

- Koskinen, K. 2000. "Institutional illusions: Translating in the EU Commission". *The Translator* 6: 49–66.
- Melby, A. K. 1996. *The Possibility of Language*. Amsterdam: Benjamins.
- Pinker, S. 1994. *The Language Instinct*. New York: William Morrow.
- Putnam, H. 1975. "The meaning of 'meaning'". Reprinted in H. Putnam, *Mind, Language and Reality. Philosophical Papers 2*. Cambridge: Cambridge University Press.
- Teubert, W. 1996. "The concept of work in Europe". In *Conceiving of Europe: Diversity in Unity*, A. Musolff, C. Schäffner and M. Townson (eds), 129–145. Aldershot: Dartmouth.
- Tognini Bonelli, E. 1996. *Corpus: Theory and Practice*. Birmingham: TWC.

Bilingual lexicography, overlapping polysemy, and corpus use

Victòria Alsina and Janet DeCesaris

1. Introduction¹

Both researchers interested in improving the quality and usefulness of dictionaries and lexicographers have welcomed the advent and availability of large computerized corpora. Representative bilingual or multilingual corpora are possible in specialized fields because in these well-defined situations set in multilingual environments the subject domains are quite restricted. Bilingual or multilingual corpora consisting of texts based either on translations produced by highly trained professionals or on comparable text production thus play an essential role in ensuring that specialized dictionaries, glossaries and terminologies actually reflect the language used in the workplace. However, the tasks and data facing the general language bilingual lexicographer are rather different in nature from the delimited contexts just mentioned: the kind of corpus which proves most useful in the construction of bilingual dictionaries is not yet well defined. While many modern monolingual dictionaries depend heavily on corpus-based data, bilingual lexicography has yet to determine what type of corpus best serves the needs of general bilingual dictionaries. This would seem to be yet another manifestation of the fact that bilingual lexicography lags behind monolingual lexicography (Hartmann and James 1998: 15).

Many researchers have noted that the typology of potential users of general bilingual dictionaries is quite varied, ranging from advanced learners to experienced translators (Al-Kasimi 1983: 154–157, Tomaszczyk 1983: 46). Bilingual dictionaries of this sort are used for both encoding and decoding by speakers of two different languages with several levels of language skills and thus must incorporate a great deal of grammatical and pragmatic information. In corpus-

based bilingual lexicography, the two alternatives previously hinted at are 'parallel corpora', which contain one set of texts in two or more languages, and 'comparable corpora', which contain texts in several languages with the same or similar composition. Teubert presents a cogent discussion of both types (1996:245–249), and concludes (rightly, we think) that "ideally, parallel corpora should be viewed as complementary to comparable corpora" (1996:252). Parallel corpora run the risk of presenting data produced under the special conditions of translation, which may be significantly different from 'regular' native-speaker production. It is a well-known fact in translation theory that "phenomena pertaining to the make-up of the source text tend to be transferred to the target text, whether they manifest themselves in a *negative transfer* (i.e., deviations from normal, codified practices of the target system), or in the form of *positive transfer* (i.e., greater likelihood of selecting features which do exist and are used in any case)" (Toury 1995:275). It is true that interference need not be seen as an undesirable trait in translation. Indeed, "its undesirability is always a function of a host of socio-cultural factors, which may therefore be said to condition our law" and "communities differ in terms of their resistance to interference, especially of the 'negative' type" (Toury 1995:277). Nevertheless, the inevitable presence of interference or transfer in translated texts does bear directly on the data to be found in parallel corpora, and makes us question the reliability of parallel corpora as the primary source of data for a general language bilingual dictionary. Comparable corpora would seem preferable for this type of dictionary project, but their use will not be addressed in this paper because we know of no such corpus data available for general purpose language in the language combinations we discuss.²

