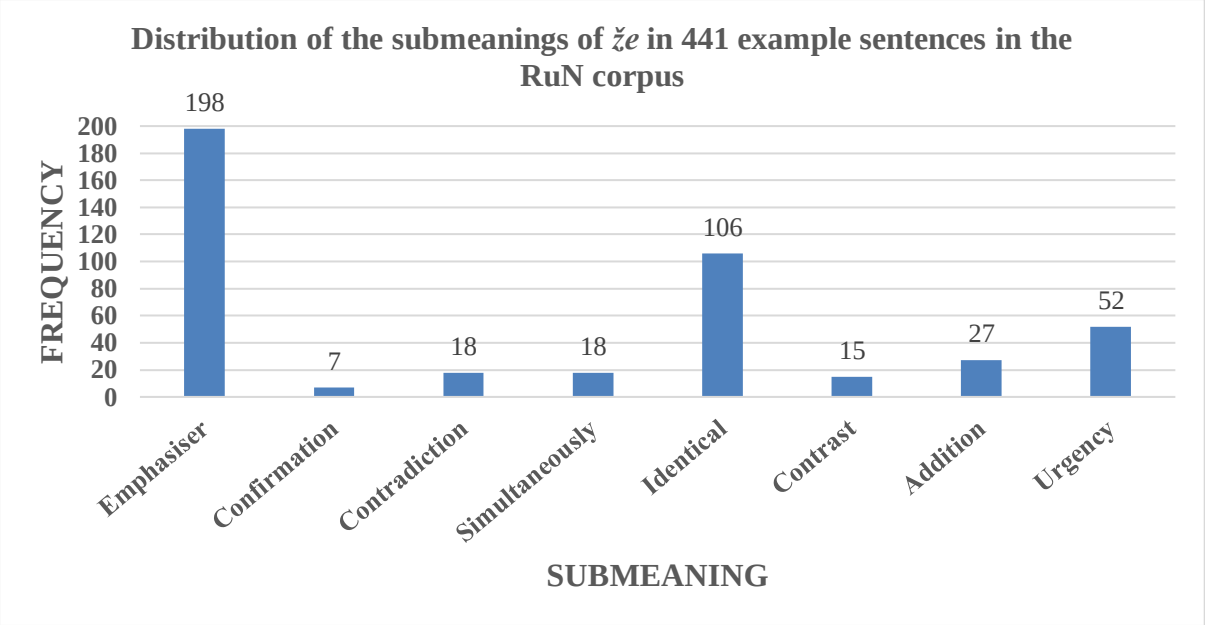

Figure 3.1: A bar chart showing the distribution of submeanings of že in 441 example sentences in the RuN corpus

I recognised EMPHASISER as the prototype for this radial category of *že*. With 198 example sentences being tagged as EMPHASISER as either the sole or main submeaning, this represents 45% of the data (see Figure 3.2).

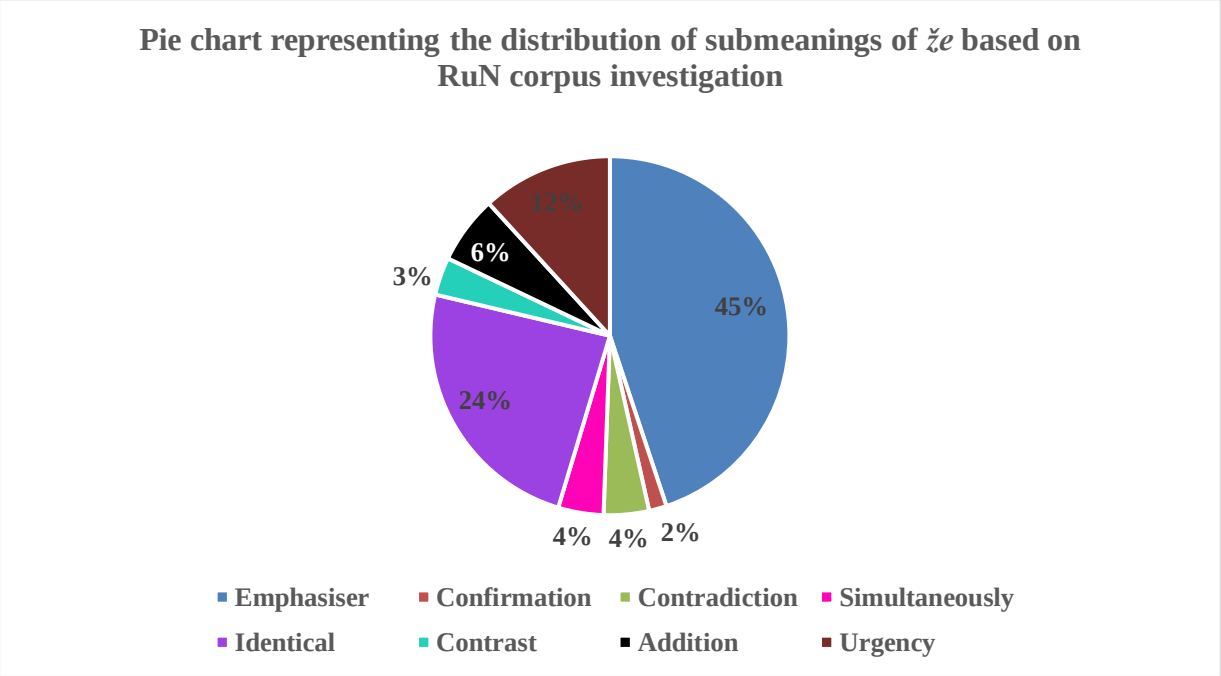

Figure 3.2: A pie chart representing the distribution of submeanings of že based on RuN corpus investigation

The high frequency of the submeaning EMPHASISER in the dataset was one reason for assigning EMPHASISER as the prototype for this radial category of *že*. When analysing the data, I recognised that the submeaning EMPHASISER appeared to motivate the other submeanings. As noted by Croft (in Geeraerts, D., 2008, p. 277), "a more schematic meaning subsuming many or all of the specific uses can arise and fit into the network". In turn, this means that the overarching schema of this radial category is emphasis to highlight the importance of something, which is in accordance with my description of the submeaning EMPHASISER. In the following sections, 3.3.4 to 3.3.11, the prototype and each of the submeanings are explained using examples from the data to illustrate their usage.

3.3.4 EMPHASISER

The name of the submeaning EMPHASISER derives from the noun *emphasis*, which can be defined as "the particular importance or attention that is given to something" (Cambridge English Dictionary, 2021). When categorised as an EMPHASISER, *že* highlights the importance that the word or phrase preceding *že* in the sentence, strengthening the speaker's intent for the recipient to be aware of this. As this notion of stress can often be highlighted through simply a change in intonation in spoken language, in many of the examples for the EMPHASISER there are no specific words in the English translation that correspond to *že*. In fact for 163 examples or 71% of the data where the submeaning or dominant submeaning is EMPHASISER, the translation does not use a word that corresponds to *že.* However, some translations were given, as in example (21) where the translation "very" is chosen to add further force to the superlative adjective "first".

(21) ***Pervye že*** *gazety dali emu pišču dlja razmyšlenij... (Piknik na l'du*)
The **very first** newspapers he looked at gave him food for thought...

In one example (22) the translator chose to represent *že* in English by adding extra punctuation, an exclamation mark, to mark emphasis:

(22) *"Segodnja* ***že*** *tol'ko tret'e fevralja ..." (Piknik na l'du*)
"Today was only February the third**!**"

One other example (23) also involved an adjustment to the punctuation. By creating two questions in English, the translator attempts to carry across the stress that the speaker wants to put on the recipient's lack of companionship.

(23) *"U tebja* ***že*** *ni ženy, ni podrugi net?" (Piknik na l'du*)
"No wife**?** No girlfriend?"

As stated, the most common finding in my analysis was that the translated English sentences did not contain a word that corresponded to *že*, although the notion of *že* as an EMPHASISER can be justified by a change in intonation when reading aloud. Two examples of this are shown in (24) and (25):

(24) "*A gde* ***že*** *ljubopytnyj tolstjak?. ." — podumal Viktor. On snova ogljanulsja po storonam. (Piknik na l'du*)
"Where," he wondered, looking around, "was nosy Fat Man?"

(25) *Ja* ***že*** *na vsjakij slučaj zatailsja v kustax. (Koronacija, ili Poslednij iz Romanov)*
I concealed myself in the bushes, just to be on the safe side.

3.3.5 ADDITION

The submeaning ADDITION describes instances where the speaker wants to introduce something that is (usually) connected to the current subject being discussed or wants to add further information. To relate ADDITION to the prototype EMPHASISER we can imagine that the speaker wants the recipient to gain some extra information. The speaker believes this added information to be noteworthy, otherwise it would have been omitted; therefore, there is an emphasis on the importance of the recipient receiving this supplementary information. *K tomu že* seems to be the most frequent construction in this submeaning, and therefore a good candidate as a prototype for the submeaning ADDITION, although other constructions are also attested, such as (26).

(26) *Na ètoj **že** stranice zametil kakuju-to kvitanciju o počtovom perevode. (Piknik na l'du)*
There was, he saw **also**, a receipt in respect of a postal draft.

The main translations for ADDITION appear to be "in addition to", "and", "also", although "moreover" was also given as the English translation of *k tomu že* in one instance. These examples are given in (27-30):

(27) "*K tomu **že** on arxitektor." (Medeja i eë deti)*
"**In addition to that**, he is an architect."

(28) ***K tomu že** ne budem zabyvat', čto mademuazel' Deklik ne imeet i togo, bez čego... (Koronacija, ili Poslednij iz Romanov*)
And also let us not forget that Mademoiselle Declique does not have those things that a respectable lady cannot manage without.

(29) "*Polli ètim zanjat'sja ne možet — on sliškom na vidu i **k tomu že** u nego polno objazannostej. (Koronacija, ili Poslednij iz Romanov*)
Paulie can't do it — he's too conspicuous and **also** he has heaps of responsibilities.

(30) *Stepanjan polučali xorošee domašnee vospitanie, francuzskomu i nemeckomu ix obučali guvernantki, **k tomu že** rannee detstvo oni proveli v Švejcarii, gde na diplomatičeskoj službe sostojal ix otec. (Medeja i eë deti*)
The Stepanyan sisters had received a good education at home and had been taught French and German by governesses. They had, **moreover**, spent their early childhood in Switzerland, where their father had held a post in the diplomatic service.

3.3.6 IDENTICAL

IDENTICAL describes cases where *že* means that something has similar or the same characteristics as something else. IDENTICAL can link to EMPHASISER in the sense that the speaker wants to reiterate the similarity of one entity or concept to another. The most common

translation in this dataset set is “[the] same”, with the constructions *to že* and *takoj že* dominating this submeaning. Examples of these are shown in (31) and (32). Other translations also classified as representing the meaning IDENTICAL consist of “as much as” from the construction *stol’ko že, skol’ko* , as well as “just like”, “equally”, and “as”, as demonstrated in examples (33-35):

(31) *Viktor i Sergej pili kofe s kon'jakom, leža na **tom že** vatnom odejale. (Piknik na l’du*)
Viktor and Sergey drank cognac-laced coffee, lying on **the same** quilted blanket.

(32) *...on pružinisto izognulsja i udaril odnogo protivnika kolenom v pax, a potom točno **takim že** manerom obošëlsja so vtorym*. *(Koronacija, ili Poslednij iz Romanov*)
...he twisted round like a spring and struck one of his opponents in the groin with his knee, and then dealt with the other in exactly **the same** manner.

(33) *No dlja sebja samogo ja stoju rovno **stol'ko že, skol'ko** nedelju ili god nazad. (Koronacija, ili Poslednij iz Romanov*)
But to myself I was worth exactly **as much as** I had been a week or a year earlier.

(34) *Viktor vygljanul v okno i uvidel v svete uličnogo fonarja dlinnuju mašinu, točno **takuju že**, kakaja byla u Miši-nepingvina. (Piknik na l’du*)
Viktor looked out, and in the light of a street lamp, saw a long car **just like** Misha-non-penguin’s moving off.

(35) *...esli Aleksej Kirillovič isčeznet **tak že** neožidanno, kak pojavilsja. (Medeja i eë deti*)
...if he [Aleksej Kirillovič] were to disappear **as** unexpectedly as he had appeared.

3.3.7 SIMULTANEOUSLY

The submeaning SIMULTANEOUSLY is used to describe *že* when it is used to depict an action happening in synchronisation with another action or actions. My data indicates that in this submeaning there are specific constructions containing *že* such as *v to že vremja* "at the same time" (36).

> (36) *...kak prevratit' ètot žanr v nečto očen' živoe, živoe i* ***v to že vremja*** *sentimental'noe, tak, čtoby daže prostoj kolxoznik, pročitav o neznakomom emu pokojnike... (Piknik na l'du*)
> Already he thought he saw how it might be vitalized, and **at the same time**, sentimentalized, so that even the simple collective farmer, never having known the late whoever-it-was he was reading about...

In this context SIMULTANEOUSLY can also be linked to IDENTICAL and EMPHASISER by the emphasis that two or more actions have the same temporal characteristics according to the speaker. In (36) it is understood that "it" is being both vitalized and sentimentalized at the same time. It should be pointed out that the temporal meaning in (20) and other instances with the construction *v to že vremja* is certainly connected to the lexeme *vremja* "time". That being said, I think it is helpful and important to separate constructions like this from the submeaning IDENTICAL to clearly demonstrate differences between different constructions and meanings of *že. Tut že* is another construction that can be categorised as the submeaning SIMULTANEOUSLY, such as example (37), where it is translated as "just as" and the translation shows that both "they" and "the penguin" went to the kitchen at the same time. A more prototypical usage of *tut že* can be seen in section 3.3.11, but this use of *tut* agrees with Nesset et al.'s radial network where the radial profile of *tut* includes a temporal meaning (Nesset et al., 2013, p.234). Such an overlap in this case could be due to the way that the translator has interpreted the sentences.

> (37) *Oni prošli na kuxnju.* ***Tut že*** *tuda prišlepal pingvin. (Piknik na l'du*)
> They went through to the kitchen, **just as** the penguin came plip- plopping that way.

3.3.8 CONTRADICTION

The submeaning CONTRADICTION is assigned to the examples of *že* when a situation or idea is in opposition with another situation or idea which has previously been mentioned, as seen in (38). CONTRADICTION links to EMPHASISER because of the speaker's desire to highlight that another idea or concept does not match the first statement. The most common English translation for CONTRADICTION in this dataset was "but", as shown in (39) and (40):

> (38) *Milord **že**, bolee ne projavljavšij ni malejšix priznakov volnenija, nespešno potjanul s ruki beluju perčatku... (Koronacija, ili Poslednij iz Romanov*)
> His Lordship, **however**, no longer displaying the slightest sign of agitation, slowly pulled a white glove off one of his hands.
>
> (39) *Po vsem ustanovlenijam za ustrojstvo koronacionnyx toržestv otvečaet moskovskij general gubernator, no ne otdavat' **že** pod sud djadju ego imperatorskogo veličestva? (Koronacija, ili Poslednij iz Romanov*)
> The person in charge of arranging the coronation festivities was the governor general of Moscow, **but** how could you bring the uncle of His Imperial Majesty to trial?
>
> (40) "*—Ty **že** znaeš', čto tam za žizn' ... — skazal on. — Strel'ba, vzryvy*" *(Piknik na l'du*)
> "**But** you know what it's like there," he said. "Shooting, explosions..."

3.3.9 CONTRAST

The submeaning CONTRAST represents the person in question's decision to deviate away from an expectation based on a previous statement. Where statement X would normally result in a reaction of Y and not Z, Z has in fact occurred. In the dataset this submeaning is primarily observed for the construction *vsë-že/vsë že*, which has been exclusively translated as "nevertheless" or "nonetheless" in these examples. I decided to not name this submeaning as "NEVERTHELESS" in order to not restrict the submeaning from expanding, should further investigations be carried out. CONTRAST is linked to EMPHASISER because of the speaker's

desire to highlight that the unexpected reaction of Z has taken place instead of the expected reaction Y. It is also related to CONTRADICTION, although differs in that CONTRAST refers to a contradiction that has taken a different path, namely reaction Z.

(41) *"— I **vse že**, mademuazel' Deklik, gde ego vysočestvo?" (Koronacija, ili Poslednij iz Romanov*)
"**Nevertheless**, Mademoiselle Declique, where is His Highness?"

(42) *Ja sejčas proiznesu slova, kotorye, vozmožno, pokažutsja vam čudoviščnymi, no **vse že** objazan ix skazat'. (Koronacija, ili Poslednij iz Romanov*)
I will say words now that might possibly seem monstrous to you, but **nonetheless** I am obliged to say them.

(43) *No na pjatyj večer ego netoroplivoj raboty gorka **vse že** končilas', i on special'no ušel čut' ran'še, ... (Medeja i eë deti*)
...but on the fifth evening of unhurried work the cabinet was **nevertheless** finished, and he specially left a little bit early...

3.3.10 CONFIRMATION

The submeaning CONFIRMATION can be used in both positive and negative situations. This submeaning is very closely related to the EMPHASISER prototype, as the speaker stresses the person or object they want to confirm or refute. The English translations use a tag question, a special construction specific to English, to represent this, as in (44) and (45). Example (46) and shows *že* as CONFIRMATION in a negative context. The original Russian sentence is depicted as a question where the speaker wants confirmation whether "he" approved the text or the subject. Example (47) uses non-standard syntax in the English translation to convey and emphasise that it was in fact "he" who assigned the role to Viktor.

(44) ***Ja že ne** obrezaju tvoi filosofskie rassuždenija, kotorye, po pravde govorja, nikakogo otnošenija k... (Piknik na l'du*)
I don't cut your philosophizings, **do I?** Even though they have, quite frankly, damn all to do with...

(45) “*— Da nu! — uxmyl'nulsja glavnyj. —* ***Ty čto že****, dumaeš’, čto ty takoj krutoj?” (Piknik na l’du*)
“You don’t say!" grinned the Chief. "See yourself as a heavy, **do you?**”

(46) *Odobrjal li on tekst* ***ili že*** *geroja teksta? (Piknik na l’du*) Though whether it was the text he approved or the subject, **was now not at all clear.**

(47) ***On že*** *provodil Viktora k redaktoru. (Piknik na l’du*)
He it was who conducted Viktor to the Editor-in-Chief.

3.3.11 URGENCY

The submeaning URGENCY refers primarily to the two constructions *srazu že* and *tut že* which can be translated as “immediately”, as shown in (48) and (49). *Srazu* and *tut* are two words that can be considered to have meanings of urgency, meaning “straight away” and “now”, respectively (*tut* can also mean “here”). This shows a clear relationship between URGENCY and the prototype EMPHASISER, as adding *že* adds emphasis to the immediateness of the action. Other translations included synonyms such as “promptly” and “at once”, as in (50) and (51).

(48) *Zakinul v vodu lesku i počti* ***srazu že*** *vytjanul serebristuju plotvičku razmerom v ladon'. (Koronacija, ili Poslednij iz Romanov*)
He tossed the line into the water and almost **immediately** pulled out a silver carp the size of an open hand.

(49) *On* ***tut že*** *poprosil sekretaršu otmenit' dal'nejšij priem i bol'še nikogo k nemu ne puskat'. (Piknik na l’du*)
...and [he] **immediately** told his secretary to cancel all his remaining appointments and admit no one else.

(50) *Katja sil'no vyrosla, obrosla koe-gde volosami, kotorye* ***tut že*** *i načala sbrivat'... (Medeja i eë deti*)
Katya had grown up markedly, sprouting hair in various places, which she **promptly** shaved off...

(51) *Èti stročki byli podčerknuty krasnym karandašom i Viktor **tut že** vspomnil o svoem poslednem razgovore s Igorem L'vovičem. (Piknik na l'du*)
This underlined in red pencil, **at once** recalled his last conversation with Igor Lvovich.

3.3.12 MULTIPLY MOTIVATED EXAMPLES

It is important to point out that there are examples in this dataset where the use of *že* is motivated by more than one subcategory. In most cases this overlap occurs with the category EMPHASISER, such as in (52)**,** where the category ADDITION is also a motivator.

(52) *Tam rjadom Bol'nica Učenyx, lečebnica u nix arenduet vremja na tomografe — **opjat' že** garantija pravil'nogo diagnoza. (Piknik na l'du*)
There's a hospital for scientists nearby, and their clinic rents time on their tomograph — an **added** guarantee of correct diagnosis.

The next example, (53) shows an instance of three submeanings being present, where IDENTICAL and SIMULTANEOUSLY are more dominant than the prototype EMPHASISER. The translation "still" is used in English to show that the world continued to function like before (SIMULTANEOUSLY) as well as stating that the world functioned in the same way as it previously did (IDENTICAL).

(53) *... nesmotrja na ego vozrosšee mnogoljudstvo i sumatošlivost', ostavalos' **vse tem že samym**, ej ponjatnym, ... (Medeja i eë deti*)
Despite being so much more crowded and having so much more hustle and bustle, the world **still** functioned in its **old way**, the way she understood, ...

3.3.13 A radial category for *že*

Figure 3.3 below shows my proposed radial category for *že*. I deduced this proposal for a radial category from the data I analysed in sections 3.3.4 to 3.3.12. EMPHASISER is the prototype of this radial category. The prototype is presented in the centre of the radial category, as this is the meaning of *že* that motivates all of the other submeaning variants. (Lakoff, 1987, p.379). Both the multiply motivated examples and the semantic properties of

the submeanings motivate the structural links between the other subcategories. Some submeanings such as IDENTICAL, SIMULTANEOUSLY, URGENCY and ADDITION are highlighted in bold as these were submeanings that had more than 15 examples in my dataset. The three remaining submeanings, CONFIRMATION, CONTRADICTION and CONTRAST are presented in a smaller font to show their lower frequency in use in this dataset. CONFIRMATION is closer to the prototype EMPHASISER because this category shares many of the same properties; CONFIRMATION differs from EMPHASISER primarily in its translation into English, where specific syntactic constructions such as tag questions are used. CONTRAST and CONTRADICTION on the other hand, have been designated the most peripheral submeanings in this radial category and are therefore further away from EMPHASISER; mainly due to the low number of examples in this dataset. Furthermore, CONTRADICTION and CONTRAST share very similar properties, mainly that they represent an occasion where there is opposition or comparison and are therefore interconnected. IDENTICAL and SIMULTANEOUSLY also share properties, shown in the multiply motivated example (53) *v to že vremja*, "at the same time" in section 3.4.9. Here, IDENTICAL properties are seen in the time at which an action occurs, and SIMULTANEOUSLY because two things are happening concurrently. Of all of the submeanings I recognised, ADDITION and URGENCY are the most independent and only share properties with EMPHASISER. This is because these submeanings consist mainly of specific constructions such as *k tomu že* "in addition to", *tut že/srazu že* "immediately".

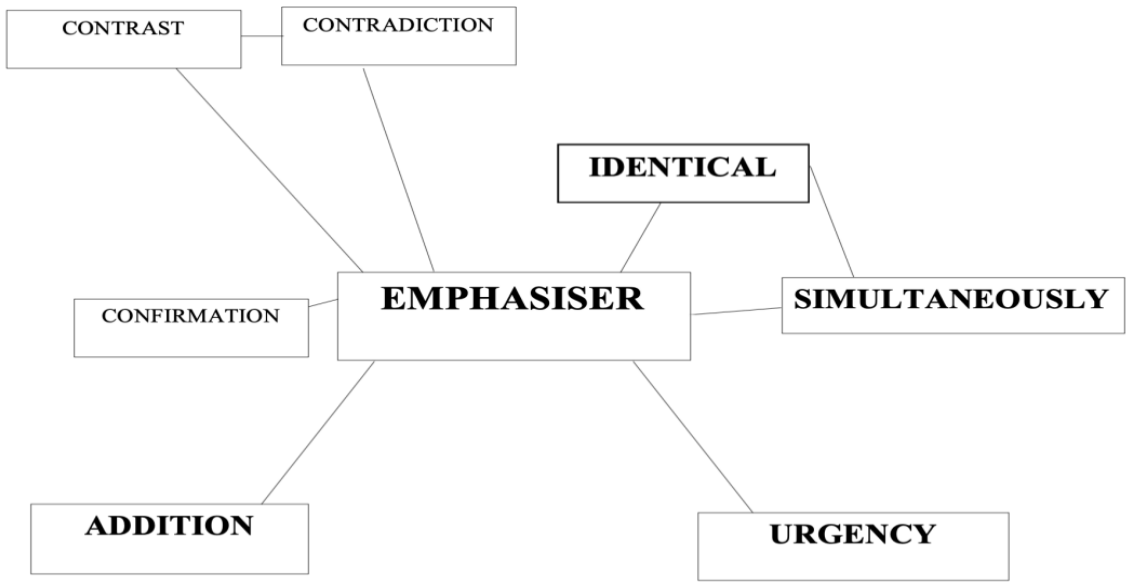

Figure 3.3: My proposal for a radial category for že based on an analysis of 441 example sentences from the RuN corpus.

The aim of this radial category is not only to represent the different meanings of *že*, but also to show how *že* is translated into English. Table 3.1 presents suggested prototypical translations for the different submeanings of *že*. I observed that the prototypical translations I suggest for ADDITION, URGENCY and IDENTICAL are connected to constructions that contain *že*. For the prototypical submeaning EMPHASISER I have not assigned a prototypical translation of *že*. There are two reasons for this: As stated in section 3.4.1, 71% of the examples for this dataset did not have a word in the translation that corresponded to *že*. Of the examples where a specific word was used to correspond to *že*, there does not seem to be a pattern, and the translation is context-specific. For the remaining four submeanings the prototypical translation is based on the most common translation from this dataset.

These prototypical translations are suggestions and not definitive. Should further studies be carried out on *že* using more data, other translations may be found. All translations are subjective, and can differ greatly based on the genre of the text.

SUBMEANING	**PROTOTYPICAL TRANSLATION**
EMPHASISER	no translation
ADDITION	in addition [to]
IDENTICAL	[the] same
SIMULTANEOUSLY	just as
CONTRADICTION	anyway
CONTRAST	nevertheless
CONFIRMATION	Tag question: Auxiliary + pronoun? (do you?, didn't he?)
URGENCY	immediately

Table 3.1: A proposal for prototypical translations for the submeanings of my proposed radial category for že.

In section 3.4 I will propose a radial category for *ved'* to facilitate comparison of *že* with *ved'*.

3.4 Ved'

To suggest a radial category for the different meanings of *ved'* I decided to carry out a corpus investigation to see how it has been translated into English. It was important to keep this

corpus investigation as similar as possible to section 3.3, to be able to show similarities and differences with the radial category investigation for *že*. I therefore used the same parallel corpus, the RuN corpus, and the genre of the texts was again restricted to literary prose.

I downloaded all examples of *ved'* with their English translations. A total of 572 examples were given. This is significantly lower than the number of examples downloaded for the investigation on *že* (4,555 examples). To retain consistency, I continued to use the same criteria as in the corpus search on *že*, such as excluding texts that were written pre-1950, in order to examine modern language only. In this parallel corpus search however, some examples were given whereby the source text was English and the Russian translation was given. In one example the source text was Norwegian and both the English and Russian translations were given. These examples were also eliminated as the focus of this search is to look at how *ved'* has been translated into English, and therefore it was paramount that the original text be in Russian.

These eliminations reduced the number to 104 examples from five texts. The distribution of examples across the five texts was: 19, 33, 46, 1, and 5. Due to the significantly fewer total number of examples in this study compared to the parallel corpus search of *že*, I decided to not eliminate any of these five texts.

Unlike *že*, *ved'* can appear word-initial in a sentence, and therefore a separate search for *ved'* with a capital "*V*" ("*Ved'*") was carried out, since the RuN corpus is case sensitive. Using the same criteria for this search, a further 47 examples from three of the five texts were added to the dataset. This gave a total of 151 examples.

The texts used were: *Piknik na l'du* by Andrej Kurkov (27 examples), *Koronacija, ili Poslednij iz romanov* by Boris Akunin (49 examples), *Žizn' i sud'ba* by Vasilij Grossman (66 examples), *Žizn' s idiotom* by Viktor Erofeev (1 example) and *Generation „П"* by Viktor Pelevin (8 examples).

3.4.1 Limitations

This part of the investigation had many of the same limitations as the investigation on *že*: the lack of variety in the genre of text, the risk of subjectivity in translation, as well as the fact that only a small number of texts were used. *Ved'* can refer to something that has previously been mentioned, so it was challenging to interpret some examples due to the sole focus on the sentence containing *ved'* and not the text as a whole. As in the study of *že*, a deeper analysis of the example texts was not carried out due to the limitations of this thesis.

3.4.2 Data analysis

As in the investigation of *že*, I first went methodically through the data and annotated how *ved'* was translated in each sentence. In many of the sentences there was no direct translation given: either it was ignored as in example (54) or the *ved'* in the Russian sentence could be conveyed in English with a change in intonation, as in example (55). The intonation change could take place in either “we’re”, “not”, or “children”, and therefore all three words are highlighted in bold.

