Missing words:
The vocabulary of BBC Spanish courses for adults

Paul M. Meara, University of Swansea, Wales, United Kingdom
Jesús Suárez García, Barnard College, New York, USA

1. Introduction

Until very recently, vocabulary played a role in the production of language teaching materials that was very definitely a secondary one. Although inevitably present in any course book, vocabulary was generally subordinated to other elements that were considered more important to the process of learning a language. Thus, in the 1950s and 1960s, when grammar translation was the dominant approach to language teaching, the lexical content of courses was considered less important than the grammatical content of textbooks. To a large extent, the vocabulary taught was determined by the words that appeared in the works of classical authors whose texts were used for teaching purposes. Later, the structural or audiolingual methods, based on behaviourist theories subordinated the vocabulary to be taught to the linguistic structures that the student was required to automatize during the process of learning. In the 1970s, communicative approaches to language teaching began to be developed, partly as a reaction to the short-comings of earlier methods. *Functions* and *Notions* came to be the main curricular focus for language teaching, and functional-notional approaches dominated classroom practice. The vocabulary which appears in courses that were developed with this method in mind simply reflects the vocabulary which is used in the context chosen to introduce the functions and notions which the student is required to master – *In the restaurant*, *At the airport*, *At the doctor's* to list but a few of the typical ones.

The common assumption that underlies all these treatments of vocabulary is the idea that vocabulary acquisition takes place naturally when people learn other more basic elements: learn grammar, or structures, or functions, and you will inevitably learn vocabulary. Vocabulary acquisition is something that just happens on its own. One consequence of this is that the textbooks of the period show a surprisingly cavalier attitude towards the vocabulary items that they choose to teach. Words are included almost at random, and their selection depends heavily on the intuition of authors.

2. The study

This paper illustrates this problem in a series of beginners' Spanish courses. Specifically, we analysed a set of six textbooks published by the BBC over a period of 30 years (1965-1995). These courses were all aimed at adult learners of Spanish, working on their own,

and were primarily intended as a companion text for radio and television broadcasts. Inevitably, however, they were widely adopted as standard textbooks for adult learners attending classes as well. It is difficult to underestimate the influence of these courses on adult education, and it is no exaggeration to say that these courses pretty much defined the syllabus for adult learners of Spanish in the UK for a period of about thirty years.

The course books analysed and their date of publication are listed below:

- *¡Oigan Señores!* (1965)
- *Starting Spanish* (1967)
- *Zarabanda* (1971)
- *¡Dígame!* (1978)
- *España Viva* (1987)
- *Sueños* (1995)

Sueños, though published in 1995, is still available, and still widely used as support material for adult learners.

Our basic data was a list of the vocabulary items included in the glossary that each of these courses provides in the back of the course book. Our analysis consisted of a comparison of the words included in each glossary, with the intention of establishing how far the glossaries coincided and differed. We used some specially written computer programs to do this. The output of these programs was a list of words which appeared in each course. We also prepared a list of words which appeared only in one course, a list of words which appeared in two courses, a list of words appearing in three courses, and so on up to six. In English, the vocabulary syllabus has been influenced for a long time by the pioneering work of Harold Palmer and Michael West, which placed a lot of emphasis on frequency counts. For this reason, English courses – at least those published in the UK – tend to be fairly uniform in the words they choose to include in their vocabulary syllabuses. For other languages, this tendency is much less marked. We thought it would be interesting to compare these influential BBC courses with a view to establishing just how much agreement there was among the editors as to what words should be included in course materials of this sort.

Before we could carry out this analysis, it was necessary to edit the word lists for each course. This editing process is more complicated than we might expect, because the different courses treat words in different ways. Specifically, the way grammatical function words were handled was not consistent. Some courses included words like *el*, *la*, *un*, *una* and pronouns in their word lists, while others did not. Furthermore, the way morphological variants of words were handled was not consistent either. Some texts listed masculine and feminine forms separately, while others did not. Some texts listed irregular forms of verbs separately. Some courses also included multi word phrases such as *ir de compras* in their glossaries, whereas others did not, and these inconsistencies also needed to be treated in a uniform way. We needed to standardise the lists in order to carry out the necessary comparisons.

