

Using corpus-based classroom activities to enhance learner autonomy

Agata Lewandowska

Adam Mickiewicz University, Poznań

alewan@amu.edu.pl

Abstract

Data-driven learning (DDL) is a relatively recent trend in language pedagogy in which class work is built around corpus data. Corpora may not be very easy to work with in the classroom, but advanced language learners should be offered opportunities to use them in order to find out what type of information can be retrieved and how to process it. Analyzing concordances under the teacher's guidance may inspire learners to try to resolve their doubts on various points of L2 grammar and usage by performing their own queries. Thus, learners develop heuristic skills which they may find useful in further stages of their language education, especially after the completion of formal instruction. The paper describes a study involving a series of DDL lessons focusing on eliminating previously identified interference errors. A few examples of DDL materials will be included, supplemented with the discussion of the ways in which reference to a corpus allows learners to access frequency and co-occurrence data and make their own language choices on the basis of that information. At the end of the project, the participants were asked to provide their feedback on the experience by means of a questionnaire which examined their attitudes towards the new tools and the idea of using them in their further language work. The paper reports and discusses the results of that survey.

1. Introduction

An important responsibility of every language teacher is to help learners become independent in their language development. The task is even more vital

in advanced classes, where learners are close to reaching the end of their formal education, and begin to prepare for the time when they will take full professional responsibility for the way they use English in their teaching, research, translation, or office work, for example. It is a teacher's task to provide such students with a range of skills and instruments that will allow them to make further progress and become truly independent language learners and users. This paper focuses on one particular instrument, that is a *language corpus*. Admittedly, there have been reservations concerning classroom use of corpus-based materials, voiced by such unquestioned authorities in language pedagogy as Cook (1998), Widdowson (2000) or Kilgarriff et al. (2008). The main objection raised by these authors is that reading concordances is too high a challenge for most language learners, as it is an advanced skill which requires high-level processing. Another issue discussed in connection with classroom concordancing is the fragmentary nature of this type of linguistic input, devoid of information concerning the context of situation and non-verbal aspects of communication. Despite these reservations, the possible benefits for advanced, motivated students seemed too appealing not to undertake the challenge and introduce the idea of corpus reference into the classroom. The English corpus used as a source for classroom materials was BYU-BNC (Davies 2004), which provides information on the type of sub-corpus (text type) for each concordance line (see Figure 2 below), revealing the possible function and level of formality of the source text. Wherever it was relevant, the information was preserved in the lesson materials, which, to some degree, compensated for the shortcomings mentioned above.

Classroom use of corpus-based materials was the subject of the author's PhD research project, whose general aim was to test the effectiveness of corpus-based teaching techniques in eliminating previously identified interference errors (cf. Lewandowska 2013). Some aspects of the study, however, are related to the area of learner autonomy and it is these that will be discussed in this paper.

2. Rationale

There are numerous ways in which the use of a language corpus in the classroom could be linked with the concept of learner autonomy. First of all, a corpus is an incredibly rich and flexible self-access instrument, which can provide learners with answers to their questions and resolve many doubts concerning their lexical, grammatical and stylistic choices. Being able to use such a resource effectively could be counted among what the *Common European framework of reference for languages* (CEFR) calls *heuristic skills*, which constitute an important aspect of their general ability to learn (i.e. *savoir-apprendre*). To be specific, heuristic skills include an ability to observe and draw conclusions from the

analysis of authentic language material, the ability to find and process new information, and “the ability to use new technologies (e.g. by searching for information in databases, hypertexts, etc.)”. When discussing these skills, the document recommends making provision “for learners to become increasingly independent in their learning and use of language” (Council of Europe 2001: 108).

As Little (2004) emphasizes, self-access instruments must not be identified with and do not guarantee learner autonomy. Yet, according to Holec (1981: 3), one of the five key elements of taking charge of one’s own learning, i.e. of autonomy, is “selecting methods and techniques to be used”. This translates into the ability and readiness to choose such resources which will be most suitable to a given learner’s needs, cognitive and learning style or, in other words, such instruments which he or she finds most effective and useful. In order to make these choices, learners need to have an awareness of what instruments are available, and whether they could benefit from them. They need to experience a variety of techniques in the classroom and test them in a friendly environment, with easily available support from the teacher. Only then will they be able to assess the usefulness of these instruments for their needs and to make informed choices. With the amount of useful language data that corpora can offer, and with their growing accessibility, it seems that introducing corpora into the language classroom is worth recommending, especially in advanced courses. Corpus-based activities, which usually involve concordances, teach learners how to process language data inductively and make generalizations. Conrad and LeVelle (2010) make the following comment about implementing data-driven learning techniques in language lessons: “Learner autonomy is increased as students are taught how to observe language and make generalizations, rather than depending on a teacher who states rules for them” (2010: 548). The authors also list other features of the activities which seem consistent with key principles of language education, that is hypothesis generating and testing, noticing and consciousness-raising, as well as emphasis on the inductive approach. Bernardini (2002: 15) talks about the concept of *serendipitous learning* which often occurs in corpus-based activities: learners discover other new facts about the target language by chance, because the data gathered in order to analyze one aspect of grammar or lexis may reveal another. Such discoveries may have a strong motivational effect on many learners, stimulating their curiosity and giving them opportunities to solve language puzzles which they pose to themselves and to their colleagues.

