



## **Corpus-Supported Foreign Language Teaching of Less Commonly Taught Languages**

### **Nives Mikelić Preradović**

Prof., Faculty of Humanities and Social Sciences, University of Zagreb, Croatia,  
nmikelic@ffzg.hr

### **Kristina Posavec**

PhD, Faculty of Education and Rehabilitation Sciences, University of Zagreb, Croatia,  
kposavec@erf.hr

### **Danijela Unić**

PhD, VERN' University of Applied Sciences, Croatia, danijela.unic@vern.hr

The study explores the implementation of corpora in teaching of the four courses of Croatian as a foreign language. Croatian, as a less-resourced and less commonly taught language, shares the same issues as other less resourced languages: studies about using corpus - are still missing. Student satisfaction and class engagement as well as teachers' course delivery methods (indirect and data-driven approach) were analysed in four groups of participants: two groups of beginner level of the foreign language learning (A1) and two groups of intermediate level (B1, B2) The experiment was conceived in three phases. In phase one, four teachers of Croatian as a Foreign Language (CFL) held an experimental lesson in their course groups. In phase two, the students were asked to complete a survey to examine degree of satisfaction with corpus-based approach in teaching CFL. In phase three, in-depth interviews were conducted with all teachers to examine their perceptions about the role of corpora in teaching, as well as their expectations regarding corpus-assisted language teaching. Based on the analysis, it was concluded that degree of students' satisfaction with experimental classes prepared using corpus was mostly positive. Experiment revealed that approach to language teaching depends largely on perception of teachers about integration and use of corpora in teaching. Corpora can be used directly in class of absolute beginners.

**Keywords:** foreign language teaching, less commonly taught languages, data-driven learning (DDL), direct use of corpora in teaching, Croatian as a foreign language

**Citation:** Preradović, N. M., Posavec, K., & Unić, D. (2019). Corpus-Supported Foreign Language Teaching of Less Commonly Taught Languages. *International Journal of Instruction*, 12(4), .

## **INTRODUCTION**

Due to the rapid advancement in both corpus linguistics and technology in recent years, many studies emphasize the remarkable potential of computer corpora for language teaching and learning. Computer and especially web corpora, provide accessibility to a large amount of authentic language data, that are used in teaching vocabulary, grammar and writing and have brought many positive effects, such as the growth in the students' understanding of usage patterns, improvement of the self-correction skills and decrease of collocational errors (Aston, 2001; Hunston, 2002; O'Keeffe & Farr, 2003; Sinclair, 2004). The use of corpus examples for language comprehension and production (Frankenberg-Garcia, 2014) as well as for grammar and vocabulary teaching (Flowerdew, 1996; Lenko-Szymanska, 2002; Liu & Jiang, 2009) has been reported by different authors over the years. Majority of the literature related to corpus-supported language learning and language teaching addresses the implementation of methodologies in teaching English as a foreign language around the world (Gabrielatos, 2005; O'Keeffe, McCarthy & Carter, 2007; Pérez-Paredes et al., 2018; Reppen, 2010; Smirnova, 2017). However, with development of corpora and tools for less commonly taught languages (LCTLs), the need to address main advantages and issues in corpus-supported language teaching of these languages became apparent.

The term 'Less Commonly Taught Languages' has several interpretations, including Minority Languages, Disadvantaged Languages, Less Commonly Spoken Languages, Endangered Languages and Less Widely Taught Languages (which are usually defined as non-Roman and non-Germanic languages). In this paper, the term LCTL covers a wide array of world languages other than the most commonly taught foreign languages which (apart from being less taught) also suffer a shortage of corpus-informed language learning resources (such as grammars, textbooks, and dictionaries) and/or language teaching applications (such as data-driven language learning).

Due to the unavailability of annotated language corpora and basic natural language processing tools, LCTLs have all, until recently, belonged to less-resourced languages. During the last decade, large corpora for most of these languages have been built (see Tymoshuk et. al., 2018 and Ljubešić & Klubička, 2014 for Slavic languages), along with projects for less-resourced languages, such as DigiSami project (Jokinen, 2018), Digital Language Diversity Project (Soria et al., 2016) and LRE Map (Calzolari et al. 2012) that made many language resources for LCTLs freely available.

Unfortunately, the use of corpora and other resources for computer-assisted language teaching and learning is found less frequently in the world of LCTLs (Ward, 2016, p. 469). Although corpus research of the LCTLs gained popularity in the last decade, studies about using corpus in foreign language teaching are still missing. As pointed out by Ward (2016), high quality resources do exist, but there is a lack of awareness amongst learners and teachers about their existence. A gap exists in research about the use of corpus-based materials in language teaching, as well as in effects of their use in context of LCTLs.

This study aims to fill these gaps by exploring possibilities and effects of using computer corpora as both indirect and direct approach to teaching Croatian as a foreign language (CFL). It explores implementation of corpora in teaching four CFL courses delivered over one academic semester. Students' satisfaction and class engagement as well as teachers' course delivery methods were analysed to observe the possibility of corpus-supported teaching of LCTLs. By examining teachers' use of computer corpora and students' reactions to this type of teaching, this paper offers an example of how foreign language teaching of LCTLs can be enriched with the use of corpora.

