Word-formation in English-French bilingual dictionaries: the contribution of bilingual corpora

Marie-Aude Lefer*

Centre for English Corpus Linguistics – Université catholique de Louvain

Research on the representation of word-formation in dictionaries is scarce and tends to be restricted to learners' dictionaries and monolingual dictionaries intended for native speakers. Nor is the issue of word-formation in bilingual dictionaries often discussed in lexicographic studies. This study, intended as a step on the way to rectifying the situation, reports the results of a comparison of the strategies adopted in four influential English-French dictionaries, focusing more particularly on derivational prefixes. The study shows that prefixes and word-initial elements in general receive very scant treatment in English-French dictionaries, which seems hardly justifiable when one thinks of the major role they play in the interpretation and translation of complex words. In my presentation I will highlight and illustrate a number of shortcomings, such as the lack of consistent criteria for the selection of affix entries and the misrepresentation of affix polysemy. More importantly, the presentation will also show how bilingual dictionary-making could benefit from bilingual corpora (both comparable and translation corpora) to improve the description of word-formation. I will propose a corpus-based list of the most productive and frequent prefixes in English and French. This list would seem to be a promising starting point for selecting more systematically and more rigorously the affixes to be included as headwords in bilingual dictionaries. To illustrate the usefulness of corpus data, I will also present a model bilingual entry for the French prefix dé– based chiefly on data extracted from an English-French translation corpus.

1. Introduction

Research on the representation of word-formation in dictionaries is scarce and tends to be restricted to learners' dictionaries and monolingual dictionaries intended for native speakers. Nor is the issue of word-formation in bilingual dictionaries often discussed in lexicographic studies. This study, intended as a step on the way to rectifying the situation, reports the results of a comparison of the strategies adopted in bilingual dictionaries, focusing more particularly on derivational prefixes, and shows how bilingual dictionary-making could benefit from bilingual corpora (both comparable and translation corpora) to improve the description of word-formation. Four influential English-French bilingual dictionaries are reviewed: Robert & Collins (RC), Hachette-Oxford (HO), Larousse-Chambers (LCh) and Harrap's Unabridged (HU). My opening assumption is that it is essential to list the most productive and frequent derivational affixes as headwords because dictionaries cannot afford to list derived neologisms exhaustively or give separate translations for each derived form.

The paper is structured as follows. The place of word-formation in lexicographic studies is broadly assessed in Section 1. Section 2 then deals with the role of bilingual corpora in bilingual lexicography. Section 3 is devoted to the place of word-initial elements in English-French bilingual dictionaries. In Section 4, I draw up a short, corpus-based list of prefixes which might be included as headwords and in Section 5, I give an example of what a model corpus-based bilingual entry might look like. Section 6 contains some concluding remarks.

2. Word-formation and lexicography

The primary focus of research on the representation of word-formation in dictionaries is on monolingual dictionaries intended for native speakers (Jackson, 2000; de Caluwe & Taeldeman, 2003; Dardano *et al*., 2006) and learners' dictionaries (Jackson, 2000; Prćić, 1999, 2005, 2008; ten Hacken *et al*., 2006) and within this research, issues related to word-formation are shown to be largely neglected.

As regards learners' dictionaries, for example, ten Hacken *et al.* (2006: 254) point out that 'the adequate representation of word formation in a learners' dictionary is an important asset in the acquisition of vocabulary' for three major reasons: the decoding of words the learner does not know, the production of new words when the learner has not yet acquired the standard word, and the creation of a tight network structure in the mental lexicon, which greatly facilitates vocabulary acquisition (*ibid.*, 243). However, research has shown that derivational affixes are not systematically included in learners' dictionaries and that information on productivity and usage is often lacking (see Prćić 1999, 2005, 2008). To address these issues, Prćić (1999) has argued that affix entries in learners' dictionaries should contain the following information: spelling (spelling variants, allomorphy of the base, use of hyphens, etc.), pronunciation, input and output units (i.e. category of the base and category of the derivative), sense distinctions (for polysemous or homographic affixes), definitions, cross-references (synonymy, antonymy), usage (e.g. register), productivity, examples and consistent terminology (use of clear labels, such as prefix, suffix, etc.).

