

# LANGUAGE TEACHING BEFORE AND AFTER 'DIGITALIZED CORPORA'. THREE MAIN ISSUES

Aquilino Sánchez  
Departamento de Filología Inglesa  
Universidad de Murcia

## ABSTRACT

*Digitalized corpora are becoming increasingly important for linguistic research and are already well known by some language teachers. The power of PCs and Internet will soon make them more popular and accessible to everybody. This paper will attempt at pointing out the role that corpora might play as complementary tools for language teaching in three areas: vocabulary lists, concordances and constellation of words. None of them is completely novel and they have been already present in language teaching materials in one way or another. Modern digitalized corpora constitute a powerful tool and allow for more varied, reliable, refined and easy to get materials in the three fields mentioned. In one of them specifically, the constellation of words, more research and practical work is still needed in order to bring the results closer to textbooks authors and teachers.*

**KEY WORDS:** *vocabulary lists, concordances, language teaching, computer-aided language teaching, lexical constellations*

## RESUMEN

*Los repertorios digitalizados de textos, o corpus lingüísticos, se muestran cada día más necesarios en los estudios lingüísticos y ya son conocidos por bastantes profesores de lenguas extranjeras. Tanto Internet como los PC los harán aún más populares y accesibles. En este trabajo se intenta poner de relieve el valor de los corpus en la enseñanza de lenguas como ayuda complementaria, especialmente en tres campos: los listas de vocabularios, las concordancias y la elaboración de constelaciones de palabras. Ninguna de estas tres áreas es totalmente nueva: todas ellas han estado presentes en los materiales docentes de una u otra manera. Los modernos repertorios de textos digitalizados constituyen una herramienta poderosa que hace posible el acceso a materiales más variados, fiables y elaborados. En uno de esos campos, el referido a las constelaciones de palabras, es preciso aún seguir investigando para hacer que los resultados puedan llegar más fácilmente a los profesores y autores de materiales didácticos.*

**PALABRAS CLAVE:** *listados de vocabulario, concordancias, enseñanza de lenguas, enseñanza de lenguas asistido por ordenador, constelaciones léxicas*

## I. INTRODUCTION

The field of Corpus Linguistics is a new one. still, much has been written on it. Many books and articles have also been published on a related subject. Corpus Linguistics in Language Teaching. My main purpose in this paper is to approach language teaching from the perspective of linguistic corpora and analyse with some detail three issues relevant to both disciplines: vocabulary lists, concordances and word association. Modern technology has made it possible for teachers to use or take advantage of data and results that large amounts of linguistic materials can offer.

The usefulness of corpora as a resource for language teachers and language learners is well illustrated in the literature (Tribble 1990; Flowerdew, 1993; Tribble & Jones, 1997; Aston 1995; Botley, Glass, McEnery and Wilson, 1996, etc.) from different perspectives.

Corpora might be useful to take some important decisions, as for example which lexical items should be taught or should be taught first, 'Economy' and 'efficiency', in fact, matters in language learning.

Corpora might be useful to illustrate linguistic patterns or linguistic usage, so that students are not always given the same nonsense examples whenever rules or models are explained.

Corpora offer an excellent basis for creating authentic exercises (fill-in, substitute, complete, reconstruct missing words or structures, etc.).

Corpora and corpus-based exercises might be useful to favour learning by discovery" in any linguistic field (grammar, vocabulary, style). In this case, learning might take the character of research, far away from offering ready-made recipes or rote learning. Even teacher has to decide what he is looking for and consequently choose which is the approach that better fits his goals.

Corpora favour the use of words and phrases within its own context, as well as the investigation and identification of which words usually keep company with other words in real communication and language use. Such 'situational lexical clusters' will be needed for creating interactive activities and texts closer to what native speakers do when they communicate.

Up till now corpora have not been easily available to the language teachers. And those having access to them may experience some difficulties in retrieving relevant and useful information from large quantities of linguistic data (several hundred thousand or several million words). The reasons for such difficulties have to do with money (expensive corpora and expensive software), technical skills (software demands too much from users or it lacks transparency) or legal problems (copyright of texts being reproduced). However, the rapid development of computer capabilities and telecommunications (internet) contribute in solving partially this problem. But it is not easy to cover all the needs language teachers have when they decide to take advantage of corpus resources.

## II. VOCARULARY LISTS

Vocabulary lists were well known by teachers and students of foreign languages before corpus linguistics became an autonomous discipline. We know about the use of vocabulary lists in language teaching since early centuries (see Sánchez 1992:31ff). Vocabulary lists and phrase books became especially popular in the sixteenth century (Sánchez 1997:49), when printing was discovered. Interest in them has remained high until nowadays. And its elaboration has also been significantly refined. The question with these lists is how reliable they are, that is, how well they represent real language usage. More often than not they just reflect the author's point of view and his own linguistic and subjective feelings about language and the frequency of the words listed. In fact, how could it be otherwise? Computer facilities for processing and manipulating huge amounts of linguistic data are quite recent and this is precisely what frequency counts require.

In 1904, J. Knowles produced the *London Point System of Reading for the Blind*. The author took 100.000 words from the Bible and other sources and extracted the 353 most common words. Pronouns, prepositions and other grammatical words made the 72 first words in the list. He also claimed that these 353 words represented 3/4 of all the words in the corpus (which matches fairly well the conclusions we might come to nowadays on the basis of large and reasonably representative corpora). A few years later, in 1914, Cook and O'Shea, in the United States, elaborated a corpus with the correspondence of thirteen adults. He also concludes that only 9 words accounted for more than a quarter of the total amount of words used: 43 accounted for a half of the total and 763 made 90% of the whole text. Those conclusions attracted the attention of lexicographers and educators. The teaching and learning of reading, spelling and general use of language might gain in efficiency if emphasis shifted to items with a heavier weight in the linguistic system.

In 1921, E. L. Thorndike published *The Teacher's Word Book*, with a list of the most frequent words in English. Those words were therefore more likely to be found by students when they engaged in reading. And that led to the conclusion that such a list was relevant for the classroom in the sense that it should be given priority in teaching and learning. Thorndike's corpus derived from 41 different sources, with a total of 4 million words. The Bible was by far the most important source, as it had already been done by Knowles; together with English classics, it accounted for 3 million words; half a million was extracted from letters, 300.000 from school readers (elementary), 90.000 from newspapers and the rest, 50.000, from other general sources. Thorndike's *Word Book* became a sound and respected reference for setting out vocabulary lists in school books and spellers (already very popular in the United States). No doubt, the resulting list pays lip service to the cultural bias of the times (the Bible and the English classics account for 3/4 of the texts on which the list is based). The objectivity of the results is filtered and constrained by the subjectivity of the criteria that underlie the design and elaboration of the corpus. In spite of it, the list was a

breakthrough in defining the frequency of words and soon became a scale against which the adequacy of vocabulary lists taught might be measured.

Vocabulary counts are not free from serious criticism (McArthur 1998:54 ff). More representativeness of the base materials is needed. Which meaning (if there are several) is implied by the most frequent words listed is also a problem to cope with (if 'bank' appears among the 700 more frequent words, which one of the different meanings of 'bank' is to be accounted for?). Thorndike was conscious of the problem and he embarked, together with Lorge, in the project of counting not just the word, but the word and the sense implied in its use. For that purpose they relied on the senses registered in the *Oxford English Dictionary*. In 1944, a new book was published by both authors: *The Teacher's Wordbook of 30,000 Words*.

Word lists gained popularity and credibility. They soon became a much praised tool for language teachers and particularly for textbooks writers and curriculum designers. Michael West worked also on vocabulary lists and wanted to incorporate them in his *New Method Readers*, prepared for foreign learners of English. West ended up with 1,490 words, which he took as 'defining items'. An *International Reader's Dictionary*, with 24,000 entries, is a practical application of his method: definitions are written using only the 1,490 most common words of English, which he had previously investigated and identified.

Methods for elaborating basic vocabulary lists were the subject of bitter discussions at that time, since full objectivity could not be achieved and subjectivity was too dangerous. A kind of compromise was reached in a new book, *A General Service List*, published by M. West in 1953. The author compiled a list of the 2,000 most frequent words of English. Following Thorndike, the list had been extracted from a 5 million word corpus. It was widely accepted as a model and considered a necessary reference and a scale for the elaboration of syllabuses for the teaching of English to foreign students. West's list incorporates important elements that had been the subject of discussion in the preceding years among 'basic vocabularists', particularly a detailed specification of the senses of each word and the percentage of uses accounting for every one of the senses. No doubt, this was an important contribution in the field.

Basic vocabulary lists were deemed useful and investigations were conducted in other languages as well, aiming at similar results. Such an undertaking was initiated in Spanish by Victor Garcia Hoz. The results were published in 1953: *Vocabulario usual, común y fundamental*. Garcia Hoz's procedure goes very much in line with Thorndike and Lorge, even if he criticises some of their variants, particularly regarding 'massive' vocabulary selection for elaborating a final list. He concludes with three different levels in word usage:

*trivial* (12,402 words),

*common* (1,971 words) and

*basic* (208 words).

