Pregledni članak
UDK: 338.48
811.111:37]378.633(497.583Split)

Primljen: 15. 6. 2016.
Prihvaćen: 9. 9. 2016.

# VOCABULARY DATA-DRIVEN LEARNING IN ENGLISH FOR SPECIFIC PURPOSES: THE CASE OF STUDENTS' VOCABULARY FILES

*Sanja Marinov*[*]
*Ekonomski fakultet Sveučilišta u Splitu*

**This paper introduces the idea of using data-driven vocabulary files as a teaching practice for the course "English for Tourism 3" taught at the Faculty of Economics, University of Split. First, the design of the vocabulary file is explained as well as the rationale behind the exercise. Then, the analysis of the students' response to the assigned exercises is presented. Finally, given the complex nature of the vocabulary file, only the vocabulary-oriented exercises are discussed while the content-oriented part of the file will be addressed in another paper. Overall, the students did not seem to have particular problems in dealing with the exercises. The outcomes of their vocabulary files reflect a range of pedagogical issues that may be discussed in relation to any other language teaching and learning activity, such as motivation, learner autonomy, learning styles, learning strategies, or group work collaboration.**

*Key words: data-driven learning, vocabulary, English for Specific Purposes, vocabulary file*

## 1 *INTRODUCTION*

This paper draws on the legacy of the corpus-based approach to linguistics and language education. The empirical nature of corpus data has made linguistic analysis more objective (McEnery & Wilson, 2001), and it has led to a series of revolutionary conclusions, as for example, that a sharp distinction between lexis and grammar cannot be made, but one can rather talk about lexico-grammar (Sardinha, 2012) or pattern grammars (Hunston & Francis, 2000). Such findings have been reflected in language education in many ways. For example, the concept of pattern grammar has become an important

[*] smarinov@efst.hr

"resource of vocabulary building in which the word is treated as part of a phrase rather than in isolation" (Hunston, 2002:106), a notion taken for granted in today's language teaching.

Today, corpus data have a dual role in language education. The indirect approach sees corpus data used in reference publishing (designing and producing dictionaries and grammars), syllabus design, materials development, and language testing. This use of corpus data is widely present, unlike the direct approach which focuses on the learners and their teachers using the corpus data directly, without mediation of course books, grammars, or dictionaries. The direct approach is today widely known as *data-driven learning* (DDL), a term coined by Tim Johns (1986) who launched this quest for viable ways of direct corpus exploitation that have been attracting attention of an increasing number of scholars and practitioners (the 12th Teaching and Learning with Corpora conference took place in July 2016 and the journal *Language learning & technology* will be dedicating a special issue to this topic in 2017).

The precision of the term, regarding what it really stands for, has been debated by researchers in the field, and questioned by Johns himself (see Boulton, 2011). Nonetheless, the term seems to be firmly established by now, but its applicability to different pedagogic contexts is open to a wide range of interpretations and is still being negotiated, and the possibilities and modes of its application are continuously researched (see Boulton & Tyne, 2014: 300-309 for an overview of the research). This study presents a case of using DDL for vocabulary development with second-year students of Tourism. It is designed to be as close as possible to ordinary teaching practice to make it more easily acceptable, a necessary step in helping DDL reach a wider audience (Boulton, 2010).

## 2 *LITERATURE REVIEW*

The direct usage of corpora in language education can take various forms and can thus be further broken down into hands-on work with corpora and paper concordancing. Hands-on corpus work assumes autonomous users handling corpus querying software and performing their own searches of raw corpus data in order to confirm a hypothesis they have formulated themselves. Paper

concordancing sees novice or less experienced/skilful users working with an edited printout of abridged or pre-processed corpus samples prepared beforehand. Leech (1997) refers to the former as the hard version and to the latter as the soft version of the approach. However, the picture is not black and white, as the proposed terminology may suggest, because "DDL activities can be plotted on a cline of learner autonomy, ranging from teacher-led and relatively close concordance-based[1] activities to entirely learner–centred corpus browsing projects" (Mukherjee, 2006:12). Boulton (2012:25) refers to this range as a continuum of DDL exercises.

One of the potential problems preventing a more widespread usage of DDL is the use of technology involved in the hands-on approach (Chambers & Kelly, 2004). Boulton and Tyne (2014:126), although strong proponents of DDL, remind us that we should not exaggerate and use something just because it is available because the usage of data, tools and methods are justified only if they contribute to learning and respond to a real need. Furthermore, Todd (2001:93) notices that when using concordances in language teaching they are commonly presented as "a carefully selected collection of instances of an item" and the learners are asked to notice patterns or draw various conclusions from it.

Independently of the level of autonomy in direct corpus use, research has suggested an extensive range of items that can be observed using DDL, such as: lexical items; lexico-grammatical features or word grammar; multi-word units; derivational morphology; grammatical patterns; comparison of items which are distinct but which present some similarities; and comparison of the same feature in different text types (Kerr, 2013: 25-27). The thusly formulated taxonomy of DDL tasks indicates an important role of lexis in DDL activities which is in line with several important developments that have influenced language teaching, e.g. a lexical syllabus (Willis, 1990) or a lexical approach to language teaching (Lewis, 1993).

Mastering the vocabulary of a language is an enormous undertaking due to a large number of words, a fact rendered more complex by the qualitative aspect of vocabulary learning, i.e. the numerous partnerships and

[1] A concordance is a list of occurrences of a searched item in a corpus with its linguistic context. Concordances can be presented in sentence format or key-word-in-context (KWIC) format.

patterns individual words create with other words (Laufer & Nation, 2011). Furthermore, one encounter with a vocabulary item is not enough to learn it. It has been estimated that the learner must view it on five to sixteen occasions in order to fully acquire it (Nation, 1990). Cobb (1997:315) observes that second language learners may not have the time they have had with their first language "for rich, natural, multi-contextual lexical acquisition to take place," and therefore suggests concordancing or data-driven learning as a way of presenting lexis in multiple contexts.

Presenting vocabulary in a set of corpora – derived contexts and/or in a concordance format, provides "contexts [that] are rich, varied and plentiful but they are also short, incomplete, and do not form a continuous storyline," but they may "offer some opportunities for contextual word learning that are not present in other more conventional text types" (Cobb, 1999:347, 348). For example, Cobb (1999) showed that in terms of defining and recognising the studied vocabulary, students who learned their vocabulary items from wordlists and dictionary explanations were equally as good as students learning from concordances and creating their own dictionary entries. However, the concordance-lexicography group was much more successful in the transfer of knowledge to novel contexts.

More recent studies have also proved DDL effective in important areas of lexical acquisition, e.g. Azzaro (2012) in a study of phrasal verb acquisition through DDL or Huang (2014) in a study of the acquisition of lexico-grammatical patterns.

