ISSN 1392-0340 (Print)
ISSN 2029-0551 (Online)
https://doi.org/10.15823/p.2020.137.11

Pedagogika / Pedagogy
2020, t. 137, Nr. 1, p. 193–207 / Vol. 137, No. 1, pp. 193–207, 2020

VYTAUTAS MAGNUS
UNIVERSITY
EDUCATION
ACADEMY

Annotating Causal Discourse Markers 'because' and 'so' for Learning English as a Foreign Language

Giedrė Valūnaitė-Oleškevičienė[1], Sigita Rackevičienė[2], Nijolė Burkšaitienė[3], Liudmila Mockienė[4], Jolita Šliogerienė[5], Dalia Gulbinskienė[6]

[1] Mykolas Romeris University, Ateities g. 20, LT-08303 Vilnius, Lithuania, gvalunaite@mruni.eu
[2] Mykolas Romeris University, Ateities g. 20, LT-08303 Vilnius, Lithuania, n.burksaitiene@mruni.eu
[3] Mykolas Romeris University, Ateities g. 20, LT-08303 Vilnius, Lithuania, sigita.rackeviciene@mruni.eu
[4] Mykolas Romeris University, Ateities g. 20, LT-08303 Vilnius, Lithuania, liudmila@mruni.eu
[5] Vilnius Gediminas Technical University, Saulėtekio al. 11, LT-10223 Vilnius, Lithuania, jolita.sliogeriene@vgtu.lt
[6] General Jonas Žemaitis Military Academy of Lithuania, Šilo g. 5A, LT-10322 Vilnius, Lithuania, dalia.gulbinskiene@lka.lt

Abstract. This paper reports on an investigation of the functions of the English causal discourse markers 'because' and 'so' and their Lithuanian counterparts and analyses the Lithuanian translation equivalents of these discourse markers using the multilingual open translation project TED Talks as a data resource. The results revealed that the discourse marker 'because' and its Lithuanian counterparts express ideational cause, followed by its rhetorical equivalent. However, the discourse marker 'so' and its Lithuanian counterparts express rhetorical consequences, followed by the rhetorical specification.

Keywords: *causal relations, discourse markers, parallel corpus, cross-linguistic discourse annotation, translation, foreign language learning.*

Introduction

In the last decades, foreign language learning through the use of corpora has been widely discussed in the research literature (Sinclair, 2004; Aijmer, 2009). It is accentuated that corpora are indispensable tools for both indirect use (teaching syllabi design) and direct use (data-driven learning) by foreign language learners (Römer, 2008; McEnery & Xiao, 2011). Bilingual parallel corpora provide benefit to foreign language teaching – they

enable the possibility of contrasting language patterns in foreign and native languages, developing the awareness of their similarities and differences, as well as analyzing and evaluating translations.

One of the most significant areas of teaching a foreign language at more advanced levels is the development of awareness of discourse modes and discourse management, as well as pragmatic features of discourse elements, i.e. abilities to comprehend and use appropriately linguistic devices to construct the coherent information representation and to express pragmatic aspects of communication. In scientific literature, such devices are referred to as discourse markers. Though their main function is to convey coherence of a message, they are also used to express the attitudes towards the message and its addressee (Crible & Degand, 2019).

Corpora provides a wide range of possibilities to develop the discourse management abilities of foreign language teachers. One of such possibilities is a manual annotation of semantic-pragmatic functions of discourse markers in parallel corpora. It reveals the extensive variation of discourse markers, enables the learner to deeper comprehend their functions and expediency in different languages, and thus enhances their usage in the learner's spoken and written language. As Valūnaitė-Oleškevičienė et al. pointed out, "discourse annotated corpora illuminate qualitative differences between the first language and the second language discourse marker use, especially in the complex cases and could be used as supplemental teaching/learning material for raising discourse management awareness of more advanced learners" (Valūnaitė-Oleškevičienė et al., 2018).

The present paper presents a case study of manually annotated English causal discourse markers (conjunctions 'because' and 'so') and their Lithuanian counterparts in a parallel corpus composed of broadcast presentations on various topics in English (TED talks) and their Lithuanian translations in the form of subtitles. The choice of the corpus sources was conditioned for several reasons. Firstly, the linguistic research on discourse markers has revealed that semantic-pragmatic variation of discourse markers is much larger in the spoken discourse "where they perform additional functions related to the management of the interaction" (Crible et al., 2019, 140). Though TED talks cannot be considered as representing a typically spoken discourse (they are prepared in advance and presented in a monologue form), they contain many features characteristic of spoken discourse and provide possibilities to perform contrastive analysis of discourse markers in several languages. Management of spoken discourse is especially relevant in the development of public speaking skills which is one of the most important objectives of foreign language teaching at more advanced levels. Secondly, even though discourse markers express similar functions in different languages, their usage schemes have important peculiarities (Zufferey & Degand, 2013). This culture-bound nature of discourse markers poses particular challenges to language learners as they have to acquire a new linguistic-cultural framework of the foreign language. Analysis of translations of discourse markers helps coping with these challenges. It sheds light on the peculiarities of discourse markers in

source and target language which cannot be discovered in monolingual analyses and enables to better understand the specificity of discourse markers in the source language. Therefore, parallel data provide added value to corpus-driven the investigation and contribute to the development of learner's discourse management skills.