The two main problems we have mentioned with bilingual corpora, the presence of interference and unavailability, do not plague monolingual corpora of English. Since reliable, contemporary corpus data is widely available for English, we decided to see how it could be used to improve the information currently provided in English/Spanish and English/Catalan dictionaries. In order to determine the possible role for monolingual corpus data in the preparation of these dictionaries, we must first identify the main problems that beset existing dictionaries. We have therefore chosen three non-derived, polysemous adjectives in English that we were sure to find amply covered in current English/Spanish and English/Catalan dictionaries and in a corpus of English: *cold*, *high* and *odd*. Existing dictionary entries for these words were analyzed to pinpoint what needed improvement, and then the *British National Corpus* was consulted to see how it might help resolve the issues resulting from the dictio-

nary analysis. We conclude that data from a monolingual corpus proves useful for addressing some of the main problems associated with providing equivalents for adjectives in a general-purpose bilingual dictionary, such as order of presentation, repetition of equivalents due to what we will define as overlapping polysemy, and decisions regarding examples, but has little bearing on the issue of delimiting possible contexts in which the equivalent provided by the dictionary is appropriate.

2. Methodology

We looked up the entries for the three adjectives in three English/Spanish bilingual dictionaries and one Catalan/English dictionary. The bilingual dictionaries consulted in the case of Spanish were *The Oxford Spanish Dictionary (OSD)*, *Larousse Gran Diccionario Español-Inglés/English-Spanish (GL)*, and *Simon & Schuster's International Dictionary English-Spanish/Spanish-English (S&S)*, which were chosen for the following reasons. First, we were interested in analyzing entries in recently published dictionaries which would reflect contemporary usage. The *OSD* in particular is noteworthy in this respect, as its first edition was published in 1994. Second, we deliberately included dictionaries produced by both British and American publishers. Third, it has been our personal experience as translators and teachers of translation that all three of these dictionaries are useful, that is to say, we ourselves use them and recommend them to our students. The English/Catalan dictionary analyzed is the *Diccionari anglès-català* published by Gran Enciclopèdia Catalana (*DAC*), which is the most comprehensive bilingual dictionary for this language combination currently available.

The entries from bilingual dictionaries were compared with those from three monolingual dictionaries, the second edition of the *Collins Cobuild English Dictionary (Cobuild)*, the *Cambridge International Dictionary of English (CIDE)* and the third edition of the *American Heritage Dictionary (AHD3)*. The choice of these particular dictionaries was also not random: *Cobuild* is the prime example of a corpus-based dictionary in English, and because it is aimed at advanced learners of the language its target audience coincides to a large extent with the users of the bilingual dictionaries under examination. *CIDE* is addressed to the same target audience, states that a corpus was used in its preparation although it does not purport to be corpus-based in the same way as does *Cobuild*, and has a very nice way of dealing with polysemy in that senses

are clearly grouped together under differentiated basic concepts. *AHD3* covers American English, a variety of the language which is explicitly included in the bilingual dictionaries and not, we feel, well represented in *Cobuild* (it is somewhat better represented in *CIDE*). Although *AHD3* is not based on a corpus, it does claim to rank the order of senses on the basis of usage, as opposed to the historical order of senses employed by other well-known American dictionaries such as *Merriam-Webster's Collegiate Dictionary*.

The corpus consulted, as mentioned above, was the *British National Corpus (BNC)*. The *BNC* "was designed to characterise the state of contemporary British English in its various social and generic uses" (Aston and Burnard 1998: 28). It includes both informative and imaginative texts, and comprises 90% written texts and 10% spoken texts. In spite of the design features of the *BNC* that might lead to controversial linguistic generalizations about general purpose English, we believe it provides a sufficiently accurate picture of British English to allow comparison of data culled from it with data from dictionaries. The number of examples of the three adjectives in this corpus are as follows:

high	28,698 examples in 3,243 texts
cold	6,438 examples in 1,592 texts
odd	4,478 examples in 1,595 texts

For the purposes of this article, we decided to examine 500 examples of each adjective randomly chosen by the search function, taken from both written and oral language, and with only one example from any given text. Although 500 is only 1.7%, 7.7%, and 11.2% respectively of the totals available for these words, this number proved workable from a practical standpoint in terms of downloading and producing a clean set of examples.

