TRAINING ESP STUDENTS IN CORPUS USE - CHALLENGES OF USING CORPUS-BASED EXERCISES WITH STUDENTS OF NON-PHILOLOGICAL STUDIES

by **Sanja Marinov**

University of Split

Split, Croatia

smarinov @ efst.hr

Abstract

This paper focuses on planning a series of activities to train learners of undergraduate, non-philological studies in using a small specialised ad hoc corpus and the results they achieved in doing them. The procedure discussed in this paper is a part of a larger project which investigates the possibility of using a small specialised corpus with ESP learners in improving their written texts. The planned activities are carefully designed to address particular issues either related to the technical aspect of using and reading corpora or to some of the language issues that are (relatively) easily observable in corpus data. The set of devised exercises is referred to as a training programme meant to prepare and guide learners towards the status of more independent corpus users. The training was carried out as a part of an obligatory language course taken by the second year university students of tourism.

The paper analyses the learners' success in dealing with each particular exercise as well as the major challenges encountered in the process. Conclusions are drawn and recommendations for future applications and researches are made.

Key words: small specialised corpus, training, ESP, corpus-based tasks, corpus-derived exercises

1. Introduction

This study analyses the process of training learners in corpus use, the main aim of which was to prepare them technically for handling the corpus and the concordance programme since they had never used it before, but also to address some particularly important language issues that can easily be observed and analysed using the given tools. Namely, raising learners' awareness of how language actually works and encouraging them to start looking at it through a slightly different looking glass seems an important step towards their autonomy in corpus use and thus contributes to making them more autonomous language learners in general.

If learners are to become autonomous corpus users or if they are only to understand the benefits of such a tool, they need to be trained through a series of guided activities. The

need for training in corpus consultation has been recognized and advocated by a number of scholars (Cheng et al., 2003; Chambers, 2005; O'Sullivan & Chambers, 2006).

2. Literature review

2.1. Data Driven Learning (DDL)

The approach to learning which makes use of corpora in the classroom is usually referred to as Data-Driven Learning (DDL) and is considered to be one of the branches of Computer-Assisted Language Learning (CALL). Nonetheless, the data-driven approach can be any attempt at raising learners' awareness of the target language by teaching them to discover the behavior of language items for themselves (Gabrielatos, 2005). Therefore, today we distinguish between text-based DDL (James and Garrett, 1991) and corpus-derived and corpus-based DDL as fathered by Tim Johns (1988).

In corpus-derived and corpus-based DDL corpora can be brought to the classroom in two different ways: either the soft or hard version (Leech, 1997). In the soft version it is the teacher who has access to corpora, the skill of using it and the software. He/she designs the tasks and takes them into classroom as printouts (Tribble and Jones, 1990; Bernardini, 2004). On the other hand, in the hard version learners, too, need to be equipped with computers, corpora and software and, above all, the skill of using them. The tasks can be either designed by the teacher (Tognini-Bonelli, 2001) or by the learners themselves (Bernardini, 2002).

The number and variety of available corpora has enabled experimenting with DDL in a number of environments. It is true that the method is often used at the university level, in particular with the students of languages or postgraduate students who are expected to be better motivated, better equipped with the research skills, and more independent and therefore yield higher quality results. For example, Paul Thompson (2004) incorporated the module on the use of corpora in applied linguistics in the graduate programme of corpus linguistics, Davies (2000) used corpora of historical and dialectal texts in a module of Spanish linguistics, and Chambers used the corpus linguistics procedures to improve her students' language skills in one of the modules of an undergraduate language programme (Chambers, 2005). The three examples may be partly illustrative of Fligelstone's (1993) three categories of corpus exploitation: teaching about corpus linguistics (Thompson, 2004), exploiting corpus resources in order to teach (Davies, 2000), and teaching students how to exploit corpus data (Chambers,

2005). However, it is very likely that any hands-on practice will be a combination of some of the three categories.

Doubts about the possibility of applying DDL at lower levels of language proficiency and in different learning environments have been challenged by a number of studies. For example, Hadley (n/a) has experimented with the first-year elementary students of the Faculty of Economics, University of Niigata (Japan). The main goal was to teach some aspects of grammar, and the results obtained dispelled the initial doubt of the success of the project.

The empirical research is not limited to English only. Kennedy and Miceli (2001, 2010) have taught their intermediate undergraduate learners of Italian to exploit a small corpus of Italian. Their analysis of learners' attitudes towards corpus use during and after the "apprenticeship", as they call the training in corpus consultation, has resulted in a number of principles as recommendations for future work. Major points to conclude are readiness to accept uncertainty of finding a satisfactory answer or recognising that the development of skills is a long-term process (Kennedy and Miceli, 2010).

2.2. Corpora uses in the language classroom

So far corpora have been used in language instruction to address a variety of language issues. This paragraph summarizes some of them but is by no means exhaustive. The list of potential applications is presented in Figure 1 and classified according to the authors who advocate them.

Tribble and Jones (1990)
- study homonyms and synonyms (also Barlow, 1992)
- lexical inference (also Gabrielatos, 2005)
- study the word grammar as well as the general grammar rules

Levy (1990)
- discover and learn phrases or typical collocations (also Willis, 1998)
- study LSP terminology
- study derivatives

Barlow (1992)
- analyse usage of lexical items
- analyse the language used in particular books and textbooks
- compare language usage (e.g. native/non-native speakers)
- study word order

Leech (1994)
- revision and critical examination of traditional ELT rules (also Gabrielatos, 2005)

Willis and Willis (1996)
- consciousness raising (increase awareness and sensitivity to language)

Cobb (1997)
- study new lexis through multiple contexts

Willis (1998)
- study potentially different meanings of some frequent words
- compare the characteristics of written and spoken discourse
- compare frequency of some language items in differente types of texts

Gabrielatos (2005)
- practice both intensive and extensive reading

Marinov (2010, 2011)
- meet some of the principles of the Lexical Approach to learning / teaching English

Figure 1: Ways of exploiting corpora in DDL.

One of the main issues addressed in the research on teaching a language with corpora is the issue of their size. Although it is generally accepted that the bigger the corpus, the more reliable it is for making generalisations about the language use (Sinclair, 1991), it is also believed that a smaller corpus is generally more useful for classroom purposes. A large corpus can result in too large a number of data, which makes it difficult to handle (Tribble, 1997). The question is how small a small corpus can be. Tribble and Jones (1990) consider a 50,000 word corpus to be useful in teaching. Aston's idea of a small corpus is within the range of 20,000 to 200,000 words (Aston, 1997). Corpora that are compiled quickly to serve a particular classroom purpose or to address a particular language issue rather than to represent

the whole language or language variation do not always meet all the strict design criteria that are considered in compiling large corpora and are referred to as ad hoc corpora (Gavioli, 2005). Bowker and Pearson (2002) are less specific about the size, claiming that corpora ranging from several thousand to several hundred thousand words are useful for studying the language for specific purposes. Aston (1997), too, makes reference to the usefulness of smaller corpora for learners who need to master the characteristics of a specific genre or type of text as is the case with teaching LSP.

