

# Investigating the Gap between L2 Grammar Textbooks and Authentic Speech: Corpus-Based Comparisons of Reported Speech

**Kevin Cancellaro**  
New Economic School

Correspondence concerning this article should be addressed to Kevin Cancellaro, Department of Humanities and Languages, New Economic School, 100 Novaya Street, Skolkovo, Moscow, Russian Federation, 143025. E-mail: kcancellaro@nes.ru

Corpus Linguistics (CL) has made significant inroads into the field of second language acquisition (SLA) and pedagogy. As more corpora have become available, researchers and teachers alike have begun to realize the importance of empirically testing ideas that have long been taken for granted and accepted as fact. This is especially true for grammar textbooks written for second language (L2) learners. Do the textbooks that are being used reflect real world grammatical usage? The current study is the first of two in which three corpora were used to examine real world usage of reported speech (RS) as compared to typical presentations of RS in popular L2 grammar U.S. textbooks as they existed in and up to the year 2007. Results show that indirect reported speech (IRS), direct reported speech (DRS) and alternative forms of RS constructions in combination are not only frequent in spoken English but also dependent on register and context. Further, simplifying RS explanations in terms of backshifting with the use of a past tense main reporting verb may be providing inaccurate information to L2 learners of American English. Results generally support, with some exceptions, the findings in previous studies which employed corpus-based analysis to study the relevance of EFL/ESL textbooks (Al-Wossabi, 2014; Barbieri & Eckhardt, 2007; Khojasteh & Shahrpour, 2014; Šegedin, 2008). A forthcoming study will examine new corpora and revised textbooks to measure the degree of change that has occurred since 2007, thereby seeking to replicate the results of a more general review on the same topic done by Khojasteh and Shahrpour (2014).

*Keywords:* Corpus Linguistics, second language acquisition, indirect reported speech, direct reported speech, alternative forms of reported speech, backshifting, corpus-based analysis

There has been a lot of interest in Corpus Linguistics (CL) and its applications for second language acquisition (SLA) in recent years, and many studies have recommended using corpus-based findings to provide accurate content for prescriptive English grammar textbooks (Biber & Reppen, 2002; Conrad, 1999; 2000; Carter & McCarthy, 1995; Frazier, 2003; Harwood, 2005; Kennedy, 2002; Lawson, 2001; Romer, 2010). Once thought of as only the domain of those interested in purely linguistic investigation, it is now seen as a useful tool for language teachers and SLA researchers for examining exactly how the language works in the real world. For language teachers in particular, this strand of research has

important implications. CL can be used in many ways when developing language materials, from using collocations to assist in vocabulary learning to expanding the list of grammatical constructions that are taught through analysis of spoken and written texts. Many teachers of English as a second or foreign language (ESL/EFL) have long noticed the gap between the material that they see in many textbooks and what is actually used by the English speaking populace. CL can be incorporated into English language teaching material development as well as in the classroom, and there has been increasing acknowledgment that this is not only desirable but necessary. Using CL to gather empirical evidence

about construction frequency, register, and discourse context, ESL/EFL materials developers and teachers can then incorporate this important information into the classroom or textbook. Many programs across the country basically leave it up to the learner to decipher colloquial constructions and the contexts in which they are found because these constructions are almost never presented in language teaching materials.

## Materials and Methods

One area where this is evident is with textbooks meant to teach grammar to second language (L2) learners. Several researchers have compared grammar descriptions that are presented in textbooks with real world language use (Al-Wossabi, 2014; Barbieri and Eckhardt, 2007; Biber and Reppen, 2002; Carter, 1998; Eckhardt, 2001; Frazier, 2003; Gilmore, 2004; Khojasteh & Shakrpour, 2014; Šegedin, 2008) and, all have noted that there is a large gap between the explanations and descriptions in textbooks and real world language use. They attribute this to several factors: 1) textbook material is not taken from empirical data about language use but rely on writer's intuition; 2) textbooks present grammatical constructions as equally generalizable and of equal communicative importance; 3) information concerning pragmatics, discourse context and register is ignored; 4) textbooks are based on written norms; 5) textbooks simplify the grammar for pedagogical purposes (Barbieri and Eckhardt, 2007; Biber and Reppen, 2002; Carter and McCarthy, 1995; Lawson, 2001). While simplifying grammar for pedagogical purposes is a worthy goal, the balance of simplification must be weighed against the difficulties students will have when confronted with real-world language use.