Aston (2001: 41) goes as far as to say that “Perhaps the greatest attraction of corpora for language pedagogy is their potential for autonomous learning”. The potential is the result of a shift in the roles of both the teacher and the learner. The former no longer needs to be the one and only authority concerning

language, and may delegate part of the responsibility to the corpus, which non-native speaker teachers may find especially appealing. Consequently, both the teacher and the students jointly endeavor to find the best answers to the questions which arise in class. The teacher will be a facilitator of the process rather than a judge. Thus, the situation evolves towards learner empowerment, one of the three principles defined by Little (1999) as fundamental to developing learner autonomy. What is even more important, the sense of empowerment can be sustained by the student, who may return to the corpus for more input in the future.

Finally, Little (1995, 2004) draws attention to the other side of the learning process, saying that it is essential for the *teacher* to be autonomous too. He goes as far as to say that teacher autonomy is a prerequisite for the development of learner autonomy. How can teachers become autonomous? It must be part of their training: "We must provide trainee teachers with the skills to develop autonomy in the learners who will be given into their charge, but we must also give them a first-hand experience of learner autonomy in their training" (Little 1995: 179). The key to autonomy in the classroom is negotiation; course content, course requirements and assessment criteria should be established in that way. According to Little, "the basis of this negotiation must be a recognition that in the pedagogical process teachers as well as students can learn, and students as well as teachers can teach" (1995: 180). Again, working with corpora may prove useful in achieving that goal, because the shift in teacher and learner roles away from authority and subordination towards co-operation and two-way exchange of ideas could open the door to such processes. Classroom work with a corpus offers both teachers and learners numerous opportunities for posing unexpected questions and finding out new, surprising facts about language. If, moreover, the teacher engages in building a small-scale learner corpus and performs interlanguage analysis for a particular group of students, he or she will prove autonomous in a very practical way, relevant to the students' situation. The data obtained in this way could provide an impetus for negotiations concerning course content, i.e. what aspects of language the course needs to address more intensively. In other words, the teacher must be an example for the students by practicing the approach that he or she advocates. The potential benefits are as follows:

- the teacher develops new skills, and at the same time obtains unique information about his or her students' needs and the areas of lexicogrammar that need more attention in class;
- the teacher demonstrates autonomy, allowing the learners to model their behavior and attitude to language study on their classroom experience (crucial in teacher training);
- students obtain feedback on their use of the target language in an interesting and informative way.

The ideas discussed above are the foundations of the study presented in this paper. Some more information about the project can be found below.

3. The study

As stated above, the paper presents part of a larger research project, whose main focus was an experimental study of the effectiveness of DDL techniques (pre-test/post-test design). The aspects of the project which are connected with the concept of learner autonomy were analyzed by means of a survey. The participants of the study took part in a sequence of corpus-based lessons, and were subsequently requested to respond to a questionnaire which addressed some autonomy-related issues and prompted them to provide feedback on the experience. The lessons carried out for the purpose of the study were part of an advanced course in English grammar, and so they focused mostly on language characterized by high levels of accuracy and formality. The major objective of the lessons was to eliminate some previously identified first language (L1) interference errors from the learners' English. The error analysis element was based on examination essays and a learner corpus compiled by the author out of students' blogs, maintained as part of another project. The learner corpus was also used in designing class materials for the experimental lessons. Through this component of the study, the author was able to demonstrate her autonomy as a teacher, and to evoke interest in the lessons on the part of the students, who found analyzing samples of their own and their peers' writing quite engaging. At a more general level, the sequence of intervention classes was intended to introduce students to the idea of using a language corpus as a source of language data and as reference.

3.1. The subjects

The study was carried out with 35 third-year students (in two groups) at Adam Mickiewicz University's teacher-training college, in the academic year 2010/2011, a year before it was closed. Their L1 was Polish, and they all majored in English, finishing their bachelor degree that year. The students' situation created a favorable context for enhancing autonomy training, because some of them were soon to complete their formal education and start their professional careers while others were planning to continue their studies at the MA level. Hence, they were interested in finding out about techniques, instruments and strategies that would allow them to become truly independent in performing such tasks as writing, translating and editing academic and non-academic formal texts. These were the kind of tasks they could expect to become part of their regular duties, either at work or in their post-graduate courses.