> (54) *Da i strannym bylo èto razdraženie, ono **ved'** suščestvovalo rjadom s ljubov'ju, rjadom s gotovnost'ju otdat' Aleksandre Vladimirovne, esli ponadobitsja, svoe poslednee plat'e, podelit'sja poslednim kuskom xleba.* (*Žizn' i sud'ba*)
> And yet, at the same time, she was ready to give her last dress away to Alexandra Vladimirovna, to share her last crust of bread with her.

> (55) *... Dementij Trifonovič, **ved'** my ne deti — vinovat, ne vinovat, kakoe èto imeet značenie...* (*Žizn' i sud'ba*)
> "**We're not children**, Dementiy Trifonovich. Whether or not he's guilty is hardly the point.” …

After annotating the translations of *ved'* in each sentence, I then attempted to categorise the translations into submeanings of *ved'*. This was not always a straightforward task as 47% of the sentences in this dataset did not appear to have a direct translation as shown in example (54). After analysing all the sentences I recognised and identified 6 submeanings:

EMPHASISER, CONFIRMATION, CONTRADICTION, REACTIVATION, CONSIDERATION and AFFIRMATION. Three of the submeanings, EMPHASISER, CONFIRMATION and CONTRADICTION, are also to be found in my radial category for *že*.

The distribution of the submeanings of *ved'* in this dataset can be seen in Figure 3.4. It is important to note that this graph represents the data where one submeaning has been assigned to the sentence. In section 3.4.9, I will discuss multiply motivated examples, where sometimes more than one submeaning can be assigned. In these cases, I have assigned what I consider to be the dominant submeaning in Figure 3.4.

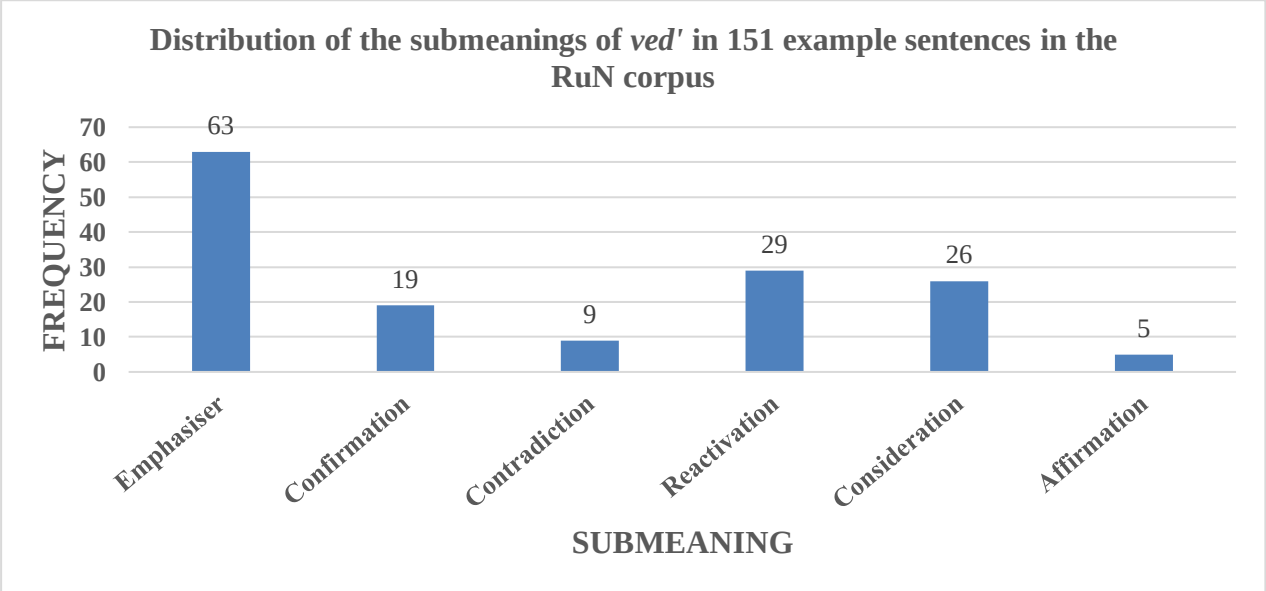

Figure 3.4: A bar chart showing the distribution of submeanings of ved' in 151 example sentences in the RuN corpus

For the radial category for *ved'*, I used the same criteria for deciphering the prototype as I used for the radial category for *že*. The submeaning EMPHASISER represents 42% of the dataset for *ved'* (see Figure 3.5), and therefore was the submeaning with the highest frequency. When analysing the data, I recognised that all the other submeanings had the same overarching schema as in the radial category for *že*: they suggest emphasis to express the importance of something, and this factor motivates their relationship to EMPHASISER. For this reason, EMPHASISER emerges as the most likely prototype for this radial category. The following

sections explain each of these submeanings using examples from the dataset to show how they are used.

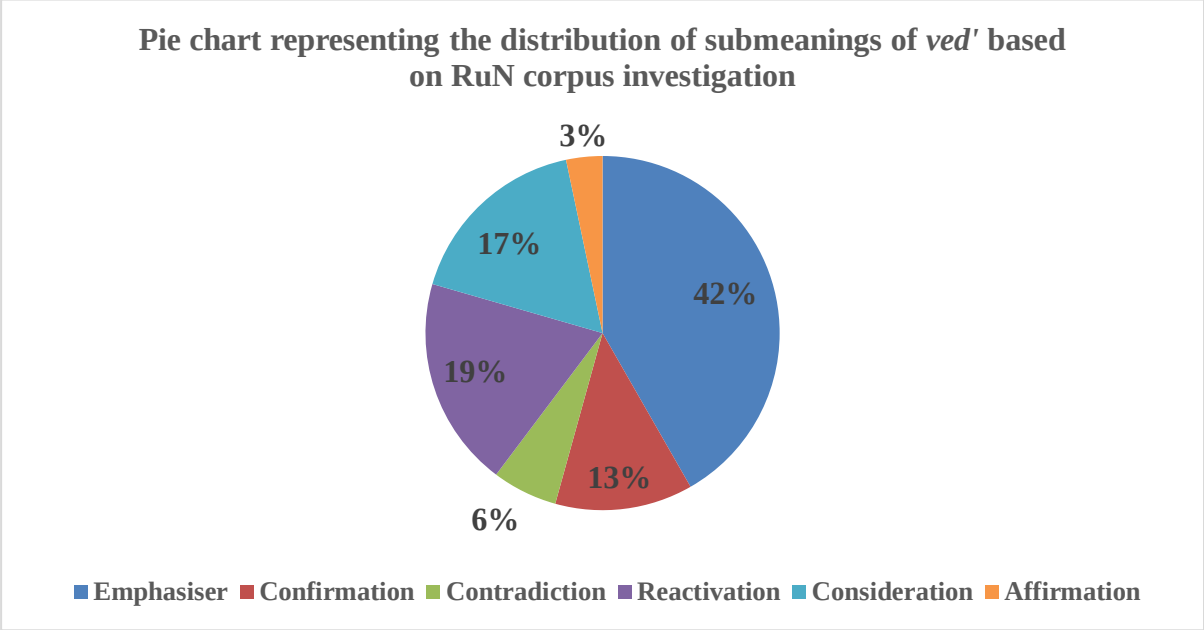

Figure 3.5: A pie chart representing the distribution of submeanings of ved' based on RuN corpus investigation

3.4.3 EMPHASISER

As stated in section (3.3.4) emphasis can be defined as "the particular importance or attention that is given to something" (Cambridge Dictionary, 2021). When assigned the submeaning EMPHASISER *ved'* seems to intensify a certain aspect of the speaker's sentence. Just as in the case of EMPHASISER as a submeaning for *že*, a direct translation corresponding to *ved'* is not given. Of the 63 examples where EMPHASISER was the submeaning or dominant submeaning, no translation was given for 43, or 72% of the examples. This can be seen in example (56), where it can be argued that *ved'* emphasises "only" to stress the fact that the flight to Stalingrad does not take a long time.

> (56) ... *kak na "jake" doberetsja do Stalingrada, vsego* ***ved'*** *neskol'ko časov, — v Rjazani možno zarjadit'sja, ...* (*Žizn' i sud'ba*)
> In a Yak he could fly to Stalingrad in only a few hours; he could refuel in Ryazan — he had a friend there who was a controller.

Another example where a direct translation was not given can be seen in example (57) where the emphasis can be portrayed by through intonation. The three highlighted words in the English translation show suggestions for where the change in intonation could occur.

(57) *No **ved'** v tot den' ego vysočestvo byl eščë živ!* (*Koronacija, ili Poslednij iz Romanov)*
"**But on that** day His Highness was still alive!"

Where translations were given, these varied, with no specific pattern allowing for any generalisations. In example (58) it can be argued that the author has translated *ved'* with "yes" to emphasise that Podchufarov's agreement to the previous statement, whereas in example (59) *ved'* has been translated with "then".

(58) *A **ved'** verno, spokojno, nikto osobenno ne bespokoit, — skazal Podčufarov.* (*Žizn' i sud'ba)*
"**Yes**, that's true enough," said Podchufarov. "No one really bothers us here."

(59) *I **ved'** ... k Èmilii vy tože neravnodušny?* (*Koronacija, ili Poslednij iz Romanov)*
"And **then** ... you are rather partial to Emilie, I believe?"

3.4.4 CONFIRMATION

The submeaning CONFIRMATION is one of the submeanings I have assigned to *ved'* that also appears in the radial category for *že*. As stated in section 3.3.10 CONFIRMATION relates to the prototype EMPHASISER because the speaker wants to highlight the person or object they are confirming. In this sense the submeaning CONFIRMATION is also closely related to AFFIRMATION (see section 3.4.8) which claims a declarative statement to be true. As with CONFIRMATION as a submeaning for *že*, the submeaning CONFIRMATION for *ved'* also uses tag questions, such as examples (60) and (61), as well as discourse markers at the end of the sentence, as in example (62).

(60) *Vy **ved'** ne kurite?* (*Koronacija, ili Poslednij iz Romanov*)
“You don't smoke, **do you?**"

(61) *... ni Linda, ni p plennikov, a tol'ko Počtal'on i ego semejstvo — kto-to **ved'** vse-taki v dome byl?* (*Koronacija, ili Poslednij iz Romanov)*
neither Lind nor the p-prisoners are here, only the postman and his family. After all, there was someone in the house, **was there not**?

(62) *Možete i sami iskat' i vybirat' — rodina **ved'** ne znaet vsex svoix geroev — ...* (*Piknik na l’du)*
Not all our country’s notables are known to it, **you see**. Many prefer it like that... "

Example (63) uses different syntax to both emphasise and confirm the fact that it is the cold that “he” likes, and this is motivated by using *ved’* in the sentence.

(63) *Naverno, ploxo, — soglasilsja Viktor. — On **ved'** xolod ljubit, a tut teplo ...* (*Piknik na l’du)*
"That’s probably it," he agreed. " **What** he likes is cold, and here it’s warm."

Finally, example (64) is a significant example to discuss as the translator has changed the syntax of the original sentence in order to recognise *ved’* in the sentence. In changing the structure of the sentence to become an interrogative question, it can be interpreted that the speaker wants someone to confirm that they understand Russian. An alternative translation using tag questions like examples (60) and (61) could have been “You understand plain Russian, don’t you?”.

(64) *Nu, vytri nos, tebe **ved'** russkim jazykom govorjat.* (*Žizn' i sud'ba*)
"Go on then! Wipe your nose! **Don't you** understand plain Russian?"

3.4.5 CONTRADICTION

As in the radial category for *že*, CONTRADICTION is a submeaning for the radial category *ved'* to represent that a situation or idea is in conflict with another, and links to EMPHASISER as the speaker wishes to highlight this discord between the two statements. In the radial category for *že*, the submeaning CONTRADICTION was most commonly translated as "but". This is also the case for three out of the nine examples where *ved'* expresses the submeaning CONTRADICTION, as in example (65):

(65) *I **ved'**, znaete, ešče včera pered snom govoril: ...* (*Žizn' i sud'ba*)
"But before he went to sleep last night, he said, …"

In example (66) the speaker does not agree that somebody can be identified. Example (67) also shows this, but from a positive perspective. The speaker contradicts a previous statement that the receiver is not brave. Examples (66) and (67) contained the word "*no*", which translates as "but", however *ved'* appears to emphasise this contrast in statements.

(66) *No kak ego opoznat', **ved'** my daže ne znaem doktora v lico?* (*Koronacija, ili Poslednij iz Romanov)*
But how can we identify him? We don't even know what he looks like.

(67) *No **ved'** ty xrabryj, ty ne poboiš'sja.* (*Koronacija, ili Poslednij iz Romanov)*
"…but you are brave; you will not be afraid."

3.4.6 REACTIVATION

The submeaning REACTIVATION refers to instances where *ved'* is used by a speaker when reactivating knowledge about a situation he/she already has, or when reactivating information that the speaker knows the receiver has or believes the receiver should have. REACTIVATION links to the prototype EMPHASISER because *ved'* is used here to emphasise that the information given is not new, and either the speaker or/and the receiver is being reminded that they already know about the topic at hand. In the dataset most sentences that were labelled with the

submeaning REACTIVATION as the sole or dominant submeaning did not have a direct translation. Example (68) was categorised as REACTIVATION as it can be interpreted that the speaker already had knowledge that the correspondent was a man. The use of *ved'* in example (69) shows that the militiaman is reactivating the receiver's knowledge that he/she already has the militiaman's telephone number from a previous event in the past.

> (68) *Net, — lixoradočno dumal on. — Korrespondent* ***ved'*** *— mužčina ...* (*Piknik na l'du)*
> No, the correspondent was a man, was his first fevered thought.

> (69) *Ne otkažus', — milicioner kivnul. — Zvonite, telefon* ***ved'*** *znaete!* (*Piknik na l'du*)
> "Wouldn't say no, " confirmed the militiaman." Just ring — you've got my number.

In some cases where *ved'* expresses the submeaning REACTIVATION, the English translation uses participle constructions. In example (70) the perfect participle is used, and example (71) uses a participial phrase. Although these constructions were only used in one of the texts, they are worthy of mention as they fit well into the submeaning of reactivating previously obtained knowledge.

> (70) *Èto moj televizor! — govorila ona i Viktor byl vynužden s ètim soglašat'sja ,* ***ved'*** *dejstvitel'no kupili televizor na ee den'gi* (*Piknik na l'du*)
> "It's my telly!" she said, which Viktor, having in fact bought it with her money, had to concede.

> (71) *Xotja, možet, i ne bylo v ètom ničego udivitel'nogo,* ***ved'*** *devočka provodila s pingvinom gorazdo bol'še vremeni, čem Viktor.* (*Piknik na l'du*)
> Although that was not surprising perhaps, seeing that she spent far more time with him than he did.

3.4.7 CONSIDERATION

The submeaning CONSIDERATION represents a situation where the speaker has considered what has been stated previously and claims that it is true. In this sense it can be argued that here, *ved'* should be given the submeaning AFFIRMATION. However, the submeaning CONSIDERATION focuses on the fact that the speaker has taken other factors into account before a statement is made. CONSIDERATION can relate to the prototype EMPHASISER because the speaker is stressing the fact that he/she has in fact considered all eventualities. In this submeaning *ved'* was almost always translated as "after all", with only one example using an alternative translation. Example (72) shows an instance where the speaker claims that someone else is not heavy and justifies this by explaining that she was light enough to be carried in the speaker's arms in a previous case. The use of *ved'* in example (73) is used to show that the speaker has considered all circumstances and concluded as to the most important secret of the doctor's power.

> (72) *S drugoj storony, xot' ona i netjaželaja (mne **ved'** uže prixodilos' nosit' eë na rukax), smogu li ja v odinočku podnjat' eë …* (*Koronacija, ili Poslednij iz Romanov*)
> On the other hand, even though she was not heavy (**after all**, I had already carried her in my arms), would I be able to carry her up such a steep slope on my own?

> (73) *A **ved'** glavnaja tajna m moguščestva doktora zaključalas' imenno v ženstvennosti.* (*Koronacija, ili Poslednij iz Romanov)*
> "But, **after all**, the most important secret of the doctor's power was precisely femininity.

There was one sentence in the dataset that primarily expressed the submeaning CONSIDERATION that did not have the translation "after all". Example (74) translates *ved'* as "just" and expresses that the speaker wants Lyuda to take everything into account and reconsider why she refuses to help beggars. This sentence is a candidate for being a multiply motivated example (see section 3.4.9) as it can be argued that this example also expresses the

submeaning REACTIVATION since the speaker is commanding the receiver to think, and therefore potentially reactivate previous knowledge.

(74) *"Ljuda, kak èto ty možeš' otkazyvat' niščim, — **ved'** podumaj: golodnyj prosit u tebja, u sytoj ... "* (*Žizn' i sud'ba*)
"Lyuda, how can you refuse beggars? **Just** think: you've got enough to eat while someone else is hungry and begging..."

3.4.8 AFFIRMATION

AFFIRMATION describes cases where *ved'* is used when the speaker makes a declarative statement, which they believe to be true. AFFIRMATION can link to EMPHASISER in the sense that the speaker wants to stress that their statement is accurate. The submeaning AFFIRMATION also links to the submeaning CONFIRMATION in the sense of confirming how true a statement is. Only five sentences in the dataset express the sole or dominant submeaning AFFIRMATION, and therefore it is not possible to make generalisations or claim any specific patterns. In example (75) it could be argued that the translator has translated *ved'* with "yes" to emphasise the speaker's statement that the receiver does not know anything about the topic of discussion. In example (76), where *ved'* is positioned word-initial, *ved'* has been translated as "and". In this instance "and" works well to both affirm and stress the added information or statement that the speaker is about to present.

(75) *Ax, vy **ved'** ničego ne znaete!* (*Koronacija, ili Poslednij iz Romanov)*
"Ah yes, you know nothing about it!"

(76) ***Ved'** dve svežie pexotnye divizii polnogo sostava pribyli iz germanskogo tyla i sosredotočeny v rajone Traktornogo zavoda, zlovešče bezdejstvujut.*
(*Žizn' i sud'ba)*
And two full-strength infantry divisions had been brought up from the rear and disposed opposite the Tractor Factory; there they remained ominously inactive.

Another sentence from the dataset shows the use of *ved'* expressing AFFIRMATION without a direct translation. In example (77) *ved'* is annotated as expressing the submeaning AFFIRMATION because the speaker's statement about Russian grand dukes is opinion-based, even though the speaker believes it to be true.

(77) *On navernjaka naročno tuda p-priexal, čtoby vysmotret' podxodjaščuju žertvu — **ved'** na Lazurnyj bereg vesnoj priezžaet stol'ko grands dues russes!* (*Koronacija, ili Poslednij iz Romanov*)
He must have gone there deliberately to seek out his future victim — so many Russian grand dukes go to the Côte d'Azur in spring!

3.4.9 MULTIPLY MOTIVATED EXAMPLES

As was the case for the radial category for *že*, there are many instances in this dataset where the use of *ved'* is motivated by more than one submeaning. The following examples show the malleability of this radial category, that different submeanings can motivate *ved'* at the same time. In most cases multiply motivated sentences involve the prototypical submeaning EMPHASISER combined with another submeaning. Example (78) shows an instance where *ved'* expresses both EMPHASISER and CONFIRMATION. It can be argued that in this sentence *ved'* is used to emphasise *dejstvitel'no*, but also confirms the claim that the speaker Alexandra and others do in fact distract "him".

(78) *Tiše, **ved'** dejstvitel'no emu mešaem, — skazala Aleksandra Vladimirovna.* (*Žizn' i sud'ba*)
"Sh!" said Alexandra Vladimirovna." We probably really do distract him."

Example (79) illustrates multiple motivation by the submeaning EMPHASISER along with two other submeanings. *Ved'* appears to stress the word "you" in the sentence (EMPHASISER), and it can be argued that the receiver has previous knowledge that he/she can be regarded as a representative of the royal family (REACTIVATION). The tag question "surely?" also gives grounds for the submeaning CONFIRMATION in this case.

(79) *Vy **ved'** možete sčitat'sja polnomočnym predstavitelem avgustejšej familii?* (*Koronacija, ili Poslednij iz Romanov)*
"**You** can be regarded as a plenipotentiary representative of the royal family, **surely**?"

There were examples in the dataset where the submeaning EMPHASISER is not one of the dominant submeanings. In example (80) *ved'* expresses the submeanings REACTIVATION and AFFIRMATION. The use of the past perfect tense in the English translation shows that the receiver had been warned of the speaker's lateness at a time in the past, and therefore this is knowledge the receiver already had (REACTIVATION). At the same time, it can be said that the speaker is making a statement that they deem to be true (AFFIRMATION).

(80) *Ja **ved'** predupredil, čto mogu opozdat', poètomu ona ne somnevalas', čto vy pojavites' pervym ...* (*Koronacija, ili Poslednij iz Romanov)*
I had warned you that I might be late, and so she had no doubt that you would be the first to arrive.

3.4.10 *Ved'* in initial position

Of the 151 examples in the dataset, there were forty-seven examples where *ved'* was word-initial. I decided to investigate these examples to see if there are any patterns to conclude from these examples. Figure 3.6 shows the distribution of the submeanings assigned to these forty-seven examples, and Figure 3.7 shows the same information in a pie-chart.

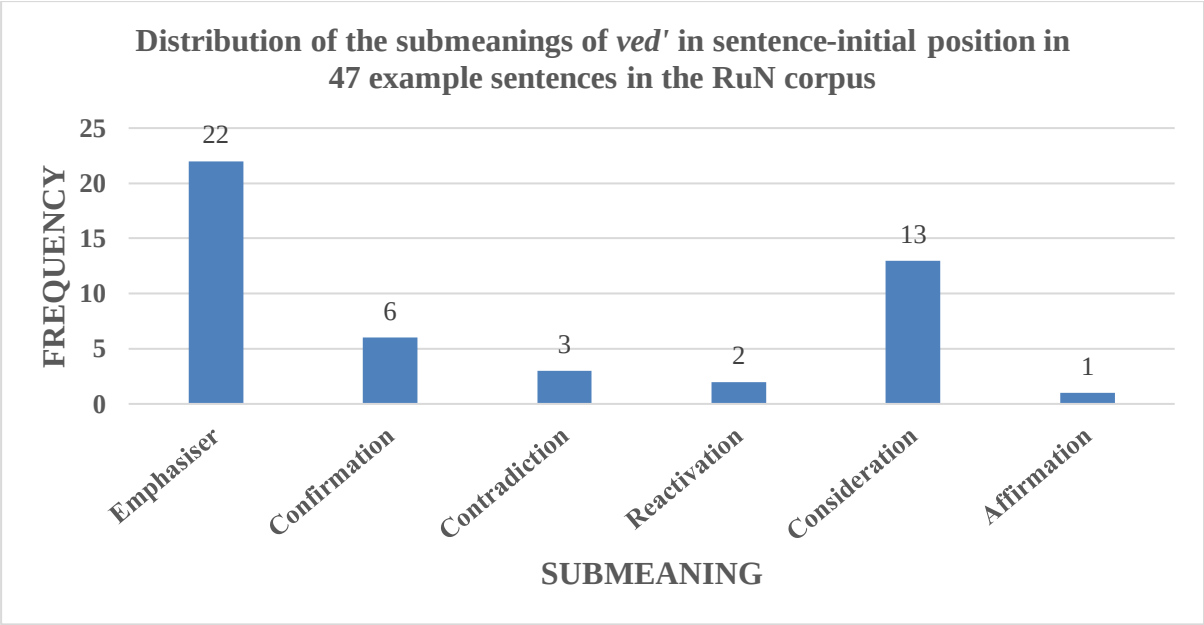

Figure 3.6: A bar chart showing the distribution of submeanings of ved' in sentence-initial position in 47 example sentences in the RuN corpus

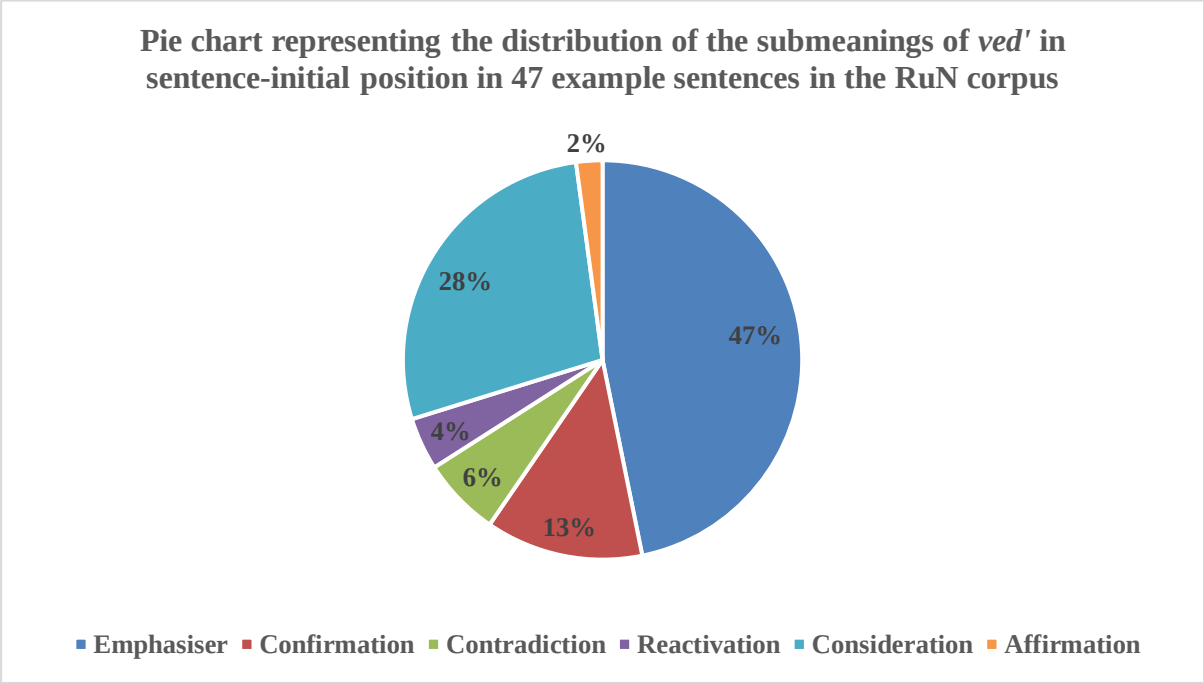

Figure 3.7: A pie chart representing the distribution of submeanings of ved' in sentence-initial position based on RuN corpus investigation

As Figure 3.7 shows, almost half of all examples of sentence-initial *ved´* express EMPHASISER as the sole or dominant submeaning. The submeaning CONSIDERATION is of particular interest

here. In the entire dataset CONSIDERATION is the submeaning expressed by *ved'* in twenty-six examples, of which thirteen involve *ved'* in sentence-initial position. All thirteen examples in sentence-initial position are translated as "after all" in English. I do not claim this to be an absolute translation of *ved'* when *ved'* is in sentence-initial position and assigned the submeaning CONSIDERATION, but the results from this dataset show evidence that this could be a potential pattern.

3.4.11 A radial category for *ved'*

Figure 3.8 shows my proposal for a radial category for *ved'*. This radial category is based on the data analysed in sections 3.4.3 to 3.4.10. Like in the radial category I proposed for *že* (see Figure 3.3), the submeaning EMPHASISER is the prototype. All the other submeanings stem from the prototype EMPHASISER because, as I have demonstrated, they all have some common properties with the submeaning EMPHASISER. The submeanings EMPHASISER, REACTIVATION and CONSIDERATION are highlighted in bold because these submeanings had the highest frequency. Similar to the radial category proposal for *že*, the submeaning CONFIRMATION is presented close to the prototype EMPHASISER in this radial category for *ved'*, which is due to the similarity in their properties. The submeanings CONFIRMATION and AFFIRMATION have been justified as being related, however the submeaning AFFIRMATION is peripheral and further away from EMPHASISER, like the submeaning CONTRADICTION, due to the low number of examples in the dataset. The submeanings REACTIVATION and CONSIDERATION stand alone, neither close to nor far away from EMPHASISER, and are two independent submeanings. As observed in section 3.4.7 the submeaning CONSIDERATION appears to translate as "after all" when in sentence-initial position.

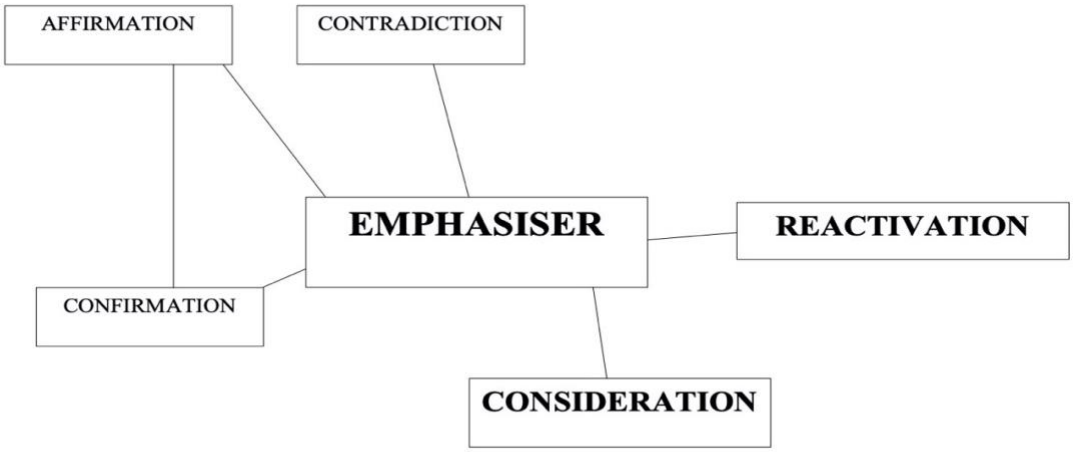

Figure 3.8: My proposal for a radial category for ved' based on an analysis of 151 example sentences from the RuN corpus.