Common and basic words together roughly equal the 3.000 words given by West in his *General Service List*. As expected, the *Vocabulario usual* is extracted following specific criteria, affecting the method and the selection of textual sources. Garcia Hoz does not agree with the need for a huge corpus: "You register many words, but not all of them can be named 'usual'. In fact, many of them are not usual" (Garcia Hoz 1953:18). He limits the corpus to 400.000 words, a figure notoriously lower than the five million words on which Thorndike based his analysis.

This amount was extracted from four different areas of 'life':

1. Family life, collected from 620 private letters.
3. 'Non specific' social life, as given in various newspapers.
3. 'Regulated' social life, as given in official documentation related to the political, religious and labour aspects of citizens.
4. Cultural life, as reflected in various books, the most widely read in those years.

Every one of those four sectors of 'normal, daily life' were assigned 100,000 words, which makes a total of 400,000. It is important to notice that García Hoz established a limit to the amount of text taken from each one of the sources. That allowed for a greater variety in the sources and a greater variety of linguistic forms, since the rate of new words in a text decreases parallel to the increase in quantity of the same text. Finally, the author selected the 13.913 more frequent items, with an average frequency of 100 occurrences. 'Common words' were selected attending to the criterion of 'words occurring in the four sectors of life and with a frequency not lower than 40 in each one of them'. Words meeting that condition are only 1.947. 'Basic vocabulary' was applied a similar criterion, increasing the occurrences in the four sectors to 400 instances or more. This condition is met by 208 words.

McArthur (1998) points out some of the problems inherent to *Basic Vocabularies*. No doubt, they must be taken as relative tools or references. The corpus underlying those lists in the first half of the century are affected by a serious lack of representativeness of the sources; texts selected are limited to a few sectors of linguistic usage, clearly biased by what was supposed to be 'good and morally accepted linguistic performance' (the Bible, the (English) classics; legal documents, religious writings). An important effort was made however to include 'family language', as shown in familiar or private correspondence, perhaps the closest you might find as a substitute for oral language, which was otherwise absent.

The relative nature of *Basic Vocabulary lists* has often been stressed as a negative ingredient of Thorndike's or Garcia Hoz's vocabularies. This is true, but we must bear in mind that this 'relative character' will always be present in any frequency lists. In fact, frequency lists can only claim to be as representative as the corpora from which they are extracted are representative. Corpora do nothing but reflect the language being used in the particular and specific moment in which they are built. In so far as linguistic usage changes or varies, frequency lists will also change and vary. Basic vocabulary lists necessarily enjoy a

rather short life. This fact can be well illustrated in the following chart. in which the lists of the most frequent words given by García Hoz are compared with the list of the most frequent words in the *Cumbre corpus*. (a corpus of contemporary Spanish) at the stage of 9 million words:

	<b>García Hoz</b> (* = not present in the 'Cumbre' corpus)	<b>Corpus 'Cumbre'</b> (** = Not present in Garcia Hoz's)
<i>most frequent 2,000 words (in alphabetical order)</i>	<b>A</b>	A
	<b>Abajo</b>	Abajo
	<b>Abandonar</b>	Abandonado*; Abandonar
	<b>Abarcar*</b>	Abandono**
	<b>Abierto</b>	Abierto
	<b>Abogado</b>	Abogado
		Abordar**
		Aborto**
		Abrazo <sup>LL</sup>
	<b>Abrazar*</b>	
	<b>Abrigo*</b>	
	<b>Abril</b>	Abril
	<b>Abrir</b>	Abrir
	<b>Absolutamente</b>	Absolutamente
	<b>Absoluto</b>	Absoluto
	<b>Absorber*</b>	
	<b>Absurdo</b>	Absurdo
		Abuelo**
	<b>Abundante</b>	Abundante
	<b>Abundar'</b>	
	<b>Aburrido*</b>	Abuso** Acá** Acabado <sup>L*</sup>
	<b>Acabar</b>	Acabar
	<b>Academia</b>	Academia
	<b>Acarrear*</b>	Académico**
	<b>Acaso</b>	Acaso
	<b>Acceder</b>	Acceder
		Acceso**
		Accidente <sup>1</sup>
	<b>Acción</b>	Acción
		Accionista**
	<b>Aceite</b>	Aceite
	<b>Acelerar<sup>x</sup></b>	
		Aceptación**
	<b>Aceptar</b>	Aceptado <sup>t*</sup>
		Aceptar
	<b>Acercar</b>	Acerca <sup>1</sup>
	<b>Acertado*</b>	Acercar
	<b>Acertar*</b>	Acero**
	<b>Acierto*</b>	
	<b>Aclarar</b>	
	<b>Acoger<sup>x</sup></b>	Aclarar
	<b>Acogida*</b>	
	<b>Acompañado</b>	

<b>Acompañar</b>	Acompañado
<b>Aconsejar*</b>	<b>Acompañar</b>
<b>Acontecer*</b>	
<b>Acontecimiento</b>	<b>Acontecimiento</b>
	<b>Acordado**</b>
<b>Acordar</b>	<b>Acordar</b>
	<b>Acorde**</b>
<b>Acostumbrar*</b>	<b>Acostumbrado</b>
	<b>Acta**</b>
<b>Actividad</b>	<b>Actitud**</b>
<b>Activo</b>	<b>Actividad</b>
<b>Acto</b>	<b>Activo</b>
	<b>Acto</b>
	<b>Actor**</b>
<b>Actual</b>	<b>Actuación**</b>
<b>Actualidad</b>	<b>Actual</b>
<b>Actualmente</b>	<b>Actualidad</b>
<b>Actuar</b>	<b>Actualmente</b>
<b>Acudir</b>	<b>Actuar</b>
<b>Acuerdo</b>	<b>Acudir</b>
<b>(Total: 49)</b>	<b>Acuerdo (Total: 57)</b>

If we arrange the first 7.000 most frequent words from both lists in alphabetical order. in the range up to 'acuerdo'. García Hoz accounts for only **49** words. while the *Cumbre* corpus includes 57. Furthermore. out of those **49** words. 16 of García Hoz's do not appear in *Cumbre*. while 73 in the *Cumbre* list are not included in García Hoz's. In this same range. therefore. only **33** words are present in both lists. which roughly corresponds to 2/3.

If we enlarge the range of comparison. many other words from *Cumbre* are not present in García Hoz:

*acusación, acusado, adiós, adolescente, adoptar, adulto, aéreo, aeropuerto, agenda, agente, agujero, ajuste, alarma, algodón, amable, amanecer, amante, asesinato, asesino, Úrea*. etc.

And many words collected by García Hoz are not found in the corpus *Cumbre*:

*adjudicar, adorar, adornar, agilidad, agrupación, ajedrez, alabar, alejar, alojar, alumbrar, amargo, anciano, apreciar, arca, armur, artillería, ascender, atento*. etc.

Some of them are clearly connected to the period in which each corpus was compiled. Among them are: *aborto, accionista, adoptar, aéreo, aeropuerto, alarma, amante, asesino*. etc. These words do not seem to belong to the core of the cultural environment of Spain in the early fifties. For a similar reason. words such as: *adorar, alabar, alumbrar, arca, artillería, ascender*. etc. (relevant in religious and military environments) do not appear in the *Cumbre* corpus. elaborated in the early nineties. when the Spanish society has deeply changed with regards to these matters. No doubt many other words are not present in both

lists as the criteria and procedure underlying both corpora differ significantly. If we just pay attention to the range of coverage, the *Cumbre* corpus (Sánchez & al. 1995) is much more ambitious and representative of the linguistic outcome and use than García Hoz's was.

The elaboration of a basic vocabulary of French was also the goal pursued by the Ministry of Education in France in the early fifties. A committee was nominated and a novel criterion was applied for building the corpus: linguistic samples were to be taken from oral language only, as produced in daily communication. The result of this investigation was *Le Français Fondamental*, a basic vocabulary of French. The procedure was as follows: a certain amount of oral samples were taped (173), which resulted in a corpus of 312.135 words. A total of 1.073 items occurred more than twenty times. Later on, this amount was reduced to a core of 805 words, since some proper nouns and other erroneous repetitions were rejected. The commission added to the list the result of a second corpus based on the 20 words given by different speakers as the most important ones in the language. The final list was taken by the CREDIF (*Centre de Recherche et d'Étude pour la Diffusion du Français*) as the standard list on which all manuals for teaching French as a foreign language should be based on.

The authority of basic vocabulary lists settled down in the teaching of foreign languages until up to the nineties and was a 'must' that could not be left out by the authors of textbooks and teaching materials. Communicative methodology overshadowed later on the leading role of basic vocabularies. Syllabuses were mainly enriched by lexical needs derived from real communicative situations. The importance of basic vocabulary lists remains however as one of the key achievements of structurally based methodologies.