Apart from the adequacy of content Aston also emphasises the advantages of methods a small corpus allows: they are more fully analysable, they are easier to become familiar with, they are easier to interpret, they are easier to construct, they are easier to reconstruct, they are more clearly patterned, their limits are clearer.

3. The present study

3.1. The aim of the study

This study is primarily a contribution to the response to the expressed need to define corpus-consultation literacy and ways of achieving it (O'Sullivan, 2007). Empirical studies have shown that some learners appreciate the benefits of corpus consultation while others find it too time-consuming or too demanding (Chambers & O'Sullivan, 2004; Chambers, 2005). This study has explored the possibilities of creating carefully targeted corpus-based tasks using a small specialised ad hoc corpus. It has faced the learners with an extensive range of corpus-based exercises and has aimed to determine which are more and which are less demanding and how much guidance learners need when tackling them.

3.2. Participants and resources

The project participants were the students of the second year of the Faculty of Economics, University of Split with the course of study in Tourism. In their third semester they take the last of the three obligatory courses in English - English in Tourism 3. There were 42 students enrolled and they all took part in the project since it was planned as an integral part of the course.

The corpus (henceforth referred to as "the Mediterranean corpus") used in the project consisted of 450,000 tokens and is therefore classified as a small corpus. It can be said to combine one register and one genre because it included only the texts from the area of

tourism, and more specifically those of the tour guides of the Mediterranean countries. The main criterion in compiling the corpus was its purpose because in the next step of the project the students would use it to improve their texts prepared for inclusion into holiday brochures and tour guides (Marinov, 2011).

The software used for corpus consultation was Intex (http://intex.univ-fcomte.fr; http://www.nyu.edu/pages/linguistics/intex/) due to its two important advantages: it is a freeware and therefore downloadable to as many computers as you need free of charge and it has an automated POS tagger incorporated, which means it automatically assigns words to their part of speech, which enables better targeted research. Intex also allows automatic lemmatization of the corpus.

Finally, although the IT resources and support are important prerequisites for using corpora in language teaching it is the investment in the users of corpora, the learners, and teachers, that poses the greatest challenge for language teaching (Kennedy and Miceli, 2001).

3.3. Design and procedure

After the Mediterranean corpus was compiled, the class lecturer analysed it first in order to design a set of tasks for students. The process was driven by four main ideas:

1. both hard and soft approach would be used,
2. exercises would address both technical issues of corpus consultation as well as language research issues,
3. simpler exercises would be combined with more challenging ones,
4. convergent tasks, in which anyone working with the same data should arrive at the same outcome, would be combined with divergent tasks, where different people working on the same data autonomously are expected to arrive at different outcomes (Bernardini, 2004).

The training was fully integrated into the curriculum and the time allocated for training was randomly set to 6 lessons (45 minutes each). Some of the activities were done in class (especially the hard version ones) while some others, the type of exercise permitting, were set for homework to save class time. The distribution classwork-homework was only partly planned in advanced and mostly decided in the process, depending on the problems and obstacles arising.

A fully equipped IT laboratory was available for class sessions throughout the process. Each learner worked on his/her on computer and had the software (Intex) installed and the corpus uploaded.

The following paragraphs describe the tasks prepared for each step and explain the rationale behind them.

Step one – Sorting concordance

The aim of this step is teaching the possibility of sorting concordance according to the left and the right context. The learners were asked to analyse the left and the right context of the node word *rooms*, one of the highly frequent words in the Mediterranean corpus. For each of the two they also had to note down ten collocations to include into their personal vocabulary files. Moreover, the learners were asked to explain what criteria they based their selection on.

The task was triggered by the problem of inadequate ways of noting down various lexical units at all levels of language acquisition. Although it is nowadays customary to note down different types of lexical units apart from the word itself, the lack of system and organisation in the process is still present, which greatly diminishes the efficiency. As Lewis (1997) suggests, we have to find out useful and accessible ways of organising the lexis if we want to make it available to learners.

Step two – Finding a lemma

The second step familiarises the learners with the possibility of finding the whole lemma of a word. Although the possibility of lemmatization may not be as interesting for English as it is for some highly inflectional languages, it may still find its applications. In this step two possible uses are illustrated with two separate tasks.

Task 1. The first task required the learners to generate (or revise) the rule about the *-y* ending in front of an inflection. The proposed words were a mixture of nouns and verbs: *valley, city, emergency, journey, currency, scenery; enjoy, ply, fly, carry, play, try*.

Task 2. The second task required the students to look for the lemma of a verb in order to check all its forms. The generated concordance was also supposed to show that some forms of a verb are more frequent than others in a certain type of genre or text. The students were asked to comment on which is the most frequent form of the verb *say* in the given corpus, and then to isolate the most frequent pattern in which this form appears.

Step three – Extending the context

Step three illustrates the possibility of exploiting small specialised corpora for traditional, linear reading. This type of search was made possible because of the familiarity with the exact structure of the texts included in the corpus. In particular, the task was triggered by the idea that the lexeme *sight* is most frequently found as a subtitle of a paragraph. This means that by double clicking on one of the concordance lines one can access a paragraph on the topic of sights of a particular place. Independently of which country the text is about, it is highly likely that such carefully targeted reading will offer a series of linguistic solutions but also ideas to talk about the topic.

The task was illustrated by an example where the principle was explained. Students then had to choose a paragraph themselves, read it and select some useful vocabulary. They could choose from the same group of *sight* paragraphs or try to find another equally interesting one on a different topic. The choice of language items was left completely to them.

Step four – Comparing synonyms

In step four the learners were shown how to observe two different lexemes in the same concordance and it was suggested this could be applied to analysing the differences between synonyms. Here the assumption was that if more attention is paid to meaning relationships, the possibility of making good lexical choices based on these relationships will be increased. One of these relationships is synonymy, defined as coexistence in a language system of two or more words with the same or similar content, but a different expression (Škiljan, 1987). Since absolute synonyms are very rare, learners should be taught how to use particular synonyms offered by dictionaries or thesaurus (Partington, 1998). In this particular task the learners were asked to analyse the difference in use and meaning between *crowded* and *packed*.

Step five – Using regular expressions

This step makes use of and emphasises the advantages offered by a POS tagged corpus. After a brief introduction to the most important symbols used by Intex the learners were given the following tasks.

Task 1. This task addressed almost the same issue as Step one but this time the learners were able to narrow down the search using a newly acquired symbol. The expression "<A>restaurant" is suggested to generate the concordance of the node word *restaurant* preceded by adjectives and students were asked to isolate ten of these adjectives according to their own choice.