Table 3.2 shows suggested prototypical translations for the different submeanings of *ved'*. For the prototypical submeaning EMPHASISER and submeanings REACTIVATION and AFFIRMATION a prototypical translation of *ved'* has not been given. In these submeanings a specific word in the English translation did not always correspond to *ved'*. In cases where *ved'* was translated for these submeanings, no specific patterns appear to occur, and the translation is context-specific. The suggested prototypical translations of the submeanings CONFIRMATION, CONTRADICTION and CONSIDERATION are based on the most common translation from the dataset. These translations are not definitive, and may change if another dataset is used.

SUBMEANING	**PROTOTYPICAL TRANSLATION**
EMPHASISER	No translation
CONFIRMATION	Tag question: Auxiliary + pronoun? (do you?, didn't he?)
CONTRADICTION	But
REACTIVATION	No translation
CONSIDERATION	After all
AFFIRMATION	No translation

Table 3.2: A proposal for prototypical translations for the submeanings of my proposed radial category for ved'.

3.5 Further analysis

3.5.1 The radial categories for *že* and *ved'*

When deriving the radial categories for *že* and *ved'* I set out to investigate each on its own terms and not link these two categories, but rather to investigate each on its own terms. When developing the radial category for *ved'* it became apparent however, that there is an overlap with the radial category for *že.* The prototype and submeaning EMPHASISER, as well as the submeanings CONFIRMATION and CONTRADICTION all appear in both proposed radial categories. In the datasets used for this investigation the submeanings CONFIRMATION and CONTRADICTION were more frequent in the radial category for *ved'* than with *že*, but the prototype and submeaning EMPHASISER had almost equal distribution in both radial categories (45% in the radial category for *že* and 42% in the radial category for *ved'*). Figure 3.9 is a Venn diagram representing the relationship between my proposed radial categories for *že* and *ved'*.

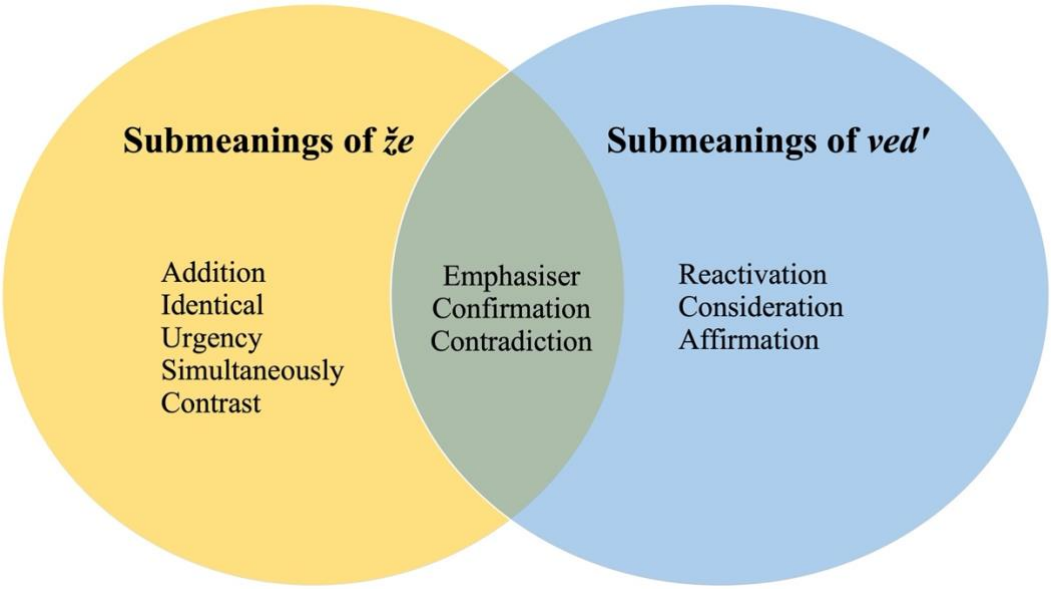

Figure 3.9: A Venn diagram showing the relationship between my proposed radial categories for že and ved'.

3.5.2 A further look at the datasets

I applied certain criteria to the datasets for *že* and *ved'*. The examples in both datasets were taken from texts written after 1950. In the dataset for *že* two texts were removed for having low sample sizes, whereas this criterion was not applied to the dataset for *ved'* as there were,

on the whole, fewer examples. After applying these criteria to the datasets for *že* and *ved'* I was left with examples from three texts for the dataset for *že*, and 5 texts for the dataset for *ved'*. Two texts appear in both the dataset for *že* and the dataset for *ved'*, namely *Piknik na l'du* and *Koronacija, ili Poslednij iz Romanov*. In the following section I will show how the submeanings of *že* and *ved'* are distributed across these specific texts.

3.5.3 *Piknik na l'du*

Whilst it is expected that there are more examples from the dataset for *že,* the difference between *že* and *ved'* here was quite significant. In the dataset for *že* there were 130 examples from *Piknik na l'du,* representing 29% of the data for *že*. For the dataset for *ved'* there were twenty-seven examples, which represents 18% of the data for *ved'*. In the examples taken from *Piknik na l'du* there were no examples where the submeaning of *že* was CONTRAST (see Figure 3.10). In the case of *ved'* there were no examples of the submeaning AFFIRMATION (see Figure 3.11). Looking further at Figures 3.10 and 3.11 it is evident that the prototype and submeaning EMPHASISER is, as expected, the dominant submeaning. It is interesting to note the two submeanings that overlap both datasets. The submeaning CONFIRMATION represents 18% of the data for *ved'*, but only 3% for *že*. The submeaning CONTRADICTION does not have a high distribution in either the dataset for *že* or *ved'*, representing 1% and 4% of the dataset, respectively. The submeanings appear to be slightly more evenly distributed in the dataset for *ved'* than the dataset for *že*, with three submeanings, CONFIRMATION, REACTIVATION and CONSIDERATION, all having similar distribution (18%, 26% and 19%).

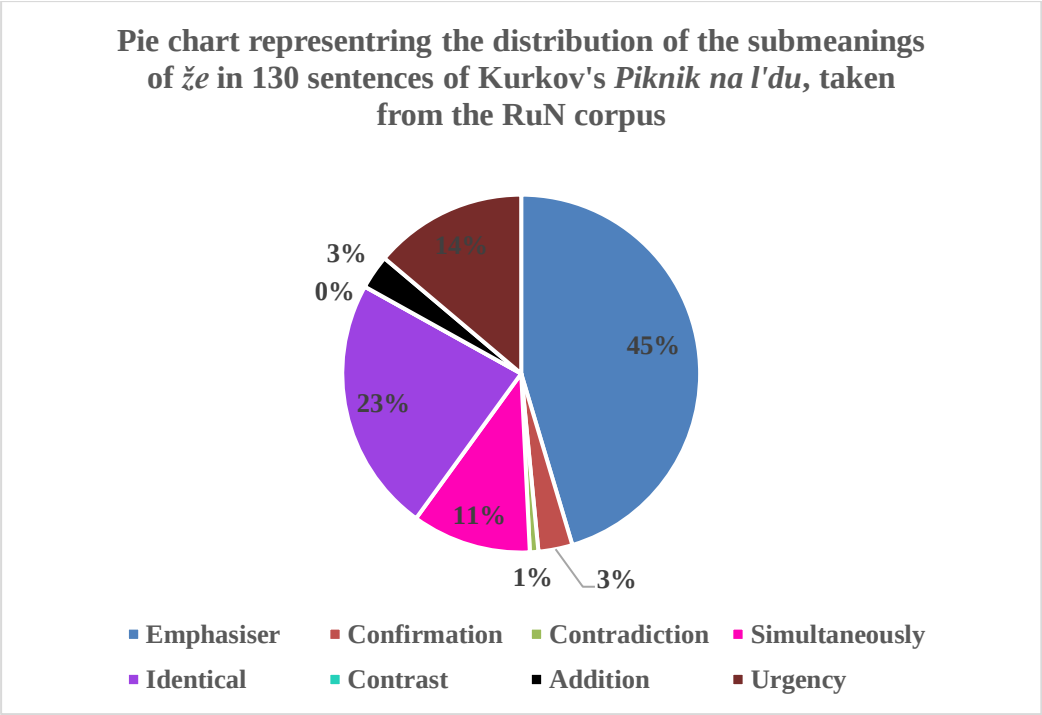

Figure 3.10: A pie chart chart representring the distribution of the submeanings of že in 130 sentences of Kurkov's Piknik na l'du, taken from the RuN corpus

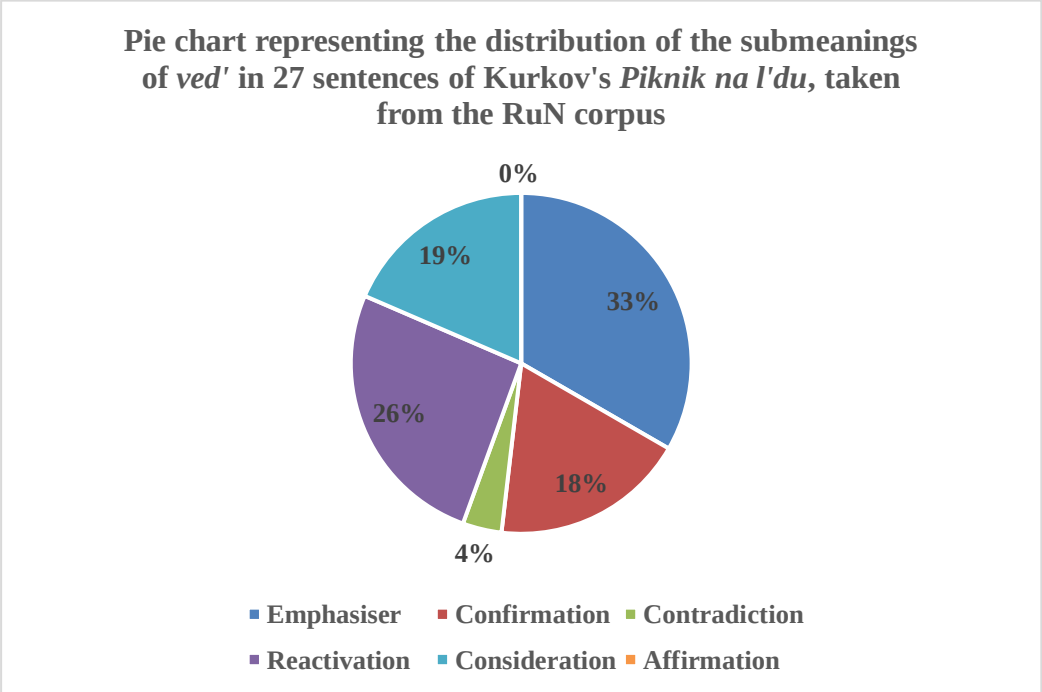

Figure 3.11: A Pie chart representing the distribution of the submeanings of ved' in 27 sentences of Kurkov's Piknik na l'du, taken from the RuN corpus

3.5.4 *Koronacija, ili Poslednij iz Romanov*

For the example sentences from *Koronacija, ili Poslednij iz Romanov* there were 141 sentences from the dataset for *že*, representing 32% of the dataset (see Figure 3.12). There were forty-nine examples of *ved'*, also representing 32% of the dataset for *ved'* (see Figure 3.13). The prototype and submeaning EMPHASISER has a much higher frequency in the dataset for *že*, with 50% of the 141 examples expressing EMPHASISER as the sole or dominant submeaning. In the case of the dataset for *ved'*, EMPHASISER was not the submeaning with the highest frequency. The submeaning EMPHASISER represents 25% of the dataset for *ved'*, whereas the submeaning CONSIDERATION represents 31% of the data. As in the observations made in section 3.5.2 for *Piknik na l'du*, the other two submeanings that overlap both datasets, CONFIRMATION and CONTRADICTION, had a higher distribution in the dataset for *ved'* than in *že*. Another noteworthy comment about the dataset for *že* is the distribution of the other submeanings. Apart from the submeaning IDENTICAL, which represents 20% of the data, all of the other submeanings have a very low, and quite even distribution.

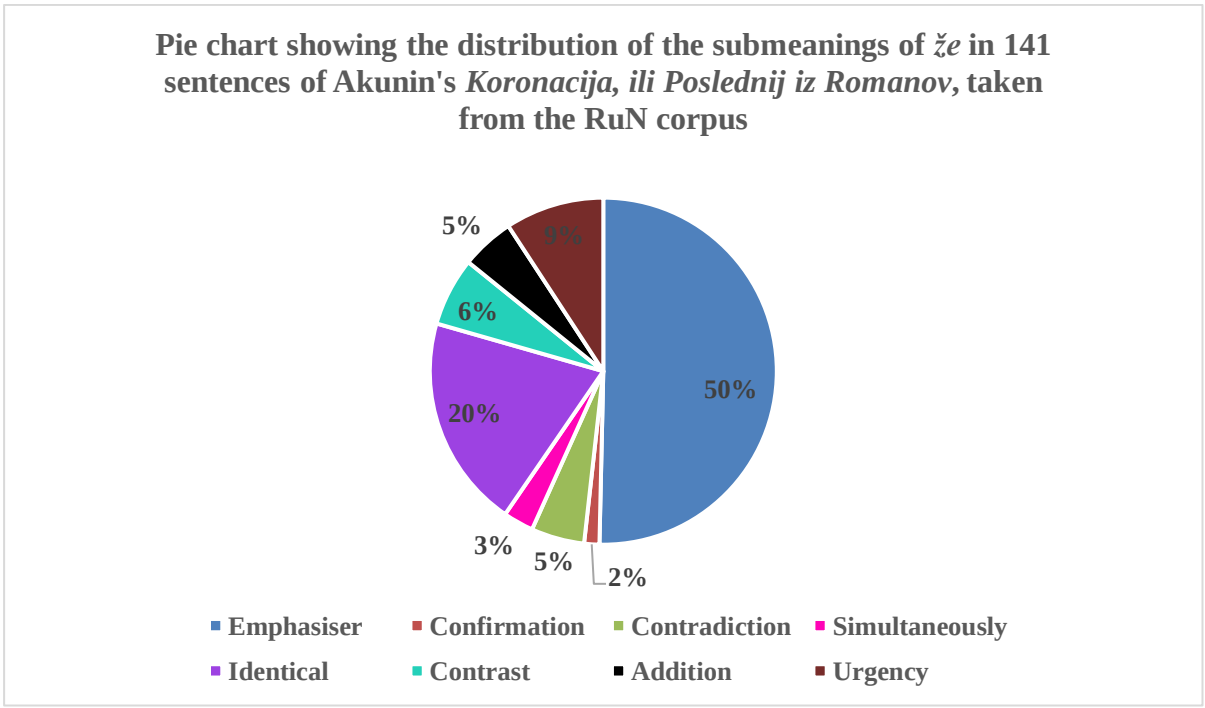

Figure 3.12: A Pie chart showing the distribution of the submeanings of že in 141 sentences of Akunin's Koronacija, ili Poslednij iz Romanov, taken from the RuN corpus

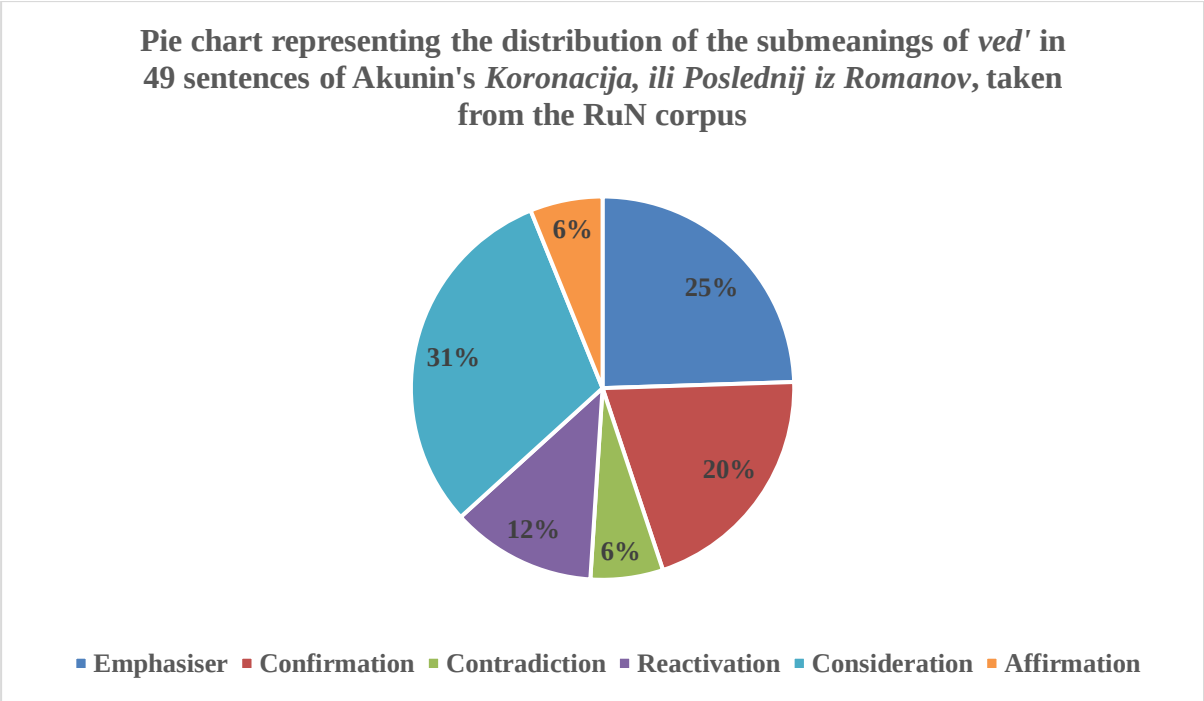

Figure 3.13: A Pie chart representing the distribution of the submeanings of ved' in 49 sentences of Akunin's Koronacija, ili Poslednij iz Romanov, taken from the RuN corpus

3.5.5 Pedagogical impact

Subsections 3.5.2 and 3.5.3 have demonstrated further how the submeanings of *že* and *ved'* are represented in specific texts. The radial categories I have proposed have relevance for second language learners (henceforth "students") and teachers. By showing students how flexible the submeanings of *že* and *ved'* can be, as well as their relation to the prototype submeaning EMPHASISER, these lexemes can become less problematic for teachers to teach, and for students to learn.

In the classroom students could be presented with texts or sentences that use *že* and *ved'*. The students' task could then be to decide which submeaning or submeanings they would assign to the sentences containing *že* and *ved'* using the radial categories I have proposed.

Alternatively, students could carry out a corpus search themselves. As claimed by Nesset and Janda (2014) this method of teaching can be both beneficial to students as well as teachers. By using real data and being able to analyse it, in this case using the proposed radial

categories to work out the meanings of *že* and *ved'*, students will obtain a real sense of achievement.

3.6 Conclusion

This aim of this investigation was to create and propose a radial category for *že* and *ved'* to attempt to better understand their meaning, as well as be able to translate them into English.

For *že* I recognised 8 submeanings: EMPHASISER, ADDITION, IDENTICAL, SIMULTANEOUSLY, CONTRADICTION, CONTRAST, CONFIRMATION and URGENCY.

For *ved'* I identified 6 submeanings: EMPHASISER, CONFIRMATION, CONTRADICTION, REACTIVATION, CONSIDERATION and AFFIRMATION.

I have demonstrated that whilst there are different submeanings connected to *že* and *ved'*, these submeanings all show common properties of emphasis and stress, which gives reason for the submeaning EMPHASISER as the prototype for both radial categories, from which all of the other submeanings stem. In addition, I have demonstrated how these submeanings are related, with the submeanings CONFIRMATION and CONTRADICTION appearing in both radial categories. I have also shown the internal structure of a radial category for *že* and for *ved'*, as well as show how these two radial categories relate. In terms of showing how *že* and *ved'* are rendered in English I have proposed prototypical translations where I deem it to be possible based on patterns I have observed in the datasets. Finally, I have demonstrated how these radial categories can be useful for learners of Russian, and suggested ways of using the findings of this investigation in a pedagogical setting to facilitate second language learners' understanding of a difficult area of Russian grammar.

These radial categories are both based on limited datasets and taken from solely fictional texts. Further empirical research into understanding the meaning of *že* and *ved'* should include a wider variety of text genres. This could lead to an expansion or change of these radial categories and could help further enhance the understanding of *že* and *ved'*. This investigation may also serve as a template for research on other Russian particles such as *-to, daže* or *eščë.*

4 What factors may influence the replaceability of *že* with *ved' ?*

4.1 Introduction

As shown in chapter 3.5 the radial categories proposed for *že* and *ved'* have some overlapping properties. This finding demonstrates the close relationship between these two lexemes. This chapter aims to investigate this relationship further and focuses on the replaceability of *že* with *ved'*. I will consider various different factors that may influence the replaceability of *že* with *ved',* using methods such as Chi-square tests, Fisher tests, Logistic Regression and cTree analysis to evaluate whether any statistical significance of replaceability exists. These statistical tests will be carried out using the statistical computer program RStudio (2015). Due to the limitations of this thesis, I will not consider the opposite possibility, namely the replaceability of *ved'* with *že*.

The potential factors influencing the replaceability of *že* with *ved'* that will be explored in the following chapter include the meaning of *že* (as proposed in chapter 3.3) and the part of speech to the left of *že*. The null hypothesis ($\mathbf{H_0}$) for this investigation was:

> $\mathbf{H_0}$= Neither the meaning of *že* nor the part of speech to the left of *že* influence the replaceability of *že* with *ved'*.

As the null hypothesis rejects two possibilities, I propose two alternative hypotheses, labelled $\mathbf{H_1}$ and $\mathbf{H_2}$ respectively:

> $\mathbf{H_1}$= The meaning of *že* influences the replaceability of *že* with *ved'*.
> $\mathbf{H_2}$= The part of speech to the left of *že* influences the replaceability of *že* with *ved'*.

4.2 Method

4.2.1 Data collection

The RNC was used to provide the data for this investigation. According to Endresen et al., *že* is tagged both as a particle and as a conjunction in the RNC (2016). To be able to look at the entire picture of *že* in the corpus, I decided to extract examples where *že* is tagged both as a particle and as a conjunction. In Russian, *že* can also be shortened to simply *ž* and is semantically identical to *že*. As in chapter 3 I decided to not include *ž* in my search of the RNC, and therefore only examples of the full form *že* were extracted.

Due to the scope of this thesis, I decided it acceptable to analyse 400 sentences. From these 400 sentences I made three datasets:

- Dataset 1: Where *že* is tagged as both particle and conjunction (400 sentences)
- Dataset 2: Where *že* is tagged as particle (200 sentences)
- Dataset 3: Where *že* is tagged as conjunction (200 sentences)

The RNC provides several pieces of information for each example, however I extracted solely the example sentence and tagged my data manually using four categories, which are explained below.

Replaceability

This category serves to describe whether *že* in the example sentence could be replaced with *ved'*. The following key was used to tag this category: Not replaceable "NR", Replaceable with word order change "RW" and Replaceable without word order change "WWR". If a sentence was tagged as "RW", this meant that *že* could be replaced with *ved'*, however in order for this to be acceptable, the position of *ved'* would have to change. If a sentence was tagged as "WWR", *že* could be replaced with *ved'* without any other changes occurring in the sentence. The replaceability of *že* with *ved'* was based on the judgement of one native speaker of Russian.

Meaning

This category provides a semantic judgement of *že* in the example sentence. This category was based on the submeanings of the radial category I proposed for *že* in chapter three and are therefore based on my own judgement. There were eight submeanings: "EMPHASISER" (emphasis and stress), "IDENTICAL" (often meaning "the same"), "URGENCY" (often meaning "immediately"), "SIMULTANEOUSLY" (often meaning "at the same time" or "still"), "ADDITION" (often meaning "as well"), "CONTRADICTION" (used to show opposition to a previous statement), "CONTRAST" (to show contrast to a previous statement) and "CONFIRMATION" (often used with tag questions). This data sample did not include any examples of the submeanings CONFIRMATION or CONTRAST from my radial category in sections 3.3.9 and 3.3.10. The distribution of the category "Meaning" across the datasets is shown in Figure 4.1:

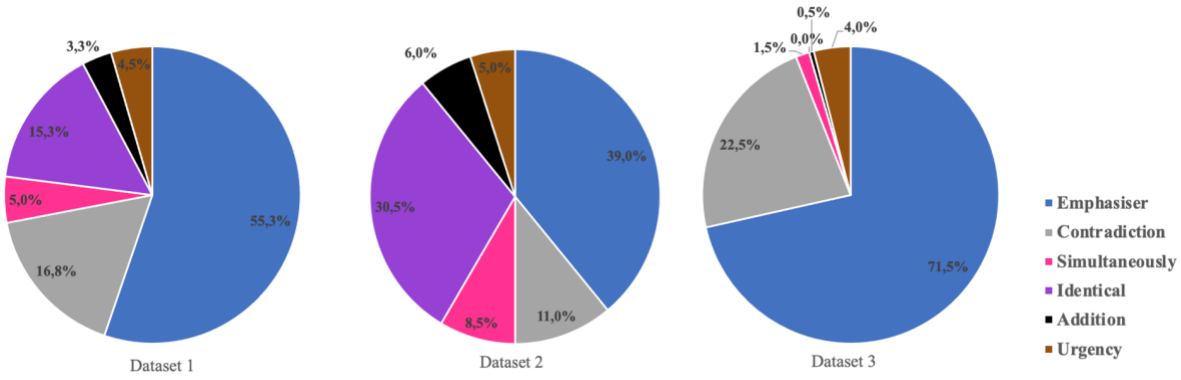

Figure 4.1: pie-charts showing the distribution of "Meaning" across dataset 1 (all), dataset 2 (že as particle) and dataset 3 (že as conjunction).

RNC_annot

This category shows how the given sentence was originally tagged in the RNC. The two options in this category were "part" (particle) and "conj" (conjunction). As Endresen et al. (2016) demonstrate, the RNC tagging system can contain errors. In my data I did not make any changes to the tags, to ensure the data remains authentic to the RNC. Tagging errors in my data are further discussed in section 4.2.2.

POS

As *že* is an clitic it is unstressed and is dependent on the preceding word. The category POS represents the "Part Of Speech" assigned to the word to the left of *že.* This tagging was

carried out manually for time efficiency, but in cases of ambiguity such as in example (81), where *kak* can be classified as both an adverb and a conjunction, the RNC was consulted.

> (81) *Stranno,* ***kak že*** *u nas na sajte est' učitelja iz russkix škol iz Turcii, Velikobritanii i Grecii.*
> "It's strange how on our website there are teachers from Russian schools who are from Turkey, Great Britain and Greece."

The tags used were: "ADJ" (adjective), "ADV" (adverb), "CNJ" (conjunction), "DPN" (demonstrative pronoun), "PART" (particle), "PN" (pronoun), "PPN" (personal pronoun), "PRE" (preposition), "SUB" (noun), "VB" (verb). The tag "PRE" did not appear in dataset 2 (*že* as particle), and the tag "PART" did not appear in dataset 3 (*že* as conjunction), as can be seen in Figure 4.2.

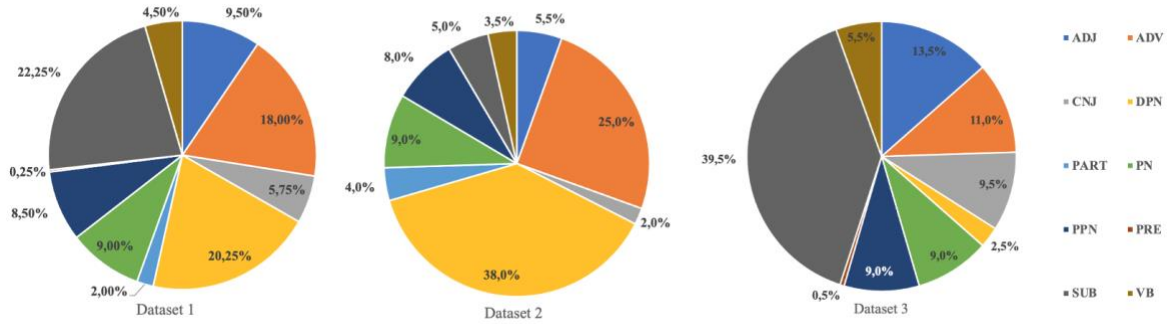

Figure 4.2: Pie-charts showing the distribution of "POS" across dataset 1 (all), dataset 2 (že as particle) and dataset 3 (že as conjunction).

4.2.2 Limitations

There were some limitations in this investigation. Whilst it would be beneficial to have more native speakers verify the replaceability of *že* and *ved'* in the sentences, only one native speaker was involved due to time constraints and the scope of this thesis.