As already mentioned, the compilation of basic vocabulary lists was based on frequency counts of running words extracted from various written texts and/or oral recordings. Lists elaborated before and after 'digitalized corpora' share this feature. Differences before and after corpora and 'corpus linguistics' refer rather to the accuracy and reliability of the final lists we might come up to. The technical means we have nowadays at our disposal play an important and decisive role in establishing differences. Thorndike's corpus, with 5 million running words, is an important achievement in the history of corpora: but frequencies had to be calculated manually and that takes too much time, too many assistants and a lot of money. Moreover, this kind of difficulties condition the design itself and might keep the attention away from other important problems, as the range, variety and representativeness of the samples gathered. Written texts were available, but not as comfortably as they are nowadays. Oral texts were extremely difficult or impossible to collect until the tape recorder made this task easy and affordable. In addition to that, design criteria were heavily biased or conditioned by non-linguistic factors (of a moral, religious or aesthetic nature). The fact that the Bible and the classics counted for 315 of Thorndike's corpus cannot be understood without the presence of those constraints. A similar case applies to García Hoz's Spanish corpus.

The procedure for compiling basic vocabulary lists from textual corpora has not varied significantly, but new technologies in the processing of texts and computation might indeed improve and refine the searches. The compilation of corpora themselves has improved. An increase in the scale of representativeness of corpora grants frequency counts more reliability. Corpora usually include a wider variety of texts, selected on more neutral and scientific grounds. Computational facilities help their manipulation and handling, allowing for a quick processing of data. This situation helps to popularise their use or to approach them with more confidence. In addition to that, the technical means available nowadays make it possible to update corpora with minor efforts, which is necessary in order to keep corpora chronologically alive, valid and reliable. Frequency lists need to run parallel to linguistic usage in the communicative process of any community. For how long a frequency list is valid is an open question and depends also on the characteristics and purpose of its users. Ten to twenty years seems a sensible period of time relevant to lexical change and therefore that would be perhaps a reasonable reference for updating vocabulary lists, especially if they are going to be used for textbooks or other teaching materials.

### III. VARIETY OF FREQUENCY LISTS IN MODERN CORPORA

Frequency lists before 'corpus linguistics' were just 'basic vocabulary' counts. Useful as they are, this is not the unique possibility we have nowadays. Computers have significantly increased the range of possibilities. Teachers and students may have access to them by simply keystnking a few commands.

We tend to believe that software applications are more limited than they actually are. Even if computational processing must adjust to some strict and logical requirements and procedures, this narrow logical frame allows for some variety and flexibility, and sometimes the nature of such a variety might be quite 'generous'. The computer is able to offer various kinds of lists based on segments of letters or 'set of letters kept between blank spaces' (which we call words).

An alphabetical list of forms can be quickly changed into a 'reverse alphabetical list' or into a list of words ordered by frequency of occurrence. With similar ease there is also the possibility of arranging together words that contain a specific sequence of letters, at the beginning, at the end or in-between the words. The classroom and the teacher can benefit from this kind of information for grammatical or orthographic purposes. The availability of examples and materials for activities and exercises is always there available. If the corpus is morphologically tagged, the possibility of searching with greater specificity (lemmas, parts-of-speech words, inflections of verbs, mood, number, etc.) increases significantly. Those are some examples readily available, taken from a Spanish corpus:

## Special Sequences:

**-ZC-**

aborrezca  
 aborrezco  
 amanezca  
 aparezcan  
 convalezco

**-RR-**

arranque  
 arranqué  
 arranques  
 Arróspide  
 Arróspide  
 cerrarse  
 cerrazón  
 cerré  
 cerrerías  
 cerril  
 cerros  
 cierra  
 cierran

**-QU-**

alambíque  
 alfeñique  
 alharaqueando  
 almanaque  
 aquí  
 Arlequin  
 arlequín

**-SP-**

asperezas  
 asperjada  
 asperjados  
 áspero  
 aspiraciones  
 aspiramos  
 aspiran  
 aspirando  
 aspirante  
 aspirantes  
 aspirar  
 aspiró

**-AD**

representatividad  
 prioridad  
 privacidad  
 probabilidad  
 productividad  
 profesionalidad  
 profundidad  
 progresividad

propiedad

**HIPER-**

hiperactividad  
 hiperbóreo  
 hiperinflación  
 hipertensos  
 hipertrofia

**RETRO-**

retroceder  
 retrocedía  
 retrocedido  
 retrocediendo  
 retrocediera  
 Retrocedió  
 Retroceso

Verbs in -AR  
 (ordered by

**frequency):**

Acabar  
 Recabar  
 Alabar  
 Grabar  
 Trabrar  
 Derribar  
 Silbar  
 Derrumbar  
 Tumar  
 Retunibar  
 Zumar  
 Embobar  
 Englobar  
 Robar  
 Jorobar  
 Probar  
 Aprobar

Specific irregular  
 verb forms (-tie-):

Abstiene  
 Acometiendo  
 Acometieron  
 Admitiendo  
 Admitiera  
 Admitieron  
 Advirtiendo  
 Advirtiera  
 Advirtieron  
 Combatiendo  
 Cometiendo  
 Conietieron  
 Compartiendo

Compartiera

Detiene  
 Detienen  
 Dimitiendo  
 Discutiendo  
 Discutieron  
 Distiende  
 Divirtiendo  
 Emitiera  
 Emitieron  
 Entienda  
 Entiende  
 Entienden  
 Entierra  
 Omitiendo  
 Partiendo  
 Partiera  
 Peniiitiendo  
 Permitiese  
 Persistieron  
 Sometiera  
 Sometieron  
 Sostiene  
 Sostienen  
 Tiene  
 Tienen

**Reverse** ordered  
 word lists:

a  
 acaba  
 acercaba  
 señalaba  
 hablaba  
 brillaba  
 pasaba  
 pensaba  
 Necesitaba  
 calentaba  
 estaba  
 Córdoba  
 biblioteca  
 dedica  
 indica  
 Bélgica  
 tónica  
 Guernica  
 Guernica  
 informática  
 música  
 Practica  
 práctica  
 blanca  
 Blanca

cerca  
 derecha  
 aprovecha  
 escucha  
 mucha  
 cada  
 enfadada  
 nada  
 preocupada  
 preparada  
 entrada  
 cansada  
 encantada  
 asustada  
 adecuada  
 situada  
 hünieda  
 bebida  
 sorprendente  
 comida  
 Avenida  
 vida  
 falda  
 espalda  
 ienda  
 Ronda  
 toda  
 izquierda  
 idea  
 línea  
 Andrea  
 desea

#### IV. CONCORBANCES AND CONCORDANCERS

Frequency lists applied to language teaching raise some serious questions when you look at them closely or when you attempt a more detailed analysis of what they imply and mean. Forms alone are a valuable help for teachers when they face the question of 'what to teach'. But mechanical counting alone fails to introduce or add any further specificity to word lists. Semantic annotations added to the most frequent words by some scholars (Thorndike & Lorge, Palmer, Ogden) help in the discrimination of senses, but they are still unable to help in relating each of the senses to its context. Context requires bringing into sight a larger quantity of text, that is words, to the right and left of the element under consideration. Lexicographers such as Johnson in the 18th century, had already isolated words within their context to extract and identify meanings. But if you have to do that manually with five million words, the work is by far too demanding for a human mind. Once again, computers and in particular linguistic computation have turned this impressive task into an extremely easy one. Concordancers might offer you thousands, hundred thousands and even millions of words with their own context in seconds: you can also make the context shorter or larger, at your own convenience; you can classify the occurrences alphabetically by the word occurring 1st, 2nd, 3rd to the right or to the left; you can search for words with specific frequencies or within a specific range of occurrence: you can pick up only a certain percentage of concordances out of the total amount of them, if this is too high and you are unable to look at every one of them...

Tribble and Jones (1990) offer some interesting data on the history of concordancing. They quote Hugo de San Charo, in the 13<sup>th</sup> century, as a pioneer in the compilation of the concordances of the Bible, with the help of 500 monks. Other instances might also be quoted in later centuries, but it always remains clear that concordances have been compiled for reasons not related to language teaching. Religious, lexicographic or various other linguistic purposes are the trigger for such burdensome work, while direct pedagogical and didactic goals are only to appear in the second half of the 20<sup>th</sup> century. Their use in language teaching and learning is parallel to the popularisation of computers and the increase in the computational power of PCs. Skeehan (1981) referred to concordancing in relation to ESP, in the context of mainframe or minicomputer applications; he considered its potential for ESP as "considerable". Higgins and Johns (1984), Ahmad et al. (1985:126) discuss the role of concordancing in Language Teaching or refer to it for exploiting texts in the classroom. Hardisty and Windeatt (1989) apply concordances to CALL. In the nineties, the amount of authors that refer to concordances as a useful tool for language teachers is continuously increasing (Flowerdew, 1993; Johns 1997; Kettemann, 1995; Barlow, 1995; Stevens, 1995; Ball, 1996; Jackson, 1997, among others).