3. Analysis of the dictionary entries and comparison with information extracted from the corpus sample

The three adjectives, *cold*, *high* and *odd*, exhibit varying degrees of polysemy, as can be seen in the summary of the definitions listed in the monolingual dictionaries in Table 1. *AHD3* consistently makes more sense distinctions than the other two monolingual dictionaries. We believe this difference may be attributed to two factors: (1) *AHD3* is a more comprehensive dictionary, with more entries than either of the other two dictionaries; and (2) unlike *Cobuild* and *CIDE*, *AHD3* is not addressed to foreign learners of English, and thus includes less fre-

quent, even uncommon, senses of words which are unlikely to be consulted by advanced learners but nevertheless are not insignificant in the context of a comprehensive dictionary for native speakers. The three dictionaries differ somewhat from one another in the number of senses assigned to each adjective, but in this respect we think the best guide is *CIDE*, which has made a noteworthy effort to limit itself to only the *basic* senses (which is particularly important in bilingual dictionaries), whereas the other two, especially *AHD3*, tend to assign a separate sense or subsense to every nuance of meaning. We will therefore be referring mainly to *CIDE* when we discuss the number of senses of each word.

Table 1. Definitions in the monolingual dictionaries

	AHD3	Cobuild	CIDE
<i>cold</i>	10 senses, divided into 18 subsenses	8 senses, plus some expressions	2 senses
<i>high</i>	13 senses, divided into 22 subsenses	15 senses, plus some expressions	7 senses
<i>odd</i>	7 senses, divided into 9 subsenses	5 senses, plus some expressions	5 senses

3.1 The adjective *cold*³

Cold is a polysemous adjective in English with two main senses: (1) 'having a low temperature', and (2) the metaphorical sense of 'unfeeling' or 'unfriendly'. Both of these senses correspond almost exactly to the two main senses of *frío*, in Spanish, and *fred*, in Catalan. Tables 2 and 3 below show the order of senses in the monolingual and bilingual dictionaries respectively.

The data in Tables 2 and 3 show that all the dictionaries consulted, even the one with the simplest structure (*DAC*), give the sense 'low temperature' first, thus reflecting the intuition of the lexicographers that it is the most frequently used sense. The sense 'unfriendly' or 'unfeeling' is generally, but not always, given in second position. Some dictionaries acknowledge up to five additional different senses, although we feel that these could be included in one of the two

Table 2. Treatment of *cold* in monolingual dictionaries

Senses of <i>cold</i>	AHD3 (10 total senses)	Cobuild (8 total senses)	CIDE (2 total senses)
low temperature	1–2, 8	1–4	1
unfriendly	3–5	6	2
cold colors	6	5	–
trail or scent	7	7	–
wrong	–	8	–
dead	9	–	–

Table 3. Treatment of *cold* in bilingual dictionaries

Senses of <i>cold</i>	OSD (4 total senses)	GL (6 total senses)	S&S (12 total senses)	DAC (Catalan) (1 total sense)
low temperature	1	1, 2	1	1
unfriendly	2	4	2, 3, 9	1
cold colors	–	5	12	–
trail or scent	1	6	5, 6	–
wrong	1	–	7	–
dead	–	3	10	–

main senses (as in *CIDE*). The varying number of senses in the dictionaries reflects two different problems: (1) exactly what constitutes a separate sense is not always clear, even to trained lexicographers; and (2) some dictionaries list highly lexicalized examples as separate senses, even though the meaning could be included as part of an earlier sense. This latter issue is particularly evident in the case of bilingual dictionaries and explains why there may be more senses listed for a word in a bilingual dictionary than in a monolingual dictionary.

It is a fact that English *cold* and Spanish *frío* / Catalan *fred* generally coincide in terms of physical reference and possible metaphorical contexts. This can be seen from the entry for *cold* in the *OSD*, in which the same equivalent (*frío*) is provided for *cold* numerous times.