Task 2. In the second part of step five the soft approach was used. A corpus-derived matching exercise addresses the problem of quantifiers (quantitatives and collectives) as a specific form of a noun phrase and an element of the grammar of structure (Willis, 2005). The same exercise underlines the existence of the so-called frames as one of the ways of generating a lexical phrase. Those singled out by this exercise fit into the frame "a + quantitative + of + N".

In the editing phase 18 different examples were carefully selected trying to achieve an unambiguous situation where each left context can be matched with only one right context, i.e. there was only one solution to the task (Table 1).

Table 1: The matching exercise on quantitatives.

1	chalet restaurant serving raclette - a block of	a	melted cheese, usually eaten with potatoes and
2	subsequently rebuilt. Inside is a collection of	b	100 buses, a comprehensive website
3	Teixeira 37; d from ?30) No-frills with a range of	c	armour and military gear, but even better a
4	used to be 10 separate houses and a stretch of	d	effort into making sure you won't want to leave
5	second to none. Students form nearly a quarter of	e	five tiny fishing villages strewn along
6	modern city. Earthquakes and sacking by a series of	f	Guinness or Murphy's (both €7), Mike's is
7	&Delos Mediterranean Europe boasts a wealth of	g	homely charm, and the shady rear courtyard is
8	-in desk, where it should be treated as a piece of	h	invaders have left little standing of the
9	For those who can't go too long without a pint of	i	luggage. Check all this with the airline well

10	pressos.pt, in Portuguese), which has a fleet of	j	oven-hot pizza, sold by weight, can cost as
11	SIGHTS & ACTIVITIES Provided you have a pair of	k	prices, although you need a mountaineering
12	for under €3, or a pizza al taglio, where a slab of	l	street, now artfully cobbled together to
13	lose yourself in Cordoba's old quarter, a maze of	m	sturdy walking shoes and don't balk at hills
14	are pretty high, and the best places put a lot of	n	the population. Orientation & Informat
15	and during July and August Greece's seas are a mass of	o	Unesco World Heritage Sites and you need only
16	motifs and spotless rooms give the Garden a dash of	p	wake and wash. The ferries come in all shapes
17	ists in summer. It's fine if you want a couple of	q	weeks of pure hedonism, but there are nicer
18	Discover the CinqueTerre (pl89), a cluster of	r	winding, cobbled alleyways, pretty plazas

After completing the exercise the students were asked to write a regular expression to get the correct answers from the corpus.

Step six - Collocation

Although some previous exercises were based on the collocational potential of particular words the two tasks in this step were designed to explore the concept slightly more systematically. Collocation is co-occurrence of two or more words in the immediate context of each other in the text (Sinclair, 1991). It is one of the most powerful principles in teaching a foreign language, in particular in the area of organising and noting down new elements since it implies entering new words with their co-text (Lewis, 1993).

Task 1. In task one the learners were asked to generate the profile of the lemma *beach* and present if graphically. This is a highly frequent word in the given corpus (499 tokens) and therefore presented a rich source of material. No strict guidelines for making the profile were given.

Task 2. The language of tourism abounds in adjectives. In front of a noun there may be more adjectives in a sequence. Grammars teach us about the order of adjectives in such cases since it is by no means arbitrary and left to the choice of individual speakers.

This task was a combination of a corpus-derived gap-fill and jumbled words exercise. 13 concordance lines having a sequence of three adjectives in front of a noun were selected. The adjectives in form of each noun were deleted and offered in random order in the right column on the same line (Table 2). The students were asked to establish the order of adjectives

preceding the nouns. Having completed the exercise peer evaluation was carried out using the corpus as the source of correct solutions. The regular expression "<A><A><A><N>" was used. Afterwards, the learners were asked to establish the rule on order of adjectives.

Table 2: Sequence of adjectives exercise.

s also vinho verde country, that wholly _____ wine, while its capital Porto is	green young addictive
7 00. Roman Forum & Palatine Hill The _____ political and religious centr	Roman ancient commercial
you can wander up into D'Alt Vila, the _____ town, with its upmarket restaura	walled old atmospheric
and thousands beheaded. In the chaos a _____ general named Napoleon Bonaparte	Corsican dashing young
urkey. SIGHTS & ACTIVITIES Mytilini's _____ museum (fsr) 2251	neoclassical archaeological excellent
; s/d €53/67.50; *?) Hotel Tourist is a_____ hotel, which has comfortable room	old fine neoclassical
capital, Athens. Drink and dine in the _____ town of Nafplio (p290), one of Gr	old Venetian gorgeous
n-Sat) Just north of the centre is this _____ cavern/tavern, complete with vaul	medieval great little
not as easily besmirched. The first and _____ writer was Homer, author of Ilia	Greek greatest ancient
built by the Knights of St John and the _____ town in Europe. Head to th	medieval inhabited largest
8.30am-1hr before sunset) is one of the _____ sights in the world. Despite	archaeological Greek major
along the rugged Adriatic coast. The _____ belt at the foot of the Dinaric	coastal Croatian narrow
y's charm. Cinema Despite a number of _____ films, such as Robert Beni	contemporary Italian popular

Step seven - colligation

Colligation is frequency of occurrence of a word with a particular part of speech, i.e. it is a collocation of a lexical and a grammatical element (Partington, 1998). For Croatian speakers of English colligational patterns are frequent sources of error. Although colligation is most frequently observed and noted in the context of verb patterns (e.g. preposition following a

verb, verb followed by a gerund or infinitive, …) it may be equally valuable in describing other parts of speech (see Step 7- Task 2).

Task 1. The aim of this task is raising awareness of some verb and preposition partnerships. This task is prepared in the soft version as a matching exercise in which the part of the line on the left finishes in a verb, and the part on the right begins with a preposition (Table 3).

Table 3: Matching exercise for colligational pattern verb + preposition.

