

Corpus linguistics: resources and activities for EFL

Antonio Pinna

This paper has a practical aim in that it intends to show how corpus linguistics can contribute to develop advanced language learners' communicative competence through some corpus-based search activities and subsequent analysis of the retrieved material. In particular, it illustrates how corpora can be exploited to study given language phenomena in order to show the connection between text and context in both cultural and functional-rhetorical dimensions.

The approach I have followed takes as its starting point the investigation of language events produced in natural, authentic situations for specific communicative purposes. It highlights how the analysis of instances of *parole* can enhance learners' awareness of the complexities inherent in acquiring communicative competence in English by revealing much about the cultural connotations or rhetorical functions associated with language choices both in general language use and in specialized contexts. This empirical method is guided by a perspective on language which considers it to be indivisible from communication and as having a dialectic relationship with the culture of the community which speaks it. Indeed, by placing tokens of language in their wider linguistic, rhetorical and cultural contexts, their rich communicative load can be brought to the fore. This is in line with Freddi's (1994: 29-30) conception of language learning and teaching objectives as being centred around four key functions of language: the communicative one, according to which language is fundamental in building up and maintaining interpersonal relationships; the pragmatic function, which sees language as an instrument of action and social regulation; the expressive function, which enables us to articulate our inner mental states; the acculturative function, by means of which language learning implies acquiring the culture of a particular community of speakers.

A focus on the social and functional dimensions of language within the context of foreign language learning and teaching is paramount. Following Hymes (1971) and his concept of communicative competence, it is plausible to hypothesize that an individual's knowledge of the language covers not only the rules of morphosyntax, but also a wider range of loosely structured, mostly unconscious linguistic and communicative information which a

speaker can make use of to produce text which is pragmatically effective and appropriate to the situational context. The study of a corpus has precise limits in that it does not guarantee that retrieved items are grammatically accurate. However, corpus studies are not concerned with what is possible in a language, but rather with what is typical in actual communication. A corpus can be used to extract information on how language items are effectively used by competent speakers and may thus contribute to enhance a learner's implicit and explicit knowledge.

The paper is divided into three sections. Section 1 is tripartite as it gives a definition of 'corpus', complementing it with an introduction to the two types of corpora which have been employed in the present study, it briefly outlines the corpus-driven approach which is illustrated in the following section and it discusses some recent models of linguistic performance. Section 2 presents two types of analysis which are based on Sinclair's (e.g. 1994: 22-25) principle of co-selection; in particular, the recurrent lexical associations of the common word forms *woman* and *man* reveal some of the cultural stereotypes at work in the English language, while the typical pragmatic associations of the adverb *awfully* in G. W. Bush's Presidential speeches show how it is preferably used in larger semi-fixed structures with precise discourse functions linked to the communicative purpose of that specific text type. Finally, Section 3 concludes by discussing some of the implications of what has been illustrated.

1. *Resources for EFL*

In this section a definition of corpus is provided and two types of corpora used in the analyses illustrated in Section 2 are described. Secondly, an inductive approach to corpus studies is rapidly sketched. Finally, the innovative role of this approach in relation to language description and its psycholinguistic dimension is put forward.

1.1 *What is a corpus?*

As a first approximation, the definition of a corpus is that of a collection of texts or text samples in an electronic format. However, the main difference between a corpus and a text archive is that the former is compiled with the objective of representing a language or a variety of a language and is mostly used in linguistic research. Since a corpus is a

primary source of language data which must be interpreted by an analyst, its content must be designed to actually mirror the variety of a language it aims to represent. A corpus created to investigate the linguistic features of the academic research article in English, for example, will include a large, balanced selection of samples of research papers published by the most important academic journals in the various disciplines so that the results which the analyst is able to uncover from the corpus can be generalized to the whole genre. The two main criteria for describing a corpus, i.e. function and content (Hunston 2002: 2), are therefore strictly interdependent: the organization of the content according to explicit rules must descend from the function for which the corpus is created.

This paper makes reference to the general and the specialized types of corpora. Among those of the first type, the Brown Corpus, compiled in 1964, contains about one million words and represents the variety of written American English produced in 1961 by means of a careful selection of text samples belonging to various registers and genres. Its British English counterpart is the Lancaster-Oslo/Bergen (LOB) Corpus. The dimensions of these first generation corpora have been surpassed by the British National Corpus (BNC), which includes over 100 million words organized in a balanced collection of written and oral British English text mainly produced in the first half of the 1990s, and by the Bank of English, a larger and continuously growing corpus which comprises British, American and Australian English texts from a variety of written and oral registers. In Section 2.1 I will make use of a collection of keyword collocates extracted from the Bank of English in 1995 to show how a general corpus can be used to explore the relationship between language and culture.