In spite of this, the fact is that at present many teachers are unfamiliar with the subject, either because they have not even heard of concordances, or because they ignore what concordances really mean or what they are for. Among the group of teachers who know about them, some still consider that they are 'too distant' from the point of view of practicality; others simply reject them because they lack the technical requirements needed for their exploitation in the classroom. In any case, it is hard to believe that teachers not familiar with computers will accept working with concordances or take advantage of their possibilities.

Concordancing software packages share some basic features: they process texts in electronic format, select the key words the user has marked and finally offer a sorted display of those words together with the accompanying context ('X' words or characters before and after, or full sentences, -in that case identified by the preceding or final stop ['.']). To these basic features some other refinements might be added, depending on the software: more or less powerful searching facilities, context definition, collocational patterns and statistical information.

Concordances are potentially useful tools for language teaching and learning. Murison-Bowie (1993:39) writes that 'concordancing gives us access to the evidence of language in use, unmediated by the grammarian, lexicographer or teacher'. This kind of 'evidence' refers mainly to authenticity and its relevance for the meaning of lexical elements within the context they are used or they belong to. Learning of vocabulary (meaning) and the right and adequate placement of words within the neighbouring context (grammar) are clearly supported by concordances. Ball (1995) illustrates the issue by using concordance programs in the learning of Old English. Johns (1988) writes in favour of concordancing in language learning because

- a. it implies authenticity.
- b. learners assume control of the learning process.
- c. learning is the result of 'research' by students, as it is understood in the so called 'data-driven learning' (DDL) (learners' competence is built by offering them direct access to language or linguistic facts).

Points b. and c. go beyond the use of concordances and in fact they might affect many other techniques used in the classroom or depend on many other issues. 'Control' and 'research' can be favoured by reading texts, by listening to authentic dialogues, etc. But key words in context provide linguistic evidence on which to base questions and discover the required answers by students themselves. The formulation of grammatical rules alone, -a deductive method of learning-, is not the best way to internalise grammar. Self discovery is



-os -os

bajos fondos

Felipe es conocido en los bajos fondos con el mote de El Pingüino.

Bajos instintos

Lumet dirige un guión del mismo autor de Bajos instintos.

baños colectivos

Cruzarían el estadio para ducharse en los baños colectivos. volverían apresurados a ponerse el uniforme.

barcos sardineros

Alerta continua sobre las operaciones de los 30 barcos sardineros y de sus tripulaciones...

-e -e

corriente favorable

El Copei se encuentra dividido en la corriente favorable a Caldera y la del ex secretario general.

creciente porcentaje

Buscan mano de obra más joven y más maleable. con un creciente porcentaje de mujeres y miembros de minorías étnicas.

creyente practicante

-¿Cómo es su relación con Dios? -Soy un creyente practicante. de misa dominical y converso con Dios a diario.

An unlimited variety of searnces affecting morphology. set phrases. prepositional verbs. etc. are quickly accessed and illustrated in any language:

-ing -ed

Of course. much encouraged by a building being **listed**. something which. until recently. was unusual.

Our historic building system is being perverted for purely political reasons.

In addition to the terminal already being **planned** for Waterloo.

Materials. when exposed. were enriched by being included within the decorative tradition.

touses in market towns are increasingly being used as offices by small firms.

Proposals depend on planning permission being **granted**. and cars circling suburban Edinburgh ...

There was no sign of being unprepared in his typically confident singing. ...

And then comes the word that Confederacy is being developed by Ivan Reitman. coming up soon ...

You cannot find information about the composers being **played**. No Canadian jokes. please. but...

depend on

How often it takes place will also depend on other factors, such as the sites that are infected...

Mr Johnston said the results would also depend on the timing of the sale of property; developments...

... and the conduction of impulses by nerves, all depend on it.

Figures vary between different centres and depend on the group of women being examined.

... might live on any planet on which life-forms reproduce and depend on supplies that can run out.

Associability: suffered by the stimulus will depend on the extent to which the stimulus is active.

Singing of the species-specific song by a bird. depend on an interaction between innate structures...

Discovering the meaning of words is an excellent 'research' exercise for learning vocabulary. Concordances for items whose meaning students may or may not know provide linguistic evidence for self discovery (see 'amenable' below) or for the identification of various senses in polysemous words (see 'book' below):

amenable:

[1. *responsible or answerable. 2. able to be controlled or influenced; responsive; submissive. 3. that can be tested by (with to)*]

1. The work's ideas, however, are better focussed and more amenable to musical realisation than those of Knot Garden or Ice Break ...

2. ... were achieved through the monumental labours of R. F. Skinner and his colleagues using such amenable subjects as the white rat and the domesticated pigeon...

3. ... who should uniformise time into an abstract geometrical co-ordinate, a continuous dimension amenable to mathematical handling" ...

4. They were surprised by his zeal and suspected trouble, but Nails was amenable and even civil during the journey, mystifying them still further.

5. Planning, in turn, assumes the existence of a system amenable to planning; a system, that is, which is coherent, consistent, rational, predictable

7. ... have considerable personal and social difficulties in addition. In some cases these may be amenable to the problem-solving approach described in Chapter 5, once drying-out has been completed.

book:

1. We were able to book a passage home on Cuiard's Duchess of Atholl, an old lady ship of whom even the ghost ...

2. It is a difficult book, but the best study of the psychology of the artist that has ever been written.

3. ...white-liberal cultivation of the African experience, where some of the best comedy in the book is located.

4. "... on top of the RCA Building as Rainbow And Stars." said Don Smith, "I told them they shouldn't book big names, they should sell the experience.

5. For Larkins the game represents a chance to book his ticket for Sunday's Nehru Trophy opener against Sri Lanka.

6. The world's second oldest league is celebrated in a new book that is an essential purchase for anyone who loves the game's nuances.

Language use may vary in different areas, regions or countries. That applies to languages spoken in various countries, as it is the case of English and Spanish. Some words and expressions are typical of regional usage. Concordances are ideal tools for identifying, pointing out or discovering distinctive semantic features, as might be the case of the Spanish word 'lindo', widely used in South American Spanish:

'lindo' in American Spanish usage

1. Hace un efecto muy **LINDO**.

7. Se encuentra en lo que el ginecólogo búlgaro César Anipión denomina "Estado de recién nacido o **LINDO** bebé".

3. **Hasta** el lector menos conocedor del **alma** femenina y de las costumbres medievales habrá supuesto que, a costa del trovador feo, la condesa y sus damas se divertían de lo **LINDO**, inventando apodos que le cuadraran bien y resaltarán lo ridículo de su aspecto o imitando su modo de andar y aplastándose con un dedo la punta de la nariz para ponerla chata.

4. Tírale una pelota de goma y verás, qué **LINDO**, cómo corre y regatea y la coge y te la lleva y mueve el rabo.

5. La sala del aeropuerto se había convertido en una verdadera pachanga, a la que pronto se unieron los niariachis con guitarrones y trompetas, cantando aquello de "México **LINDO** y querido, si muero lejos de ti, que digan que estoy dormido y que me traigan aquí".

6. Era un **viejo LINDO**, envuelto en una alegre bata de franela a cuadros, con pantuflas a juego y un gorro de lana de dormir como los de los cuentos, rojo, largo y acabado en punta con un pompón.

7. El mulato **LINDO**, qué lástima de hombre.

It might also be necessary or advisable to contrast the use of words with meanings which can be easily confused, as is the case of *'exercise/execute'*:

execute:

1. Corrupted by his own 'incurable sophistication', he yearns to execute new possibilities.

2. There are dark castles with brooding lords and proud clan leaders who live only to execute vengeance; rustic Cinderellas who turn out to be noble ladies...

3. ... to an infant does not vest the legal estate in him, but operates only as an agreement to execute a settlement in his favour.

4. This means that the grantor must execute a principal vesting deed and trust instrument, holding the land meanwhile as trustee for the ...

5. ... No special form is necessary for the appointment of an agent, except that an agent who is to execute documents under seal in the name of his principal must be appointed by a 'power of attorney'....

6. Where the settlor is himself insolvent, the trust is invalid, and creditors can execute against all the settlor's property including that over which he had purported to set up a ...

7. ... continues to be one of deciding whether to attack or to stay back, and to discover how to execute a charge successfully.

exercise:

9. ...opening Allego's second subject into a miniature slow movement?), and the composer needed to exercise an iron control in order to draw all the intellectual and emotional threads together.

10. What more appropriate than that a performance should exercise similar technical control and let the emotional content, so to speak, take care of itself.

11. ...owing the now notorious faulty tanning-lotion incident. it is common knowledge that Goss must exercise some degree of caution about exposing his skin to loud noises.
12. ... the length of this month was not determined by any precise rule. the pontiffs were left to exercise their discretion, and they frequently abused this power for political ends.
13. Peretz gives only one small hint of the little man's power to exercise choice or change his fate.
14. The homicides and postmortems in the book permit the new religion of science to exercise its power: but they also occasion the necrophile broodings which esude from Dyer.
15. ...he or she will channel the need to exercise control onto some other aspect of behaviour such as food intake.