1	the extradition to the Hague of Croatian generals accused	a	at protecting the Mediterranean monk seal. It is a welcome innov
2	ou go, you will find easy-going, tolerant people, accustomed	b	to Jupiter, Juno and Minerva; and the Piazzale delle Corpora
3	Park of Alonnisos, encompassing seven islands, is aimed	c	for; specialist vegetarian restaurants are few and far between
4	tainment Nightclubs in Rabat - some of which are attached	d	to the upscale hotels - normally charge at least Dh60 for ent
5	k Bulvan 3) If you are planning a picnic or just bored	e	of war crimes remains highly controversial. RELIGION Croats
6	Vegetarians and vegans are not particularly well catered	f	with beer, pop into this great wine store and choose from the ra
7	et, with 22 entrances and 4500 shops, permanently crammed	g	to as piccola Roma ('little Rome'). Leave your preconcepti
8	ilt by Agrippa; the forum and Capi-tolium temple, dedicated	h	to welcoming visitors and proud of the country they fought
9	tolia and part of Thrace, was born. Atattirk then embarked	i	after one of its inhabitants - the ris (lynx). Dense forests
10	Thus an American citizen living in Madrid is not entitled	j	for its postcard sunsets, is less hectic than Fira and a must-vi
11	, the flawless village of Oia (pronounced ee-ah), famed	k	for its high-profile gay scene. Although it's not quite so he
12	ak National Park near Delnice, east of Rijeka, is named	l	to a VAT rebate on items bought in Paris, while an EU passport
13	d his tragedy. In centuries past, Verona was even referred	m	on a rapid modernisation programme: establishing a secular de
14	of the Mediterranean's most fashionable resorts, renowned	n	with punters, promoters, pickpockets and policemen. It's a fun

Task two. While exploring the corpus high frequency and very easily observable patterns of the lexeme *worth* were noticed. There are exactly 150 lines in which the lexeme occurs. Since the exercise was approached in the hard version, the learners were guided by a number of questions to make it less challenging. The idea of turning this step into a more carefully

guided exercise was a change to the original idea, based on the experience gained in the course of training (Table 4).

Table 4: Guiding questions provided for analysing the 150 concordance lines of the lexeme *worth*.

1	Make a concordance of the word *worth*.
2	Note down its frequency. Does it make it worth studying?
3	Sort the concordance to the right and answer the following questions:
a	What part of speech is it usually followed by (noun, verb, pronoun…)?
b	What form does the word following it take (singular, plural, article or no article, verb form)?

Step eight – Word class

The linguistic concept of word class is of central importance to a language system and therefore worth considering in the process of language teaching and acquisition. It creates the connection between grammar and lexis and makes them inseparable. Its importance is most evident in the grammar of patterns since the words we find in a particular pattern usually have a similar meaning and function, i.e. the words we can find in the same pattern belong to the same word class (Willis, 2005). Accordingly, the aim of this task was to raise learners' awareness of such a relationship between meaning and pattern.

The task observes the pattern "verb + object + to infinitive" obtained with the regular expression "<V>(him+her+you+them)to<V:W>". 32 lines are retrieved and used in the task without any further editing (Table 5). The learners were first asked to complete the lines with the missing verbs. This was done in order to encourage intensive reading. The answers were checked against the corpus and the repeated pattern was established. Finally, the students were asked to group the verbs that appear in the established pattern according to their meaning.

Table 5: Word class gap-fill exercise.

Complete the concordance lines with the following verbs:
avoid / believe / call / construct / drop / gasp / get / have / have / pause / pay / prove / receive / remember / see / sleep / stay / transform / travel / waive / watch / withdraw /
When you are done check your results against the corpus. The pattern used to obtain the lines from your task was the following: **<V>(him+her+you+them)to<V:W>**.
Does a certain pattern emerge from the examples? Do the verbs that trigger the pattern have something in

common? Could you classify them according to a similarity in meaning?

get, Europcar and Hertz have the big advantage of offering one way rentals that **allow you to** _____ the car off at any one of their many stations in Croatia

e budget, you should be able to manage on between €70 and €80 a day. This would**allow you to** _____ in a cheap hotel or pension, eat in a modest restaurant

 tickets or car hire) without you having to carry great wads of cash. They also **allow you to** _____ cash at selected banks or from the many ATMs that are l

r so. Some of the more expensive companies charge extra for a video system that **allows you to** _____ Croatian soap operas during your trip...unmissable. If

 If you have to claim later make sure you keep all documentation. Some policies **ask you to** _____ back (reverse charges) to a centre in your home country where

r   LEGAL MATTERS   French police have wide powers of search and seizure, and can **ask you to** _____ your identity at any time. Foreigners must be able to prove t

or more. She sat at the mouth of a chasm that emitted fumes. These she inhaled, **causing her to**_____, writhe and shudder in divine frenzy. After sacrificing

ber that Netscape and Internet Explorer both have mail modules). Most ISPs also **enable you to** _____ your emails through its website, which only requires y

offices can make arrangements for an official guide (Dh250 per day), which will **enable you to** _____ and experience much more in the medina. The guides do k

i; by Frederic Damgaard, a Danish gallerist promotes the work of local artists, **encouraging them to**_____ their traditona; thuya techniques into more saleab

eliable of the lines, HML also officially accepts Eurail and Inter-Rail passes, **entitling you to** _____ free in deck class (paying a €15 supplement in July an

 home urgently, a free ticket home is exceedingly unlikely as the embassy would **expect you to** _____ insurance. If you have had all your money and documents

metimes 50% or even 100% more for a one-night stay, although you may be able to **get them to** _____ the surcharge if you arrive in the low season. Prices for pr

 Santa Maria della Neve and built by Pope Liberius in 352 after the Virgin Mary **instructed him to**_____ a church on the spot where the next snow fell. Its m

ut its long boulevards, open squares and historic Barrio Humedo (Damp District) **invite you to** _____ for a day or two.   Sights & Activities   Leon'

ople about the circumstances surrounding the 11 March train bombings in Madrid, **leading them to**_____ ETA was the author of the attack when all along evide

 Galaxy Internet Cafe (g 038-213 5637; Rruga Taulantia; 200 leke).   Let waves **lull you to** _____at Hotel Besani (g 068-203 5781; Skembi i Kavajes; s US$2

uberges de la Jeunesse (LFAJ; far. 014416 78 78; www ,auberges-de-jeunesse.com) **need you to** _____ or purchase a Hostelling International (HI) card or a

 health expenditures. The former option is generally preferable, as it does not **require you to** _____ out

> of pocket in a foreign country.   WARNING Codein
>
> ost ISPs also enable you to receive your emails through its website, which only **requires you to** _____
>
> your account name and password. It pays to become fam
>
> cia still boasts many ancient buildings - do not listen to locals elsewhere who **tell you to** _____ the
>
> place completely. If nothing else, it's worth a stop
>
> from selling carpets to full-scale rip-offs; it's probably easier chatting than **telling them to** _____ stuffed
>
> straight off, but do not let yourself be ste

3.4. Results and findings

Step one – Sorting concordance

Although the hard version was applied, no particular problems in carrying out the task were reported. The learners approached the task in many different ways with equally various outcomes. Generally, they did not appear to have a highly efficient strategy of noting down the observed vocabulary. Only rarely did they show a high level of observation, analysis and generalisation skills, e.g. one student noticed a higher frequency of the adjective *decorated* but also noted down a series of adverbs that accompany the given adjective or generalised the observed data in form of a pattern "are + adjective", "can be + adjective". Similar generalisations were noticed in only 6 out of the total of 42 learners.

Finally, the students pointed out five different criteria that guided them in the choice of vocabulary. These are shown in Table 6 along with the respective frequencies.