*cold*¹ ... *adj* 1 <water/weather/drink> *frío*; I'm ~ tengo *frío*; my feet are ~ tengo los pies *fríos*, tengo *frío* en los pies; it's ~ today/in here hoy/aquí hace *frío*; the soup is ~ la sopa está *fría*; I'm getting ~ me está entrando *frío*; it's getting ~ está empezando a hacer *frío*; you dinner's getting ~ se te está enfriando la comida; the water has gone ~ el agua se ha enfriado; the engine starts straight from ~ without fail el motor arranca en *frío* sin fallar; the trail has gone ~ se han borrado las huellas; the news was already ~ la noticia ya estaba pasada or añeja; no, you're still ~, getting ~er (in game) no, *frío*, más *frío*; ⇒ *blow*² *vi* 1(a)
 2 (a) (unfriendly, unenthusiastic) <person/stare/color> *frío*; I got a very ~ reception me recibieron con mucha frialdad or muy *fríamente*, la recepción que me dieron fue muy *fría*; to be ~ TO or WITH sb tratar a alguien con frialdad, estar*/ser* *frío* con algn; to go ~ on sth; I went ~ on the idea (colloq) la idea dejó de hacerme gracia (fam); to leave sb ~: that leaves me ~ (colloq) (eso) me deja *frío* or tal cual (fam), (eso) no me da ni *frío* ni calor (fam) (b) (impersonal) <logic> *frío*; keeping to the ~ facts ... ateniéndose únicamente a los hechos ...
 3 (unconscious) ⇒ *out*² 1(b)
 4 (without preparation) sin ninguna preparación; I came to the job ~ empecé el trabajo sin ninguna preparación; I was expected to start from ~ esperaban que empezara sin ninguna preparación.

Figure 1. Entry for *cold* in the *OSD*

Such repetition is in no way limited to the *OSD*, but quite commonplace in entries that represent what we call *overlapping polysemy*. In overlapping polysemy, a word in one language is polysemous, and there exists an equivalent word in the other language that, by and large, exhibits the same polysemy. Overlapping polysemy is a manifestation of what Sinclair (1996: 179) termed “parallels between the textual environment of a word in one language and a word that is used to translate it in another.” At this point we are not as concerned with the causes behind overlapping polysemy as we are with its effects on bilingual lexicography, although we might speculate that a cognitive linguistics approach to metaphor in language could be enlightening. We have found examples of overlapping polysemy in English on the one hand and Spanish and Catalan on the other in all word classes, for example: verbs, Eng. *run* /Sp. *correr*, Cat. *córrer* (‘walk quickly’ and ‘run a risk’); prepositions, Eng. *before* /Sp. *antes*, Cat. *abans* in both spatial and temporal contexts; nouns, Eng. *dough* /Sp. and Cat. *pasta* referring both to a mixture of flour and water and to money; and adverbs, Eng. *naturally* / Sp. *naturalmente*, Cat. *naturalment* meaning in a natural (as opposed to unnatural) way or expressing the expectedness of an outcome. The existence of overlapping polysemy has not gone unnoticed in the literature; for example, Tognini-Bonelli (1996: 207–214) discusses a case of overlapping polysemy with reference to English *real* /Italian *reale* in some detail from the perspective of using corpora to identify translation equivalents.

Given the overlapping polysemy exhibited by *cold/frío-fred*, we might have expected that the entries in the bilingual dictionaries would have taken advantage of the overlap and would hence turn out to be simpler and shorter than those in the monolingual dictionaries. In fact, however, several situations occur: the structure in the *DAC* is quite simple; the structure in the *OSD* is somewhat more complex, but is less so than that of either *GL* or *S&S*, both of which contain long entries with many senses. In fact, in these latter two dictionaries it appears that little to no attempt has been made to organize the material.

After our initial analysis of the entries for *cold* in the bilingual dictionaries, we are now in a position to identify areas in which the dictionaries differed from one another and which are, perhaps, potential points for improvement:

- the design of the entry, to take advantage of overlapping polysemy when it exists;
- the criteria for ordering the equivalents in the entry; and
- decisions determining which set phrases or idioms should be afforded equivalent translations in the entry.