Sequences of related functional elements are readily at hand:

No... ni:

1. Los servidores del amo y no de quien los elige. [no tendremos aquí, ni] democracia. ni justicia. ni una economía...
2. No sólo no tenemos pensiones sino. que [no tendremos, ni] siquiera, sueldo.
3. Porque, si esto no hubiera funcionado. ni antena ni nada.
4. Los sitios de apariciones. que en aquel momento [no había videntes, ni] había, digamos. apariciones. sino que..
5. Sobre todo. cuando [no has sido ni] pastor. ni oveja.
6. Eso [no es, ni] como ha dicho el señor Sánchez Drago. ni como ha dicho la prensa...
7. ... y unos dolores espantosos que [no se pueden ni] contar. ¿no?
8. Yo llevaba plantillas. [no podía andar ni] un paso. incluso nie querían poner en un cachecito...

Either ... or:

1. ...Too many of the minor roles become cameos, and a cameo is [either an ornament or] a blot. but in either case tends to detract from the momentum of a film.
2. ... there is nothing radical or dynamic in [either the vocalisation or] the movement.
3. The most favoured option is to pick a well-known name and [either anthologise him or] produce what is now grandly referred to as a biodrama.
4. ... between those who were willing to prolong the applause indefinitely and those who [either disappeared discreetly halfway through or] scarpereed quickly at the end.
5. BY NOW, the name of Peter Greenaway functions [either as a reliable guarantee or] a reliable warning.
6. A nostalgia has become apparent, after an era of aggressive atheism, [either for the great organized beliefs or] for substitutes chiming better with the would-be freer spirit of the times.
7. ... so they must do away with the persons who [either made him or] could save him.
8. ... with this one. he [either becomes God by coinicidn acclaim, or] his faiths. feeling deceived. flood the second-hand racks within weeks.

Stevens (1990:5) is right in affirming that concordancing is "economical in terms of time to implement because it requires only a program plus a text base. where the text base could be the concatenated sum (subset or superset) of all the texts used for text reconstruction". It is economical because any teacher (or even students) will get dozens of examples of authentic linguistic production in seconds and because they will also have the possibility of applying the technique to a great variety of activities and exercises. I mentioned above a few models related to lexical learning. The amount of variety is almost open-ended.

up to the user creativity and eagerness in looking for new sequences. Any kind of mechanic exercise can be based on concordances, particularly substitution or fill-in types. With some more effort on the teacher's side, multiple options can be added as well. Cloze exercises on concordances are recommended, particularly in DDL, since they favour the use of context in order to find out the meaning of the missing items. Moreover, cloze sentences can be accompanied by a set of concordances through which students will find valuable help to fill in the gap.

Where to find concordancers is another issue, particularly if their use is limited to the classroom. In this case any of the commercially available ones (*Monoconc*, *Wordsmith*) is enough. 'Monoconc' for example is easy and friendly and texts to be processed need only be in ANSI format. It loads and runs quickly, the sorting of results is virtually instantaneous and the definition of search strings is very transparent. A contrastive review of commercially available concordancers is given in Higgins (1991a) and Tribble (1997), among others.

Regarding texts or corpora, some are also available ('Microconcord', by Oxford University Press, for example). Digitalized encyclopaedias (Encarta, British Encyclopaedia) are also in the market and Internet offers the teacher an infinite amount and variety of authentic texts for classroom use. Facing students with real language in the right way, at the right moment and in the right amount will help them internalise the true flavour of the language they are learning, far away from discouraging patterns, isolated words, or abstract grammatical rules.

## V. ASSOCIATION OF WORDS AND LEXICAL CONSTELLATIONS

Word lists elaborated in the first part of the XX century were already influenced by what McArthur (1998:57) calls 'the subjective and the objective view'. The objective method was well represented by Thorndike: Palmer and M. West favoured the subjective one. While the 'objective view' is limited to word frequency as it is given in a corpus, the 'subjective view' adds 'a subjective sense' or subjective and reasonable criteria to assist pure objective counting. Such a 'subjective sense' is based on the association of meanings, which Palmer called 'constellation'. "The lesson in which the word *eat* occurs for the first time contains also the word *drink*". Palmer says (McArthur 1998:60). *Eat* and *drink* are linked by 'thought association', i. e., one usually brings with the other one. Frequency of occurrence in those instances is the final product and should not be taken as the 'promoting factor' for including a word in a list or not. Palmer did not develop his idea of 'constellation'. His lists are groups of words occurring whenever a specific topic is in focus and words which apparently always

co-occur with other words. The co-occurrence of words has been perceived by Palmer and many other textbook authors along the history of language teaching and it must be acknowledged that such co-occurrence is sometimes taken care of in some textbooks and teaching materials. Dialogue books since the thirteenth century, natural methods or situational methods, for example, are centred around a communicative node which demands word association. This same linguistic evidence has found its scientific counterpart in what is nowadays called 'collocation'. In its broadest sense, 'collocation' is more or less equivalent to 'recurrent word combination'. In the Firthian tradition it is generally used in a stricter sense: two or more words which have a strong tendency to be used together. A classical example illustrates the issue: in English you say '*burning ambition*' instead of '*firing ambition*'. Sinclair (1991:170) defines collocation as 'the occurrence of two or more words within a short space of each other in a text'. There is no contradiction between both views, at least for the purpose of what we want to bring forward here. Learners of foreign languages are interested in learning not only words in isolation, but words in context, at least within the sentence range and adequately related and connected to the situation in which they are used. Such a claim reflects a basic linguistic fact and it should be a reminder of what goes on in the communicative process: words are never alone in real communication, that is, words always go in the company of other words. And the question immediately arises: which words accompany every other word? Is it relevant for learners to be aware of such a linguistic reality? Yes, indeed. The subjective perception about the association of words, very often present in teachers and manuals, has yielded relevant materials in this respect regarding the inclusion of lists of semantically related words, arranged by topics, situations or the like. We find such materials already in the 'Vocabularies' printed in the sixteenth century (Sánchez 1997:49 ff.), which offered teachers and students lists of the more usual words together with dialogues based on daily communicative situations. Subjective perception however does not solve the problem adequately and with accuracy. The words which accompany other words cannot be satisfactorily extracted by mere frequency counts based on huge quantities of textual samples, since 'company words' refer to other single words which 'need' or 'require' them, but the mutual attraction is not always direct. Very often words are attracted by 'intermediaries', that is, by words attracted which at the same time attract other words themselves. And the process might repeat several times within a sentence. Cantos and Sánchez (forthcoming) analyses the attraction of the Spanish word 'mano' (*hand*) in one of its senses (*a layer of paint*) and they discover that in the examples analysed the words attracted are '*pintura, aplique, tono, intensidad, deseada*'. Only '*pintura*' and '*aplique*' are directly attracted by '*mano*', while '*tono*' is attracted first by '*pintura*' and '*intensidad*' first by '*tono*'. All of them shape a kind of constellation governed by a hierarchical dependency

scheme. Mechanical counting of isolated words alone will be unable to cope with such a degree of semantic complexity.

Even if language teaching cannot yet benefit from these investigations, concordances are an important step forward for extracting useful and reliable information on the company that words keep, since the range of text covered by the search is limited to the sentences in which the key form occurs. Once concordances for an item are available, we only need to ask the computer for a frequency list of all the words in them, that is the words used around the key form searched. If we take the word *teléfono* in Spanish (ca. 2.700 occurrences in a twenty million word corpus, as *Cumbre*) we get the following first items in frequency:

*llamar, tener/recibir información, hablar, dar (algo), charlar, decir (algo), preguntar, pagar, recibir, escuchar, comunicar, conectar, contestar, ...*

All those words seem to match quite well the true function of a 'telephone': to call other people, to get/give information, to talk, to chat, to say (something), to ask, etc.

We might expect that different languages behave similarly regarding the company of words they keep. It is not exactly so. The same item, telephone, in English goes along with the following words (according to a 4 million corpus of English):

*service\*, have, call, car\*, time\*, say, open\*, use\*, allow\*, box\*, number\*, access, cellular\*, mobile\*, public\*, technology\*, etc.*

Words with an '\*' do not match the Spanish list. Both corpora are not equivalent and the results cannot be taken as fully reliable, but they illustrate a basic fact: association of words take place along the lines of what speakers think and do in specific situations of real life. Speakers of Spanish seem to use the telephone for 'talking, communicate with others and pay', while the English apparently refer to the telephone as a 'service', for 'calling, saying, using', relevant in terms of 'time', implying a 'number', giving 'access' to others, with a 'technological' change towards the 'cellular, mobile' phone...

The examples above are a necessary complement to vocabulary lists based on frequency alone. The word 'teléfono' is included among the first 1,000 most frequent words of Spanish. But some of the words that usually go with it are not. Still, they are more likely to appear whenever 'teléfono' is there because the communicative context will require their presence. Students will therefore communicate with more ease if they know not only the most frequent words of a language, but the words accompanying them as well. To take frequency counts as the unique criterion for fixing a basic list for language teaching or learning (the 'objective view') is not a utopia and should be refined and completed with the addition of their 'linguistic friends'. It helps no doubt in limiting and making a first selection

of the most relevant words used in communication. but frequency counts should not be dissociated or disconnected from the set or constellation of items that each 'frequent' word brings with it. If we want to have more accurate, reliable and useful results, raw frequency counts need some correcting factors and the constellation of words is one of them.