Table 6: Criteria applied by students in the selection of collocations with lexeme *rooms.*

Criteria	Example	Number of students	Percentage
frequency	airy, comfortable, decorated	42	100%
unfamiliar word	cramped, immaculate, impeccable, no-frills, well-appointed	21	50%
new collocational potential	cheerful	15	35.71%
useful word	overlooking	3	7.14%
sound	stylish	1	0.42%

Table 6 clearly shows that the criterion of frequency is in the first place and was applied by everyone. Half of the learners selected unknown vocabulary and looked up their meanings. Whenever they approached the teacher to ask for the meaning, they were directed to generate a new concordance for that particular word and try to guess the meaning from the context. Results also show that part of the students avoided dealing with new vocabulary and "recycled" the familiar lexis. Another valuable criterion is the one where the students became

aware of the new collocational potential of an already familiar word. Finally, a small number of students were guided by the criteria of usefulness and pleasure.

Step two – Finding a lemma

As regards task 1, all learners successfully noted down plurals of the nouns as well as all forms of the verbs. Also they concluded that –*y* changes into –*i* in front of the ending when –*y* is preceded by a consonant, while this does not happen when it is preceded by a vowel. However, only 9 students mentioned that this is not the case in front of the –*ing* ending when all these verbs keep their –*y* (*plying, trying, carrying*). It cannot be said that all students were autonomous in their analysis and they certainly did not work at the same pace. It was noticed that they frequently exchanged ideas in the process of analysis.

This exercise triggered another search, that of the verb *ply*, which was an unknown item. Teacher guidance was required to carry out the analysis. The learners found out that various means of transport are mentioned in the left context (*vessels, hydrofoil service, bus*) and in the right context the location where the action of the verb takes place is often indicated (*the Mediterranean, between .... and...., airport bus route, coastal road, main street, major international routes*). Although it was hard to establish the exact meaning it was concluded that it is something about how means of transport move.

When reflecting on task 2, it can be stated that having generated the concordance of the lemma *say* it was not difficult for students to notice that the most frequent form in this corpus was *said*. Figure 2 shows all the lines in which the most frequent form of *say* (*said*) occurs. In this particular concordance the searched pattern is also easily noticeable and no students had problems recognising it. However, a very small number of students (only 7 of them) tried/were able to express this pattern in a general form as *is said to be + past participle* and *is said to have + part participle*.
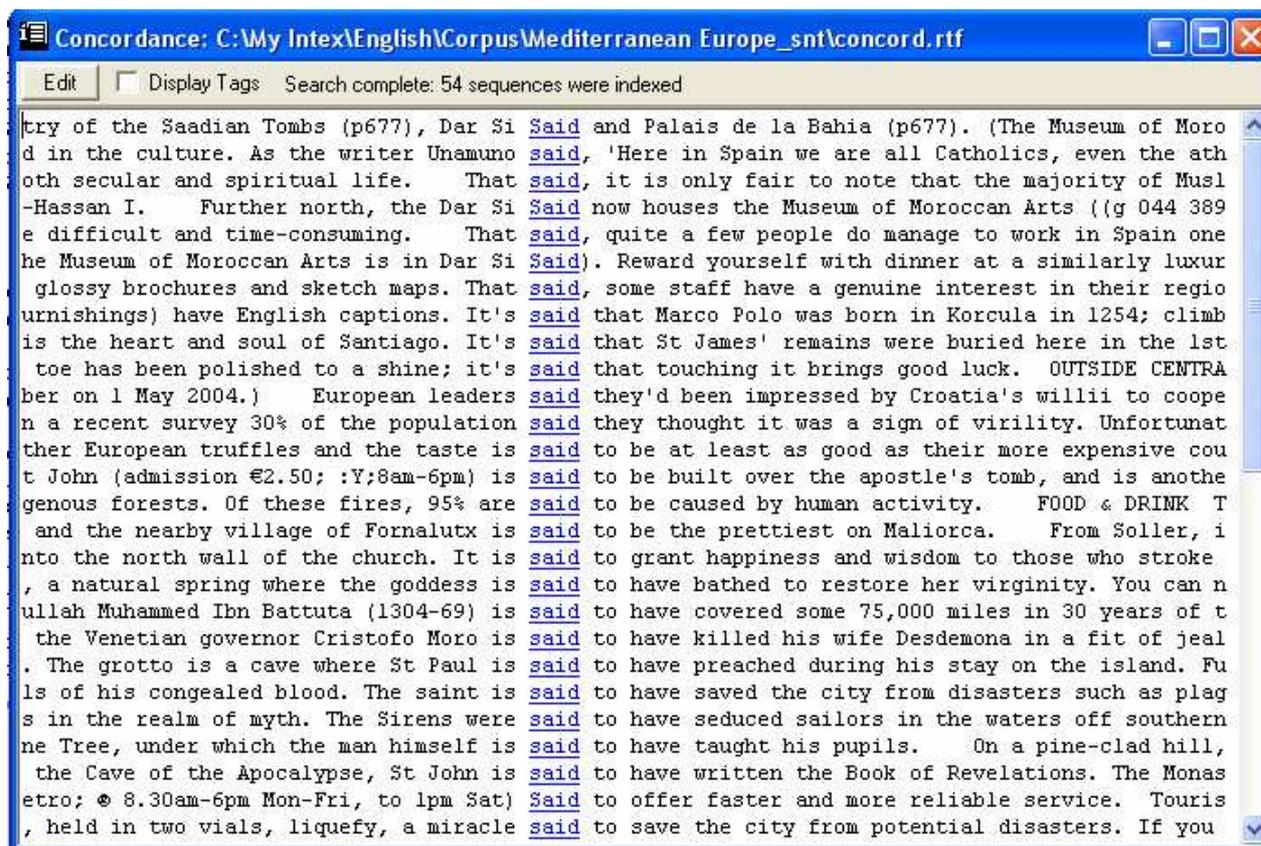
Figure 2: The most frequent form of the lemma *say* in the Mediterranean corpus.

Step three - Extending the context

When asked to read the paragraphs on a particular topic and select useful vocabulary, most learners chose to study the *sights* topic, the one that was set as an example and did not venture in searching for similar possibilities. However, each learner had a different set of items extracted from the paragraphs and the choice of selected vocabulary ranged from single words to whole sentences. The items were just listed and no attempt at grouping them or organising them for future reference was made.

Step four – Comparing synonyms

The comparison of synonyms *crowded* and *packed* proved to be extremely difficult for most students and they reported they "feel there is some kind of difference but cannot describe it" (in the exact words of one of the students). Teacher guidance was necessary and further research was prompted by a couple of questions: *What can be crowded/packed? What can something be crowded/packed with?* Students were also reminded of the possibility of

left/right sorting which produced some results. The students also spontaneously opted for group work.