## **LITERATURE REVIEW**

### **Theoretical Framework**

#### **Using computer corpora in foreign language teaching**

According to Reppen (2010), a computer corpus can greatly contribute to language learning because it provides a source of natural or authentic texts for language learning. Furthermore, Miangah (2012) states that a computer corpus can be used in teaching morphology or exploring the structure of words through their frequency in various morphological forms, as well as in analysis of syntax and in cultural studies, where a corpus can serve to research specifics that differ from one culture to the other, such as human habits. According to Römer (2006), corpora provide pedagogically significant discoveries of contrastive language data and can be used to gain useful information for creating bilingual and learner dictionaries. Römer (2008) argues that corpora could also aid in decision making about when and what will be taught in language teaching process, but teachers and students can also access corpora directly during class and complement the teaching process. There are many benefits of corpus versus textbooks in teaching foreign languages. It is important to emphasize that use of corpora does not imply the absence of textbooks in teaching, but rather the integration of computer corpora into language teaching during preparation of teaching materials or integration of corpora in classes, where teachers act as mediators between students and corpora. Corpus analysis is an ideal tool for making a principled decision about what material to present to students from the foreign language textbook (Barbieri & Eckhardt, 2007). Biber & Conrad (2010) list three types of corpus results that are important in learning and mastering grammar of a language: frequency information, register comparisons, and association between grammatical structures and words. Similarly, Lawson (2001) states that corpora provide information on frequency of occurrences in language used in everyday communication, so that one can inspect the most commonly used words or phrases that actually appear, while textbooks often contain artificially created, fictitious language examples. Furthermore, corpora provide information on how the use of special linguistic features differs in a particular context or situation. Also, corpora provide information on linguistic features of the language in use, and, finally, information on the discursive properties of certain linguistic features such as collocations, lexical grammatical associations, etc.

### **Direct and indirect corpus use in language teaching**

Leech (1997) points out that language learning and corpora have 3 points of contact: indirect use of corpora in teaching (reference sources, teaching materials and language tests), use of corpora in teaching and development of corpora for teaching purposes (learner corpora).

Indirect use of corpora in language teaching refers to design of syllabi and development of teaching materials and is closely related to the content being taught. Here, a corpus can be used by teachers as an excellent source of authentic examples of a language. The indirect approach is focused on analysis and research of corpora, on the corpus evidence and the effect it has and may have on syllabus design.

Data-driven learning (DDL) is a type of learning where students study the language themselves with the help of a corpus. Direct use of corpora in language teaching provides practical knowledge to students (McEnery, Xiao & Tono, 2006), as students search through a corpus in order to satisfy their own learning needs. Main goal of this type of language learning is to develop student autonomy, mastering of the language skills and acquiring of new technology skills. It is a way of learning in which students become researchers and self-explore the language (Boulton, 2012). As Johnes (1991) points out, the use of a computer is in the centre of the DDL approach. The computer corpus is not a substitution for a teacher, but rather a resource. Teachers role does not imply their exclusion from the teaching process, but rather a new role of a mediator in language teaching.

The key to a successful corpus-based approach to language learning is an appropriate level of teacher orientation or pedagogical mediation that depends on the age, experience and previous knowledge of students. McEnery & Xiao (2010) emphasize that direct use of corpora is limited by several factors related to student language proficiency level and experience, time constraints, curriculum requirements, knowledge and skills of teachers essential for the corpus analysis and pedagogical mediation between the corpus and the student, as well as access to resources for creation of materials, such as computers and corresponding software or a combination of these factors.

The advantage of direct use of corpora, according to McEnery & Xiao (2010), is in the possibility to provide word frequency information (quantitative analysis) and to display accurate descriptions of use of lexical terms in an annotated corpus. A very important and significant feature of using corpora in teaching foreign languages is that they provide more realistic examples of language use that reflect the complexity and nuances of the native language. Therefore, the use of a large corpus can provide an important insight into learner's foreign language experience (Tucker, 2009). In addition to quantitative, computer corpora also allow for qualitative analysis (McCarten, 2007), i.e. the possibility to display a large number of texts that provide students and teachers with a context in which a particular word or a phrase appears and how it is used in different contexts. Apart from mastering the vocabulary, corpora can also help students in learning collocations - a type of word combinations that is fixed to some degree or

words that co-occur at a certain distance (Nesselhauf, 2005). Collocations enable students to speak fluently, make their speech more understandable and help them write and sound more like native speakers (Fan, 2009). Finally, corpus concordances allow for an overview of the context within which a particular term appears, i.e., they help students understand the association between words (Miangah, 2012).

### **Indirect approach and two data-driven approaches to foreign language learning (FLL)**

There are three approaches for teachers to integrate computer corpora into their foreign language teaching. Indirect, where teachers gather data from corpus searches, prepare teaching materials and students work with the material. The second and third approaches are direct approaches (Götz, 2012). The first direct approach refers to learner-corpus interaction where students independently study the language and try to come to conclusions, while their learning does not necessarily coincide with the curriculum. Here, the role of a teacher is secondary, and consists in guiding students when encountering a problem or ambiguity in the research. One of the features of this approach to language learning is spontaneous or accidental learning, since the acquisition of one language feature may trigger the acquisition of another one. This approach is motivating for learners; it can deepen the student's cognitive process and contribute to a better understanding and retention of information (Boulton, 2012). The second direct approach refers to teacher-corpus interaction where it is not necessary for students to come into direct contact with a corpus, but they learn and interpret the concordances in class that were prepared by teachers who used the corpus. The main feature of this approach is learning focused on a specific part of the curriculum, where the teacher eliminates all other information that could distract students from what is considered important for them to adopt. This approach can be implemented in all language learning levels regardless of the student's previous knowledge and can be integrated into the existing curriculum and learning topics (Kreyer, 2010).