3.2. The procedure

The project was designed into a sequence of activities so as to allow students to benefit the most from the experience. Some weeks before the experimental lessons began, the students had a lesson in a computer lab, during which they had an opportunity to become familiar with corpora available online. The participants had hands-on access to a few corpora, and different interfaces for them. The teacher first provided them with a few example queries, telling them exactly what settings to apply and what to write in the search window. Later in the session, they were supposed to define their own questions on English lexicogrammar, and try to answer them by using a corpus. The teacher was there to advise them on how to formulate their queries in order to obtain the most relevant corpus data. The aim of the session was to let students understand how a corpus works, and what types of information can be retrieved from it. Some guidance was offered in interpreting the results and reading concordances. This involved both the visual display options in the results (e.g. sorting the entries to the right or to the left of the 'node word') and instructions relevant to particular queries (where to look for patterns to find a good answer to a given problem, e.g. what part of speech might be relevant in the concordance lines obtained from a query).

The next stage of the project was the core of the study: the two groups of students were given a sequence of lessons which addressed some of the interference-related issues identified through error analysis of the learner corpus. As mentioned above, the study involved a pre-test/post-test design. Therefore, before the experimental (corpus-based) and control (conventional) lessons began, students were given a test which measured the error rate for each problem on the scale of 1 to 5. The score was based on five items in a translation test for each of the six language problems selected for the study. The error rate was expected to decrease in the post-test, after the treatment. Each group had three corpus-based lessons and three control lessons, so that all students could undergo the experimental treatment, develop opinions on the experience and provide feedback through a questionnaire afterwards.

3.3. The lessons

Each of the experimental classes made use of sets of printed concordances from a native speaker corpus (the *British National Corpus* – the BNC, Davies 2004) and the learner corpus created by the author. Additionally, in some lessons parallel corpus data were included, as they were especially useful in dealing with interference problems. Most often a lesson would start with an example sentence in Polish and

elicitation of its translation into English, which usually provided a starting point for addressing the expected problems. Alternatively, erroneous translations encountered in students' writing and other work (blogs, spoken English, etc.) were used. All materials were printed and made available to students in class. Introducing the self-access format was not possible for reasons connected with research design. It was important that all students be exposed to the same data, in the same amounts, so that measurements of effectiveness would be comparable for all participants in the study.

3.3.1. DDL lesson materials

All the experimental lessons depended strongly on various forms of concordances from native speaker corpora, parallel corpora, and learner corpora (error analysis). In some cases, other corpus data were also included, such as word lists and graphs, collocate frequency lists, syntactic frames and patterns, as well as word frequency data across different sections of a corpus representing different styles.

Working with such materials requires students to modify their approach to learning. Since they are dealing with real-life data, there may be examples which go against the general trend, so learners need to develop a certain degree of tolerance of ambiguity. They must be looking for trends rather than straightforward rules, since there are bound to be 'fuzzy' cases and blurred lines. Yet, when working with paper printouts rather than direct access, the teacher can, or even must, select the concordances to be used and so can eliminate some confusing cases. The materials should not, however, be artificially biased towards a target rule. It is advisable to allow some ambiguity into the data to keep the language authentic and the processing more complex. The questions that should be asked during a corpus-based lesson are: "What is better/more natural?" rather than "What is right/wrong?", because this is the kind of information users can obtain from a corpus. The rules are derived from use, from what native speakers choose to say or write in a given context.

3.3.2. Example materials

One of the experimental lessons focused on the problems that Polish students have when reporting research results. The trigger sentence for the lesson was one that can be very often encountered in popular science and even academic texts in Polish:

Amerykańscy naukowcy stwierdzili, że...
(*American researchers have found that...*)

The initial learner corpus analysis proved that it is quite common for Polish learners even at such an advanced level to choose the verb 'state' here. This could be classified as an interference error of underdifferentiation. The participants

found it difficult to distinguish between such verbs as ‘state’, ‘find’, ‘establish’ or ‘say’ in sentences like this, and chose ‘state’, probably on the basis of phonetic similarity of the Polish verb ‘stwierdzić’. The resulting sentence often reads: “American researchers have *stated* that...”. It must be said that the participants could fall into some other traps of the lexicogrammar involved, like the use of the Present Perfect tense rather than Simple Past, or sometimes the problematic choice between ‘scientists’ and ‘researchers’, for example.

Three corpus-based activities were chosen to address the issue, each of which is briefly presented below. For the sake of space, only samples of the original lesson materials have been provided here. First some parallel corpus material was presented, in which fragments with the Polish verb ‘stwierdzić’ in different forms were found and matched with their English equivalents (see Figure 1). The task was to find the corresponding verbs in the English fragments, so that students could see what choices are available (and most common) in such contexts.

Find and underline translations of forms of the word ‘stwierdzić’ in the fragments below.

