

The Use of Corpus Linguistics to Teach Cognates to Spanish-speaking English Language Learners ¹

MARÍA DEL CARMEN BARRERA COBOS, UNIVERSIDAD DE LAS AMÉRICAS PUEBLA²

Abstract

The rapid growth of technology has made the development of better, cheaper, and more accessible computers possible. As a consequence, the information stored on computers is now more easily available to teachers and students, such as the data obtained from corpora (Wichmann, Fligelstone, McEney, & Knowles 1997). However, despite these rapid advances in technology, the use of corpora in the L2 classroom is still a practice many teachers are not familiar with. Hence, the purpose of this study is to integrate corpora into the L2 classroom, and to get students to act as language researchers through the analysis of concordance listings obtained from a corpus. The participants in this study were native Spanish speakers learning English as a second language at an English language teaching institution. They were encouraged to analyze the data obtained, to come up with their own hypotheses about how language works and behaves, and to interpret and describe the language. The participants were guided through the process of analyzing data obtained from a corpus, they analyzed and interpreted the data without much difficulty, and most of the hypotheses they formulated were confirmed.

Resumen

El acelerado crecimiento de la tecnología ha hecho posible el desarrollo de computadoras más baratas, avanzadas y accesibles. Como consecuencia, la información almacenada en las computadoras, como la que se obtiene a partir de corpus, se ha vuelto accesible a estudiantes y maestros (Wichmann, Fligelstone, McEney, & Knowles 1997). Sin embargo, y a pesar del acelerado avance tecnológico, el uso de corpus en la enseñanza de una segunda lengua es una técnica que muchos maestros no conocen aún. Por lo tanto, el propósito de este estudio es integrar el uso de corpus en el salón de clases de segunda lengua y lograr que los alumnos actúen como investigadores del idioma a través del análisis de concordancias obtenidas de un corpus. Los participantes de este estudio eran hablantes nativos del español, estudiantes de inglés como segunda lengua en una institución de enseñanza del inglés. En este estudio, los participantes analizaron la información obtenida de un corpus, formularon sus propias hipótesis sobre el funcionamiento y comportamiento del idioma, y fueron inducidos a interpretar y describir el mismo. Los participantes fueron guiados en el proceso de análisis de información, y la mayoría de las hipótesis que propusieron fueron correctas.

Introduction

Nowadays, computers have become an important part of our lives. They are becoming smaller, cheaper, and more accessible to teachers and students. As a consequence, the amount of information stored on computers is now more easily available, such as the data obtained from corpora (Wichmann, Fligelstone, McEney, & Knowles 1997). However, despite the rapid growth of technology, the use of computers and corpora in the L2 classroom is not a practice many teachers are

¹ This is a refereed article.

² maria.barreracs@udlap.mx

familiar with. In order to promote the integration of corpora into the L2 classroom, it is necessary for teachers to encourage students to exploit all the tools available to them that can help them learn a language. Students should try to go beyond the language they learn in the classroom and in their books, and one useful tool that can help them learn a language in a more autonomous way is corpora. Students can analyze words, collocations, and sentence structure, and they can get information about the language through corpora.

Johns (as cited in McEnery et al. 2006, p. 99), is believed to be one of the first to realize the potential that corpus linguistics could have in language learning. He argues that research is too serious to be carried out by researchers only, and that language learners should be encouraged to act as researchers with access to linguistic data in order to carry out their learning. According to Kennedy (as cited in McEnery et al. 2006, p.99), the learning of a language is a process of learning knowledge explicitly with awareness. This process requires students to be exposed to language data. There are two types of data-driven learning (DDL): teacher directed or learner led (i.e., when students discover learning), but DDL is mainly learner centered. Leech (as cited in McEnery et al. 2006, p. 99) argues that the DDL is an autonomous learning approach that gives the students the opportunity to act as researchers and to want to make discoveries by making an individual contribution.

Johns (as cited in McEnery et al. 2006 p. 99), also identified the three phases of inductive reasoning with corpora in DDL: observation (of concordances obtained from a corpus), classification (of relevant features of the language), and generalization (of the rules of the language).

