A multi-method approach to estimating subjectivity of causal connectives: the case of *poetomu* and *tak chto* in Russian

E. A. (Elena) Savinova

Research thesis

Research Master's in Linguistics

Utrecht University

June 29, 2021

Supervisors:

Dr. W. M. (Pim) Mak

Prof. Dr. T. J. M. (Ted) Sanders

Table of contents

Acknowledgements

First of all, I would like to thank my supervisors, Dr. Pim Mak and Prof. Dr. Ted Sanders, for supporting me in the development of my research ideas and letting me work on the topic of my interest in my native language. I learned a lot from the critical discussions we had about different aspects of connective specialization in subjectivity and I also enjoyed these discussions very much. I would like to express my gratitude personally to Pim for investing his time and effort into giving me detailed feedback on many aspects of this work and for always being available for a timely discussion. And I want to thank Ted for introducing me to the field of discourse processing in the first place and for supporting my interest in subjectivity and causality, which I hopefully will be further working on during my PhD. Besides, I would like to thank Ted for letting me assist with teaching his course, which was a very pleasant and useful experience, as well as for helping me out during my search for the future career in academia. I would also like to thank my family, who played a big role in the success of this thesis, and who always agreed to help me with piloting my experiments, checking my Russian stimuli and promoting my research to potential participants. Finally, I want to thank my research master peers and friends, Agnė and Zhou, for supporting me throughout the thesis semester and for making the experience of writing this thesis less stressful and more enjoyable.

Abstract

Cross-linguistic evidence suggests that languages often employ specialized causal connectives to express subjective versus objective relations. Although there are both established and emerging methodologies used for investigating this phenomenon, little attention has been paid to obtaining converging evidence from several methods, at the same time discerning the different aspects of connective specialization in subjectivity that they elucidate. The present study addressed this issue by developing and testing a multi-method approach to investigating subjectivity profiles of causal connectives in a language and their effects on processing expectations. This approach was applied to the Russian forward causal connectives *poetomu* and *tak chto*, which have not been studied in this respect. In Study 1, the traditional corpus analysis and an innovative method of online connective insertion task were combined to characterize subjectivity profiles of the Russian causal connectives. In Study 2, an online sentence continuation task was used to tap into processing expectations triggered by the Russian causal connectives. The results of all the three methods provide converging evidence that *tak chto* expresses subjective relations and triggers subjective continuations more than *poetomu*. The differences between the results of different methods are discussed. The present study proves the importance of combining different methodologies for investigating subjectivity profiles of connectives in a language and contributes to the cross-linguistic field of research on subjectivity reflected in causal coherence markers.

1. Introduction

One of the important aspects of understanding discourse is the ability to constantly keep track of the changing perspectives, i.e., sources of information, and to build coherent representations of who did, said and thought what (Graesser et al., 1997; Zwaan et al., 1995; Zwaan & Rapp, 2006). Correctly identifying sources of information and the degree to which these sources are involved in the information representation, i.e., the subjectivity of certain utterances, is a crucial skill for correct interpretation of spoken and written texts. Previous literature suggests that subjectivity, among other things, is especially relevant in the interpretation of causal relations (Pander Maat & Sanders, 2001; Sanders et al., 1992; Sanders et al., 2009; J. Sanders et al., 2012). More specifically, according to the literature on subjectivity and causality in discourse processing, causal relations vary in their degree of subjectivity (Pander Maat & Sanders, 2001; Sanders, 1997; Sweetser, 1990; Sanders et al. 2009). In general, two types of causal relations can be identified: objective relations that are observable in the objective reality and are simply reported by the speaker, and subjective relations that are construed by some conscious mind, most often the speaker. A number of cross-linguistic studies provide evidence that in several languages across the world, such as Dutch (Degand, 2001; Degand & Pander Maat, 2003; Pit, 2007; Sanders & Spooren, 2015), French (Degand & Pander Maat, 2003; Pit, 2003; Zufferey, 2012), German (Pit, 2003, 2007), Turkish (Çokal et al., 2020) and Mandarin (Li et al., 2013; Xiao et al., 2021), specified causal connectives are systematically used to mark subjective versus objective relations. These findings suggest that the distinction between subjective and objective causality is an important cognitive distinction that is reflected on the level of the surface linguistic marking. This distinction is also reflected in discourse processing: several studies on different languages show that subjectivity in causal relations is associated with an online processing cost (Canestrelli et al., 2013; Traxler, Bybee, et al., 1997; Traxler, Sanford, et al., 1997). Namely, as soon as the relation is identified as subjective, there is a delay in processing. Importantly, in languages with specialized subjective and objective causal connectives, this processing delay related to subjectivity is observed immediately after subjective connectives (Canestrelli et al., 2013; Wei et al., 2019), while at the same time the processing of the upcoming relation is facilitated (Canestrelli et al., 2013; Li et al., 2017; Wei et al., 2021). This means that specialized connectives can serve as cues towards subjectivity or objectivity of the upcoming relation in processing.

In terms of determining the specificity of causal connectives, the empirical studies cited above mostly employ the method of corpus analysis in order to identify the subjectivity profiles of certain causal connectives or coherence markers in different languages. Corpus analysis implies selection of cases where the connectives of interest are used from large corpora not annotated for discourse relations and annotation of these cases for the type of causal relation. Although this method provides an insight into the natural use of causal connectives, it also has several drawbacks. Firstly, the approach where the occurrences of causal connectives rather than causal relations are extracted may require time-consuming iterative selection of the cases for analysis, simply because causal connectives under investigation usually have other non-causal meanings or can be of low frequency. Secondly, certain connectives or relations may be underrepresented in the corpora, which can affect the general picture: for example, it may result in ostensible absence of prototypical connectives for the underrepresented relations. Finally, corpus analysis involves manual annotation of the selected relations for their degree of subjectivity. This is not only laborious and requires more than one annotator but is also conceptually difficult due to the absence of "clear" subjective or objective cases. Connectives are not only used to express one or another type of relation but they also play a role in the creation of the relation itself, adding subtle meanings and even changing the degree of subjectivity (Sanders et al., 2009, J. Sanders et al., 2012).

One of the recent studies exploring subjectivity profiles of causal connectives used a different methodology, namely, a crowdsourcing experiment with a connective insertion task, to investigate the phenomenon in Spanish (Santana et al., 2021). According to the authors, this method resulted in a more effective and less time-consuming categorization of causal connectives in terms of subjectivity as compared to the previous corpus studies (Santana et al., 2017, 2018) because it let them quickly gather the data representing actual interpretations of naïve speakers. This method has another important conceptual advantage in comparison to the corpus analysis. Namely, it allows for pre-selection of prototypical subjective and objective causal relations rather than instances of connectives. Such an approach helps to avoid the issues related to the frequency imbalance of connectives or relation types discussed above among corpus analysis disadvantages.

Although the connective insertion task may appear as a more effective substitute to the traditional corpus analysis, it is important to discern the different aspects related to subjectivity that these methods elucidate. The traditional corpus analysis allows researchers to look at the natural use of causal connectives and takes into account the influence of higher-level discourse

representations, such as genre, or discourse type (Stukker & Sanders, 2014; Webber, 2009). This is important because previous studies show that subjectivity profiles of causal connectives may be unstable across genres (Li et al., 2013). Moreover, the distribution of absolute corpus frequencies representing how often certain relations are exemplified by certain connectives can provide information about the effectiveness of these connectives as processing cues for the type of the upcoming relation across discourse types. At the same time, the picture provided by the corpus analysis is affected by higher-order discourse characteristics, such as genre or modality (Zufferey, 2012). A good example is the prevalence of subjective over objective relations in discourse in general (Santana et al., 2018), and in spoken discourse specifically (Sanders & Spooren, 2015). This imbalance between the types of causal relations may result in the ostensible underspecification of a connective: it may turn out that a connective exemplifies the majority of the objective relations in a spoken corpus but at the same time is encountered in a considerable number of subjective relations, merely due to the prevalence of the latter in speech. In the connective insertion task, on the contrary, the number of different relation types can be controlled. Therefore, this method allows for quick and direct evaluation of connective specialization by showing how prototypically subjective and objective cases are disentangled with the use of different connectives by the native speakers. Thus, the current study argues that it is important to approach the problem of identifying subjectivity profiles of causal connectives in a language and their role as processing cues using a combination of different methodologies.

In terms of processing effects of specialized connectives, previous processing studies show that prototypically subjective and objective causal connectives enable readers to anticipate the degree of subjectivity of the upcoming relation (Canestrelli et al., 2013; Li et al., 2017; Traxler, Bybee, et al., 1997; Traxler, Sanford, et al., 1997; Wei et al., 2019, 2021). More specifically, specialized connectives facilitate tracking of the information source and provide a cue whether the upcoming content should be attributed to a subject of consciousness (SoC), i.e. considered subjective (Wei et al., 2019). In processing, this is reflected in an immediate delay after subjective as compared to objective connectives (Canestrelli et al., 2013; Wei et al., 2021) and in faster processing of the upcoming congruent relations (Canestrelli et al., 2013; Li et al., 2017; Wei et al., 2021). At the same time, speakers' expectations should also be guided by higher-level discourse type characteristics, such as, for instance, the prevalence of subjective relations. However, it is still unclear how exactly these two factors, namely, the specialization of connectives in terms of subjectivity and the higher-level discourse type characterisitcs, contribute to predictive processing. In other words, it is unclear whether these two factors

interact with each other or have independent influence on steering processing expectations in discourse with respect to subjectivity. In order to investigate this issue, the present study employs a sentence continuation task that targets the speakers' expectations directly.

To sum up, although there are both established and emerging methodologies used for investigating the general phenomenon of subjectivity reflected in causal coherence markers, little attention has been paid to obtaining converging evidence from several methods, at the same time discerning the different aspects of the subjectivity in causality phenomenon that they elucidate. The general aim of the present research is to address this issue by developing an integrative multi-method approach to investigating the subjectivity of causal connectives in a language, both in terms of specialization and in terms of processing expectations, and applying this approach to the Russian language that has not yet been studied in this respect. Naturally, investigation of subjectivity profiles in a new understudied language that is typologically different from the languages studied so far also serves as an important contribution to the research field. From a more detailed perspective, the current research paper aims to answer two research questions, namely:

> 1)   Are subjective profiles of Russian forward causal connectives reflected in the same way in a corpus analysis and a connective insertion experiment?

> 2)   Are Russian forward causal connectives good processing cues towards the subjectivity/objectivity of the relation and is this modulated by the discourse type?

To address the first research question, Study 1 combining the traditional corpus analysis and a novel crowdsourcing connective insertion task with prototypical subjective/objective stimuli was conducted focusing on the two Russian forward causal connectives *poetomu* and *tak chto*. To answer the second research question, Study 2 combined evidence from the corpus analysis and an original online crowdsourcing experiment with a sentence continuation task manipulating connective and discourse type.

This paper is organized as follows. In the next section, previous cross-linguistic research on subjectivity and causality, as well as several methodological issues will be discussed. Subsequently, the goals and hypotheses of the present study will be outlined. Section 3 describes the results of Study 1, including both the corpus analysis of the Russian forward causal connectives and the online connective insertion experiment. In section 4, the results of Study 2 employing an online sentence continuation experiment are reported. Finally, Sections

5 and 6 are devoted to the general discussion of the findings of both studies and conclusions, respectively.

2. Theoretical background

2.1. Subjectivity and causality: cross-linguistic evidence

Different theoretical approaches to causality in discourse distinguish between at least two types of causal coherence relations: objective and subjective relations (Le groupe λ-l, 1975; Sanders, 1997; Stukker & Sanders, 2012; Sanders et al., 2009; Sweetser, 1990). Objective relations are those that hold between events in objective reality and are simply reported by the speaker, while subjective relations are construed by the speaker in the current discourse, and thus, can only be interpreted in relation to the speaker, or more precisely, the subject of consciousness (SoC), which is the speaker in the most subjective cases. Several approaches propose a more refined distinction between the types of causal relations with respect to subjectivity (Pander Maat & Sanders, 2001; Sanders & Spooren, 2015; Stukker et al., 2009). Within the objective type, they distinguish between non-volitional causal relations that simply report the observed causal connections between the events in the world (1a), and volitional relations that involve volitional decisions of an SoC (1b). Within subjective causal relations, other two types are distinguished: epistemic causal relations that are construed as inferences made by an SoC and should be interpreted as relating to the SoC's conclusions about reality (1c), and speech act relations where an illocutionary speech act is connected to the reason for its utterance (1d).