Specialized corpora are created for ESP (English for Special Purposes) research. Among these, the Lancaster/IBM Spoken English Corpus (SEC), containing about 50,000 words, was compiled for the detailed analysis of the prosodic aspects of oral English; the Helsinki Corpus is a collection of written texts produced in the 750-1700 period designed for the diachronic investigation of British English; the Child Language Data Exchange System (CHILDES), which contains approximately 20 million words, was originally developed for the study of those factors which influence children's language acquisition. In Section 2.2 I will illustrate some pragmatic associations retrieved from the specialized corpus of G. W. Bush's Presidential speeches (2001-03) and compare them with those emerging from some general corpora. This will enable some considerations concerning Bush's rhetorical style to be made.

1.2 *What is corpus linguistics?*

Corpus linguistics can be used to study different language phenomena such as lexis, syntax, semantics and pragmatics. When the analysis is based on pre-existing theories of language, it essentially works as a methodology which is particularly effective in quantitative investigations aiming to show the proportions with which certain predefined categories are present in a given class of texts. Biber (1988), for example, demonstrates that differences between written and oral English vary in relation to the registers he examined, are multi-dimensional, i.e. are not based on such simple dichotomies as formal-informal, and may be described in terms of quantitative combinations of various language exponents which, via a complex statistical procedure, he correlates to a set of basic dimensions of variation. This corpus-based, quantitative approach is also adopted by Morley (2004a, 2004b) in his analysis of persuasion in English journalism. In both papers he defines the categories of phenomena which are assumed to be at work in editorials vis-à-vis the less persuasion-oriented genre of news reports and then carries out a quantitative corpus study of the two genres to demonstrate that a distinction between them can be diagnosed by statistical means.

It is also possible for the analyst not to adopt any specific pre-existing theory of language and rather choose an inductive procedure on the basis of which the formulation of theoretical statements is the result of repeated cycles of data observations, specification of hypotheses and their verification, expression of generalisations which are empirically testable (Tognini-Bonelli 2001: 2). This corpus-driven approach is suitable for qualitative analyses which aim to identify and describe language phenomena on the basis of their recurrent associations with their contexts. The importance of this approach for emerging descriptions of language is dealt with in Section 1.3.

Exploring language via corpora is particularly important for functional analyses of language, since it enables the analyst to focus on the communicative dimension. Learners can benefit from this approach which leads them to investigate occurrences of language in real contexts. The study of the behaviour of individual words or sequences of words in their typical co-texts can show associative preferences which mark a competent and efficient use of the language for genuine communicative purposes. The principle underlying the associative behaviour of words and a theoretical model of word-context association is expounded in the following section.

1.3 *The principle of co-selection and the central role of lexis in communicative competence*

The qualitative approach in corpus studies has produced new paradigms in our understanding of language. Their theoretical bases, however, are rooted in conceptions formulated in the past. Tognini-Bonelli (2001: 157-164), for example, explains how an inductive, corpus-driven procedure is consistent with a contextual theory of meaning as elaborated by Firth. In this theory (Firth 1957: 193-196), a contextual approach to the study of word meaning must take into account a word's typical associative relationships with other words, i.e. its preferred company. Firth's concept of collocation is based on the idea that the meaning of a lexical item is determined by its function in context and is conventionalized by its repeated use to achieve certain pragmatic purposes in routine contexts. While Saussure's approach to meaning is founded on abstract paradigmatic relationships within the language system, Firth laid the ground for studying meaning as a result of recurrent syntagmatic relationships. Sinclair (1991, 1996, 1998) has demonstrated the widespread tendency of words to co-occur within limited co-texts. While his definition of collocation is restricted to lexical co-occurrences, Sinclair (1998: 14-16) places the principle of co-selection at the centre of his theory of language and uses it to delineate a model of meaningful syntagmatic relationships between words based on increasing levels of abstractions. His extended unit of meaning (Sinclair 1996, 1998; Tognini-Bonelli 2001: 101-110; Stubbs 2001), which is illustrated in Section 2.2, is best defined as a recurring sequence of co-occurring language items centred on an invariable core word and characterized by an identifiable pragmatic function. This functionally complete unit is an important mechanism in the creation of textual meaning. According to Sinclair (1998: 14-16) co-selection operates at the lexical level, in cases where individual words co-occur in a limited co-text with a higher frequency than what is expected on the basis of their random distribution in the corpus; it also operates at the grammatical level, through the repeated co-occurrence of words belonging to specific syntactic classes, at the semantic level, when a given word is shown to co-occur with words belonging to specific semantic areas, and at the pragmatic level, if the analyst can identify a specific discourse or interpersonal function which is repeatedly carried out by the co-selected words.