The usability of the RNC created challenges when tagging the parts of speech. As the tags are not available when the data is downloaded, this had to be done manually, which introduced the risk of human errors.

Further supporting Endresen et al.'s proposal for a new tagging system in the RNC, there was an example in my data where the *že* was incorrectly tagged in the RNC (examples (82) and (83)). In example (82) *že* is tagged as a particle but tagged as a conjunction in example (83).

(82) *"U nas **že**, vo-pervyx, sil'noe lobbi so storony tabačnyx kompanij,…"*
"First of all, we have a strong lobby on the part of tobacco companies…"

(83) *"U nas **že** daže zapadnye licenzionnye igry učat tol'ko tomu, čto vyigrat' praktičeski nevozmožno,…"*
"In our country, even Western licensed games only teach you that it is almost impossible to win, …"

4.2.3 Data examples

The following three examples are all taken from dataset 2 (*že* as particle) to show how my data analysis is put into practice.

(84) *"Ja **že** predlagaju vspomnit' otmennogo personaža i vpolne sebe interesnuju trilogiju o žestkom oxotnike na vampirov."*
"I suggest recalling the excellent character and quite an interesting trilogy about the tough vampire hunter."

In example (84) *že* is tagged as EMPHASISER for meaning, as a particle in the RNC, the POS to the left of *že* is a personal pronoun, and in this sentence *že* can be replaced with *ved'* without word order change. This means that example (84) is interchangeable with example (85):

(85) *"Ja **ved'** predlagaju vspomnit' otmennogo personaža i vpolne sebe interesnuju trilogiju o žestkom oxotnike na vampirov."*

Example (86) shows a different possibility for replacing *že* with *ved'*:

(86) *"Počemu **že** ona rascenivaetsja kak vozmožnost', a te primery net?"*
"Why is that regarded as a possibility, and those examples are not?"

In example (86) *že* is tagged as CONTRADICTION for meaning, as a particle in the RNC, and the POS to the left of *že* is an adverb. In this sentence *že* can be replaced with *ved'* with word order change, as in (87).

(87) ***"Ved'** počemu ona rascenivaetsja kak vozmožnost', a te primery net?"*
** "Počemu **ved'** ona rascenivaetsja kak vozmožnost', a te primery net?"*

I will finally show an example (88) from my dataset where it was not considered possible to replace *že* with *ved'*:

(88) *"V to **že** vremja poroj mestnye kompanii vyigryvali tendery na postavku oborudovanija i programmnogo obespečenija v drugie regiony."*
"At the same time, sometimes local companies won bids for the supply of equipment and software to other regions."

In example (88) *že* is tagged as SIMULTANEOUSLY for Meaning, as a particle in the RNC, and the POS to the left of *že* is a demonstrative pronoun. In this sentence *že* cannot be replaced with *ved'*, with or without word order change, as shown in (89) and (90):

(89)* *"V to **ved'** vremja poroj mestnye kompanii vyigryvali tendery na postavku oborudovanija i programmnogo obespečenija v drugie regiony."*

(90) ** "**Ved'** v to vremja poroj mestnye kompanii vyigryvali tendery na postavku oborudovanija i programmnogo obespečenija v drugie regiony."*

4.3 Statistical tests and analysis

4.3.1 Adjustment of data

Due to the restrictions on some of the following tests, I decided to collapse the tags “RW” and “WWR” in the category “Replaceability” and retag them all as replaceable “R”. The distribution of the tags “NR” and “R” are shown in Figure 4.3:

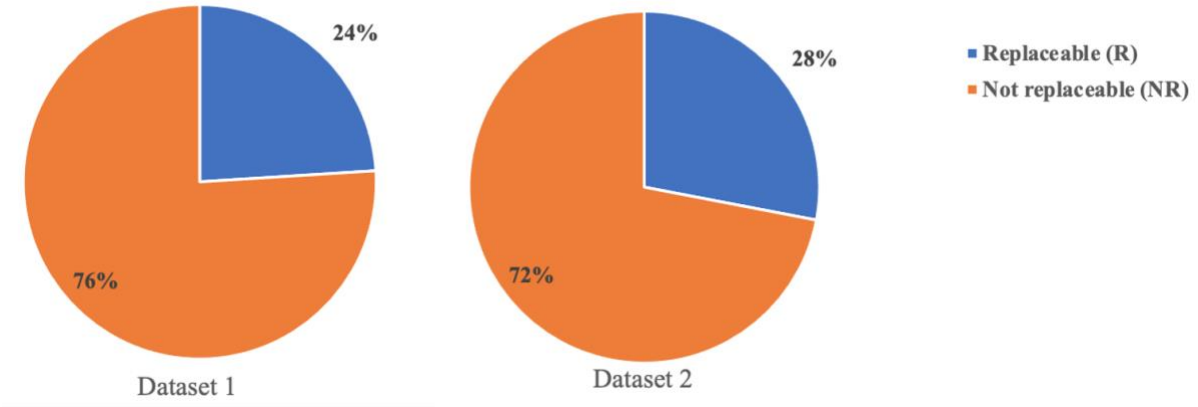

Figure 4.3: Pie-charts showing the distribution of “Replaceability” of že with ved’ in dataset 1 (all) and dataset 2 (že as particle).

I have previously stated that I made three datasets: dataset 1 consisted of all 400 example sentences, dataset 2 consisted of the 200 example sentences where *že* was tagged as a particle in the RNC, and dataset 3 consisted of the 200 example sentences where *že* was tagged as a conjunction in the RNC. Due to the scope of this investigation, I only carried out the statistical tests on dataset 1 and dataset 2. As discussed in chapter two there has previously been discussion in the literature to eradicate particle as a part of speech (Zwicky 1985, Endresen et al., 2016). For this reason, it is interesting to add extra focus on dataset 2.

4.3.2 Chi-square and Cramér’s V: Replaceability vs. Meaning

The aim of this investigation is to determine which factors may influence the replaceability of *že* with *ved’*. I decided to first see if my null hypothesis could be rejected. I started by carrying out a Chi-square test comparing “Replaceability” to “Meaning” to see if there was any significant deviation, as well as a Cramer’s V to determine the effect size. To check if a Chi-square was possible, I tested for the expected values. For a Chi-square to work, the

expected values for each cell should be five or higher (King et al., 2011, p.369). Unfortunately, this was not the case when comparing "Replaceability" to the meanings ADDITION, SIMULTANEOUSLY and URGENCY, so they had to be taken out of this test. The three remaining categories for "Meaning" (CONTRADICTION, EMPHASISER and IDENTICAL) made up 87% of the data for dataset 1 (all), and 81% of the data for dataset 2 (*že* as particle). I decided that this was an acceptable amount of data to retain.

When analysing the Chi-square test the p-value is of interest to report any significant deviation. It shows the probability of getting such an extreme distribution or more extreme between "NR" vs "R" and the three selected categories of "Meaning". For the Chi-square test I used the add-on package "vcd" (Meyer et al., 2020). The p-value for dataset 1 was 1.047e-05 and 6.334e-10 for dataset 2. These p-values mean that both dataset 1 and 2 have a statistically significant effect. A Cramér's V calculation was carried out on the same data to show the effect size, with a range from 0 to 1. A guideline is that $0.1 \leq < 0.3$ indicates a small effect size, $0.3 \leq < 0.5$ indicates a medium effect size, and ≥ 0.5 indicates a large effect size (Levshina, 2015, p.209). Dataset 1 had a Cramér's V of 0.256 and dataset 2 had a Cramér's V of 0.513. This shows that the effect size of "Replaceability" vs "Meaning" in dataset 1 is small to medium and large for dataset 2.

4.3.3 Fisher test: Replaceability vs. Meaning

Whilst the Chi-square test gives us a general overview of the data we are looking at, a Fisher test takes a closer look at the data. As the data used are categorical and not of a large number, it is acceptable to carry out a Fisher test. I carried out a Fisher test on dataset 1 (all) and included all six meanings of *že* that appear in the dataset. The results of the Fisher test are shown below in Table 4.1. Boxes highlighted in green represent attraction, and boxes highlighted in red represent repulsion. As the matrix in this case contains only two rows, the results were the same for each row, but the attraction differed.

	ADDITION	CONTRADICTION	EMPHASISER	IDENTICAL	SIMULTANEOUSLY	URGENCY
Not replaceable with *ved'*	**[+] 0.7417**	**[-] 0.7559**	**[-] 1.583e-07**	**[+] 5.651e-06**	**[+] 1.304e-07**	**[+] 6.232e-07**
Replaceable with *ved'*	**[-] 0.7417**	**[+] 0.7559**	**[+] 1.583e-07**	**[-] 5.651e-06**	**[-] 1.304e-07**	**[-] 6.232e-07**

Table 4.1: The results of the Fisher test look at Replaceability vs. Meaning for dataset 1 (all).

Overall, Table 4.1 shows that *že* tends not to be replaceable with *ved'*. The submeanings IDENTICAL, SIMULTANEOUSLY and URGENCY had the strongest attractions to "NR" (not replaceable), and the strongest repulsions to "R" (replaceable). This table indicates that in sentences where *že* is tagged as IDENTICAL, SIMULTANEOUSLY and URGENCY, *že* is not replaceable with *ved'*. EMPHASISER, on the other hand, had the strongest attraction to "R" and strongest repulsion to "NR", indicating that in sentences where *že* is tagged as EMPHASISER, *že* is replaceable with *ved'*. These findings are consistent with and strengthen my findings in chapter two, as the submeaning EMPHASISER is one of the three submeanings that does overlap in my radial categories for *že* and *ved'*, and therefore where *že* is an EMPHASISER, we can see that it can be replaced with *ved'*. The submeanings IDENTICAL SIMULTANEOUSLY and URGENCY were amongst the submeanings that did not overlap with *ved'*, and this statistical finding is in accordance with my observation in chapter two. The meanings ADDITION and CONTRADICTION did not give significant results, and therefore no absolute conclusions can be made from this dataset.

4.3.4 Chi-square and Cramér's V: Replaceability vs. POS

To further test the null hypothesis, a Chi-square test was carried out, focusing on "Replaceability" vs "POS". Like the Chi-square test in section 4.3.2 with "Meaning", a test of the expected values revealed that only three of the tags in the category "POS" had an expected value of five or more for all cells: "Adverb", "Demonstrative pronoun" and "Pronoun". For dataset 1 (all) this amounted to 72% of the data, but only 47% for dataset 2 (*že* as particle). The p-value for dataset 1 was 2.501e-08 and 4.408e-09 for dataset 2. In terms of Cramér's V scores, the results vary slightly from the Cramér's V carried out on "Replaceability" vs. "Meaning". The Cramér's V for dataset 1 is 0.43, showing a medium effect size, and 0.518

for dataset 2, showing a large effect size. This shows that dataset 2 is more significant than dataset 1.

4.3.5 Null hypothesis rejection

To summarise the findings of the statistical tests carried out so far on my data, the null hypothesis that "Neither the meaning of *že* nor the part of speech to the left of *že* influence the replaceability of *že* with *ved'*" can be rejected. By comparing "Replaceability" with "Meaning" the results of the Chi-square test and the Cramér's V showed a small effect size for dataset 1, and a large effect size for dataset 2. It appears that there is some relationship between these two factors. Investigating further the results of the Fisher test showed the sentences with "NR" and IDENTICAL as well as "R" and EMPHASISER showed strong attractions. I then compared "Replaceability" with "POS" and the results of both the Chi-square test and Cramér's V showed significant and similar results.

The results of these tests indicate that both the meaning of *že* and the part of speech to the left of *že* influence the replaceability of *že* with *ved'*. Whilst I have rejected the null hypothesis, the results so far do not clearly show which factor, "Meaning" or "POS", may influence the replaceability of *že* with *ved'* more. Further investigation using a Logistic Regression Model and a cTree was carried out to try to answer this question.

4.3.6 Logistic Regression Model

A logistic regression model (henceforth LRM) can be used in statistical modelling to show how multiple factors are associated with the outcome of a dependent variable (Baayen et al., 2013, p. 257). In the case of this investigation, there are two possible outcomes (NR and R), and this is a binomial model (Levshina, 2015, p.253). For this model I used dataset 1 (all). The aim is to predict the "Replaceability" of *že* with *ved'* based on a combination of the predictors "Meaning", "RNC_annot" and "POS" (as explained in section 4.2.1). The optimal situation for an LRM is to find a model with the fewest predictors, but the one that best suits the data (Baayen et al., 2013, p.257). To measure for this, I used Akaike's Information Criterion (AIC) using different combinations of predictors with the aim of finding the lowest AIC values possible. I tested six combinations and the best result was:

Replaceability ~ Meaning + RNC_annot + POS (AIC value: 339.3361)

		Model Likelihood Ratio Test		Discrimination Indexes		Rank Discrim. Indexes	
Obs	400	LR chi2	133.53	R2	0.425	C	0.853
NR	304	d.f.	15	g	3.280	Dxy	0.706
R	96	Pr(> chi2)	<0.0001	gr	26.583	gamma	0.727
max |deriv|	0.001			gp	0.259	tau-a	0.258
				Brier	0.124		

	Coef	S.E.	Wald Z	Pr(>|Z|)
Intercept	-5.4275	1.1785	-4.61	<0.0001
Meaning=contradiction	1.8886	0.9287	2.03	0.0420
Meaning=emphasiser	1.8235	0.8671	2.10	0.0355
Meaning=identical	-1.0503	1.1299	-0.93	0.3526
Meaning=simultaneously	-8.0459	44.3696	-0.18	0.8561
Meaning=urgency	-8.3115	43.5316	-0.19	0.8486
RNC_annot=part	1.5701	0.3283	4.78	<0.0001
POS=ADV	2.1659	0.8168	2.65	0.0080
POS=CNJ	2.6223	0.8992	2.92	0.0035
POS=DPN	1.2839	0.9847	1.30	0.1923
POS=PART	-8.6292	70.7620	-0.12	0.9029
POS=PN	2.9387	0.8227	3.57	0.0004
POS=PPN	3.5331	0.8361	4.23	<0.0001
POS=PRE	-6.9939	200.1331	-0.03	0.9721
POS=SUB	1.9000	0.8075	2.35	0.0186
POS=VB	3.2565	0.9015	3.61	0.0003

Figure 4.4: The results of the Logistic Regression Model for dataset 1 (all)

For the LRM I used the add-on packages “rms” (Harrell Jr., 2020) and “car” (Fox and Weisberg, 2019). The LRM showed a number of statistics and I will report the most important (see Figure 4.4). The LRM shows the total number of observations (400 in this case) and their distribution (304 examples of “NR” and 96 of “R”). The “Model Likelihood Ratio Test” tells us whether the model is significant. With a p-value Pr(>chi2) <0.0001, this reports a good level of significance and tells us that at least one predictor (“Meaning”, “RNC_annot” or “POS”) is associated with the outcome. Another statistic worth reporting is the Nagelkerke pseudo R-2. For this LRM the pseudo-R2 was 0.425. The other statistic to report is the concordance index *C* under “Rank Discrimination Indexes”. In this dataset it is certainly worth reporting, with a score of 0.853, showing excellent discrimination (Levshina, 2015, p.259).

The table of coefficients should also be reported. The first estimate, “intercept”, represents all predictors at their reference levels. The reference levels for each predictor in the LRM in R are organised alphabetically. This means that the reference level for “Meaning” was “ADDITION”, for “RNC_annot” it was “Conj” and for “POS” it was “ADJ”. To find out the reference level for “Replaceability” the following coding was used:

```
levels(dat$Replaceability)
[1] "NR" "R"
```

This means that “NR” is the reference level and therefore goes in the denominator of the odds ratio: R/NR. This means that for logit values, + values favour “R” and - values favour “NR”. The coefficient score for “intercept” was -5.4275 and its p-value was 0.0001. This score tells us that for this dataset, when the meaning of *že* is ADDITION, the RNC tags the example of *že* as a conjunction and the POS to the left of *že* as an adjective, there is a strong tendency for *že* not to be replaceable with *ved’*. I will highlight the most interesting coefficient scores for each predictor.

Whilst none of the levels for the “Meaning” variable are particularly significant, EMPHASISER and CONTRADICTION gave the best p-values: 0.0355 and 0.0420 respectively. The coefficient scores were 1.8235 and 1.8886 and these scores tell us that EMPHASISER and CONTRADICTION increase the probability of *že* being replaceable with *ved’* in comparison with ADDITION.

For the variable “POS” the parts of speech to the left of *že* that increase the probability of *že* being replaceable with *ved’* in comparison with “ADJ” were “ADV”, “CNJ”, “PN” and “PPN”.

For the variable “RNC_annot” there are only two levels, and the level “part” (particle) gave a significant p-value of 0.0001 and a coefficient score of 1.5701. As this was the only level to compare against “conj” (conjunction) I used exponentiation to obtain the simple odds ratios. The odds ratio score was 4.807129, which tells us that for this dataset, when the reference levels are ADDITION for “Meaning” and the part of speech to the left of *že* is an adjective, the

odds of *že* being replaceable with *ved'* in cases where the RNC has tagged *že* as a particle are 4.8 times higher than in sentences where the RNC has tagged *že* as a conjunction.

4.3.7 cTree analysis

The aim of a cTree is to show which of the factors in dataset 1 (all) give the optimal sorting of the data. A cTree analysis works well with few factor levels, which makes it an ideal statistical tool for this investigation (Baayen et al., 2013, p.264, p.267). For the cTree my dependent variable was "Replaceability", and the independent variables were "Meaning", "RNC_annot" and "POS". Before carrying out the cTree I checked the variable importance for replaceability, as shown in Figure 4.5. Figure 4.5 presents the variable importance and shows that POS appears to be the most influential factor, and the meaning of *že* seems to be the least influential factor, although it competes well with the RNC annotation. For the cTree I used the add-on packages "party" (Hothorn et al., 2006), "lattice" (Sarkar, 2008), and "Hmisc" (Harrell Jr et al., 2020). The cTree is presented in Figure 4.6.

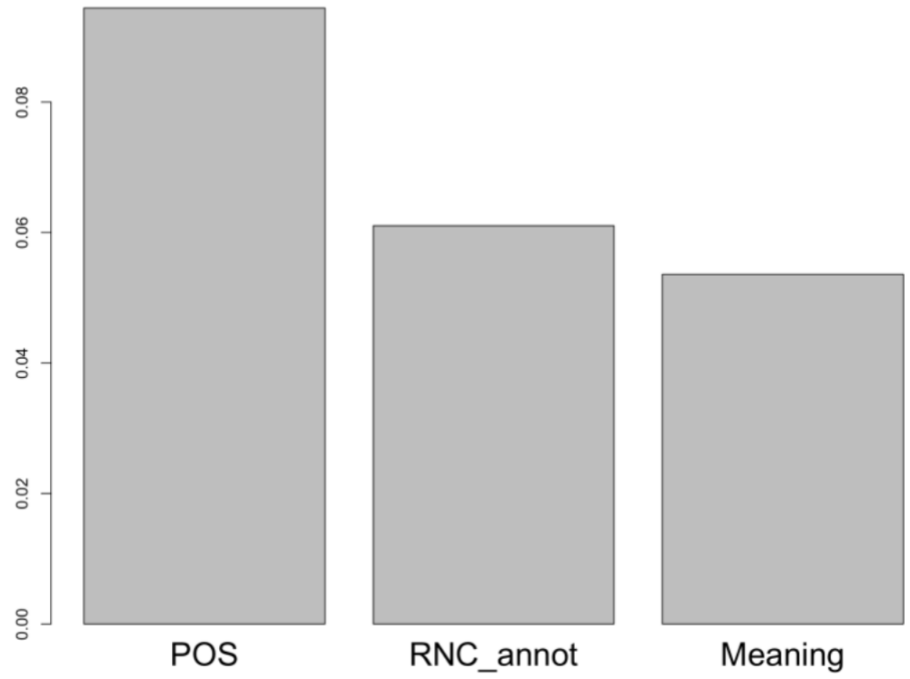

Figure 4.5: A bar chart showing the variable importance of replaceability for the cTree analysis in Figure 4.6

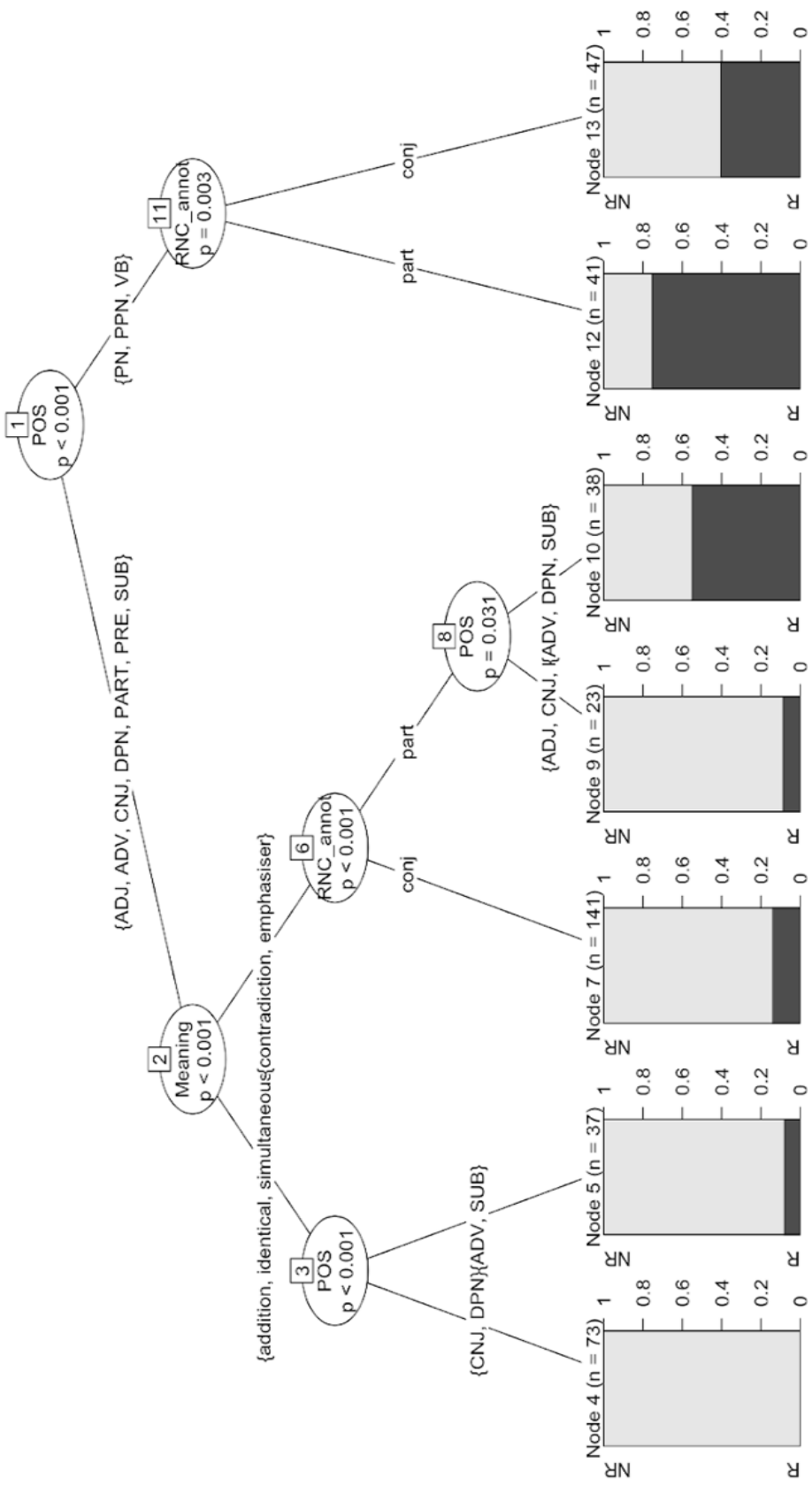

Figure 4.6: A cTree showing the optimal sorting of dataset 1 (all).

At the top of the cTree in node 1 is "POS", which shows that "POS" is the first predictor with the best split when "Replaceability" is considered. The first predictor is commonly interpreted as the most important predictor, although this is not always the case (Baayen et al. 2013, p.265). To the left there are several other splits, and to the right there is one other split. I will briefly summarise some of the nodes seen in Figure 4.6.

Node 4 (n=73) represents 18% of the data. It shows that when the "POS" is a conjunction or demonstrative pronoun, and the meaning is either ADDITION, IDENTICAL, SIMULTANEOUSLY or URGENCY, then *že* cannot be replaced with *ved'*.

Node 5 (n=37) represents 9% of the data and shows that when the "POS" is an adverb or a noun, and the meaning is either ADDITION, IDENTICAL, SIMULTANEOUSLY or URGENCY, then *že* cannot be replaced with *ved'* in 34 cases (91%).

Node 7 (n=141) represents 35% of the data and indicates that when *že* is tagged as a conjunction in the RNC and the meaning is either CONTRADICTION or EMPHASISER, *že* is not replaceable with *ved'* in 121 cases (86%).

Node 9 (n=23) represents 5% of the data and indicates that when the "POS" is an adjective, conjunction or particle, and *že* is tagged as a particle in the RNC, and the meaning is either CONTRADICTION or EMPHASISER, *že* is not replaceable with *ved'* in 21 cases (91%).

Node 10 (n=38) accounts for only 9.5% of the data but shows that when the "POS" is an adverb, demonstrative pronoun or noun, and *že* is tagged as a particle in the RNC, and the meaning is either CONTRADICTION or EMPHASISER, *že* is replaceable with *ved'* in 21 cases (55%).

Node 12 (n=41) represents 10% of the data and is also the node with the most concentrated cases where *že* is replaceable with *ved'*. Here *že* is replaceable with *ved'* in 31 cases (71%) when *že* is tagged as a particle in the RNC, and the "POS" is either pronoun, personal pronoun or a verb. The category "Meaning" is not relevant to this node.

Node 13 (n=47) represents 11% of the data. Here *že* is replaceable with *ved'* in 19 cases (40%) when *že* is tagged as a conjunction in the RNC, and the "POS" is either pronoun, personal pronoun or a verb. The category "Meaning" is not relevant to this node.

This summary of the cTree demonstrates that the part of speech to the left of *že* plays a more important role in influencing the replaceability of *že* with *ved'* than the meaning of *že*. In fact, whether the RNC has tagged *že* as a particle or a conjunction seems to play a more central role than the meaning of *že*.

4.4 Conclusion

The aim of this chapter was to identify which factors may influence the replaceability of *že* with *ved'*. I created a dataset based on 400 example sentences of *že* taken from the RNC. Using a Chi-square test, Cramér's V and Fisher test I proved that the null hypothesis (that no factors influence the replaceability of *že* with *ved'*) had to be rejected.
By focusing on the meaning of *že* and the part of speech to the left of *že* as potentially influential factors, the Chi-square test and Cramér's V indicated that these factors did seem to influence the replaceability of *že*.

To investigate this further I carried out a Logistic Regression Model with different reference levels to look at the interaction between different factors. Finally, I made a cTree. The results of the cTree analysis seem to indicate that the meaning of *že* is not as significant as the part of speech to the left of *že*. In fact, it appeared to indicate that the way *že* is tagged by the RNC was more decisive than the meaning in deciphering the replaceability of *že*.

The hypothesis that the meaning of *že* is associated with the replaceability of *že* with *ved'* does not seem to be as significant as the part of speech to the left of *že*. However, I will not discard the meaning of *že* as an important factor to the replaceability of *že* with *ved'*. I will, however, not look further at the way in which *že* is tagged in the RNC due to the scope of this thesis.

In this chapter I have investigated the replaceability of *že* with *ved'* from a statistical perspective. As noted in section 4.2.2 only one native speaker of Russian was asked to

evaluate the replaceability of *že* with *ved'* in the 400 examples. Combining the meaning of *že* with the part of speech to the left of *že* I wanted to observe how more native speakers react to *že* being replaced by *ved'*.

5 A questionnaire-based investigation of *že* and *ved'*

5.1 Introduction

In chapter three I proposed two radial categories for *že* and *ved'* where I presented the different meanings of these two lexemes and how they are translated in English. In my investigation I demonstrated cases where *že* and *ved'* appear to be independent of each other, as well as instances where they seem to share common properties. The polysemy of *že* and *ved'* is one obstacle for learners of Russian to overcome when using these types of lexemes. In chapter 4 I used statistical methods to examine what factors may influence the replaceability of *že* with *ved'* and recognised that the part of speech to the left of *že* played a more important role than the meaning of *že*. In this chapter I will expand on this observation in chapter 4, and consider the following:

1. Does context, in particular the context of the word preceding *že*, play a role in the appearance of *že* in a clause?
2. Based on the fact that *že* and *ved'* share some of the same semantic properties, is it possible to replace *že* with *ved'* based on context?

To investigate these questions, I carried out two small experiments using texts from the RNC to attempt to determine whether native speakers of Russian follow any specific patterns when using *že* and *ved'*.

As already stated in chapter 2 *že* is a clitic, does not bear stress and is therefore dependent on the preceding word for stress. Thus, *že* forms a single phonological word with that preceding word, and therefore *že* cannot be positioned word-initial in the sentence. *Ved'*, on the other hand, can appear as the first word in a sentence, as well as post-initially. For this reason, and due to the limitations of this thesis, I will limit this investigation to consider only instances where the position in the sentence allows both *ved'* and *že* to appear. This investigation will not consider cases where *ved'* can be found in sentence-initial position.