(1)   a. It is raining, so the streets are wet.

b. It was raining, so we entered a café.

c. He does not pick up the phone, so he is probably busy.

d. It might rain later today, so please take an umbrella with you.

Although in the examples above all types of the relations were expressed with the same English causal connective 'so', cross-linguistic evidence suggests that languages often employ specific causal coherence markers to express different types of causal relations. For example, German backward causal connectives *weil* and *denn* (Pit, 2003, 2007) and French connectives *parce que* and *car* (Degand & Pander Maat, 2003; Pit, 2003) are used to express prototypically subjective and objective causal relations, respectively. In Mandarin, forward causal connectives *yīn'ér* and *yúshì* prefer non-volitional and volitional domains, respectively, while *kějiàn* mostly expresses epistemic relations (Li et al., 2013). In Dutch, the subjectivity profiles of both forward and

backward causal connectives show clear prototype structure: *daardoor* and *doordat* are restricted for non-volitional uses, *daarom* and *omdat* are used in volitional cases, and *dus* and *want* express subjective relations in general (Degand & Pander Maat, 2003; Stukker & Sanders, 2012; Sanders & Spooren, 2015; Stukker et al., 2009). Thus, subjectivity in general and more detailed distinctions related to subjectivity are important parameters of cognitive categorization in the domain of causality, and therefore, they often get reflected in the preferential use of different causal connectives.

There are also languages that do not show systematic categorization of causal connectives based on subjectivity, such as English (Andersson & Sundberg, 2021; Knott & Sanders, 1998) or Spanish (Santana et al., 2021). For example, in Spanish, there is only one specific subjective connective *puesto que*, while the other connectives *porque* and *ya que* can express both subjective and objective relations. In English, in addition to the underspecified connective *so*, there is *therefore* exhibiting preference for subjective relations and *as a result* preferring objective relations (Andersson & Sundberg, 2021). However, as Andersson and Sundberg describe, usage differences between these connectives do not reflect systematic categorization of causal connectives in terms of subjectivity but rather pertain to other variables, such as register and rhetorical purposes. Thus, the conceptual distinction between subjective and objective relations is not always reflected in the lexicon of causal connectives and further cross-linguistic research is needed to understand how subjectivity can be systematically signaled in languages.

2.2. Forward causal connectives in Russian

Russian has a large number of coherence markers for forward and backward causality. Most of them are multiword, so it is often not easy to draw a line between connectives and less grammaticalized coherence markers. According to Russian grammars, such as, for example, the corpus grammar (Apresyan & Pekelis, 2012), there is only one proper conjunction of consequence in Russian, that is *tak chto*. However, several sources describing Russian markers of cause and consequence agree on the fact that there is another prominent marker of consequence relation, i.e. *poetomu*. Apresyan and Pekelis (2012) define it as a pronominal adverb because it can be used as an adverb and in combination with other conjunctions (2).

(2)   *Monopolizaciya – eto ser'eznoe prepyatstvie dlya ekonomicheskogo razvitiya. Imenno **poetomu** monopolii zapreshcheny zakonom.*

'Monopolization is a serious obstacle for economic development. Exactly **poetomu** (=for this reason) monopolies are prohibited by law.'

Other works, for example Haag (2004), consistently call *poetomu* a pronominal conjunction. Despite different approaches to the part of speech status of *poetomu*, the most important information for the present paper is that all approaches converge on the fact that *poetomu* is often used as a forward discourse connective linking clauses of cause and consequence. The absolute frequencies of *poetomu* and *tak chto* in the Russian National Corpus (RNC, 321 million words) reveal that *poetomu* (97 206) is even more frequent than *tak chto* (50 702). However, these numbers are not very representative because in the case of *tak chto*, they also include instances where *tak* and *chto* are two different adjacent words rather than a multiword connective, and in the case of *poetomu*, the figures include adverb uses. Unfortunately, the query options in the RNC do not allow for more precise automatic estimation.

To the best of my knowledge, Russian causal connectives have not yet been studied with respect to their subjectivity/objectivity. However, preliminary introspective and corpus observations suggest that for the two forward causal connectives described above, there may be usage preferences that coincide with the subjective/objective distinction. More precisely, *poetomu* seems to fit well in non-volitional (3a) and volitional (3b) causal relations, while *tak chto* is naturally used in epistemic (3c) and speech act cases (3d).

(3)   a. *Noch'yu byl dozhd', **poetomu** vezde luzhi.*

       'It was raining at night, **so** there are puddles everywhere.'

       b. *Byl sil'nyj dozhd', **poetomu** my zashli v kafe.*

       'It is raining hard, **so** we entered a café.'

       c. *On ne beret trubku, **tak chto** on, navernoe, zanyat.*

       'He does not pick up the phone, **so** he is probably busy.'

       d. *Segodnya budet dozhd', **tak chto** voz'mi, pozhalujsta, zontik.*

       'It will rain today, **so** please take an umbrella.'

Haag (2004) notes that the semantics of *poetomu* underlines explanation of the reason for the consequence expressed in the clause this connective attaches. Such observation partially captures the nature of objective causality: there is a clear cause that leads to a consequence in the observable reality. On the contrary, in subjective causality, the consequence-cause

connection between the events in the real world is less salient. For instance, in example (3c), the real-world connection is reversed with respect to the surface forward argument-claim order: the fact that he is busy can lead to the fact that he is not picking up the phone, and the speaker of (3c) makes an inference, i.e. claim, about a possible cause of the event in the first clause. As example (3c) suggests, *tak chto* can be naturally used in such cases, so it does not require clear pragmatic accent on the objective reason expressed in the first clause.

2.3. Identifying subjectivity profiles of connectives: experimental vs. corpus approach

The studies devoted to investigating the subjectivity profiles of causal connectives in different languages usually employ the method of corpus analysis (Andersson & Sundberg, 2021; Çokal et al., 2020; Li et al., 2013; Sanders & Spooren, 2015; Xiao et al., 2021). The procedure underlying this method is the following: firstly, a number of occurrences of different causal connectives are selected in a corpus or several corpora representing different discourse types and then these occurrences are analyzed by two or more annotators according to the most relevant characteristics related to subjectivity. Such characteristics are the type of relation (non-volitional, volitional, epistemic, speech act), presence and explicitness of the SoC, type of the SoC, among others (Çokal et al., 2020; Li et al., 2013; Sanders & Spooren, 2015). Based on these parameters, it is possible to identify the subjectivity profiles of the causal connectives under question.

Corpus analysis as a method of identification of the subjectivity profiles has one main advantage: it allows researchers to capture the language use as it is and to investigate how different discourse types, such as spontaneous conversation, newspaper discourse, narratives, etc., influence the use of causal connectives. This method also allows to capture other linguistic aspects relevant to the choice between different causal connectives, which are obviously not restricted to the distinction in the degree of subjectivity. However, corpus analysis as a method comes with certain disadvantages as well. First of all, the selection of the occurrences of connectives rather than causal relations of different subjectivity type, requires a lot of time and effort, as in many languages the discourse connectives under investigation are not exclusively used as causal connectives linking two clauses. For example, as discussed above, the Russian connective *poetomu* can serve as an adverb or as a connective adding a unit smaller than a full clause (comparable to *therefore* in English: *He is good in biology, and therefore, in natural sciences in general*). Exclusion of such irrelevant cases is often left to manual selection. Moreover, low frequencies of certain connectives or relations in the corpora may lead not only to a long process of manual selection of the cases for analysis, but also affect the analysis itself.

For instance, if the speech act causal relations are underrepresented in the corpus, it will be difficult to find that a certain causal connective is prototypically used to express this type of relation. Moreover, annotation of the selected cases for various parameters requires more than one annotator, which is laborious and time-consuming. Furthermore, the concept of the type of causal relation is a theoretical construct and one could think of prototypical examples of subjective or objective relations. However, in practice, many causal relations encountered in the corpus or in speech are often far from prototypical, which makes the task of annotation challenging. It is also important to note that the perceived differences in the subjectivity profiles of discourse connectives can be employed by the language users in order to create various linguistic effects, from shifting perspective to "objectifying" causal inferences (Sanders et al., 2009, J. Sanders et al., 2012). In general, connectives not only express, but also create the type of relation. This makes the instances of causal relations encountered in the corpora less easily attributable to one certain relation type, which also influences the analysis. As soon as the connective is removed for the purpose of defining the relation between the two remaining clauses, the original meaning of the sentence is partially lost, which can affect the type of causality, too.

A possible alternative method for identifying subjectivity profiles of causal connectives that has been used in one recent study on Spanish and is also adopted in the current research (Study 1) is a connective insertion task (Santana et al., 2021; Scholman & Demberg, 2017). The procedure that such an experiment employs is rather simple and straightforward. Participants are presented with sentences, where two parts are causally connected but the connective is omitted, and are asked to fill in the missing connective. The procedural advantage of such an approach is that it is relatively simple to administer: the connective insertion task can be easily implemented online via crowdsourcing platforms (Scholman & Demberg, 2017; Yung & Scholman, 2019). Moreover, an experimental approach allows for selection of examples of prototypically subjective or objective relations as stimuli. This eliminates the difficulties related to selection of appropriate examples expressing causality in the corpus, annotation of "fuzzy" cases and connective or relation type imbalance, explained above. Additionally, the stimuli set for an experiment of this kind, consisting of the prototypical examples for every type of relation, can be translated to various languages and reused in future experimental research because the concept of the type of causal relation is language-independent.

Despite all the advantages of the experimental method, it also has several drawbacks. Firstly, the fact that connectives contribute to the creation of the relation and to its type is

equally important for the connective insertion task. Even when presented with a relation designed to look prototypically subjective or objective without a connective, participants can create new interpretations by choosing to insert one or another connective. Another major drawback of the experimental methodology proposed is the fact that it ignores possible differences in the usage profiles of connectives across different discourse types and modalities, such as spoken or written discourse. Notably, the ability to compare different discourse types is one of the advantages of the corpus method.

To sum up, both traditional corpus analysis and the innovative experimental method of connective insertion task have been used in the previous studies to characterize subjectivity profiles of causal connectives in a language. However, the results that they provide have not been directly compared, and thus, it is still unclear whether the two methods converge on the same usage profiles of causal connectives with respect to subjectivity, and whether the new connective insertion task is a valid method for investigating this issue. This is especially important given that these two methods shed light on slightly different aspects related to the phenomenon of specialization of connectives in expressing subjective/objective relations, and therefore, are to some extent complementary to each other. As such, the combination of the two methods is beneficial for a comprehensive approach to identifying subjectivity profiles of causal connectives in a language, and for this purpose, as well as for the purpose of validation of the connective insertion task, the present study will employ both methodologies to explore subjectivity of the Russian forward causal connectives.

2.4. Connectives as processing instructions

Previous experimental studies devoted to processing of subjective versus objective causal relations provide evidence that understanding subjectivity comes at a cost (Canestrelli et al., 2013; Traxler, Bybee, et al., 1997; Traxler, Sanford, et al., 1997). In order to understand that the utterance is subjective, the speakers need to add an extra layer to their interpretation that represents the source of information, or, in other words, a conscious mind who is responsible for the content (Wei et al., 2019). This additional layer of interpretation takes more time and effort in comparison to objective relations, which is reflected in processing. For example, it was found that in English, the processing time of the second clause following *because* is slower for subjective relations (4b) than for objective relations (4a) (Traxler, Sanford, et al., 1997). A subsequent study on the same stimuli (Traxler, Bybee, et al., 1997) further clarified that the delay occurs at the main verb region (*left her purse*), i.e., when it becomes clear that the relation is not objective.

(4)   a. Susan lost her money and credit cards because she left her pure at the bus stop.

b. Susan was careless with money and credit cards because she left her pure at the bus stop.