Thus, the presented case study has a ***twofold aim***: to establish pragmatic aspects of the causal markers 'because' and 'so' and their Lithuanian counterparts in the compiled ad hoc parallel corpus of Ted Talks and to analyze the Lithuanian translation variants of the English discourse markers. In order to achieve the given aim, the following ***objectives*** were pursued:

- presentation of the research framework and research methodology;
- annotation of pragmatic domains and functions of the selected discourse markers using the framework developed by Crible (Crible & Degand, 2017);
- quantitative analyses of the annotated values and interpretation of their results;
- quantitative analysis of the lexical equivalents of the English discourse markers in the Lithuanian translations (subtitles) and insights into the reasons for the received results;
- developing a model for students' autonomous investigation of discourse markers;
- recommendations to foreign language teachers to develop the awareness of discourse markers in the classroom.

The authors' pedagogical experience shows that the case studies of this kind require learners' foreign language skills of more advanced levels, as well as analytical abilities to tackle a functional variety of discourse management devices, to identify their functions, to systematize and generalize the collected data. Despite the complexity of the task, the case studies of this kind serve as a significant tool in the development of foreign language learners' discourse management abilities, as well as their linguistic awareness. Data-driven learning enables students to discover themselves linguistic diversity in real usage of languages, get insights into regular patterns, understand the importance of the mode of discourse, and the role of the recipients in the use of linguistic devices and their translation.

Theoretical background

As it was mentioned above, discourse markers play a special role in a spoken discourse where they perform various pragmatic functions. Crible defines them as a special category of pragmatic markers the role of which is "to function on a metadiscursive level as a cue to situate the host unit in a co-built representation of on-going speech" (Crible, 2014, 3). Their functions range from signaling relations between different information units to expressing the speaker's relation to the information being presented and to the audience. As Crible and Degand state, they help "speakers convey not only the

coherence of their intended message but also their attitude towards this message and towards the interlocutor" (Crible & Degand, 2019, 3–4). According to Sun, discourse markers enable speakers "to make their presence felt in the text, to give guidance to the audience as to how the text is organized, what processes are being used to produce it, and what the speaker's intentions and attitudes are regarding the subject matter, the readers, and their text" (Sun, 2013, 2137). Thus, discourse markers "do not contribute much to the propositional content of a message but modify it in various subtle ways" (Buysse, 2010).

Usually, discourse markers are grouped into relational discourse markers (pragmatic connectives), such as *and, but, because, actually*, and non-relational discourse markers (pragmatic particles), such as *well, I mean, you know* depending on their function in context. The main function of relational discourse markers is to signal a two-place relation between a host unit and its context while non-relational discourse markers perform various metadiscursive functions related to structuring and punctuation, interpersonal management, etc. (Crible, 2014, 15–17).

Thus, discourse markers play an important role in spoken communication, the efficiency of which depends heavily on their effective use. Therefore, their acquisition of paramount importance in foreign language learning, especially at more advanced levels. The acquisition of discourse markers enables foreign language learners to better understand and interpret the speech of native speakers and to produce more fluent and coherent speech themselves. However, polyfunctionality of discourse markers and their culture-bound nature make their acquisition a challenging task for foreign language learners.

In the last decade, quite a few studies have been conducted on the acquisition of discourse markers by foreign language learners. In his doctoral dissertation, Buysse (2010) examined the use of the discourse markers by Belgian learners of English in spoken interviews. The researcher compiled and compared two learner sub-corpora containing interviews with undergraduates majoring in English linguistics and in commercial sciences, as well as contrast both learner sub-corpora with the native speaker corpus. The results revealed that the learners prefer markers with more structural functions *(so, well)* and neglect the ones with more interpersonal functions *(you know, I mean, sort of)*. The most frequent discourse marker among all investigated speakers appeared to be the marker 'so'. However, it was used significantly more often by foreign language learners than by native speakers (Buysse, 2010). Polat (2011) reported the results of the study on the use of discourse markers *you know, like, well* by a naturalistic learner of English using developmental learner corpus reflecting the progress of the learner during a year. The results of the study revealed very different developmental patterns for three discourse markers and indicated the usefulness of developmental learner corpora in the studies of pragmatic acquisition. According to the results of the study, the marker *you know* was heavily overused by the naturalistic learner at the beginning of the study and its usage decreased during the year, the usage of *like* was fluctuant while the marker *well* was not used at all during the study. The presented results of the studies by Buysse and Polat

reveal that acquisition of discourse markers may be very different among classroom and naturalistic learners.