I will begin this chapter by presenting and discussing a small pilot experiment (section 5.2) that led to the development of the main focus study (section 5.3). The first experiment (henceforth Experiment 1) was a pilot experiment to observe where Russian native speakers

would place *že* or/and *ved'* in a sentence. Experiment 2 is a more comprehensive experiment where I observe native speakers' flexibility in using *že* and *ved'*.

5.2 Pilot experiment

Experiment 1 is a pilot experiment to this section and is valuable to discuss. Using the RNC I chose ten sentences whereby five contained *že* and five contained *ved'*. The main criterion for selecting a sentence was based on my own linguistic abilities in Russian so I was able to understand the sentences in question. It was fundamental that the sentences could grammatically contain both *že* and *ved'*. Therefore, sentences with fixed constructions such as "*takoj že*" and "*k tomu že*", to name but a few, were not included because *ved'* never functions as a synonym in these cases.

As previously discussed in chapter two (McCoy, 2003a, p.125), both *že* and *ved'* may function as a way for the speaker to activate knowledge that he/she believes the receiver has already. *Že* and *ved'* may also refer to something already previously stated in the text. To try to give participants of Experiment 1 the opportunity to gain a fuller picture of the context of the focus sentences, and therefore aid them in using their native speaker intuition to decide whether *že* and/or *ved'* should be placed in the selected sentence, two to three preceding sentences were given to the participants.

As part of Experiment 1 I was very interested in seeing how the participants rated the original sentence. I had two hypotheses:

> 1) The participants will always rate the original sentence from the RNC as *otlično* "great" and all other alternatives as *nevozmožno* "impossible", including the option that does not include *že* or *ved'*.
> 2) The participants will give varied responses based on the text as a whole as participants will interpret the focus of the sentences in different ways. This includes the variant without *že* and *ved'*.

The lexemes *že* and *ved'* add a sense of subjectivity to a sentence. *Že* can be interpreted as the speaker wishing to emphasise a particular aspect of a sentence, and *ved'* can be used by a

speaker to stress an event or idea that has previously been mentioned. Therefore, I predict that the option available to the participants that does not include either *že* or *ved'* will be just as viable an alternative as the other options.

In preparation for the pilot experiment, a native speaker who did not take part in the experiment assisted in deciding the possible insertion points of *že* and *ved'*. It was important that only the lexemes in question, *že* and *ved'*, would change position in the sentence; all other words in the sentence remained in the same position as in the original sentence.

It is important to state that all 11 participants in Experiment 1 were Russian native speakers and researchers within the field of linguistics. This may therefore have influenced the way they answered the questions in terms of their sensitivity to grammar. All participants were recruited by personal communication and the experiment was carried out online.

5.2.1 Limitations

The development of Experiment 1 was the springboard to Experiment 2, which will be discussed in section 5.3, and is therefore valuable to discuss. From a scientific perspective Experiment 1 has its limitations. The results of Experiment 1 are not statistically significant due to the low number of participants (eleven). Other variables were also not controlled for, such as the type of text chosen or its formality. These factors may have influenced the use and position of both *že* and *ved'* in the sentences. Also, there were no filler sentences. This was however a conscious decision. As the examples used were quite long, my main goal was that the participants would complete all the questions. In section 5.2.2 I present three examples from Experiment 1. As this was a pilot experiment and due to lack of funding, a limited survey website (SurveyMonkey) was used. This meant that only a limited amount of data could be extracted, and I therefore only extracted the three most relevant examples.

5.2.2 Data analysis

Example (91) shows one of the ten texts given to the participants. The participants had to decide on the position of *že* and/or *ved'* in a sentence. The target sentence was indicated with three red question marks: "**???**". Participants were then given alternative answers and asked to rate each alternative answer in one of three ways: "*otlično*" (great), "*dopustimo*"

(acceptable) or "*nevozmožno*" (impossible). One alternative answer always excluded both *že* and *ved'.* Other possible options were given with *že, ved'* and if grammatically possible, both *že* and *ved'*. The original sentence from the RNC was always included as an alternative.
In example (91) alternative a) was the original example from the RNC where *ved'* was positioned sentence-initial. Alternative b) removed *ved'* and added *že*, and to make the sentence a grammatically correct option *že* was positioned after the pronoun *èto*. Alternative c) removed both *ved'* and *že* and alternative d) included both *ved'* and *že,* whereby *ved'* was positioned as in alternative a) and *že* was positioned as in alternative b).

(91) *Stiven Spilberg. Každyj raz, kogda slyšiš' ego imja, na um srazu prixodjat takie velikie fil'my, kak «Čeljusti», «Nazad v buduščee», «Park Jurskogo perioda», «Spasti rjadovogo Rajana» i dr. I v očerednoj raz Spilberg dokazal, čto možet snimat' ne tol'ko vysokobjudžetnye blokbastery, no i sdelat' xorošee, kačestvennoe kino za malye den'gi.* ***???****. No čto kasaetsja samogo fil'ma, to on namnogo dorože, čem ego bjudžet.*

The alternative answers here included:

a) ***Ved'*** *dlja amerikanskogo fil'ma 50 mln $ èto ne den'gi.*
b) Dlja amerikanskogo fil'ma 50 mln $ èto ***že*** *ne den'gi.*
c) Dlja amerikanskogo fil'ma 50 mln $ èto ne den'gi.
d) ***Ved'*** *dlja amerikanskogo fil'ma 50 mln $ èto* ***že*** *ne den'gi.*

The aim was that all participants would rate all the available alternatives. Either due to technical difficulties or a lack of understanding of the task, not all participants answered every question. In the example given all eleven participants rated alternative a), ten participants rated alternative c), and nine participants rated alternatives b) and d). However, as this is a pilot experiment, I decided not to disregard alternatives that were not answered.

In the case of example (91) the results were nevertheless interesting. All participants rated alternative a), the original sentence from the RNC, as *otlično*. For all the other alternatives only one or two participants rated them as *otlično*. The participants tended to judge alternatives b) – d) as *dopustimo*, but there were some participants who rated alternatives b) –

d) as *nevozmožno*. In fact, for alternative c), the sentence where neither *ved'* or *že* were included, four participants rated it as *nevozmožno*.

	Отлично	Допустимо	Невозможно	Responses
Ведь для американского фильма 50 млн $ это не деньги. Count Row %	11 100.0%	0 0.0%	0 0.0%	11
Для американского фильма 50 млн $ это же не деньги. Count Row %	1 11.1%	6 66.7%	2 22.2%	9
Для американского фильма 50 млн $ это не деньги. Count Row %	2 20.0%	4 40.0%	4 40.0%	10
Ведь для американского фильма 50 млн $ это же не деньги. Count Row %	2 22.2%	6 66.7%	1 11.1%	9
Totals Total Responses				11

Figure 5.1: Figure showing the results of example (91) of Experiment 1

One participant used the comments section of the questionnaire to explain the reasoning behind their choice:

> "*Ved'* is a must because it links this sentence to the previous one ... *že* is a little more colloquial ... The two middle sentences are linguistically correct, but do not really fit in".

This comment supports the idea that this topic does not have one concrete answer. Whilst one participant does not believe that alternatives b) and c) are acceptable options, respectively seven and six participants deemed these alternatives as either *otlično* or *dopustimo*.

Based on the observation that alternative a) was considered as the best answer, it is notable to compare alternative a) to alternative d). Whilst both examples contain *ved'* in the same position in the sentence, alternative d) also contains *že*. It seems that the addition of *že* affected the acceptability of alternative d) for most of the participants, with one participant rating it as *nevozmožno*.
Another example which gave an interesting result can be seen in example (92).

(92) *Nikolaev i Golubovič obvinjalis' v tom, čto oni jakoby nanesli neskol'ko udarov drevkom flaga sotrudniku milicii. Pri takix xarakterizujuščix dannyx — a u oboix oni otličnye— nikogo i nikogda ne arestovyvajut i ne sažajut.* ***???***.

The alternative answers here were:

a) Zdes' ***že*** *Tverskoj sud Moskvy dal im po tri goda lišenija svobody!*

b) ***Ved'*** *zdes' Tverskoj sud Moskvy dal im po tri goda lišenija svobody!*

c) Zdes' Tverskoj sud Moskvy dal im po tri goda lišenija svobody!

In example (92) the participants were asked to rate only three examples because it was decided that it was not grammatically possible to use both *že* and *ved'*. Alternative a) was the original sentence from the RNC where *že* was used. Alternative b) had *že* removed, and instead *ved'* was added. In this case *ved'* changed position and was sentence-initial. Finally alternative c) included neither *že* nor *ved'*.

All eleven participants rated alternative a), and ten participants rated alternatives b) and c). Ten participants rated alternative a) as *otlično* and one participant rated it as *dopustimo.* All participants reacted negatively to alternative b). This indicates that *ved'* is not acceptable in this sentence, and one participant even justified their choice by commenting:

"[There is] a different meaning of *že* here, so *ved'* cannot be substituted*".*

To expand on the participant's comment, it can be interpreted that the use of *že* in alternative a) is as an EMPHASISER, where the focus is on the word *zdes'* "here". In this case, a translation of *že* into English could consist of an emphasis on the word when spoken. Whilst *ved'* can also function as an EMPHASISER, it acts more as a CONTRADICTION in this case, with a possible English translation of "but" or "however". Therefore, this gives a different meaning to the sentence which does not fit with the rest of the given text.

It is interesting to note how important the participants regard *že* in this sentence. Whilst six out of ten participants rated alternative c) as *dopustimo*, none rated the alternative that did not

include *že* or *ved'* as *otlično*. This example further indicates that *že* plays an important role in this sentence.

	Отлично	Допустимо	Невозможно	Responses
Здесь же Тверской суд Москвы дал им по три года лишения свободы! Count Row %	10 90.9%	1 9.1%	0 0.0%	11
Ведь здесь Тверской суд Москвы дал им по три года лишения свободы! Count Row %	0 0.0%	0 0.0%	10 100.0%	10
Здесь Тверской суд Москвы дал им по три года лишения свободы! Count Row %	0 0.0%	6 60.0%	4 40.0%	10
Totals Total Responses				11

Figure 5.2: Figure showing the results of example (92) of Experiment 1

The final example from Experiment 1 that will be presented here can be seen in example (93).

> (93) *"Vot vidiš', Ivanovna, kak xorošo teper' v apparatnoj. Nikto ne otvlekaetsja na tvoju malyšnju". A ja emu govorju: "Da, da, konečno, Nikolaj Grigor'evič". A on govorit:* ***???****. Ja govorju: "Ja ubrala ved' uže". A potom kak-to večerom s raboty idu, smotrju-u pod"ezda zjateva mašina stoit. Ja bystree pošla.*

The alternatives in example (93) were:

> *a) "**Ved'** ty **že** sama znaeš'-u nas ne položeno".*
> *b) "Ty **ved'** sama znaeš'-u nas ne položeno".*
> *c) "**Ved'** ty sama znaeš'-u nas ne položeno".*
> *d) "Ty **že** sama znaeš'-u nas ne položeno".*
> *e) "Ty sama znaeš'-u nas ne položeno".*

Example (93) is an interesting due to the number of alternative responses. Alternative a) uses both *že* and *ved'*, with *ved'* in sentence-initial position. Alternatives b) and c) both use only *ved'*, but with a syntactic difference: *ved'* is non-initial in alternative b) and sentence-initial in alternative c). Alternative d) is the original sentence from the RNC where *že* was used, and

alternative e) excludes both *že* and *ved'*. In example (93) eleven participants rated alternative d), and ten participants rated all of the remaining alternatives. This could indicate that one participant misunderstood the task and only rated the sentence they believed to be correct.

	Отлично	Допустимо	Невозможно	Responses
"Ведь ты же сама знаешь-у нас не положено". Count Row %	7 70.0%	2 20.0%	1 10.0%	10
"Ты ведь сама знаешь-у нас не положено". Count Row %	8 80.0%	2 20.0%	0 0.0%	10
"Ведь ты сама знаешь-у нас не положено". Count Row %	6 60.0%	4 40.0%	0 0.0%	10
"Ты же сама знаешь-у нас не положено". Count Row %	9 81.8%	2 18.2%	0 0.0%	11
"Ты сама знаешь-у нас не положено". Count Row %	7 70.0%	3 30.0%	0 0.0%	10
Totals Total Responses				11

Figure 5.3: Figure showing the results of example (93) of Experiment 1

In this example the participants reacted positively to almost all examples. In fact, a single participant rated one alternative (alternative a)) as *nevozmožno*. 82% of participants rated alternative d), the original sentence from the RNC, as *otlično*. The most interesting observation with example (93) is that overall, the participants rated all alternatives as *otlično,* and in fact no participant rated the other alternatives as *nevozmožno*. The results of example (93) are visualised below in Figure 5.4 to further highlight the similarity in distribution of answers by the participants.

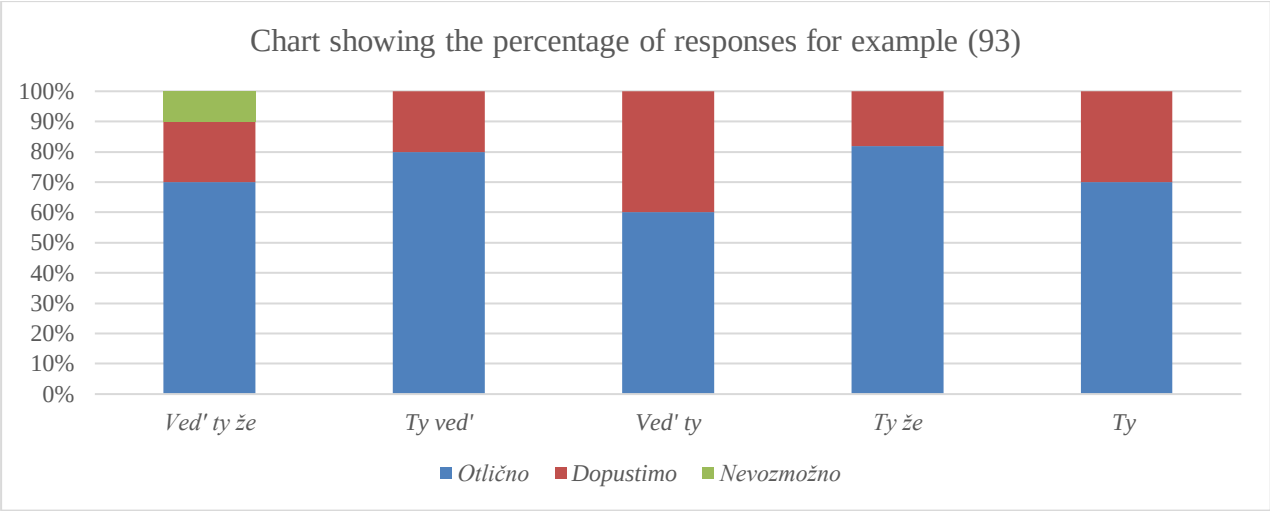

Figure 5.4: A bar chart representing the distribution of the participants' responses to the different alternatives in example (93)

The results of this example further demonstrate the flexibility of the lexemes *že* and *ved'*. It is also interesting to highlight alternatives b) and d). In alternative b) *ved'* appeared in second position and in alternative d) *že* appeared in second position. It appears that *že* and *ved'* act synonymously in this case as the participants rated these alternatives almost identically. The switch in word order in alternatives b) and c) also seemed to make little difference for the participants. Finally alternative e) could indicate that *že* is not as vital to this sentence as was the case of example (93), because seven out of ten participants rated the alternative that contained neither *že* or *ved'* as *otlično*, and three as *dopustimo*.

5.3 From pilot experiment to questionnaire

The pilot experiment (Experiment 1), demonstrated by the examples discussed in section 5.2.2, shows that participants did not always respond with the same choice of word order as the original example from the RNC. As clearly shown by example (93) the participants interpreted the examples in different ways.

In Experiment 1 the participants were given relatively large portions of text to maximise the context the participants would have for evaluating the sentences. But is a great amount of context paramount to deciding whether a sentence can contain *že* or *ved'*? Can other factors, such as the part of speech to the left of *že*, i.e. the word that *že* is dependent on, play a role?

Can the submeanings of *že*, as presented in my proposal for a radial category for *že* in chapter 3, also play a role?

To attempt to answer these questions and shed more light on this topic, I created a questionnaire based on the dataset from chapter 4.

5.4 Method

5.4.1 How I collected my data

For this questionnaire I chose example sentences from the dataset I developed and analysed in chapter 4. Two variables in this dataset included the part of speech to the left of *že* (henceforth POS), and the meaning of *že*. The POS variable was tagged manually, but in cases of ambiguity such as the lexeme *kak* (adverb and conjunction), the original tag in the RNC was consulted. The meaning variable was based on the radial category for *že* that I developed in chapter 3. There were nine options for POS and six meanings to choose from. Due to the scope of this thesis, it was not realistic to compare every option, as this could have amounted to fifty-four different combinations. The three POS and meanings with the highest frequency in the dataset were selected. For POS I selected NOUN, ADVERB and DEMONSTRATIVE PRONOUN. For meaning I selected EMPHASISER, IDENTICAL and CONTRADICTION. I decided to use a Latin Square Design format, a grid or matrix containing the same number of rows and columns (Richardson, J. T. E., 2018). The goal was to focus on sentences that involve all combinations of the three POS and three meanings. Unfortunately, two combinations were not present in my dataset: NOUN/ IDENTICAL and DEMONSTRATIVE PRONOUN/CONTRADICTION, as shown in Table 5.1, where green represents combinations where sentences were available in the dataset, and red represents combinations that were not present in this dataset. Whilst the combinations NOUN/IDENTICAL and DEMONSTRATIVE PRONOUN/CONTRADICTION are not attested for in my dataset, I do not claim that these combinations are not possible.

	EMPHASISER	IDENTICAL	CONTRADICTION
NOUN			
ADVERB			
DEMONSTRATIVE PRONOUN			

Table 5.1: A Latin Square Design showing the combinations to be used in the questionnaire

For each of the seven combinations that are attested in my data, I randomly selected six sentences, giving a total of forty-two sentences. I used six sentences for each combination because for the combination ADVERB/EMPHASISER, there were a total of six sentences available.

I used the survey website www.survio.com to carry out my questionnaire. The target group for this questionnaire was Russian native speakers. All participants were asked three questions before taking part in the questionnaire. The first question was whether Russian was the participant's native language. Originally, I intended on developing and carrying out this questionnaire in Russia. This would have guaranteed responses from native speakers, as well as given me the opportunity to analyse the use of *že* in discourse. This was not possible due to the outbreak of the COVID-19 pandemic, and therefore the questionnaire I made was shared via social media channels, as well as by asking personal contacts to participate.

The second question asked participants to state their age. All participants were over the age of 18. I was able to obtain the age ranges of the questionnaire:

- Age 18-29 (47,7%)
- Age 30-44 (40,9%)
- Age 45-59 (11,4%)

Finally, I asked participants about their gender, with the options male, female, other, do not want to answer.

- 33 females (75%)
- 11 males (25%)

5.4.2 Limitations and questionnaire instructions

In total forty-five participants completed the questionnaire. However, one participant was removed when analysing the results because they answered "no" to the question "Is Russian your mother tongue?". It is likely this participant was Ukrainian or Belarusian and had other

motivations for answering “no”, but this could not be confirmed, and therefore I analysed the results of the remaining forty-four participants. One participant did not answer eight of the questions, however this still meant that the participant completed 80% of the task, and therefore this participant’s answers were retained.

For the questionnaire participants were presented with forty-two sentences. *Že* was removed from these sentences, but the clause that originally contained *že* was shown in bold. Participants were made aware that some sentences contained only one clause, and therefore the entire sentence was highlighted in bold. The instructions were presented to the participants in the following manner:

- *Pročitajte 42 korotkix predloženija. V každom predloženii est' čast', vydelennaja polužirnym šriftom. V nekotoryx primerax vsë predloženie budet vydeleno polužirnym šriftom.*
- *Vam neobxodimo rešit', možno/objazatel'no nužno/nel'zja dobavit' časticu «že» v ljubom meste v vydelennoj časti predloženija.*
- *Esli Vaš otvet «Nel'zja», to Vy dolžny perejti k sledujuščemu predloženiju.*
- *Esli Vaš otvet «Možno» ili «Objazatel'no nužno», to Vy dolžny rešit' možno li v ètoj že samoj vydelennoj časti predloženija upotrebit' časticu «ved'».*

For the first part of the question participants were asked to read the sentence and decide whether they could add *že* in the highlighted clause by answering the question “*Možno li dobavit' že?*” (“Can you add *že*?”). The response options were *možno* “yes”, *objazatel'no nužno* “you must” and *nel'zja* “no”. If the participant answered *možno* or *objazatel'no nužno*, then they were asked whether they could add *ved’* in the same highlighted clause. If the participant answered *nel'zja*, they were instructed to go on to the next question.

For the question about *ved’* the participants were asked to rate the question *Možno li dobavit' ved’?*” (Can you add *ved’*?”) with following options: *soglasen/soglasna* “agree”, *častično soglasen/soglasna* “partly agree”, and *ne soglasen/soglasna* “do not agree”.

Despite the specific instructions to move on to the next question if the participants answered *nel'zja* for the question about adding *že*, most participants answered the question about adding *ved'*.

Human error meant that one sentence was left out and another sentence was repeated. The repeated sentence first appeared as question 11, and then as question 32, and the results were different. This indicates that participants have not understood the task, or, as suspected, the use of *že* and *ved'* can be so subjective that participants interpreted the sentence differently the second time around. However, this only happened in the case of one sentence, allowing me to make only this assumption. Therefore, question 32 was not included in the analysis. This means that a total of 41 sentences were analysed and the combination NOUN/CONTRADICTION included five sentences.

As with the pilot experiment no filler sentences were included. I madet his decision because I did not want the survey to be too long: participants were not rewarded for their participation and therefore a simple and short task was a priority to get the most amount of data. Following the rule of thumb of the Central Limit Theorem (King et al., 2011, p.176) in order for the sampled distribution of the data from this analysis to approximate a normal distribution I needed more than thirty participants to complete the task. This led me to prioritise participation quantity over other factors, such as filler sentences.
Two participants commented that the survey was difficult to do because of a lack of context. This was to be expected as *že* and *ved'* can often refer to something or an event that has previously been mentioned in the text (McCoy, 2003a), and the participants were given no other context than the one sentence from the text.

5.5 Data analysis and discussion

In the following section I will present the findings from the questionnaire. I will discuss each combination, and focus on specific sentences that gave insightful results. At the end of each subsection, I show the combined results for that combination and offer general observations.

5.5.1 NOUN/EMPHASISER

Figure 5.5 shows a bar chart of the results for question one. In this sentence the original highlighted clause with *že* was "*Teatr* ***že*** *dolžen byt' čestnym po otnošeniju k sebe i k svoemu delu, ...*". The orange bars show the respondents' answers to the question "*Možno li dobavit' že?*" ("Can you add *že*?"). The blue bars show the respondents' answers to the question "*Možno li dobavit' ved'?*" ("Can you add *ved'*?"). The bar chart shows that respondents decided that both *že* and *ved'* can be added to the highlighted clause. Interestingly a total of 84.1% of participants said that *že* could be added or had to be added, and 84.1% of participants agreed or partially agreed that *ved'* could be added to the highlighted clause.

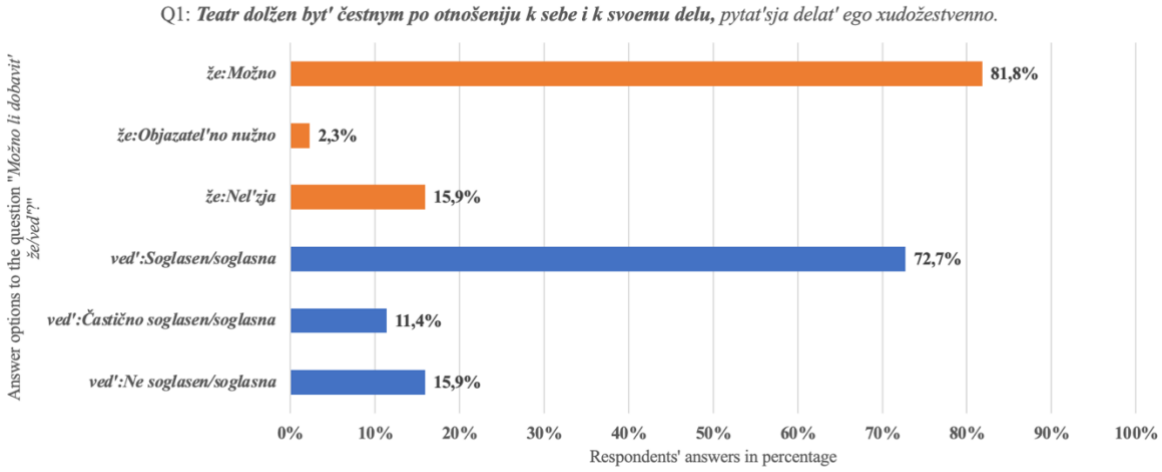

Figure 5.5: A bar chart showing the results of the participants' responses to question 1 of the questionnaire

The highlighted clause in question two, "***Antona zabyla***", was very short and presented varied results, as shown in Figure 5.6. It is noteworthy to mention that for the question about *že* over two-thirds of respondents said that *že* could be added, yet 27.3% of participants decided that it was not possible to add *že* anywhere in the highlighted clause. For the question about *ved'* the results were more mixed. Less than half of participants fully agreed that *ved'* could be added in the clause *Antona zabyla*, and a third of participants did not agree with this.

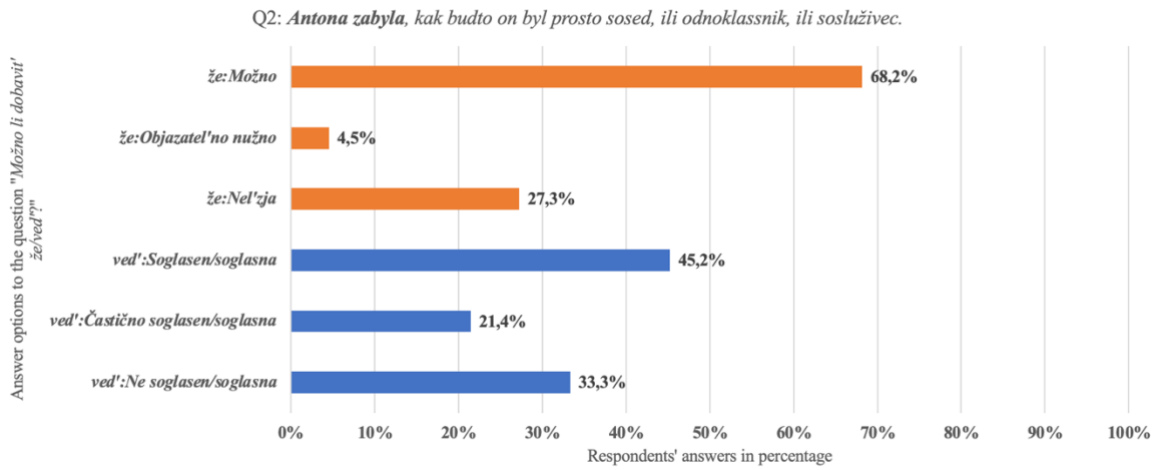

Figure 5.6: A bar chart showing the results of the participants' responses to question 2 of the questionnaire

Another interesting observation from the sentences with the combination NOUN/EMPHASISER was question 9, shown in Figure 5.7. For this question there was one clause, and therefore the entire sentence was highlighted. In general participants reacted negatively, with 65.9% of respondents stating that *že* cannot be added to the clause, and 78% of respondents did not think that *ved'* could be added. However, 34.1% of respondents agreed that *že* could be added, showing a stark contrast in the way participants interpreted the use of *že* in this sentence.

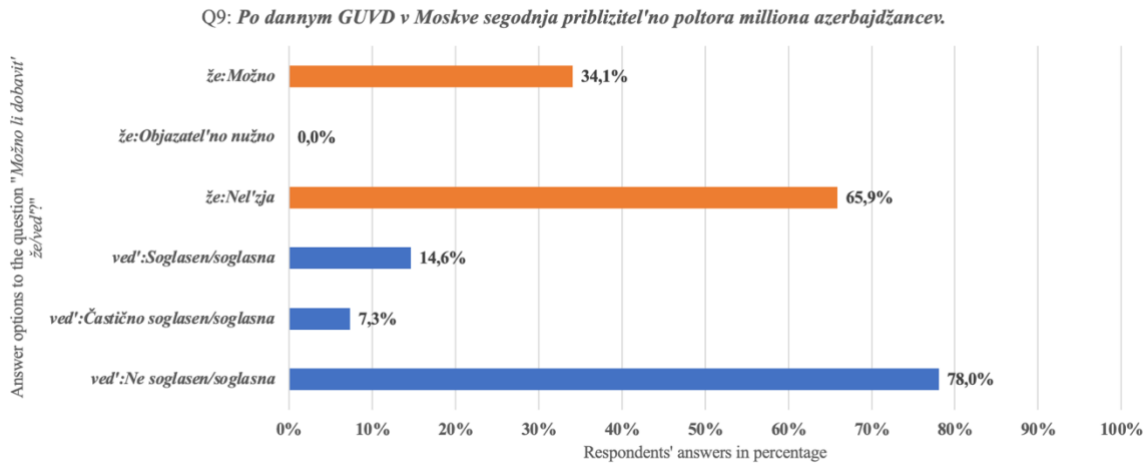

Figure 5.7: A bar chart showing the results of the participants' responses to question 9 of the questionnaire

Figure 5.8 shows the combined results of the question *Možno li dobavit' že?* for the sentences with the combination NOUN/EMPHASISER. In general, the participants were positive towards adding *že* to the clause in question. Question 9 was the only sentence in this combination

where this was not completely the case and is the only sentence where more than 30% of participants responded with *nel'zja*. On the whole respondents did not highly evaluate *že* as being mandatory in these sentences, with 0% of participants responding with *objazatel'no nužno* in questions 8 and 9.