One more example will help illustrating the issue. The word *economía*, in the corpus *Cumbre* offers the following frequency count of the first 92 words:

*mundo, gobierno, país, mercado, política, ministro, española, año, determina, fuente, millones, nacional, crecimiento, económico, tiene, ministerio, años, desarrollo, empresas, crisis, sector, países, prensa, ser, banco, empleo, hacienda, inversión, trabajo, público, decir, mundial, internacional, comisión, situación, sistema, capital, producción, sociedad, precios, recuperación, proceso, comercio, inflación, europea, medidas, mercados, sectores, industria, dinero, acuerdo, estados, gasto, unión, plan, déficit, importante, programa, poder, tipos, aumento, nación, paro, pasado, período, recursos, actividad, real, consejo, financieros, pesetas, empresa, consumo, ministros, crear, diario, ahorro, coyuntura, cunihio, problema, respecto, bienes, ley, fin, fiscal, poco, domésticas, plazo.*

In this list, we find the following nouns:

*mundo, gobierno, país, mercado, determinar, ministro, millones, nacional, crecimiento, económico, ministerio, desarrollo, empresas, crisis, sector, prensa, banco, empleo, inversión, trabajo, mundial, internacional, comisión, sistema, capital, producción, sociedad, precios, recuperación, proceso, comercio, inflación, medidas, mercado, sectores, industria, dinero, acuerdo, estados, gasto, unión, plan, déficit, programa, poder, tipos, aumento, nación, paro, período, recursos, actividad, consejo, pesetas, empresa, consumo, ministros, ahorro, coyuntura, cambio, problema, bienes, ley, fin, plazo.*

Adjectives are significantly shorter in number:

*española, nacional, económica, económico, público, mundial, internacional, europea, importante, pasado, real, financieros, diario, políticas.*

And the amount of verbs is reduced to only eight:

*determinar, tener, ser, decir, poder, crear, pesar, reducir*

Intuitively, we can conclude that learners of Spanish wanting to learn the word 'economía' must also know that whenever this word occurs there are other words that systematically and persistently combine with it. If language users want to understand the meaning of this item and the sentence or context in which it occurs, the words around

'economía' are certainly relevant. And this includes the fact that nouns are significantly higher in number (92), contrasted against only eight verbs.

This is even more relevant if the association of words, as mentioned above, are not equivalent in different languages. A native speaker of English will have the tendency to associate 'economy' to the constellation of words of this item in his native language, which do not match with the constellation of Spanish 'economía'. The case would be similar in the opposite direction. Native speakers of English need to know the constellation of 'economía' in order to adjust to the Spanish linguistic system.

Compared to Spanish, the word *economy* takes with the following 'lexical company':

Nouns:

*market, sect, rates, recession, year, policy, chancellor, news, party, society, budget, figures, interest, people, state, way, cent, markets, page, programme, rest, century, deficit, demand, government, growth, part, rise, stock, surplus, time, value, work, world, argument, bank, base, Britain, change, class, concept, country's, dollar, economics, effect, employment, form, etc.*

Adjectives:

*British, political, economic, high, foreign, last, national, modern, real, created, financial, international, monetary, new, own, recent, single, strong, considerable, first, etc.*

Verbs:

*said, are, have, make, may, created, growing, keep, rise, based, declared, etc.*

If we compare the constellation of 'economía' and 'economy' in the range of the Grst most frequent words in each language, we have the following results (only words with an '\*' are common in both languages):

Spanish	English
<b>Nouns (arranged by frequency):</b> * <i>mundo, *gobierno, país, *mercado, *política, determinar, *ministro, millones, nacional, *crecimiento, económico, *ministerio, *desarrollo, empresas, crisis, *sector, *prensa, *banco, *empleo, inversión, *trabajo, mundial, internacional, comisión,</i>	<b>(arranged by frequency)</b> * <i>market, *sector, *rates, recession, year, *policy, *chancellor, *news, party, *society, budget, *figures, interest, people, *state, way, cent, *markets, page, programme, rest, century, deficit, demand, *government, *growth, part, rise, stock, surplus, time, value, *work,</i>

Spanish	English
<i>sistema. capital. producción. *sociedad. precios. recuperación. proceso. comercio. inflación. medidas. *sectores. industria. etc.</i>	<i>*world. argument. *bank. etc.</i>
<b>Adjectives:</b> <i>española. *nacional. *económico. público. mundial. *internacional. europea. importante. *pasado. *real. *financieros. diario. *políticas</i>	<b>Adjectives:</b> <b><i>British, *political, *economic, high, *foreign, last, *national, modern, *real, created, *financial, *international, monetary, new</i></b>
<b>Verbs:</b> <b><i>determinar, "tener, "ser, "decir, "poder, *crear, pesar, reducir</i></b>	<b>Verbs:</b> <b><i>*said, *are, *have, mnke, *may, *create, growing, keep</i></b>

Items found in both languages are

*mercado, política. ministro. crecimiento. ministerio. desarrollo. empresas. sector. prensa. banco. empleo. trabajo. comisión. sociedad*

*\*market. \*sector. \* rates. \*policy. \*chancellor. \*news. \*society. \*figures. \*state. \*markets, \*government. \*growth. \*work. \*world. \*bank.*

while these are language exclusive:

Spanish:

*determinar. millones. nacional. crisis. inversión. mundial. internacional. sistema. capital. inversión.*

English:

*recession. year. party. budget. interest. people. way. cent. page. programme. rest. century. deficit. demand. part. rise. stock. surplus. time. value. change. class. concept. dollar. economics.*

Shared words reflect common interests in both societies. while items exclusive of each one of the languages reveal differences in the organisation of the world around the topic and notional field of 'economy'. The English seem to emphasise:

*recession. budget. interest. programme. rest. deficit. demand. stock. surplus. value. change. dollar. etc.*

While the Spaniards consider that economy:

'determines, is important, reduces' (*determina, pesa, reduce*).

The English tend to think of it as something that

'makes, grows, keeps, rises, declares, bases'.

## VI. CONCLUDING REMARKS

Corpora, regarding the three issues dealt with here (vocabulary lists, concordances and association of words in communication), are a novel and valuable tool for linguistic research and offer a wide range of possibilities for applying the product of such a research to language learning and teaching:

- a) Authors and language teachers will find in them an inexhaustible source of materials for investigating and applying the products of this investigation to textbooks or in the classroom.
- b) Teachers may use corpora in order to gather relevant texts for reading, designing and elaborating complementary exercises and so enrich daily routines with variety and linguistic samples taken from real usage. Samples might refer to any linguistic topic: grammar, morphology, orthography, syntax, vocabulary, etc.
- c) Corpora offer the basis for establishing reliable conclusions on the adequacy of texts being used in the classroom. Frequency counts as they appear in corpora can be contrasted against the vocabulary offered by manuals. The frequency of the Spanish verb *tener* and its inflected forms has been investigated both in a corpus and a textbook for teaching Spanish as a foreign language. The results are worthwhile quoting here: the most frequent forms of *tener* account for a frequency of 0.96% in the corpus (Sánchez et al. 1995a). Within this global share, the inflected form *tiene* counts for 30% of all the occurrences, *tengo* for 8.5% and the forms for the indicative present of *tener* (*tengo, tienes, tiene, tenemos, tenéis, tienen*) reach up to 66%.

In *Cumbre I* (Sánchez et al. 1995b), a textbook for teaching Spanish as a foreign language, *tener* counts for 0.77 % of all the words used in the book. *Tiene* takes 30% of all the different inflected forms of this verb, *tengo* 14%: all the variants of *tener* for present indicative reach 66% (see below):

<b>Corpus <i>Cumbre</i>: <i>tener</i></b>	<b>Textbook for Teaching Spanish as a FL: <i>Cumbre-I</i>: <i>tener</i></b>
Total: 0.96%	Total: 0.77%

- <i>tiene</i> : 30%	- <i>tiene</i> : 30%
- <i>tengo</i> : 8.5%	- <i>tengo</i> : 14%
- <i>Present indicative</i> : 66%	- <i>Present indicative</i> : 66%

Figures do not match exactly in all instances, but they are reasonably close to each other. In that sense we can state that the textbook adjusts fairly well to real language use regarding the frequency of *tener*. The analysis could still be enlarged and include other aspects connected to frequency or association of words. Such results make a sound objective tool for taking important decisions affecting the selection of teaching materials.

A corpus is not a magic tool that might solve all the problems teachers have to face. But it helps in solving some of them and might bring into the classroom real language usage in connection to what modern technology can offer in assisting teaching.