Table 7 is the sample answer offered by one of the groups. Not all groups were equally successful but they all answered the questions used to facilitate the analysis. All answers lack the final conclusion about the actual difference between the two lexemes, i.e. the analyses was successful but the final synthesis was not. The corpus context did allow to notice the fine difference between the two synonyms but the final observation about the meaning of *packed* being stronger than that of *crowded* is entirely intuitive and is not supported by the actual evidence from the text.

Table 7: Difference between synonyms crowded + packed – a successful learner worksheet.

CROWDED

What can be crowded?

city, town, restaurant, buses, camping ground, beach, bar, areas, places, office, public transport, streets, quay

What is it crowded with?

Only four examples say what it is crowded with:

- beaches get crowded <u>with holiday-makers</u>
- It's regularly crowded <u>with diners munching contentedly on house specialities</u>
- the old city crowded <u>with 18th-century townhouses</u>
- Malta has a fascinating history and is crowded <u>with physical and cultural reminders of its past</u>

In other examples it is probably always people although it is not clearly said.

In the left context we find these words:

can be, can get, hot and, a bit, too, least, less, often, sometimes, it gets, always

PACKED

What can be packed?

Many different places are described as packed: beaches, town, accommodation, sights, shelves, camping grounds, shopping malls, main square, restaurant, theatres, tables, backstreets, major international routes, cafes, village, dance floor, resort, menu, …

What is it packed with?

- in many cases "packed" is followed by "with"
- things and places can be packed not only with people but also things and similar, for example: restaurants, shops, churches, pensions and hotels, sights, healthy food, boutiques, practical information, furniture, …

In the left context we also noticed:

is, are, absolutely, always, densely, generally, often, thoroughly, tightly, usually, jam-

It is difficult to talk about the difference in meaning. Maybe «packed» sounds a bit «stronger», like «very crowded».

Step 5 – Using regular expressions

Task 1 encouraged learners to narrow down the search for adjectives preceding the noun *restaurant* using the symbol for adjectives (<A>). The task was easy to carry out for all 42 students and they all noticed that narrowing down the search in this way (106 lines instead of 317) simplifies the search.

After completing the first exercise the students were asked to write the regular expression which would generate such a concordance in which they would be able to check their answers. This was too difficult for all the participants so they were offered two possibilities: "a<N>of" and "a<MOT>of" ("N" being a symbol for a noun, and "MOT" being a symbol for any word) and using one of these two ways they accessed the answers to their task. They were asked to peer-evaluate their colleagues' work by checking it against the corpus solutions. Some students preferred self-evaluation to peer-evaluation.

Step six – Collocation

Some of the learners' profiles of the lemma *beach* produced in task 1 were better thought out and more carefully presented while others indicated the lack of skill in tackling this type of exercise. The choice of what to include was either guided by frequency or the learners tried to note down as much information as possible. The selected collocations were classified into the following groups: adjectives, nouns, verbs, prepositions, interesting expressions, and the like. The graphic presentation also varied from simple *Word* tables to *Microsoft Visio* presentations. One of the most successful presentations was used for discussion and analysis of possible improvements (Marinov and Pašalić, 2009: 188).

Students described task 2, in which they were supposed to order the three adjectives in front of nouns, as very complex because all combinations seemed to be acceptable for them. The task was rendered difficult by: long concordance generated for the regular expression "<A><A><A><N>", errors in morphosyntactic tagging, and the students' inability to classify the adjectives into groups according to meaning. Teacher intervention consisted in increasing the number of lines to be observed (from the original 13), however, without looking into the full concordance and offering Carter & McCarthy's (2006) terminology for classifying adjectives according to meaning (evaluation, size, physical quality, shape, age, colour, participle, origin, material, type, purpose).

A new task now was to classify the adjectives in the groups proposed by the grammar and establish regularity in their ordering. Table 8 presents the joint results of the whole class

efforts. The examples sufficed to establish the following order of adjectives: evaluation + size + age + origin + type. The students could only guess about other classes since there were no sufficient examples to illustrate all possible combinations.

Table 8. Classification of adjectives into the CGE groups of meanings –the final result of the whole class efforts.

addictive young green	evaluation + age + colour
adjacent early Christian	physical quality + age + origin
ancient Roman commercial	age + origin + purpose
atmospheric old walled	evaluation + age + participle
dashing young Corsican	evaluation + age + origin
excellent neoclassical archaeological	evaluation + origin + type
fine old classical	evaluation + age + origin
fine old neoclassical	evaluation + age + origin
gorgeous old Venetian	evaluation + age + origin
great little medieval	evaluation + size + origin
greatest ancient Greek	evaluation + age + origin
largest inhabited medieval	size + participle + origin
largest living medieval	size + participle + origin
lovely Italian baroque	evaluation + origin + type
major Greek archaeological	evaluation + origin + type
major international rental	evaluation + origin + type
modern Turkish pop	evaluation + origin + type
narrow Croatian coastal	physical quality + origin + type
popular contemporary Italian	evaluation + age + origin

small daily vegetarian	size + age + type

Step seven – Colligation

The matching exercise proposed in task 1 of this step is a well-known type of exercise learners are accustomed to. As such, they found it relatively easy to complete. Although the students did recognise some colligational patterns *(bored with, accused of)*, these were used only as a confirmation of their choice and not as a starting point for the exercise. It was only upon completion of the exercise, in the process of getting feedback, that the regularity of verb being in the left part of the exercise and the preposition in the right part was actually established.

Even though the concordances of *worth* in task 2 were 150 lines long, their analysis presented no problems for students. The weaker students worked on its paper version and took it for homework while stronger students were able to complete it in class. Table 9 is an example of the type of answers students gave in response to the guided task.

Table 9: Outcome of a student's analysis of the 150 lines long concordance of the lexeme *worth*.

The word WORTH appears 150 times in this corpus.

It's worth studying because there are many examples on how to use the word "worth", which is good to know when writing some texts.

It's usually followed by verbs (catching, investigating, considering, noting, paying, remembering) and articles followed by nouns (a look, the effort, the price, the visit).

The verbs are in -ing form and the articles are followed by nouns in singular.

Step eight – Word class

The learners found this task very demanding, in particular some of the task requirements. After some dictionary consultation the first part of the exercise (gap-fill) was completed. The pattern of the highlighted chunks was also established although with some difficulties in formulating/wording it. However, the following step in which they were asked to group the key verbs according to their meaning resulted in half of the students giving up altogether. The rest of them (already working in self-created groups) compared their results and came up with one answer as shown in Table 10.

Table 10: Classification of verbs into groups according to meaning – learner worksheet.

make it possible: allow, enable, entitle, (encourage),

ask/require sth from someone: ask, expect, need, require, tell, instruct, (invite)

make: cause, encourage, get, invite, lead, lull, (instruct)

4. Discussion and practical implementation

It is evident from the analysis of learners' approach to the tasks and the results achieved that different tasks presented different levels of difficulty, which can be taken into consideration in the future similar projects or design of classroom materials.