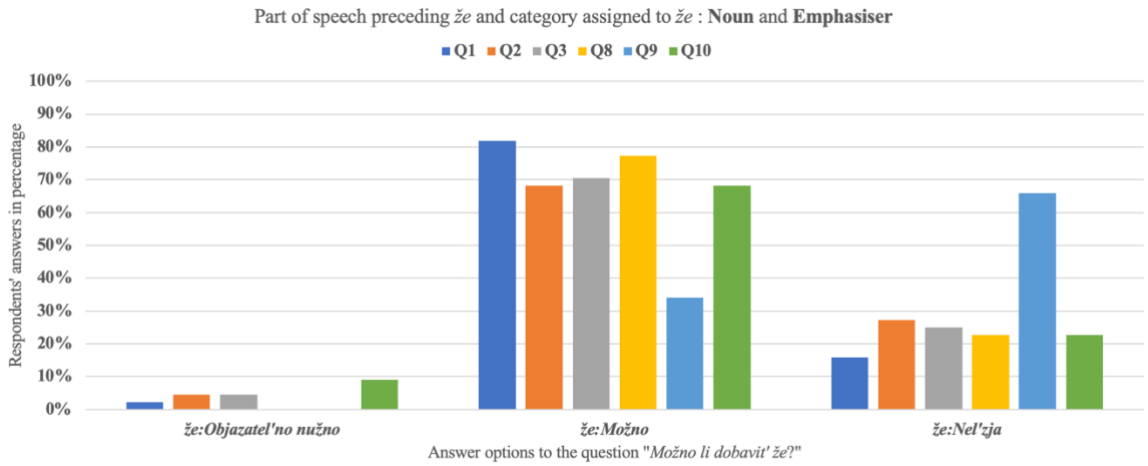

Figure 5.8: The combined results of the question Možno li dobavit' že? for the sentences with the combination NOUN/EMPHASISER

Figure 5.9 shows the results of the same sentences with the combination NOUN/EMPHASISER and the question *Možno li dobavit' ved'?*. Compared to the way participants responded to adding *že* to the clause, the response for *ved'* was slightly more varied, with the participants responding quite evenly to *soglasen/soglasna* and *ne soglasen/soglasna*. Questions 8 and 9, however, show conflicting trends in comparison with the other sentences. Whereas under 40% of participants rated questions 1, 2, 3 and 10 as *ne soglasen/soglasna*, for questions 8 and 9 the results were 72.7% and 78% respectively.

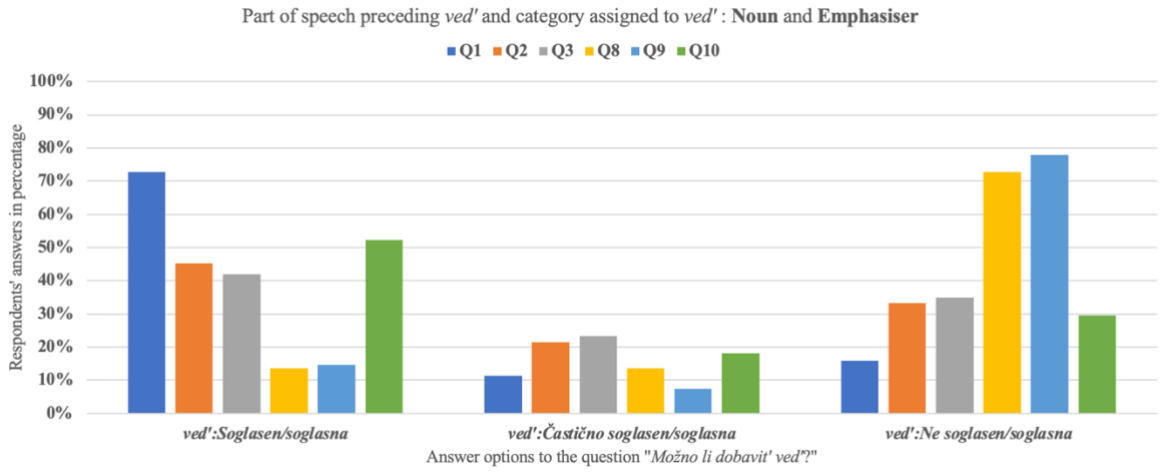

Figure 5.9: The combined results of the question Možno li dobavit' ved'? for the sentences with the combination NOUN/EMPHASISER

For sentences with the combination NOUN/EMPHASISER participants generally favoured *že* more than *ved'*. It should be pointed out that the responses demonstrate that participants did not always completely reject *ved'* in cases where they decided that *že* could be added to the highlighted clause. In general, the percentage of respondents who valued the addition of *ved'* as *ne soglasen/soglasna* was not significantly higher than the percentage of respondents who decided the addition of *že* to the clause was *nel'zja*.

5.5.2 NOUN/CONTRADICTION

As explained in section 5.4.2 a human error led to one sentence being repeated twice and therefore one sentence was absent from the questionnaire. The absent sentence belonged to the combination NOUN/CONTRADICTION, which means for this combination there are only five sentences, and not six as originally intended. Figure 5.10 shows responses to question 4. The most notable observation for this question is the way in which respondents answered the question about *že*. It is striking that whilst almost half of participants agree that *že* can be added to the highlighted clause (45.5%), a further 47.7% believe that *že* must be added. For *ved'* more than half of respondents did not agree that *ved'* could be added, and just over 25% fully agreed that *ved'* could fit in the clause.

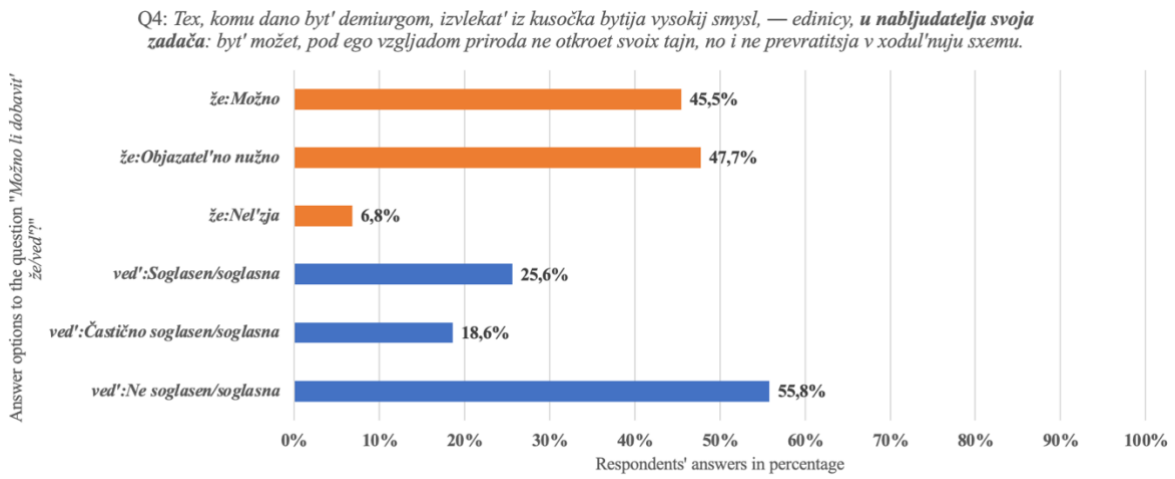

Figure 5.10: A bar chart showing the results of the participants' responses to question 10 of the questionnaire

Figure 5.11 shows the results of question 7 and presents different results to question 4. Whilst more participants agreed that *že* could be added to the highlighted clause in question 7 than in question 4, no respondents deemed *že* necessary in question 7, unlike in question 4. These conflicting results indicate that participants reacted differently to these two sentences, despite them both belonging to the NOUN/CONTRADICTION combination. Further comparing the results of these two questions, the participants reacted contradictorily when asked whether *ved'* can be added to the highlighted clause. Whereas in question 4 the majority response was *ne soglasen/soglasna* "do not agree", 60.5% of respondents answered with *soglasen/soglasna* "agree" in question 7. Therefore, participants were generally positive to adding *že* or *ved'* to the highlighted clause in question 7.

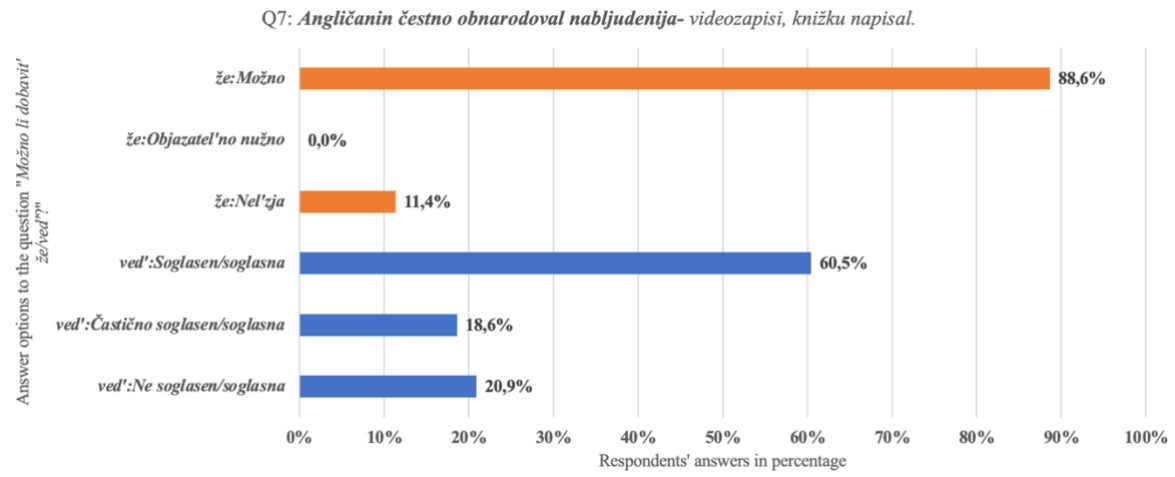

Figure 5.11: A bar chart showing the results of the participants' responses to question 7 of the questionnaire

Figure 5.12 shows another example from the combination NOUN/CONTRADICTION. Similar to question 4, question 12 shows competing findings for the responses *možno* and *objazatel'no nužno* in the question about *že*. The distinct observation in question 12 is the responses to whether *ved'* can be added to the highlighted clause, where 90.7% of participants answered with *ne soglasen/soglasna*.

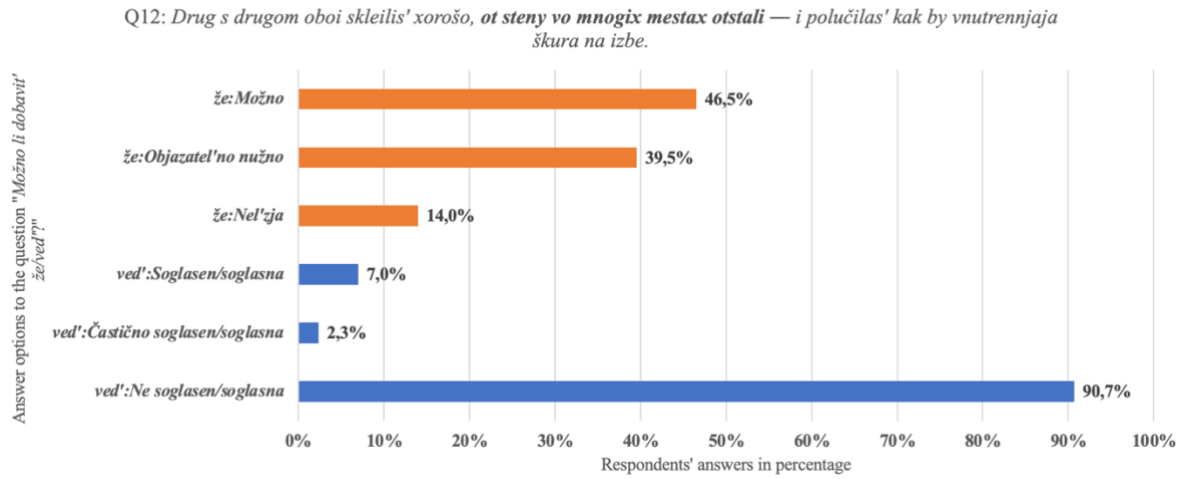

Figure 5.12: A bar chart showing the results of the participants' responses to question 12 of the questionnaire

The combined results for the results of *že* in Figure 5.13 show a similar trend to the results of the combination NOUN/EMPHASISER in Figure 5.8. Generally, respondents reacted positively to *že* being added to the highlighted clause, and in fact the response *objazatel'no nužno* was more prevalent here. In comparison to Figure 5.8 the response *nel'zja* was less widespread.

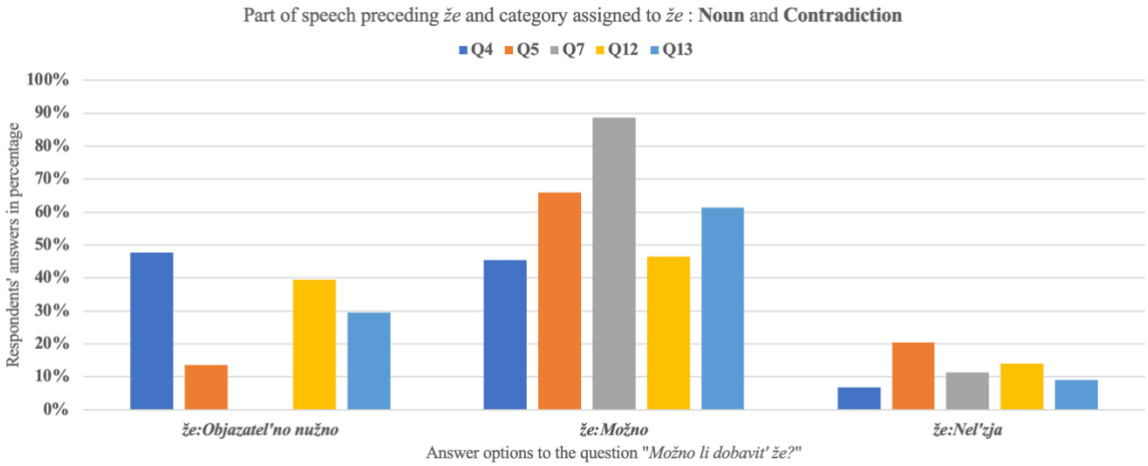

Figure 5.13: The combined results of the question Možno li dobavit' že? for the sentences with the combination NOUN/CONTRADICTION

Figure 5.14 shows the combined results for *ved'*. Apart from the one outlier, question 7, these findings show that participants did not agree that *ved'* could be added to sentences where the combination was NOUN/CONTRADICTION.

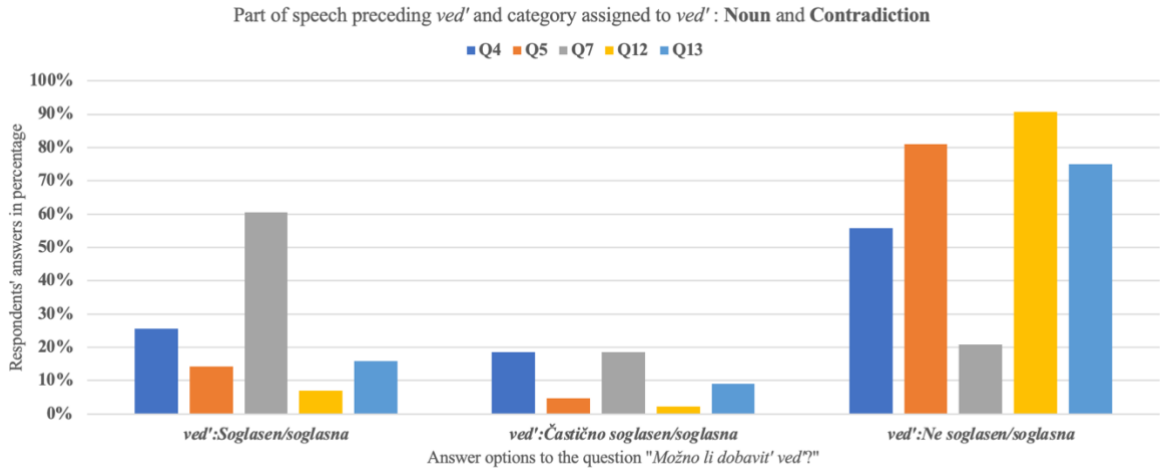

Figure 5.14: The combined results of the question Možno li dobavit' ved'? for the sentences with the combination NOUN/CONTRADICTION

5.5.3 ADVERB/EMPHASISER

Figure 5.15 shows the respondents' answers for question 6. The positive results for *že* were not surprising in this case. It is not uncommon for the adverb *opjat'* to be followed by *že*. A search in the RNC shows the construction *opjat' že* appears 5,445 times. This is significantly

greater than a search for *opjat' ved'* or *ved' opjat'*, which appeared thirty-two and ninety-six times respectively. It is therefore interesting that the results for *ved'* were so conflicting in this example, with 40.9% of participants agreeing that *ved'* can be added, and 45.5% of participants disagreeing. As participants were not asked to decide the insertion point of *že* and *ved',* they may have valued the sentence based on other words, such as the noun *učebnikov*.

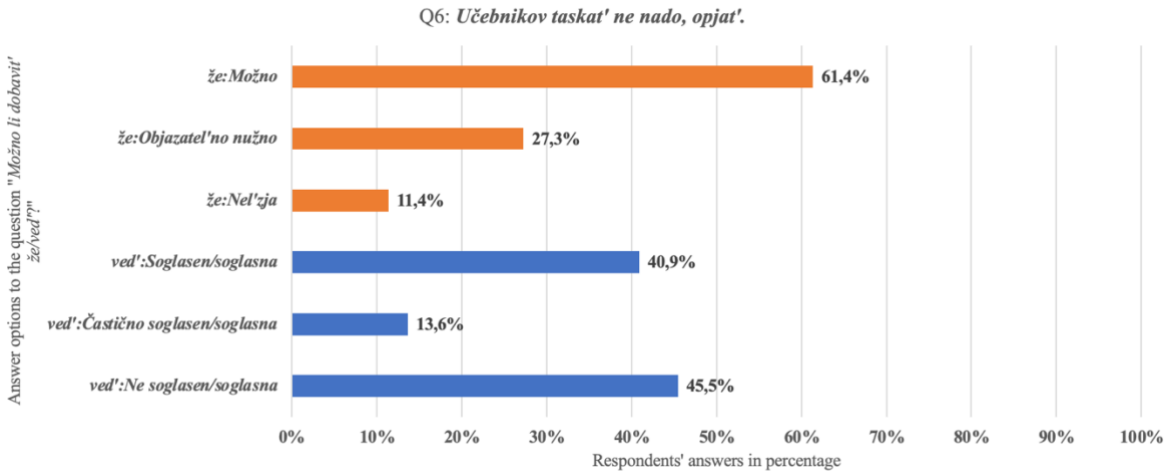

Figure 5.15: A bar chart showing the results of the participants' responses to question 6 of the questionnaire

Figure 5.16 shows how the participants answered question 14. Despite 36.4% of participants stating that *že* can be added to the sentence this question is interesting to report because the participants in general rejected adding either *že* or *ved'*, with 63.6% rejecting *že* and 58.5% of respondents rejecting *ved'*.

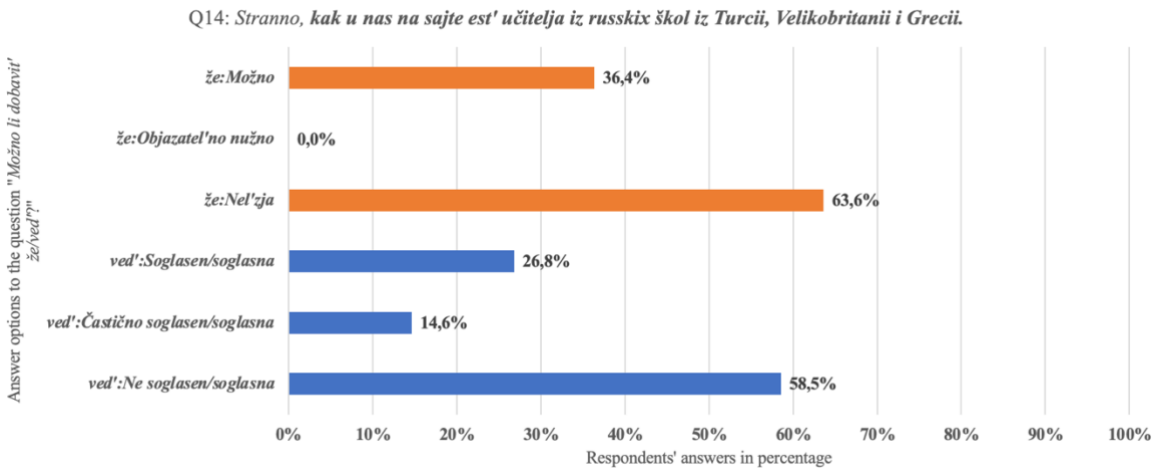

Figure 5.16: A bar chart showing the results of the participants' responses to question 14 of the questionnaire

Figure 5.17 shows the results for how participants answered the question "*Možno li dobavit' že*?" for the combination adverb/EMPHASISER. On the whole participants responded positively the addition of *že*, although compared to the combinations NOUN/EMPHASISER and NOUN/CONTRADICTION (Figures 5.8 and 5.13), the response *nel'zja* was more prominent here.

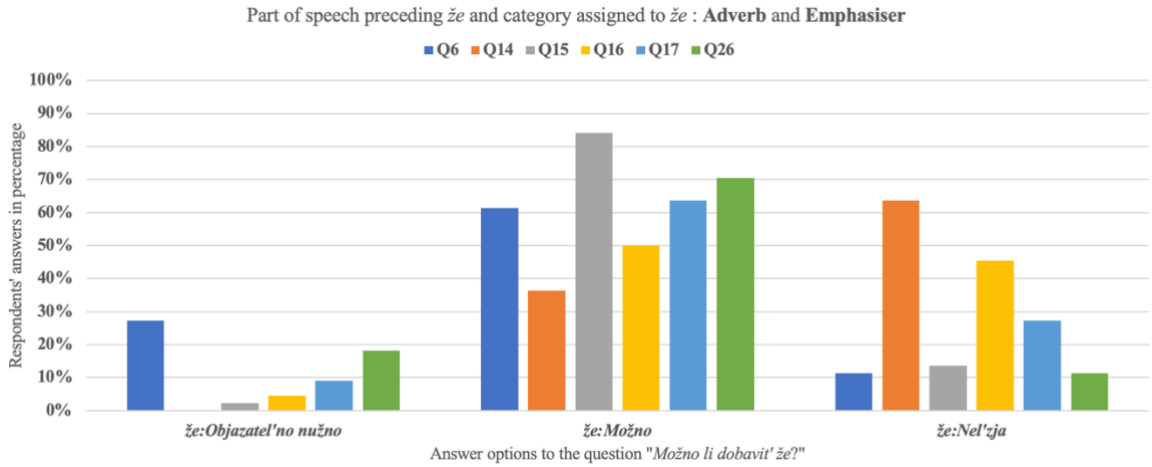

Figure 5.17: The combined results of the question Možno li dobavit' že? for the sentences with the combination ADVERB/EMPHASISER

The results of the participants' response to adding *ved'* to the sentences are combined in Figure 5.18, with little to no significant findings to report. It can be concluded that *ved'* was predominantly favoured by participants for the combination ADVERB/EMPHASISER.

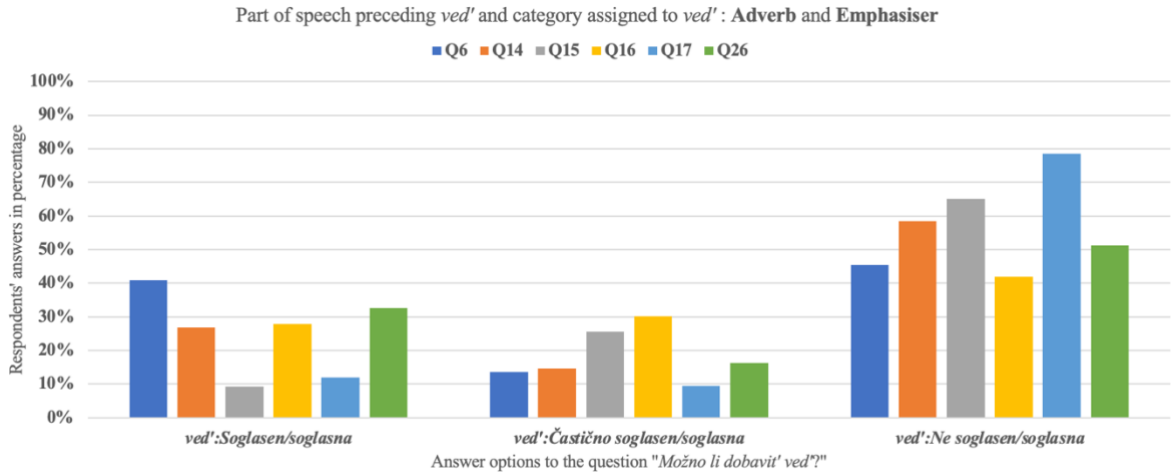

Figure 5.18: The combined results of the question Možno li dobavit' ved'? for the sentences with the combination ADVERB/EMPHASISER

5.5.4 ADVERB/IDENTICAL

Figure 5.19 shows the results for question 21. The findings for both *že* and *ved'* seem striking at first glance. 75% of respondents decided that *že* must be added to the highlighted clause, whereas 97.7% did not agree that *ved'* could be added. Although the participants were not given this information, the original sentence placed *že* after the adverb *stol'ko*. A brief search in the RNC showed 5,964 cases of the construction *stol'ko že*, compared to 156 instances of *ved' stol'ko* and 8 of *stol'ko ved'*. In addition to this, the construction *stol'ko že, skol'ko* "(just) as much/many as" appear 967 times in the RNC. This gives reason to speculate that participants interpreted the insertion point of *že* the same as the original.

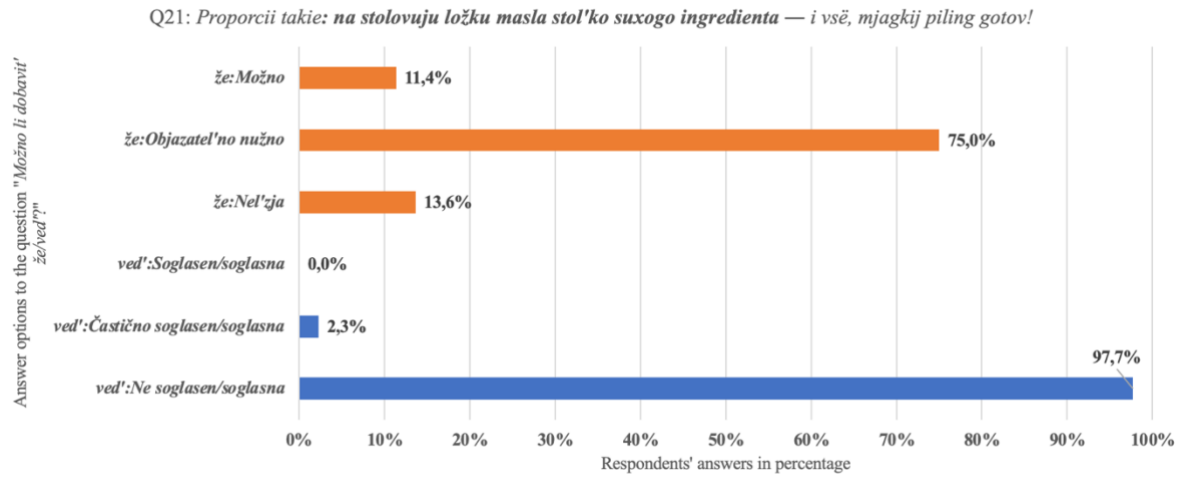

Figure 5.19: A bar chart showing the results of the participants' responses to question 21 of the questionnaire

In contrast to question 21 (Figure 5.19), the results of question 31 (Figure 5.20) are more positive to adding *ved'* to the highlighted clause. In general participants once again preferred *že* with 65.9% of respondents stating that *že* could be added and 27.3% claiming that *že* must be added. However, in comparison to 0% in question 21, 53.5% of participants agreed that *ved'* could be added to question 31, and only 27.9% did not agree.