The changes that we implemented to achieve this standardisation mainly consisted in removing items which were not single words, or were morphological variants of other words. These included:

- Multi word expressions which could not be considered as single lexical items (*hace frío, ponerse al teléfono*);
- regular feminine forms (*vieja, gorda, hermana*);
- plurals;
- non-infinitive forms of verbs;
- superlative adjectives.

In all these cases, only the base form of the word was counted.

We also standardised variant forms of words which appeared in different forms in different texts, e.g. *espárrago/espárragos*, *verdura/verduras*, etc. In these cases, we used a single form.

Although some of these changes are not straightforward to implement, they allowed us to simplify the word lists, and once this was done it was a relatively routine process to run comparisons between them. We also used other sources to throw light on the vocabulary of the BBC courses. These were the vocabulary lists contained in *Un Nivel Umbral* (Slagter, 1979), and *Diccionario de Frecuencias de las Unidades Lingüísticas del Castellano* (Alameda, Cuetos, 1995).

Un Nivel Umbral is a Spanish adaptation of the *Threshold Level* vocabulary developed by a team led by Peter Slagter with the support of the Council of Europe. In this list, vocabulary is taught as part of a notional and functional syllabus aimed at beginning adult learners. The aim, both in the Spanish list, and in the other languages for which a *Threshold Level* has been elaborated, is to establish a basic vocabulary which will allow a person travelling in Europe to communicate with speakers of the local language. Bearing this aim in mind, it seemed reasonable to compare the vocabulary that appears in the BBC courses with the vocabulary that is listed in the *Nivel Umbral* syllabus.

At the time we did this work, the *Diccionario de Frecuencias de las Unidades Lingüísticas del Castellano* was the most recent and the most complete work of its kind that had been published in Spanish, although a number of other word counts have appeared more recently. We used this word count to determine which levels of frequency covered the vocabulary included in each of the BBC courses, and this allowed us to determine how far frequency of use was indeed a factor in the selection of words for inclusion in these courses.

3. Results and analysis

The first analysis that we carried out was a comparison of the six textbooks against each other. The results of this analysis are reported in *Table 1*.

In *Table 1*, the first column lists the six textbooks under review, while *Column Total* shows the number of different words that each of them contains. These totals vary considerably. *¡Oigan Señores!* teaches a basic vocabulary of only 675 words, while *Sueños*, the text with the largest vocabulary teaches a basic vocabulary of 1,956 words – almost three times as many. *Table 1* also allows us to make a comparison with the number of words listed in the lexical component of the *Nivel Umbral* syllabus (1,158 words). *Table 1* shows that two of the BBC course books have vocabularies which are considerably larger than the vocabulary in the *Nivel Umbral* list. *Dígame* aims at 1,549 words, while *Sueños* aims at 1,956 words. The other courses all have considerably smaller vocabularies.

Course	Total	n1	n2	n3	n4	n5	n6
¡Oigan Señores!	675	174	122	113	94	74	98
Starting Spanish	941	149	135	152	189	218	98
Zarabanda	753	101	99	109	143	203	98
Dígame	1,549	454	286	259	231	221	98
España Viva	951	154	170	165	168	196	98
Sueños	1,956	865	316	240	219	218	98
Nivel Umbral	1,158						

Table 1: Number of words in each textbook (Total) *and words appearing in more than one textbook* (n1-n6)

We might imagine that there would be a large degree of overlap in the basic vocabulary taught by the six courses. Surprisingly, this turns out not to be the case, and we can see this in the right hand columns of *Table 1*. The numbers in this part of the table require some explanation. *Column n1* shows the number of words that are unique to each course. Thus, *¡Oigan Señores!* contains 174 words which do not appear in the word lists provided by any of the other courses. *Column n2* shows the number of words that each course shares with one other book. Thus, *¡Oigan Señores!* shares 122 words with one other course – but these 122 words are shared out among the other five courses, rather than all appearing in a single text. *Columns n3* to *n6* show the number of words that each course shares with 2, 3, 4 and 5 other courses. Thus, *¡Oigan Señores!* contains 113 words which appear in three other textbooks, 94 that appear in four courses, and 74 that appear in five courses. Only 98 words appear in the word lists of all six courses – a figure which is surprisingly small. This list of 98 words is shown in *Table 2*.