In this paper (section "Context, Participants and Data-Gathering Instruments"), we provide examples of all three approaches (the indirect and two data-driven approaches) in teaching Croatian as a foreign language.

### **Empirical Research on Data-driven Learning**

For the past several years, researchers have been investigating the data-driven learning in the language learning classroom (Boulton & Pérez-Paredes, 2014; Leńko-Szymańska & Boulton, 2015). It was revealed that data-driven learning in the language classroom has a positive influence on vocabulary acquisition (Karras, 2016) and collocations (*Chan & Liou, 2005; Daskalovska, 2015; Frankenberg-Garcia, 2014; Vyatkina, 2016*)

Meta-analysis of results from different studies has confirmed that DDL brings significant benefits to learners and outperforms traditional language teaching methods (Boulton and Cobb, 2017). Moreover, some of these studies report on the positive learners' perception of data-driven learning. However, apart from the study for German as a foreign language (Vyatkina, 2016), the research studies for languages other than English are rare.

## **METHOD**

### **Objectives**

Croatian, as a less-resourced and less commonly taught foreign language, shares the same issues as other LCTLs: although annotated language corpora and natural language processing tools have become available for LCTLs in the last decade, studies about language teaching applications (such as data-driven foreign language teaching) are still missing.

This study explores the implementation of corpora in teaching four courses of Croatian as a foreign language (CFL) at the different language learning levels and attempts to answer the following research questions:

- To what extent are students satisfied with the corpus-based approach in teaching a foreign language?
- Is there a difference in attitudes towards the direct use of corpus in class between the students at the beginner level and the intermediate level of foreign language learning?
- What are the teachers' perceptions about the use of corpora at the specific language learning levels and student engagement?

### **Context, Participants and Data-Gathering Instruments**

The experiment was conceived in three phases.

In phase one, four teachers of Croaticum (Centre for Croatian as Second and Foreign Language at the University of Zagreb), which is considered the largest and central institution for teaching CFL, held an experimental lesson in their course groups (groups were formed from the level A1 to B2 of the Common European Framework of Reference for Languages - CEFR). For the preparation of the experimental lesson, teachers have used a Croatian computer corpus of their own choice and were able to choose whether the corpus will be used to prepare teaching materials and exercises, or they would try to integrate the corpus into direct teaching.

In the second phase, after the experimental lessons, the students in all four groups were asked to complete a survey to examine the degree of their satisfaction with the corpus-based approach in teaching CFL.

In the third phase, the in-depth interviews were conducted with all four teachers, recorded and transcribed for analysis. None of the teachers had prior experience with the direct or indirect use of the computer corpora, although they have completed training on corpus use and were informed about possibilities of corpus-assisted language teaching. The aim of the interviews was to discover teachers' perceptions about the role of corpora in teaching, as well as their expectations regarding corpus-assisted language teaching.

#### **Group 1 (CEFR level B2)**

Eighteen students in upper intermediate (B2) level. They use the Croatian language independently and skilfully and possess a higher level of language knowledge. Majority

of the students had Polish as their mother tongue, with other languages including Russian, German and Portuguese.

For the preparation of the teaching unit (experimental lesson), hrWaC corpus (Ljubešić & Klubička, 2014), the Croatian web corpus and currently the largest computer corpus of Croatian language with 1.9 billion tokens was used. The corpus for preparation of the lesson on the prefixed verbs, more precisely, verbs with the prefix *ob-/op-* and *od-/ot-*. Set of exceptions was elaborated and the most common Croatian verbs with these 2 prefixes were found in the corpus. The students gained insight into the most common verbs that were discussed during the lesson and were given the opportunity to get acquainted with the frequency data. In addition, students were presented with the most common noun collocations that co-occur with these prefixed verbs and with their context in the concordances that were selected by the teacher from the hrWaC corpus. Selected concordances were also used to produce reading comprehension exercises. Students were presented with printed teaching material to the that contained examples of prefixed verbs extracted from the corpus. Based on the corpus material, students have mastered new grammatical rules and adopted new vocabulary. The concordances were also used as a reading exercise where each student gained insight into the context of the prefixed verbs. Students were introduced to the concept of a corpus to explain its meaning and role in language teaching. She did not use the corpus directly in the classroom.

### **Group 2 (CEFR level A1)**

The students belonged to the beginner level A1. The group consisted of 17 participants. The majority of students had Spanish as their mother tongue, followed by German, Korean, Italian and Portuguese. The corpus was not used in direct teaching nor mentioned because it was felt to be a further burden and could confuse students who are absolute beginners. This approach was based solely on teacher's own preparation of the teaching materials (only exercises) from examples found in the hrWaC corpus additionally adapted to the class. lessons were held in which the students were taught the dative case. Exercise after the lesson did not differ from the usual exercises in its structure, although it was based on the corpus examples. It was believed that the lecture could not have been prepared in any other way, since his students possessed a low level of knowledge of Croatian language and a very limited vocabulary.