RIBLIOGRAPHY

- Ahmad, Khurshid, Greville Corbett, Margaret Rogers, and Roland Sussex. (1985). *Computers, language learning, and language teaching*. Cambridge. London: Cambridge University Press.
- Aston G (1995). "Corpora in language pedagogy: matching theory and practice" in Cook G & B Seidlhofer (eds) *Principie and practice in applied linguistics: studies in honour of HG Widdowson*, Oxford. OUP
- Aston, Guy. (1996). "The British National Corpus as a Language Learner Resource" in Botley, Glass, McEnery & Wilson (eds.). ( 1996). pp. 178-191.
- Aston, Guy. (1997). "Ennching the leaming environment: Corpora in ELT" in Wichmann, Fligelstone, McEnery and Knowles (eds.). pp. 51-64.
- Ball, C. (1996). *Tutorial Notes: Concordances and Corpora*. Washinton DC. Georgetown University. Also available HTTP: <http://www.georgetown.e.../corpora/tutorial2.html> [1997. 03/12].
- Ball, Catherine N.. (1995). *Hwæt! A Corpus-based Hypermedia Resource for Old English Lexical Acquisition*. Washington DC. Georgetown University.
- Barlow, Michael. (1997). *MonoConc for Windows*, Athelstan. Houston TX
- Barlow, Michael. (1992). "Using Concordance Software in Language Teaching and Research". In W. Shinjo et al. *Proceedings of the Second International Conference on Foreign Language Education and Technology*. (LLAJ & IALL).
- Barlow, Michael. (1995). "Corpora for Theory and Practice". *International Journal of Corpus Linguistics*, 1(1).
- Barlow, Michael. (1996). "Parallel Texts in Language Teaching" in Botley, Glass, McEnery & Wilson (eds.) 1996. pp. 45-56.
- Bazerman C. (1994). *Constructing Experience*, Carbondale. Southern Illinois University Press
- Biber, D., S. Conrad & R. Reppen (1994). "Corpus-based approaches to issues in Applied Linguistics". *Applied Linguistics* 1512:169-188
- Bongers, H. (1947). *The history and principles of vocabulary control*. Holland. Woopi-Woerden.
- Botley, Simon, Julia Glass, Tony McEnery & Andrew Wilson. (1996). *Proceedings of Teaching und Language Corpora*, 1996 (UCREL Technical Papers Volume 9).
- Cuadernos de Filología Inglesa*, 9(1), 7000. pp. 5-37

- Braddock, R., Lloyd-Jones, R., and Schoer, L. (1963). *Research in Written Composition*. Urbana, IL: NCTE.
- Butler, Jonathan. (1991). "Cloze procedures and concordances: The advantages of discourse level authenticity in testing expectancy grammar". *System*, 19, 1/2:29-38.
- Capel, Annette. (1993). *Concordance Samplers 1: Prepositions*, London. Collins Cohuild.
- Carne, Chris. (1996). "Corpora. Genre Analysis and Dissertation Writing: an Evaluation of the Potential of Corpus-based Techniques in the Study of Academic Writing" in Botley, Glass, McEnery & Wilson (eds.). 1996. pp. 127-137.
- Carpenter, E.. 1993. *Confusable Words*, London. HarperCollins.
- Celce-Murcia, M.. (1990). "Data-based language analysis and TESL" in Alatis, J.E. (ed.) *Georgetown University Round Table on Language and Linguistics* .1990 . Washinton DC. Georgetown University Press.
- Chalker, S. (1994). "Pedagogical grammar: Principles and problems" in M. Bygate, A. Tonkyn, and E. Williams. (Eds.). *Grammar and the Language Teacher*. (Pp. 31-44). London: Prentice Hall.
- Cobb, Tom. (1997). "Is there any measurable learning from hands-on concordancing?". *System* Vol. 25 No. 3. pp. 301-315.
- Coniam, David. (1996). "Using Corpus Word Frequency Data in the Automatic Generation of English Language Cloze Tests" in Botley, Glass, McEnery & Wilson (eds.) 1996. pp. 29-44.
- Cook, G. (1997). "Key concepts in E.L.T: Schemata". in *ELT Journal*. vol. 51. no. 1. 86.
- Facchinetti, Roberta. (1996). "The Exploration of Diachronic English Software by Foreign Language Students" in Botley, Glass, McEnery & Wilson (eds.) 1996. pp. 150-159.
- Feldmann, Ute and Brigitte Stemmer. (1987). "Thin- aloud a retrospective da-in C-Te-taking: Diffè languages diff-learners-sa-approaches?". in Faerch, C. and G. Kasper (Eds.). (1987). *Introspection in second language research*. NY. Clevedon. pp. 351-367.
- Fernandez-Villanueva, Marta. (1996). "Research into the Functions of German Modal Particles in a Corpus" in Botley, Glass, McEnery & Wilson (eds.). pp. 83-93.
- Fligelstone, Steve. (1993). "Some reflections on teaching, from a corpus linguistics perspective". *ICAME Journal* 17 pp. 97-109.
- Flowerdew, John. (1993). "Concordancing as a tool in course design". *System* 21/3. pp. 331-243.

- Flowerdew. Lynne. (1996). "Contextualised business vocabulan" in D. Gardner & L. Miller (eds.) *Tasks for Independent Language Learning*. TESOL. pp. 127-129.
- Fox. Gwyneth. (1991). "Context dependency and transitivity in English". in Johns. Tim. and Philip King (Eds.). *Classroom concordancing*. ELR Journal 4. University of Birmingham. pp. 17-35.
- García Hoz. Victor. (1953). *Vocabulario usual, vocabulario común y vocabulario fundamental*. CSIC, Madrid.
- Garton. James. (1991). "Question and answer: Concordancing in analysis of classroom discourse". *ON-CALL* 5. 4:9-11.
- Gavioli. Laura. (1996). "Corpus di testi e concordanze: un nuovo strumento nella didattica delle lingue straniere". *Rassegna Italiana di Linguistica Applicata*. 3.
- Gavioli. Laura. (1997). "Exploring texts through the concordancer: Guiding the Learner" in Wichmann. Fligelstone. McEnery and Knowles (eds.) *Teaching and Language Corpora*, London. Longman. pp. 83-99.
- Gledhill. Chris. 1996. "Science as a Collocation. Phraseology in cancer research articles" in Botley. Glass. McEnery & Wilson (eds.). 1996. pp. 108-126.
- Goodale. M. (1995). *Concordance Samplers 2: Phrasal Verbs*, London. Collins.
- Goodale. M.. (1995). *Concordance Samplers 3: Tenses*. London. Collins.
- Hardisty. D. and S. Windeatt. (1989). *CALL*. Oxford. OUP.
- Higgins. John and Tim Johns. (1984). *Computers in Language Learning*. London. Collins.
- Higgins. John. (1991a). "Which concordancer? A comparative review of MS-DOS software". *System* 19. 1/2:91-117.
- Higgins. John. (1991b). "Looking for patterns". in Johns. Tim. and Philip King (Eds.). *Classroom concordancing*. ELR Journal 4. University of Birningham. pp. 63-70.
- Holliday. Lloyd. (1993). "Solutions to KWIC concordancing using word processors in CALL classes". *ON-CALL* 811. pp. 28-31.
- Jackson. Howard. (1997). "Corpus and Concordance: Finding out about Style" in Wichmann. Fligelstone. McEnery and Knowles (eds.) *Teaching and Language Corpora*. London. Longman. pp. 224-239.
- Jappy. Tony. (1995). "Concordancing and practical grammar" *The journal of TESOL France*. 2.2 pp. 193-209.

- Jennings, K. and Doyle, T. (1996). "Curriculum innovation, teamwork and the management of change". in J. Willis and D. Willis. (Eds.). *Challenge and Change in Language Teaching* (p 169-177). Oxford: Heinemann.
- Johns T.. (1986). "Micro-Concord, a Language Learner's Research Tool". *System*. 14(2), 151-162
- Johns. Tim. (1988). 'Whence and Whither Classroom Concordancing?' in Bongaerts. T., P. de Haan. S. Lobbe. and H. Wekker (eds.). *Computer Applications in Language Learning*, Foris. pp. 9-27.
- Johns T.. (1991). "From print out to handout: Grammar and Vocabulary teaching in the context of data-driven learning". in Johns T. and Ph. King. eds.. *Classroom Concordancing, ELR-Journal* 4, 27-45
- Johns. Tim. (1993). "Data-driven learning: an Update". *TELL & CALL* 1993/2, pp. 4-10
- Johns. Tim. (1997). "Contexts: the Background, Development and Trialling of a Concordance-based CALL Program". in Wichmann, Fligelstone, McEney and Knowles (eds.), pp. 100-115.
- Johns. Tim. (1997). "Kibbitzing One-to-Ones". (Web version of notes for presentation at BALEAP meeting on Academic Writing, University of Reading, 29th November 1997).
- Karpati. Ildiko. (1995). *Concordance in Language Learning and Teaching*. University of Pecs.
- Kerr, J.. (1992). "Using concordancing software with ESL learners". *ON-CALL* 613 pp. 2-7.
- Kettemann. B. (1995). "On the use of concordancing in ELT". in TELL&CALL no. 4, 4-15. Also Available HTTP: <http://gewi.kfunigraz.ac.at/~ketteman/conco.html> [1997 02125].
- Kettemann. Bernhard. (1995). "On the Use of Concordancing in ELT" *TELL&CALL* 199514, pp. 4-15.
- Ketteniann. Bernhard. (1996). "Concordancing in English Language Teaching" in Botley, Glass, McEney & Wilson (eds.) 1996, pp. 4-16.
- Kucera, H. & W. Nelson Francis. (1967). *Computational analysis of present day American English*. Providence Rhode Island. Brown University Press.
- Leech. Geoffrey. (1997) "Teaching and Language Corpora: a Convergence" in Wichmann, Fligelstone, McEney and Knowles (eds.), pp. 1-24.