Sadeghi & Yarandi (2014) performed a study in which they examined the impact of explicit teaching of discourse markers on their usage in spoken language by Iranian learners of English. The researchers examined the usage of discourse markers in two groups of students with the same level of English proficiency. Prior to the examination, the first group was given explicit teaching on discourse markers in five classes throughout a month, while the second group did not get any explicit instruction on this topic. During the research examination, the same conversation texts were given to both groups and the students were asked to retell the conversations in their own wording. The findings indicated that explicit teaching had a significant impact on the use of discourse markers: the first group demonstrated much better results than the second one even though they did not always use them appropriately.

The conducted studies show that the acquisition of discourse markers requires special endeavors of foreign language learners. It is also evident that "some of the problems originate from insufficient information about discourse markers" (Sadeghi & Yarandi, 2014). Thus, knowledge of discourse markers, their types and functions contribute significantly to the development of awareness of discourse markers, their comprehension, and usage.

The data-driven learning method provides the learners with the possibility to acquire knowledge on discourse markers, their types and functions, through a corpus-driven bilingual case study.

Two discourse markers, the primary function of which is to signal causal relations, were chosen for the case study. The selection of the discourse markers was conditioned by their role in the comprehension and construction of coherent information representation in a text. Causal discourse markers enable to express logical relations between discourse segments and formulate argumentative messages; therefore, they play a central role in argumentative writing and speaking. Moreover, cognitive linguistic research has shown that causally related information is remembered better and processed faster (Mulder, 2008, 12–13). Thus, causal discourse markers could be seen as devices enhancing the learning process. The enumerated qualities make them especially important in learning a foreign language. Usually, teaching concentrates on the usage of discourse markers in written language. However, their use in spoken discourse has significant peculiarities, the awareness of which is important for the development of speaking skills of foreign language learners.

Research methodology

As given in the objectives, the present case study was conducted attempting to develop a model for students' autonomous investigation of discourse markers which would enable

them to gain knowledge on discourse markers, their functions and use. Such data driven learning, in contrast to explicit teaching, provides much more effective results as it allows the learners themselves to raise their own awareness of linguistic phenomena, to deeper understand them and apply more effectively in communication in a foreign language.

For the manual annotation of the functions of the extracted discourse markers, the modified functional taxonomy for spoken data developed by Crible was chosen (Crible & Degand, 2017). The taxonomy covers functions of discourse markers within four domains: (1) ideational domain which refers to the events happening in the real world and to semantic relations between such events; (2) rhetorical aspect concerns the speaker's subjective attitude towards the information being presented; (3) sequential domain which refers to "the structuring of discourse segments, both at macro- and micro-level", and (4) interpersonal domain concerns "the interactive management of the exchange, in other words to the speaker-hearer relationship" (Crible, 2014, 18; Crible & Degand, 2017, 5). Discourse markers can perform 15 functions which can be expressed in any aspect depending on the context. Schematically the taxonomy is presented in Table 1 below.

Table 1
Crible's revised taxonomy

Ideational		**Rhetorical**		**Sequential**		**Interpersonal**	
[addition]	[alternative]	[cause]	[closing]	[concession]	[condition]	[consequence]	
[contrast]	[enumeration]	[opening]	[punctuation]	[resuming]	[temporal]	[topic-shift]	[specification]

(Adapted from Crible & Degand, 2017, 18)

Such taxonomy allows the annotators to choose "to start at domain-level or function-level, to annotate both levels simultaneously or independently, and could even decide to stop at one level if a particular domain DM token is under-specified for the other level" (Crible & Degand, 2017, 20).

In the process of the manual annotation of the English and Lithuanian discourse markers, the additional technique of translation spotting was employed to deal with complicated cases. This technique enables to analyze ambiguous markers in source language using parallel data from target language (Cartoni et al., 2013). Besides, this technique enables to analyze differences between discourse markers used in foreign and native languages, which is particularly important in cases where there are no equivalents (Danlos & Roze, 2011).

Manual annotation is usually carried out by several annotators assigning a label from a list of functions and domains for each discourse marker. The inter-annotator agreement is important to validate the annotations thus it is advisable to have a team of annotators working on the set of data.

For the study, 60 cases of 'because' and 90 cases of 'so' and their Lithuanian counterparts were extracted from the compiled parallel corpus of TED talks and manually annotated. The received data were analyzed by quantitative approach. First, the functions performed by the annotated English and Lithuanian discourse markers were analyzed; secondly, the Lithuanian translation equivalents of a parallel corpus of Lithuanian transcripts of TED talks of the English discourse markers were investigated. All the examples that were analyzed and presented in the article were taken from the corpus of TED talks transcripts in English and Lithuanian. Research findings are presented below.