Sinclair's definition of co-selection is textual, since it operates, together with morphosyntax, as a key mechanism in the organization of discourse. Partington (1998: 18) and Hoey (2004, 2005) hypothesize a psycholinguistic

dimension of the phenomenon. Partington (1998: 18) suggests that the principle may be at work in text production and interpretation, so that it must be integrated in the system of communicative competence, specifically in relation to performance and context-related tasks. Hoey (2004, 2005) proposes the theory of *lexical priming* to explain the psycholinguistic mechanism responsible for text cohesion. He maintains that competence is the result of a speaker's knowledge of the typical behaviour of lexical items in their co-texts of use. Put differently, an essential part of what we know about a given word, including its meaning, is based on its typical associations.

In conclusion, the theoretical frameworks examined above place lexis at the centre of linguistic and communicative competence. They also point to the importance for language learners of becoming acquainted with the most typical associations between words and their co-texts in order to boost both their knowledge of English and their performance.

2. *Activities for EFL*

In this section two activities based on lexical and pragmatic associations between words are illustrated. First of all, the recurrent lexical associations of *man* and *woman* emerging from a collection of keyword collocations drawn from the Bank of English in 1995 are examined to highlight some of the cultural stereotypes linked with these words. Secondly, the pragmatic associations of the adverb *awfully* in G. W. Bush's Presidential speeches are analyzed to show what they reveal about their recurrent discourse functions and the President's oratorical style.

2.1 *Lexical associations of the words man and woman*

A representative corpus of English can yield reliable information about the collocates of a given word, that is those words which co-occur with statistical significance (Barnbrook 1996: 94-100). Recurrent lexical associations of cultural keywords often encode stereotypes which are shared or simply known by members of a language community (e.g. Stubbs 1996: 157-195). This is the case of words denoting social groups in terms of ethnicity, age or gender. Krishnamurthy (1996), for example, considers the words which are most frequently associated with the adjectives *ethnic*, *racial* and *tribal* in the Bank of English and notes that these near-synonyms are

actually used to denote different types of human groups, often following a stereotype of Western superiority. His analyses show that *ethnic* denotes human groups in Europe and is also employed as a neutral term in academic discourse; *racial* carries connotations of discrimination and is mostly used in texts produced in or concerning South Africa; finally, *tribal* is used to denote African groups and carries with it connotations of primitivism and exoticism. Similarly, Stubbs (2001: 453-454) examines the lexical, semantic and pragmatic associations of words used to denote groups of people on the basis of their age and concludes that *underage*, *youth* and *teen* are often associated with words indicating illegal activities, violent practices or low quality objects. Finally, Tognini-Bonelli (2001: 125-128) studies the collocation profile of the Italian word *donna* in a corpus of Italian newspapers and demonstrates how the word is often associated with the semantic areas of physical appearance, sexual attraction, reproduction and marriage.

Gender is certainly a sensitive issue in contemporary Western societies and the words denoting human males and females are loaded with connotations that bear witness to the cultural stratification of the concepts of masculinity and femininity. Sigley and Holmes (2002), for example, demonstrate how the use of the word *girls* to indicate adult women (a practice which is shown by approximately 50% of all the instances of this word in the LOB corpus) has an ideological implication as it construes women as vulnerable, subordinate to men and exposed to be the object of male sexual desire. From a cognitive perspective, Koller (2004) demonstrates how adult female referents may be construed as sharing typically masculine connotations in the specific discourse of business magazine features when taking top managerial positions in a field traditionally dominated by men. Since the conceptual metaphor BUSINESS IS WAR is shown to be central in describing corporate relationships, Koller's data prove that the most frequent lexical associations employed to construct the identity of women managers are dependent on the construct BUSINESSWOMEN ARE WARRIORS.