## 3 *THE PRESENT STUDY*

### 3.1 *Aims*

The aims of this paper are to explain the design of a set of vocabulary-oriented DDL exercises and to study how students approached the tasks, how efficiently they processed the information, and how they presented the results.

### 3.2 *Participants and resources*

The subjects of this study were sixty two second-year students of Tourism at the Faculty of Economics, University of Split. They were involved in the studied activity as a part of their compulsory English in Tourism 3 course in

the 3rd semester of their studies. The study was conducted in the autumn/winter semester of the academic year 2015/2016. The course curriculum includes five main topic areas: Tourist attractions, Tour guiding, Benefits and drawbacks of tourism, Sustainable tourism, and Marketing in tourism. The students who participated in this study were already familiar to some extent with DDL exercises since they were occasionally used in the teaching process in their 1st and 2nd semesters.

The exercises were prepared using Sketch Engine, a corpus tool that provides access to ready-to-use corpora and also allows users to build, upload and install their own corpora (Kilgarriff et al., 2014). The corpora can be browsed and queried using a vast range of functions, of which only the core ones were applied in preparing the current material. Sketch Engine for Language Learners (SkELL), one of the most pedagogically oriented offspring of Sketch Engine, was also adopted in the design of the exercises. It is a user-friendly language learning website where all the information is corpus-based and is presented in three different formats: word sketch, examples and similar words (Kilgarriff et al., 2015).

### 3.3 *Design and procedure*

The vocabulary-oriented DDL exercises analysed in this paper were a part of a comprehensive activity designed to accompany and supplement the topics of Tourist attractions and Tour guiding. The whole procedure started with a pre-test that checked the students' knowledge of several key content and vocabulary items related to the topics. It was followed by students' autonomous group work on a set of corpus-based exercises, i.e. *the DDL vocabulary file*. Autonomous, however, does not imply individual, because the students were engaged in collaborative group work and worked in 12 groups of 5 to 6 students in each. The progress the students achieved while studying the files was checked by a post-test.

The DDL vocabulary file was designed around a number of principles:

i *The exercises were all related to one topic*. Exercises related to the area of students' main interests should be more motivating (Tribble, 1997) although research does not always obtain that kind of feedback from students (Kabalin Borenić, Marinov, Mencer Salluzzo, 2013).

ii *The exercises were a combination of content-study and vocabulary-study*. The prevailing idea is that in teaching a language for specific purposes (LSP) the teacher's role is to provide language and the participants provide the context since they are the experts in the field. Teaching experience has taught us that this is often not true in the first and second year of study. Students are still novices in their field(s) of study and may lack the content that can be mediated through the language course content.

iii *The exercises favoured paper-based exercises*. Paper-based exercises were favoured as a less threatening approach than a hands-on approach for the students whose main interest is not the study of language. Hands-on exercises were only applied to some items of the vocabulary-related section.

iv *In some exercises the corpus data was presented in the form of full sentences and in the others as key-word-in-context*. In their previous encounters with the key-word-in-context (KWIC[2]) presentation of data, students were at times concerned about the lack of context in truncated concordance lines. The choice was led by the purpose of the exercise.

v *The exercises were guided by questions and/or instructions*. The introductory instruction and guidelines to the exercises were presented on two levels: (i) *guided exercises* asked specific questions that had to be answered based on the observed data; (ii) *self-study exercises* allowed more freedom to the students, i.e. they were asked to observe the data and make conclusions about the usage of a particular language item or select and classify data according to their own criteria.

Out of the total of sixteen exercises in the DDL vocabulary file eight were vocabulary-related (five guided and three self-study exercises). This paper focuses on the students' work with eight vocabulary-related exercises, while the content-related exercises and the results of the pre-test and the post-test are to be addressed and published in another paper.

[2] Key-word-in-context (KWIC) is a way of presenting corpus extracts/samples by placing the key item (called the node) in the middle of the display and the context to the left and to the right is defined by a determined number of spaces and not by the beginning and end of a sentence.

The students first downloaded their DDL vocabulary files from Moodle (a course management system used at the Faculty of Economics in Split) and then worked in groups of five or six to complete them. The DDL vocabulary files were completed as home assignments and the time period allowed for their completion was two weeks. The students were allowed to organise/distribute the work in any way convenient to them, but the final outcome had to be submitted as a single document with all group members familiar with its contents. Finally, individual groups were invited to attend an interview about their results and the problems they encountered while working on the vocabulary files. This was also an opportunity for the subject teacher to provide them with the necessary feedback.

The choice of vocabulary items to present in these exercises was led by the task that lied ahead of the students, i.e. to research a particular Split city centre sight and to prepare a 10-minute video on the assigned sight. Therefore, the presented items had to be easily recognised by the students as possibly useful for their own presentations. An additional criterion was to focus on the items that were perceived as either a potential source of error or underused by Croatian speakers. Part of the data was derived from the British National Corpus and part from the SkELL tool, as will be described in detail in the *Results and findings* section.

## 4 *RESULTS AND FINDINGS*

The process of creating exercises is research in itself and even more so if it involves browsing corpora and extracting the relevant data. Therefore, these processes and the corresponding results regarding each particular exercise are presented in the *Results and findings* section of this paper along with the observation of students' responses.

### 4.1 *Designing the guided vocabulary exercises*

#### 4.1.1 *Exercise 1*

The vocabulary item introduced in the first exercise was the verb *to house*. The long-standing observation of the students of tourism, who are all native speakers of Croatian, has shown that they often have a stronger receptive than productive knowledge of this verb and therefore tend to underuse it. The

cross-linguistic influence theory recognises this phenomenon as *avoidance* which stands in contrast to *preference* indicating that L2 users prefer some language items over others (Scott & Pavlenko, 2008). The exercise thus aimed at raising awareness of the presented item.

The number of instances of this item in BNC (112,181,015 tokens) is overwhelming (4,250 or 37.88 per million). Therefore, the selection of data was facilitated by a Sketch Engine's feature from which it derives its name, i.e. Word Sketch. This is a "one-page, automatic, corpus-derived summary of a word's grammatical and collocational behaviour" which "presents lists of collocates in columns of grammatical relations (gramrels) or colligations" (Thomas, 2015:161). It draws attention to the most frequent words accompanying the searched item and may facilitate or direct our choice of what to present.

The noun *museum* was found to be a highly frequent subject of *to house* and the first exercise was built on this information. The exercise was presented in a 15-line KWIC format. The students were asked to notice what the *museums* commonly *house* (e.g. *an art exhibition*), to identify the modifiers of these items (e.g. a *temporary* art exhibition), and to present the findings visually for later easier retrieval for study reasons.