In this small-scale exploratory study, the use of corpora in the teaching of vocabulary in English will be implemented in order to promote a more autonomous learning style. In this study, students will be encouraged to act as researchers of the language, and to formulate hypotheses about how the language works through the analysis of data obtained from corpora.

Research questions

Given the growth of technology today, it is necessary that we, as teachers of languages, keep up with its development and try to implement it in the language classroom as much as possible. Hence, the purpose of this study is to incorporate corpora into the L2 classroom, and to get students to act as language researchers. The present study is a small-scale exploratory study that addresses the question of the potential effectiveness of the use of corpus linguistics in the L2 classroom to teach a specific vocabulary building activity using cognates. The main research question that this study will address is:

How effective is the use of corpus linguistics in the L2 classroom?

In an attempt to answer this question, some subsidiary questions will also be addressed:

How effective is it to analyze data obtained from a corpus in the L2 classroom without using computers?

How difficult was the interpretation of the data for the students?

Are students capable of successfully interpreting concordances? Are they able to formulate correct hypotheses about how English works and behaves?

Is it feasible to use corpora to create activities for the L2 class?

The expected answers to all the research questions are affirmative. However, for questions 3 and 4, considering that students are not familiarized with corpus linguistics, they have difficulty interpreting the data and formulating correct hypotheses about how the language works and behaves. Nevertheless, in order to facilitate students' analysis of the data, they will be carefully guided through the instructions on how to use the corpus chosen for this study.

Review of the Literature

With the rapid growth of technology, computers have become cheaper, smaller, and more accessible to people (Wichmann et al. 1997). The growth of technology makes the development of computers that now offer massive storage and an increased processing power at an affordable price possible (McEnery, et al. 2006). According to Wichmann et al. (1997), computers have also become more accessible to teachers and students. As a result, the information stored on computers, such as the data obtained from corpora, is now more easily available to them.

What is a corpus?

McEnery et al. (2006), define a corpus in modern linguistics as "a collection of sampled texts, written or spoken, in machine-readable form which may be annotated with various forms of linguistic information" (p. 4). That is to say, a corpus is a compilation of natural texts stored into a computer that contains interpretative linguistic information that is useful for the analysis of a language. According to Cook (2003), through the systematic analysis of corpora, it is possible to observe the different patterns and regularities of language use.

As Cook (2003) states, before computers existed, printed materials were collected in order to study language. These materials were read laboriously, and facts were recorded by hand. In recent years, however, corpus linguistics has evolved with the help of electronic and automatic searching. Nowadays we are able to search millions of words within seconds to obtain information about word combinations and frequencies.

According to Kennedy (1998), Corpora are compiled for different purposes. Some corpora have been designed for general descriptive purposes in linguistic research. That is to say, they have been designed so that they can be examined in order to answer questions regarding different linguistic levels, such as the lexis, grammar, prosody, pragmatics and discourse patterns of a language. Some other corpora have been designed for specialized purposes, such as deciding which words and word meanings should be included in a dictionary for learners; finding out which words or meanings are most commonly used in a certain domain (e.g. economics); or to discover how language is used differently in a specific geographical, social, historical or work-related context.

There are two main types of corpora, as stated by Kennedy (1998). These are:

General: a text base for unspecified linguistic analysis. It generally includes text of various genres, domains, forms. They are sometimes also called *core corpora*, and they are to be used mainly for comparative studies.

Specialized: corpora designed with a particular research project in mind; e.g., training and test corpora, dialect and regional corpora, spoken and written corpora, and learner corpora.

A brief description of some of the main corpora that can be accessed today is provided below:

Corpora in English:

12. American National Corpus. This is a general corpus available for research and educational purposes. It contains over 22,000,000 words of written (72%) and spoken (28%) American English. It can be accessed at: <http://americannationalcorpus.org/OANC/index.html>
13. Michigan Corpus of Academic Spoken English. Academic, spoken corpus made up of approximately 1.8 million words. This corpus gives access to 152 transcripts, 1,848,364 words and over 190 hours of recorded material. It can be accessed at: <http://quod.lib.umich.edu/m/micase>
14. British National Corpus. This is a general British English corpus. It includes a 100 million word collection of samples of both written (90%) and spoken (10%) language from different sources. It can be accessed at: <http://www.natcorp.ox.ac.uk>
15. Collins, The Bank of English. General corpus that contains 524 million words and continues to grow with the constant addition of new spoken and written material. It can be accessed at: <http://www.collins.co.uk/Corpus/CorpusSearch.aspx>
16. Child Language Data Exchange System. Specialized corpus that contains about 20 million words from spoken language, and it keeps growing. It can be used to study first language acquisition, and language development in children. It can be accessed at: <http://childes.psy.cmu.edu>