An interesting activity which could be carried out with EFL learners could be based on searching for and retrieving information on the words *man* and *woman* as a preliminary investigation into the connotations associated with these concepts in the English-speaking world. The following is a possible analytic procedure. I have included in Tab. 20 the most important collocates of *man* and *woman* taken from Cobuild English Collocations 1.1, a collection of collocates of the 10,000 most frequent

words in the Bank of English in 1995. For reasons of space I will concentrate on the associations between the two most frequent evaluative adjectives and the two keywords, i.e. *good man* and *beautiful woman*.

Table 1. The 20 most important collocates of *man* and *woman*

Man	Woman
Young	Man
Old	Young
Said	Old
Like	Said
No	First
Woman	Another
Another	Like
Very	Year
Made	No
Only	Black
Black	Any
Himself	Everyday
White	Only
Good	Very
Any	Life
Whose	Right
Called	Beautiful
Right	Married
Great	Says
Every	Whose

An analysis of the contexts in which the associations *beautiful woman* and *good man* are used (Tables 2 and 3) reveal a view of women as objects of male desire, since they are mostly defined on the basis of evaluations concerning their physical appearance; men, on the contrary, are evaluated according to their social roles as productive and efficient members of the community.

Table 2. Twenty concordance lines for the association *beautiful* + *woman* in Cobuild English Collocations.

1	leggy, curvaceous, dark and beautiful	woman,	29. Call me now on 0839 102216.
2	a still handsome man and a beautiful	woman	- Lucky that Henry didn't want me!
3	The stream took on the form of a	woman	-a beautiful virgin innocently tempting
4	Then you see her, a beautiful	woman,	about 30, sitting on a comfortable
5	had fallen in love with a beautiful	woman	and, after transferring most of the
6	our wife, I know: she is a beautiful	woman	but she's hard and cold like the cou
7	ESS Antonia de Sancha is a beautiful	woman.	But her affair with Heritage Secret
8	--that this was a beautiful	woman	but it also had to have a feeling of
9	lack guy seeks attractive, beautiful	woman	for lasting relationship. Age immate
10	With the portrait of a beautiful	woman	found five years ago, dubbed the Mon
11	you find a picture of a beautiful	woman	having a wash in Hitler's bath! I
12	`A very beautiful, very terrible	woman.	I think he might have killed her if
13	. `She's the most beautiful	woman	in the world - More beautiful than A
14	perience. The result is a beautiful	woman	of substance," says society photogr
15	an's friend has just won a beautiful	woman,	one he would have liked for himself
16	in contrast, was a thin, beautiful	woman	past middle age who ruled her large
17	eting an exquisitely beautiful young	woman	says how pleased she is that the tow
18	in a long white dress, a beautiful	woman;	the other of a two-year-old girl wi
19	is neighbours, and she's a beautiful	woman,	true, but she talks like a street k
20	Sometimes when you meet a beautiful	woman,	you know their beauty is more than

Analyzing the lines in Tab. 2 it may be noticed that the combination *beautiful woman* is frequently associated with a description of the woman's physical appearance, with superlatives (ex. 13), intensifiers (ex. 12 and 17), and other evaluative adjectives (ex. 1, 2, 3, 6, 9, 12 and 16), some of which denote sensuality (ex. 1 and 3). There is also a tendency to indicate the woman's age, either in a precise or approximate way (ex. 1, 4, 16, and 17), as if the two most important features in referring to a woman were her age and physical appearance. Both parameters co-operate in construing a social identity of women as objects of male desire; therefore, age is also essential in determining their potential appeal. A third aspect is shown by the presence of the contrastive conjunction *but* (ex. 6, 7, 8, 19). This is preceded

by a concession expressing a positive evaluation of a woman's physical appearance which is immediately contrasted with a negative evaluation concerning her social dimension, i.e. her incapacity to love (ex. 6, 8, 12) or to behave according to certain social rules (ex. 7, 19). Therefore, the stereotypical view of women as objects of male desire is accompanied by the conventional idea of insensitive, or even dangerous, beauty, as ex. 11 seems to indicate by making the connection between *eros* and *thanatos*.

Table 3. Twenty concordance lines for the association *good + man* in Cobuild English Collocations.