#### 4.1.2 *Exercise 2*

As it often happens with browsing through corpora, one exercise easily leads into another, triggered by the language noticed in the process of developing the previous one. Therefore, the following exercise, consisting of 20 KWIC lines, kept the focus on the verb *to house* but with another frequent subject highlighted by Word Sketch, namely the noun *building*. Changing the noun from *museum* to *building* in the left context of the node[3] caused the right context to also change, and the students were asked to define the change. The main task was accompanied by two more questions: one about the forms of the verb *to house* and what they depended on (some grammar was thus revised) and one about the use of the word *formerly* (used twice). Finally,

[3] Node is the term describing a key language item (either a single word or a longer phrase) that is used as a search item in browsing corpora and is therefore displayed in the centre of the KWIC display.

students were asked to contextualise their findings by writing three sentences containing the studied items and speaking about their own experience.

### 4.1.3 *Exercise 3*

The third exercise was also indirectly the outcome of the first two because the noun *exhibition* (the node word of this concordance) was frequently found in the context of *museum houses*. However, the focus was on the prepositions that follow the noun *exhibition* since "prepositions are a common source of problems for learners in a number of languages" (Chambers, 2005) and the patterns including prepositions can be well displayed in a KWIC concordance. The use of prepositions *at*, *of*, and *on* after *exhibition* (identified using Word Sketch) was illustrated by 30 KWIC concordance lines. The students' task was to identify the meaning acquired by the three prepositions in the presented contexts and to visualise the findings.

### 4.1.4 *Exercise 4*

The fourth exercise took the noun *sculpture* as a key vocabulary item since there are several important sculptures in the Split city centre. The material for the exercise was drawn from SkELL which, as explained above, was made to bring corpora closer to novice and less proficient users, building a bridge between dictionary usage and corpora consultation. SkELL has a very user-friendly interface that, based on a frequency analysis, selects the frequent word partnerships and classifies them into several categories (Figure 1).

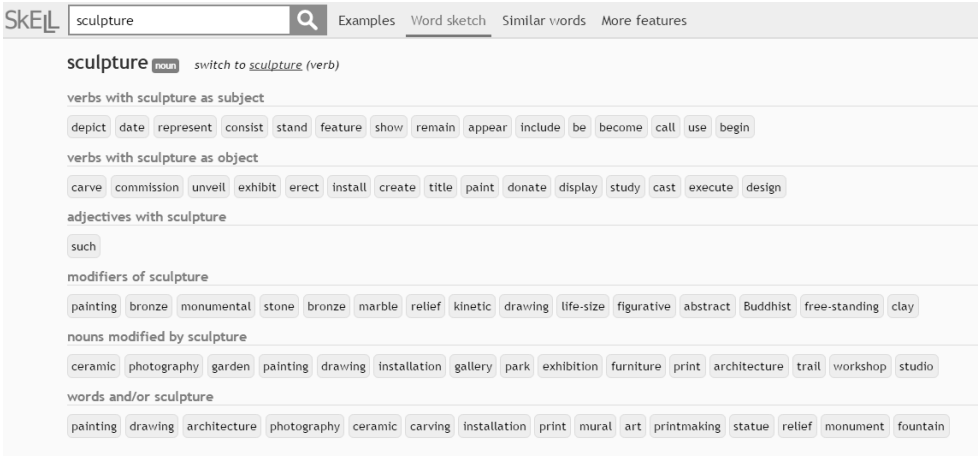


*Figure 1: The Word Sketch of the noun sculpture as presented by the SkELL interface*

From SkELL's Word Sketch for *sculpture* we selected 5 verbs used with sculpture as subject (*depict, represent, feature, date, stand*), 5 verbs used with sculpture as object (*dedicate, unveil, install, cast, erect*) and 5 modifiers of the noun (*bronze, figurative, free-standing, Romanesque, monumental*). Each item was illustrated by two sentences from SkELL's examples (Figure 2). The 30 sentences thus obtained were presented in random order. Students were asked to identify the collocates and to categorise them in the same categories from which they were selected. Finally, they had to contextualise some of these items by writing two example sentences for each category.

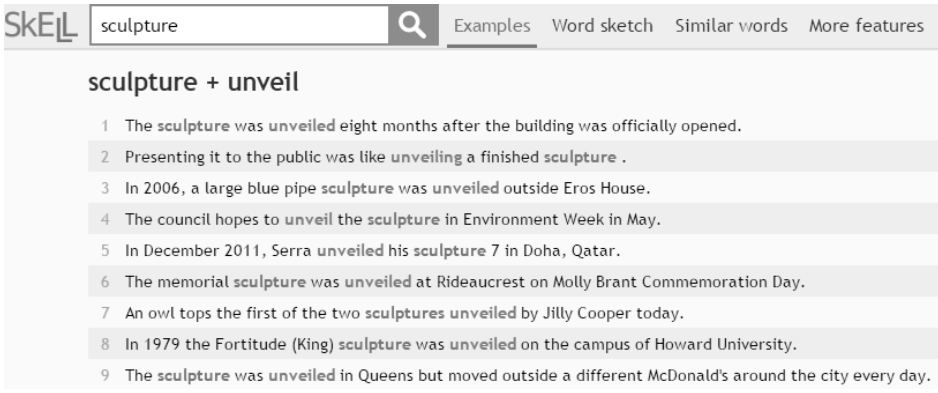


*Figure 2: Extract from SkELL's examples page for sculpture + unveil*

4.1.5 *Exercise 5*

The design of the last exercise in this section focused on the noun *square* because squares are the heart and soul of the city of Split. SkELL was used again as the starting point, and the following word partnerships with square were selected: *enclose*, *overlook*, *line*, *desert* (*square* as object); *become*, *face* (*square* as subject); *public*, *main* (modifiers). There were 16 sentences illustrative of these word partnerships (two for each item). The students were first asked to identify the key word being observed, and then they were asked to answer 7 questions guiding them in noticing particular items. Finally, they had to contextualise them by setting illustrative examples/sentences describing their own environment.

### 4.2 *Designing self-study vocabulary exercises*

The three exercises described in this section are self-study only relative to the previous five which were more guided. The students were either asked to observe the data and make conclusions or to extract useful language from a larger amount of data. However, no specific questions were asked to guide them through their observation in the first exercise, and they were given the freedom of choosing the useful vocabulary to include in their vocabulary files in the second and third exercise.

#### 4.2.1 *Exercise 1*

The first exercise addressed the usage of the verb *to date* when used to indicate that something has its origin at a particular time or that it has existed since a particular time. It is reasonable to assume that it would often be used in tour guiding and the teaching experience indicates it is an underused item in Croatian students' production, although they have no problems recognising it. The Word Sketch makes difference between *back* being used with *date* in transitive (Figure 3) and intransitive (Figure 4) ways. This is indeed interesting for teaching and comparison because the two fall into different frames and are followed by different types of information. Finally, the concordance featuring *date from* was also added. This resulted in three sets of 15 KWIC concordance lines with the node *date* illustrative of the above three usages. The students were asked to find out what the three sets had in common and how they differed.