### **Group 3 (CEFR level A1)**

The students in the group 3 also belonged to the beginner level A1. The group consisted of 17 participants. The majority of students had Polish as their mother tongue, followed by Turkish, English, Arabic, French, Spanish, Albanian, Chinese and Korean.

hrWaC corpus was used to prepare the lesson about food and drink vocabulary. Additionally, students were taught about collocations and collocational range for the adjective + noun. The lesson began with a presentation where students learned about the significance of the corpus and the existing Croatian computer corpora that are

available online. The two largest corpora for the English language: BNC<sup>1</sup> and COCOA<sup>2</sup> were also mentioned. The reasons for using corpus in foreign language teaching were discussed with students and the evidence of the use of corpora in teaching collocations, vocabulary, grammar and reading exercises was provided. After that, students learned about the concept of collocations and collocation range and were provided with the examples they will encounter in the class. In addition to preparing teaching materials and exercises, corpus was used in direct teaching instructing the students how to correctly form the corpus query language (CQL) query. The students also learned to make a concordance list and frequency counts and interpret them.

#### **Group 4 (CEFR level B1)**

The students in the group 4 belonged to the intermediate level B1. The group consisted of 14 participants. The majority of students had Spanish as their mother tongue, while other participants were native speakers of Albanian, Japanese, Macedonian and French. Different approach to teaching was used. A study of the lexical nest of the verb *to play* using the hrWaC corpus was conducted. The corpus was used in direct teaching for students to learn about the possibilities the corpora can provide and to teach the ways of searching the corpus. At the beginning of the lesson the students learned about the importance of the corpus, and immediately started the direct search of the corpus. They were instructed how to form the CQL query to find the lexical nest. the results of CQL queries, concordance lists and KWIC (Key Word in Context) format of a concordance (where the search word is displayed in the centre with some context to the right and some context to the left of it) were clarified. students and teacher went through concordance lists and analysed the context of the verbs that belong to the lexical nest of the verb *to play*. The corpus provided students with the context of use and helped them to learn the difference in the meaning of certain verbs. In the next step, the frequency counts were created where students learned about the most frequent verbs belonging to the lexical nest of the verb *to play*. Teachers involvement consisted mostly of analysing concordance lists and studying the context of the surrounding content. Based on the concordance lists, together with the students, the teacher analysed the meaning of individual verbs and their context to enable the students to notice the differences in the use of verbs in a particular context. The emphasis of the lesson was on discovering the meanings, context and verb conjugation, and on gathering grammatical and lexical knowledge.

### **FINDINGS**

#### **Research Question 1**

*To what extent are students satisfied with the corpus-based approach in teaching a foreign language?*

---

<sup>1</sup> British National Corpus

<sup>2</sup>The Corpus of Contemporary American English



The analysis of the satisfaction of the participants was conducted on the overall sample of all four groups of participants. The total number of participants in the research was 66.

Only one (A1 level) out of the four groups of participants was not made familiar with the concept of the corpus, and it was not emphasized to the participants of that group that materials for the exercise were prepared using the corpus.

The group at the B2 level did not use the corpus directly, but the lesson was based on the corpus examples prepared by the teacher and students were made familiar with the concept of the corpus.

The participants in the third and the fourth group (A1 and B1 level) had the possibility to use the hrWaC corpus directly in class and were trained by the teacher to search the corpus using CQL queries.

Out of the total number of participants in the research, 62.5% of them were female, while there were 37.5% male students. The age structure of the students was as follows: 55% of the students were between 20 and 29 years old, 12.5% of them were less than 20 years old, 12.5% of the students were between 30 and 39 years old, while 12.5% of them were between 40 and 49 years old. Finally, 5% of the participants were between 50 and 59 years old, while only 2.5% of them were over 60 years old. Most students were from Spain (28.2% of them reported Spanish as their mother tongue). The second most represented mother tongue was Polish (17.9%), 7.7% of participants reported German as their mother tongue, 5.1% of participants reported Portuguese, English, French, Turkish, Korean or Albanian as their mother tongue, while 2.6% of students reported Russian, Italian, Macedonian, Chinese, Arabic or Japanese as their mother tongue.

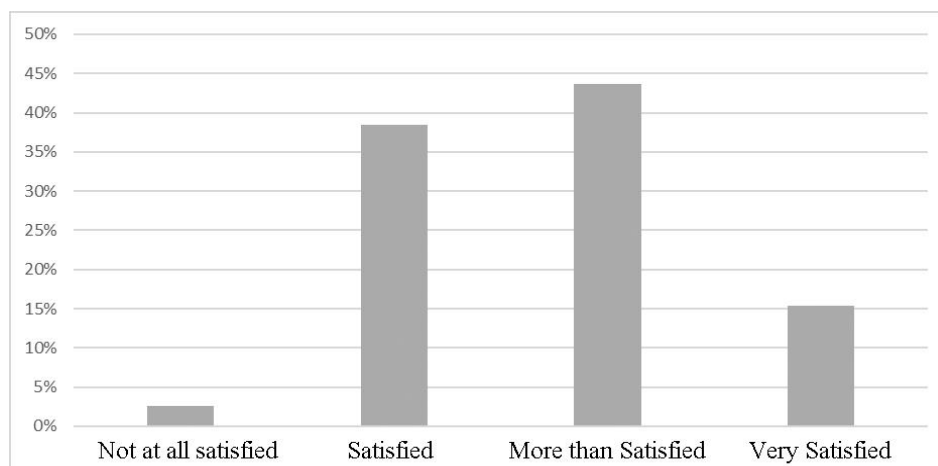


Figure 1  
The Degree of Satisfaction of the Croaticum Students with the Experimental Lecture Prepared using the Corpus

On a scale from 1 to 4 (1 = not at all satisfied and 4= very satisfied), 59% of the participants were more than satisfied or very satisfied with the experimental lectures, 38.5% of the students were satisfied, while only one student expressed dissatisfaction with the experimental lecture (Figure1).