As regards bilingual dictionaries, studies are scarce. A notable exception is Cartoni (2008), which has demonstrated that productive derivational affixes are not regularly included in Italian-French and Italian-English dictionaries. Cartoni also deplores the lack of information on morphological productivity and the semantics of prefixes.

3. Bilingual corpora and bilingual lexicography

As Williams (2008: 261) points out, corpora are not yet standard practice in bilingual lexicography. However, Teubert (1996: 240) rightly argues that an unbiased analysis of corpora yields considerably more insight than an evaluation restricted to citations selected by lexicographers. Teubert goes on to say that 'by exploiting corpora, bilingual and multilingual lexicography can reach a new quality level, a level that was just not possible without corpora' (*ibid.*).

Comparable corpora, i.e. corpora of original texts in two or more languages matched by criteria such as register, genre, domain, time of publication, and size, are essential resources for corpus-based bilingual lexicography. They can be used to establish frequency-based lists of potential headwords, thereby improving headword selection, to investigate polysemy, identify the range of semantic contexts in which the equivalents listed by the bilingual dictionary can be used, choose the order in which to present equivalents, determine which fixed expressions, compounds and collocations to include in entries, etc. (see e.g. Roberts, 1996; Roberts & Cormier, 1999; Alsina & DeCesaris, 2002; Sharoff, 2004; Szende, 2004; Ferraresi *et al.*, 2008). However, comparable corpora are not enough. Lexicographers also need to have access to translation corpora to examine possible translation equivalents on the basis of authentic data rather than their own bilingual competence.

Translation corpora, i.e. corpora consisting of original texts and their translations into one or several languages, can also be used to enrich bilingual dictionaries in that it is possible, on the basis of translation data, (1) to determine which translations are frequent and which are unusual, thereby empirically identifying translation equivalents, and (2) to validate or invalidate lexicographers' intuitions (see e.g. Salkie, 2002 & 2008; Corréard, 2005; Krishnamurthy, 2005). Translation corpora are 'repositories of translation units and their equivalents in the target language' (Teubert, 2002: 193). As such, they provide authentic examples of the translation strategies adopted by a wide range of (supposedly) highly competent bilinguals. However, dictionary compilers seem to shy away from using translation

corpora for the compilation or revision of bilingual dictionaries. A possible reason for this is that 'few commercial dictionary-writing schedules allow time for the lexicographers to browse through data from parallel [i.e. translation] corpora' (Atkins, 2002). Some lexicographers therefore claim that unless tailor-made multilingual tools are devised specifically for lexicography, translation corpora cannot contribute to commercial bilingual dictionary-making (Atkins, 2002). These technical issues, which are directly linked to the time constraints of commercial lexicography, will have to be solved if progress is to be made. In addition, it is worth noting that most translation corpora are unfortunately too small for studying low-frequency words or patterns and tend to be confined to a limited number of genres and text types (e.g. EU documents, fiction, journalese). Finally, it should be borne in mind that translations often contain traces of source text interference (translationese) and translation norms. Translation data should therefore be used with caution.

In view of the respective strengths of comparable and translation corpora, some scholars have convincingly argued that the two types of corpora should be used in tandem in bilingual lexicography (e.g. Teubert, 1996; Roberts, 1996; Roberts & Cormier, 1999; Lefer, 2009). This is also the line taken in this study.

4. The place of prefixes in English-French bilingual dictionaries: overview

4.1. Initial word-parts listed as headwords
A search on the CD-ROM versions of the major English-French bilingual dictionaries[1] has revealed that there is wide variation across dictionaries as to which word-parts are listed as headwords (see Table 1). It appears from Table 1 that the four dictionaries reviewed in this paper contain very different repertoires of word-parts (between 20 and 46 in French and between 26 and 106 in English). Moreover, there are large discrepancies within dictionaries as regards the two directional halves. The RC, HO and HU list twice as many English word-parts as French elements. As demonstrated in Lefer (2009), there is no empirical justification for this unbalance, as English and French have very similar stocks of derivational affixes at their disposal (c. 100 prefixes and c. 150 suffixes in each language).

---

[1] The search was carried out with the help of the *Search Full Text* option (search words *préfixe/prefix*, *composé/compound*) in the *Hachette-Oxford*, *Larousse-Chambers* and *Harrap's Unabridged* dictionaries. It was carried out manually (using the scroll-down menu of entries) in the *Robert & Collins*.