We now turn to the corpus data relating to *cold*, to see if it bears on any or all of these issues. We were able to use 489 of the 500 examples containing *cold* that were downloaded from the *BNC*; in 11 examples the context provided by the search was not explicit enough to determine the sense of *cold* being used. As seen in Table 4, the sample showed that the 'low temperature' sense is by far the most frequently used in English.

Table 4. *Cold* in the BNC sample

Senses of <i>cold</i>	Number of examples in BNC sample
low temperature	358
unfriendly	54
giving impression of low temperature	6
expression: <i>cold war</i>	35
expression: <i>in the cold light of morning/day/dawn</i>	10
expression: <i>cold comfort</i>	6
expression: <i>cold feet</i>	6
expression: <i>cold shoulder</i>	6
other collocations	8

The group of 358 includes several collocations in which the sense 'low temperature' was clear to us (e.g. *cold sweat*), and 22 figurative uses such as that exemplified by the expression *reality closed its cold hand around her* in which the sense of 'low temperature' was still evident. The sample included a significant number of lexicalized collocations, which are particularly important for bilingual dictionaries since they can constitute exceptions to the almost perfect equivalence between *cold* and *frío/fred*. The most frequent of these collocations was *cold war* (35 cases) — the equivalent of which is the loan translation *guerra fría / guerra freda* — and this large number of cases no doubt reflects the fact that much of the textual base of the *BNC* is journalistic and thus concerned with politics. There is also a third, metaphorical, sense of *cold* which we identify as 'giving the impression of low temperature'. This sense is used in contexts in which *cold* is applied to nouns referring to color, light or appearance (e.g. *cold grey/gleam/outlines/full moon*).

All of the above-mentioned senses exemplified in the corpus data, and others which did not appear in our 500 examples but which no doubt would have turned up in a larger sample from the corpus, such as *cold sore*, are present in the dictionaries examined. In addition, the dictionaries assigned the most frequent meaning to the first sense in the entry. The only exception to these observations is *in the cold light of day*, which surprisingly is not present in any of the dictionaries in spite of being the second most frequent lexicalized collocation.

Moreover, the meaning of this expression is opaque and would not be immediately understood by foreign speakers.

3.2 The adjective *high*

Tables 5 and 6 show the order of senses for the polysemous adjective *high* in the monolingual and bilingual dictionaries consulted. We may first note that, as in the case of *cold*, *AHD3* makes more sense distinctions than the other two dictionaries. The first two or three senses refer to physical height in all three dictionaries; what is perhaps surprising is the variation in the order of the other senses — for example, the sense referring to the foul smell of meat is third out of thirteen in *AHD3*, last of seven in *CIDE*, and absent from *Cobuild*.

Table 5. Treatment of *high* in monolingual dictionaries

Senses of <i>high</i>	AHD3 (13 total senses, with 22 subsenses)	Cobuild (15 total senses)	CIDE (7 total senses)
distance	1	1, 2, 3	1
above average	8	4, 5, 7	2
important	2, 6, 7	8, 9	3
mental state	10	16	4
sound	4	14	5
education			6
bad smelling	3		7

Table 6. Treatment of *high* in bilingual dictionaries

Senses of <i>high</i>	OSD (6 total senses)	GL (27 total senses)	S&S (18 total senses)	DAC (Catalan) (1 total sense)
distance	1	1, 2	1	1a
above average	2	3, 5, 12, 21, 23, 24		1b
important	1	1, 7, 8	4, 5, 9	1b
mental state	4	27	17	1b
sound	1	10, 11	13	1b
education				1b
bad smelling	6	17, 19	11	

Table 6 clearly shows that the treatment of *high* in the bilingual dictionaries is even more diverse. At one extreme there is *GL*, which distinguishes 27 different senses, not counting the large number of examples provided; for its part, *S&S* includes 18 senses. We note that in both these dictionaries the number of senses

is greater than that given by the monolingual dictionaries, although it must be said that neither of these two bilingual dictionaries includes subsenses, which might have lowered the number of main senses substantially. *OSD* divides its entry into 6 senses, and the English/Catalan dictionary, *DAC*, as it did with *cold*, lists only one basic sense divided into two subsenses, one for the physical sense and the other for the metaphorical sense. Although the structure of the entry is much simpler in the *DAC* than in the other dictionaries, the entry itself is not much shorter because there are many examples under each subsense. And, like the monolingual dictionaries discussed above, the bilingual dictionaries list equivalents for the physical reference of the adjective first and then differ as to the order of the derived senses presented.