The students have also evaluated the exercises prepared by teachers on a scale from 1 to 4. Only one student was dissatisfied with the structure of the exercise, perhaps because he did not understand it, 64.1% of the students were more than satisfied or very satisfied, while 33.3% of participants were satisfied with the structure of the exercise.

Fourteen students left their own comments expressing the positive attitude towards corpus examples and considering the exercise with the corpus search as an interesting and useful way of learning that they can use independently. But, two intermediate level learners expressed the need for the class to last longer than one school period and stated that they were overwhelmed by the existence of a huge number of examples.

Furthermore, the survey aimed to explore the students' opinion about the most effective ways of mastering the Croatian language. The questionnaire offered five categories that the participants could choose from. Each student was able to select one or more categories. The participants selected communication as the most effective method for language learning (64.9%), while listening was selected by 40.5% of students, followed by writing (43.2%). Moreover, 24.3% of participants considered internet usage as an effective way of language learning, while only 16.2% of participants considered interactive computer applications as an effective way of language learning (Figure2).

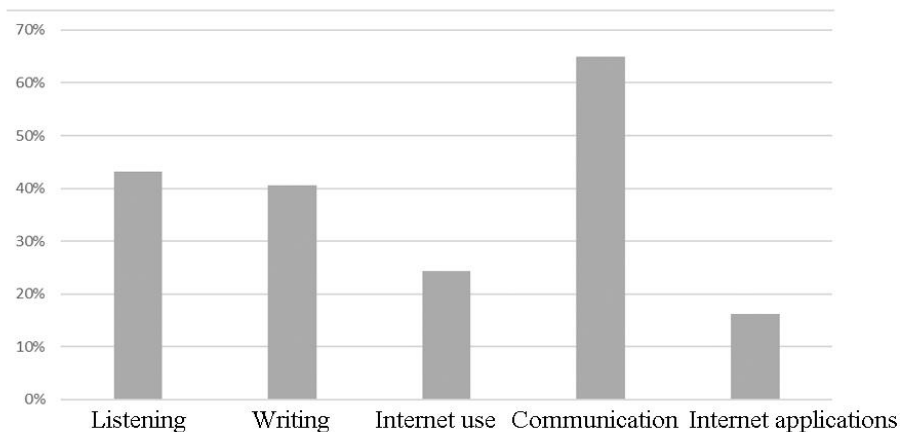


Figure2  
Students' Attitudes on Effective Ways of Learning Croatian as a Foreign Language

Also, the participants were surveyed about the activity they find the most interesting when learning a foreign language. Each participant was able to select one or more categories.

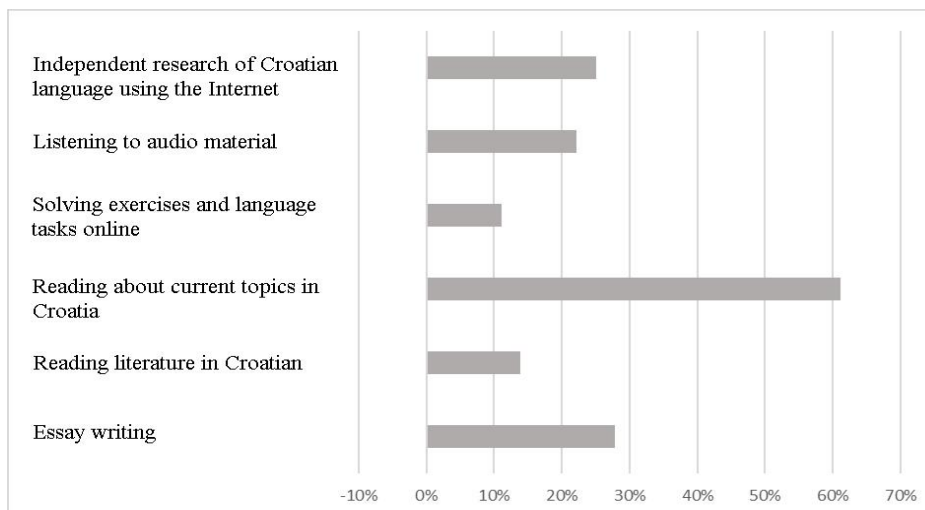


Figure 3  
Students' Attitudes about the Most Interesting Activities for Learning Croatian Language

The participants selected *reading about current topics in Croatia* as the most interesting language learning activity (61.1%), while *essay writing* was selected by 27.8% of them. Furthermore, *independent research of Croatian language using the Internet* was selected by 25% of students, while *listening to audio material* was selected by 22.2% of them and 13.9% of the participants were interested in *reading literature in Croatian*. Finally, only 11.1% of the participants were interested in *solving exercises and language tasks online* (Figure 3). Apart from the already offered categories, 19.4% of the participants left their own comments about what they considered interesting when learning a language. The comments were: communication with the native speakers, conversational exercises, multimedia presentations of learning topics, interactive dialogues, vocabulary contexts and concrete examples from everyday life.

### Research Question 2

*Is there a difference in attitudes towards the direct use of corpus in class between the students at the beginner level and the intermediate level of foreign language learning?*

The separate analysis of responses of the beginner level students and the intermediate level students of CFL showed that beginners were more inclined to think that the direct use of corpus in class would be beneficial for their language learning (61.9%), while the students in the intermediate level, who had greater knowledge of the Croatian language and were therefore more independent in their work, felt that the direct use of corpus in the class would not benefit their language learning (57.14%).