Research findings

The results of the performed analyses are presented in two sections. Section 4.1 presents the results of the quantitative and qualitative analyses of the manual annotation of the English discourse markers 'because' and 'so' and their Lithuanian counterparts which were labelled with a domain and a function from the chosen functional taxonomy (Crible & Degand, 2017). Section 4.2. presents the results of the quantitative analysis of the Lithuanian translation variants of the English discourse markers.

4.1 The use of the discourse markers 'because' and 'so' and their Lithuanian equivalents

Discourse marker 'because' and its Lithuanian counterparts

The results of the current research reveal that the discourse marker 'because' and its Lithuanian translation equivalents are pertinent to causative meaning in most cases in the sample, 60% of the cases convey ideational cause in both languages (Figure 1).

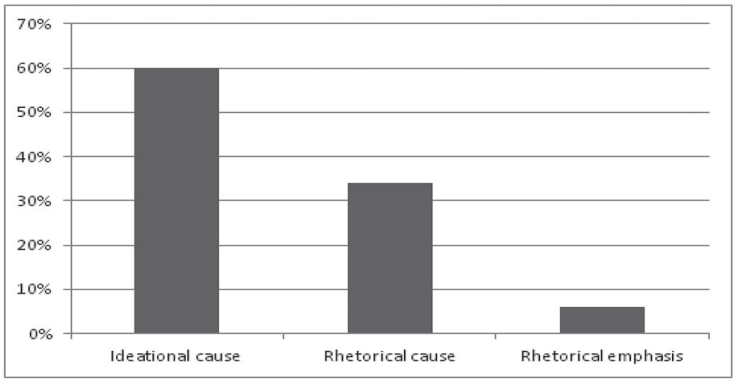


Figure1. The annotated values of the discourse marker 'because' and its Lithuanian counterparts

While analyzing the results, it is important to emphasize that based on the notion that the ideational aspect is related to events, ideational cause reveals causative meaning based on facts, for example (1):

(1) I am not sure, because I was not frank with him.

(1) Aš nesu tikras, nes nebuvau atviras su juo.

It is also identified that 34% of the cases in the annotated examples are related to the rhetorical cause, and as the rhetorical cause is connected to the person's expressed subjectivity the functional meaning of the discourse marker creates the effect of a subjective cause expressed by the speaker, for example (2):

(2) [Because] I believe that technology is so vital that it offers us unlimited opportunities.

(2) [Kadangi] tikiu, kad technologija yra tokia svarbi, kad atveria mums neribotas galimybes.

While analyzing the particular example above it could be seen that the rhetorical aspect is linked to the whole statement and in this case, the phrase used in the argument *I believe* may add to the subjectivity of the discourse marker 'because' in this situation. In addition, it could be observed that the rhetorical meaning of 'because' could be also connected to a broader situation, which expresses subjective connotation, see (3) and (4):

(3) So this is my most frequently used webpage, not least [because] it was created by psychologists.

(3) Tai yra mano dažniausiai lankomas tinklalapis, ir tikrai ne [dėl to, kad] buvo sukurtas psichologų.

(4) We must claim our beliefs in films, [because] if we conceal our true stories and make them populist, we will fail.

(4) Privalome susigrąžinti savo tikėjimą filmais, [nes] jei nuslėpsime savo tikras istorijas vardan populizmo, mes pralaimėsime.

In the annotated sample the discourse marker 'because' is not related to the causal function in 6 % of the cases as it is used for rhetorical emphasis. The example below illustrates one of the cases of rhetorical discourse structuring. Interestingly, in the particular example, causal marker (*[Because] for* me) is also related to the surrounding situation expressing speaker's point of view, see (5).

(5) [Because] for me, figures aren't just figures.

(5) [Matot], man skaičiai – ne šiaip skaičiai.

Summarizing the research results it is evident that the cases in the sample mainly reveal the ideational cause and are related to the ideational factual domain.

Discourse marker 'so' and its Lithuanian translation equivalents

Different results are revealed concerning the analysis of the discourse marker 'so', which demonstrate that 'so' and its Lithuanian translation equivalents are related to the rhetorical aspect and are used to express consequence, followed by the sequential aspect and only a tiny 3% of the annotated cases functioning in the ideational domain (Figure 2).