The study of the 500 examples taken from the *BNC* (492 of which we were able to use) confirms that *high* is indeed a highly polysemous word. Table 7 summarizes the distribution of senses we found in our sample. Perhaps the most striking aspect of this information is the fact that the historically original sense of 'extending (relatively) far upwards' or 'placed at a great distance from the bottom', i.e. the sense giving a physical description that is listed first by all the dictionaries examined, is obviously not the most frequent, although it is by no means rare. The most frequently found sense was that of 'situated at the top part of the scale' when applied to nouns describing objectively measurable qualities, such as *pressure* (13), *rate(s)* (18), *degree* (10), *price* (10), *cost(s)* (10), *level* (29), *proportion* (9). A closely related sense, but with the further component of 'good' or 'positive', is also frequently found, and occurs when the adjective is paired with nouns describing qualities involving a subjective assessment: *quality* (13), *class* (3), *standard(s)* (17), *reputation* (4), and *performance* (3). Yet another frequent meaning is that of 'important', 'above others in its class', which we found in collocations such as *high court* (16), *high school* (12), *high commissioner* (7), and *high priest* (3). There are a few cases in which *high* is

Table 7. *High* in the BNC sample

Senses of <i>high</i>	Number of examples in BNC sample
top of scale	198
top of scale + good	76
top of musical scale	7
important, above others	71
physical height	82
expression: <i>high street</i>	25
expression: <i>high profile</i>	6
other collocations	27

applied to sound as in *high pitch*. Finally, *high* is present in a number of collocations such as *high street*, *high profile*, *high priority*, *high time*, etc.

To summarize up to this point, we have seen that the bilingual dictionaries examined take little notice of overlapping polysemy, even when it is almost complete (the case of *cold/frio/fred*). The main exception to this observation is the *DAC*, which gives one main equivalent and then several expressions. The bilingual dictionaries differ from one another with regard to the order of equivalents, especially in the case of figurative senses, and with regard to which equivalents are included. And, finally, relatively small samples from the corpus contained frequent fixed expressions and collocations that are legitimate candidates for translation equivalents. Not all of these expressions are included in the dictionaries under examination.

3.3 The adjective *odd*

To judge from the number of senses listed in *AHD3*, *odd* is not as polysemous as either *high* or *cold*, and examination of the bilingual dictionaries shows us that it exhibits practically no overlapping polysemy with its equivalents. The five descriptors used by *CIDE* for *odd*, viz. 'strange', 'separated', 'numbers', 'not often', and 'approximately' (as an affix), do not correspond to either one or two adjectives in Spanish or Catalan. Since overlapping polysemy does not really

Table 8. Treatment of *odd* in monolingual dictionaries

Senses of <i>odd</i>	AHD3 (7 total senses)	Cobuild (5 total senses)	CIDE (5 total senses)	(Historical)
strange	1	1	1	5
separated	3, 4	5	2	1
numbers	5	4	3	2
not often	6	2	4	4
approximately	2	3	5	3

Table 9. Treatment of *odd* in bilingual dictionaries

Senses of <i>odd</i>	OSD (4 total senses)	GL (19 total senses)	S&S (9 total senses)	DAC (Catalan) (5 total senses)
strange	1	13	7, 8	1
separated	3	5, 6	1	4
numbers	3	1	2	2
not often	2	8	6	5
approximately	4	3	3	3

come into play with this adjective, we must concentrate on the other two issues that we have identified as needing improvement: order of equivalents and inclusion of fixed expressions. Table 8 above shows the order of senses in the monolingual dictionaries (we have included the historical order of senses (according to the *Oxford English Dictionary*) in the final column as a point of comparison), and Table 9 the order of equivalents in the bilingual dictionaries.