Corpora in Spanish:

17. Corpus de la Real Academia Española. General corpus that contains over 200 million words from spoken and written language. It can be accessed at: <http://corpus.rae.es/creanet.html>
18. Corpus del Español. General corpus that contains approximately 100 million words from written language dating from the 1200s to the 1900s. It can be used to study the evolution and development of Spanish throughout the years. It can be accessed at: <http://www.corpusdelespanol.org>

Procedures used in corpus analysis

There are different procedures that can be used to obtain information from a corpus, to search a corpus, or to show, classify and categorize the data that is being investigated. The following are the most commonly used formats:

Wordlists

Kennedy (1998) defines wordlists as lists of the word forms in a certain corpus, displayed in alphabetical order. The number of times each word appears on the corpus (occurrences) is added and presented next to the word. Biber, Conrad, & Reppen (1998) also talk about lists of all the words in a corpus, but refer to them as frequency lists. These lists show the number of times each word occurs in a corpus, and they can be sorted in order of frequency.

Concordances

Concordance listings, or concordances, are displays of the occurrences of a certain word with the context that surrounds it. The chosen word and the context that surrounds it are presented in a single line, in the form of a sentence. The chosen word is displayed in the middle of the line, and the context on each side of it. Usually, concordances display several lines of context. Through concordance listings, it is possible to see the meanings and words related to the word being investigated, as well as how it behaves in a context (Biber et al. 1998).

Statistics

Basic descriptive statistics on the number of word forms, the length of sentences, the number of sentences in a text, and the number of words contained in particular sentences, are often provided in corpora. Statistics are useful for identifying certain features associated with particular text types (Kennedy 1998).

What is a corpus-based approach?

Kennedy (1998) mentions that a great amount of effort has been put into the development of corpora during the last ten years. Corpus-based approaches have introduced new methods to language description through quick and accurate analyses carried out by computers.

According to Biber et al. (1998), a corpus-based approach is empirical; it analyzes authentic samples of language use in natural texts; it uses a corpus in order to carry out the analysis; it uses computers extensively to analyze the data using interactive and automatic techniques; and it depends on qualitative and quantitative techniques for the analysis.

Biber et al. (1998) mention that there are many advantages of using a corpus-based approach. To begin with, computers are able to recognize and analyze large amounts of language than could be done by hand. Also, computers are reliable and consistent since they do not get tired during the analysis. Another advantage is that the human analyst and the computer can work in an interactive way: while the analyst makes difficult linguistic judgments, the computer keeps record of the analysis.

With corpus-based approaches it is now possible to analyze great amounts of data, making it easier for linguists to carry out more studies about how language works. Corpus-based studies can also be applicable to the area of educational linguistics. New materials and classroom activities can be designed with the help of corpora, allowing teachers to provide students with real language that is used in different natural settings (Biber et al. 1998).

The present study is an example of the application of a corpus-based approach to the area of educational linguistics. The activity created for this study is an example

of how concordances can be explored and analyzed in order to identify the meaning of words. Concordance listings were used in this study to help students determine whether some words in English and Spanish were cognates or false cognates.

What are cognates?

Spanish and English are two languages that have many similarities. According to Rodriguez (2001), one of the most evident similarities is that both English and Spanish use the Roman alphabet. Also, these two languages share many cognates. Frunza and Inkpen (2008) define cognates as words in two or more languages, which have a similar meaning and spelling.