### Research Question 3

*What are the teachers' perceptions about the use of corpora at the specific language learning level and student engagement?*

All four teachers who participated in the in-depth interviews emphasized that the structure of the student groups was very heterogeneous with respect to their mother tongue, age, sex and level of formal education, which required a different approach to teaching CFL in each group and became an issue in the direct use of the corpus in class. The teachers claimed that DDL approach would have been easier to implement if they have had more homogeneous groups and students with similar mother tongues in the specific group.

Each teacher had his own perception about the use of corpora in teaching at the specific language learning level.

The first teacher at A1 (group 2) level claimed that corpora cannot be used in the beginner level of language learning and refused to use the corpus in class. According to the teacher, the students were somewhat interested in teaching materials, while their engagement in class was low.

The other teacher at A1 level (group 3) claimed that her group of students was particularly interested in information on the frequencies and possibilities of independent use of the corpus. In her lesson, she included teaching collocations and collocation span, which is not the part of the curriculum at the beginner level. According to the teacher, the engagement of her students was high, and their understanding of collocations was very positive. They were open to communication and active participation in the class, which is exceptionally good for the beginner's level of language learning. The teacher proved that corpora can be used in the beginner level of language learning, which was confirmed by the high degree of engagement and satisfaction of her students.

The teacher at B2 level (group1), who also had experience in teaching A1 level, claimed that corpora could be used in the beginner level of language learning because students possess little knowledge of Croatian and therefore it is of the utmost importance to provide them with an insight into the frequency of the vocabulary that they will learn beforehand. The same teacher believes that corpora can be useful for learners in the higher levels of language learning, but also concludes that raw corpus examples could be too complicated for lower level of language learning and that they should be chosen carefully by teachers to match the level of knowledge of the learners. According to the teacher, students in her group (B2 level) had a high level of engagement in class and were open to communication but, compared to the previous generations of learners, had a much lower level of language competence and were a larger group, which were the main reasons why she opted for the indirect use of corpora in teaching.

The last of the four teachers who taught B1 level students (group 4) claimed that corpora are very useful in the higher levels of language learning, such as her group (B1 level), because students at that level possess enough knowledge of Croatian language. She advocated the use of corpora at the higher levels of language learning to elaborate materials, such as connectors, adverbs, morphosyntactic cases and idioms. According to the teacher, students in her group (B1) were very interested in teaching materials and corpus use, while their engagement in class was very high. They were open to communication and active participation in the class.

Based on these findings, we can conclude that the chosen approach to corpus-supported foreign language teaching depends largely on the perception of teachers about the integration and the use of corpora in teaching. It is also evident that corpora can be used directly in class with absolute beginners. It is important for teachers to discover the most appropriate way in which corpora can be incorporated in their teaching units.

### **Limitations**

It is important to emphasize that there are certain limitations of our experiment. The teachers have arbitrarily decided to use a single corpus (hrWaC), each of them has experimented with a single teaching unit and they had no time to subsequently assess if students have really mastered the material. The reason for this is that the teachers have a strict curriculum that must be followed, and there is not much time left for them to include the additional teaching material or methods.

### **DISCUSSION**

Based on the analysis of the student survey, it can be concluded that the degree of students' satisfaction with the experimental classes prepared using the corpus was mostly positive. Also, the students expressed interest in the corpus examples and the exercises they took part in during the class.

Our findings are in line with the results of some of the latest studies that revealed that students mostly felt positively towards using corpora in language teaching and learning (Chujo & Oghigian, 2008; Girgin, 2011; Kayaoglu, 2013; Koo, 2006; Liu & Jiang, 2009). This suggests that, regardless of the foreign language they learn, students have positive attitudes toward computer corpora use in foreign language teaching. Our results are also in agreement with findings of some other studies (Yoon & Hirvela, 2004), showing that students at the beginner language learning levels appreciate the benefits of corpora more than those in the advanced language learning levels. Kayaoglu (2013) revealed that although most of the intermediate level students enjoyed discovering patterns in the corpus, some of them were overwhelmed by the existence of huge number of examples, which is also in agreement with our findings.

Regarding the analysis of students, our research has revealed several significant facts related to the structure of the language class, the possibilities for the corpus use in different levels of language learning, and the level of students' satisfaction with the classes prepared with the use of corpus. The structure of the students at Croaticum is very heterogeneous with respect to their mother tongue, age, sex and level of formal education, which requires a different approach to teaching CFL.

In our survey, individual participants have submitted their own comments, which further clarified their attitudes towards the experimental lecture and its structure. The most frequent comments were related to the lack of time (i.e. some participants believed that more time should be spent on corpus search to master the corpus tools). The next most frequent set of comments was related to a positive attitude towards corpus tools that enabled them to see more contextual examples of language use. A few students stated that there should be more classes of this kind because they are interesting, useful for

further learning and because a corpus is always available for independent learning at home. One participant, although considering the use of corpus tools as a good learning method, reported that the CQL query was complicated and that he would benefit from the simpler method of corpus search. Two students reported that their class was complicated and that the example sentences were full of unknown words.

The comments show that participants were mostly satisfied with the lessons that were prepared and conducted using the corpus, but they also show that teachers should tailor their teaching materials for the specific language learning level and the structure of the specific group.

The results of the students' satisfaction survey were corroborated with the teachers' perceptions about the positive impact of corpus-supported foreign language teaching on student engagement in class, which is in line with previous studies (Bernardini, 2004; Comelles et al., 2012) claiming that students' engagement increases and teacher's role changes when using corpora in the language learning classroom.