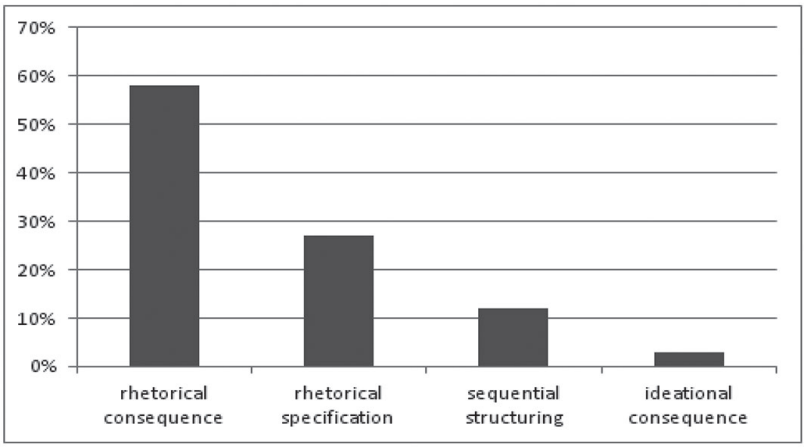


Figure 2. The annotated values of the discourse marker 'so' and its Lithuanian counterparts

According to figure 2, in both languages 58% of the annotated cases in the sample are related to rhetorical consequence and 27% of cases convey rhetorical specification. The findings reveal that the person's opinion is conveyed and the effect of subjectivity is produced. The example below can illustrate the function of rhetorical consequence, see (6):

(6) And [so], to try to convince you that mathematics is fantastic, I will give you my some examples.

(6) [Na] ir norėdama jus įtikinti, kokia išties nuostabi yra matematika, pateiksiu jums keletą pavyzdžių.

The research results also reveal that 12% of the cases in the annotated sample are related to the sequential sphere, which implies that 'so' and its Lithuanian equivalents fulfill the function of the sequential structure, see (7):

(7) [So] the math proves that first of all, you should cross out the potential partners.

(7) [Taigi] matematika įrodo, kad pirmiausia jūs turėtumėte išbraukti potencialius partnerius.

The findings reveal that the cases of 'so' with the meaning of ideational consequence are rare (3%).

(8) If you don't listen to mathematicians, [so] you have to reject everybody and die lonely.

(8) O jeigu neklausysite matematikų, [taigi] turėsite visus išbraukti ir mirsite vienas.

The research results identify that the discourse marker 'so' in the annotated sample mainly functions in the rhetorical and sequential domains rather than the ideational one.

4.2. Translation variants of the discourse markers 'because' and 'so'

Translations of discourse marker 'because'

The second stage of the research revealed what Lithuanian equivalents are used for the translation of 'because'. The translation values are demonstrated in Figure 3.

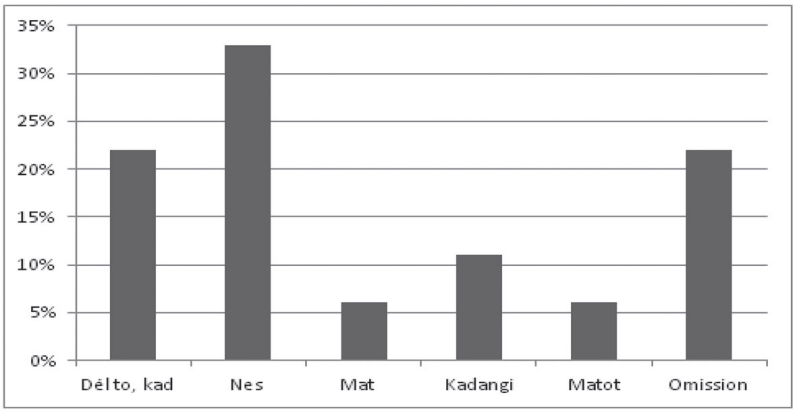


Figure 3. Translation values of the discourse marker 'because'

The research reveals that the most frequent translation choices into Lithuanian for 'because' comprise *nes* (33%) and *dėl to, kad* (22%), *kadangi* (11%). It could be observed that all the variants mentioned above are the choices provided by Ted talk subtitle cases, see (9), (10) and (11):

(9) [Because] the ones who hate you are going to hate you anyway.

(9) [Nes] tiems, kurie jūsų nekenčia, vis tiek jūsų nekęs.

(10) It is just [because] our society is so smart and so wonderful.

(10) Taip yra [dėl to, kad] mūsų visuomenė yra labai protinga ir nuostabi.

(11) Now, [because] they're psychologists, they gather information on human beings.

(11) [Kadangi] jie psichologai, jie renka informaciją apie žmones.

It is also interesting to observe that 12% of Lithuanian equivalents distinguished in the research sample comprise the particle *mat* (6%) and the parenthetical verb *matot* (6%). It should be admitted that both translation choices are appropriate because the chosen variants convey the rhetorical meaning of the discourse marker 'because' in the given context in the source language. It reveals that the interpreters emphasized the pragmatic function of the text, see (5) above and (12):

(12) But I am also sure that mathematics has got a lot to offer us [because], it is all about the study of cases.

(12) Ir taip pat žinau, jog matematika turi daug ką mums pasiūlyti, [mat] matematika yra apie atvejų studijavimą.