Table 1. Word-parts listed as headwords in English-French bilingual dictionaries

French to English			
	Robert & Collins	**Hachette-Oxford**	**Larousse-Chambers**	**Harrap's Unabridged Pro**
Word parts listed as prefixes	*anglo-, anti-, archi-, auto-, bi-, contre-, cyber-, déca-, demi-, e-, équi-, ex-, franco-, gréco-, hyper-, hypo-, inter-, judéo-, kilo-, macro-, méga-, mi-, micro-, milli-, mini-, mono-, multi-, nano-, narco-, néo-, nitro-, poly-, post-, pré-, pro-, proto-, quasi-, semi-, sous-, super-, supra-, sur-, trans-, tri-, ultra-, vice-* (46)	*afro-, bio-, ex-, kilo-, maxi-, méga-, mi-, mini-, néo-, pan-, photo-, politico-, post-, pré-, pro-, pseudo-, super-* (17)	*anglo-, centi-, déca-, déci-, ex-, franco-, hyper-, inter-, méga-, mi-, mini-, mono-, nano-, néo-, neuro-, non-, paléo-, para-, pico-, proto-, pseudo-, quasi-, semi-, simili-, super-, vice-* (26)	*angio-, anglo-, anti-, archi-, auto-, bi-, bio-, cyber-, déca-, éco-, électro-, euro-, ex-, extra-, ferro-, franco-, géo-, giga, gréco-, hyper-, inter-, macro-, méga-, mi-, micro-, mini-, mono-, nano-, néo-, neuro-, non-, paléo-, para-, photo-, pico-, poly-, proto-, pseudo-, psycho-, quasi-, rétro-, semi-, simili-, super-, vice-* (45)
Word parts listed as parts of compounds		*demi-, hecto-, quasi-* (3)		
Total	46	20	26	45

English to French			
	Robert & Collins	**Hachette-Oxford**	**Larousse-Chambers**	**Harrap's Unabridged Pro**
Word parts listed as prefixes	*aéro-, Anglo-, ante-, anthropo-, anti-, arch-, astro-, audio-, Austro-, auto-, bi-, biblio-, bio-, cardio-, circum-, co-, contra-, counter-, cyber-, de-, demi-, di-, e-, eco-, electro-, equi-, ex-, extra-, ferro-, Franco-, gastro-, geo-, Greco-, gyro-, hemat-, hemato-, hemo-, hetero-, Hispano-, hydro-, hyper-, hypo-, Indo-, inter-, intra-, intro-, iso-, macro-, magneto-, matri-, mega-, meta-, micro-, mid-, milli-, mini-, mono-, multi-, nano-, narco-, neo-, neuro-, nitro-, non-, oleo-, omni-, ortho-, osteo-, over-, paleo-, pan-, para-, petro-, phono-, photo-, physio-, poly-, post-, pre-, pro-, proto-, pseudo-, psycho-, pyro-, quadri-, quasi-, re-, retro-, Russo-, semi-, sino-, stereo-, sub-, super-, supra-, techno-, tele-, thermo-, trans-, tri-, tribo-, ultra-, un-, under-, uni-, vice-*	*over-, under-* (2)	*eco-* (1)	*all-, Anglo-, anti-, auto-, bi-, cardio-, co-, contra-, crypto-, eco-, electro-, Euro-, ex-, extra-, ferro-, Franco-, geo-, giga-, Graeco-, haema-, haemato-, haemo-, hand-, he-, hemi-, hetero-, homeo-, homo-, hyper-, hypo-, Indo-, inter-, intro-, iso-, Italo-, leuco-, litho-, little-, macro-, mega-, meta-, micro-, milli-, mini-, mono-, nano-, near-, necro-, neo-, new-, nitro-, non-, oft-, omni-, out-, over-, paleo-, pan-, para-, phato-, photo-, physio-, politico-, poly-, pre-, pro-, proto-, pseudo-, pyro-, quadri-, quasi-, quick-, retro-, Russo-, self-, semi-, she-, Sino-, stomato-, sub-, Turco-, über-, ultra-, under-, vice-, worst-* (86)
Word parts listed as parts of compounds		*all-, Anglo-, anti-, arch-, counter-, crypto-, demi-, dial-, e-, eco-, Euro-, ever-, ex-, Franco-, Graeco-, hyper-, macro-, medico-, mega-, micro-, mid-, mini-, mono-, much-, multi-, nano-, near-, neo-, non-, oft-, pan-, petro-, poly-, post-, pro-, pseudo-, quasi-, semi-, super-, tele-, ultra-* (41)	*all-, Anglo-, anti-, bi-, co-, crypto-, Euro-, ex-, extra-, Franco-, Graeco-, near-, new-, non-, oft-, over-, paleo-, pro-, pseudo-, self-, semi-, sub-, ultra-, under-, vice-* (25)	
Total	106	43	26	86