1	milar, straight-forward, good-looking	man,	21-35. Pot (returned). Box 27185 YO
2	to deport him. `He's a lovely gentle	man	and a good Christian who wants to save
3	ncies. One example concerns that	man,	and great editor, John W. Campbell J
4	good president of our division # and a	man	at that. When it came my turn, I began
5	good at odds. Ekman appeared a good	man,	but he was homosexual and a drug
6	family me to the outer door. He was a good	man,	courteous, compassionate. I hoped he
7	the saying goes you can't keep a	man	down, or out of the game even, and
8	good being, as he puts it, `a very gentle	man	/ Even-tempered and good natured",
9	his name ooh he's a good actor this	man.	`Jus He were was a good actor. Well they
10	was very kind. I thought what a	man	he was and remembered that, in the be
11	good at a nice guy. Hansen: Good man,	man	- If you're going to use a mobile pho
12	good here's a \$5,000 reward, 'cause a good	man	is hard to find when he -- or she --
13	By now, shooting is too good for the	man,	Mayle's fellow - villagers now want
14	him anyhow - But he a good	man	nuh - No trouble yourself no more 'b
15	Christian Mrs O'Rourke. Well, dear, your	man's	away on a business trip, is he?" sa
16	good Amounts to the same thing, right?	Man's	no fuckin' good to us, either way -
17	woman there took a good look at	man's	face. She could identify him, which
18	the work for the common good of	man.	That observer might also believe that
19	everyday surprisingly good job playing straight	man	to the comic antics of the four leads
20	I remember him as a tall, virile	man	with rugged good looks who might
			have

Tab. 3 provides a series of concordance lines centred on the combination *good + man*. The adjective is rarely used to evaluate a man on the basis of his physical appearance (ex. 1 and 20). On the contrary, it mostly expresses opinions concerning his social dimension, mostly in terms of professional capacities (ex. 3, 4, and 9), of conventional relationships with other people (ex. 2, 5, 6, 8, 10 and 14) and usefulness with respect to the objectives of a group (ex. 12 and 16). The social identity of men, as it is encoded in the language, is thus preferably defined in terms of institutional relationships with other members of the community, with a special emphasis on their membership in certain groups (ex. 2 and 14) or the assumption of roles that are positively evaluated (ex. 3, 4, 5, 6, and 9).

In conclusion, the analyses of the concordance lines presented in Tables 2 and 3 have shown a procedure which may be followed by EFL learners in the realization of a project concerning the cultural stereotypes at work in the definition of gender in English and how recurrent lexical associations in the language can contribute to construe the social identity of men and women.

2.2 Pragmatic associations: the role of *awfully* in G. W. Bush's Presidential speeches

The topic of this section follows results in recent research studies on the pragmatic functions of extended units of meaning centred around intensifiers in the context of G. W. Bush's Presidential speeches (Pinna 2004, 2006). The studies have demonstrated how they can be described in a systematic way and attributed a discursive function. In a corpus of G. W. Bush's Presidential speeches (2001-03) containing about 1.9 million words, *awfully* is one of the most frequent intensifiers with 30 instances of use. Tab. 4 compares the frequencies of this adverb per million words between Bush's corpus and three general corpora: the BNC (British English), the Brown and Frown corpora (American English).

Tab. 4 Frequencies of *awfully* per million words in some corpora of English

Corpora	Frequencies of <i>awfully</i> per million words
G. W. Bush's Presidential speeches (2001-03)	16.66
British National Corpus	3.78
Brown Corpus + Frown Corpus	4.00

The data in Tab. 4 show a clear preference for the use of *awfully* in Bush's speeches with respect to its general use in the language. What follows is an analysis of its use in the U.S. President's discourse to uncover some details of his rhetorical style.

2.2.1 The pragmatic association centred on 'MAKE *it awfully* + adjective'

Table 5 presents a selection of concordance lines for *awfully* taken from Bush's corpus.

Table 5. The combination 'MAKE *it awfully* + adjective'

1	fight off the junk lawsuits that are making it	awfully	difficult for people to expand their businesses
2	on. (Applause.) And the death tax made it	awfully	difficult for you to make that decision.
3	They're scattered everywhere. It makes it	awfully	difficult to hold anybody accountable.
4	scattered all over Washington. It makes it	awfully	hard to set priority and to hold people accounta
5	lot of hurt, death, displacement. It makes it	awfully	hard for a society to function that is at war with
6	everywhere in Washington, which makes it	awfully	hard to hold anybody to account.
7	lawsuits cause your bill to go up, it makes it	awfully	hard for people in some parts of your state to
8	they're scattered everywhere, which makes it	awfully	hard to align authority and responsibility.
9	you've got so many agencies, it makes it	awfully	hard to hold people accountable. I believe in
10	And, yet, our tax code makes it	awfully	hard for people to realize that dream.