Whitley (1986) distinguished 3 different origins of cognates in Spanish and English:

1. Inheritance. Some words in English and Spanish were inherited from Indo-European languages, e.g. mother-madre, six-seis, name-nombre.
2. Coincidence. Some words are not true cognates if their similarity is only a coincidence, e.g. have-haber, much-mucho, other-otro.
3. Borrowing. Occurs when Spanish borrows a word from English, when English borrows a word from Spanish, or when both languages borrow a word from a third language.
 - a. English → Spanish: estándar, boicot, sándwich, láser, líder.
 - b. Spanish → English: ranch, vista, canyon, patio, vanilla, guitar.
 - c. Both from Latin: application-aplicación, exact-exacto.
 - d. Both from French: hotel, control, menu-menú.
 - e. Both from Italian: piano, soprano, bank-banco.
 - f. Both from Greek: map-mapa, diploma, planet-planeta.

Rodriguez (2001) identified 7 different types of cognates in English and Spanish:

1. Words which have an identical spelling in both languages, e.g. hospital, fatal, actor.
2. Words whose spelling is almost the same, e.g. contamination-contaminación, evidence-evidencia.
3. Words whose similarities are not as apparent, e.g. sport-deporte, perilous-peligroso.
4. Words which are cognates in spoken speech (oral cognates) rather than in written speech (written cognates). In other words, they sound more similar than they appear in their written form, e.g. pleasure-placer, peace-paz.
5. Words that have more than one meaning and which are cognates for one meaning, but not for the other, e.g. letter-letra (letter of the alphabet), letter-carta (written correspondence).
6. Words that are similar, and that can be used as reference to teach other words, e.g. disappear-desaparecer, appear-aparecer.
7. False cognates: words which are similar in both languages, but whose meaning is not related, e.g. succeed-tener éxito (not suceder); embarrassed-avergonzado (not embarazada).

Why is it important to know when words are cognates or false cognates?

According to Malkiel (2009), identifying when a word is a cognate or a false cognate is vital for second language learning. Since cognates are words which have the same meaning in two languages, they provide students with what could be referred to as "free" vocabulary. That is to say, students could acquire these words in the second language without much effort.

Teaching cognates

Rodriguez (2001) argues that Spanish speakers know more English than they are aware of. He believes students know more words in their second language than they realize, due to the fact that they are similar in form and in meaning in both languages. Therefore, it is important that cognates and false cognates are taught in the L2 classroom, so that students can become aware of the similarities in vocabulary in both languages, and acquire words in the L2 almost effortlessly.

Rodriguez (2001) suggests that cognates should be used to teach students to guess the meaning from context in a text whenever they come across a word that is similar in form in both English and Spanish. This way, students will analyze the language and will make sense of a text.

Finally, cognates should be used to scaffold students' learning. In other words, teachers should take advantage of what students already know about their L2 without being aware of it.

Methodology

Location

This study was carried out at an English language teaching institution with over 65 years of experience in teaching English as a second language. There are 33 courses in total in this institution, each of them consisting of 30 hours of class work. These courses have been grouped into an introductory level and 6 cycles:

Introductory Level

1. Basic Cycle: levels Basic 1 (B1) to Basic 6 (B6).
2. Intermediate Cycle: levels Intermediate 1 (M1) to Intermediate 6 (M6).
3. Advanced Cycle: levels Advanced 1 (A1) to Advanced 6 (A6).
4. Higher Studies Cycle: levels Higher Studies 1 (HS1) to Higher Studies 6 (HS6).
5. English Mastery Cycle: levels English Mastery 1 (EM1) to English Mastery 4 (EM4).
6. English Proficiency Cycle: levels English Proficiency 1 (EP1) to English Proficiency 4 (EP4).

Participants

For this study, a group of Advanced 3 students was selected. The group consisted of 5 students; however, only 3 agreed to take part in this study. The participants were 3 Mexican students of English as a second language at the chosen institution. From these participants, one was a male student and two were female students. The male student was 20 years old, and the female students ranged in age from 50 to 55 years. All three students were studying the third level of the Advanced cycle

(Advanced 3). At this level, students will have completed 450 hours of effective studying, and will have finished the B1 level of the Common European Framework of Reference for Languages (CEFR). The Advanced 3 level introduces students to the B2 level of the CEFR, and prepares them for B2 level examinations such as the First Certificate in English (FCE).

Materials

Two handouts were created for this study. Handout 1 (see Appendix 2) consisted of 3 pages. Page 1 included some general information about what cognates/false cognates are; it also contained a brief explanation of what corpus linguistics is, a description of the corpus used in this study (Collins, The Bank of English), and a short explanation of what concordances are. Pages 2 and 3 included a step-by-step guide on how to use the Collins corpus. Each step on the handout included instructions, and was illustrated with an image of the Collins corpus website, so that students could see what the webpage looks like.