(Traxler, Sanford, et al., 1997: p. 91)

In languages that distinguish between prototypically subjective and objective connectives, this subjectivity-related delay is observed earlier, namely, at the connective region or immediately after it (Canestrelli et al., Wei et al., 2021). These findings suggest that specialized connectives serve as processing cues towards the subjectivity of the unfolding relation. Furthermore, experimental studies in Mandarin (Li et al., 2017; Wei et al., 2021) show that clauses following the specialized subjective connective *kějiàn* 'so' are read faster than clauses following a neutral causal connective *suoyi* 'so'. This facilitatory effect of subjective connectives on processing of subjective relations reflects the incremental nature of processing and the role of specified connectives as processing instructions. Subjective forward causal connectives guide readers to expect subjective content, such as judgements and speech acts (Sweetser, 1990), whereas objective connectives indicate that the upcoming content is objective. While the online effects of the specialized connectives on reading times can be observed with eye-tracking while reading paradigms, such as the ones cited above, the properties of the expectations triggered by the specialized connectives can be studied using other methods, such as a sentence continuation task adopted in Study 2 of the present paper.

On the other hand, several corpus studies on specialization of causal connectives in terms of subjectivity show that subjectivity profiles of causal connectives are not always robust across higher-level discourse categorizations, such as genre (Li et al., 2013; Li et al., 2016; Stukker & Sanders, 2014) or register (Andersson & Sundberg, 2021). For example, the prototypically subjective Dutch connective *dus* 'so' is used more often in non-prototypical objective cases in feature articles as compared to news reports (Stukker & Sanders, 2014). Other examples are the underspecified Mandarin forward causal connectives *yīncǐ* and *suǒyǐ* that tend to express subjective relations in argumentative and informative genres and, on the contrary, objective relations, in the narrative genre (Li et al., 2013; see similar results for backward causal connectives in Li et al., 2016). Although the studies cited above suggest that discourse type can influence subjectivity profiles of causal connectives, it is still unclear how this influence is reflected in the effectiveness of connectives as processing cues. Moreover, it is still unclear

whether differences in distribution of various causal relation types across genres (Sanders & Spooren, 2015; Stukker & Sanders, 2014), such as the prevalence of subjective relations in spoken discourse, play an important role in guiding processing expectations. In other words, it has to be studied to what extent the specialization of connectives in terms of subjectivity and the higher-level discourse characteristics contribute to speakers' expectations about subjectivity of the upcoming relation.

2.5. *The present study*

The goal of the present study is two-fold. Firstly, from the methodological point of view, the study aims to test a multi-method approach to identifying subjectivity profiles in a language, both in terms of specialization and in terms of processing instructions. This is achieved by combining corpus and experimental methodologies that elucidate different aspects related to distinction in the degree of subjectivity between causal connectives. Secondly, the study aims to contribute to the research field on cross-linguistic categorization of subjectivity in causality by investigating the phenomenon in Russian, which is typologically different from the languages studied so far.

As for the subjectivity profiles, based on introspection and preliminary corpus observations, the hypothesis is that Russian forward causal connectives are specialized in terms of subjectivity, with *tak chto* predominantly expressing subjective relations and *poetomu* mostly used in objective relations. This hypothesis is tested in Study 1 with a combination of two methodologies, namely, the traditional corpus analysis, aggregating spoken and newspaper discourse, and the novel connective insertion task. As for processing instructions, no specific hypotheses can be developed as they depend on the outcome of Study 1. However, taking into account the possible dominance of subjective causal relation type in spoken discourse (Sanders & Spooren, 2015; Santana et al., 2018), expectations triggered by Russian causal connectives are hypothesized to depend on discourse type. This hypothesis is tested in Study 2 in a sentence continuation task manipulating connective and discourse type.

3. Study 1: Subjectivity profiles of Russian forward causal connectives

The first study employed both the traditional corpus analysis and an innovative experimental method of online connective insertion task (Santana et al., 2021) to investigate how specialized Russian forward causal connectives are in terms of subjectivity. As described in Section 2.3, a connective insertion task was adopted to investigate categorization of causal connectives by naïve speakers and quickly assess the specialization of the connectives *poetomu* and *tak chto*

by looking at the insertions in most prototypical subjective versus objective cases. The corpus analysis was also used to investigate how stable the subjectivity profiles of Russian causal connectives are across discourse types (newspaper and spoken) and to obtain an estimation of how frequencies of different relations across discourse types may affect the effectiveness of connectives as processing cues for the degree of subjectivity.

3.1. Method

3.1.1. Corpus analysis

For the corpus analysis, occurrences of the two Russian forward causal connectives *tak chto* and *poetomu* were selected from two different corpora: the Newspaper corpus (305 million words) of the Russian National Corpus (RNC) and the Spoken corpus of the RNC (13 million words). The frequencies of the two connectives in both corpora are presented in Table 1. Note that the numbers for both connectives include instances where they are not used as forward causal connectives linking two clauses (around 10-20 %). The choice of different media is important as it represents different types of discourse. Spoken and newspaper discourse varies in the degree of subjectivity and spontaneity, which can influence the usage profiles of discourse connectives across different discourse types. From each of the two corpora, 100 occurrences of each connective were selected for analysis.

Table 1

Frequencies of the Russian forward causal connectives across corpora

Corpus	Frequency, instances per million	
	tak chto	*poetomu*
Newspaper	174	442
Spoken	345	648

Following the previous corpus studies on the subjectivity of causal connectives (e.g., Li et al., 2013), we used the paraphrase test for defining one of the four types of relation: non-volitional, volitional, epistemic or speech act (Table 2).

Table 2

Paraphrase test for the type of relation

Type of relation	Paraphrase
Speech act	The fact that P leads to the speaker SoC asking/suggesting/advising/commanding the addressee that Q.
Epistemic	The fact that P leads to the SoC's conclusion/claim that Q.
Volitional	The fact that P leads to the SoC's intentional act that Q.
Non-volitional	The fact that P leads to the fact that Q, no intentionality involved.

Of all 400 fragments for analysis, 80 fragments (20%) were coded independently by the author and another Russian speaker with educational background in linguistics, following the procedure of partial overlap coding of coherence relations (Spooren & Degand, 2010). The inter-rater agreement measured by Cohen's kappa was 0.76, suggesting that the coding was sufficiently reliable. All discrepancies between the raters were discussed. The rest of the fragments were coded by the author.

3.1.2. Experiment

Participants

A total of 43 participants took part in the online experiment (age: *M* = 36, *SD* = 16, range 18-68; 27 female; level of education: 37 higher, 4 incomplete higher, 1 secondary, 1 vocational). All participants were native speakers of Russian. Before the start of the experiment, all participants read the information letter about the experiment and signed an online informed consent form by ticking the appropriate box. There was no option to proceed with the experiment, if a participant did not give consent. Participants did not receive financial compensation for completion of the experiment. The study was approved by the Faculty Ethics assessment Committee - Humanities (FEtC-H) of Utrecht University.

Materials

The materials for the experiment consisted of 40 experimental items and 40 fillers. The materials were piloted on two native speakers of Russian not taking part in the main experiment. Their task was to fill in missing connectives, but, in contrast to the actual experiment, there was no choice of answer options. Experimental sentences in which the pilot participants inserted

non-causal connectives, had been changed. Several filler sentences had been changed after the pilots in order to get a more balanced distribution of the two types of filler connectives. The experimental items consisted of Russian sentences with forward causal relations, where the causal connective linking the two clauses was omitted. There were four conditions reflecting the four types of causal relations: non-volitional, volitional, epistemic and speech act. In each condition, there were 10 sentences exemplifying prototypical relations of the corresponding type (in contrast to Santana et al., 2021, the prototypical stimuli were invented and not extracted from corpora). Examples of the sentences used in the experimental conditions are presented in Table 3. The fillers consisted of Russian sentences with contrastive and concessive relations, where the connective linking the two clauses was omitted. The connectives that could be used in such sentences were *no* ('but') and *hotya* ('although'). There were 20 sentences for *no*, and 20 sentences for *hotya*, although both connectives could be used in several items due to their similarity in meaning in the context of contrastive-concessive and concessive-contrastive relations. The stimuli, including experimental items and fillers, were manually pseudo-randomized: there were no more than two experimental items or fillers in a row and items of the same condition could not follow one another. Different conditions were equally distributed across the list. Based on this pseudorandomized order, two experimental lists were created by reversing the order of the stimuli. Each participant saw only one experimental list.

Table 3

Examples of the items across experimental conditions translated to English

Condition	Example
non-volitional	During an overnight hurricane, a tree fell on the railroad tracks, _____ the first morning train was cancelled.
volitional	Kostya does not like to cook himself, _____ he always orders food delivery.
epistemic	Even our excellent student got 4 in physics, _____ physics is definitely the most difficult subject.
speech act	Yesterday I went to take out the trash, _____ maybe you can do it today?

Experimental sentences were created to exemplify prototypically subjective (epistemic or speech act) and prototypically objective (non-volitional or volitional) relations. This approach inevitably resulted in several confounds related to the natural differences between the four types

of relations. For instance, all speech act cases were spoken, while in other conditions, there were sentences that could be encountered in both spoken and written discourse. Similarly, in the non-volitional condition, many sentences were examples of written encyclopedic discourse because it represents one of the prototypical uses of non-volitional relations. These idiosyncratic patterns of differences between the conditions are not possible to avoid in the chosen experimental design as different types of causal relations also differ on other conceptual dimensions, such as prototypical type of discourse or propositional attitude of one of the clauses, etc.

Procedure

The experiment was created and administered online using the Qualtrics software (https://www.qualtrics.com). The participants could access the experiment by clicking on a publicly available link. They were instructed to read sentences consisting of two parts, in which the connectives were omitted, and to insert the connective that, according to their opinion, served best for connecting the parts in each particular sentence. There were four multiple choice options for every sentence: *tak chto, poetomu, no* and *hotya*. On average, the task took around 15 minutes to complete. The participants could return to the task after partial completion within a one-week period from the start of the experiment. Only two participants used this option.

3.2. Results

3.2.1. Corpus analysis

The distribution of the connectives across the types of causal relations in the two corpora is presented in Figure 1 and Table 4. In Table 4, the percentages without brackets sum up to 100 horizontally and reflect the proportion of different relations expressed by each connective in each corpus. The percentages in brackets sum up to 100 vertically for each corpus and show the relative proportions of occurrence of the two connectives for each relation in that corpus.

Table 4

The percentage of occurrences of Russian forward causal connectives across different relations in the spoken and the newspaper corpus

Corpus		Non-volitional	Volitional	Epistemic	Speech act
Newspaper	*poetomu*	26 (63.4)	35 (77.8)	35 (35.4)	4 (26.7)
	tak chto	15 (36.6)	10 (22.2)	64 (64.6)	11 (73.3)
Spoken	*poetomu*	20 (54.0)	33 (78.6)	38 (38.3)	9 (40.9)
	tak chto	17 (46.0)	9 (21.4)	61 (61.6)	13 (59.1)

We analyzed the corpus data using *glm* function from the built-in statistical package in the R software (R Core Team, 2020). A log-linear analysis of frequency counts was conducted to examine the relationship between connective, type of relation and discourse type. The two-level variables of connective and discourse type were coded using scaled sum contrasts (-0.5, 0.5). The four-level variable of relation type was coded using treatment contrasts with epistemic type as the reference level. First, the full model containing all main effects and interactions between the three factors was built. Then the model best fitting the data was selected using backward stepwise selection algorithm based on the Akaike information criterion (*step* function). The final model selected contained main effects of connective and relation type and their interaction, suggesting that there was no significant effect of discourse type on the subjectivity profiles. The results of the model are presented in Table 5. There was a significant interaction between the type of relation and connective. In fragments with *tak chto*, there were significantly more epistemic relations than non-volitional ($z = -3.30$, $p < .001$) and volitional ($z = -6.08$, $p < .001$) relations as compared to *poetomu*-fragments. The distribution of epistemic and speech act relations was similar across both connectives ($z = -0.08$, $p = 0.84$), which was expected since both relations belong to the general subjective domain. Thus, the results reveal that *tak chto* mostly exemplified subjective (149 out of 200 or 75%) relations while the majority of relations with *poetomu* were objective (114 out of 200 or 57%).
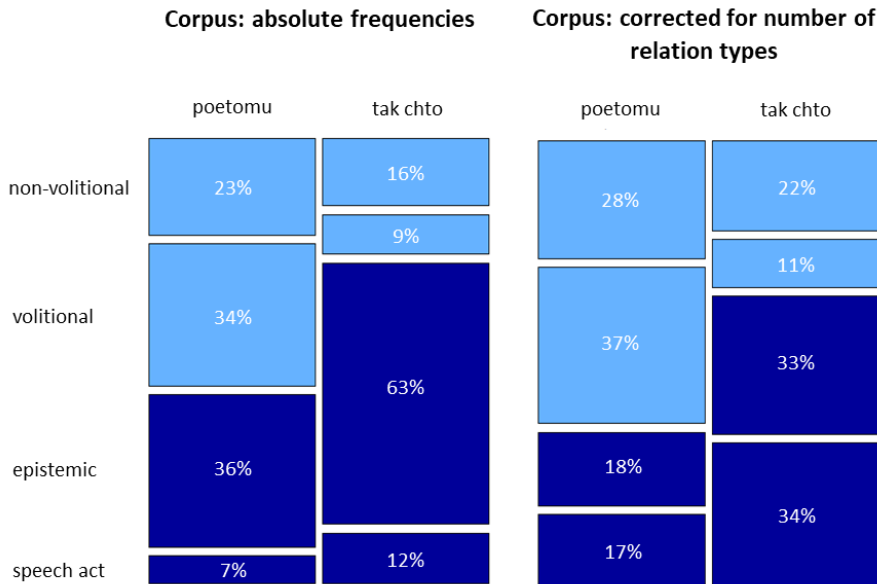
Table 5