The lines in Tab. 5 yield a general schema which can be summarized as follows:

1. there is a recurrent lexical association between the core word *awfully*, the lemma MAKE and the pronoun *it* in the co-text immediately to the left of the core word;
2. there is a recurrent grammatical association between the combination highlighted above and an adjective in the co-text immediately to the right of the core word;
3. there is a semantic preference for the type of adjective which is recurrently selected in the co-text immediately to the right of the core word. It is, indeed, an adjective which expresses negative evaluation along the easy-difficult parameter as far as reaching an objective is concerned.
4. there is a recurrent pragmatic association between the semantic schema shown above and two abstract entities which occur in its immediate left and right co-texts: the objective which is difficult to obtain is presented in the *to*-infinitive clause following the evaluative adjective, while the entity that hinders its achievement precedes the lemma MAKE.

A schematic representation of this pragmatic association is shown in Tab. 6.

Table 6. The pragmatic schema centred on the combination 'MAKE *it* awfully + adjective'

Evaluated entity		Evaluative category	(Objective)
Noun group		<i>awfully</i> + adjective	' <i>to</i> -infinitive' clause
lawsuits			to get medical care
scattered (governmental offices)	MAKE <i>it</i>	<i>awfully hard/difficult</i>	to hold anybody to account
tax code			to realize that dream
a baby out of wedlock			

The discourse function of this extended unit of meaning is to propose the negative evaluation of an entity or situation that hinders the achievement of an objective (which is thus implicitly evaluated in a positive way). The relationship between the evaluated entity and the evaluative category is realized by the combination MAKE *it* which establishes a causal connection between the hindering situation and the objective that the President deems desirable. It must also be noticed that the pronoun *it* in the combination MAKE *it* does not have a semantic role but a grammatical one as it anticipates what is later presented in the *to*-infinitive clause.

In conclusion, the schema outlined in Tab. 6 is a typical rhetorical device used by Bush to criticize a situation which needs to be corrected by means of new governmental initiatives. This is part of a strategy which organizes those parts of Presidential speeches which introduce legislative proposals according to a problem-solution structure whereby the problem part may be expressed using the pragmatic association illustrated above. The relationship between the extended unit of meaning and its context is a functional one and, as it participates in the orator's attempt to win people's consensus, it is definitely in line with the general communicative purpose of political speeches.

2.2.2 The pragmatic association centred on 'awfully kind'

A second type of pragmatic association based on *awfully* is shown in Tab. 7.

Table 7. The combination '*awfully kind* + *to*-infinitive'

1	orado and Tennessee are here. You all are	awfully	kind to come. Thank you for being here, as we
2	nor being here today, Governor Wise. It's	awfully	kind of him to come and say hello to the Presi
3	us put this event on every year. And it's	awfully	kind of you to do this. It's good to see you.
4	ney General from the state of Nevada. It's	awfully	kind of him to come all the way over here.
5	ighborhood Center for hosting us. You're	awfully	kind to have us. Behave yourself. (Laughter.) I
6	Thank you for opening up your hall. It's	awfully	thoughtful of you to let us come in here.

Tab. 7 shows that *awfully* is typically associated with an adjective which positively evaluates someone's behaviour in a structure which is used to express gratitude. The person whose behaviour is assessed is either expressed in the prepositional phrase *of* + personal pronoun or in the noun group operating as the subject of the clause. The praised behaviour is presented in the *to*-infinitive clause to the right of the core word. The intensifier *awfully* is thus part of a pragmatic unit which can be realized in the following two variants:

Table 8a. The pragmatic association '*awfully kind* + *to*-infinitive'

Evaluation carrier		Evaluative category	Evaluated entity
Nominal group		<i>awfully</i> + adjective	' <i>to</i> -infinitive' clause
You	are	<i>awfully kind</i>	to come to have us

Table 8b. The pragmatic association '*awfully kind* + *to*-infinitive'

	Evaluative category	Evaluation carrier	Evaluated entity
It is	<i>awfully</i> + adjective	Prepositional phrase	' <i>to</i> -infinitive' clause
<i>It's</i>	<i>awfully kind/thoughtful</i>	<i>of</i> + pronoun	to do this to let us come

The pragmatic function of the unit is to express gratitude on the part of the President. It is mostly used on those occasions where Bush visits parts of the country. The unit of meaning is used to thank specific people in the audience for being present at the meeting or inviting the President. It is typically found close to the beginning of a speech. Thanking formulae are a direct way of establishing a friendly social atmosphere and are thus an important device in winning people's good-will right at the beginning of a speech (Pinna forthcoming). The function of the units exemplified in Tab. 8a and 8b are thus clearly supportive of the persuasive purpose of Presidential speeches.