POL: W przypadku gdy Komisja <i>stwierdzi</i> , że różnica między przepisami ustawowymi, wykonawczymi lub administracyjnymi Państw Członkowskich narusza warunki konkurencji na rynku	ENG: Where the Commission finds that a difference between the provisions laid down by law, regulation or administrative action in Member States is distorting the conditions of competition in the internal market
POL: Jeżeli <i>stwierdzi</i> , że doszło do naruszenia, proponuje środki właściwe do jego zaprzestania.	ENG: If it finds that there has been an infringement, it shall propose appropriate measures to bring it to an end.
POL: Jeżeli Rzecznik Praw Obywatelskich <i>stwierdzi</i> przypadek niewłaściwego administrowania, przekazuje sprawę do	ENG: Where the Ombudsman establishes an instance of maladministration, he or she shall refer the matter to
POL: kiedy któregoś dnia pani Bovary ośmieliła się <i>stwierdzić</i> , że chlebodawcy powinni dbać o pobożność służby,	ENG: once even, Madame Bovary having thought fit to maintain that mistresses ought to keep an eye on the religion of their servants,
POL: <i>Stwierdził</i> , że przepada za piknikami nawet w upalne letnie dni, kiedy roi się od os i much, ale najbardziej lubi jesień.	ENG: He said that he loved picnics even on a hot summer’s day when there were wasps and flies, but he much preferred the autumn.
POL: Państwo lub Państwa Członkowskie zastosują wszelkie właściwe środki w celu wyeliminowania <i>stwierdzonych</i> niezgodności.	ENG: the Member State or States concerned shall take all appropriate steps to eliminate the incompatibilities established.

Figure 1: Parallel corpus concordance for the Polish lemma ‘stwierdzić’.

The next step was an analysis of native speaker corpus data (the BNC, Davis 2004-), where participants were asked to pay attention not only to the grammatical structures used in the fragments but also to the level of formality

of the text and the origin of the source material provided on the left margin. The style must be defined as formal or even official, and the texts are within the area which could be classified as the humanities, with more accurate labels being political, legal, educational, and one technical; some are academic and others are not. This aspect of the use of the verb 'state' needs to be emphasized in light of the learner corpus data given in Figure 4 below.

W_ac_polit Law Edu	that it should be put into effect. The agreement stated that " in consideration of such desire " the
W_ac_polit Law Edu	liable for the injury to the plaintiff. The court stated that " negligent conduct is more likely to break
W_ac_polit Law Edu	old paradigm is replaced with a new one . Kuhn stated the idea boldly , making paradigms very general
W_news_p_oth_er Com	unamended form applicable at the relevant time stated : "Where a term of a contract purports to exclude
W_non_ac_hum Arts	known criminals . In 1852 an official report stated that the " uselessness of the force , as at present
W_non_ac_polit Law Ed	international financial community . Fujimori stated , however , that the loan had been approved on
W_non_ac_polit Law Ed	foreign nationals as " human shields " , and stated emphatically that " the sovereignty of Kuwait is
W_non_ac_polit Law Ed	High Commission for Refugees (UNHCR) stated on Dec. 5 that more than 200,000 Kurds had fled
W_non_ac_polit Law Ed	official joint statement Shar " and Genscher stated only that they had discussed " the situation in the
W_non_ac_tech Eng	position . However , for reasons previously stated these have not been used in the present design .

Figure 2: BNC concordance for the verb 'stated' (with errors indicated).

In order to demonstrate the difference in use between the Polish verb 'stwierdzić' and the English verb 'to state', the next set of data presented to the students was a concordance from the *National Corpus of Polish* (NKJP, Pęzik 2012), with the lemma 'stwierdzić' as a key word (see Figure 3). The task administered with this material was aimed to find the best verb in English as an equivalent appropriate for a given context (partial translation).

How would you translate the key word in each example? (NKJP data)

Sam uczestniczyłem we mszy w obrządku wschodnim i	stwierdziłem	, że jest ona po prostu trochę inna.
Scenę Mimów przy Operze Kameralnej. Po kilku latach	stwierdził	, że aby się rozwinąć, musi być sam.
przez jordanowskich policjantów. Funkcjonariusze	stwierdzili	, że mężczyzna nie posiada wymaganego przy
. Szkoła jednak okazji. Oglądając biegi na 800 m	stwierdziłem	, że na pewno zakwalifikowałbym się do
". Wezwana karetka zabrała go do szpitala. Lekarze	stwierdzili	, że poparzeniu uległo 85 proc. jego ciała.
Wrócą czasy grzecznej telewizji? "Naukowcy	stwierdzili	, że prawdopodobieństwo zajścia w ciążę lub
taka drobna różnica. W czym? Otóż pan prezydent	stwierdził	, że Rywina zna: "Siwiec, Szymczycha czy
wywiesi flagę Unii Europejskiej. Uczestnicy spotkania	stwierdzili	, że sportowcy ze Starego Kontynentu już od
innych takich producentów w Polsce. Kontrolujący	stwierdzili	, że takie rozwiązanie wybrano na podstawie
Wymiotło sklepy! Byłem wczoraj w Berlinie i	stwierdziłem	, że zamknęły znane mi sklepy fotograficzne!
jest sukcesem, zawodnicy po zakończonym turnieju	stwierdzili	, że zrobili więcej niż sami oczekiwali. : A
weszli na stację orbitalną w maskach tlenowych i nie	stwierdzili	aby znajdowały się w jej atmosferze znajdowały
okazała się ponysłą dla Gromowskiej. Wojewoda	stwierdził	nieważność uchwały, o której dyskutowano na
były celowo fałszowane. W pobranych próbkach biegli	stwierdzili	obecność kilkuprocentową frakcji lekkich
W toku czynności sprawdzających	stwierdziliśmy	to, co znalazło się w uzasadnieniu o odmowie

Figure 3: An NKJP concordance for the lemma 'stwierdzić'.