For one grammatical construction in particular, Reported Speech (RS), the weaknesses in instruction materials listed above can have a large effect on the ability of an L2 learner to acquire this construction in an accurate and natural way. The RS construction as it is described in most grammar textbooks is very complex and one that all learners find difficult to master. The first step in this process requires a student to understand how to use "question word", "yes/no", "that" and "command" noun clauses for which various changes occur depending on which is used. RS constructions are special types of noun clauses. Some types of RS, the types that are emphasized in many textbooks, require further changes such as backshifting and pronoun substitution. This adds complexity to an already complex construction. It is not the case that the RS construction is a rare one that can be ignored in any program of grammar study; on the contrary, it is very common in both written and spoken language. If

language teachers and students are required to spend the extensive time and effort needed to master RS, the grammar rules presented in textbooks should be based on empirical evidence about how RS is actually used in real world language.

An examination of five grammar textbooks (Bland, 1996; Eastwood, 1999; Elbaum, 2001; Fuchs and Bonner, 1995; Thewlis, 2001) and one textbook written for English language teachers (Parrot, 2000), which were in common use in 2007, show that there are some weaknesses in the presentation of RS and little or no consensus about a standard for teaching this construction. All the books focus on Indirect Reported Speech (IRS) and offer little instruction about the usage and possible forms of Direct Reported Speech (DRS). IRS is reported as being the spoken version of DRS. Verbs used to introduce RS are presented in the past tense, implying very strongly that the past tense is the preferred tense for this construction. Changing sentences from DRS to IRS by backshifting and pronoun change in for all forms of statements, and questions with IRS is the main focus, with command forms using the infinitive construction. Grammar Dimensions, one of the most popular textbooks for ESL learners describes reported speech in this way after briefly defining DRS and IRS:

Because we are describing something that has already occurred (speaking or thinking), we need to change the time frame of the verb phrases that we are reporting. (Thewlis, 2007, p. 402)

Parrot (2000), in his textbook for English teachers, supplies examples that are all in the past tense, without mentioning the various possible permutations that can occur for both DRS and IRS in different registers and contexts. Three textbooks mention that backshifting is not necessary when the main reporting verb is in the present, but offer no further information about when the reporting verb should be used in the present. None of the textbooks offer information about differing uses in pragmatic, discourse or register contexts. Alternative DRS constructions such as "to be like", "to be all" and "go" are not covered. The teachers instruction textbook does mention that DRS and IRS are both used in spoken English but offers the following assessment of alternative forms:

Learners may also come across common, very informal equivalents to said (*which we would very rarely need to teach*). (Parrot, 2001, p. 225) (Italics are mine).

It is odd, to say the least, that while admitting that alternative forms are common, Parrot implies that they should be taught only rarely. Several studies have employed CL to investigate the apparent gap between RS in real and textbook contexts (Al-Wossabi, 2014; Barbieri, 2005a; Barbieri, 2005b; Barbieri & Eckhardt, 2007; Blyth, Recktenwald and Wang, 1999; Eckardt,

2001; Gilmore, 2004; Khojasteh & Shakrpour, 2014; Šegedin, 2008). In the seminal study on this topic, Barbieri and Eckhardt (2007) examined both IRS and DRS in different registers and modes and came to some general conclusions which are worth briefly restating here.

1. It is not necessary or desirable to teach RS as a transformation mechanism from DRS to IRS.
2. IRS should be taught in conjunction with samples of newspaper writing, since it is much more frequent in newspaper writing than in casual conversation.
3. DRS should be taught in the context of casual conversation.
4. For IRS, teach past-past tense sequences but introduce other widely used sequences.
5. For less widely used sequences like past-present or present-present, point out discourse functions.
6. For DRS, teach the *say*, but also introduce alternatives like “to be like” and explain how they are different and the contexts in which they are used.