agradable	cenar	desde	llevar	parecer	salir	tienda
ahí	cerca	dormir	luego	pasar	seguir	trabajar
avión	cerrar	empezar	mano	película	sello	trabajo
azul	cerveza	entrada	manzana	periódico	semana	tren
bailar	ciudad	habitación	marido	piso	sin	último
barato	claro	iglesia	mediodía	playa	sitio	ver
bastante	comida	invierno	mes	plaza	sobre	verano
billete	cosa	jugar	mismo	precio	sol	verdad
cabeza	creer	lado	necesitar	preferir	solo	vez
caja	cuarto	lavar	niño	pueblo	sopa	viajar
carne	cuenta	leche	novio	puerta	sólo	viaje
caro	dejar	leer	nuevo	quedarse	tercero	viejo
casi	demasiado	lejos	número	querer	terminar	volver
cena	descansar	llamar	pagar	saber	tiempo	ya

Table 2: The 98 words appearing in all six course books

Clearly, there is little evidence here of a principled selection underlying the choice of words to be taught. One is tempted to note that the list allows women to talk about their husbands, but does not allow men to talk about their wives, that blue is the only colour in the list, although the list does not contain any noun that is likely to be blue, that the list contains *semana* and *mes* but not *día* or *noche*. The presence of *tercero* (but not *primero* or *segundo*) is also surprising. This view that the list of shared words is a bit odd is reinforced by the data shown in *Table 3* which compares the *Nivel Umbral* vocabulary with the vocabulary of the six textbooks. Of the 1,158 words that appear in the *Nivel Umbral*, 157 also appear in one of the textbooks, 164 appear in two, 188 appear in three courses, 188 appear in four, 180 appear in five courses, and only 95 appear in all six. 285 words which appear in the *Nivel Umbral* do not figure in any of the six textbooks.

	Total	1 text	2 texts	3 texts	4 texts	5 texts	6 texts
N Umbral	1,158	157	164	188	188	180	95

Table 3: Where the vocabulary of Nivel Umbral *appears in the texts.*

Several additional points are worth making here. The first point concerns the very large differences in the vocabulary learning load which the different courses expect of their students. The mean number of vocabulary items is 1,136 but there is substantial variation around this figure. It strongly suggests that there is no real agreement here about what might be considered a reasonable vocabulary learning load for adult beginners. Given that the courses all apparently have the same objectives – to prepare the students to communicate in Spanish and to introduce them to the typical aspects of Spanish culture – this lack of consensus among the courses is particularly striking.

The second point that stands out is that only 98 words appear in the glossaries of all six courses. These words include a number of items which are not among the 2,000 most frequent words in Spanish – *sello* (stamp), *avión* (aeroplane), *sopa* (soup), for instance – though, of course, the relevance of these words to tourists will be readily apparent. More importantly, perhaps, the list does not contain a number of highly frequent words which we might expect to find in a list of common words. *Pequeño* (small), *ahora* (now), *hombre* (man), *mujer* (woman) are all absent from this list, despite the fact that they are listed among the 500 most common words in Spanish. Surprisingly, perhaps, of the 98 words which appear in all the textbooks, only 60 come from the first thousand most frequent words in Spanish. Ten percent of these shared words are 'unusual' words that do not appear in the two thousand most frequent words in Spanish.

The texts vary very widely in the frequency of their selected vocabulary, but in general, the vocabulary they teach does not seem to pay much attention to frequency as a selection criterion. The best text in this regard is *Zarabanda*, where 40% of the vocabulary taught falls within the 1K frequency band. For the other texts, the figures are much worse.