2.2.3 How is *awfully* used in general corpora of English?

So far I have shown how *awfully* is functionally linked to its immediate co-text so as to be an integral part of extended units of meaning which are strategically deployed to achieve certain discursive or interactional objectives within the general communicative purpose of Bush's speeches. It is however fair to ask whether these uses of the adverb are peculiar to the

U.S. President's rhetorical style. A hint at a possible answer is already given in Tab. 4 which attests considerable use of the adverb in Bush's corpus, a clear signal of the importance of its functional roles in his speeches. To get a fuller answer, we can start by observing the behaviour of *awfully* in the Brown and Frown corpora. The eight instances of *awfully* retrieved from these American English collections of texts are shown in Tab. 9.

Table 9. Concordance lines for *awfully* in the Brown and Frown corpora

1	ou'll find plenty of blankets there". "You're	awfully	kind", the girl said. "We'll pay you back if
2	hadn't gone out to fetch her. "That was an	awfully	two weeks".
3	with- well, you don't know him but he's	awfully	nice- and we went to a couple of places- I don't
4	She smiled at me, but it was an	awfully	sad smile. She was even more miserable than me.
5	a lesser work might be crippling. It is	awfully	talky, for instance, and not all of the talk is
6	artley's work as "old world, old souled, and	awfully	fatigued," McBride sounded a criticism that he
7	I'm usually a very strong woman, but I'm	awfully	tired". "And hungry", he said. "Start in. It's not
8	she wants you to know that she's	awfully	grateful". Lucy did not believe him; Myra

A quick glance is sufficient to notice that there are no instances of the *awfully hard/difficult* combination and only one example of the *awfully kind* combination. A further investigation in the BNC yields 378 instances of *awfully*. The 13 *awfully hard/difficult* combinations are shown in Tab. 10.

Table 10. Concordance lines for *awfully difficult/hard* in the BNC corpus

1		Awfully	difficult to get it in the car wasn't it?
2	She said they had an	awfully	difficult job and we ought to help them as
3	That would be	awfully	difficult to do," she said.
4	ort of service he's going to make life	awfully	difficult for the Forest de central defence
5	"It's so	awfully	difficult". I'd be bound to get myself in a
6	s of course special clothing and then	awfully	difficult and making it people's education
7	Erm they're awfully	awfully	difficult to cast because the art of bronze
8	rong, normally yielding water can be	awfully	hard!)
9	You have to hit	awfully	hard.
10	It's	awfully	hard to ride this thing with numb legs!
11	it's	awfully	hard to straighten them you know?
12	It's	awfully	hard to get information about registered
13	doesn't look as if anybody's working	awfully	hard to get them, and I'd like to know why

None of the attested examples in Tab. 10 presents the MAKE *it awfully hard/difficult to*-infinitive schema. This makes it possible to reach the preliminary conclusion that the extended unit of meaning described in Section 2.2.1 is typical of Bush's rhetorical style. This is further supported by the results of a search for the adverb in two corpora containing contemporary political speeches. Blair's 2001-03 corpus contains approximately 250,000 words but no instance of *awfully*. Clinton's corpus of Presidential speeches delivered in 1993 contains about 751,000 words and six examples of the adverb, but it is never used in any of the pragmatic associations highlighted in Tab. 6 and Tab. 8a/b.

The *awfully kind* combination is attested twice in the BNC in the examples below:

1. That really is awfully kind of you. (Fiction)
2. Oh I wouldn't mind sir, thanks a lot, that's awfully kind of you. (Conversation)

These show that this combination is typically found as part of an extended unit of meaning used to express gratitude in general English. However, its use in Presidential discourse implies its re-contextualization as a formula which contributes to manage the interaction between the President and the audience; in particular, it aims to win their good-will in a strategic part of Bush's typical speech.

3. Conclusion

Corpus linguistics offers important resources and a useful analytical approach for the development of EFL learners' communicative competence. The theoretical principles of co-selection, lexical competence and priming illustrated in Section 1.3 substantiate a conception of language as function in context in which there is a close link between competence and performance, since these two dimensions can influence each other. Furthermore, these principles emphasize the central role of lexis in language learning.