Our findings from the interviews with four CFL teachers revealed that teachers support the independent use of corpora at the advanced levels of language learning to upgrade the student's knowledge base. Apart from the use of corpora in class, teachers believe that students can benefit from independent corpora use at home where they can explore corpora and learn the language at their own pace.

## **CONCLUSION AND RECOMMENDATIONS**

This research investigated the role of computer corpora in teaching Croatian as a foreign language. Croatian, as a less-resourced and less commonly taught language, was used as a basis to investigate the possibility of corpus-supported teaching of less-resourced languages through indirect and data-driven approach, as well as students' satisfaction with such course delivery methods.

The indirect and data-driven approach to language instruction in the L2 Croatian context were equally interesting for students in this study (as judged by teachers), thus confirming other findings (e.g. Boulton, 2012 and Mukherjee, 2006) about choosing the approach with the local conditions in mind (e.g. the group's structure).

When measured on a sample of four groups of students (N=66) who participated in corpus-supported foreign language teaching, the students' satisfaction was mostly positive, while their class engagement was significant (according to their teachers).

We can conclude that corpora can assist in the preparation of teaching materials for teaching less commonly taught foreign languages because they allow teachers to extract and teach repetitive linguistic patterns.

Further research could be directed towards the assessment of the knowledge of students who participate in data-driven language learning to find out whether students' results would improve compared to the traditional approach. It would be useful to explore in detail the use of corpus in direct teaching with students. The research that would enable students to use corpora independently in the classroom would certainly provide the

insight into a new way of foreign language teaching. Apart from using computer corpora to support teaching of Croatian as a second and foreign language, it would be useful to conduct research on teaching Croatian language in primary and secondary schools to explore to what extent and how computer corpora can contribute to the improvement of language acquisition among native speakers of a LCTL.

Finally, the last area relates to the role of lexicographers and authors of grammar and vocabulary textbooks, since it is important to explore their opinions and attitudes on the use of corpora in the design of the textbooks.

#### REFERENCES

- Aston, G. (2001). Learning with corpora: An overview. In G. Aston (Ed.), *Learning with corpora* (pp. 6- 45). Houston, TX: Athelstan.
- Barbieri, F., & Eckhardt, S. (2007). Applying corpus-based findings to form-focused instruction: The case of reported speech. *Language Teaching Research* 11(3), 319-346.
- Bernardini, S. (2002). Exploring new directions for discovery learning. *Language and Computers*, 42(1), 165-182.
- Bernardini, S. (2004). Corpora in the classroom. An overview and some reflections on future development. In J. M. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 15-38). Amsterdam: John Benjamins Publishing Company.
- Biber, D., & Conrad, S. (2010). *Corpus Linguistics and Grammar Teaching*. [Pearson Education Monograph Series]. White Plains NY: Pearson Education.
- Boulton, A. (2012). What data for data-driven learning? *Eurocall Review*, 20(1), 23-27.
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348-393. doi: 10.1111/lang.12224.
- Boulton, A., & Pérez-Paredes, P. (Eds.). (2014). ReCALL special issue: Researching uses of corpora for language teaching and learning Editorial Researching uses of corpora for language teaching and learning. *ReCALL*, 26(2), 121-127.
- Calzolari N., Del Gratta R., Fracopoulo G., Mariani J., Rubino F., Russo I., & Soria C. (May, 2012). The LRE Map. Harmonising Community Description of Resources. In *Proceedings of LREC 2012* (1084-1089). Istanbul, Turkey 23-25 May 2012.
- Chan, T., & Liou, H. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-noun collocations. *Computer Assisted Language Learning*, 18(3), 231-250.
- Chujo, K., & Oghigian, K. (2008). A DDL approach to learning noun and verb phrases in the beginner level EFL classroom. In A. Frankenberg-Garcia, T. Rkibi, M.R. Cruz, R. Carvalho, C. Dreito, & D. Santos-Rosa (Eds.), *Proceedings of TaLC 8-Lisbon, 8th Teaching and Language Corpora Conference* (pp. 65-71). Lisbon: Associação de Estudos e de Investigação do ISLA-Lisboa.

- Comelles, E., Laso, N. J., Forcadell, M., Castaño, M.E., Feijóo, S., & Verdaguer, I. (2012). Using online databases in the linguistics classroom: dealing with clause patterns. *Computer Assisted Language Learning* 26(3), 282-294.
- Daskalovska, N. (2015). Corpus-based versus traditional learning of collocations. *Computer Assisted Language Learning*, 28(2), 130-144.
- Fan, M. (2009). An exploratory study of collocational use by ESL students – A task-based approach. *System*, 37(1), 110-123.
- Flowerdew, J. (1996). Concordancing in language learning. In M. Pennington (Ed.), *The power of CALL* (pp. 97-113). Houston, TX: Athelstan.
- Frankenberg-Garcia, A. (2014). The use of corpus examples for language comprehension and production. *ReCALL* 26(2), 128-46.
- Gabrielatos, C. (2005). Corpora and language teaching: just a fling or wedding bells? *Teaching English as a Second or Foreign Language*, 8(4). Retrieved from: <http://www.tesl-ej.org/ej32/a1.html>.
- Girgin, U. (2011). *Corpus-based activities at lower levels of EFL proficiency: The effectiveness of using concordance lines on grammar learning* (Unpublished master thesis). Bilkent University, Ankara.
- Götz, S. (2012). Testing task types in data-driven learning: Benefits and limitations. In: K. Biebighäuser, M. Zibelius, & T. Schmidt (Eds.), *Aufgaben 2.0 - Konzepte, Materialien und Methoden für das Fremdsprachenlehren und -lernen mit digitalen Medien* (pp. 249-276). Tübingen: Narr.
- Hafner, C. A., & Candlin, C. N. (2007). Corpus tools as an affordance to learning in professional legal education. *Journal of English for Academic Purposes*, 6(4), 303-318.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Johnes, T. (1991). Should you be persuaded - two samples of data-driven learning materials, Classroom Concordancing. *ELR Journal*, 4, 1-16.
- Jokinen, P. K. (May, 2018). Researching less-resourced languages: the digisami corpus. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (pp. 3382-3386). Miyazaki, Japan, 7-12 May 2018.
- Karras, J. (2016). The effects of data-driven learning upon vocabulary acquisition for secondary international school students in Vietnam. *ReCALL*, 28, 166-186.
- Kayaoğlu, M. N. (2013). The use of corpus for close synonyms. *The Journal of Language and Linguistic Studies*, 9(1), 128-144.
- Koo, K. (2006). *Effects of using corpora and online reference tools on foreign language writing: A study of Korean learners of English as a second language* (Unpublished doctoral dissertation). Iowa City: The University of Iowa.