Finally, the participants were to address the issue from a teacher's point of view. Figure 4 presents a concordance of the lemma 'state' (as a verb) from

the learner corpus (with the original spelling preserved). The lines which have been underlined here were classified as errors in the error analysis performed for the needs of the study. The material was used in class for an error recognition activity (without the underlining), in which students were to recognize which of the uses of the word were appropriate and which were not, and then to recommend alternatives to the inappropriate ones.

the right person to do it. All I wanted to state is what I think about the situation as such
on that you care for him/her. Other people state that each day is a good day for expressing
k about three years ago I would definitely state that it's all rubbish. Whereas now, I feel
use of different reasons, it's not time to state them now and to make you feel sorry for me.
port he has but, for me, it's too early to state whether he is capable of bringing back the
ide you with some more music later on as I stated in one of my posts;) Among the songs that
en establishing good rapport with them. As stated in the article, another reason for discipl
oblematic nature. pavlow Welcome! Hi! As stated in the title this blog is entirely devoted
American president who is not white. Obama stated that he would improve the American economy
nent figures in political sociology. He is stated to be the creator of neoliberalism (neocla

Figure 4: Learner corpus concordance for the verb lemma 'state' (with errors indicated).

The teacher's task during these activities was to guide students in their analysis and elicit their observations, focusing on the type of text and style, the absence of reports of academic findings in the native speaker corpus data with the verb 'state', and on the complementation patterns. The latter element is a convenient example of an opportunity for serendipitous learning mentioned above; although this is not the 'target' problem here, the absence of a comma after the verb should be highlighted in the data to prevent or eliminate common punctuation errors in noun clauses with 'that' as a conjunction. The most likely source of the problem is the Polish rule which requires a comma in equivalent contexts (in front of 'że'). This case shows another advantage of corpus-based materials: they prove to be useful in providing negative evidence, which is so rare in natural language acquisition. The language data in combination with quantitative information constitute valuable input in cases where a form popular with learners does not in fact appear in native-speaker language use, or appears only under certain limited conditions.

DDL has a model of language data processing which is described by Johns (1991: 4) in the following three steps:

- IDENTIFY – learners find out from the data what language problem they are to address;
- CLASSIFY – learners are supposed to decide which category of grammar a particular language form represents;
- GENERALIZE – on the basis of the data provided learners try to establish a pattern and formulate a rule.

A corpus might be seen as just another type of reference material, which students with high aspirations concerning the quality of their target language output should be able to benefit from. Classroom work with corpus-based materials could be treated as a form of training, so that learners would be able to perform similar analyses of the issues in the target language which provoke their interest or cause some doubts. In class the teacher is there to guide the learners in analyzing data, recognizing patterns and making comparisons between various sets of data. Afterwards, students should be ready to make their own queries and use the tool independently.

4. Results

Following the sequence of corpus-based lessons, students were asked to share their reflections and impressions with the author by means of a questionnaire composed of nine questions. The aim of the survey was to obtain students' feedback on DDL activities in general, on the effectiveness of different kinds of corpus-based activities, and on their own use of corpora outside the classroom. Responses to three of the questions were found most relevant to problems of autonomy. They provide information on the students' attitude to corpus-based language learning and their own possible future use of the tool.

The questionnaire was written in Polish, the students' L1, in order to prevent any confusion about the meaning of the questions and to make responding to them as undemanding as possible. There were 35 respondents, 27 females and 8 males, most of them 21-22 years of age. The questionnaire was not anonymous, because the project included a correlational study, which required the responses to be analyzed against students' test scores. They were requested, however, to answer the questions in all honesty and without any positive (or negative) bias. The first question was (Lewandowska 2013: 462):

Which of the ten adjectives listed below express your opinion on corpus-based activities?

- boring
- confusing
- convincing
- difficult
- interesting
- effective
- overwhelming
- thought-provoking
- time-consuming
- clear and understandable¹

¹ All translations from Polish mine, AL.

The adjectives were listed in a random order to prevent any patterned or automatic responses, but the results were divided into two groups of adjectives: those with positive and those with negative associations (note that students could choose as many adjectives as they felt appropriate). Figures 5 and 6 represent the two sets of data respectively. The raw numbers for the first set (positive adjectives) were as follows: ‘convincing’ – 27, ‘interesting’ – 24, ‘effective’ – 23, ‘thought-provoking’ – 30, and ‘understandable’ – 26. The raw numbers for negative adjectives (see Figure 6) are much lower: ‘boring’ – 4, ‘confusing’ – 3, ‘difficult’ – 1, ‘overwhelming’ – 7, and ‘time-consuming’ – 7. Since the survey was a supplementary element of the study, the statistical significance of these results was not calculated. The raw numbers do indicate that the negative opinions are definitely outweighed by the positive ones.