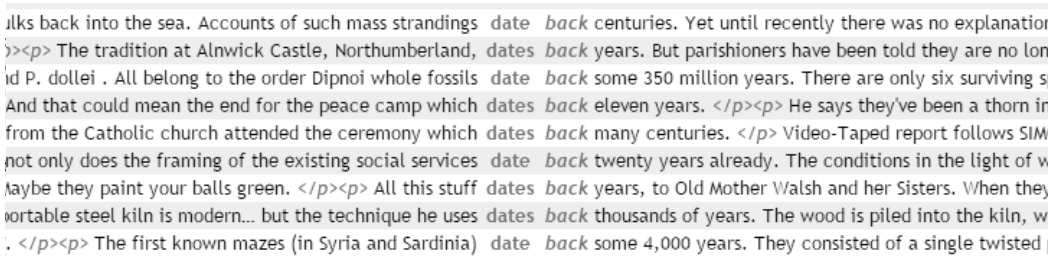


*Figure 3: Transitive use of* date *followed by* back

e people angry - that's my game.'  The booing dates *back* to Cooper's early days at his native Durham Wasps. He
date.  When in 1961 an Anglo-Kuwaiti accord dating *back* to 1988 expired, Kuwait became independent. Iraq (itself
abolition of rates fulfils an ambition of Mrs Thatcher's dating *back* to 1974. But the interventions, each one usually more
RIC HOUSES IS NOBLE. LIVING IN ONE IS SUBLIME.  Dating *back* to medieval times, Britain's historic houses are one of
Abbey used by an order of Cistercian Monks.  Dating *back* to the 18th Century the main house has been converted
The Grange  The oldest part of this spacious house dates *back* to 1720, with two wings added later. The rooms are high-
ded smuggling, lace-making and quarrying; the latter dates *back* to Roman times.  (b) Hooken undercliff Spectacu
:ester is also a well known hunting centre with a hunt dating *back* to the late 1700s. The Army established its ordnance depo
rs was one of the oldest guilds in London with records dating *back* to 1180. In the 1300s it was renamed the Guild of Grocers

*Figure 4: Intransitive use of* date *followed by* back

### 4.2.2 *Exercise 2*

The second of the three self-study exercises required the students to use a computer with an Internet connection in order to exploit SkELL as a language resource or reference. At the moment SkELL is freely available, and its user-friendly interface makes it particularly suitable for autonomous self-study. The targeted item was the noun *street*. Students were asked to explore the 6 categories in which the collocates of *street* are distributed and to study a minimum of 3 items per category. The selected items were to be accompanied by the illustrative sentence(s) provided by SkELL. In the exercise description the students were informed that the end product would be their personalised profile of the word *street*. The criteria for selection were to be exclusively the students'.

### 4.2.3 *Exercise 3*

The last of the three self-study exercises repeated almost the same procedure of the previous exercise but for the noun *saint*. It should not be surprising that the topics of Tour guiding and Tourist attractions dedicated some space to this item. A lot of legends, stories, churches, and museum exhibits involve referring to one or more saints important for the visited location. The city of Split has a Cathedral named after one of the martyrs of Diocletian's persecutions – St. Domnius, and the procession that takes place on May 7th every year is not only a long-standing tradition in the city but also one of the central cultural and entertainment events. The exercise afforded the students an opportunity to "collect" useful vocabulary they could apply in talking about this topic. It was slightly less binding than the previous one because the students were not limited in the number of categories they had to address or the number of items they had to select. The choice was exclusively theirs.

### 4.3 *Students' response to the guided vocabulary exercises*

#### 4.3.1 *Exercise 1*

This exercise can be described as rather straightforward, requiring the skill of noticing the recurrent pattern by "vertical reading" (as opposed to the regular linear reading) of the KWIC concordances and some organisation of the language noticed. Students' presentations of the findings still differed in approach but mostly in the effort put into formulating the answers. If a classification should be made we can distinguish between:

(i) **Valid and well-organised responses**: responses listed (most often) five nouns indicating what *museums house* (collections, treasures, exhibitions, display, paintings) and the corresponding modifiers underneath each. Group 3 did it in a table format bolding the nouns; group 5 highlighted the nouns in red and provided a dictionary definition; group 10 provided a list of nouns highlighted in red with the corresponding adjectives in brackets; group 12 modified the concordance itself by reducing the first column containing less important content and by extending the part starting with *museum*, putting in evidence the required elements and then extracting the findings into a three-column table. Finally, group 8 did not separate modifiers from nouns but still offered a clear visual presentation of their findings (as shown in Figure 5).


*Figure 5: An example of a visualisation of what* museums house *(by group 8)*

(ii) **Valid but superficially/badly organised**: group 7 identified the most frequent items *housed* by *museums*, they also made a list of modifiers and even provided monolingual dictionary explanations of each, but the findings were not clearly organised; groups 9 and 11 gave two separate lists: a list of nouns (items housed) and a list of modifiers, indicating no connection between them.

(iii) **Fail to respond to the exercise requirements**: group 1 copied several examples; group 6 highlighted in red some of the modifiers in the concordance list; group 4 listed several examples and grouped them according to the form of the verb. Finally group 2 made some valid observations but these did not respond to the exercise requirements (they wrote about the parts of speech that preceded or followed the verb).

4.3.2 *Exercise 2*

The students observed they were at first confused by the question about the difference in the second exercise. It took them some time and discussion with colleagues to draw conclusions, which is partly reflected in their answers that can be grouped as follows:

(i) **Answers with correct explanations**: Six groups managed to identify the difference and explained it in their own words. Some explanations were formulated more successfully than the others but they made the same point.

(ii) **Dictionary definition**: Answers from group 1 and group 5 (Figure 6) clearly show that the students resorted to monolingual dictionaries to help them work out the difference they were looking for, but the examples were selected from the two given exercises. This means they started from the examples and then looked for the best dictionary entry defining this usage.