This study also raised the question of register and the alternative DRS “to be like” as they relate to the age of the speakers. This issue was not investigated, but it was generally stated that “to be like” appears to be expanding its influence in American English but is still restricted to speakers under the age of 40. In addition to this, the study did not examine the precise role that narrative discourse plays when using DRS in any form. Al-Wossabi (2014) comes to a similar conclusion in a study that targeted only EFL learners and focused on comparing only two sources: Oxford Pocket English Grammar (OPEG) and Longman Grammar of Spoken and Written English (LGSWE), a grammar book based on corpus findings but does not itself directly examine corpora. Šegedin (2008) finds a similar pattern with EFL textbooks being used in Croatia in elementary and secondary schools. In order to ascertain whether the suggestions and results above merit further study and to shed light on register and DRS narrative discourse variations as they appear in real speech,

the current study examined three spoken corpora taken from transcripts of the TV shows *Friends* (909,000 words), *Frazier* (990,000 words) and from interviews (290,000 words) with jazz musicians from the documentary, *Jazz*. The age of the speakers varies from 20’s-30’s, late 30’s- 50’s and 60’s to 80’s respectively. Corpora were examined using Textstat.

### Research Questions

1. How are IRS and DRS used in spoken English?
2. How common are alternative RS constructions? How are they used? (to be like, to be all, to go)
3. Are there register/context differences in usage of different RS constructions?

### Results

In Table 1, we see the totals from the occurrences of the verbs *say* and *tell* in present tense and past tense for the three corpora (See Figure 1). The past tense main reporting verb *said* showed the highest frequency for all three corpora for IRS, although *say* was also used fairly frequently when compared to *said*. The jazz corpora showed nearly equal usage of both *said* and *say* as the main reporting verb, with *Frazier* and *Friends* showing about half and a third difference in usage between the two. The past tense reporting verb *told* was less frequently used than *said* in IRS in all three corpora. The present tense form shows a surprisingly high frequency, especially for the *Friend’s* corpora. An examination of the data showed that this was due to the extremely high frequency of the construction ‘Don’t tell me’ + embedded ‘that’ NP.

In most cases, the frequency of both verbs in both tenses was lower for DRS than for IRS especially for *said* (See Figure 2) and dramatically in the case of DRS *told* and *tell*. Interestingly, DRS *say* was as frequent as IRS *say* in the *Jazz* and *Friends* corpora and half as frequent as the *Frazier* corpus. Perhaps most interesting was the extremely high frequency of DRS *said* in the *Jazz* corpus. This corpus was a collection of interviews and, unlike the other two corpora, most

Table 1  
*IRS and DRS Totals for Say, Said, Tell and Told*

	IRS	IRS	IRS	IRS	DRS	DRS	DRS	DRS
	TotSAID	TotSAY	TotTOLD	TotTELL	TotSAID	TotSAY	TotTOLD	TotTELL
<i>Frazier</i>	241	152	189	192	83	65	10	8
<i>Friends</i>	253	90	153	237	101	89	8	8
<i>Jazz</i>	85	72	30	42	514	95	14	10

Table 2  
*Alternative DRS Totals*

	DRS be like	DRS be all	DRS go
Frazier	6	0	0
Friends	95	22	5
Jazz	22	2	0

of the speakers talked for long periods with very little turn taking. DRS *said* seems to be a feature of narrative discourse and not as commonly used in rapid turn taking style dialogues like those in the other corpora.

Alternative DRS constructions (See Figure 3) were used with the most frequency in the Friends corpus with *to be like* being the most frequent. While this confirms the idea that Alternative DRS forms are a feature of younger speakers, it also shows that this construction may be expanding in American English. The *to be like* construction was also used in the jazz corpus, whose participants were in the 60's to 80's age range. The DRS *go* was infrequent in all cases.

Another interesting finding was that of tense variation combinations between the main reporting IRS verb and the embedded verb (See Figure 4). For IRS, one rule that is constantly emphasized is that if the main reporting verb is in the past, the embedded clause's verb must be backshifted except for special cases. Although not mentioned often, as explained

above in the review of grammar textbooks, if the main reporting verb is in the present tense, no backshifting should take place. For IRS *Said*, this seems to be supported somewhat - but not completely - by the data. But for IRS *say*, it appears that this rule is not as reliable as once thought. In the Jazz corpus, instances of present-past combinations, which should really occur in IRS, outnumber the "correct" form of present-present. In the Frazier corpus, the same violation occurred at a fairly high 46 instances compared to 106 instances of the "correct" form. While the "still true" rule accounts for much of the data for IRS *said*, no such rule accounts for the present-past combination for IRS *say*, which seems to be showing properties of DRS where tense is free because the speech is quoted exactly.