Handout 2 (see Appendix 3) contained printed concordance listings for each of the words used in this study: *career*, *realize*, *resume*, *eventually*, *actually*, *notable*, *splendid*, *criticize*. In order to facilitate students' analysis, the 10-11 most comprehensible sentences obtained from the corpus were chosen. For the creation of the second handout, the Collins, The Bank of English corpus was used. The Collins, The Bank of English corpus is a collection of modern English language, and can be accessed at: <http://www.collins.co.uk/Corpus/CorpusSearch.aspx> . It was founded in 1991 by the University of Birmingham in the UK, and by Collins. The Bank of English corpus contains 524 million words, and it is continuously growing with the addition of new data. The corpus includes written texts from magazines, fiction and non-fiction books, reports, newspapers, websites, and brochures. It also includes texts from spoken material that comes from conversations, discussions, television, radio broadcasts, interviews, and meetings. The Bank of English is mainly used by Collins lexicographers and linguists that analyze patterns of word combinations, the frequency of words, and the uses of some words in particular, in order to include this information in dictionaries. However, The Bank of English can also be used by language teachers, linguists, translators and students as a tool for their studies and professional activity (HarperCollins Publishers Ltd, 2004).

Procedure

In order not to take up class time, the participants were asked to turn up at the institution 1 hour before their class. A 60-minute lesson was prepared for this study (see Appendix 1), and it was carried out at the institution in order to make students feel at ease in their usual learning environment.

At the beginning of the class, some words in English that could be either cognates or false cognates were written on the board so that students could guess their meaning. The words used in this study were: *career*, *realize*, *resume*, *eventually*, *actually*, *notable*, *splendid*, and *criticize*. From these words, *career*, *realize*, *resume*, *eventually*, and *actually*, were false cognates, and *notable*, *splendid*, and *criticize*, were cognates. The false cognates used were chosen due to the fact that they are problematic words for students. Students use those words very frequently, but they use them as if they were cognates; therefore, they attribute the wrong meaning to

them, and the sentences they produce turn out to be incorrect. The cognates used in this study were randomly chosen.

Students were encouraged to guess what the words presented meant. Their answers could be either a definition in English, or an equivalent word in Spanish. Once students had come up with a meaning for each word, they were asked if they thought the words looked like some words in Spanish. Students agreed that in fact the words were similar to some Spanish words, but they did not know what they were called. In order to familiarize students with the terms *cognates* and *false cognates*, a special handout that included information about cognates was given to them (see Appendix 2). The students went through the first part of the handout, and the teacher explained what cognates/false cognates are. The teacher also raised students' awareness of the importance of knowing when a word is a cognate/false cognate, since they can help them learn more words in English. If students know a word is a cognate, it might be easier for them to remember its meaning. As a result, their range of vocabulary becomes wider. Students were then told that one way in which they can discover the meaning of cognates and false cognates is with the help of corpus linguistics. Handout 1(see Appendix 2) also contained some information about corpus linguistics. Students read a definition of what a corpus is, and a brief description of the corpus (Collins, The Bank of English) that was used to obtain the data needed for the analysis. The last section on the first page of Handout 1(see Appendix 2) included a short explanation of what concordances are, and how they should be interpreted. The teacher and the students went through this last section, and the teacher further explained how to read concordances to the students.

Once students had been given some background knowledge about what a corpus is, they were referred to pages 2 and 3 on Handout 1 (see Appendix 2), and they were guided through the 5 steps of the step-by-step instructions on how to use the Collins corpus to obtain concordances. Students were reminded that the purpose of the lesson was to know whether the words previously written on the board were cognates or false cognates, and that they would use the concordances from the Collins corpus in order to find it out. The teacher gave out Handout 2 (see Appendix 3), which contained the printed concordances, and asked students to work together to analyze and interpret the data obtained from the corpus. They were given 35 minutes to complete this task. The students read the concordances for each word, and tried to identify whether they were cognates or false cognates. When students agreed that a word was a false cognate, they were asked to guess what it meant based on the context that surrounded it. This activity led students to use concordances as a resource to get information about the language, and to formulate hypotheses about how the language works.