The results of the corpus analysis

Fixed effects	Estimate	SE	z-value	p-value
Intercept	3.87	0.07	52.49	<.001
Relation: non-vol	-0.91	0.14	-6.68	<.001
Relation: vol	-0.98	0.15	-6.55	<.001
Relation: sp-a	-1.69	0.19	-9.01	<.001
Connective: *tak chto*	0.54	0.15	3.65	<.001
Relation: non-vol; Connective: *tak chto*	-0.90	0.27	-3.30	<.001
Relation: vol; Connective: *tak chto*	-1.81	0.30	-6.08	<.001
Relation: sp-a; Connective: *tak chto*	-0.08	0.37	0.20	0.84

Note. Formula: frequency counts ~ type of relation*connective. *SE* – standard error; *non-vol* – non-volitional; *vol* – volitional; *sp-a* – speech act.

Although the analysis of absolute frequencies of causal relations types per connective in the corpus shows that there are certain dependencies, it may seem from the overall subjective/objective ratio that *poetomu* is less specialized than *tak chto*. However, this is partially due to greater prevalence of subjective over objective relations in the corpus in general (235 to 165). If absolute frequencies are corrected for the number of relation types (see Figure 1), *poetomu* turns out to be as specialized (65% of objective relations) as *tak chto* (67% of subjective relations).

Figure 1

The distribution of Russian forward causal connectives across different types of causal relations in the spoken and the newspaper corpus
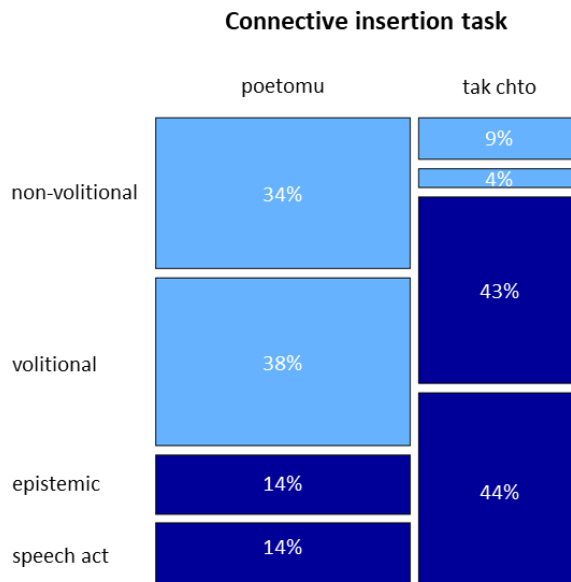

3.2.2. Experiment

The responses of all 43 participants who completed the experiment were included in the analysis (including the two participants who completed the experiment in more than one session). Responses where non-causal (filler) connectives were inserted in the causal experimental sentences were excluded from the analysis (78 data points, which is 4.5% of the data). The distribution of the two causal connectives across conditions is presented in Figure 2 and Table 6. The width of the bins in Figure 2 represents the proportion of insertions of each connective: in general, the participants preferred to insert *poetomu* more than *tak chto* (1009 vs. 633 insertions, respectively). In Table 7, the percentages without brackets sum up to 100 horizontally and reflect the proportion of different relations expressed by each connective. The percentages in brackets sum up to 100 vertically and reflect the proportions of choosing one connective over another for each relation.

Table 6

The percentage of insertions of the two causal connectives across conditions

	Non-volitional	Volitional	Epistemic	Speech act
poetomu	34.3 (85.4)	38.3 (93.3)	13.4 (33.4)	14.0 (33.7)
tak chto	9.3 (14.6)	4.4 (6.7)	42.5 (66.6)	43.8 (66.3)

Figure 2

The distribution of the two causal connectives across experimental conditions

Data analysis was performed using the *lme4* package (Bates, Mächler, et al., 2015; version 1.1.26) in R software (R Core Team, 2020). Considering the choice of a connective as a binary dependent variable, we built two separate generalized linear mixed-effect models: one with the four-level predictor for the type of causal relation and one with the two-level predictor, where the four conditions were collapsed into two general categories, i.e. subjective versus objective relations. The two-level variables were coded using scaled sum contrasts (-0.5, 0.5). The four-level predictor variable was coded using repeated contrasts with the help of package *hypr* (Rabe et al., 2020; version 0.1.11). This coding scheme allowed for comparison of the neighbouring levels: namely, non-volitional versus volitional, volitional versus epistemic, and epistemic versus speech act relations. Maximal random effect structure was gradually reduced to achieve

convergence following the procedure described in Bates, Kliegl, et al. (2015). The results of the two models are presented in Table 7.

Table 7

The results of the online connective insertion experiment

Model 1				
Fixed effects	*Estimate*	*SE*	*z-value*	*p-value*
Intercept	-0.95	0.18	-5.30	<.001
Non-volitional vs. volitional	0.66	0.49	1.37	0.17
Volitional vs. epistemic	-3.90	0.43	-9.02	<.001
Epistemic vs. speech act	0.06	0.31	0.19	0.85
Random effects	*Variance*	*SD*		
Items (intercept)	0.30	0.55		
Subjects (intercept)	0.54	0.74		
Non-volitional vs. volitional	0.93	0.97		
Volitional vs. epistemic	0.44	0.66		
Epistemic vs. speech act	0.25	0.50		
Model 2				
Fixed effects	*Estimate*	*SE*	*z-value*	*p-value*
Intercept	-0.91	0.17	-5.29	<.001
Subjective vs. objective	3.44	0.30	11.33	<.001
Random effects	*Variance*	*SD*		
Items (intercept)	0.36	0.60		
Subjects (intercept)	0.53	0.73		
Subjective vs. objective	0.98	0.99		

Note. Formula 1: connective ~ type of relation (4 levels) + (1 | item) + (1 + type of relation | subject). Formula 2: connective ~ type of relation (2 levels) + (1 | item) + (1 + type of relation | subject). *SE* – standard error; *SD* – standard deviation.

In the first model, there was a significant difference between epistemic and volitional conditions ($z = -9.02$, $p < .001$), suggesting that in epistemic relations *tak chto* was inserted more often than in volitional relations. The contrasts between the two subtypes of objective relations, i.e. non-volitional and volitional ($z = 1.37$, $p = 0.17$), and the two subtypes of subjective relations, i.e. epistemic and speech act ($z = 0.19$, $p = 0.85$), turned out to be non-significant. This means that the ratio in the usage of the two connectives was similar across the two subtypes within each type of subjectivity. The second model revealed a significant main effect of subjectivity on the connective choice ($z = 11.33$, $p < .001$): *tak chto* was inserted more in subjective relations than in objective relations, as expected.

3.3. Discussion

According to previous empirical studies, languages often carve up the conceptual domain of causality by employing different surface markers, i.e. causal connectives, to express different types of causal relations with respect to subjectivity (Çokal et al., 2020; Degand, 2001; Degand & Pander Maat, 2003; Li et al., 2013; Pit, 2007; Sanders & Spooren, 2015; Xiao et al., 2021). Study 1 investigated whether this also applies to the Russian language, and more specifically, whether the Russian forward causal connectives *poetomu* and *tak chto* differ in the degree of subjectivity. The aim of the study was to test whether the subjectivity profiles of Russian causal connectives are reflected in a similar way in both corpus analysis and connective insertion experiment.

The results of Study 1 show that both corpus analysis and online connective insertion experiment converge on finding significant differences between the two Russian forward causal connectives in terms of subjectivity. Namely, the connective *tak chto* is predominantly used in subjective causal relations, while the connective *poetomu* prefers to express objective relations, in line with the hypothesis. However, the strength of the specialization of both connectives was different between the two methods, with the corpus analysis showing less pronounced specialization. For *poetomu* in particular, the corpus analysis revealed relatively low specialization on objective relations (57%), which raises a question whether this connective is indeed specialized and whether it serves as a cue for objectivity in processing. To clarify this issue, Study 2 using a connective insertion task was conducted. The corpus analysis did not reveal any significant difference in the subjectivity profiles of the Russian connectives between spoken and newspaper discourse types.

The subjectivity profiles of the Russian connectives *poetomu* and *tak chto* can be compared to the profiles of forward causal connectives in other languages. In contrast to Dutch, where the connective *daarom* is specialized in volitional relations (52% of causal relations expressed; Pander Maat & Degand, 2001) and the connective *daardoor* is specialized in non-volitional relations (100% of causal relations expressed; Pander Maat & Degand, 2001), or to Mandarin, where the objective connective *yīn'ér* is restricted to non-volitional domain (48% of causal relations expressed; Li et al., 2013), the Russian connective *poetomu* seems to be generally objective (57% of relations expressed in the corpus), without any preference for the specific domain within objective causality (23% non-volitional vs. 34% volitional relations). The preference of *poetomu* to express objective relations in general is also reflected in the corrected corpus figures (65% of relations expressed: 28% non-volitional vs. 37% volitional), as well as in the connective insertion task (72% of relations expressed: 34% non-volitional vs. 38% volitional). Similarly, *tak chto* appears to be a prototypical connective for both epistemic and speech act relations: if the low frequency of speech act relations in the corpus is accounted for, *tak chto* is used in 33% and 34% of epistemic and speech act relations, respectively (compare with the similar picture in the connective insertion task: 43% vs. 44%, respectively). The subjective profile of *tak chto* could be comparable to the Dutch connective *dus* and the Mandarin connective *kějiàn*, however, the corpus studies on these languages did not correct for the low frequency of the speech act relations or intentionally disregarded this type of relations due to low frequency, which makes direct comparisons difficult (Li et al., 2013; Pander Maat & Degand, 2001).

Thus, the profiles of the Russian forward causal connectives *poetomu* and *tak chto* seem to be sensitive to the major distinction between subjective and objective causality. In the experimental results, *poetomu* was preferred in both non-volitional and volitional relations, while *tak chto* was dominant in both speech act and epistemic relations. Nevertheless, in the corpus results, the distribution of the absolute frequencies shows that *poetomu* can be used in epistemic cases (36%) as much as in volitional cases (34%), and in general, *poetomu* exemplified objective relations only in 57% of cases. Moreover, *poetomu* was generally inserted more than *tak chto* in the experimental task and accounted for 61% of all insertions. These observations could potentially indicate that, despite a preference for expressing objective relations of cause-consequence, *poetomu* is more pragmatically diverse and, consequently, less specialized than *tak chto*. However, the ostensible underspecification of *poetomu* is largely due to the prevalence of subjective relations in the corpus (59% of all relations; compare with 75%

in Santana et al., 2018). Corpus analysis corrected for the frequency imbalance between subjective and objective relations reveals more clear specialization of *poetomu* on objective relations (65%). The corrected figures for *poetomu* are closer to the results of the connective insertion task (72% of objective relations expressed), however, the specialization in the experimental results is still more pronounced. This issue will be further addressed in the general discussion in Section 5.