The sample from the *BNC* shows that *odd* is much more common in the sense of 'strange' than in any of the other senses; the mathematical sense and the sense of 'not matched' or 'not part of a pair', which are historically older and, we believe, still important senses of this word, were quite few in number. The figures from the corpus sample are given in Table 10.

Table 10. *Odd* in the *BNC* sample

Senses of <i>odd</i>	Number of examples in <i>BNC</i> sample
strange	257
occasional	151
numbers = not even	24
approximately	31
not matched	8
expression: <i>odd jobs</i>	11
expression: <i>odd man out</i>	8

The expression *odd man out*, which is present in the bilingual dictionaries and explicitly explained in both *Cobuild* and *CIDE*, turns out to be less frequent than the phrase *odd jobs*. If we compare these findings with the entries in the bilingual dictionaries, we see that *OSD* and *DAC* list the 'strange' equivalent first, but note that *S&S* begins with *suelto* 'unmatched', historically the first meaning, while *GL* begins with *impar* 'not even' and gives the reader the 'strange' sense in 13th position.

This example brings up an important issue in relation to corpus-based lexicography, namely how to evaluate senses with relatively few occurrences in the corpus. The 'not matched' sense of *odd* was relatively infrequent in our sample, which might lead lexicographers to omit it from a bilingual dictionary, but we believe that would be an error because this meaning cannot be derived from other information and it is perceived by speakers as a basic sense of the word. Frequency data *alone* are not enough to determine the inclusion of a sense in a general purpose dictionary. By contrast, the same number of occurrences should be interpreted quite differently when dealing with an expression. Our sample shows that the expression *odd jobs* is quite frequent and always used in

the plural (with the meaning of 'several unrelated, not regular jobs'), and this information can help guide lexicographers in their choice of examples.

4. Conclusions

A major problem with equivalents in bilingual dictionaries is the identification of the range of semantic contexts in which the equivalent provided by the dictionary can be used. Adjectives that have more than one sense are used in a variety of lexical contexts, so it might seem that the starting point for determining which equivalents should be included in a bilingual dictionary is the information, and specifically the sense distinctions, provided by a monolingual dictionary. Several bilingual dictionaries covering the language combinations we have considered here are based on the information from monolingual dictionaries, although this is not openly stated. For instance, it seems that the order of equivalents presented in *Simon and Schuster's International Dictionary* corresponds to the order of senses as presented in *Merriam Webster's Collegiate Dictionary of English* (which is historical). However, the sense distinction in a monolingual dictionary, whether it be addressed to native speakers or to advanced-level foreign-language learners, may, in practice, not be right in the context of a bilingual dictionary precisely because both languages are not taken into account from the very beginning. In our opinion, the fact that bilingual dictionaries do not seem to be conceived of as bilingual, contrastive works but instead take as their starting point a monolingual dictionary or, at most, a description of only one of the languages does not always yield optimal results. Since, by and large, bilingual dictionaries are not written from the perspective of both languages, they do not take into account phenomena such as overlapping polysemy, which can only make sense when there are two languages involved. The prime role played by both languages in identifying translation equivalents explains why data from a monolingual corpus will not be relevant to this important issue for bilingual lexicography, because a monolingual corpus is simply not constructed from the standpoint of two languages.

Although we do not believe that a monolingual corpus can resolve problems resulting from overlapping polysemy, we have seen that it can play an important role in two other, equally important, issues facing the bilingual lexicographer, namely choosing the order in which to present equivalents and determining which fixed expressions to include in a specific entry. In a case like that of *odd*, the fact that the word is used much more often in the sense of

'strange' than in the mathematical sense of 'not even' can be used as an argument for the equivalent of the more frequent sense coming first (this argument seems to have been used by *OSD* and *DAC*). We are not arguing explicitly for the order of equivalents always to be based on corpus data, but the *GL* entry that buries the 'strange' sense in the 13th position (out of a total 19) is not, in our opinion, as useful to the reader as it could be. In the case of *cold*, the expression *in the cold light of day/morning/dawn* is relatively frequent in a large corpus of English, yet contrastive analysis shows that it is opaque to Spanish and Catalan speakers; here the corpus data helps us to determine which expressions should warrant translations in the dictionary entry.