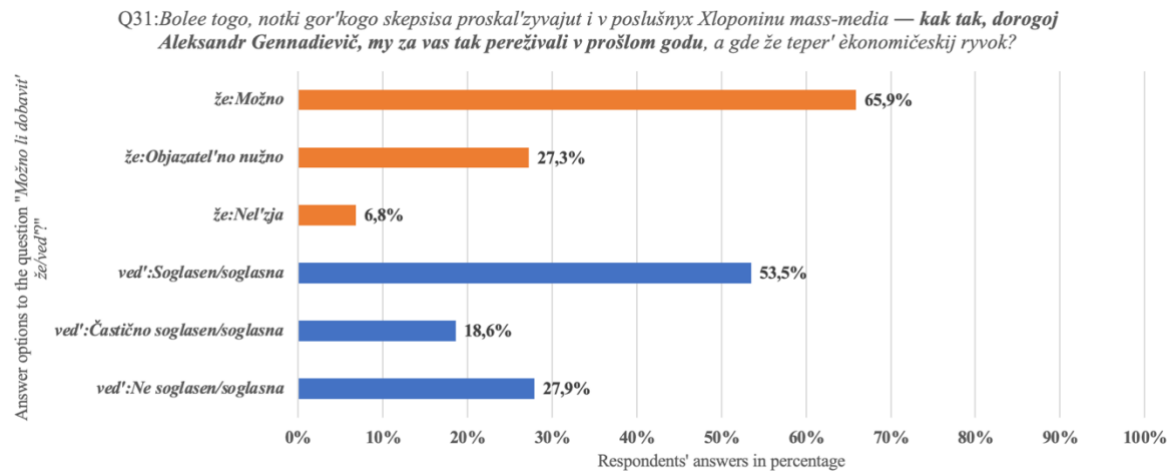

Figure 5.20: A bar chart showing the results of the participants' responses to question 31 of the questionnaire

The results of *že* in Figure 5.21 represent the combination ADVERB/IDENTICAL. Once again, the response *nel'zja* "no" to the question about whether *že* can be added to the highlighted clause showed low frequency. The most significant finding to report here is the frequency of responses of *objazatel'no nužno* "you must". This indicates that the participants rated the need for *že* higher in sentences where the part of speech to the left of *že* is ADVERB, and where, according to my proposed radial category, the submeaning of *že* is IDENTICAL.

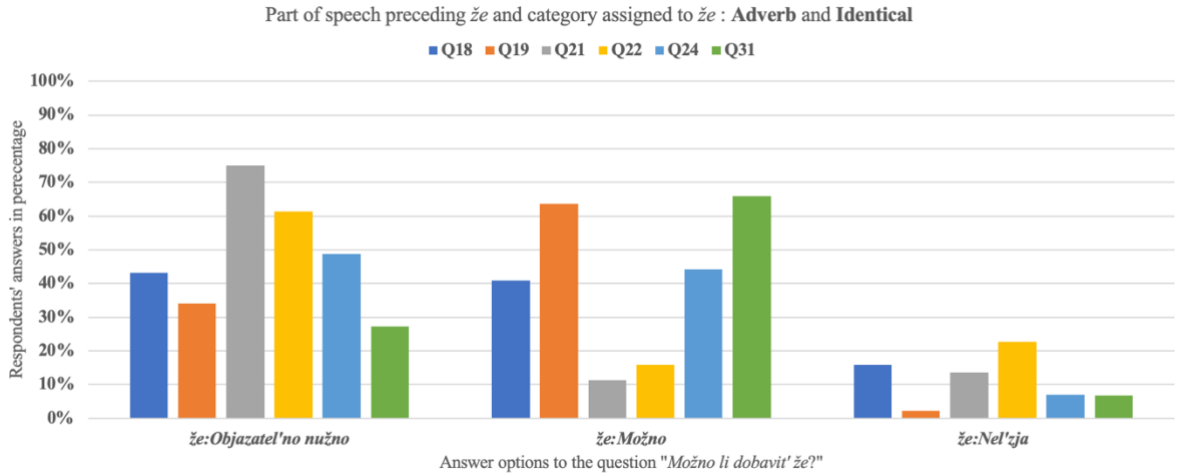

Figure 5.21: The combined results of the question Možno li dobavit' že? for the sentences with the combination ADVERB/IDENTICAL

Figure 5.22 shows the combined results for *ved'* for questions of the combination ADVERB/IDENTICAL. It is evident here that the previously discussed question 31 (Figure 5.20)

is not representative of this combination. It can be concluded that on the whole participants did not agree to adding *ved'* to the highlighted clause in these sentences.

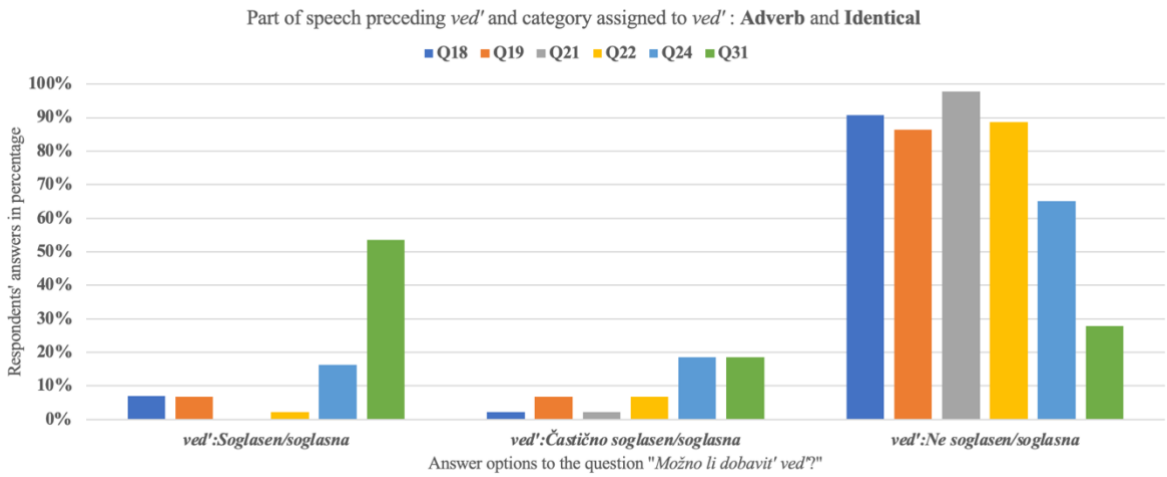

Figure 5.22: The combined results of the question Možno li dobavit' ved'? for the sentences with the combination ADVERB/IDENTICAL

5.5.5 ADVERB/CONTRADICTION

Question 20, shown in Figure 5.23 shows an example of a sentence with the combination ADVERB/CONTRADICTION. As in Figures 5.12, 5.15 and 5.20, the findings for *že* in Figure 5.23 follow a similar pattern: most participants decided that *že* can be added, followed by those who decided *že* must be added, and finally the least frequent response was *nel'zja*. For *ved'* this was not the case. Every participant chose the answer *ne soglasen/soglasna* when asked whether *ved'* could be added to the highlighted clause.

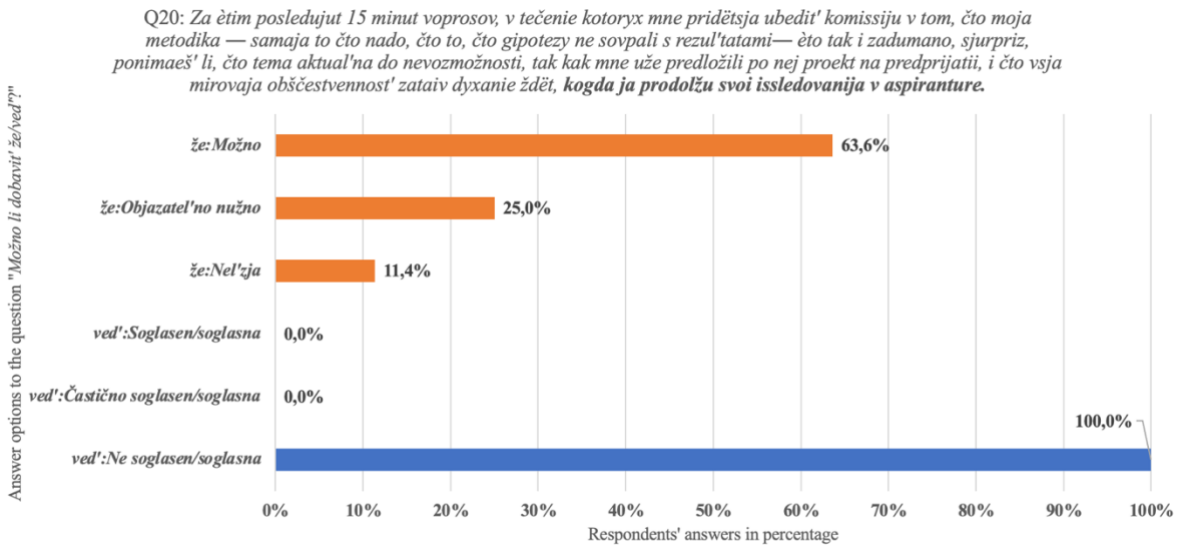

Figure 5.23: A bar chart showing the results of the participants' responses to question 20 of the questionnaire

Another question from this combination was question 28, shown in Figure 5.24. Here, the results were different to question 20 (Figure 5.23). 90.9% of participants agreed that *že* can be added to the highlight clause (in this case the entire sentence), but no participants decided that *že* was obligatory. For *ved'* the findings were mixed. Once again, the most frequent response was *ne soglasen/soglasna*, but in this question 30.2% of participants agreed, and 16.3% partially agreed that *ved'* could be added.

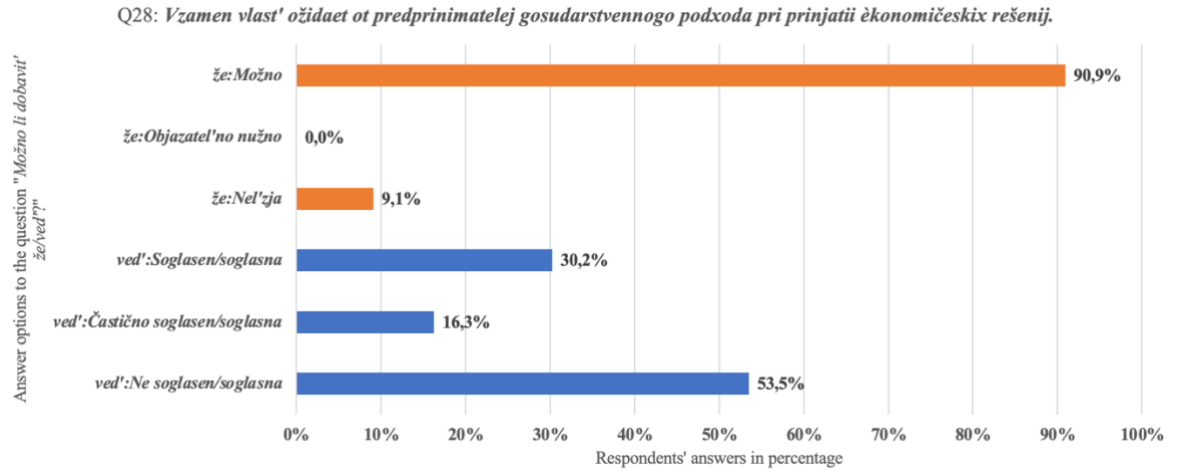

Figure 5.24: A bar chart showing the results of the participants' responses to question 28 of the questionnaire

To summarise the results of *že* in Figure 5.25 it is evident that *možno* was the dominant answer for all questions. The responses for *objazatel'no nužno* were very mixed, with

responses ranging from 0% to just over 30%. Whilst the response *nel'zja* in question 11 deviated from the norm, most other responses represented ~10% of the answers for those questions.

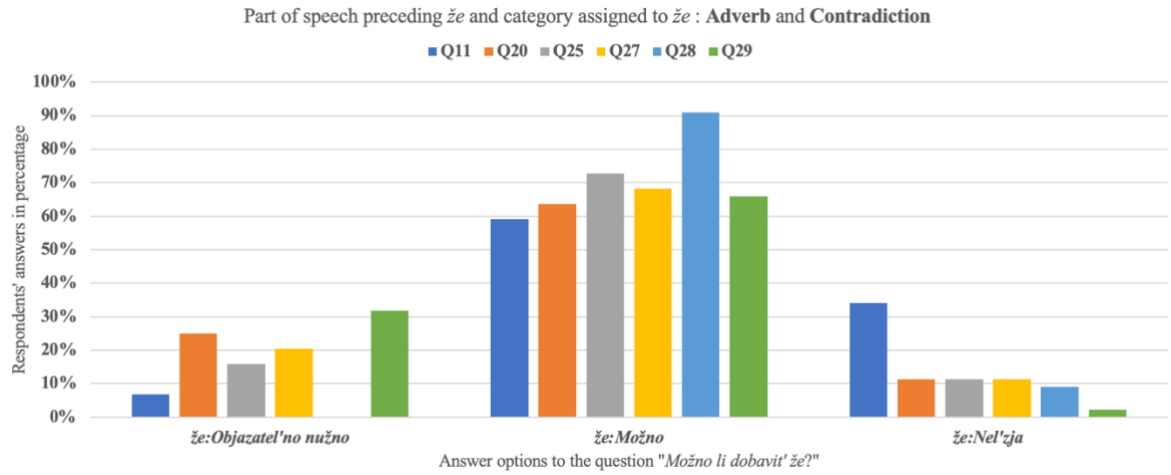

Figure 5.25: The combined results of the question Možno li dobavit' že? for the sentences with the combination ADVERB/CONTRADICTION

The responses to whether *ved'* can be added to the highlighted clause show a clear pattern; the majority of participants did not agree that *ved'* could be added (Figure 5.26). It is interesting to note that although the participants did not know if or/and where *že* was placed in the original sentence, they reacted in a positive way to *že* and reacted in such a contrasting way to *ved'*.

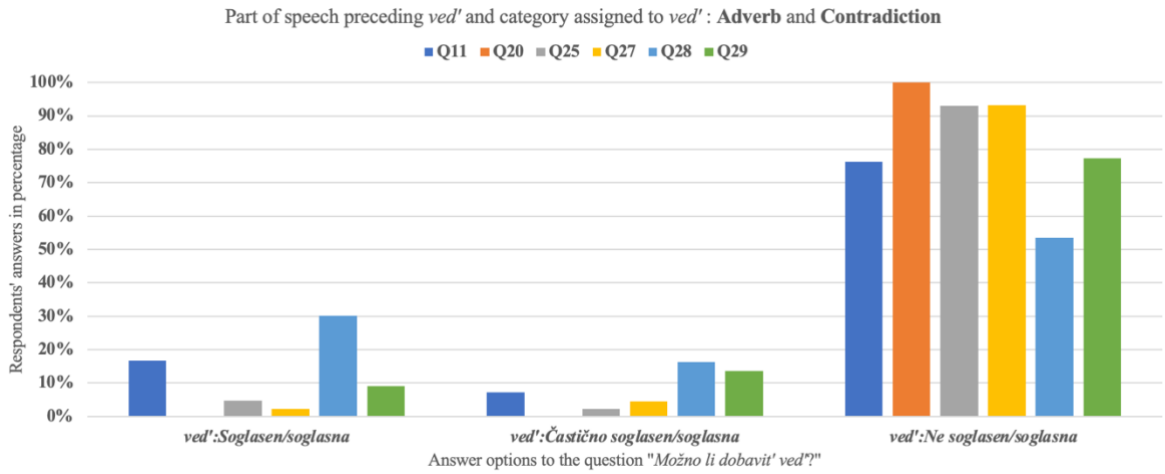

Figure 5.26: The combined results of the question Možno li dobavit' ved'? for the sentences with the combination ADVERB/CONTRADICTION

5.5.6 DEMONSTRATIVE PRONOUN/EMPHASISER

The first question that appeared in the questionnaire with the combination DEMONSTRATIVE PRONOUN/EMPHASISER was question 23, shown in Figure 5.27. The results in Figure 5.27 show a likeness between the way the participants responded to the question about *že* and the question about *ved'*. Here, participants favoured the addition of *že* and *ved'* with 95.5% of responses for *že* as *možno*, and 84.1% of responses for *ved'* as *soglasen/soglasna*.

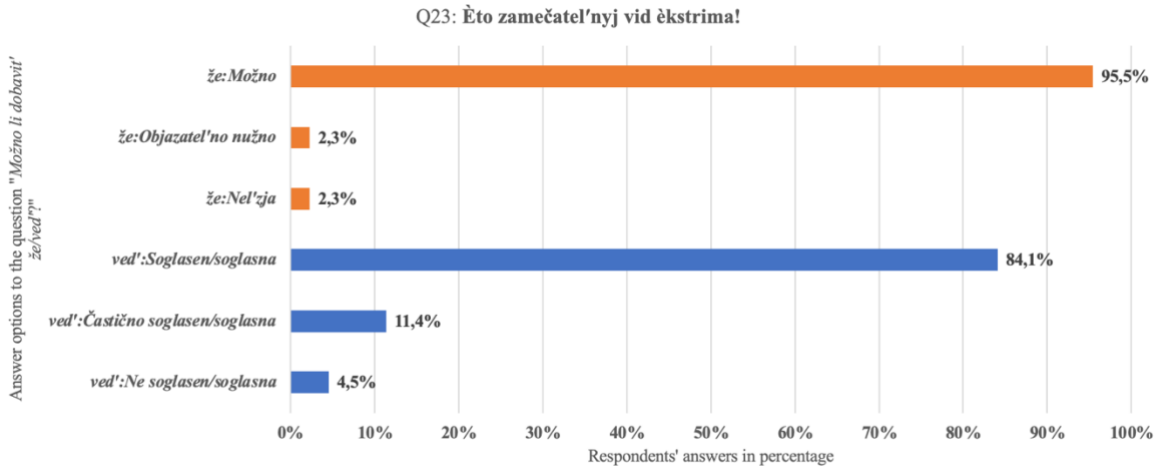

Figure 5.27: A bar chart showing the results of the participants' responses to question 23 of the questionnaire

This pattern was not mirrored in other sentences of the combination DEMONSTRATIVE PRONOUN/EMPHASISER. Question 33, shown in Figure 5.28, shows that the participants had a much stronger tendency towards *že* in the highlighted clause of this sentence, with 70.5% of participants deciding that *že* was *objazatel'no nužno*. In contrast, 93.2% of respondents did not agree that *ved'* could be added here, with the remaining 6.8% of respondents only partially agreeing.

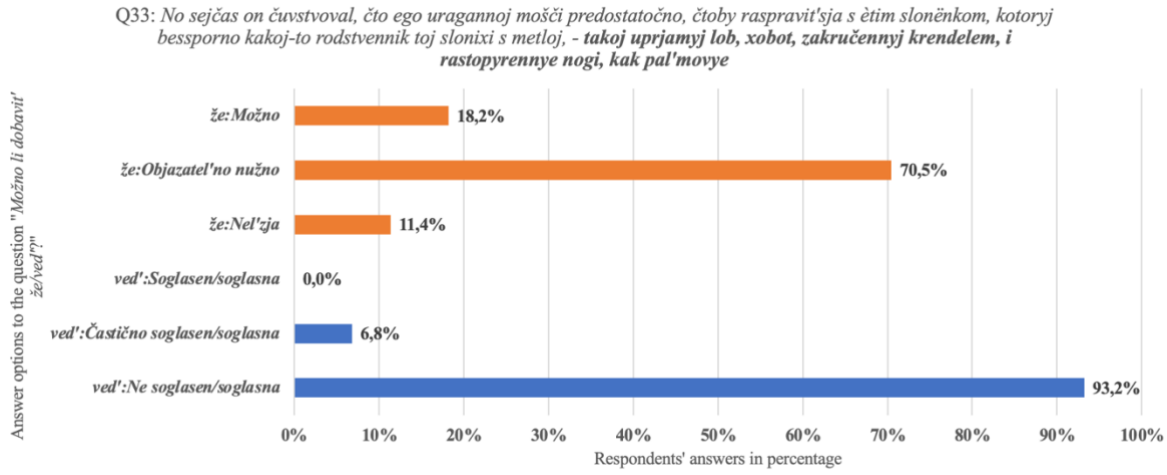

Figure 5.28: A bar chart showing the results of the participants' responses to question 33 of the questionnaire

The combined results for the responses for *že* with the combination DEMONSTRATIVE PRONOUN/EMPHASISER are shown in Figure (5.29). These results show that participants had a strong reaction to question 34 as well as question 33 (see Figure 5.28). Question 34 is in fact the sentence where the most participants decided that *že* had to be added to the highlighted clause, namely 86.4%. Otherwise, respondents favoured *možno* and were positive to adding *že*.

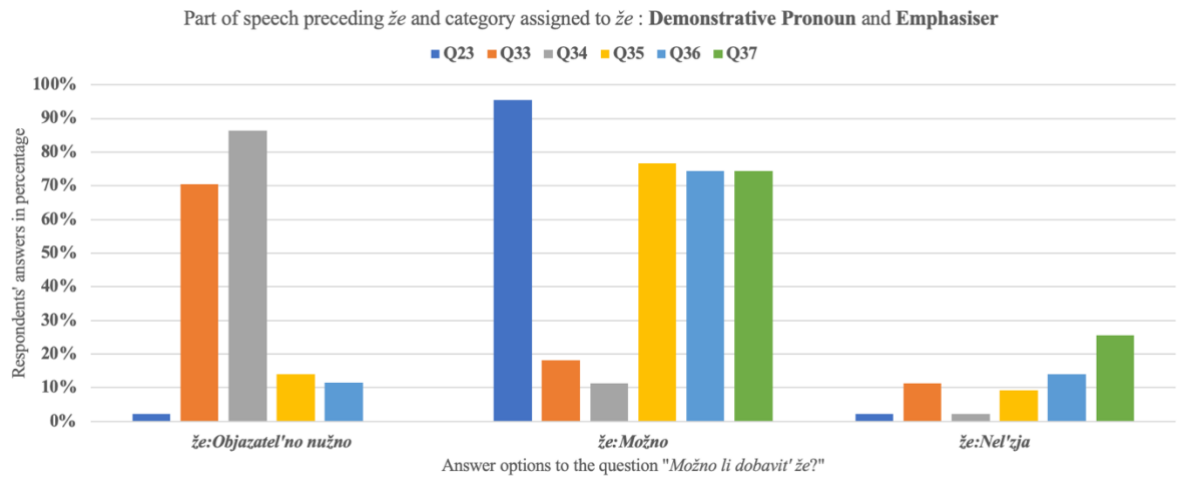

Figure 5.29: The combined results of the question Možno li dobavit' že? for the sentences with the combination DEMONSTRATIVE PRONOUN/EMPHASISER

Despite the participants' positive reaction to adding *ved'* in question 23 (see Figure 5.27), the general consensus from the respondents for the combination DEMONSTRATIVE PRONOUN/EMPHASISER was that *ved'* cannot be added to the highlighted clause. A possible explanation for different reaction to question 23 compared to questions 33-37 may be the relationship between *ved'* and the demonstrative pronoun *èto* when it means "it/this is".

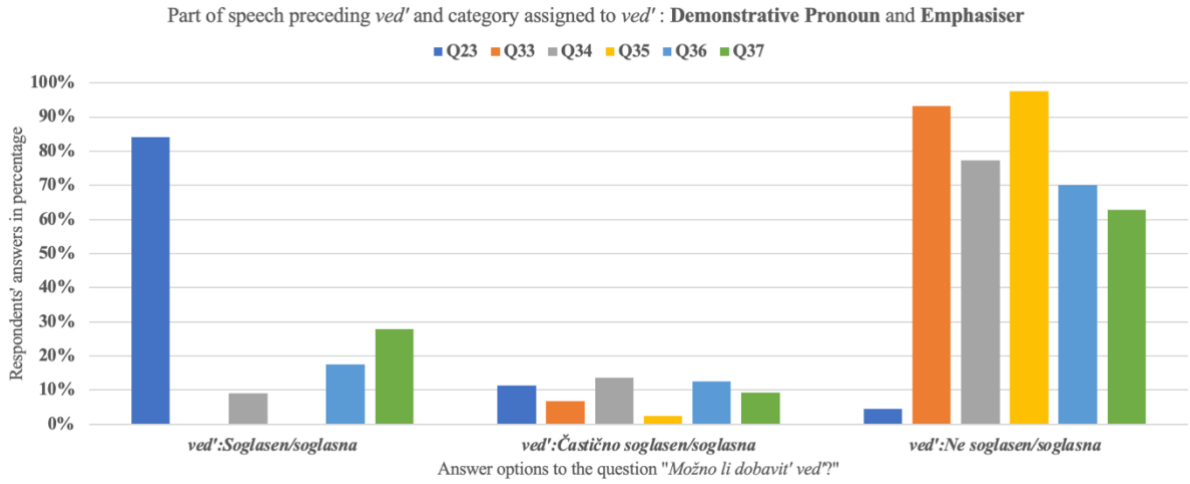

Figure 5.30: The combined results of the question Možno li dobavit' ved'? for the sentences with the combination DEMONSTRATIVE PRONOUN/EMPHASISER

5.5.7 DEMONSTRATIVE PRONOUN/IDENTICAL

The results of question 40 in Figure 5.31 show once again contrasting results between *že* and *ved'*. 44.2% of respondents decided that *že* could be added to the highlighted clause, and 51.2% said that *že* had to be added. This was not the case with *ved'*: 97.7% of participants did not agree that *ved'* could be added.

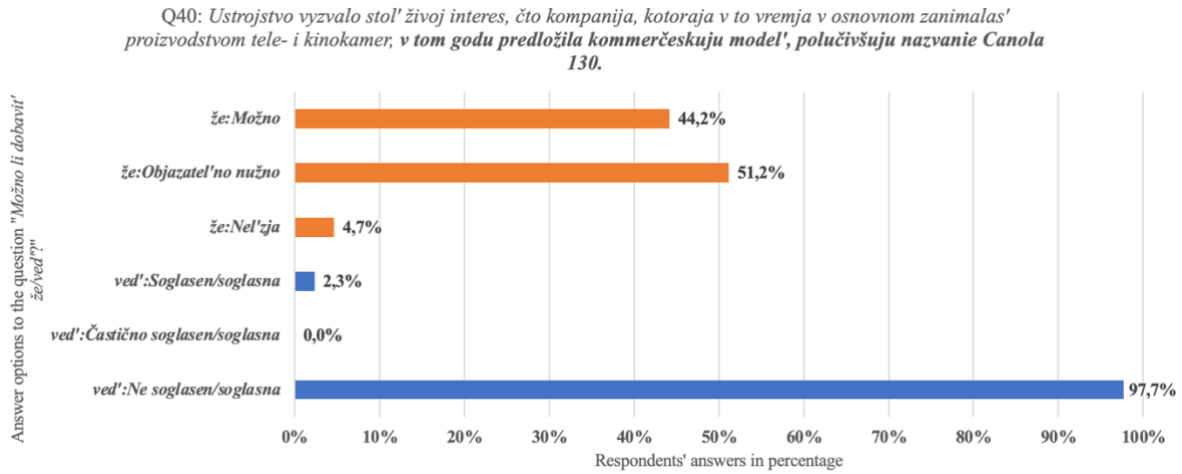

Figure 5.31: A bar chart showing the results of the participants' responses to question 40 of the questionnaire

However, for question 41 the results were not as clear-cut. In Figure (5.32) whilst it is evident that *že* was favoured by the participants, they did not rate *že* as necessary as in question 40 (Figure 5.31). Further comparing question 40 and 41 it is evident that the respondents reacted more positively to adding *ved'* to the highlighted clause in question 41, where almost a third of participants responded with *soglasen/soglasna*.

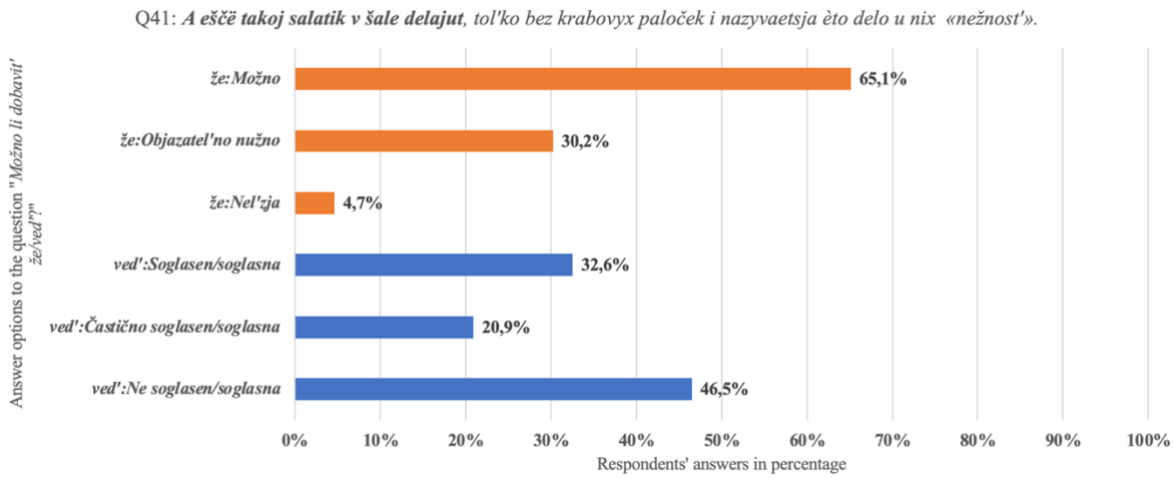

Figure 5.32: A bar chart showing the results of the participants' responses to question 41 of the questionnaire

The results for the *že* in the combination DEMONSTRATIVE PRONOUN/IDENTICAL can be seen in Figure 5.33. Similar to the combination ADVERB/IDENTICAL (Figure 5.21), participants in general rated *že* in the combination DEMONSTRATIVE PRONOUN/IDENTICAL as important: more than 50% of participants decided that *že* was *objazatel'no nužno* in three questions. Otherwise, the majority response was *možno*.