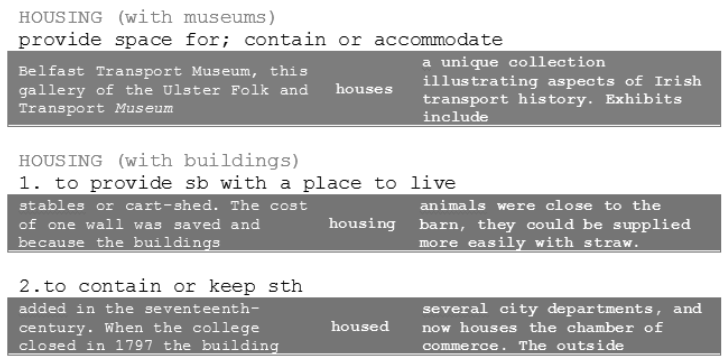


*Figure 6: Dictionary definitions and exercise examples by group 5*

(iii) **Wrong answers**: One group did not make an attempt at an answer, and three groups failed to notice or explain the difference. Two groups thought it was the verb that changed, but they did not provide any explanations, and one group thought the difference lied in the absence of modifiers in the "building examples".

As for the rest of the questions, none of the groups had any problems defining the meaning of *formerly*. However, none of the groups performed the grammatical observation successfully. They all correctly identified the tenses (past simple and present simple) but frequently confused the gerund for the continuous.

In addition, almost all of the students' illustrative sentences provided a good context for the studied item. It was noticed, however, that some followed more closely the examples set by the exercise (e.g. *City Museum of Split houses the Emanuel Vidović gallery.*), while the others relied on their own intuition producing a more metaphorical sense (e.g. *Diocletian's Palace houses the City Museum of Split.*). The students rarely failed to provide a valid context (e.g. *Splitsko-dalmatinska region houses more and more tourists every year.* or *Faculty of economics in Split often houses various projections about different subjects.*) most often using *to house* where *to host* would have been more appropriate.

### 4.3.3 *Exercise 3*

Three out of twelve groups can be said to have failed to respond adequately to the exercise requirements. One group emphasised all the prepositions

preceding the noun and did not attempt any explanations, another one isolated several examples but offered no explanation, and one offered dictionary explanations for the usage of the prepositions but did not limit it to the actual use in this particular context (e.g. *on - some specific place, day of the week, to indicate the state of something*). In six groups the findings were well sorted to be visually more accessible, and more easily retrievable. Visualisations ranged from simple tables and systematic bullet use to smart art graphics. One of these is illustrated by Figure 7.

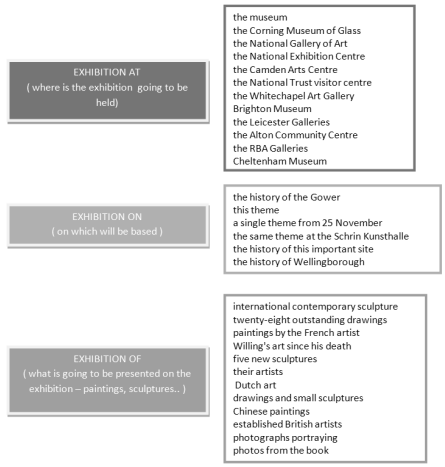


*Figure 7: Visualisation of the use of the prepositions* at, on, *and* of *after the noun* exhibition *(by group 8)*

### 4.3.4 *Exercise 4*

Sorting the *sculpture* collocations looked like a relatively simple task, not much different from a traditional one. However, the students reported that they were mostly confused by the metalanguage used in the exercise instructions, namely the requirement to distinguish between the verbs that are used with *sculpture* as a subject and those used with sculpture as an object. This was also reflected in their responses.

Although there were many various attempts it can be concluded that most groups did not complete the exercise successfully. Some were closer to identifying all the verbs but also added some other verbs from the context that do not collocate with *sculpture*, and the majority did not group them correctly in relation to *sculpture* as a subject or a verb. Four groups succeeded in getting

the classification right (with some minor errors), and they also provided relevant example sentences.

As for the rest of the illustrative sentences it was noticed that they did not always reflect the students' reality. More importantly, even when they were illustrative of the students' environment they did not provide examples of the items observed in the exercise, but they used other verbs or modifiers to describe the chosen sculpture(s). The sentences were either invented by students or copied from some Internet sources.

It is interesting to note that one group, as they explained in the interview, while working on the self-study exercises that actually required them to use the SkELL interface, noticed that the categorisation was the same as the one required in this exercise and searched for *sculpture* in SkELL which provided them with the right answers.

#### 4.3.5 *Exercise 5*

The analysis of the response to this exercise focused on three out of seven questions that guided the students. The first one asked for an explanation of how the verbs *enclose*, *overlook*, *line*, and *desert* were used in relation to the key word. More specifically, the students were asked to paraphrase them or find synonyms. The other two required students to apply the words *enclose* and *overlook* to the context of their local environment.

In defining and paraphrasing, the students, in most cases, resorted to monolingual dictionary definitions or websites providing synonyms. Using these resources, however, yielded different results. Some groups listed too many synonyms, often unrelated to the context being observed (e.g. *enclose - surround, block off, encase, encircle, wrap, bound, cage, circle, confine, enfold, fence; overlook – fail to notice, disregard, ignore; desert - ruin, demolish, occupy*). One group attempted to provide their own explanation based on their observation, and one group did not provide any explanations at all. The verb *line* seems to have caused the most difficulties, so we got e.g. *"line means that something takes place on the square"* when students tried to provide their own explanation and *"line - fill, surround, cover, decorated"* or *"market square lined with a - to place a layer of something over the inside surface of something (synonym: coat)"* when they

poorly chose a dictionary explanation, or *"line - an indication of demarcation; boundary; limit"* where the part of speech was not correctly identified.

A good combination of personal observation and dictionary usage was provided by group 5, who resorted to a dictionary but carefully selected adequate definition corresponding to the given context (*to enclose – to surround or close off on all sides; to overlook – to have a view of from above, to look over on; to line – to stand or to be positioned at intervals along; to desert – to abandon)*.

As for the example sentences applying *overlooking* and *enclose* the students also partly relied on the Internet as a source of information but mostly produced valid, correct and informative examples indicating exactly what some of the most important squares in their city were enclosed by, and which accommodation in Split overlooked some of the most central squares. This is in contrast to the problems they had in defining them.

The exercise did not ask for an example of the verb *to line* which is now seen as a drawback due to the problems the students had in defining it.

### 4.4 *Students' response to the self-study vocabulary exercises*

#### 4.4.1 *Exercise 1*

In defining what *date back* (transitive), *date back* (intransitive) and *date from* have in common, students generally agreed that they all contained the word *date*, they referred to the past, they shared the same meaning (of how old something is), and all sentences described a context of time.

As for the differences there was a great variety of responses. Two groups offered no response, and two groups managed to express the exact difference in what seemed to be their own words. Below is an example of a very detailed account of the observation (Table 1).