Criticising textbooks for not taking account of objective frequency criteria is probably a little unfair. At the time these texts were produced there were very few analyses of vocabulary frequencies in Spanish, and the tradition of counting word frequencies in English had had little impact on the teaching of Spanish, at least in the UK. The most widely available frequency count at the time was Juilland and Chang Rodríguez (1964), a text which was not widely known outside academic circles, and was in any case primitive by comparison with the standard English word lists. It was based on a very restricted set of genres – mainly fiction published between 1920 and 1940 – and the corpus on which the count was based on was a relatively small one – a mere 500,000 words. The application of corpus linguistics to teaching Spanish was still a long way in the future (Sánchez et al., 1995) and would take a long time to influence Spanish teaching in the UK.

What is more surprising is that the authors of these course books seemed to be unaware of the curriculum work that was being carried out by the Council of Europe over the period of publication that we are looking at. The *Threshold Level* syllabus was explicitly developed for adult learners travelling in Europe, precisely the target audience that these BBC courses were aimed at. We might expect that there would be a considerable overlap between the words taught in these courses and the emerging functional syllabuses. Surprisingly, this does not appear to be the case.

Table 4 shows the extent of the overlap between the BBC courses and the vocabulary of the *Nivel Umbral*. Large overlapping vocabularies are found for *Dígame* and for *Sueños*, but this appears to be an accidental result of the fact that these two courses teach large vocabularies anyway. There is no real evidence that the Council of Europe material influenced the vocabulary lists of these courses, and no real evidence that the publication of the *Nivel Umbral* in 1979 marked a watershed in the production of materials aimed at adult learners.

course book	words only in course book	shared words	words only in *Nivel Umbral*
¡Oigan Señores!	338	337	821
Starting Spanish	393	548	610
Zarabanda	274	479	677
Dígame	860	689	469
España Viva	434	517	641
Sueños	1,261	695	463

Table 4: Overlap between the six courses and the vocabulary of Nivel Umbral

4. Conclusion

This brief study suggests that the lexical syllabuses adopted by the BBC in their courses for adult learners of Spanish were largely unsystematic and unprincipled. The choice of vocabulary was mainly driven by ad hoc principles, particularly the random occurrence of words that just happened to appear in a particular context. All words are treated equally, and it is not the case that highly frequent words are singled out or given special treatment.

The very small number of words that appear in all six courses suggests that this ad hoc approach to vocabulary selection may not always be successful. It raises a number of interesting questions about the sort of performance we would expect from learners who have followed each of these courses to their conclusion. What would happen if we put together six students each of whom had followed one of these courses? Would they actually be able to communicate with one another? And how would they each perform if we took them to Spain at the end of their courses, and set them loose on an unsuspecting Hispanic speaking word? Would students who had followed a vocabulary rich course (like *Sueños*) be better able to communicate than students who had followed one of the lighter courses? Indeed, we might add: are students who follow one of the vocabulary rich courses less likely to complete these courses, simply because they are being overloaded with words?

Nation (1990) identifies several reasons why a systematic approach to vocabulary selection seems worthwhile:

- Whichever method of vocabulary teaching we adopt, teaching vocabulary is inevitable;
- there is a large body of research about how vocabulary can be taught effectively and which vocabulary is most suitable for learners;
- there is a lot of empirical evidence which suggests that some vocabulary learning methods are more effective than others;
- students themselves almost always report that vocabulary is their biggest single stumbling block in learning an L2.

The results of this study suggest that ad hoc selection methods do not necessarily provide a good vocabulary learning environment for students. They do, however, lead to textbooks which contain a lot of vocabulary which is of limited value, and fail to provide contexts which teach learners some of the basic vocabulary that they need.

Notes

[1] It is perhaps worth noting here that the *Nivel Umbral* list is rather smaller than the basic vocabularies we find in other versions of the *Threshold Level* materials published by the Council of Europe, which typically contain about 2,000 words. It is not clear why the *Nivel Umbral* is smaller than this typical value.