

## Non-equivalence of delexicalised verbs in bilingual dictionaries

### Abstract

This paper deals with the treatment of delexicalised verbs in bilingual dictionaries. In particular, it aims to highlight the problems related to translation equivalence in the light of recent multilingual corpora research and how bilingual dictionaries can improve their description of delexicalised verbs on the basis of corpus data. Importantly, attention is drawn to the fact that bilingual lexicography concerned with two typologically different languages, as opposed to mainstream European practice, also poses challenges to researchers of machine translation. Specific reference is made to English-Turkish bilingual dictionaries and recent corpus research on Turkish.

Keywords: delexicalization, delexical verbs, translation equivalence

### 1. Introduction

Corpus research has shown that the commonest meanings of words are not usually those supplied by introspection and consequently by dictionaries. It has also been demonstrated that

"[t]here is a broad general tendency for frequent words, or frequent senses of words, to have less of a clear, and independent meaning than less frequent words or senses. These meanings of frequent words are difficult to identify and explain; and, with the very frequent words, we are reduced to talking about uses rather than meanings. The tendency can be seen as a progressive *delexicalization*, or reduction of the distinctive contribution made by that word to the meaning" (emphasis added) (Sinclair 1991:113).

These frequent words without independent meaning commonly operate in the construction of normal texts (Sinclair and Renouf 1988), and it is unhelpful to attempt to analyse them grammatically (Sinclair 1991:113). However, *delexicalization* has not been studied extensively as a linguistic phenomenon, and limited literature is available on a small number of delexicalized nouns and prepositions (e.g. Sinclair 1989), intensifiers (Partington 1993), and verbs and adjectives (Sinclair and Renouf 1988, Sinclair et al. 1996). The notion of delexicalization is central to the study of collocation where the delexical word acts in conjunction with other words and shares their meaning (Sinclair and Renouf 1988, Partington 1993). Delexicalization has also been neglected in monolingual and bilingual lexicography (Pérez-Hernández 1996), and delexical structures have not been considered in terms of equivalence in bilingual dictionaries. Previously, in an investigation of Serbo-Croatian and English verb+noun collocations, Benson (1989) also pointed out that the contrast between collocations in different languages was striking but not adequately dealt with in bilingual dictionaries. For the purposes of this paper I will discuss only two of these verbs in English, *do* and *make*, and two such verbs in Turkish, *yapmak* and *etmek*.

## 2. Delexicalized verbs

Typically, *have*, *take*, *give*, *do* and *make* can be listed among the most delexicalized verbs in English (Collins Cobuild English Grammar 1990:147). "The primary function of *make*, for example, is to carry nouns like *decision/s* [...] thereby offering the alternative phraseology 'make your own decisions' to 'decide on something' [...] and so on. Which of the two formulations to choose is obviously a strategic matter in text creation, but the delexical option is firmly there" (Sinclair and Renouf 1988).

Another area of research where this common category of verbs has not found a satisfactory explanation is Language Contact. Studies on various languages in contact show that these verbs operate largely in borrowing although their delexical function as such has not been recognised and they have been varyingly called *auxiliary verbs* (e.g. Backus 1996:236 Turkish, Tamis 1986:169 Greek), *semi-auxiliary verbs* (e.g. Silva-Corvalán 1986 Spanish), *compound verbs* (Kachru 1978 Hindi), *operators* (Romaine 1989:131 Panjabi), and so on.

Recent corpus-based monolingual dictionaries of English (e.g. *Collins Cobuild English Dictionary 1995*) deal with the delexical function of such verbs. For example, the user finds out straight away that "DO is often used instead of a more specific verb" as in *do the washing up*, *do the garden* and so on (p. 486) and MAKE "is used with a wide range of nouns" which refer to an action as in *make a phone call* and so on (p. 1006). However, the bilingual dictionaries examined in this paper have not caught on to this fact yet (see also Pérez-Hernández 1996 on Spanish-English dictionaries).

Not surprisingly, standard grammars and dictionaries of Turkish do not have very comprehensive treatments of *etmek* and *yapmak*, supposedly the equivalents of *do* and *make*, and often refer to *etmek* as the 'compound verb' or 'auxiliary verb'. It is specified that most loan words from Arabic and Persian operate in Turkish as nouns in *noun+etmek* construction, regardless of their grammatical class in the original language; and there is hardly any mention of *yapmak* with reference to its delexical function (or any other verb for that matter). For example, Underhill's grammar (1976:246) touches briefly on this issue:

"Many verbs in Turkish are compounds, formed by a noun indicating an action, followed by the auxiliary verb *etmek*. The noun is usually, although not always, of Arabic, Persian, or European origin [...]. Thus, with the word *telefon* 'telephone' we get the verb *telefon etmek* 'to telephone'; the noun remains invariable, and the verb is conjugated as any other verb".