*Table 1: Observation about the difference between the three usages of* date *by group 2*

| |
|---|
| "The first set of concordances uses the verb date to refer to a known time in the past. For example, a particular year of century when something took place. The second set uses the verb date together with the amount of time that passed since a certain event. This usage of the verb doesn't have to be precise. In the 3rd set, date is used to describe a rough estimate of the time in the past a certain event took place. It refers to a century or a longer period of time when an event it thought to have happened." |

The remaining 8 groups noticed, e.g., that the observed items were followed by a different preposition, that they were used in a different tense, or they resorted to a dictionary if they were not able to make conclusions based on sentences/item observation. For example, group 7 made use of the definition provided by Oxford Dictionaries Online:[4] "establish or ascertain the date of an object or event". This same dictionary makes a difference between *date* used with and without object, which might have helped to explain the differences presented in this exercise, but the students failed to notice it. The same group also used the MacMillan's entry[5] *"date back to something - begun at a particular time in the past; date from - to be made at a particular time in the past"*, but did not manage to turn it into a correct explanation of the difference.

Group 12 quoted the explanation taken from Cambridge Dictionaries Online[6]: "*date back – to have existed for a particular length of time or since a particular time"*, but they, too, failed to indicate the difference between the two usages (with or without *to* following), although it was contained in the definition.

### 4.4.2 *Exercise 2*

The results of the analysis of students' work with SkELL can be broken down into two groups: (i) students' preferred vocabulary, i.e. the items they selected for the profile of *street*; (ii) findings about the actual classification of collocates for *street* in SkELL. The former is presented in table format (Table 2) for easier reference and the latter is illustrated with a number of examples.

*Table 2: Items selected for inclusion into the word* street *profile and the respective frequency*

| | |
|---|---|
| verbs with *street* as object | line (6x), cobble (2x), patrol (5x), desert (3x), wander (1x), cross (4x), pave (3), walk (7x), flood (1x), crowd (4x), number (2x) |
| verbs with *street* as subject | vend (1x), race (2x), fight (7x), shout (4x), light (5x), chant (1x), protest (4x), border (2x), run (5x), number (2x), surround (5x) |
| adjectives with *street* | narrow (9x), empty (5x), adjacent (2x), safe (4x), quiet (3x), clean (5x), smart (2x), full (1x), alive (4x), busy (1x), fair (1x), wide (2x) |

[4] http://www.oxforddictionaries.com/definition/english/date
[5] http://www.macmillandictionary.com/dictionary/american/date_2#date_33
[6] http://dictionary.cambridge.org/us/dictionary/essential-british-english/date-back-to-something

| | |
|---|---|
| modifiers of *street* | main (5x), crowded (6x), busy (2x), one-way (3x), shopping (6x), residential (2x), quiet (1x), cobblestone (3x), tree-lined (2x), paved (1x), city (5x) |
| nouns modified by *street* | corner (5x), gang (6x), lamp (3x), lighting (2x), fighting (1x), performer (6x), address (5x), racing (1x), parking (8x), cred (1x), prostitution (1x) |
| words and/or *street* | sidewalk (8x), alley (2x), square (6x), avenue (5x), lane (4x), highway (3x), boulevard (4x), park (5x), plaza (1x), road (3x) |

Observing the table it is difficult to establish any particular criteria that led to the students' choices, and they could not explain it themselves in the conversations we had about their work. However, the selected items make it clear that they either did not pay enough attention to the categories they were choosing from, or they did not understand the metalinguistic information contained in the names of the categories (e.g. *verbs with street as object* or *verbs with street as subject*). They relied on the categorisation offered by SkELL which showed to have some drawbacks both in terms of the classification of collocates and illustrating their usage.

Although some confusion can ensue from the verbs that are often used in the passive form, and thus have the same form as in cases when they act as modifiers, this is rather well handled by SkELL. For example: *crowd* (verbs with *street* as object), as in *The streets were more* ***crowded*** *than usual,* and *crowded* (modifier of *street*) as in *He could not bear the* ***crowded*** *streets.*

There is some repetition or questionable distribution of items caused by two categories that may overlap: *adjectives with street* and *modifiers* (e.g. *busy* is repeated in both categories and *residential*, *quiet*, and *main* are classified as modifiers and not as adjectives). These, however, were not expected to mislead the students in dealing with their exercises.

The most problematic category is that of *verbs used with street as subject*. Many verbs listed under this category did not actually have *street* as a subject but formed compounds with it as is evident from the examples in Table 3. As such they would be more appropriately listed as words modified by *street*.

*Table 3: Misleading examples in SkELL - not illustrative of a verb with* street *as subject but a word modified by* street

| |
|---|
| Street Karate is a funny, dynamic street ***fighting*** game. |
| A working paper describing attempts to formalize street ***vending*** in Mexico. |
| You should also make sure you report any information or instance of illegal street ***racing*** to the police. |

Finally, as for the adjectives two out of the fifteen selected by the students were noticed to be out of place in this category. *Fair*, in the sentences that illustrate it, is a noun that forms a compound with *street* as in *The mid-Lent fair is the largest street* **fair** *in Lincolnshire and one of the largest in the country.* On the other hand *smart* forms a compound with *street* as in *Instead, Nancy was brazen and street* **smart** *– at times heading straight into a potentially dangerous situation.*

### 4.4.3 *Exercise 3*

Unlike in the previous exercise, in this one the students were free to choose the number of items they considered relevant for their topic, as well as the categories from which they chose, which was reflected in the findings. However, it was noticed that some groups had simply followed the established system of choosing an item from each category so, e.g., as many as 8 groups included the only adjective offered by SkELL (*such*) which, although it may be frequent, does not appear to be indispensable for the profile of the observed noun.

Overall, the number of selected items per group ranged from four to sixteen, with an average of eight words. The largest number was selected from the category of *verbs with saint as object* (Table 4).

*Table 4: Items selected for inclusion into the word* saint *profile and their respective frequencies*

| | |
|---|---|
| verbs with *saint* as object | commemorate (3x), celebrate (2x), martyr (5x), honour (2x), celebrate (4x), worship (4x), canonize (1x), glorify (2x), depict (3x), revere (1x) |
| verbs with *saint* as subject | live (5x), appear (2x), go (1x), die (5x), stand (2x), look (2x), write (1x) |
| adjectives with *saint* | such (8x) |

| | |
|---|---|
| modifiers of *saint* | blessed (2x), Catholic (3x), patron (8x), Christian (1x), holy (1x) |
| nouns modified by *saint* | day (9x), Peter (1x), angel (3x) blessed (1x) |
| words and/or *saint* | sinner (2x), bishop (5x), relics (3x), prophet (1x) |

In this exercise the students were also required to contextualise some of the selected items in sentences when speaking about their reality in order to indicate the purpose of the whole exercise. In nine out of twelve groups the selected items were successfully illustrated while in the remaining three they were not. It was interesting to notice that the groups lacking their own illustrative examples are those who selected most items to include in the profile (9, 13, and 16 items respectively).