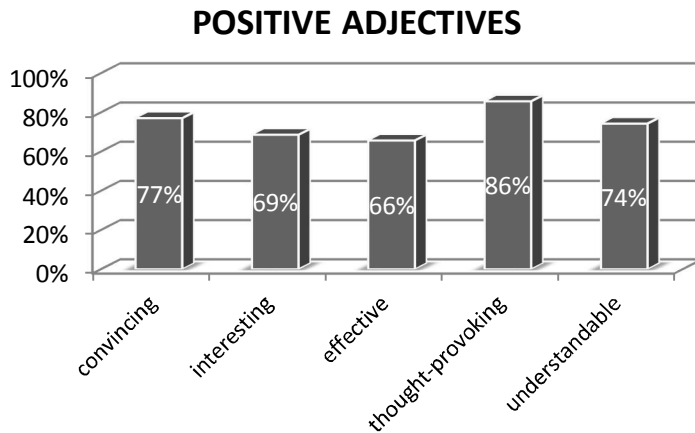


Figure 5: Responses to Item 1 – positive adjectives.

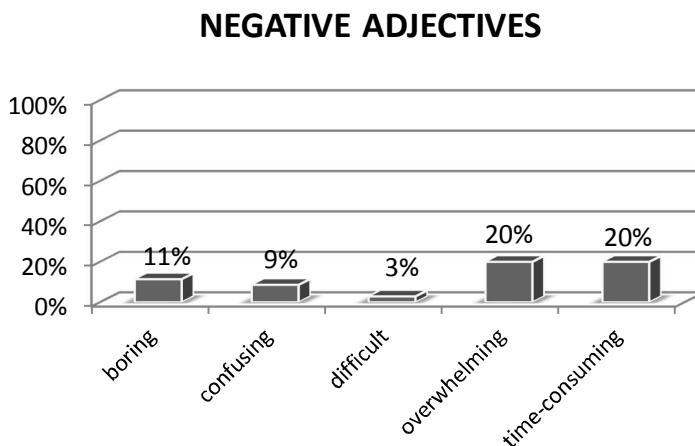


Figure 6: Responses to Item 1 – negative adjectives.

It is quite clear from the responses provided that generally the lessons based on corpus materials left students with positive impressions and did not discourage them. The negative adjectives that had the highest scores were 'overwhelming' and 'time-consuming', which is quite understandable: the amount of data is, indeed, quite large and may prove difficult to process, especially at the beginning. Time-economy is also a concern, but if the problems to be solved are particularly persistent, then perhaps it is worth spending some more time on them.

The change in perception of DDL activities was the subject of the next item in the questionnaire reported here, which was as follows (Lewandowska 2013: 464):

How did you like working with corpus-based materials?

- (1) I did not like using such materials from the start and have not changed my mind.
- (2) I liked using such materials from the start but have changed my mind.
- (3) I did not like using such materials from the start but have changed my mind.
- (4) I liked using such materials from the start and have not changed my mind.

The question was intended to reveal what were the participants' first impressions of corpus-based activities, as well as whether their evaluation changed as more opportunities to participate in them were provided. Figure 7 shows the distribution of responses to that question, and it can easily be observed that the largest group of respondents had a negative first impression. A large majority of these students, however, changed their mind later on, after becoming more familiar with the techniques and gaining more confidence.

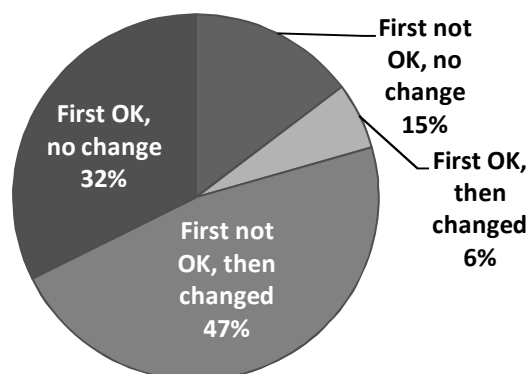


Figure 7: Responses to Item 5 – students' opinion and how it changed.

The subjects of the study were all students of English at a teacher-training college, and so it was interesting to find out whether the corpus project had any impact on their future work as teachers, or at least on their ideas of it at that

point. This was the reasoning behind the next item in the questionnaire, which requested them to confirm or deny their willingness to use corpora and DDL techniques in their future teaching, if they ever became teachers. A ‘compromise’ option was provided, which allowed the respondents to limit the intended use of corpus-based techniques to certain types of courses (the question was open-ended so that students could specify what courses would be suitable). Figure 8 shows students’ responses to that question, and they are, one must say, very encouraging. A vast majority of students expressed their interest in using corpus-based activities with their prospective students, though 31% of them limited the type of learner in some ways (the most common suggestions here were advanced or specialized courses like ESP or EAP). Over half of the respondents were ready to use DDL techniques without any restrictions.