Nevertheless, the prevalence of subjective over objective relations in the corpus should not be neglected since the corpus serves as a proxy for the actual language use. The fact that *poetomu* seems to have less strong preference for expressing exclusively objective relations, given the relation type imbalance, is especially interesting with respect to discourse processing. Previous processing studies on Dutch and Mandarin suggest that even those connectives that show inconsistencies in specialization across corpus studies and discourse types (Li et al., 2013; Stukker & Sanders, 2012) can serve as processing cues towards objectivity of the relation (Canestrelli et al., 2013; Wei et al., 2019). For example, the Dutch backward causal connective *omdat* has been shown to have rather inconsistent profile across studies (Stukker & Sanders, 2012): it expressed 95% of objective relations in the newspaper corpus according to Sanders and Spooren (2015), but only 52% of relations expressed by *omdat* were objective in the newspaper corpus in the study by Degand and Pander Maat (2003). Nevertheless, the study by Canestrelli et al. (2013) showed that *omdat* triggered objective expectations: when *omdat* was inserted in subjective relations, readers experienced serious processing difficulties in the end of the second clause. Another example is the Mandarin connective *yīn'ér* (Li et al., 2013) that expressed 65% and 76% of objective relations in the informative genre and in the narratives, respectively, but only 44% of relations with *yīn'ér* were objective in the argumentative genre. Although the processing studies on Mandarin did not take into account different genres, *yīn'ér* was found to guide speakers' attention away from the subject of consciousness, i.e., to trigger objective expectations, in the visual word paradigm study by Wei et al. (2019). To sum up, based on the previous literature, it is rather difficult to estimate whether the fact that *poetomu* expresses 57% of objective relations in the present corpus study leads to the interpretation that this connective is not a reliable cue towards objectivity of the relation.

Furthermore, Stukker and Sanders (2012) argue that cross-linguistically, objective causal connectives often have less robust profiles across discourse types than subjective connectives. For example, several studies in German (Keller, 1995) and French (Zufferey, 2012) suggest that objective connectives are prone to subjectification in diachronic

development, and especially in spoken discourse (Stukker & Sanders, 2012; Zufferey, 2012). This means that discourse type is an important factor that can affect both the subjectivity profiles of causal connectives and the processing instructions triggered by them. Although there was no effect of discourse type on the subjectivity profiles of Russian connectives in the present corpus study, it could be the case that *poetomu* is still a less strong cue towards objectivity in the spoken than in the newspaper discourse. On the contrary, *tak chto* as a subjective connective that showed a strong preference for subjective causal relations across discourse types is assumed to be a strong cue towards subjectivity of the relation. To test these processing hypotheses, Study 2 was conducted.

4. Study 2: Russian causal connectives as processing instructions for the degree of subjectivity

The second study used an online sentence continuation experiment to investigate whether the Russian forward causal connectives *poetomu* and *tak chto* serve as processing cues towards objectivity and subjectivity of the upcoming relation, respectively. The sentence continuation experiment manipulated connective (*poetomu* vs. *tak chto*) and discourse type (newspaper discourse vs. spoken conversations) to test whether the expectations triggered by the two connectives differ across discourse types. Taking into account the results of the corpus analysis of absolute frequencies in Study 1 and previous literature on the influence of discourse type on subjectivity profiles (Stukker & Sanders, 2012; Zufferey, 2012), it was hypothesized that the difference between the continuations triggered by the Russian connectives would be smaller in the spoken as compared to the newspaper discourse. To test this hypothesis, the sentences that resulted from participants' continuations were analyzed for the type of relation they expressed.

4.1. Method

Participants

The participants of the online sentence continuation experiment were 40 native speakers of Russian (age: M = 35, SD = 14.7, range 18-67; 33 female; level of education: 28 higher, 5 incomplete higher, 5 vocational, 2 secondary). As with the connective insertion task, all participants read the information letter and signed an online informed consent form before the start of the experiment. There was no option to proceed without the consent. Participants were compensated with 450 rubles (5 euros) for the completion of the experiment. The study was approved by the Faculty Ethics assessment Committee - Humanities (FEtC-H) of Utrecht University.

Materials

The experiment employed a two-by-two design with 10 items per condition: there were two different causal connectives (*poetomu* vs. *tak chto*) and two types of discourse (spoken vs. newspaper). The experimental items were 40 three-sentence long stories where the last sentence was incomplete in such a way that the second clause directly following the forward causal connective was missing. An example of an experimental item is provided in (5). In the newspaper discourse type conditions, the items resembled extracts from newspapers and were presented in the written modality. In the spoken discourse type conditions, the items were presented as audio recordings. The stories were created in such a way that the participants could think of a possible continuation without much cognitive effort. For example, possible continuations of the spoken item in (5) could be "*it will take me more time that usual to get there*" (objective) or "*please, begin without me*" (subjective). Possible continuations of the newspaper item could be "*buses often miss the schedule*" (objective) or "*the mayor made a good decision*" (subjective). Another important criterion used in the creation of the stimuli was making sure that the most logical possible continuations were not already stated in the preceding context of the story.

(5) **Spoken:**

Privet! Da, ya znayu, chto vy vse uzhe tam. YA uzhe edu, pravda, navigator pokazyvaet, chto v centre vse stoit. V takoe vremya v gorode vsegda probki, **poetomu/tak chto** *…*

'Hi! Yes, I know you are all already there. I'm on my way, though the navigator shows that the center is stuck. At this time of the day there are always traffic jams in the city, **poetomu/tak chto** …'

Newspaper:

Mer prinyal reshenie vydelit' otdel'nuyu polosu dlya obshchestvennogo transporta na Dmitrovskom shosse. Eto pomozhet uluchshit' rabotu obshchestvennogo transporta, osobenno v utrennie chasy. V takoe vremya v gorode vsegda probki, **poetomu/tak chto** *…*

'The mayor decided to allocate a separate lane for public transport on Dmitrovskoe highway. This will help improve public transport connection, especially in the early hours. At this time of the day there are always traffic jams in the city, **poetomu/tak chto** …'

The spoken items were pre-recorded using two female Russian-speaking volunteers. There were also 40 filler items (20 spoken and 20 newspaper stories). The filler items copied the structure of the experimental items but used different connectives: *no* 'but' or *hotya* 'although'. The materials were piloted on two native Russian speakers not taking part in the experiment. This led to adjustment of several items where the pilot participants found the stories too difficult to continue. The stimuli were distributed across four experimental lists according to the Latin square design. For each of the four lists, the order of the stimuli was pseudo-randomized in such a way that: 1) there were no more than two experimental items or fillers in a row, 2) items of the same condition could not follow one another, 3) different conditions were equally distributed across the list. By reversing the four pseudo-randomized orders, eight experimental lists were created to avoid an effect of stimulus order. Each participant saw only one experimental list.

Procedure

The experiment was created using the Qualtrics software (https://www.qualtrics.com) and distributed online via social media platforms. The participants were instructed to read short excerpts from the newspapers or listen to the short excerpts from the spoken conversations and write down a possible continuation of the last sentence in each story. On average, the experiment took 60 minutes to complete. The participants could return to the task after partial completion within 72 hours from the start of the experiment. This option was used by seven participants. After completion of the experiment, the participants were redirected to another Qualtrics survey asking for their personal data and bank account numbers required for financial compensation. This was done to make sure that: 1) only those participants who completed the experiment could be compensated; 2) participants' personal data was not associated with their responses in the experiment.

4.2. Results

Responses of all 40 participants who completed the experiment were included in the analysis (including the seven participants who took more than one session to complete the task). The sentences resulting from participants' continuations were annotated as either subjective or objective following the same procedure as in the corpus analysis (see Table 2). Unintelligible responses and responses starting with the connective not matching the condition were excluded from the analysis (8 data points, which is 0.5% of the data). The distribution of subjective and objective continuations across conditions is presented in Table 8 and Figure 3. In general, the

participants inserted more subjective than objective continuations (837 vs. 755 continuations, respectively). Across discourse types, *tak chto* triggered more subjective than objective continuations (61%), and *poetomu* triggered more objective than subjective continuations (56%).
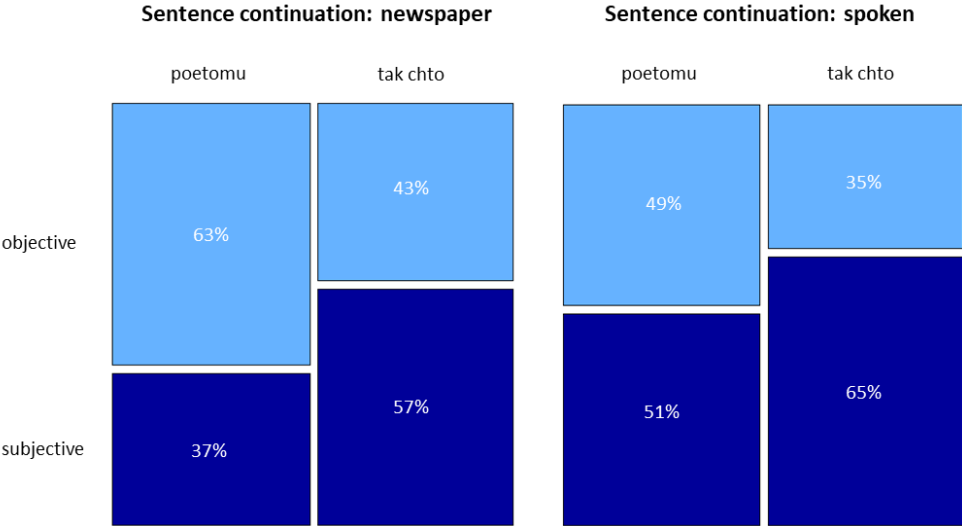
Table 8

The distribution of subjective and objective continuations across conditions

		Objective	Subjective
Newspaper	*poetomu*	253 (63.3)	147 (36.7)
	tak chto	170 (42.8)	227 (57.2)
Spoken	*poetomu*	194 (48.6)	205 (51.4)
	tak chto	138 (34.8)	258 (65.2)

Note. Percentages in the brackets some up to 100 horizontally per connective per discourse types.

Figure 3

The percentage of subjective and objective continuations for each connective across discourse types

To perform the statistical analysis, we built a generalized linear mixed-effect model with the relation type of the continuation as a binary dependent variable using *lme4* package (Bates, Mächler, et al., 2015; version 1.1.26) in R software (R Core Team, 2020). The model included

an interaction of connective and discourse type, both coded using scaled sum contrasts (-0.5, 0.5). Maximal random effect structure was gradually reduced to achieve convergence following the procedure described in Bates, Kliegl, et al. (2015). The results of the model are presented in Table 9.

Table 9

The results of the online sentence continuation experiment

Fixed effects	Estimate	SE	z-value	p-value
Intercept	0.14	0.12	1.17	0.24
Poetomu vs. *tak chto*	0.93	0.22	4.27	<.001
Newspaper vs. spoken	0.58	0.22	2.58	0.01
Connective : discourse type	-0.39	0.44	-0.88	0.38
Random effects	Variance	SD		
Items (intercept)	1.03	1.01		
Items (discourse type)	1.15	1.07		
Subjects (intercept)	0.13	0.36		
Subjects (discourse type)	0.11	0.33		

Note. Formula: relation type ~ connective*discourse type + (1 + discourse type| item) + (1 + discourse type| subject). *SE* – standard error; *SD* – standard deviation.

There were significant main effects of connective (z = 4.27, p <.001) and discourse type (z = 1.17, p = 0.01) on subjectivity of continuations, meaning that the odds of *tak chto* triggering subjective continuations were higher compared to *poetomu* and the odds of spoken extracts triggering subjective continuations were higher compared to newspaper extracts. The interaction between connective and discourse type turned out to be non-significant (z = -0.88, p = 0.38), suggesting that the difference between expectations triggered by the two causal connectives was not altered in spoken as compared to newspaper stories. In general, the analysis showed that more subjective continuations were inserted after *tak chto* than after *poetomu* and more subjective continuations were inserted in spoken extracts than in newspaper extracts.