Less incorrectly categorised collocations were noticed with this item. Students' choices drew attention to one: *angel* is categorised as *nouns modified by saint* while the illustrative sentences show it should be in the category of *words and/or saint*.

## 5 *DISCUSSION AND PRACTICAL IMPLEMENTATION*

In terms of designing DDL activities the above analysis has shown that frequently, in light of new data, one activity triggers another. Such was the case with the activities that started from the verb *to house*. We have also seen that it is useful to have sophisticated tools (e.g. Sketch Engine) to facilitate identifying the target items but also less complex, user-friendly tools for direct consultation by students (e.g. SkELL).

Students' own responses/solutions can help us find new directions too. For example, we may wonder why it was more difficult for students to establish the change that takes place in the right context once the word *building* replaces *museum* in the left context of the node word. Is it just the difficulty that seems to be imbedded in the vertical reading of the KWIC presentation of the concordance? It did not seem to be a problem in finding the items *housed* by *museums*. Is the answer perhaps in the lack of critical thinking and "daring" to draw one's own conclusions? If so, these skills should be practiced and, as O' Sullivan (2007) claims, DDL is particularly suitable for developing procedural knowledge as it encourages practicing a wide range of cognitive

skills. In terms of research, a comparison could be made with full sentence presentation to find out if this makes the task easier to either support or dismiss the results of research that has shown that KWIC concordance can sometimes provide better results than full sentence context (Boulton, 2009).

Related to the above are the findings in the *date exercise* and also *exhibition + preposition exercise*, where the students often resorted to the authority of a dictionary to provide them with the correct explanation rather than relying on their own observation and the ideas resulting from it. As was shown in the analysis, the dictionary findings were not always used well in explaining the observed items. In terms of students' preference for references, Frankenberg-Garcia (2005) showed that her translation students also preferred specialist-mediated resources (dictionaries) over materials that involved more learner autonomy (corpora). The results of her study call for more training in integrating skills at using different resources together. Therefore, seen from the pedagogic perspective, having students refer to the available dictionaries is also one of the benefits of DDL exercises and involves them in the process of relating two different sources of information about language.

Some students showed a superior ability to notice and express what was noticed as shown e.g. in Table 1. Moreover, this group's observation was so detailed that it may be said to be at a point of overgeneralisation based on a rather small sample of language. This is, however, a permanent threat of reading and interpreting corpus data, but it should not discourage us from letting students engage in thinking and drawing their own conclusions which can later be negotiated and used as a starting point for a discussion.

Although this study did not measure the actual learning outcomes it nonetheless observed that the Nesselhauf's (2004: 140) view that DDL is "particularly useful for points which have already been covered in the classroom, possibly even repeatedly, but which the learners nevertheless still get wrong" was only partly supported by the evidence from the students' vocabulary files. For example, while dealing with *to house* and *exhibition + preposition* proved to be fairly easy, *date back* proved to be more complex. It is not difficult to conclude that such results reflect the level of difficulty of the task itself where the former required only observing and classifying the

observed and the latter required some conclusions to be drawn and properly worded.

As planned in the DDL vocabulary file design, the technology was introduced in a very unthreatening way by using SkELL. In SkELL the corpus information is already pre-processed and presents "a feast of information on the word" (Kilgarriff et al., 2015:67), and the users only need to help themselves to what they need. Students' selections were varied and also showed that they selected a noticeably smaller amount of items in the 3rd (when they were not limited by the number) than in the 2nd self-study vocabulary exercise. It will be interesting to see in the post-test how good their choice was and how well the students chose their vocabulary and how efficient they were in applying it. For the time being we can only conclude that the students did not cope well with the large amount of autonomy that was allowed in the two SkELL exercises that required them to make their own free choice in creating a word profile. In the future it might be advisable to set more specific and more teacher-guided exercises for students who would then get familiar with the type of questions that can be answered by SkELL data and eventually may become able to query it themselves, as suggested by Kilgariff et al. (2015).

Finally, the metalinguistic information posed some problems in classifying the collocations of the noun *sculpture* or recognising miscategorised items in the SkELL. This means that it is either the best to avoid the metalinguistic information with students who are not (future) language experts or to use the problems that arise from such obstacles as a starting point for explanation and clarification of the problem areas.

## 6 *CONCLUSION*

Although fully aware of the fact that the primary idea of DDL is to consult corpora to find answers to language problems that arise, we have taken the reverse approach in designing the proposed set of tasks. Within the given learning and teaching reality this approach has adapted DDL to the needs of the target users and to the course curriculum.

The design of this study started from the premise that DDL can be quite a revolutionary practice if attempted in its original, most learner-led format,

since it requires a high level of language awareness, motivation, and some IT skills in handling the required software that may not be found in many language learning communities. In such learning environments we have the option of choosing from the "softer part" of the continuum of DDL exercises which makes the approach look less revolutionary and this was the path that was explored in this study.

Overall, the students did not seem to have any particular problems with dealing with the exercises, although the DDL vocabulary file was assigned as an out-of-class activity in order to save class time. The students' response also showed that they reacted in a more confident way when dealing with the guided exercises, while they found the high level of autonomy in the self-study exercises confusing. This supports Gilquin and Granger's (2010) claim that those assignments that are carried out at home may have to be accompanied by more detailed instructions.

The outcomes of the students' vocabulary files reflect a range of pedagogical issues that may be discussed in relation to any other language teaching and learning activity, such as e.g. motivation, learning autonomy, learning styles, learning strategies, or group work collaboration. These were, however, not studied on this occasion but may be the subject of future studies.

The ease of preparing simple exercises as illustrated by the analysed set is encouraging, and it might be used as a way of motivating teachers who have not used it yet to try it themselves. The tools used in this study can also be highly recommended.

To conclude, this paper focused on the analysis and the qualitative explanation of the results of students' engagement in a specific type of DDL. Due to the large number of exercises and submitted assignments the analysis was limited to studying the vocabulary part of the DDL vocabulary file. The analysis of the content-oriented part of the assignment is in preparation, as well as the comparison of the pre-test and the post-test results which will allow a combination of a quantitative and qualitative approaches.

## *REFERENCES*


Azzaro, G. (2012) Phrasal verbs through DDL. *Journal of theories and research in education* 7 (2), 1-24.

Boulton, A. (2009) Testing the limits of data-driven learning: language proficiency and training. *ReCALL* 21 (1), 37-51.

Boulton, A. (2010) Data-driven learning: On paper, in practice. In Harris, T. and Moreno Jaén, M. (eds) *Corpus linguistics in language teaching*. Bern: Peter Lang, 17-52.

Boulton, A. (2011) Data-driven learning: the perpetual enigma. In Goźdź-Roszkowski, S. (ed) *Explorations across Languages and Corpora*. Frankfurt: Peter Lang, 563-580.