When students finished analyzing all the concordances, the teacher checked their hypotheses. The students said whether the words were cognates or false cognates, and what their meanings were. Students were reminded of the meanings they had attributed to each word at the beginning of the class, and they compared them with the meanings they came up with after the analysis. The teacher gave students feedback on their answers, and told them if the words were actually cognates or false cognates, and what they really meant.

Once students' answers had been checked, they were encouraged to reflect on the use of corpora to examine the way language works. The teacher asked some reflection questions about using the corpus and the concordances to analyze language:

- How did you feel while using the concordances from the corpus?
- Do you think using a corpus could help you learn English?
- Did you find it easy to use?
- Apart from guessing the meanings of words, what else could you use the corpus for?

At the end of the lesson, students were reminded of the importance of the use of all the tools available to learn a language. The teacher encouraged students to become language researchers, and to think about the way the language works, rather than simply looking up a word in a dictionary and getting the meaning without analyzing how the word behaves in a context or structure.

Results

At the beginning of the class, students were asked to work in a group and say what they thought the words used in this study meant. Some of the answers students provided were definitions in English, and other were equivalent words in Spanish. These were students' answers:

- Career: what you study at university. In Spanish: una carrera.
- Notable: distinguished. In Spanish: notable.
- Realize: to do something. In Spanish: realizar.
- Resume: to summarize. In Spanish: resumir.
- Criticize: In Spanish: criticar.
- Eventually: very often.
- Actually: in the present, in this moment. In Spanish: actualmente.
- Splendid: a generous person.

It can be seen from the definitions above that students assumed all the words were cognates. Therefore, the meanings students attributed to the false cognates were all wrong. However, it should be mentioned that one of students' answers was quite unanticipated. The meaning students attributed to the word *splendid*, was not the expected one. *Splendid* is translated as *espléndido* in Spanish. However, the word *espléndido* can have two meanings:

1. Splendid, magnificent (HarperCollins Publishers Ltd, 2005).
2. Lavish, generous (HarperCollins Publishers Ltd, 2005).

The definition that students were expected to provide was the first one. Nevertheless, students first thought of the second meaning of the word *espléndido*. This led students to believe that the word *splendid* was a false cognate. However, students were told that the word *splendid* is translated into Spanish as *espléndido*, and its meaning is something which is *magnificent*. Depending on the context in which the word *espléndido* is used, the word *splendid* can either be a cognate or a false cognate.

The definitions of what cognates and false cognates are, and the step-by-step instructions on how to use the corpus (see Appendix 2), were carefully explained in order for students not to be overwhelmed by so much information. However, as the teacher presented all the new information, students seemed a little confused: they were frowning and looking at each other and at the teacher, as if they could not understand what the teacher was saying. Nevertheless, when Handout 2 (see Appendix 3) was given out, and students were asked to analyze the words, they understood the instructions, and started working without difficulty. During the 35 minutes in which the analysis took place, students interpreted the data obtained from the corpus, and used concordances as a resource to get information about the language. Students were talking to each other about the possible answers, they were very engaged in the activity, and they negotiated the meaning of words. Each student participated actively in the task, and they all expressed their own opinions and hypotheses about the way words behaved in a specific context. The conversations that students had, and the decision-making process that they went through were both very interesting. Unfortunately, the class was not recorded and students' discourse could not be analyzed.

Students' answers and hypotheses about the language were checked once they had finished analyzing the words. The effectiveness of the use of concordances to teach cognates was measured based on how many correct hypotheses students formulated. Since students seemed a little confused at the beginning of the analysis, the possibility of students' answers being incorrect was considered a potential result. Surprisingly, and contrary to expectations, most of the students' hypotheses and answers were correct. Students were able to identify which words were cognates, and which words were false cognates. However, in some cases it was difficult for them to say what the exact meaning of the words was. The answers that students provided after analyzing the data were the following:

- Career: false cognate. Meaning in English: an occupation. Meaning in Spanish: una profesión o vida profesional.
- Notable: cognate. Meaning in Spanish: notable.
- Realize: false cognate. Meaning in Spanish: darse cuenta de algo.
- Resume: false cognate. Students were not able to identify the meaning of this word.
- Criticize: cognate. Meaning in Spanish: criticar.
- Eventually: false cognate. Students were not able to identify the meaning of this word.
- Actually: false cognate. Meaning in Spanish: de hecho.
- Splendid: cognate/false cognate depending on the meaning in Spanish. Splendid is a cognate when it means espléndido in Spanish, but a false cognate when it means una persona generosa.
- The teacher helped the students with the two words that they could not guess the meaning of, by going through the concordances again. Students' attention was drawn to the two or three most comprehensible sentences in the concordance listings that provided evident clues. One more time, students were encouraged to try to guess the meaning of the words.

Students kept talking to each other and discussing the possible meanings of both words. However, they were still unable to come up with a final answer. The teacher then helped students by telling them what the words really meant:

- Resume: false cognate. Meaning in English: to start again. Meaning in Spanish: retomar, reanudar.
- Eventually: false cognate. Meaning in English: finally. Meaning in Spanish: finalmente.

The teacher reminded students of the meanings they had attributed to the words at the beginning of the class, and they compared them to the meanings they provided after they had analyzed the concordances. The students were really surprised by how different the meanings were, but because they witnessed how the words behaved in a context, they were convinced that their answers were correct, and they realized that they do not always have to depend on the teacher for getting the correct answer. Students were also surprised because they realized they had used the words incorrectly for a long time. At the same time, they were happy that they had now learnt the correct meaning of each word, and how to use them.

At the end of the lesson, the teacher asked some questions in order to encourage students to reflect on the use of corpora to analyze the way language works. Students mentioned that at first they were a little nervous about the analysis because they thought it was going to be a difficult task to carry out. However, once students started reading the concordances, they felt comfortable because they understood what the sentences said, and they were able to analyze them without difficulty.

When asked if they thought using a corpus could help them learn English, students mentioned they thought it could be very useful. Students said they found concordances very helpful, since they are examples of real language produced by native speakers of English. They also found the analysis of concordances to be more interesting than simply looking up a word in a dictionary. By analyzing concordances, students said they were able to look at a word in context, and to guess what it meant. They mentioned that the analysis of concordance listings could help them remember words more easily because they become aware of how they are used in the language, instead of just being given the meaning of words as in dictionaries.

Finally, students were asked if they could think of something else they could use the corpus for, and they contributed with some of their ideas. Students said they could use the corpus to see how phrasal verbs are used, to get more examples of comparative and superlative adjectives, and to find out the meanings of new words without looking them up in a dictionary.

Conclusion

The purpose of this small-scale exploratory study was to integrate corpora into the L2 classroom, and to get students to act as language researchers through the analysis of concordance listings obtained from the Collins, The Bank of English corpus. In order to achieve the purpose of this study, a one-hour lesson to teach cognates and false cognates in English was prepared.

There were 5 research questions that this study addressed. The main question was:

How effective is the use of corpus linguistics in the L2 classroom?

The results obtained in this study suggest that the use of corpus linguistics in the L2 classroom is effective, indeed. Despite the fact that the students who participated in this study had never heard of corpus linguistics before, they felt comfortable analyzing concordance listings and hypothesizing about how words behave in the language. Students understood the instructions, and analyzed the data without difficulty.

This study also addressed some subsidiary questions that attempted to answer and elaborate on the main question:

How effective is it to analyze data obtained from a corpus in the L2 classroom without using computers?

Even though corpora are collections of texts stored into computers, no computers were used in this study due to the fact that there were not enough computers for students to work with at the English language teaching institution where they study. However, having the step-by-step instructions on how to use the corpus, and the concordances printed out, turned out to be very effective for the vocabulary activity that was prepared for this study. In order to access the Collins, The Bank of English corpus, students need to have an internet connection. This could lead students to feel overwhelmed when asked to access the corpus online, due to the amount of information available to them on the web. Nevertheless, since all the materials used in this study were printed out, the amount and the type of information presented to students was a 100% controlled. Even though students did not access the corpus online, they understood how to interpret concordances with the help of Handout 2 (see Appendix 3) and the teacher's explanation; they were engaged in the activity; they paid attention throughout the class, and were successful in carrying out the activity.