The analysis has shown that it is not always necessary to start with the convergent and then move on to divergent tasks (Bernardini, 2004) because the exercise outcome is always influenced by a whole range of factors. For example, the training started by a very simple divergent task (Step 1) which caused no problems, while a convergent task of Step 6 (Task 2) proved to be very demanding and required the teacher's intervention both in terms of extending the task and bringing in more reference material as well as the actual learner guidance through the task.

It is the complexity of the task that has proved to be decisive for the outcomes. The corpus-driven exercises may be rendered complex by: the length of the concordance, by the lack of context in the truncated concordance lines, difficult context of the unabridged, unedited, authentic language material, and by the level of analytical skills required to come to the solutions.

Although lengthy concordances may be difficult to handle, this problem was partly taken care of by working with a small specialised corpus, which is more easily manageable (Aston, 1997). Lengthy concordance outputs were an advantage, not a limitation, in cases where the task was simple and straightforward (Step 5, Task 1), when patterns were easily observable and/or where the tasks were well guided by an additional set of questions (Step 7, task 2).

It is difficult to remedy for a lacking context (unless learners work in the hard version) but difficult contexts (unknown vocabulary) were tackled by using dictionaries and the Internet.

The major component of the task complexity, however, is the type of analytical skills it requires (Step 2, Task 2; Step 4; Step 6; Step 8). Poorer results and often a loss of

motivation are evident in the tasks which require the ability of analysis and synthesis in order to draw conclusions or establish rules as was the case with synonym comparison, adjective sequence or word class categories. This does not mean that such exercises are to be avoided but that they are to be more carefully guided.

An example of a successfully guided task is the one of colligation with *worth*. This case also illustrates the need for and the benefit of flexibility in corpus-consultation processes where the procedures should constantly be adapted to the type of learners, their responses and reactions. Corpus materials allow such changes, adaptations, and improvements to be made (as was the case with Step 4 or Step 6, Task 2).

The tasks in which the students were asked to exercise more autonomy often lacked a good strategy and organisation. They revealed greater differences in individual approaches as, for example, the willingness to generalise the data and organise it more systematically (e.g. in profile of *beach* or searching through *sights* paragraphs).

In terms of the outcome such tasks have indicated that practicing DDL with this type of learners can be rendered more beneficial if the tasks contained more detailed directions of what learners are required to do. In the case of Step 6, Task 1, learners' profiles of the lemma *beach* might have been more informative, correct, and complete had they been given specific instruction of what had to be included in the profile (part of speech, number of items, position in relation to the node word, grouping of collocates according to meaning). Learners might have also benefited more from Step 3 had their extensive reading of *sight* paragraphs been more carefully designed content wise around the precise concept they were looking for (e.g. time expressions, locations, activities, historical periods, ...) and language wise with the specific instruction about what kind of lexical items to select (e.g. besides isolated words they are asked to note down longer meaningful stretches of the sentences). In loosely set tasks longer concordances also become more difficult to handle because apart from managing a massive amount of language students also have to establish the strategies for organising meaningful output (Step 6, Task 1).

In DDL, as in language teaching in general, getting the correct answer is not the only purpose of indulging into particular activities. The knowledge learners acquire en route may be equally crucial or even more important because it establishes foundation for further learning. There is often no such thing as a definite answer or a solution to the corpus-derived exercise as is the case with some more traditional exercises. Therefore, although the use of corpus usually emphasises the result of the linguistic analysis, i.e. the declarative knowledge,

it is important to remember that the process of corpus analysis also develops the so-called procedural knowledge where essential learning skills are developed (Aston, 1996; Cohen, 2003). While working with corpus data students are trained to develop a process-oriented view of language learning which encourages their active involvement (O'Sullivan, 2007).

In terms of procedural knowledge a number of linguistic as well as pedagogical benefits were obtained. The learners were actively involved in rule generation by the process of deduction (Step 2, Task 1; Step 6, Task 2; Step 8). They were motivated to discuss their results with their peers, exchange opinions and join efforts in finding and presenting the final result. The learners also engaged in self-evaluation and peer-evaluation, were able to notice the benefits and limitations of particular reference materials (corpus vs. grammar). In Step 6, Task 2, for example, the corpus did not offer sufficient information for generation of a complete rule but it offered a significant number of examples contextualised in the corpus characterised by authenticity of text and purpose (Johns, 1988). On the other hand, the grammar offered an exhaustive rule based on a much larger corpus but a smaller number of examples atypical of the language of tourism. Thus, the ground for more individual, independent, and creative work was provided. Such exercises, although yielding poorer and less structured results from the traditional point of view, should be seen as a driving force towards more autonomy in learning and are a necessary part of the full package.

The analysis has shown that problems that arise from dealing with particular exercises may also provide ground for discussion. An interesting discussion was raised about the difference between *famed for* and *renowned for* when a learner suggested there might be some difference in the connotative meaning of the two. The concordance generated to compare the two lexemes dismissed the hypothesis. Deduction of meaning was not always successful but studying the newly generated concordances the learners got an insight into some contexts in which the word appears and were motivated to look it up in a dictionary (Step 2, Task 1). This example pointed out the possibility of deducing meaning from multiple contexts but also its limitations.

Many of the exercises designed illustrate the benefits of small specialised corpora as envisaged by Aston (1997). They are more fully analyzable and easier to become familiar with. In the proposed set of exercises this advantage is especially evident in Step 3. The idea of paragraph search came as a result of familiarity with the structure and the content of the texts contained in the corpus. Although the main idea of corpus use and concordance output reading is studying the immediate context of the key word, i.e. intensive reading in vertical

direction, this exercise has shown that a small corpus enables one to design exercises for a well targeted and rather productive extensive reading. If one wants to develop learners' good writing skills, extensive reading of texts similar to those we want to write should help subconscious acquisition of such a skill (Lewis, 1997).

The advantage of using a POS tagged corpus has, at this point, become providing the opportunity for designing better targeted exercises containing various patterns and frames. Learners' awareness of this benefit has remained at the level of narrowing down collocational patterns with different parts of speech, especially adjectives (Step 5, Step 6). Nevertherless, it would be too ambitious to try and teach them how to actually write their own regular expressions.

5. Conclusion

This study engaged 42 learners of tourism in Data-Driven Learning and analysed their responses to it. The participants had never been introduced to the idea of language corpus or the possibilities of using it as a source of language examples or reference material in their language acquisition before. In order to enable them to use it autonomously in improving their written work in the later stages of the project (Marinov, upcoming) they had to be equipped with some technical knowledge for handling concordance programmes but also with some practical linguistic input in order to direct their attention to the kind of language phenomena that can easily be observed in a corpus. The tasks included in the eight steps of the training programme reflect these two goals.