4.3. Discussion

Study 2 set out to investigate how specialization of Russian forward causal connectives in terms of subjectivity is reflected in processing expectations that they trigger depending on different

types of discourse. Based on the outcome of the corpus analysis part of Study 1, it was hypothesized that the difference between the expectations evoked by *poetomu* and *tak chto* may depend on the discourse type. In Study 2, this effect would be reflected in the interaction between connective and discourse type in the sentence continuation task.

However, the results of Study 2 revealed no interaction between connective and discourse type suggesting that *tak chto* triggered significantly more subjective continuations than *poetomu* across newspaper and spoken discourse types. Instead, two main effects of connective and discourse type were found in Study 2, which means that both connective specialization and higher-order discourse characteristics have an independent effect on the speakers' expectations about the subjectivity of the upcoming content. Along with the specialization profiles of the two Russian connectives discovered in Study 1, *tak chto* evoked more subjective expectations than *poetomu*. At the same, more subjective relations are also expected in spoken discourse than in the newspaper discourse (58% vs. 47% of subjective relations, respectively). The prevalence of subjective continuations in the spoken items corroborates the hypothesis developed in Sanders and Spooren (2015) about the difference in salience of the Deictic Center of Communication between spoken and written discourse. While written discourse, such as newspaper articles, is rather detached from the Deictic Center of Communication (Sanders et al., 2009), spoken discourse with its direct availability of speaker and hearer can be considered more "subjectively grounded" (Sanders & Spooren, 2015: p. 65).

Although the difference between *tak chto* and *poetomu* in the percentage of subjective relations triggered turned out to be stable across discourse types, absolute figures for each connective reflect how discourse type impacts the effectiveness of each connective as a processing cue separately. In spoken discourse with its expected abundance of subjective causal relations, *poetomu* becomes a less effective cue towards objectivity of the relation (only 49% of objective continuations) than in the newspaper discourse (63% of objective continuations). In newspaper discourse, on the contrary, more objective relations are expected, and therefore, the expectations for subjectivity triggered by *tak chto* are attenuated (57% of subjective continuations evoked compared to 65% in spoken stories).

Although the sentence continuation task provides some insights into discourse processing by evaluating sentence continuations that speakers create on the spot, this method does not allow for direct investigation of online processing. Participants of the online sentence continuation experiment were asked to write down their responses, which gave them additional opportunities to process input, think over possible continuations and edit their responses

(Scholman et al., 2020). This means that Study 2 assessed rather offline interpretations than online processing, whereas the effect of specialized connectives on the online reading times as measured by the eye-tracking paradigms (Canestrelli et al., 2013; Li et al., 2017; Traxler, Bybee, et al., 1997; Wei et al., 2021) could be more salient. Further research employing the online processing paradigms mentioned is needed to identify whether this is the case and whether the effects of immediate processing delay and facilitation of the expected type of relation observed for specialized connectives in other languages are also present in Russian. Future research using eye-tracking while reading paradigms could also shed light on the role of discourse type in online processing effects of connective specialization.

5. General discussion

Previous literature on subjectivity in causality shows that languages often employ specialized connectives to express subjective versus objective relations (Degand, 2001; Degand & Pander Maat, 2003; Çokal et al., 2020; Li et al., 2013; Pit, 2007; Sanders & Spooren, 2015; Santana et al., 2021; Xiao et al., 2021). These specialized connectives serve as processing cues for the degree of subjectivity and as such, facilitate the processing of the upcoming content (Canestrelli et al., 2013; Li et al., 2017; Traxler, Bybee, et al., 1997; Traxler, Sanford, et al., 1997; Wei et al., 2019, 2021). Previous cross-linguistic research on this topic mainly employed corpus studies to investigate the subjectivity profiles of causal connectives. At the same time, there are emerging experimental methodologies, such as the connective insertion task, that can address the same research question (Santana et al., 2021). For several languages, there is also some work devoted to processing effects of the specialized connectives. However, it remains unclear how the results of all these methods relate to each other and to what extent they converge on providing the same picture with respect to subjectivity profiles of causal connectives in a language. The present study set out to address this issue by testing a multi-method approach to investigation of subjectivity profiles of causal connectives in a language, which involved the traditional corpus analysis, the connective insertion task and the sentence continuation task. The sentence continuation experiment as a method also helped to connect the typical corpus research with processing by looking at how causal connectives trigger different expectations with respect to subjectivity of the upcoming content depending on different discourse types. Testing the multi-method approach on Russian causal connectives that have not yet been studied with this respect provided a valuable contribution to the cross-linguistic research on subjectivity in causality.
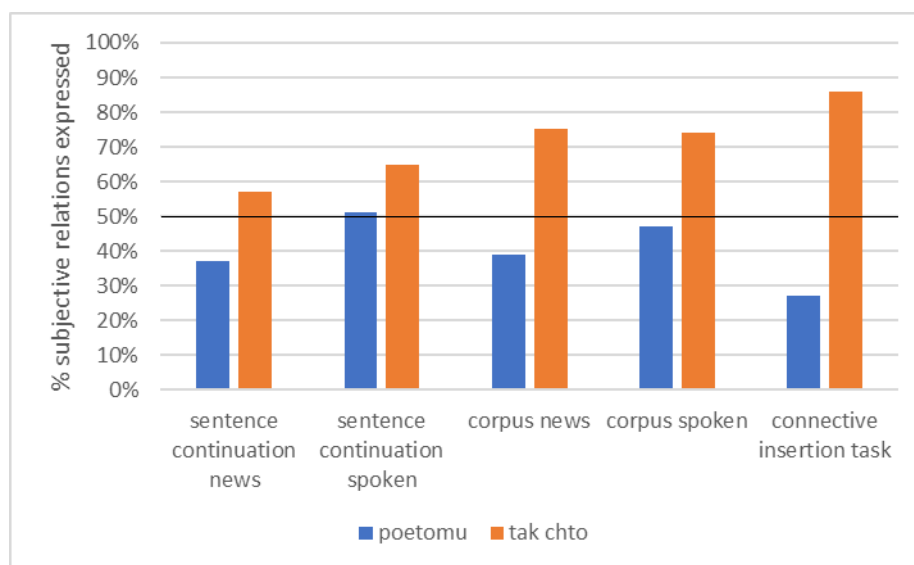
5.1. Comparison of different methods

The three methodologies used in the present study, i.e. the traditional corpus analysis, the online connective insertion task and the online sentence continuation task, provide converging evidence that the Russian causal connectives *poetomu* and *tak chto* differ with respect to the degree of subjectivity they encode. More specifically, *tak chto* is used in subjective relations more than *poetomu* and evokes more subjective expectations than *poetomu*. Thus, all the three methods show that Russian connectives do have different subjectivity profiles, where one is preferred in subjective relations more than another. This is an important conclusion that means that all the methods employed are valid methods of identifying whether certain connectives in a language are specialized in terms of subjectivity.

However, the results of the three methods also differ in the strength of the observed specialization of connectives. Figure 4 below reflects these differences by combining the results of the three methods in the same graph. The results of the corpus analysis and the sentence continuation task are split across discourse types.

Figure 4

Specialization of Russian causal connectives across the three methods

As Figure 4 shows, the connective insertion task resulted in the strongest specialization profiles of Russian causal connectives in terms of subjectivity. This strong specialization could be explained by the nature of the task. Firstly, this task takes clear prototypical subjective and objective cases as stimuli for connective insertions. On the contrary, the examples of causal relations in the corpora and the relations built in the sentence continuation task do not represent

pure subjective versus objective cases, since they reflect actual language use, where many other linguistic and situational factors are at stake (Graesser et al., 1997; Zwaan et al., 1995; Zwaan & Rapp, 2006). Because of this rarity of purest exemplars of causal relation types in the language (Stukker & Sanders, 2012), researchers conducting corpus analysis, for example, need to discuss their annotations of relation types among each other and sometimes come up with strategies on how to deal with unclear cases (Li et al., 2013).

Secondly, the connective insertion task requires participants to make a forced choice between the two causal connectives for a given causal relation. This procedure allows for little freedom of expression on the participants' part compared to the sentence continuation task, where participants can create continuations based on their interpretation of the context, their world knowledge or other factors that are not related to the given connective. Similarly, corpus examples consist of relations created by language users who are not restricted in the choice of linguistic expressions at all compared to the participants of the connective insertion experiment. Thus, little freedom of expression imposed by the forced binary choice task in combination with the use of prototypical pure cases of relation types as stimuli resulted in relatively large preferences for each connective to express its corresponding type of relation in terms of subjectivity.

At the same time, both in the corpus and in the sentence continuation task, the specialization of Russian causal connectives turned out to be less pronounced. This is again due to the fact that the relations in these two methods are examples of everyday language use and therefore do not represent prototypical subjective and objective cases. In general, an important difference between the connective insertion task on the one hand, and the corpus analysis and the sentence continuation task on the other hand is the approach to investigating connective specialization. The connective insertion task looks into which causal connectives are used in each type of relation, whereas the corpus analysis and the sentence continuation task examine which types of causal relations are possible for each connective. The approach where connectives are chosen for the predefined prototypical subjective/objective relations gives much clearer specialization profiles (Figure 4) for the reasons discussed in the previous paragraph.

The approach where types of relations are identified for each connective gives a less clear picture because it incorporates the influence of other factors that are not related to the distinction between subjective and objective connectives. In the sentence continuation task in particular, participants could base their expectations for the subjectivity of the upcoming

relation not only on the given causal connective, but also on the previous context of the story, on their world knowledge, or on higher-order discourse characteristics such as discourse type (Graesser et al., 1997; Zufferey et al., 2018; Zwaan et al., 1995; Zwaan & Rapp, 2006). In fact, the results of the sentence continuation task confirm statistically that discourse type was one of the factors that affected processing expectations: in spoken speech, the information is always presented by the speaker SoC and directly connected to the speaker (Sanders & Spooren, 2015), hence more subjective relations are expected and produced. Due to the relative prevalence of subjective relations in speech, the specialization of *poetomu* in spoken language is less pronounced (53% of objective relations in the spoken corpus) and hence its effectiveness as a cue is decreased (49% of objective relations in the sentence continuation task), which is clear from the comparison of spoken versus newspaper discourse type figures both for the corpus analysis and for the sentence continuation experiment. The less clear specialization of Russian connectives in the outcome of the corpus analysis and the sentence continuation task should not be considered a methodological disadvantage. On the contrary, these two methods provide a distribution pattern that is closer to the actual language use, which is influenced by many factors.

In general, an important conclusion of the present study is that no method can be considered the best for investigating subjectivity of the causal connectives in a language. Different methods shed light on slightly different aspects of connective specialization in a language and have their own advantages and disadvantages. The connective insertion task is a method that allows for the clearest, the fastest and the most effective estimation of subjectivity-related distinctions in the lexicon of causal connectives (Santana et al., 2021). It is easily administered, allows for collection of large datasets, and furthermore, the prototypical subjective and objective stimuli for this task can be easily translated and applied to other languages. Thus, a connective insertion experiment could be considered a good exploratory paradigm, with which the investigation of the subjectivity profiles of causal connective should start for a language that has not been previously studied with this respect. However, one needs to keep in mind that probable sharp contrasts in preferences between the connectives that will result from this task should be interpreted with caution since they are based on the most prototypical cases.