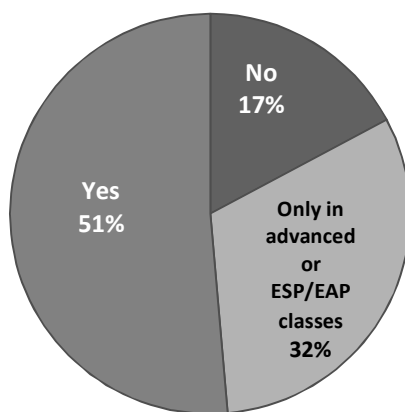


Figure 8. Responses to Item 7 – declared use of DDL techniques in future teaching.

The results obtained from the previous question should be seen mostly as positive feedback on the corpus-based lessons conducted as part of the project. The participants did not have any experience of teaching such lessons themselves, so how their impressions are going to translate into their teaching practice and whether concordances will become part of their teaching repertoire is impossible to establish on the basis of this project.

Finally, from the point of view of learner autonomy it was of particular interest to find out whether using corpora in class affected the students’ own study techniques, and whether they included corpora among their reference tools. Such was the aim of Item 8 in the questionnaire. The results are provided in Figure 9, and it must be said that the very positive outcome for the previous item loses some of its effect when this last set of data is examined.

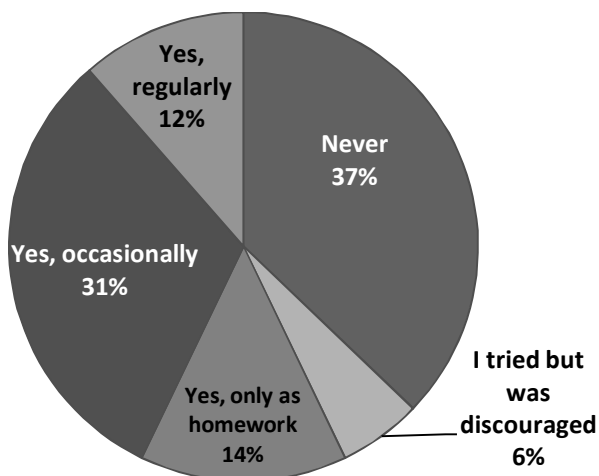


Figure 9. Responses to Item 8 – students' own use of corpora ('Do you ever use a corpus on your own, to check the use of those English words or expressions you have doubts about?')

Admittedly, the results are not very encouraging, with only 12% of the participants declaring regular corpus use for the needs of their studies, e.g. writing their assignments. However, another 31% claim they do it occasionally, so they may still become regular users of the tool, when they find themselves pressured to produce highly accurate English by their supervisor at the university or at work. They know how to access reliable online corpora and how to find the information they need. These two groups, then, can be said to have benefitted from the corpus-based classes in terms of learner autonomy. The others will probably not return to corpora unless another teacher in another course encourages them to do so. The question that remains to be asked is: Was it worth it?

5. Conclusions and implications

The lessons which were part of this project were more focused on language and language data than on the technical aspects of corpus use. They were intended to encourage students to test these instruments on their own and learn how to ask good questions to obtain data that would be most useful to them. Only printed materials were used in class, because that was required by the design of the study. The participants did have an opportunity, however, to become familiar with corpus tools beforehand, so they understood how the materials were generated and where they could find more information if they were intrigued by some of the data they encountered in class.

The survey which was carried out at the end of the project was intended to reveal whether this aim was achieved. From the responses provided one

could conclude that the novel lesson activities were generally successful with the students, and most of them had positive feelings about the lessons. The results also reveal, however, that there are quite a few students (about 1/5 in all) who do not find corpus-based activities useful, and would rather study in more conventional ways. As with all classroom activities, students have their preferences and there is no one ideal format that will please everybody. It is important, however, to provide students with a wide range of choices and allow them to test different techniques, so that they would develop self-awareness in terms of what suits their cognitive and learning styles. Part of learner autonomy is the ability and willingness to make informed choices about their learning. In order to make such choices, they need to have an awareness of the instruments available and of their own abilities. The classroom, then, could be seen as a testing lab, where learners can try out different learning techniques and choose those which they find most appealing and effective. Since corpus-based activities support data analysis and inductive thinking, which are perhaps less frequently employed in the language classroom, it is worth including them in teachers' repertoire. Those students who do not want to be given ready answers and prefer to arrive at solutions by themselves will appreciate the addition.

The number of students who became interested in corpora and started using them on their own should perhaps be bigger to qualify as a return on the time invested in the activities in class, but still there are other benefits: the focus of the lessons themselves, i.e. eliminating particular interference errors. Even though the lessons did not turn out to be *more* effective than the conventional ones (cf. Lewandowska 2013), they did show an improvement in the use of the target forms, so their primary objective was achieved. It seems that at least in the case of this experiment, most of the more skeptical researchers' objections were not confirmed. The participants did not find the activities difficult, and most of them reacted to the lessons positively. It may be that corpus-based activities must be given enough time and focus in class, as was the case here, and should not be used *ad hoc*, as an additional element in a standard language lesson. Another reason why they were generally well received might lie in the status of the students involved: they were not average EFL course participants, but students who had chosen English as their major in a teacher-training college and who were in the final year of their studies. They had enough academic background to face the challenge.