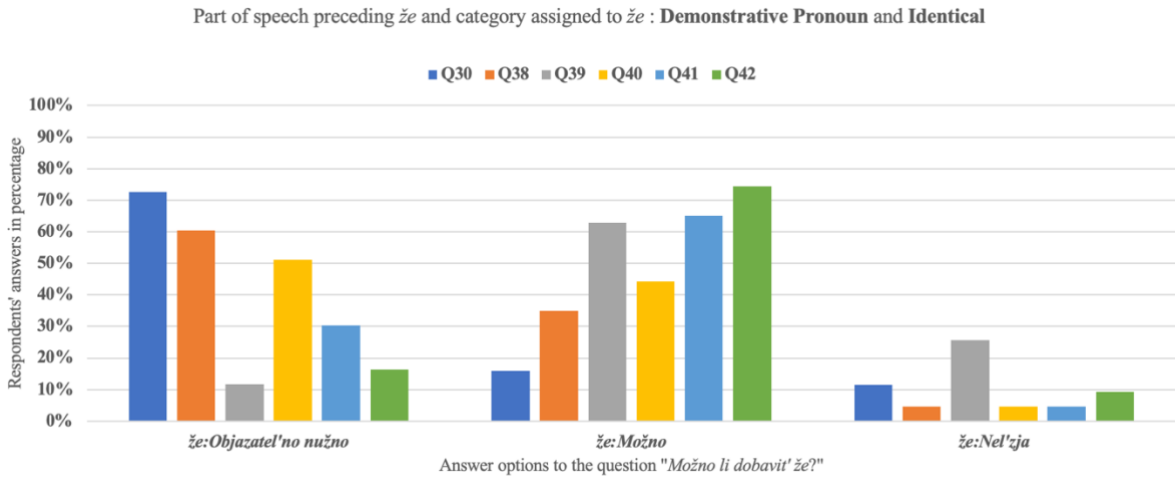

Figure 5.33: The combined results of the question Možno li dobavit' že? for the sentences with the combination DEMONSTRATIVE PRONOUN/IDENTICAL

Figure 5.34 shows the results for *ved'*. These results do not differ significantly from the results of the other combinations. On the whole *ved'* was not favoured by the participants. Question 41 appears to be the most *ved'*-friendly sentence, although acceptability is still very

low: only 32.6% of respondents answered *soglasen/soglasna* to adding *ved'* to the highlighted clause.

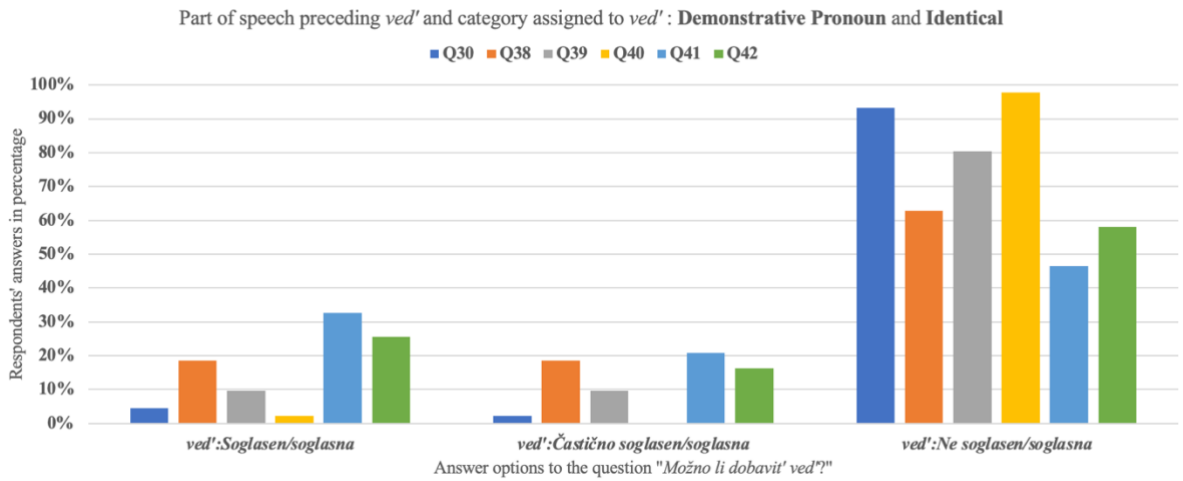

Figure 5.34: The combined results of the question Možno li dobavit' ved'? for the sentences with the combination DEMONSTRATIVE PRONOUN/IDENTICAL

5.5.8 Observations of the questionnaire

Based on the graphs presented in the present section, some conclusions can be drawn. In general participants agreed that *že* could be added to the highlighted clause in every combination. As the original sentence contained *že* it is not surprising that the participants reacted so positively to the addition of *že.* There were some combinations where participants reacted very positively to adding *že* to the sentence. By this, I mean that either the response *objazatel'no nužno* or/and the response *možno* had a high distribution. For *že* the strongest combinations were: NOUN/EMPHASISER, ADVERB/IDENTICAL, DEMONSTRATIVE PRONOUN/EMPHASISER and DEMONSTRATIVE PRONOUN/IDENTICAL. Of these combinations, ADVERB/IDENTICAL showed the strongest attraction to adding *že*. It was interesting to observe that in some instances where participants decided that *že* could not be added to the sentence. This observation demonstrates the level of subjectivity involved in using *že*, as seemed to be the case in Experiment 1.

On the whole participants did not agree that *ved'* could be added to the highlighted clause. Table 5.2 shows a ranking for how participants responded to the seven combinations.

	EMPHASISER	IDENTICAL	CONTRADICTION
NOUN	✈		✈
ADVERB	✈	✈ ✈ ✈	✈ ✈
DEMONSTRATIVE PRONOUN	✈	✈	

✈ = least negative reaction ✈ = negative reaction

✈ ✈ = very negative reaction ✈ ✈ ✈ = extremely negative reaction

Table 5.2: A ranking of how participants responded to adding ved' to the highlighted clause in a given sentence

The combinations NOUN/IDENTICAL and DEMONSTRATIVE PRONOUN/CONTRADICTION are highlighted in red as they were not included in the experiment. There were two combinations where the participants did not always react negatively towards adding *ved'* to the sentence: NOUN/EMPHASISER and ADVERB/EMPHASISER. These responses indicate that participants were most positive towards adding *ved'* to sentences where the meaning was EMPHASISER. This observation further supports the overlapping submeanings in my Venn diagram of my proposed radial categories for *že* and *ved'* in section 3.5 (see Figure 3.).
Furthermore, participants had mixed reactions to sentences where adverb was the part of speech to the left of *že*, with ADVERB/EMPHASISER being one of the combinations most positive to *ved'* being added to the sentence, and ADVERB/IDENTICAL the most negative. From this, it appears the meaning of *že* plays a stronger role than the part of speech to the left of *že,* when participants decided whether *ved'* could be added to the sentence. This observation differs from the statistical analysis in chapter four, where the part of speech to the left of *že* appeared to play a more important role than the meaning of *že*.

To shed more light on this, I will use the data from the questionnaire to carry out statistical analysis to see if any further conclusions can be made.

5.5.9 Statistical analysis

In order to test how significant the data I collected from the questionnaire is, I wanted to carry out a logistic regression. As Levshina states, "this technique [logistic regression] is particularly popular in probabilistic multifactorial models that explain and predict the

speaker's choice between two or more near synonyms or variants on the basis conceptual, geographic, social, pragmatic and other factors" (Levshina, 2015, p. 253). Whilst this should mean that my data is a good candidate for creating a Logistic Regression Model (LRM), this was not possible. One participant did not answer eight of the questions and some participants followed the instructions and did not answer the question about *ved'*. This meant that the data contained some blank responses, and therefore a logistic regression was not possible as this type of statistical model does not react well to missing data.

I was, however, able to create a Classification Tree (cTree). A cTree's aim is to sort the data in the best way possible according to the factors in the dataset. When data has few factor levels, such as the present analysis, a cTree is a fitting statistical method to use (Baayen et al., 2013, p. 267). In the following section I will present and interpret my cTree model of data collected from my questionnaire.

5.5.10 cTree analysis

For the cTree my dependent variable was "Response_VED", i.e., the participants' response to whether *ved'* could be added to the highlighted clause in a sentence which originally contained *že*. The independent variables were "Preceding_lexeme" (the part of speech to the left of *že*) and "Meaning" (the submeaning of *že* derived from my proposed radial category in chapter 3). My cTree is presented in Figure 5.35. Node 1 in this cTree is "Meaning" and is the predictor with the first split when Response_VED is taken into account. This predictor is commonly interpreted as the most important predictor, although this is not always the case (Baayen et al. 2013, p.265). Node 1 is the first split and divides the meanings, where the meaning EMPHASISER is further split based on the "Preceding_lexeme". I will briefly summarise the leaf nodes 3,4,7,8 and 9.

Node 3 (n=571) represents 31% of the dataset. It shows that when the combination was ADVERB/IDENTICAL or ADVERB/CONTRADICTION, 112 participants (20%) agreed that *ved'* could be added to the sentence.

Node 4 (n=479) represents 26% of the data. It shows that when the part of speech to the left of *že* was a demonstrative pronoun or noun, and the meaning of *že* was IDENTICAL or CONTRADICTION participants decided that *ved'* could be added in 145 cases (30%).

Node 7 (n=264) represents 14% of the data and shows that when the combination was ADVERB/EMPHASISER, participants decided that *ved'* could be added to the sentence in 111 cases (42%).

Node 8 (n=261) represents 14% of the data. Node 8 shows how participants responded to adding *ved'* to the sentence when the combination was DEMONSTRATIVE PRONOUN/EMPHASISER. Here, it was possible to add *ved'* in 84 cases (32%).

Node 9 (n=264) represents 14% of the data. It shows that when the combination is NOUN/EMPHASISER *ved'* could be added to the sentence in 145 cases (55%).

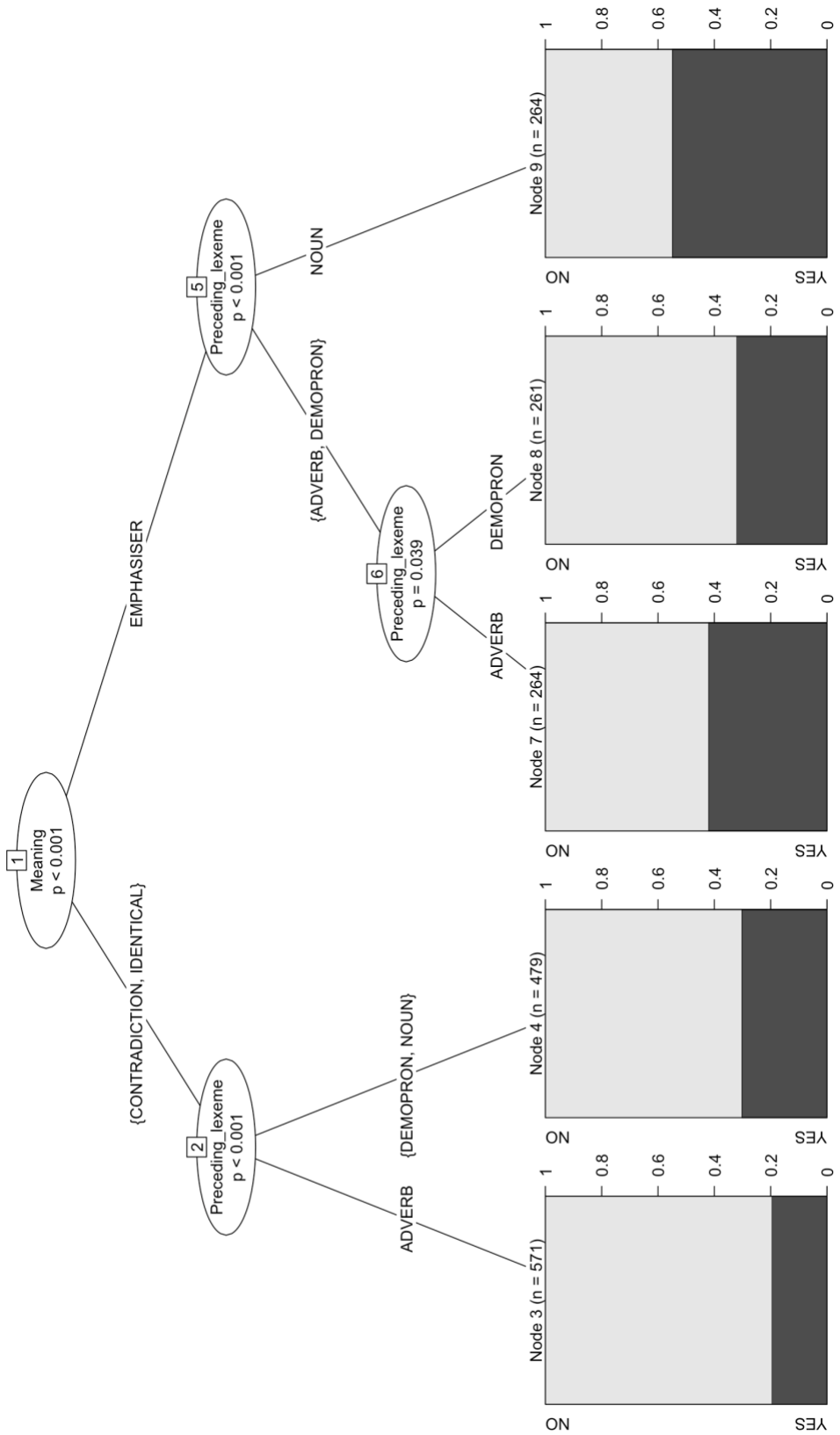

Figure 5.35: A cTree showing the optimal sorting of data from the questionnaire based on the participants' response to adding ved' to the highlighted clause

At first glance it appears that the cTree supports the observation made in section 5.5.8 that the meaning of *že* plays the strongest role in influencing whether participants decided that *ved'* could be added to the sentence. However, nodes 2, 5 and 6 show that the part of speech to the left of *že* was the variable where several splits occurred, which may indicate that this variable is more important than the meaning of *že* in this dataset. I measured the importance of the variables and present the results in Figure 5.36. Here it is evident that the variable "Preceding_lexeme", or part of speech to the left of *že*, is more important than the meaning of *že*. This finding is in accordance with the statistical analysis carried out in chapter four.

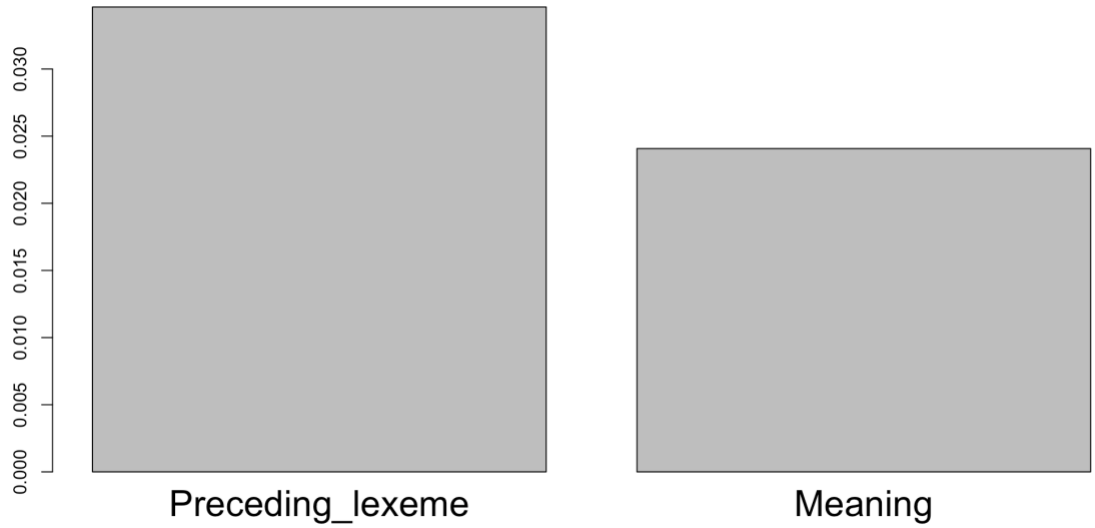

Figure 5.36: A bar chart showing the variable importance of replaceability for the cTree analysis in Figure 5.35

5.5.11 Further reflections

In the questionnaire instructions, the participants were told to only decide whether they would add *že* or not because from the pilot experiment I thought it was interesting that there were so many different answers. My intention was not to look at the insertion point, and for this reason it was removed. Should I carry out this questionnaire again, or carry out a similar study, it would be interesting to look further at the insertion point. A possible method would be that participants could see the sentence in the same way as in the present study, with the clause containing *že* highlighted in bold. The participant's task would be to re-write the highlighted clause adding *že* if and where they see fit. The same could then be repeated for *ved'*. Alternatively, an interactive task could be created where participants can drag and drop *že* and *ved'* into the sentence as they see fit. Tasks such as these would be time-consuming for

annotating the participants' answers but would be an effective way of understanding the syntactic varieties of *že* and *ved'*. To increase chances of obtaining more statistically significant data and use statistical tools such as an LRM, I would make sure that there were no missing combinations in the dataset.

Carrying out this study has been a learning experience. There is room for improvement in terms of the building of the questionnaire, such as including fillers and ensuring that there are no errors. However, I have never carried out a questionnaire to such an extent before, and it is noteworthy that this investigation is built on my own findings. To my knowledge such empirical studies have never before been carried out focusing on *že* and *ved'*, and therefore this investigation serves as an excellent foundation for further research.

5.6 Conclusion

At the beginning of this chapter, I set out to investigate two questions:

1. Does context, in particular the context of the word preceding *že*, play a role in the appearance of *že* in a clause?
2. Based on the fact that *že* and *ved'* share some of the same semantic properties, is it possible to replace *že* with *ved'* based on context?

I began my investigation by presenting a pilot experiment where participants were asked to rate different word order combinations involving *že* and *ved'*. The contradictory findings of this small experiment led me to take this further. I decided not to focus so much on specific word order, but rather on whether there are factors that influence whether *že* or *ved'* can be added to a sentence. By using dataset 1 that I analysed in chapter four, I developed a questionnaire consisting of seven sentence combinations, based on the word preceding *že* and the meaning of *že*. Participants were asked to decide whether *že* and *ved'* could be added to the highlighted clause in a given sentence. The original sentence contained *že,* and therefore it was not surprising that participants reacted positively to the addition of *že*, particularly when the combination was ADVERB/IDENTICAL.

As shown in my analysis participants did not generally favour the addition of *ved'*. When

analysing the different combinations, it appeared that participants rejected the addition of *ved'* the most when the combination was ADVERB/IDENTICAL, and were most positive towards the combination ADVERB/EMPHASISER. This indicated that the meaning of *že* played a more decisive role for participants when favouring the addition of *ved'*. However, this does not seem to be the case in the cTree analysis, where the part of speech to the left of *že* appears to be more important.

The results of this questionnaire did not show ground-breaking results to answer the research questions of this chapter. Whilst participants generally reacted positively to the addition of *že* in the given sentence, there were cases where participants reacted negatively. This indicates that the word preceding *že* may play a role in the addition of *že* to a clause, but this is not absolute. Whilst participants generally reacted negatively to the addition of *ved'* in the given sentence, there were instances where participants reacted positively. This shows that there is potentially some level of synonymy between *že* and *ved'*, the extent to which is not clear from this investigation.

One thing is certain, this investigation has applied empirical methods not previously carried out on *že* and *ved'*, and therefore lays the foundations for further empirical research on the relationship between these two lexemes.

6 Conclusion

This thesis was motivated by a drive to gain a better understanding of a notoriously difficult topic of Russian grammar. Noticing that little research using modern empirical methods has been carried out on *že* and *ved'*, I wanted to use cognitive linguistics, corpus data and statistical analysis to investigate these two lexemes.

In chapter three I used the RuN parallel corpus to investigate the meaning of *že* and *ved'* as well as how they are translated to English. I developed submeaning categories and used cognitive linguistic methods to suggest a radial category network for each lexeme. My proposed radial categories appeared to show a relationship between *že* and *ved'*, as some submeaning categories overlapped.

Chapter four used corpus data from the RNC and statistical tools to investigate the relationship between *že* and *ved'*. Here I investigated which factors, if any, may influence the replaceability of *že* with *ved'*. The factors I focused on were the meaning of *že* (based on my proposed radial category in chapter three), the part of speech to the left of *že*, and how the RNC had tagged *že*. The results of the statistical tests carried out indicated that the part of speech to the left of *že* was the most influential factor.

The findings in chapter four went against my original assumptions. Therefore, I decided to investigate the relationship between *že* and *ved'* further. I first carried out a pilot experiment where participants were asked to rate the position of *že* and *ved'* in a given sentence. Following this, I decided not to focus on the insertion point, but rather on the context of the word preceding *že* and the meaning of *že*. I carried out a questionnaire on native speakers of Russian where participants had to evaluate whether *že* and *ved'* could be added to a given clause in a sentence. The cTree analysis of the questionnaire results indicated that the word preceding *že* was the most important factor when adding *ved'* to the given sentence. However, the results were not conclusive.

Some of the investigations in this thesis have not shown definitive findings, whereby clear conclusions can be drawn. This does not, however, diminish the value and importance of this

thesis. To my knowledge such methods using corpus data and statistical tools have not been carried out on *že* and *ved'* before, and for research in this area of Russian grammar to move forward, I believe it was fundamental to carry out the investigations presented in this thesis.

This thesis may act as a springboard for further investigations on this topic. Further work on *že* and *ved'* could involve looking closer at their meaning, use and replaceability on more data, to gain a bigger picture of the relationship between these two lexemes.

Future research could also involve other Russian lexemes that are often classified as particles, such as *-to* and *a* to investigate their relationship with *že* and *ved'*. It would be of interest to researchers and learners of the Russian language to see how these lexemes are related and yet also differ from each other.

Finally, to return to my personal motivation as a learner of Russian to understand the meaning and use of *že* and *ved'*. Can I now confidently use these lexemes and understand them? *Konečno že!*

References

Baayen, R. H., Endresen, A., Janda, L. A., Makarova, A., Nesset, T. (2013) Making choices in Russian: pros and cons of statistical methods for rival forms. *Russian Linguistics 37*(3): (p. 253–291). https://doi.org/10.1007/s11185-013-9118-6

Bast, R., Endresen, A., Janda, L. A., Lund, M., Lyashevskaya, O., Mordashova, D., Nesset, T., Rakhilina, E., Tyers, F. M., Zhukova. V. (2021). The Russian Constructicon. An electronic database of the Russian grammatical constructions. Available at https://constructicon.github.io/russian/.

Cambridge Dictionary: https://dictionary.cambridge.org/dictionary/english/emphasis (Accessed: 26/1/21).

Cambridge Dictionary: https://dictionary.cambridge.org/dictionary/english/particle (Accessed: 10/3/21).

Croft, W. (2001). *Radical Construction Grammar. Syntactic theory in typological perspective*. Oxford.

Croft, W. (2008). The role of domains in the interpretation of metaphors and metonymies. In Geeraerts, D. (Ed.) *Cognitive Linguistics: Basic Readings*, (p. 269–302). Berlin, New York: De Gruyter Mouton. https://doi.org/10.1515/9783110199901

Dunn, J., & Khairov, S. (2009). *Modern Russian Grammar A Practical Guide*. Abingdon: Routledge.

Endresen, A., Janda, L. A., Reynolds, R., & Tyers, F. M. (2016). Who needs particles? A challenge to the classification of particles as a part of speech in Russian. *Russian Linguistics*, *40*(2), (p. 103–132). https://doi.org/10.1007/s11185-016-9160-2

Fox, J., & Weisberg, S. (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Goldberg, A.E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences. 7*(5), 219-224 https://doi.org/10.1016/S1364-6613(03)00080-9

Hagstrom, P., & McCoy, S. (2002). Presuppositions, *Wh*-Questions, and Discourse Particles: Russian *ŽE.* Paper presented at *Formal Approaches to Slavic Linguistics (FASL-11),* University of Massachusetts-Amherst.

Harrell Jr, F. E., Dupon, C., et al. (2020). Hmisc: Harrell Miscellaneous. R package version 4.4-0. https://CRAN.R-project.org/package=Hmisc

Harrell Jr, F. E. (2021). rms: Regression Modeling Strategies. R package version 6.1-1. https://CRAN.R-project.org/package=rms

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, *15*(3), 651-674.

King, B. M., Rosopa, P. J., & Minium, E. W. (2011). *Statistical reasoning in the behavioral sciences* (6th ed.). New York: Wiley.

Lakoff, G. (1987). *Women, Fire, and Dangerous Things. What Categories Reveal about the Mind.* Chicago: Chicago University Press.

Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

McCoy-Rusanova, S. (2008). Decomposing particles in combination: colloquial Russian –*to, že,* and *ved*. In Antonenko, A., Bailyn, J.F., Bethin, C.Y. (Ed.), *Annual workshop on Formal Approaches to Slavic Linguistics: the Stony Brook meeting*, 2007, (p. 282–296). Ann Arbor: Michigan Slavic Publications.

McCoy-Rusanova S. (2017) Scalar Implicatures, Presuppositions, and Discourse Particles: Colloquial Russian –to, že, and ved' in Combination. In: Lee C., Kiefer F., Krifka M. (Ed.)

Contrastiveness in Information Structure, Alternatives and Scalar Implicatures. Studies in Natural Language and Linguistic Theory, vol 91. Springer, Cham. https://doi.org/10.1007/978-3-319-10106-4_6

McCoy, S. (2003a). Unifying the meaning of multifunctional particles: the case of Russian ŽE. *University of Pennsylvania Working Papers in Linguistics*, *9*(1), 10.

McCoy, S. (2003b). 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.

Meyer, D., Zeileis, A., and Hornik, K. (2020). vcd: Visualizing Categorical Data. R package version 1.4-8.

Nesset, T., Endresen, A., Janda, L. A., Makarova, A., Steen, F., & Turner, M. (2013). How 'here' and 'now' in Russian and English establish joint attention in TV news broadcasts. *Russian linguistics*, *37*(3), 229-251. https://doi.org/10.1007/s11185-013-9114-x

Nesset, T., & Janda, L. A. (2014). Korpuslingvisten i klasserommet: lingvistiske profiler i fremtidens språkundervisning. *Acta Didactica Norge*, *8*(2), Art. 5, 16 sider. https://doi.org/10.5617/adno.1128

Padučeva, E. V. (1988). La particule že: sémantique, syntaxe et prosodie. In *Les particules énonciatives en russe contemporain 3*, (p.11-44).

Parrott, L.A. (1997). *Discourse Organization and Inference: The Usage of the Russian Particles Že and Ved'* (Doctoral dissertation). Harvard University, Cambridge, Massachusetts.

Richardson, J. T. E. (2018). *Latin Square Design*. In: Frey, Bruce B. (Ed.), *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*, (p. 949). SAGE Publications. https://doi.org/10.4135/9781506326139.n382

Sarkar, D. (2008). *Lattice : Multivariate Data Visualization with R* (1st ed. 2008. ed., Use R!). New York, NY: Springer New York.

Vallduví, E., & Vilkuna, M. (1998). On Rheme and Kontrast. In Culicover, P. W., & McNally, L. (Ed.), *The Limits of Syntax* (p. 79-108). San Diego: Academic Press.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the national academy of sciences*, *104*(19), 7780-7785.

Zwicky, A. M. (1985). Clitics and particles. *Language*, *61*(2), 283–305.

Electronic resources

RStudio Team (2015). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/. (last accessed 26/02/21).

RuN corpus: https://www.hf.uio.no/ilos/english/research/projects/run/corpus/ (last accessed: 10/1/21).

Russian National Corpus (RNC). Available at www.ruscorpora.ru. (last accessed 13/9/20).

SurveyMonkey: www.surveymonkey.com (last accessed 10/11/19).

Survio survey website: www.survio.com. (2021) Link to questionnaire in chapter five: https://www.survio.com/survey/d/V7A9J1F3Q9B8A7J8B (last accessed: 12/2/21).

TROLLing repository: https://dataverse.no/dataverse/trolling (last accessed 11/5/21).