Similarly, we find such explanations as the following in *The Concise Oxford Turkish-English/English-Turkish Dictionary* (p. 98):

"ETMEK is the verb most commonly used to make a composite verb, chiefly with Arabic nouns, eg *zannetmek*, to think; *sarfetmek*, to spend; when the noun is of two syllables, the noun and the verb are usually written separately, eg *hizmet etmek*, to serve; *telefon etmek*, to telephone".

In these entries, differences between spoken and written texts of Turkish are not dealt with, either. This is largely due to the fact that a systematic study of their frequency counts has been lacking (Kurtböke 1998). The only comparative study of spoken and written Turkish corpora goes back to research (Pierce 1961, 1963) conducted between 1957-60 in Turkey. In Pierce's

corpus of 140,000 words of spoken Turkish, *yapmak* ranked 14th (1,264 occurrences) among the most frequent 20 items, and in his 100,000-word written corpus *etmek* was the 3rd (1,944 occurrences) and *yapmak* was the 10th (650 occurrences).

## 2.1. Translation Equivalence

Translation equivalence has always been an intriguing issue for compilers as well as the user of dictionaries:

"Traditionally, bilingual dictionaries do not include a specific explanation of the reasons why a word is translated in one way rather than another [...] there is usually just a list of words in the target language that constitute possible translation equivalents under certain circumstances, and very little information about when to use which" (Sinclair et al 1996:177).

In fact, the most commonly used English-Turkish/Turkish-English dictionaries do exactly that. For example, in *The Oxford English-Turkish Dictionary*, we have a list of a number of possible equivalents without any reference as to which equivalent is used when and the user is expected to work it out without contextual information:

"DO Yapmak, etmek; kılmak, bitirmek; başarmak; tanzim etm; düzeltmek; (mesafe) kat etmek; bir rolü oynamak; (arg.) aldatmak, kafese koymak; elverişli olm., uygun gelmek, yakışmak. be done - yapılmak; tamamlanmak; (et) kafi pişirilme; bitkin bir hale gelmek; (arg.) aldanmak" (p. 151).

MAKE Yapmak, etmek, kılmak; yaratmak, imal etm.; husule getirmek; teşkil etm.; kazanmak" (p. 327).

*The Oxford Turkish-English Dictionary* likewise, provides the following equivalents for *do* and *make*:

"YAPMAK Do; make; create, give rise to; build; construct; constitute; arrange; repair; apply; set to rights; make ready" (p. 501).

"ETMEK Do; make; cost; be worth; fetch (a price): (+neg.adv.) live, exist, manage: (-i), reach, find: (-den), do without; deprive of: (-e), do to; make; (*child.*) relieve o.s. *As aux v it is used with adjectives and nouns, combining with monosyllables* (BERBAT ETM., ZANNET-MEK)" (p. 169-170).

These entries both in *The Oxford English-Turkish* and *The Oxford Turkish-English* dictionaries include a number of idiomatic uses later on in the same column which will not be considered here. However, the selection of these idiomatic expressions seems to have been based on the lexicographers' intuition rather than on other criteria. This is not surprising as the major bilingual English-Turkish dictionaries rely on their previous editions which were not based on large corpora (Kurtböke 1996).

## 2.2. Contextual information from English and Turkish corpora<sup>1</sup>

The missing information in the entries above is the frequent noun phrases which co-occur with these verbs and carry most of the meaning. Although it is not always predictable whether a user will go to the noun or to the delexical verb in the dictionary to find such structures (see for example Atkins and Varantola 1997), consistency in their treatment is important. The user may not be aware that some verbs may be nominalized and used in a delexical structure with a similar meaning (Sinclair et al 1990:147, see also Kurtböke 1997). For example, in *The Oxford English-Turkish Dictionary* the delexical structure *make+a+decision*, which is the alternative to the verb *decide*, is not included with the noun (*decision*) or the verbs involved (*decide, make*). Similarly, a Turkish user cannot retrieve the *make a decision* option by looking up the Turkish equivalent of *decision* (*karar*) or *decide* (*karar vermek*).