Boulton, A. (2012) What data for data-driven learning? *The EUROCALL Review. Proceedings of the EUROCALL 2011 Conference* 20, 23-27. Available at http://eurocall.webs.upv.es/documentos/newsletter/papers_20(1)/07_boulton.pdf (13.03. 2016.)

Boulton, A. and Tyne, H. (2014) Des documents authentiques aux corpus: démarches pour l'apprentissage des langues. Paris: Didier Coll.

Chambers, A. (2005) Integrating corpus consultation in language studies. *Language Learning & Technology* 9 (2), 111-12. Available at http://llt.msu.edu/vol9num2/chambers/ (16. 08. 2016.)

Chambers, A. and Kelly, V. (2004) Semi-Specialized Corpora of Written French as a Resource in Language Teaching and Learning. *Teanga* 21, 114-133.

Cobb, T. (1997) Is there any measurable learning from hands-on concordancing? *System* 25 (3), 301-315. Available at http://www.lextutor.ca/cv/Hands_on.html (12.03. 2016.)

Cobb, T. (1999) Breadth and Depth of Lexical Acquisition with Hands-on Concordancing. *Computer Assisted Language Learning* 12 (4), 345-360. Available at http://www.lextutor.ca/cv/breadth_depth_99.pdf (16.08. 2016.)

Frankenberg-Garcia, A. (2005) A peek into what today's language learners as researchers actually do. *International Journal of Lexicography* 18 (3), 335-355.

Gilquin, G. and Granger, S. (2010) How can data-driven learning be used in language teaching? In O'Keeffe, A. and McCarthy, M. *The Routledge Handbook of Corpus Linguistics*. Routledge: London, 359-370.

Huang, Z. (2014) The effects of paper based DDL on the acquisition of lexico grammatical patterns in L2 writing. *ReCALL* 26 (2), 163-183.

Hunston, S. (2002) *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hunston, S. and Francis, G. (2000) *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.

Johns, T. (1986) Micro-Concord: A Language Learner's Research Tool. *System* 14 (2), 151-162.

Kabalin Borenić, V., Marinov, S., and Mencer Salluzzo, M. (2013) Special and spotless? Or maybe not? *Language myths in tourism. Acta Turistica* (25). Zagreb: University of Zagreb, Faculty of Business and Economics, 119-216.

Kerr, B. (2013) Grammatical description and classroom application. Theory and practice in data-driven learning. *Bulletin VALS-ASLA* 97, 17-39.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. and Suchomel, V. (2014) The Sketch Engine: ten years on. *Lexicography: Journal of ASIALEX* 1 (1), 7–36.

Kilgarriff, A., Marcowitz, F., Smith, S., and Thomas, J. (2015) Corpora and Language Learning with the Sketch Engine and SKELL. *La Revue Française de Linguistique Appliquée* XX-1, Pub. Linguistiques, 61-80.

Laufer, B. and Nation, I. S. P. (2011) Vocabulary. In Gass, S. M. and Mackey, A. (eds) *The Routledge Handbook of Second Language Acquisition*. London and New York: Routledge, 163-176.

Leech, G. (1997) Teaching and language corpora: A convergence. In Wichmann, A., Fligelstone, S., McEnery, T. and Knowles, G. (eds) *Teaching and language corpora*. New York: Addison Wesley Longman, 1-23.

Lewis, M. (1993) *The Lexical Approach: The State of ELT and a Way Forward*. Hove, UK: Language Teaching Publications.

McEnery, T. and Wilson, A. (2001) *Corpus linguistics* (2nd edition). Edinburgh: Edinburgh University Press.

Mukherjee, J. (2006) Corpus linguistics and language pedagogy: The state of the art – and beyond. In Braun, S., Kohn, K. and Mukherjee, J. (eds) *Corpus technology and language pedagogy: New resources, new tools, new methods* [Special issue]. English Corpus Linguistics, 3. Frankfurt: Peter Lang, 5-24.

Nation, I. S. P. (1990) *Teaching and learning vocabulary*. Boston: Newbury.

Nesselhauf, N. (2004) Learner corpora and their potential for language teaching. In Sinclair, J. (ed) *How to Use Corpora in Language Teaching*. Amsterdam & Philadelphia: Benjamins, 125-52.

O'Sullivan, Í. (2007) Enhancing a process-oriented approach to literacy and language learning: the role of corpus consultation literacy. *ReCALL* 19 (3), 269-286.

Sardinha, T. B. (2012) Lexicogrammar. *The Encyclopedia of Applied Linguistics*. Blackwell Publishing ltd.

Scott, J. and Pavlenko, A. (2008) *Crosslinguistic influence in language and cognition*. New York and London: Routledge.

Thomas, J. (2015) *Discovering English with the Sketch Engine*. Brno: Laptop Languages.

Todd, R. (2001) Induction from self-selected concordances and self-correction. *System* 29 (1), 91-102.

Tribble, C. (1997) Improvising corpora for ELT: Quick and dirty ways of developing corpora for language teaching. In Lewandowska-Tomaszcyk, B. and Melia, P. L. (eds) *Practical applications in language corpora*. Lodz, Poland: University Press, 106-117.

Willis, D. (1990) *The Lexical Syllabus*. London: Collins.

# UČENJE VOKABULARA PRIMJENOM KORPUSNO UTEMELJENIH ZADATAKA U NASTAVI JEZIKA STRUKE: SLUČAJ UPORABE KORPUSNO UTEMELJENE VJEŽBENICE

U ovome se radu razmatra ideja uvođenja samostalnoga studentskog rada s korpusno utemeljenim zadatcima (engl. *data-driven learning*) u redovitu praksu kolegija Engleski jezik u turizmu na Ekonomskom fakultetu u Splitu. U radu prvo objašnjavamo sadržaj i način oblikovanja korpusno utemeljene vježbenice. Zatim slijedi analiza odgovora studenata na zadane vježbe. S obzirom na složenu prirodu cjelokupne vježbenice ovaj se rad zadržava samo na vježbama leksika, a sadržajno se motivirani i orijentirani dio vježbenice analizira u zasebnom radu. Studenti nisu imali osobitih problema s izradom zadataka koji bi se mogli pripisati upravo takvoj vrsti zadataka. Ishodi njihovih zadataka odražavaju niz pedagoških pitanja koja se mogu vezati za bilo koju drugu aktivnost u nastavi stranoga jezika, kao što su na primjer motivacija, autonomija u učenju, stil i strategije učenja ili skupni rad i suradničko učenje.

*Ključne riječi: korpus u nastavi stranoga jezika, leksik, Engleski jezik u turizmu, korpusno utemeljena vježbenica.*