The corpus analysis and the sentence continuation task, on the contrary, provide a better insight into the influence of other discourse characteristics, such as discourse type, on the behavior of causal connectives with respect to encoding subjectivity. Naturally, the subjectivity

profiles resulting from these methods are less pronounced. The sentence continuation task in particular appears to be the least sensitive measure for connective specialization. A possible reason for this is that the participants of this task could not change the connective given in every story but could adjust their continuations to fit their expected relation type based on the other contextual and textual factors, even if this relation type was incongruent with the profile of the given connective. Several excluded cases where participants started their continuation with another causal connective that was not given in the condition support this idea. In the corpus and in the connective insertion task, on the contrary, using a certain connective was someone's "deliberate" choice, which partly reflects his or her categorization of the causal connectives in terms of subjectivity. Despite decreased sensitivity to the subjectivity profiles of the sentence continuation method, it can provide insight into processing instructions that are triggered by the causal connectives, which is an important phenomenon related to connective specialization. To sum up, there is a trade-off between proximity to the actual language use with its complex interplay of various factors and a quick evaluation of specialization of connectives on the dimension of subjectivity exclusively. Therefore, I would like to stress the importance of using several methods for investigating the subjectivity profiles of causal connectives, especially when these have been understudied in a language (Santana et al., 2021).
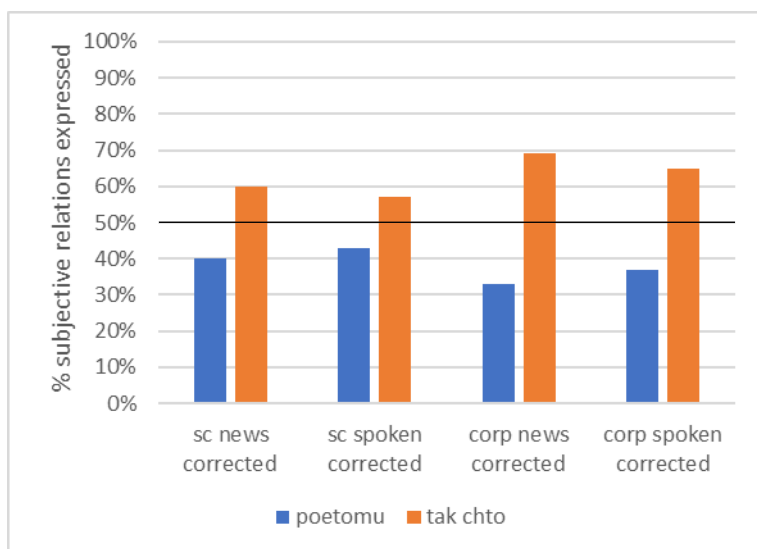
5.2. Correction for the imbalance between relation types

An important issue relevant for discussion of the methodological differences is the issue of imbalance between subjective and objective relation types across discourse types. In the present corpus study, there were more subjective than objective relations among the corpus examples in general, and this prevalence of subjective relations was relatively larger in the spoken than in the newspaper discourse. Similar overall relative prevalence of subjective over objective relations was observed in the previous corpus research on Spanish (Santana et al., 2018). In the sentence continuation task of the present paper, more subjective than objective relations were inserted in the spoken stories, but more objective than subjective relations were inserted in the newspaper extracts. In other words, when the number of relation types is not controlled for, like in the corpus analysis or in the sentence continuation experiment, there may be imbalance between the different types of causal relations, which can be caused by specific properties of the example selection, by discourse type or other possible factors. This imbalance is not accounted for in the analysis based on absolute frequency distributions in the corpus and in the sentence continuation task. To account for this relation type imbalance, the absolute frequencies of occurrence of each connective expressing specific relation type need to be corrected for (i.e.,

divided by) the number of occurrences of this relation type in this type of discourse. Figure 5 presents the outcome of the two methods if such correction is applied.

Figure 5

Specialization of Russian connectives based on frequencies corrected for the number of subjective/objective relations per discourse type

Note. *Sc* – sentence continuation task; *corp* – corpus analysis.

The comparison between Figures 4 and 5 reveals an interesting pattern: corrected figures are somewhat closer to the clearest specialization pattern observed in the connective insertion task. This is not surprising since the correction eliminates the influence of the relation type bias across discourse types. More specifically, correction for relative prevalence of subjective relations, which is the case in both spoken and newspaper corpora as well as in spoken stories in the sentence continuation task, results in a more specialized picture for *poetomu* and a less specialized picture for *tak chto*. In fact, the specialization of both connectives becomes comparable: for instance, in the corrected spoken corpus results (Figure 5), the preference of *poetomu* for expressing objective relations is 63% and the preference of *tak chto* for subjective relations is 65%, while in the spoken corpus without correction the figures are 53% and 74%, respectively (Figure 4). Indeed, if more subjective relations are expected and produced in spoken discourse, the fact that *poetomu* still prefers to be used in a considerable number of objective cases renders this connective intuitively more specialized, while for *tak chto* that is already preferring the dominant type of relation the logic is reversed. On the contrary, when objective relations are more common, like in the newspaper stories in the sentence continuation task, the informativeness of *tak chto* as a connective signaling subjectivity is enhanced, whereas

the effectiveness of *poetomu* should be decreased since it triggers expectations about already preferred relation type. To sum up, although absolute frequencies shed light on the natural usage profiles of causal connectives, the specialization in subjectivity as such is better reflected when the relation type bias is controlled for.

5.3. Subjectivity in Russian causal connectives

The findings of the present study demonstrate that Russian causality lexicon includes at least two connectives, *poetomu* and *tak chto*, that are specialized in terms of subjectivity. Study 1 shows that *tak chto* prefers to express subjective relations more than *poetomu* that is mostly used in objective relations. These preferences are also reflected in processing, as Study 2 suggests: namely, *tak chto* triggers more subjective expectations than *poetomu*. At the same time, according to the results of Study 2, people's expectations for subjectivity of causal relations are also influenced by discourse type: more subjective relations are expected in spoken conversations, while more objective relations are expected in newspaper discourse. This affects the subjectivity profiles of the connectives in such a way that *tak chto* becomes a less efficient cue for subjectivity in newspaper discourse (57% of subjective continuations), and *poetomu* turns out to be non-specialized and thus not a good cue for objectivity in spoken discourse (49% of objective continuations).

The findings of the sentence continuation task performed in Study 2 deserve special attention. As discussed above, they showed the least specialized picture for both Russian connectives in comparison to the corpus analysis and the connective insertion task. One possible reason for such an outcome is connected to the phenomenon of subjective construal (Verhagen, 2007): one and the same relation can be categorized as subjective or objective by different speakers based on their particular interpretation of the situation or their communicative purposes. The possibility of subjective construal leads to the abundance of relations that are not prototypical and not easy to categorize among the responses of participants of the sentence continuation task, which inevitably affects the results. A closer look at the participants' responses corroborates this reasoning. For instance, there was a considerable number of responses with *nuzhno 'need'*, which is a modal predicate roughly meaning one of the two things: 1) either *'have to'*, indicating external pressure of the circumstances, and as such, conveying more objective meaning, 2) or *'should'*, expressing the speaker's subjective point of view or advice. An example of a story with a continuation using *nuzhno* is presented in (6).

(6) *Poslushaj, ne bespokojsya, ya mogu tebe pomoch' ustroit'sya k nam. YA obyazatel'no pogovoryu s nashim menedzherom naschet tebya, kak tol'ko poyavitsya vozmozhnost'. Sejchas takoj vozmozhnosti net, **poetomu tebe nuzhno nemnogo podozhdat'.***

'Listen, don't worry, I can help you to get into our company. I will definitely talk to our manager about you as soon as there is a possibility. Now this is not possible, **poetomu you need to wait.'**

The continuation in (6) can either be interpreted as "you have to wait", which leads to objective relation, or as "you should wait/I advise you to wait", which leads to subjective relation. It is unclear which of the relations the respondent intended to produce and therefore, categorization of this relation into one or another type by the annotator may not coincide with the respondent's interpretation.

The participants of the sentence continuation task could also use various compensation strategies to compromise between the given connective on the one hand, and the type of relation they would like to construe based on the story on the other hand. For instance, in story (7), the respondent could have intentionally used an explicit speech act (*I am asking you to care*) instead of an implicit speech act (*please, care*) in order to adjust to the objectivity of *poetomu* but at the same time preserve the subjective relation that he or she wanted to insert. Examples like (7) are in line with the theory of prototypical structure of causal relations (Stukker & Sanders, 2012), according to which causal relations expressed by incongruent connectives in terms of subjectivity are non-prototypical instantiations of their relation type. In (7), the explicit speech act is a non-prototypical example of speech act relations and subjective relations in general, and therefore its co-occurrence with *poetomu* is not surprising.

(7) *Privet! YA zvonyu, chtoby poprosit' tebya pomoch' s nashimi cvetami, kak v proshlyj raz. My v sredu uezzhaem v otpusk na more na dve nedeli, a za cvetami pouhazhivat' nekomu. Im nuzhno postoyannoe vnimanie, **poetomu proshu tebya uhazhivat' za cvetami poka my budem v ot''ezde**.*

'Hi! I'm calling to ask you to help us with our flowers like last time. We are leaving on Wednesday for vacation at the seaside for two weeks, and there is no one to look after the flowers. They need constant attention, **poetomu I am asking you to care for them while we are away.'**

Another similar situation is illustrated with the common pattern of co-occurrence of *tak chto* with an explicit SoC phrase *I think* in the continuations (8). Stating an SoC of the subjective relation explicitly makes the relation less prototypically subjective (Stukker & Sanders, 2012). This strategy could have been used by the respondents who were forced to change their initially planned objective continuation into subjective relation because of *tak chto*.

(8)   *- Ty reshila, chto budesh' delat' so svoim razbitym steklom na ajfone? Pora kak-to etim zanyat'sya, a to skoro on sovsem perestanet rabotat'.*

*- Da, ya znayu. Normal'nyj remont mozhet stoit' bol'she 10 tysyach rublej, **tak chto, ya dumayu, mozhet, luchshe novyj kupit'.***

'- Have you decided what to do with your broken glass on your iPhone? It's time to do something about it, otherwise it will soon stop working at all.

- Yeah, I know. Good repair can cost more than 10 thousand rubles, **tak chto I think it is maybe better to buy a new one.**'

Another important remark about subjectivity of the Russian causal connectives is that the results of all the three methods used in the present study reflect only the significance of the difference between the two connectives, which is preserved across methods and discourse types. However, if we consider specialization of each connective separately, without reference to the other connective, a different analysis should be performed. Namely, it needs to be tested whether the distribution of subjective and objective relations expressed by each connective is significantly different from chance distribution (p = 0.5). In order to test this, additional four binomial tests were conducted per connective per discourse type. The corpus data was used to perform these tests, since only these data satisfy the requirement for independent observations. According to the binomial tests, *tak chto* expressed significantly more subjective than objective relations in both spoken (74%, $p <.001$) and newspaper (75%, $p <.001$) discourse, and therefore can be considered clearly specialized on subjective relations. *Poetomu*, in its turn, was clearly specialized in expressing objectivity only in the newspaper corpus (61%, $p = 0.04$), whereas in the spoken corpus, the distribution between objective and subjective relations expressed by *poetomu* was not significantly different from chance (53%, $p = 0.62$).

The specialization of the two Russian causal connectives in terms of subjectivity can be compared to other languages. In general, although there seems to be a division of labour between *poetomu* and *tak chto* with respect to expressing subjective versus objective relations, the differences in the degree of subjectivity encoded are rather relative and gradual. This

resembles the situation with backward causal connectives in Turkish (Çokal et al., 2020), where the objective connective *için* exemplified 31% of subjective relations across genres while the subjective connective *çünkü* expressed subjective relations in 54% of cases (compare to the difference of 43% vs. 75% subjective relations expressed in the corpus for *poetomu* and *tak chto*, respectively). However, in contrast to Turkish, where the objective connective seems to be more specialized than the subjective connective, the situation with *poetomu* and *tak chto* in Russian rather resembles the general pattern observed by Stukker & Sanders (2012) for languages like Dutch, French and German. The pattern is the following: while subjective connectives seem to have a robust preference for expressing subjective relations across discourse types, the objective connectives show less consistent profiles that vary across discourse types and studies (Degand & Pander Maat, 2003; Pit, 2003; Stukker & Sanders, 2012; Zufferey, 2012). Russian connective *poetomu* fits in this pattern: the corpus analysis and the sentence continuation task show that in the spoken discourse, this connective does not have strong preference for expressing objective relations as compared to the newspaper discourse. Similar situation is observed with backward causal connectives in French, where the connective *parce que* prefers objective relations only in the written discourse, while in the spoken discourse it expresses more subjective relations and gradually replaces the subjective connective *car*, which is extremely rare in the spoken language (Zufferey, 2012). In general, diachronic studies suggest that causal connectives are prone to subjectification (Traugott, 1995), which means that the tendency of objective connectives to become more subjective, especially in the inherently subjective discourse types such as spontaneous spoken conversations, is in line with the common language development pattern. In contrast to French, however, the Russian subjective connective *tak chto* is not being replaced by *poetomu* as evidenced by its moderately high frequency of occurrence in the spoken corpus.