How difficult was the interpretation of the data obtained for the students?

At first students seemed a little confused and overwhelmed by so much information that was presented to them. However, once students read the concordances and realized they were able to understand the sentences, they started interpreting the data without difficulty. This can be partly attributed to the fact that the best concordance lines were chosen beforehand, and were printed out instead of using the raw data from the computer. The only problem students encountered was that they could not identify the meaning of two of the given words.

Are students capable of successfully interpreting concordances? Are they able to formulate correct hypotheses about how English works and behaves?

Because students were not familiarized with corpus linguistics at the beginning of the lesson, there was a probability that students might find it difficult to interpret the data and formulate correct hypotheses. Nevertheless, the majority of students' answers were correct, and the hypotheses they formulated were confirmed.

Is it feasible to use corpora to create activities for the L2 class?

It is highly recommended that specific language activities are created when using corpora in the L2 classroom, since they guide students throughout the activity, they draw the attention of students to the task, and they engage them throughout the class. When using corpora in the L2 classroom, it is crucial that teachers have a clear objective that they want to achieve by the end of the class. If no activities are created, students will be overwhelmed by the amount of information they will come across on the internet. This could lead students to feel frustrated, and to not want to learn the language anymore.

The integration of corpora into the L2 classroom demands the teacher to create specific language activities, rather than just giving students free access to the information. A corpus is a complex database that requires some skills in order to interpret the data obtained, e.g. one needs to know how to read frequency lists, concordances, and statistics. Therefore, it is necessary to guide students through the process by which they will be able to analyze data (Gavioli 1997). That is to say, in order for students to be successful at interpreting the data, they need to be taught how to read concordances, frequency lists, and statistics.

This study intended to get students to act as language researchers through the analysis of concordance listings obtained from a corpus. Corpora provide data that needs to be read, to be analyzed, or to be interpreted. Therefore, the purpose of this study was to encourage students to analyze the data, and to come up with their own hypotheses about how language works and behaves. Through the analysis of concordances, students were led to interpret and describe the language, rather than just looking up the words in a dictionary and getting the meaning without making any effort to understand how language functions.

The use of corpus linguistics in the L2 classroom is an important issue for future research. This small-scale exploratory study addressed the question of the potential effectiveness of the use of corpus linguistics in the L2 classroom to teach a specific vocabulary activity. In this study, the use of corpora to teach a vocabulary building activity turned out to be effective. The students were successful in carrying out the activities using data obtained from a corpus. However, it is recommended that further studies are undertaken in order to develop a complete teaching approach or method that language teachers can follow when using corpora in the L2 classroom. Finally, it should be taken into consideration that in order for teachers to be successful at integrating corpora into the L2 classroom, it is necessary that they are trained on how to use a corpus so that they can transfer those skills to students.

References

- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Cook, G. (2003). *Applied linguistics*. Oxford University Press.
- Frunza, O., & Inkpen, A. (2008). Disambiguation of partial cognates. *Language Resources and Evaluation*, 42(3), 325-333.
- Gavioli, L. (1997). Exploring texts through the concordancer: Guiding the learner. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and language corpora* (pp. 83-99). Longman: London and New York.
- HarperCollins Publishers Ltd. (2004). *The Bank of English, 2004* [electronic corpus]. Available from Collins Web Site, <http://www.collins.co.uk/books.aspx?group=153>

- HarperCollins Publishers Ltd. (2005). Espléndido. In *Collins Spanish-English dictionary* (Eight Edition, p. 423). HarperCollins Publishers.
- Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.
- Malkiel, B. (2009). Translation as a decision process: Evidence from cognates. *Babel*, 55(3), 228-243.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-Based language studies: An advanced resource book*. New York: Routledge.
- Rodriguez, T. A. (2001). From the known to the unknown: Using cognates to teach English to Spanish-speaking literates. *Reading Teacher*, 54 (8), 744-746.
- Whitley, M. S. (1986). *Spanish/English contrasts: A course in Spanish linguistics*. Washington, D.C.: Georgetown University Press.
- Wichmann, A., Fligelstone, S., McEnery, T., & G. Knowles (Eds.). (1997). *Teaching and language corpora*. Longman: London and New York.