Pairs like *to decide* and *make a decision* in English, however, do not always have an equivalent pair in Turkish. Although there is a superficial similarity, the Turkish pair *kararlaştırmak* and *karar vermek* function differently as the independent verb form (*kararlaştırmak*) in Turkish encapsulates *noun+reciprocal suffix+causative suffix* and acquires a different meaning (*to arrange to...which involves the participation of at least two people in the decision*). In *The Oxford English-Turkish Dictionary*, both *kararlaştırmak* and *karar vermek* are given as the equivalents of *decide*<sup>2</sup> but the distinction in their meaning (due to the suffixation in the independent verb) is not clear.

Similarly, in *The Oxford Turkish-English Dictionary* the equivalent for *karar vermek* is given as *decide*. In the Turkish corpus, on the other hand, we observe that not all occurrences of *karar vermek* translate as *decide* but it also means *come to the conclusion that...* as in 8 below. Interestingly, 16 out of 45 occurrences have a *nominalization+dative suffix* form immediately to the left of *karar +vermek* as in 1-8:

1.	hangisi olduğuna	karar vermeden	önce kullanılan bir
2.	Katılmaya	karar verirken	şahsi ilgilerini takip
3.	kalmayı mı istediğinize	karar vermeniz	gerekecektir Eğer
4.	hakkı olduğuna	karar verilenler	programdan
5.	için hükümetin atmaya	karar verdiği	adımların özeti aşağıda
6.	bir daire kiralamaya	karar vermişti	Son beş yıldır kendisi
7.	yolun hangisi olduğuna	karar verir	Sağlık Hizmetleri Şikayet
8.	sakatlığı olmadığına	karar verdi	ve başvurusunu reddetti

While in English the verb *decide* is simply followed by the infinitive marker *to* as a base for the following verb (e.g. *he decided to leave*), the equivalent Turkish construction uses a nominalized verb followed by a dative case marker. This information is clearly important for the purposes of machine translation although it is hardly dealt with in manual lexicography.

Let us now turn to *delexical do+noun* pattern. Corpus evidence shows that the nouns *damage* (154) and *harm* (162) collocate frequently with *do*. As in the case of *decide* and *make a decision*, there are pairs *to damage* and *do damage*, *to harm* and *do harm*. But with their Turkish equivalent *zarar*, which is an Arabic loan, the choices are limited as it cannot function as a verb on its own. In *The Oxford English-Turkish Dictionary* the equivalent of *to damage* and *to harm* is given as *zarar vermek*. Interestingly, this information helps us discover yet another delexicalised verb in Turkish, *vermek*. This illustrates the fact that

standard dictionary and grammar book information on *etmek* (as an auxiliary which chiefly co-occurs with Arabic nouns) should be reconsidered; apparently there are others which have the same function. In the Oxford Turkish-English bilingual dictionary *give* is the equivalent of *vermek*, which, in turn, is another delexicalised verb in English. This, then, points to the fact that their collocates do not run parallel in English and Turkish and further investigation is necessary to establish the equivalents.

Corpus evidence shows that *zarar+vermek* pattern is often preceded by the dative suffix in Turkish (10-13). This is again a significant construction which is not dealt with in manual lexicography but of interest for machine translation processes:

9.	ve psikolojik yönden	zarar verir	Eğer siz de bu tür şiddet
10.	İnsanların kendilerine	zarar vermek	istedikleri inancında
11.	ya da başkalarına	zarar verme	riski olması durumunda
12.	gazetesi topluma	zarar vermeye	yönelik yayınlarından
13.	ters düşecek onlara	zarar verecek	sözler söylemek
14.	vatanına yarar ya da	zarar verebilecek	bir konumda olması

Again, in the Turkish-English volume the user cannot find the delexical construction *do damage/harm* by looking up the noun *zarar* or the verb *vermek*.

Let us now examine *yapmak* and *etmek*, the Turkish equivalents of *do* and *make*. Concordance lines from the Turkish corpus of *yapmak* show that a considerable number of the preceding nouns end in the suffix *-lik/lık/lük/luk*<sup>2</sup> and acquire the meaning of *work as a...*, *play the role of...* as in the examples 15-30.<sup>4</sup> Such strong *suffix+verb* co-selection supports the view that traditional categories of grammar should be treated with caution when the machine translation process involves at least one agglutinative language, as the suffix carries or shares the semantic load with the noun:

15.	bu işte önderlik	yaptı	
16.	olan ekibe önderlik	yapmak	kabiliyetine ve Avustralyada
17.	takdimcilik	yapacak	kabiliyetiniz var mı
18.	önünde takdimcilik	yapma	tecrübesine sahip olmanız veya
19.	konularda danışmanlık	yapmakla	sorumludur tazminat
20.	konusunda danışmanlık	yapmaktır	
21.	konularında danışmanlık	yapmaktan	sorumlu olacaktır nitelikler
22.	yıldan beri Video'culuk	yapmaktadır	Mesleğine dört elle
23.	çocuklara özel hocalık	yapmak	veli committesine veya okul
24.	ve büyükbabalık	yaparak	ve dinlenince eğitim ya da
25.	Avustralya'da eczacılık	yapabilmek	için eyalet veya bölgedeki
26.	kaydolmadan eczacılık	yapmak	Yasalarca suç olarak kabul
27.	buyana yıl Başkanlık	yapan	ve arası hükümsüz iktidar olan
28.	meclisinde Başkanlık	yapmıştır	Kraliçenin doğum günü
29.	incelemeye başkanlık	yapmaktadır	Bu konuda fikir bildirmek
30.	bize tercümanlık	yapıyor	yardım ediyor diye' de onları

The frequent co-occurrence of this suffix with *yapmak* shows that this tendency is significant and should be included in the treatment of the delexical structure of *yapmak*. This type of information does not appear in the dictionaries consulted here.

The last verb under examination is *etmek*. A Turkish noun which is not of Arabic origin and collocates frequently with *etmek* is *göç* (migration), meaning *migrate*. According to the dictionary, *göç* is also used as an independent verb (*göçmek*) with the same meaning although it has no instances in the Turkish corpus<sup>5</sup> and that the delexical form is much more common:

32.	adına Avustralya'ya	göç edecek	olanlara mesleksel
33.	Avustralya'ya	göç edebilme	bu ülkede çalışma ya da
34.	kanunsuz olarak	göç edenlere	kanuni müsaadeleri almaları
35.	ailelerinden	göç edenlerin	beraberce denenmesi ve
36.	Avustralya'ya	göç ederek	gelmiş ve yıldan beri
37.	Amerika'ya	göç eden	göçmenleri çok geçmeden
38.	gibi Avustralya'ya	göç eden	Türklerin resmi yoldan ilk

An English-speaking user with limited knowledge of Turkish can easily be misled by such information as it is not clear whether *göçmek* and *göç etmek* are interchangeable and equally common or not in Turkish. In fact, the corpus evidence shows that *göçmek* is not common at all.

The last point I want to make is that there are delexical structures which are restricted with respect to the form of the following noun. That is, the noun is mainly used in the plural form as for example in *do+repairs*. Interestingly, the Turkish equivalent of the singular form *tamir* (repair) is used with *etmek* but the plural form *tamirat* (repairs) attracts *yapmak* and there is also a difference in meaning. This is again important information about usage which is not available to the user in the dictionary.

### 3. Conclusion

The most commonly used bilingual dictionaries today, such as *The Oxford English-Turkish/Turkish-English Dictionary*, still rely on the previous editions (Kurtböke 1996) which were certainly not based on large corpora. However, recent developments in the areas of multilingual lexicography and translation show that contextual information is essential in the identification of translation equivalents, which can only be accessed through large corpora. One of the intriguing areas identified as a result of ongoing research projects on translation equivalence (Sinclair et al 1996) is *delexicalization*. Initial experiments on a limited number of language pairs (e.g. Spanish-English) lead to the observation that delexical verbs, which do not perform the action but share the semantic load with the following noun phrase, are common across languages. Consequently, a detailed study of their co-occurrence tendencies in large corpora is essential for the new generation of bilingual dictionaries and machine translation. In this paper, the equivalence of English and Turkish delexical verbs is explored as an area awaiting further investigation in multilingual lexicography and translation projects.<sup>6</sup>

#### 4. Notes

- <sup>1</sup> The corpora used are the Bank of English, Cobuild, Birmingham and the *Ozturk Corpus*, a collection of 1000 newspaper texts, compiled as part of a large-scale study of Turkish-English language contact in Australia.
- <sup>2</sup> Due to the differences in the word order, Turkish constructions appear in the reverse order. That is *noun+verb* as opposed to *verb+noun* construction in English. Naturally in passive constructions the word order will be the opposite.
- <sup>3</sup> The choice of the vowel (i, ı, u or ü) will be constrained by the preceding syllable, in accordance with the vowel harmony rule.
- <sup>4</sup> The other *noun+lik* or *adjective+lik* forms found in the corpus in the company of *yapmak* are: *değişiklik, göçmenlik, temizlik, hazırlık, kolaylık, yaramazlık, fenalık, pazarlık, tanıklık, fedakarlık, ayırimcılık, yanlışlık, azgınlık, ikiyüzlülük, yolculuk, yolsuzluk*.
- <sup>5</sup> This is in spite of the fact that *Ozturk Corpus* has been built in the context of immigration (Turkish in Australia), and *göç* and its various forms rank among the most frequent words.
- <sup>6</sup> I would like to thank Jeremy Clear, Liz Potter at Cobuild and Phillip King at EISU, University of Birmingham for their comments and suggestions.