The study revealed that *because* was not translated in 22% of cases, see (13):

(13) I found it difficult to say it, [because] I didn't wish to be judged.

(13) Man buvo sunku tai ištarti. Nenorėjau būti pasmerktas.

These cases are classified into two groups. The first group of omissions comprises the cases when multiple discourse markers or more than one discourse marker introduce the statement in the source language. Such a feature as abundant use of discourse markers is particularly characteristic to spoken discourse. In such cases, the translator choice was

to translate only one marker into Lithuanian, which helps to draw a conclusion that the translator's choice could have been predisposed by the requirements of synchronizing the subtitles to make them laconic. The example (14) analyzed below reveals the translator's decision to translate *when* by *kai* and to omit 'because' keeping the temporal meaning and also observing the requirements of the synchronization, see (14):

(14) [Because] [when] people select their photos for their website, they choose the attractive ones.

(14) [Kai] žmonės pasirenka savo nuotraukas internetiniam puslapiui, jie pasirenka patraukliausias.

More cases of omission were established where it was chosen to translate the text into Lithuanian by using different grammatical structures. The example (16) demonstrates a successful translator choice to change the statement in order to convey the overall meaning, which leads to the omission of the discourse marker 'because', see (16):

(16) I convinced her it was [because] I was gaining life experience, not that I was seeking for popularity.

(16) Įtikinau ją, kad šie pokyčiai – tai dalis mano gyvenimiškos patirties, o ne populiarumo paieškos.

The above discussed observations lead to the conclusion that most translation choices were appropriate in rendering both the semantic and pragmatic values of the marker.

Translations of discourse marker 'so'

The findings from the second stage of the study of the translations of 'so' showed similar results to those of the discourse marker 'because' which proved that the most common variants of translating 'so' into Lithuanian were *taigi* (35%) and *tam, kad* (2%), which are the choices found in English-Lithuanian dictionaries (see translations for the discourse marker 'so' in Figure 4).

In the rest of the cases, 'so' was translated by particles *tai* (6%), *na* (4%) and the parenthetical verb *tarkim* (6%). The latter results were established when the translator chose to use the translation strategy of transposition, i.e. grammatical forms in the source and target languages differ. Example (17) below presents the case in which the discourse marker 'so' was translated by the parenthetical verb *tarkim* that performs the function of rhetorical specification in Lithuanian translation:

(17) [So] if you take Tom, for example, everybody notices that Tom is a very handsome man.

(17) [Tarkim], pateikime Tomą kaip pavyzdį, visi pastebės, kad Tomas yra gražus vyras.

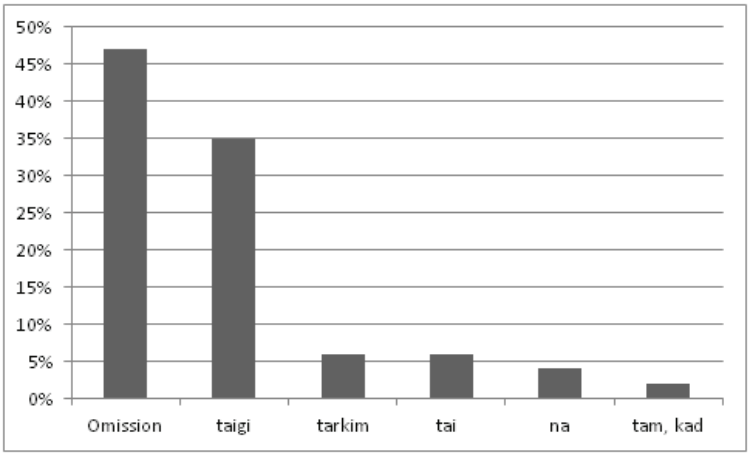


Figure 4. Translation values of the discourse marker 'so'

The research revealed that the translator used the translation strategy of omission in 47% of cases for the discourse marker 'so'. The omission technique was applied in the cases where 'so' in the source language was used in the rhetorical sphere or for sequential structures, which is characteristic of spoken language, see (18):

(18) [So] let's say, that you start going out at the age of 16 and, you'd like to marry by the age of 30.

(18) Tarkime, kad pradedate draugauti, kai jums 16, ir norėtumėt susituokti trisdešimties.

This phenomenon of the translator choices could have the explanation that the translation is synchronized so that the subtitles can be quickly and easily read. The guidelines for making subtitles indicate that the subtitles have to be of adequate length for the audience to read when they are shown on the screen. (TED Translators Wiki, 2017).

Conclusions

Foreign language teachers should be aware that manual annotation of the functions of discourse markers and their analysis requires foreign language skills of more advanced levels, as well as abilities to categorize the data according to the given taxonomy and analyze them using quantitative and qualitative methods. Foreign language teachers should help their students to acquire the necessary skills and abilities. Though the process is not straightforward and requires extra efforts, it provides unique possibilities to the learner to obtain linguistic knowledge on discourse markers by discovering autonomously regularities in the foreign language and comparing them with the patterns in one's native language.