The awareness of the advantages a small specialised corpus may have in LSP teaching Mediterranean corpus specifically triggered this study, and the corpus was exploited to teach and learners were taught to exploit it.

Overall, the training process showed it is possible to introduce students of non-philological studies to corpus consultation although they may approach some tasks with more or less motivation and enthusiasm. This finding has confirmed other empirical studies and is also true of other classroom activities.

Some specific recommendations for future application and research can be made as follows:
- the soft approach should/can be combined with the hard approach,
- simple exercises should be combined with more challenging ones,
- convergent exercises should be combined with divergent ones,

- more guidance should be provided for more analytically demanding tasks if all learners are to be involved and not only the more skilful and creative ones,

- both individual and collective/collaborative approach to tasks should be encouraged,

- the teacher should exercise flexibility in adapting the tasks and creating more corpus queries if the situation requires it (at this stage and with this type of learners it cannot be left to them),

- classwork should be combined with homework because corpus work is very time-consuming.

This study has been able to provide more qualitative that quantitative data due to the lack of controlled situation in which students could be observed. Since the learners combined classwork with homework and individual with collaborative efforts, their individual performance cannot be measured. Based on the qualitative data collected some future studies can choose to focus on only one of the observations and try to study it in a more controlled fashion. For example, a guided approach to the same exercise can be contrasted to the unguided one, the abilities to generalise and establish rules for different types of language items can be compared, etc.

References

Aston, G. (1996). What corpora for ESP?. Paper presented at a conference on ESP at the University of Pavia, November 1996. Available at: http://home2.sslmit.unibo.it/~guy/pavesi.htm

Aston, G. (1997). Small and large corpora in language learning". in J. Melia and B. Lewandowska-Tomaszczyk (eds.) *PALC '97 Proceedings. The First International Conference: Practical Applications in Language Corpora,* 51-62. Łódź: Łódź University Press. available at: http://www.sslmit.unibo.it/~guy/wudj1.htm.

Bernardini, S. (2002). Exploring new directions for discovery learning. In B. Kettemann and G. Marko (eds.). *Teaching and Learning by Doing Corpus Analysis*. Proceedings from the Fourth International Conference on Teaching and Language Corpora (pp. 165-182). Amsterdam: Rodopi.

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future developments. In J. Sinclair (ed.), *How to Use Corpora in Language Teaching* (pp. 15-36). Amsterdam: Benjamins.

Bowker, L. and Pearson, J. (2002). *Working with Specialised Language: A Practical Guide to Using Corpora.* London: Routledge.

Carter, R. and McCarthy M. (2006). *Cambridge Grammar of English* (CD – ROM). Cambridge: Cambridge University Press.

Chambers, A. (2005). Integrating corpus consultation procedures in language studies. *Language Learning & Technology 9*(2), 111-125.

Chambers, A., & O'Sullivan, Í. (2004). Corpus consultation and advanced learners' writing skills in French. *ReCALL, 16*(1), 158-172.

Cheng. W., Warren, M. and Xun-feng, X. (2003). The language learner as language researcher: Putting corpus linguistics on the timetable. *System, 31*(2), 173-186.

Cohen, A. (2003). Strategy training for second language learners. ERIC Digest. Available at: http://www.cal.org/resources/digest/0302cohen.html.

Davies, M. (2000). Using multi-million word corpora of historical and dialectal Spanish texts to teach advanced courses in Spanish linguistics. In L.Burnard and T. McEnery (eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 173-185). Frankfurt: Peter Lang

Fligelstone, S. (1993). Some reflections on the questioning of teaching, from a corpus linguistics perspective, *ICAME, 17*, 97-109.

Gabrielatos, C. (2005). Corpora and language teaching: Just a fling or wedding bells? *TESL-EJ 8*(4). Available at: http://writing.berkeley.edu/TESL-EJ/ej32/a1.html.

Gavioli, L. (2005). *Exploring Corpora for ESP Learning*. Amsterdam: Benjamins.

Hadley, G. (n/a). Sensing the winds of change: An Introduction to Data-Driven learning. available at: http://www.nuis.ac.jp/~hadley/publication/windofchange/windsofchange.htm

James, K. and Garrett, P. (eds.) (1991). *Language Awareness in the Classroom*. London: Longman.

Johns, T. (1988). Whence and whither classroom concordancing? In Bongaerts et al. (eds.) *Computer Applications in Language Learning* (pp. 9-33). Dordrecht: Foris Publications.

Johns, T. (1991). Should you be persuaded: Two examples of data driven learning. In Johns, T. and King, P. (eds.), *Classroom Concordancing, 1-16.* ELR Journal 4. Birmingham: University of Birmingham.

Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology 5*(3), 77-90.

Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning & Technology* 11(3), 28-44

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

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

Lewis, M. (1997). *Implementing the Lexical Approach: Putting Theory Into Practice*. Hove: Language Teaching Publications.

Marinov, S. (2010). Improving students' vocabulary using a corpus. 10th International Educational Technology Conference Proceedings Book, Volume II. available at: www.iet-c.net/publications/ietc2010-2.pdf, 2010

Marinov, S. (2011). The role of small specialised corpus in teaching ESP. Unpublished master's thesis, University of Zadar.

Marinov, S. (upcoming). Can students use corpora to enhance their writing?. in Lenko-Szymanska, A. (ed.) *TaLC10: Proceedings of the 10th International Conference on Teaching and Language Corpora*. Warsaw: Institute of Applied Linguistics, University of Warsaw. available at: http://talc10.ils.uw.edu.pl/proceedings/

Marinov, S., Pašalić, M. (2009). Bringing corpus to classroom. In *Strategies for optimising the quality standard in higher education ESP* (pp. 183-192). Cluj- Napoca.

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

Partington, A. (1998). *Patterns and Meanings*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Silberztein, M. *Intex*. available at: http://intex.univ-fcomte.fr/ and http://www.nyu.edu/pages/linguistics/intex/

Sinclair, J.M. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Škiljan, D. (1987). *Pogled u lingvistiku*. Zagreb: Školska knjiga.

Thompson, P. (2004). The University of Reading master's program in applied linguistics (http://www.reading.ac.uk/)

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Tribble, C. & Jones, G. (1990). *Concordances in the Classroom*. London: Longman.

Tribble, C. (1997). Improvising corpora for ELT: Quick and dirty ways of developing corpora for language teaching. In Lewandowska-Tomaszczyk, B. and Melia, P.L. (eds.), *Practical Applications in Language Corpora*. Łódź: Łódź University Press.

Willis, D. (2005). *Rules, Patterns and Words – Grammar and Lexis in English Language Teaching*. Cambridge: Cambridge University Press.