In general, the presence of the subjectivity-related distinction in the lexicon of the Russian causal connectives puts Russian in line with the languages like Dutch, French, German, Mandarin and Turkish (Degand & Pander Maat, 2003; Çokal et al., 2020; Li et al., 2013; Pit, 2007; Sanders & Spooren, 2015; Santana et al., 2021; Xiao et al., 2021). Naturally, Russian is less similar in this respect to languages like Spanish (Santana et al., 2021) and English (Andersson & Sundberg, 2021; Knott & Sanders, 1998), where the distinction between the degree of subjectivity is not systematically encoded in the connective choice. As for processing, although the results of the present study cannot be directly compared to the eye-tracking studies conducted for other languages in the previous literature, they do indicate that *poetomu* and *tak*

chto provide different processing instructions with respect to subjectivity. This can be considered in line with the processing research on Dutch and Mandarin that shows how the difference in processing instructions of the specialized connectives is reflected in online reading times and in looks at the subject of consciousness (Canestrelli et al., 2013; Li et al., 2017; Wei et al., 2019, 2021). To conclude, the results for Russian obtained in the present study contribute to the body of literature corroborating importance of cognitive categorization into subjective and objective relations in the domain of causality.

6. Conclusion

The present study looked at the subjectivity profiles of Russian forward causal connectives across discourse types and methods. More specifically, this study aimed to investigate whether different methods provide converging evidence on the subjectivity profiles of causal connectives in a language. To address this issue, a multi-method approach combining the traditional corpus analysis, the connective insertion task and the sentence continuation task has been applied to identify the subjectivity profiles of two Russian forward causal connectives across discourse types. The results of the three methods provided converging evidence suggesting that the Russian lexicon of causal coherence markers contains connectives varying in the degree of subjectivity. Further research is needed to give a broader overview of subjectivity-related distinctions in Russian causality lexicon, both in terms of usage preferences and in terms of processing instructions. The evidence from Russian provided by the present study contributes to the field of cross-linguistic research on subjectivity in causality. More importantly, the discussion of the differences between the results obtained with different methods can serve as guidance for the future research in the field. The present study stresses the importance of using various methods to research the phenomenon of subjectivity reflected in discourse coherence markers.

References

Andersson, M. & Sundberg, R. (2021): Subjectivity (Re)visited: A Corpus Study of English Forward Causal Connectives in Different Domains of Spoken and Written Language. *Discourse Processes.* Advance online publication. https://doi.org/10.1080/0163853X.2020.1847581

Apresyan, V. Yu., & Pekelis, O. E. (2012). Podchinitel'nye soyuzy [Subordinate conjunctions]. In *Materialy dlya proekta korpusnogo opisaniya russkoj grammatiki* (http://rusgram.ru).

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. Article arXiv:1506.04967v2.

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Canestrelli, A. R., Mak, W. M., & Sanders, T. J. (2013). Causal connectives in discourse processing: How differences in subjectivity are reflected in eye-movements. *Language and Cognitive Processes, 28*(9), 1394–1413.

Çokal, D., Zeyrek, D., & Sanders, T. J. (2020). Subjectivity and objectivity in Turkish causal connectives? Results from a first corpus study on çünkü and için. In D. Zeyrek & U. Özge (Eds.), *Discourse Meaning: The View from Turkish* (pp. 223-248). De Gruyter Mouton.

Degand, L. (2001). *Form and function of causation: A theoretical and empirical investigation of causal constructions in Dutch*. Peeters.

Degand, L., & Pander Maat, H. (2003). A contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale. In A. Verhagen & J. van de Weijer (Eds.), *Usage based approaches to Dutch*, (pp. 175–199). LOT.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. In J. T. Spence, J. M. Darley, & D. J. Foss (Eds.), Annual Review of Psychology (Vol. 48, pp. 163–189). Annual Reviews.

Haag, E. O. (2004). *Funktsional'naya tipologiya i sredstva vyrazheniya prichinno-sledstvennykh otnoshenij v sovremennom russkom yazyke* [Functional typology and means of expression of cause-consequence realtions in contemporary Russian] [Doctoral dissertation, Tartu University]. TÜ väitekirjad alates 2004 - Theses, PhD, MSc, ETD. http://hdl.handle.net/10062/906

Keller, R. (1995). The epistemic *weil*. In D. Stein & S. Wright (Eds.), Subjectivity and Subjectification (pp. 16–30). Cambridge University Press.

Knott, A., & Sanders, T. (1998). The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of pragmatics, 30*(2), 135–175.

Le groupe λ-l. (1975). Car, parce que, puisque. *Revue Romane*, 10, 249–280.

Li, F., Evers-Vermeul, J., & Sanders, T. J. (2013). Subjectivity and result marking in Mandarin. *Chinese Language and Discourse, 4*(1), 74–119.

Li, F., Mak, W. M., Evers-Vermeul, J., & Sanders, T. J. (2017). On the online effects of subjectivity encoded in causal connectives. *Review of Cognitive Linguistics. Published under the auspices of the Spanish Cognitive Linguistics association, 15*(1), 34–57.

Li, F., Sanders, T. J. & Evers-Vermeul, J. (2016). On the subjectivity of Mandarin reason connectives - Robust profiles or genre-sensitivity? In N. Stukker, W. Spooren, & G. Steen (Eds.), *Genre in language, discourse and cognition* (pp. 15–49). De Gruyter Mouton.

Pander Maat, H. L. W., & Degand, L. (2001). Scaling causal relations and connectives in terms of speaker involvement. *Cognitive Linguistics, 12*(3), 211–245.

Pander Maat, H., & Sanders, T. J. (2001). Subjectivity in causal connectives: an empirical study of language in use. *Cognitive Linguistics, 12*(3), 247–273.

Pit, M. (2003). *How to express yourself with a causal connective? Subjectivity and causal connectives in Dutch, German and French*. Rodopi.

Pit, M. (2007). Cross-linguistic analyses of backward causal connectives in Dutch, German and French. *Languages in Contrast, 7*(1), 53–82.

R Core Team. (2020). *R: A language and environment for statistical computing* (Version 4.0.3). [Computer software]. https://www.R-project.org/.

Rabe, M. M., Vasishth, S., Hohenstein, S., Kliegl, R., & Schad, D. J. (2020). hypr: An R package for hypothesis-driven contrast coding. *Journal of Open Source Software, 5*(48), Article 2134.

Sanders, T. J. (1997). Semantic and pragmatic sources of coherence: On the categorization of coherence relations in context. *Discourse Processes*, *24*(1), 119–147.

Sanders, T. J., Sanders, J., & Sweetser, E. E. (2009). Causality, cognition and communication: A mental space analysis of subjectivity in causal connectives. In T. Sanders & E. Sweetser (Eds.), *Causal categories in discourse and cognition* (pp. 19–60). Mouton de Gruyter.

Sanders, J., Sanders, T. J., & Sweetser, E. E. (2012). Responsible subjects and discourse causality: how mental spaces and perspective help identifying subjectivity in Dutch backward causal connectives. *Journal of Pragmatics, 44*(2), 191–213.

Sanders, T. J., & Spooren, W. P. (2015). Causality and subjectivity in discourse: The meaning and use of causal connectives in spontaneous conversation, chat interactions and written text. *Linguistics, 53*(1), 53–92.

Sanders, T. J., Spooren, W. P., & Noordman, L. G. (1992). Toward a taxonomy of coherence relations. *Discourse processes, 15*(1), 1–35.

Santana, A., Nieuwenhuijsen, D., Spooren, W., & Sanders, T. (2017). Causality and subjectivity in Spanish connectives: Exploring the use of automatic subjectivity analyses in various text types. *Discours. Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics, 20.*

Santana, A., Spooren, W. P., Nieuwenhuijsen, D., & Sanders, T. J. (2018). Subjectivity in Spanish discourse: Explicit and implicit causal relations in different text types. *Dialogue & Discourse, 9*(1), 163–191.

Santana, A., Spooren, W., Nieuwenhuijsen, D., & Sanders, T. J. (2021). Do Spanish causal connectives vary in subjectivity? What crowdsourcing data reveal about native speakers' preferences. *Text & Talk, 41*(2), 211–237.

Scholman, M., & Demberg, V. (2017). Crowdsourcing discourse interpretations: On the influence of context and the reliability of a connective insertion task. In N. Schneider & N. Xue (Eds.), *Proceedings of the 11th Linguistic Annotation Workshop* (pp. 24–33). Association for Computational Linguistics.

Scholman, M. C., Demberg, V., & Sanders, T. J. (2020). Individual differences in expecting coherence relations: Exploring the variability in sensitivity to contextual signals in discourse. *Discourse Processes, 57*(10), 844–861.

Spooren, W., & Degand, L. (2010). Coding coherence relations: Reliability and validity. *Corpus Linguistics and Linguistic Theory, 6*(2), 241–266.

Stukker, N., & Sanders, T. J. (2012). Subjectivity and prototype structure in causal connectives: A cross-linguistic perspective. *Journal of Pragmatics, 44*(2), 169–190.

Stukker, N., & Sanders, T. J. (2014). Subjectivity and Prototype Structure in Causal Connective Use across Discourse Contexts. In M. Taboada & R. Trnavac (Eds.), *Nonveridicality and Evaluation: Theoretical, Computational and Corpus Approaches* (pp. 153–187). Brill.

Stukker, N., Sanders, T. J., & Verhagen, A. (2009). Categories of subjectivity in Dutch causal connectives: a usage-based analysis. In T. J. Sanders & E. Sweetser (Eds.), *Causal categories in discourse and cognition*, (pp. 119–171). Mouton de Gruyter.

Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure.* Cambridge University Press.

Traugott, E. C. (1995). Subjectification in grammaticalization. In D. Stein & S. Wright (Eds.), *Subjectivity and Subjectivisation: Linguistic Perspective* (pp. 31–54). Cambridge University Press.

Traxler, M. J., Bybee, M. D., & Pickering, M. J. (1997). Influences of connectives on language comprehension: Eye-tracking evidence for incremental interpretation. *Quarterly Journal of Experimental Psychology, 50*(3), 481–497.

Traxler, M. J., Sanford, A. J., Aked, J. P., & Moxey, L. M. (1997). Processing causal and diagnostic statements in discourse. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(1), 88–101.

Verhagen, A. (2007). Construal and perspectivisation. In D. Geeraerts & H. Cuyckens (Eds.), *Handbook of Cognitive Linguistics* (pp. 48–81). Oxford University Press.

Webber, B. (2009). Genre distinctions for discourse in the Penn TreeBank. In K. Y. Su, J. Su, J. Wiebe, & H. Li (Eds.), *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP* (pp. 674–682). Association for Computational Linguistics.

Wei, Y., Evers-Vermeul, J., Sanders, T. J., & Mak, W. M. (2021). The Role of Connectives and Stance Markers in the Processing of Subjective Causal Relations. *Discourse Processes*. Advance online publication. https://doi.org/10.1080/0163853X.2021.1893551

Wei, Y., Mak, W. M., Evers-Vermeul, J. & Sanders, T. J. (2019). Causal connectives as indicators of source information: Evidence from the visual world paradigm. *Acta Psychologica*, 198, 102866.

Xiao, H., Li, F., Sanders, T. J., & Spooren, W. P. (2021). How subjective are Mandarin reason connectives? A corpus study of spontaneous conversation, microblog and newspaper discourse. *Language and Linguistics, 22*(1), 166–210.

Yung, F., Demberg, V., & Scholman, M. (2019). Crowdsourcing discourse relation annotations by a two-step connective insertion task. In A. Friedrich, D. Zeyrek & J. Hoek (Eds.), *Proceedings of the 13th Linguistic Annotation Workshop* (pp. 16–25). Association for Computational Linguistics.

Zufferey, S. (2012). "Car, parce que, puisque" revisited: Three empirical studies on French causal connectives. *Journal of Pragmatics, 44*(2), 138–153.

Zufferey, S., Mak, W. M., Verbrugge, S., & Sanders, T. J. (2018). Usage and processing of the French causal connectives 'car' and 'parce que'. *Journal of French Language Studies 28*(1), 85–112.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(2), 386–397.

Zwaan, R. A., & Rapp, D. N. (2006). Discourse comprehension. In M. A. Gernsbacher & M. J. Traxler (Eds.), *Handbook of psycholinguistics* (pp. 725–764). Elsevier.