The research results reveal that the meaning that 'because' and its Lithuanian equivalents express is causative, because more than fifty percent of all instances in the analyzed

corpora indicate ideational cause. We also found out that in terms of discourse management more than thirty per cent of the instances have a close association with the rhetorical domain as personal opinions of the speakers were signified by them. As concerns discourse management and the discourse marker 'so', the research showed that in the analyzed corpora this discourse marker and its Lithuanian equivalents impart meaning and are used mainly in the rhetorical aspect, which is connected with conveying the speaker's subjectivity. We also concluded that the discourse marker 'so' played a role in structuring the discourse, which was evident in several instances of sequential structuring applied to open, resume, or close the topic. In contrast, the research also disclosed that the discourse marker 'so' expresses ideational consequence only in a few cases.

The analysis of translations of the discourse markers 'because' and 'so' revealed that the discourse marker 'because' was most frequently translated as *nes, dėl to, kad,* and *kadangi*, each of them being translation variants presented in a number of English–Lithuanian dictionaries. These translation variants found in dictionaries made up 66% of the sample. Likewise, the discourse marker 'so' was most frequently translated into Lithuanian as *taigi*, which is the usual translation variant presented in a number of bilingual English-Lithuanian dictionaries.

Next, the research results demonstrated that in order to impart the pragmatic aspect of the spoken discourse in translations of the discourse markers 'because' and 'so' into Lithuanian, translators would in some cases opt for a particle or a parenthetical verb which leave out the causative meaning in Lithuanian.

Lastly, we found a considerable number of omissions in translations of the analyzed discourse markers, which implies that translators applied this technique to comply with the guidelines for making subtitles shorter.

The contrastive analysis of discourse markers provides more detailed data on the functions of discourse markers both to language teachers and translators. Awareness of the polyfunctionality of discourse markers gives deeper insight into their use and translation. Case studies of this type may serve as an effective method to develop students' linguistic awareness as well as analytic abilities and are especially relevant for students majoring in English philology and/or translation.

References

Aijmer, K. (Ed.) (2009). Corpora and language teaching. *Studies in Corpus Linguistics*, 33. Amsterdam: John Benjamins publishing company.

Buysse, L. (2010). Discourse marker 'so' in native and non-native spoken English. *PhD Thesis*. Leuven: university of Leuven.

Cartoni, B., Zufferey, S., & Meyer, T. (2013). Annotating the meaning of discourse connectives by looking at their translation: the translation spotting technique. *Dialogue and Discourse, 4*(2), 65–86. Retrieved from: http://dad.uni-bielefeld.de/index.php/dad/article/view/2823/3542

Crible, L. (2014). *Identifying and describing discourse markers in spoken Corpora*. Louvain-la-Neuve: université Catholique de Louvain.

Crible, L., & Degand, L. (2017). Reliability vs. granularity in discourse annotation: what is the trade-off? *Corpus Linguistics and Linguistic Theory, 14*(2), 1–29. Retrieved from: https://doi.org/10.1515/cllt-2016-0046

Crible, L., & Degand, L. (2019). Domains and functions: A two-dimensional account for discourse markers, *Discours,* 24, 3–35. Retrieved from: https://journals.openedition.org/discours/9997

Crible, L., Abuczki, Á., Burkšaitienė, N., Furkó, P., Nedoluzhko, A., Rackevičienė, S., Valūnaitė Oleškevičienė, G., & Zikánová, Š. (2019). Functions and translations of discourse markers in TED talks: A parallel corpus study of underspecification in five languages. *Journal of Pragmatics, 142*, 139–155. Retrieved from: https://doi.org/10.1016/j.pragma.2019.01.012

Danlos, L., & Roze, C. (2011). Traduction (Automatique) des connecteurs de discours. *Proceedings of TALN 2011*, France: Montpellier. Retrieved from: http://utilisateurs.linguist.univ-paris-diderot.fr/~danlos/Dossier%20publis/TradConn-TALN'11.pdf

McEnery, T., & Xiao, R. (2011). What Corpora can offer in language teaching and learning? *Hinkel, Eli (Ed.). Handbook of research in second language teaching and learning, 2*, 364–380. New York: Routledge. Retrieved from: https://doi.org/10.4324/9780203836507

Mulder, G. (2008). Understanding causal coherence relations. *Doctoral Thesis*. Utrecht: university of Utrecht.

Polat, B. (2011). Investigating acquisition of discourse markers through a developmental learner corpus. *Journal of Pragmatics, 43*(15), 3745–3756. Retrieved from: https://doi.org/10.1016/j.pragma.2011.09.009

Römer, U. (2008). Corpora and language teaching. *Lüdeling, A., & Merja K. (Eds.). Corpus Linguistics. An International Handbook*, [HSK series], 1, 112–130. Berlin: Mouton de Gruyter.