### Discussion and Conclusion

Several of the findings in this study contradict the suggestions of Barbieri and Eckhardt (2007), specifically the first three. The first suggestion, that it is not necessary or desirable to teach RS as a transformation mechanism from DRS to IRS, is not supported because of the frequency of both in spoken American English. The basic rule of DRS present tense to IRS past tense backshifting should be modified but not eliminated altogether, as this skill seems to be necessary in real world speech. The most common tense combinations

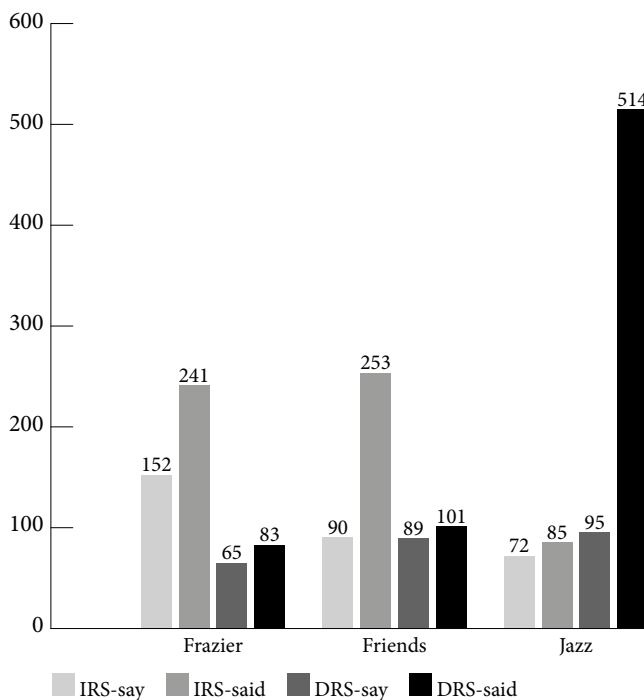


Figure 1. IRS, DRS, Say and Said.

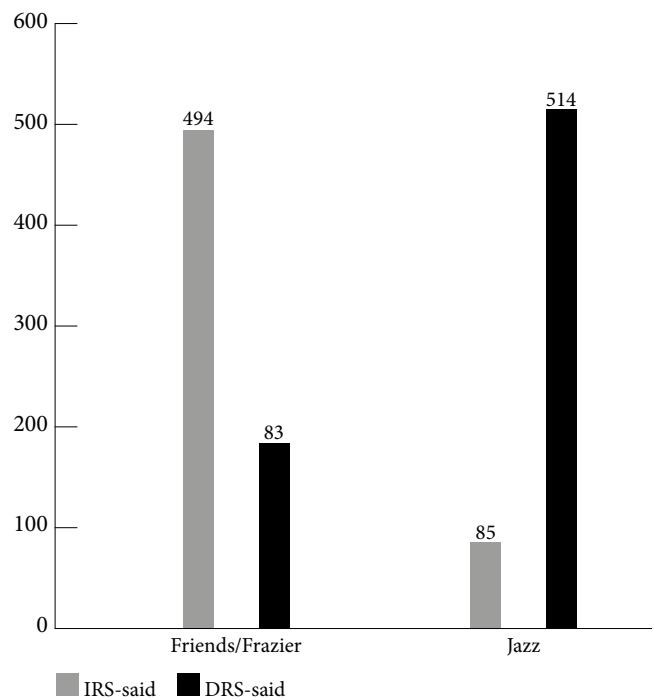


Figure 2. IRS and DRS Totals for Said.

should be taught and then followed by alternative tense combinations. The data in this study shows that the normal DRS to IRS backshifting when the IRS said is used is by far the most common of the tense combinations shown (See Figure 4).

Suggestion two, that IRS should be taught in conjunction with newspapers as it is more frequent in writing than in casual conversation, is not supported by the data in this study. While no direct comparison with written English was done, IRS in both present and past are clearly being used frequently in spoken English. The suggestion that DRS should be taught in the context of casual conversation should therefore also be adjusted. DRS is not the only construction speakers use to report speech in casual conversation according to the data collected in this study. It seems clear that speakers across registers and contexts use a complex variety of DRS, IRS and alternative DRS in various combinations. The reason why this is so is probably due to sociolinguistic or pragmatic factors that should be investigated further but are beyond the scope of this paper. However, the often near total absence of pragmatic information in prescriptive textbooks has been noted previously (Vallenga, 2004). Presenting DRS as the form that should be used in casual conversation would be just as misguided as present IRS past as the main form of reported speech. What seems apparent from the findings, here, is that DRS past is very common for narrative type discourse and not as common in regular conversation. This

narrative discourse function also seems to apply to a lesser degree to the *to be like* construction, as evidenced by its use by the speakers in the Jazz corpora. Even though the speakers in this corpora are at an age where this construction should not be used, there were instances of its use, perhaps showing its influence in narrative discourse across all registers.