#### 5. References

- Alderson A. D. and Fİz 1978 *The Oxford English-Turkish Dictionary*. 2nd edition.
- Alderson A. D. and Fİz 1959 *The Concise Oxford Turkish-English/English-Turkish Dictionary*.
- Alderson A. D. and Fİz 1984 *The Oxford Turkish-English Dictionary*. 3rd edition.
- Atkins, B.T.S. and K. Varantola 1997 Monitoring Dictionary Use. *International Journal of Lexicography*. 10:1. pp. 1-38.
- Backus, A. 1996 *Two in one: Bilingual Speech of Turkish Immigrants in The Netherlands*. Ph.D. Thesis. Tilburg University Press.
- Benson, M. 1989 The Collocational Dictionary and the Advanced Learner. In Tikoo, M.L. (ed) *Learners Dictionaries: The State of The Art*. SEAMEO Regional Language Centre. pp. 84-92.
- Collins Cobuild English Language Dictionary* 1995. Harper Collins Publishers.
- Kachru, B.B. 1978 Toward Structuring Code-Mixing: An Indian Perspective. *International Journal of the Sociology of Language*. 16. pp. 27-47.
- Kurtböke, P. 1996 The Impact of Corpus Planning on Bilingual Dictionaries. In Gellerstam, M. et al (eds) *Euralex 96 Proceedings*. pp. 591-596.
- Kurtböke, P. 1997 Towards a Corpus of Australian Turkish. In *The University of Melbourne Working Papers in Linguistics and Applied Linguistics* 16. pp. 151-158
- Kurtböke, P. 1998 Delexicalised verbs in Turkish from a Corpus perspective. Paper to be presented at *9th Intl. Conference on Turkish Linguistics*. The University of Oxford. 12-14 August.
- Kurtböke, P. forthcoming. Delivery yapılır: Single occurrence or co-selection. In *Proceedings of 1st Intl. Symposium on Bilingualism*. University of Vigo, Spain.
- Partington, A. 1993 Corpus Evidence of Language Change - The Case of the Intensifier. In Baker, M. et al (eds) *Text and Technology. In Honour of John Sinclair*. John Benjamins, Amsterdam. pp. 177-192.

- Pérez-Hernández, C. 1996 A Pilot Study on Translation Equivalence between English and Spanish. In Sinclair et al. pp. 218-237.
- Pierce, J. 1961 A Frequency Count of Turkish Affixes. *Anthropological Linguistics*. 3:9. pp. 31-42.
- Pierce, J. 1963 A Statistical Study of Grammar and Lexicon in Turkish and Sahaptin. *International Journal of American Linguistics*. 29 pp. 96-103.
- Romaine, S. 1989 *Bilingualism*. Blackwell, Oxford. 2nd ed.
- Silva-Corvalán, C. 1986 Bilingualism and Language Change: The extension of ESTAR in Los Angeles Spanish. *Language* 62. pp. 587-608.
- Sinclair, J. 1989 Uncommonly Common Words. In Tikoo, M.L. (ed) *Learners Dictionaries: The State of The Art*. SEAMEO Regional Language Centre. pp. 135-152.
- Sinclair, J. 1991 *Corpus Concordance Collocation*. OUP.
- Sinclair, J. and A. Renouf 1988 A lexical syllabus for language learning. In R. Carter and M. McCarthy (eds) *Vocabulary and Language Teaching*. Longman. pp. 140-160.
- Sinclair, J. et al 1990 *Collins Cobuild English Grammar*. Collins.
- Sinclair, J., Payne, J. and C. Pérez-Hernández (eds.) 1996 Corpus to Corpus: A Study of Translation Equivalence. *International Journal of Lexicography*. Special Issue 9:3.
- Tamis, A. 1986 *The State of Modern Greek Language as Spoken in Victoria*. Ph.D. Thesis. The University of Melbourne.
- Underhill, R. 1976 *Turkish Grammar*. The MIT Press, Cambridge, Mass.