Sadeghi, B., & Yarandi, M. (2014). Analytical study on the relationship between discourse markers and speaking fluency of Iranian EFL students. *International Journal of Linguistics and Communication, 2* (2), 101–123.

Sinclair, J. (Ed.) (2004). How to use Corpora in language teaching. *Studies in Corpus Linguistics*, 12. Amsterdam: John Benjamins publishing company.

Sun, W. (2013). The importance of discourse markers in English learning and teaching. *Theory and Practice in Language Studies, 3*(11), 2136–2140. doi:10.4304/tpls.3.11.2136-2140

TED Translators Wiki. (2017). *How to compress subtitles?* Retrieved from: http://translations.ted.org/wiki/How_to_Compress_Subtitles.

Valūnaitė-Oleškevičienė, G., Puksas, A., & Gulbinskienė, D. (2018). TED-MDB multilingual discourse-annotated corpus for language learning and teaching. *Pedagogika*, *131* (3), 99–109.

Zufferey, S., & Degand, L. (2013). Annotating the meaning of discourse connectives in multilingual Corpora. *Corpus Linguistics and Linguistic Theory, 13*(2), 399–422. doi: 10.1515/cllt-2013-0022.

Priežastinių diskurso žymeklių *because* ir *so* anotavimo metodo naudojimas anglų, kaip užsienio kalbos, mokymuisi

Giedrė Valūnaitė-Oleškevičienė[1], Sigita Rackevičienė[2], Nijolė Burkšaitienė[3], Liudmila Mockienė[4], Jolita Šliogerienė[5], Dalia Gulbinskienė[6]

[1] Mykolo Romerio universitetas, Ateities g. 20, 08303 Vilnius, gvalunaite@mruni.eu
[2] Mykolo Romerio universitetas, Ateities g. 20, 08303 Vilnius, n.burksaitiene@mruni.eu
[3] Mykolo Romerio universitetas, Ateities g. 20, 08303 Vilnius, sigita.rackeviciene@mruni.eu
[4] Mykolo Romerio universitetas, Ateities g. 20, 08303 Vilnius, liudmila@mruni.eu
[5] Vilniaus Gedimino technikos universitetas, Saulėtekio al. 11, 10223 Vilnius, jolita.sliogeriene@vgtu.lt
[6] Generolo Jono Žemaičio Lietuvos karo akademija, Šilo g. 5A, 10322 Vilnius, dalia.gulbinskiene@lka.lt

Santrauka

Viena reikšmingiausių užsienio kalbos mokymosi pažengusiesiems sričių yra diskurso analizė ir diskurso valdymas. Anotavimo metodas suteikia besimokančiajam išskirtinę galimybę įgyti kalbinių žinių apie diskurso žymeklius, surasti dėsningumus užsienio kalboje ir palygti juos su gimtąja kalba. Straipsnyje pateikiamas angliškų priežastinių diskurso žymeklių *because* ir *so* bei lietuviškų jų atitikmenų sričių ir funkcijų tyrimas; taip pat analizuojami šių diskurso žymeklių vertimo į lietuvių kalbą variantai, naudojant daugiakalbį atvirojo vertimo projektą TED Talks kaip duomenų šaltinį. Tyrimas buvo atliekamas dviem etapais. Pirmiausia buvo anotuotos ir palygintos anglų ir lietuvių kalbų žymeklių funkcijos. Po to buvo analizuojami žymeklių vertimai į lietuvių kalbą. Išvados parodė, kad diskurso žymeklis *because* ir jo lietuviški atitikmenys dažniausiai išreiškia objektyviąją priežastį, po jos eina retorinis atitikmuo, per kurį kalbėtojai išreiškia savo subjektyvumą. Tačiau diskurso žymeklis *so* ir jo lietuviški atitikmenys daugeliu atvejų išreiškia retorinę pasekmę, po kurios eina retorinis detalizavimas. Taip pat buvo nustatyta, kad, verčiant abu diskurso žymeklius, buvo dažnai naudojamas praleidimas. Anglų kalbos diskurso žymeklių sričių ir funkcijų anotacija bei jų vertimai į lietuvių kalbą yra naudingi ne tik vertėjams, bet ir užsienio kalbų mokytojams, nes leidžia geriau susipažinti su teksto rišlumu bei pragmatiniu diskurso aspektu.

Esminiai žodžiai: *priežastiniai ryšiai; diskurso žymekliai; lygiagretusis tekstynas; daugiakalbio diskurso anotacija; vertimas; užsienio kalbų mokymasis.*

Gauta 2020 02 10 / Received 10 02 2020
Priimta 2020 04 21 / Accepted 21 04 2020