Overall, the data collected here show a much more complicated picture of real world RS usage than is presented in textbooks for L2 learners and in previous studies. While it is true that grammar must be simplified to some extent to match the lower proficiency levels of L2 learners and to aid in comprehension and acquisition of new constructions, presenting information that is wrong (i.e., IRS is used for spoken English with a past tense reporting verb, DRS is for written English, backshifting applies to each case in a uniform way, alternative constructions should not be taught), is clearly bad and confusing for the students, who must take this partial knowledge and then adapt to the sometimes vastly different input that they encounter in the real world of language use.

With larger and larger spoken corpora that reflect many registers and discourse contexts coming online, empirical testing of spoken grammar norms as presented in grammar textbooks can soon reliably be undertaken for all grammatical constructions. It is highly likely that some new constructions can also be found that merit inclusion in the L2 grammar

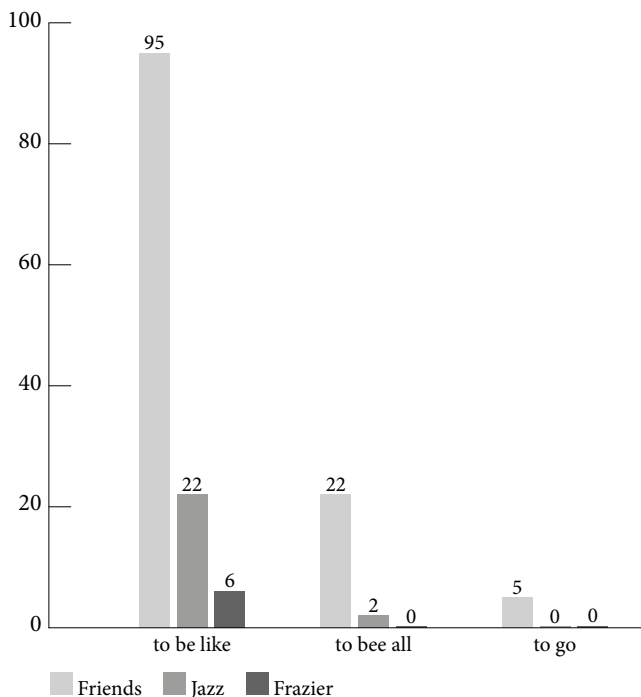


Figure 3. Alternative RS Construction Frequency.

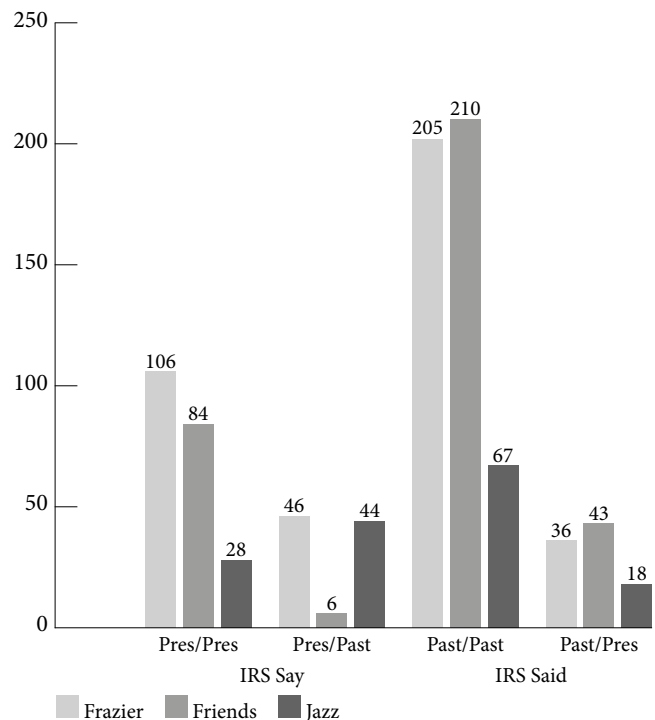


Figure 4. Tense Variations/Combinations for IRS Main and Embedded Verbs.

textbook. CL is clearly a tool that can directly impact the quality of SLA instruction and research now and in the future. In a forthcoming study, the number of corpora will be greatly expanded and current popular ESL/EFL grammar textbooks will be examined with the hope that publishers and developers will begin to incorporate more accurate and useful grammatical explanations in their textbooks.