We thus conclude by saying that a monolingual corpus does have a role to play in the preparation of a general purpose bilingual dictionary. Bilingual dictionaries that do not take frequency data into account in the organization of their entries, such as *S&S* and *GL*, have been criticized in this article for the way they arrange the information the lexicographers have chosen to include. However, we should like to point out that we ourselves often use these dictionaries because they provide a wealth of information, especially in the form of translation equivalents. The information is there but it is poorly organized, and translators, whose professional obligations require them to search for the right equivalent, are willing to spend the time and effort necessary to plod through the entries. Conversely, a bilingual dictionary with significant gaps in coverage, no matter how well organized the information may be, is going to be found lacking. That is precisely our experience with the *DAC* — a dictionary with well-structured entries, as we have seen in the cases of *cold* and *high*; nevertheless, in our opinion the dictionary contains too few translation equivalents for the level of advanced learners, not to mention that of translators. Those of us who teach know that second-language learners, some of the main users of general purpose bilingual dictionaries, often do not read whole entries; that is why the structure of these reference books needs to be the best possible. We hope to have suggested at least two ways in which use of a monolingual corpus can be fruitful in this task.

Notes

1 Work on this paper was supported by grant PB96-0305 from the Spanish Ministry of Education and Culture, which we hereby acknowledge.

2 We note that the institute we are affiliated with, the Institut Universitari de Lingüística Aplicada of Pompeu Fabra University, is currently building a multilingual corpus for some

languages for specific purposes that includes parallel production in English, Spanish and Catalan.

3 In order to show the range of senses for the three adjectives as clearly as possible, we present the data from the monolingual dictionaries first, followed by that from the bilingual dictionaries, although in carrying out our research we started with the bilingual dictionaries because as teachers of lexicography and translation we knew the entries would prove to be different from one another.

References

- Al-Kasimi, A. M. 1983. "The interlingual/translation dictionary. Dictionaries for translation". In *Lexicography: Principles and Practice*, R. R. K. Hartmann (ed.), 153–162. London: Academic Press.
- American Heritage Dictionary*, 3rd ed. 1992. Boston: Houghton Mifflin.
- Aston, G. and Burnard, L. 1998. *The BNC Handbook*. Edinburgh: Edinburgh University Press.
- Cambridge International Dictionary of English*. 1995. Cambridge: Cambridge University Press.
- Collins Cobuild English Dictionary*, 2nd ed. 1995. London: HarperCollins.
- Diccionari anglès-català*. 1983. Barcelona: Gran Enciclopèdia Catalana.
- Hartmann, R. R. K. and James, G. 1998. *Dictionary of Lexicography*. London and New York: Routledge.
- Larousse Gran Diccionario Español-Inglés/English-Spanish*. 1983. Mexico City: Larousse.
- Merriam Webster's Collegiate Dictionary*, 10th ed. 1996. Springfield, Mass.: Merriam Webster.
- Oxford English Dictionary*, 2nd ed. 1991. Oxford: Oxford University Press.
- The Oxford Spanish Dictionary*. 1994. Oxford: Oxford University Press.
- Simon & Schuster International Dictionary English-Spanish/Spanish-English*. 1971. New York: Simon & Schuster.
- Sinclair, J. 1996. "An International Project in Multilingual Lexicography". *International Journal of Lexicography* 9: 179–196.
- Teubert, W. 1996. "Comparable or Parallel Corpora?". *International Journal of Lexicography* 9: 238–264.
- Tognini Bonelli, E. 1996. "Towards Translation Equivalence from a Corpus Linguistics Perspective". *International Journal of Lexicography* 9: 197–217.
- Tomaszczyk, J. 1983. "On bilingual dictionaries. The case for bilingual dictionaries for foreign language learners". In *Lexicography: Principles and Practice*, R. R. K. Hartmann (ed.), 41–51. London: Academic Press.
- Toury, G. 1995. *Descriptive Translation Studies and Beyond*. Amsterdam and Philadelphia: John Benjamins.