The activities presented in Section 2 may be included in class work inspired by the communicative approach. In particular, the first activity (Section 2.1) has shown an important connection between examples of performance and their cultural context. This type of activity offers EFL learners the possibility to become aware of the complexities inherent in language learning and explore in depth the association between language and culture. Corpora are a source of input which students can exploit at different levels of competence, though perhaps advanced learners are those

who can benefit the most. In addition, work on corpus data provides learners with opportunities to reflect on the ways specific language phenomena are used. This is particularly true for the type of activity illustrated in Section 2.2 which demonstrates the expressive and functional dimensions of certain pragmatic schemata in relation to an overarching communicative purpose. Corpus analytical techniques can therefore be used to explore language variation and rhetorical stylistics both in a quantitative and qualitative fashion that can only benefit the EFL learner.

Bibliography

- Barnbrook, G. (1996) *Language and Computers*, Edinburgh, Edinburgh University Press.
- Biber, D. (1988) *Variation across Speech and Writing*. Cambridge, Cambridge University Press.
- Cobuild (1995) *Collins COBUILD English Collocations on CD-ROM*. London: HarperCollins.
- Firth, J. (1957) *Papers in Linguistics*. London, Oxford University Press.
- Freddi, G. (1994) *Glottodidattica. Fondamenti, metodi e tecniche*. Torino, UTET.
- Hoey, M. (2004) "The textual priming of lexis", in Aston, G., Bernardini, S., and D. Stewart (eds.) *Corpora and Language Learners*. Amsterdam, John Benjamins, 21-41.
- Hoey, M. (2005) *Lexical Priming. A New Theory of Words and Language*. London, Routledge.
- Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge, Cambridge University Press.
- Hymes, D. (1971) *On Communicative Competence*. Philadelphia, University of Philadelphia Press.
- Koller, V. (2004) "Businesswomen and war metaphors: 'Possessive, jealous and pugnacious?'" *Journal of Sociolinguistics*, 8 (1): 3-22.
- Krishnamurthy, R. (1996) "Ethnic, racial and tribal: The language of racism?", in Caldas-Coulthard, C. R. and M. Coulthard (eds.) *Texts and Practices. Readings in Critical Discourse Analysis*. London, Routledge, 129-149.
- Morley, J. (2004a) "The sting in the tail: persuasion in English editorial discourse", in Partington, A., Morley, J. and L. Haarman (eds.) *Corpora and Discourse*. Bern, Peter Lang, 239-255.
- Morley, J. (2004b) "Modals in persuasive journalism: an example from the *Economist*", in Facchinetti, R. and F. Palmer (eds.) *English Modality in Perspective. Genre Analysis and Contrastive Studies*. Frankfurt, Peter Lang, 67-82.
- Partington, A. (1998) *Patterns and Meanings. Using Corpora for English Language Research and Teaching*. Amsterdam, John Benjamins.
- Pinna, A. (2004) "The discourse prosody of 'deeply' in G. W. Bush's Presidential speeches". *Linguistica e Filologia*, 18: 169-194.

- Pinna, A. (2006) "Discourse prosody of some intensifiers in G. W. Bush's Presidential speeches", in Hornero, A. M., Luzon M. J., and S. Murillo (eds.) *Corpus Linguistics. Applications for the Study of English*. Bern, Peter Lang, 93-111.
- Pinna, A. (forthcoming) "Pragmatic markers of relevance, mitigation and solidarity in G. W. Bush's Presidential speeches", in Douthwaite, J., Jottini, L. and G. Del Lungo (eds.) *Atti del XXII Convegno dell'Associazione Italiana di Anglistica*. Università degli Studi di Cagliari (15th-17th September 2005).
- Sigley, R., and J. Holmes (2002) "Looking at 'girls' in corpora of English". *Journal of English Linguistics*, 30 (2): 138-157.
- Sinclair, J. (1991) *Corpus, Concordance, Collocation*. Oxford, Oxford University Press.
- Sinclair, J. (1994) "Trust the text", in Coulthard, M. (ed.) *Advances in Written Text Analysis*. London, Routledge, 12-25.
- Sinclair, J. (1996) "The search for units of meaning". *Textus*, 9: 75-106.
- Sinclair, J. (1998) "The lexical item", in Weigand, E. (a cura di), *Contrastive Lexical Semantics*, Amsterdam, John Benjamins, 1-24.
- Stubbs, M. (1996) *Text and Corpus Analysis*. Oxford, Blackwell.
- Stubbs, M. (2001) "On inference theories and code theories: Corpus evidence for semantic schemas". *Text*, 21 (3): 437-465.
- Tognini-Bonelli, E. (2001) *Corpus Linguistics at Work*. Amsterdam, John Benjamins.