## References

- Al-Wossabi, S. (2014). Spoken grammar: An urgent necessity in the EFL context. *English Language Teaching*, 7(6), 19-26. doi:10.5539/elt.v7n6p19
- Barbieri, F. (2005a). Quotative use in American English. A corpus-based, cross-register comparison. *Journal of English Linguistics*, 33(3), 222-56.
- Barbieri, F. (2005b). Quotative *be like* in American English: Ephemeral or here to stay? *Paper presented at SHEL 4*, Northern Arizona University, Flagstaff, AZ, 30 September-1 October 2005.
- Barbieri, F., & Eckhardt, S. E. B. (2007). Applying corpus-based findings to form-focused instruction; the case of reported speech. *Language Teaching Research*, 11(3), 319-346. doi: 10.1177/1362168807077563
- Biber, D., & Reppen, R. (2002). What does frequency have to do with grammar teaching? *Studies in Second Language Acquisition*, 24, 199-208.
- Blyth, C. J., Recktenwald, S., & Wang, J. (1990). I'm like, 'Say what?!': A new quotative in American oral narrative. *American Speech*, 65(3), 215-27.
- Carter, R., & McCarthy, M. (1995). Grammar and spoken language. *Applied Linguistics*, 16(2), 141-58.
- Conrad, S. M. (1999). The Importance of corpus-based research for language teachers. *System*, 27, 1-18.
- Conrad, S. M. (2000). Will corpus linguistics revolutionize grammar teaching in the 21st century? *TESOL Quarterly*, 34(3), 548-60.
- Eckhardt, S. E. B. (2001). *Reported Speech: Empirical corpus findings compared with ESL/EFL textbook presentations* (Unpublished Master's thesis). Ames, IA: Iowa State University.
- Frazier, S. (2003). A corpus analysis of would - clauses without adjacent *If*-clauses. *TESOL Quarterly*, 37(3), 443-46.
- Gilmore, A. (2004). A comparison of textbook and authentic interactions. *ELT Journal*, 58(4), 363-71.
- Harwood, N. (2005). What Do We want EAP teaching materials for? *Journal of English for Academic Purposes*, 4, 149-161.
- Kennedy, G. (2002). Variation in the distribution of modal verbs in the British national corpus. In R. Reppen, S. Fitzmaurica & D. Biber (Eds.), *Using Corpora to Explore Linguistic Variation* (pp. 73-90). Amsterdam, Netherlands: John Benjamins.
- Khojasteh, L., & Shokrpour, N. (2014). Corpus linguistics and English language teaching materials: A review of recent research. *Khazar Journal of Humanities and Social Sciences*, 17(3), 5-17.
- Lawson, A. (2001). Rethinking French grammar for pedagogy: The contribution of French corpora. In R. C. Simpson & J. M. Swales (Eds.), *Corpus Linguistics in North America: Selections from the 1999 symposium* (pp. 179-194). Ann Arbor, MI: The University of Michigan Press.
- McCarthy, M. (1998). 'So Mary was saying': Speech reporting in everyday conversation. In M. McCarthy (Ed.), *Spoken Language and Applied Linguistics* (pp. 150-175). Cambridge, UK: Cambridge University Press.
- Parrot, M. (2000). *Grammar for English Language Teachers*. Cambridge, UK: Cambridge University Press.
- Romer, U. (2010). Using general and specialized corpora in English Language teaching: Past, present and future. In M. C. Campoy, B. Belles-Fortuno & M. L. Gea-Valor (Eds.), *Corpus-based Approaches to English Language Teaching* (pp. 18-38). London, UK: Continuum.
- Šegedin, D. (2009). Reported speech: Corpus-based findings vs. EFL textbook presentations. *ELOPE*, 5(1/2), 199-213.
- Thewlis, S. (2007). *Grammar dimensions 3. Form, meaning and use*. Boston, MA: Heinle-Cengage ELT.
- Vellenga, H. (2004). Learning pragmatics from ESL & EFL textbooks: How likely? *The Electronic Journal for English as a Second Language*, 8(2), 2-16.