- Kreyer, R. (2010). How much wheat is there in the chaff? Issues concerning the use of concordances in the classroom. *Language Forum* 36(1-2), 155-173.
- 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.
- Leech, G. (1997). Teaching and language corpora: a convergence. In A. Wichmann, S. Fligelstone, T. McEnery and G. Knowles (Eds.), *Teaching and Language Corpora*. (pp. 1-23). London: Longman.
- Lenko-Szymanska, A. (2002). How to trace the growth in learners' active vocabulary: A corpus-based study. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus linguistics* (pp. 217-230). Amsterdam: Rodopi.
- Leńko-Szymańska, A., & Boulton, A. (2015). *Multiple affordances of language corpora for data-driven learning*. Amsterdam, Netherlands: John Benjamins.
- Liu, D., & Jiang, P. (2009). Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL. *The Modern Language Journal*, 93(1), 61-78.
- Ljubešić, N., & Klubička, F. (2014). {bs, hr, sr} WaC – web corpora of Bosnian, Croatian and Serbian. In F. Bildhauer & R. Schäfer (Eds.), *Proceedings of the 9th Web as Corpus Workshop* (pp. 29-35). Association for Computational Linguistics, Stroudsburg.
- McCarten, J. (2007). *Teaching Vocabulary: Lessons from the corpus, lessons for the classroom*. Cambridge University Press.
- McEnery, A., & Xiao, R. Z. (2010). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 364-380). London & New York: Routledge.
- McEnery, T., Xiao, R. & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.
- Miangah, T. M. (2012). Different aspects of exploiting corpora in language learning, *Journal of Language Teaching and Research*, 3(5), 1051-1060.
- Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art – and beyond. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus technology and language pedagogy: New resources, new tools, new methods* (pp. 5-24). Frankfurt am Main, Germany: Peter Lang.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins.
- O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom. Language use and language teaching*. Cambridge: Cambridge University Press.

- O'Keeffe, A., & Fair, F. (2003). Using language corpora in initial teacher education: Pedagogical issues and practical applications. *TESOL Quarterly*, 37, 389-418.
- Pérez-Paredes, P., Ordoñana Guillamón, C., & Aguado Jiménez, P. (2018). Language teachers' perceptions on the use of OER language processing technologies in MALL. *Computer Assisted Language Learning*, 31(5-6), 522-545. <https://doi.org/10.1080/09588221.2017.1418754>.
- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge: Cambridge University Press.
- Römer, U. (2006). Pedagogical applications of corpora: Some reflections on the current scope and a wish list for future developments. *Zeitschrift für Anglistik und Amerikanistik*, 54(2), 121-13.
- Römer, U. (2008). Corpora and language teaching. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An International Handbook - volume 1* (pp. 112-130). Berlin: Mouton de Gruyter.
- Sinclair, J. M. (2004). *How to use corpora in language teaching*. Amsterdam: Benjamins.
- Smirnova, E.A. (2017). Using corpora in EFL classrooms: The case study of IELTS preparation. *RELC Journal*, 48(3), 302-310.
- Soria, C., Russo, R., Quochi, V., Hicks, D., Gurrutxaga, A., Sarhimaa, A., & Tuomisto, M. (May, 2016). Fostering digital representation of EU regional and minority languages: The digital language diversity project. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation* (pp. 3256-3260). Portoroz, Slovenia 23-28 May 2016.
- Tucker, G. H. (2009). Exposure, expectations and probabilities: Implications for language learning. In A. McCabe, R. Whittaker & M. O'Donnell (Eds.), *Advances in language and education* (pp. 239-253). London: Equinox.
- Tymoshuk, R., Blagoeva, D., & Sosnowski, W. P. (2018). New Bulgarian, Polish, and Ukrainian phraseology and language corpora. *Cognitive Studies*, 18, 1-13.
- Vyatkina, N. (2016). Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning & Technology*, 20(3), 159-179.
- Ward, M. (2016). CALL and less commonly taught languages – still a way to go. In S. Papadima Sophocleous, L. Bradley & S. Thouésny (Eds), *CALL communities and culture – short papers from EUROCALL 2016* (pp. 468-473). Research-publishing.net. <https://doi.org/10.14705/rpnet.2016.eurocall2016.608>
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13(4), 257-283.