Admittedly, the low results in the section of the questionnaire concerning negative evaluation may be attributed to the fact that the survey was not anonymous. The respondents may not have wanted to admit that they had problems completing the tasks. If this were the case, however, the results for all negatively biased adjectives would have been the same or similar. The results indicate that

the word 'difficult' was the least commonly chosen of all the evaluative adjectives, so one can assume that there is some value to the result. The questionnaire revealed another important aspect of using DDL activities: students' first reactions to them may not be very positive, but this may change after a few lessons. Therefore, the teacher should not give up after first signs of dissatisfaction from students, but let them practice more and develop the skills necessary to analyze language in this way. The results indicate that this is one of those things that take time.

When students leave the classroom for good, they must be ready to find their way around the ever-changing environment of the English language. Corpus reference may be *a* solution, or even *the* solution, for some of them, so they should be given an opportunity to learn how to use it. Further research with larger groups of participants would be needed to establish with more certainty how successful DDL activities really are and whether they enhance learner autonomy to a significant degree.

References

- Aston, G. (ed.). 2001. *Learning with corpora*. Bologna: Athelstan.
- Aston, G. 2001. "Learning with corpora: An overview". In Aston, G. (ed.). 2001. 7-45.
- Bernal, E. and DeCesaris, J. (eds). 2008. *Proceedings of the XIII EURALEX International Congress. 15-19 July*. Barcelona: Universitat Pompeu Fabra.
- Bernardini, S. 2002. "Exploring new directions for discovery learning". In Kettemann, B. and Marko, G. (eds). 2002. 165–182.
- Council of Europe. 2001. *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Conrad, S. M. and LeVelle, K. R. 2010. "Corpus linguistics and second language instruction". In Spolsky, B. and Hult, F. M. (eds). 2010. 539-556.
- Cook, G. 1998. "The uses of reality: A reply to Ronald Carter". *ELT Journal* 52. 57-63.
- Davies, M. 2004. *BYU-BNC. Based on the British National Corpus from Oxford University Press*. (<http://corpus.byu.edu/bnc>) (date of access: 25 Aug 2014).
- Holec, H. 1981. *Autonomy and foreign language learning*. Oxford: Oxford University Press.
- Johns, T. 1991. "Should you be persuaded: Two examples of data-driven learning". In Johns, T. and King, P. (eds). 1991. 1-13.
- Johns, T. and King, P. (eds). 1991. *Classroom concordancing*. Birmingham: English Language Research Journal.
- Kettemann, B. and Marko, G. (eds). 2002. *Teaching and learning by doing corpus analysis*. Amsterdam: Rodopi.
- Kilgarriff, A., Husák, M., McAdam, K., Rundell, M. and Rychlý, P. 2008. "GDEX: Automatically finding good dictionary examples in a corpus". In Bernal, E. and DeCesaris, J. (eds). 2008. 425-432.
- Lewandowska, A. 2013. "The effectiveness of data-driven learning techniques in eliminating Polish advanced EFL learners' interference errors". Doctoral dissertation, Adam Mickiewicz University, Poznań, Poland. (<http://hdl.handle.net/10593/5762>) (date of access: 25 Aug 2014).
- Little, D. 1995. "Learning as dialogue: The dependence of learner autonomy on teacher autonomy". *System* 232. 175-181.
- Little, D. 1999. "Developing learner autonomy in the foreign language classroom: A social-interactive view of learning and three fundamental pedagogical principles". *Revista Canaria de Estudios Ingleses* 38. 77-88.
- Little, D. 2004. "Constructing a theory of learner autonomy: Some steps along the way". In Mäkinen, K., Kaikkonen, P. and Kohonen, V. (eds). 2004. 15-25.
- Mäkinen, K., Kaikkonen, P. and Kohonen, V. (eds). 2004. *Future perspectives in foreign language education. Oulun yliopiston kasvatustieteiden tutkimuksia* 101.

- Pęzik, P. 2012. "Wyszukiwarka PELCRA dla danych NKJP" [PELCRA search engine for the NCP]. In Przepiórkowski, A., Bańko, M., Górski, R. and Lewandowska-Tomaszczyk, B. (eds). 2012. 253-273. (<http://hdl.handle.net/11089/1326>) (date of access: 25 Aug 2014).
- Przepiórkowski, A., Bańko, M., Górski, R. and Lewandowska-Tomaszczyk, B. (eds). 2012. *Narodowy Korpus Języka Polskiego* [The National Corpus of Polish]. Warszawa: Wydawnictwo Naukowe PWN.
- Spolsky, B. and Hult, F. M. (eds). 2010. *The handbook of educational linguistics*. Chichester, England: Blackwell Publishing.
- Widdowson, H. G. 2000. "On the limitations of linguistics applied". *Applied Linguistics* 21. 3-25.