- Legenhäusen, L. & D. Wolff, (1991). "Zur Arbeit mit Konkordanzen im Englischunterricht". *Der Fremdsprachliche Unterricht: Englisch* 25/4. pp. 24-39.
- Levy, Mike (1990). "Concordances and their integration into a word-processing environment for language learners". *System* 8/2. pp. 177-188.
- Lewis, M. (1994). *The Lexical Approach*. London: Langurige Teaching Publications.
- Long, M. (1988). "Instructed interlanguage development". in L. Beebe. (Ed.). *Issues in Second Language Acquisition: Multiple Perspectives*, Rowley, Mass.: Newbury House.
- Louw, Bill. (1991). *Classroom concordancing of delexical forms and the case for integrating language and literature*. In Johns, Tim. and Philip King (Eds.). *Classroom concordancing*, ELR Journal 4. University of Birmingham. pp. 151-178.
- Ma, Bruce C.K.. (1993). "Small Corpora Concordancing in ESL Teaching and Learning". *Hong Kong Papers in Linguistics and Language Teaching*, 16. pp. 11-30.
- McArthur, T. (1998). *Living words. Language, lexicography and the knowledge revolution*. Exter. University of Exeter Press.
- Meyer, Renee, Mary-Ellen Okurowski and Therese Hand. (1996). "Using Authentic Corpora and Language Tools for Adult-centred Learning" in Botley, Glass. McEnery & Wilson (eds.) 1996. pp. 171-177.
- Milton, John. (1996). "Exploiting L1 and L2 corpora for Computer. Assisted Language Learning Design: the role of an interactive hypertext grammar". in Botley, Glass. McEnery & Wilson (eds.) 1996. pp. 233-243.
- Mindt, Dieter. (1986). "Corpus, Grammar, and Teaching English as a Foreign Language". in Gerhard Leitner (ed.). *The English Reference Grammar Language and Linguistics*. Writers and Readers (Tübingen). pp. 125-139.
- Mindt, Dieter (1997). "Corpora and the Teaching of English in Germany". in Wichmann. Fligelstone. McEnery and Knowles (eds.). pp. 40-50.
- Mpartusa, Cynthia, Alison Love, and Andrew Morrison. (1991). "Bringing concord to the ESP classroom". In Johns, Tim. and Philip King (Eds.). *Classroom concordancing*, ELR Journal 4. University of Birmingham. pp. 115-134.
- Munson-Bowie, S.. 1993. *Micro-concord manual: an introduction to the practices and principles of concordancing in language teaching* (Oxford University Press: supplied with Microconcord program).
- Murison-Bowie, S.. (1996). "Linguistic corpora and language teaching" in *Annual Review of Applied Linguistics*, 16. pp. 182-199.

- Ogden. C.K. (1930). *The Basic Dictionary*. London. Psyche Miniatures.
- Ogden. C.K. (1933). *The Basic Words: A Detailed Account of their Uses*. London. Psyche Miniatures.
- Ogden. C.K. (1940). *The General Basic English Dictionary*. London. Evans.
- Owen. Charles. (1996). "Do concordances require to be consulted?". *ELT Journal*. 50/3. pp. 319-224.
- Palmer. H. M. West and L. Faucett. (1936). *Interim Report on Vocabulary Selection*. The Camegie Conference Report. London. King & Son.
- Quiao. H.L. and R. Sussex. (1996). "Using the Longman Miniconcordancer on tagged and parsed corpora. with special reference to their use as an aid to grammar learning". *System* 3411. pp. 41-61.
- Sampson. Geoffrey. (1993). "Analysed corpora of English: A consumer guide". in Pennington. Martha C. and Vance Stevens (Eds.). *Computers in Applied Linguistics*. Clevedon. Avon: Multilingual Matters. 181-200.
- Sánchez. A. R. Sarmiento. Pascual Cantos y d. Simón (1995a). *CUMBRE. Corpus lingüístico del español contemporáneo. Fundamentos, metodología y análisis*. Madrid. SGEL s.a.
- Sánchez. A. & Al. (1995b). *CUMBRE*. (método integral para la enseñanza del español como LE. (Niveles 11. 31 y 31. Libro del Alumno). Madrid. SGEI,.
- Sánchez. A. (1997). *Los métodos en la enseñanza de idiomas. Evolución histórica y análisis didáctico*. Madrid. SGEI, s.a.
- Sánchez. A. (1997). "Utilidad de los corpus lingüísticos en el aula de español para extranjeros". *CARABELA*. 42. 1997. pp. 39-59.
- Sánchez. A. and P. Cantos (forthcoming). "A computational Approach for the Construction of lexical Constellations.
- Scott. M. (1997). *WordSmith Tools*. Oxford. OUP.
- Sinclair. John. (1986). "Basic computer processing of long texts". In Leech. Geoffrey. and Christopher N. Candlin (Eds.). *Computers in English Language Teaching and Research*. Harlow. Essex: Longman. pp.185-203.
- Sinclair. J. McH. and Renouf. A.J. (1987). "A lexical syllabus for language learning". In McCarthy. M.J. and Carter. R.A. (Eds.). *Vocabulary in Language Teaching*. London: Longman.

- Sinclair. John. (1991). *Corpus, concordance and collocation*. Oxford. OUP.
- Skeehan. Peter. (1981). "ESP teachers, computers and research". in *ELT documents*, 112 - The ESP teacher: Role, development and prospects. London: The British Council.
- Stevens. Vance. (1990). "Text manipulation: What's wrong with it, anyway?". *CAELL Journal* 1. 2:5-8.
- Stevens. Vance. (1991a). "Classroom concordancing: vocabulary materials derived from relevant, authentic text". *English for Specific Purposes* 10. pp. 10-15.
- Stevens. Vance. (1991b). "Create your own concordancer with a one-line DOS command. MS-DOS". *Newsletter* 5. 1:2-3.
- Stevens. Vance. (1993). "Concordances as enhancements to language competence". *TESOL Matters* 3. 6:11.
- Stevens. Vance. (1995). "Concordancing with Language Learners: Why? When? What?". *CAELL Journal* 6:3. pp. 2-10.
- Thorndike. E. L.. (1921). *The Teacher's Word Book*. Columbia University. Teachers College.
- Thorndike. E. L.. (1932). *A Teacher's Word Book of the 20,000 Words Found Most Frequently and Widely in General Reading for Children and Young People*. Columbia University. Teachers College.
- Thorndike. E. L. and Irving Lorge. (1944). *The Teacher's Word Book of the 30,000 Words*. Columbia University. Teachers College.
- Tribble. C. & G. Jones (1990). *Concordances in the Classroom. A Resource Book for Teachers*, London: Longman.
- Tribble. C. (1996). "Corpora, concordances and ELT" in *IATEFL Newsletter* no. 130. 35-26.
- Tribble C. (1997). "Corpora, Concordances and ELT". In Roswood T. (ed). *New ways of using computers in language teaching*. TESOL. Alexandria VA
- Tribble. C & G Jones. (1997). *Concordances in the Classroom: using corpora in language education*. (new edition). Houston TX. Athelstan.
- Turnbull. Jill. & Jack Rurston. (1998). "Towards independent concordance work for students: Lessons from a case study". *ON-CALL*. 12:2. pp. 10-21.
- West. Michael. (1935). *New Method Dictionary*. London. Longman.
- West. Michael. (1953). *A General Service List of English Words*. London. Longman.
- Cuadernos de Filología Inglesa*. 9(1). 2000. pp. 5-37

- Wichmann, Anne, Steve Fligelstone, Tony McEnery & Gerry Knowles (eds.). (1997). *Teaching and Language Corpora*. London. Longman.
- Willis, D.. (1990). *The Lexical Syllabus*. London. Collins.
- Willis, D.. (1993). "Syllabus, Corpus and Data-driven Learning". *IATEFL Annual Conference Report Plenaries*. pp. 25-31.
- Wilson, Andrew, and Tony McEnery (eds.). (1994). *Corpora in Language Education and Research a selection of papers from Talc94*. (UCREL Technical Papers Volume 4. University of Lancaster).
- Witton, Nic. 1993. "Using the Mini-Concordancer with Languages other than English". *ON-CALI*, 7/2. pp. 19-30.
- Wu, M.H.. (1992). "Towards a contextual lexico-grammar: an application of concordance analysis in EST teaching". *RELC Journal* 23(2). pp. 18-34.