4.2. Word-part labelling

In addition, as can be seen from Table 1, the meta-language used in the dictionaries reviewed is inconsistent. HO and LCh, for instance, list many English word-parts under the heading 'part of compound' while they list the majority of French word-parts under the label 'prefix'. This is not to say that dictionaries should necessarily use the linguistically appropriate terms to refer to word-parts, but they should at least be consistent in the two directional sections as to which terms they use to label these word-forming elements.

Despite the fuzzy boundaries between derivational affixes and compound parts, it is nevertheless possible to suggest a number of defining criteria for affix status in English and French, especially when one is ready to adopt a prototypical view of 'affixhood', i.e. an approach where some affixes are seen as more prototypical of the category 'affix' than others. The following defining criteria could be used by lexicographers as a touchstone of affixhood, bearing in mind that some less prototypical affixes may turn out to fulfil only three or four of the five criteria outlined below:

(a) *Position*: prototypical derivational affixes are either word-initial (prefixes) or word-final (suffixes) but, unlike combining forms (elements of Latin or Greek origin, such as –*hydr*–), they cannot occur in both positions. The only exceptions to this criterion are English particles such as *out*–, *over*–, etc., which are often included in inventories of derivational affixes despite their ambiguous status as word-initial elements in derivatives and word-final elements in phrasal and prepositional verbs (note, however, that they meet the other defining criteria listed below);

(b) *Syntactic autonomy*: prototypical derivational affixes are bound lexical morphemes. However, a small number of free grammatical morphemes (and to a lesser extent, free lexical morphemes) can develop affix status (e.g. –*free*);

(c) *Meaning*: as a rule, derivational affixes convey a stable semantic content. In addition, they tend to have a weaker referential meaning than free forms (compare, for example, *re*– 'again' and *biblio*– 'book'). Derivational affixes having a corresponding free form (e.g. *under*–) can be considered as real derivational affixes on condition that (i) they convey meanings other than the ones found with the corresponding independent word or (ii) they convey a more restricted range of meanings than the corresponding free forms. In addition, initial combining forms that (i) show a tendency to play a subordinate role with respect to the root; (ii) are more general or abstract in meaning; and (iii) express cognitively relevant, recurring components of meaning which are organised in paradigmatic relations with the meanings expressed by other prefixes, can be considered as prefixes (see Iacobini, 1998) (these combining forms can be treated as prefixes on condition that they also fulfil criterion (d), e.g. *pseudo*–). Combining forms which convey a clear and concrete referential meaning such as *aero*– 'air' or *hydro*– 'water' should be excluded from the category of affix;

(d) *Nature of the base to which the affix attaches*: derivational affixes typically attach to free bases (derivatives with a bound root are all lexicalised derivatives, e.g. *conceive, receive, deceive*). Elements of Greek or Latin origin should be classified as affixes in so far as they combine with free bases (and meet criterion (c), e.g. *pseudo*– in *pseudo-scientific*). Elements of Greek or Latin origin attached to combining forms should be considered as combining forms forming neoclassical compounds and not as affixes (e.g. *hydrogen*, *aerodrome*);

(e) *Frequency*: derivational affixes tend to be used repetitively in the formation of words (e.g. *e*– as in *e-book*, *e-ticket*, *e-money*, etc.).

These five criteria make it possible to draw a sharper line between derivational affixes, compound parts and combining forms and could thereby constitute an interesting tool for lexicographers in that they could help them systematise word-part labelling in dictionaries.

4.3. Content of prefix entries

Closer examination of the content of the prefix entries reveals a number of weaknesses, which are discussed and illustrated below.

First, many entries simply consist of the prefix in the source language and one formally congruent prefix in the target language, without any example or further comment. This is illustrated in Figure 1. On the other hand, a number of entries do not list the formally congruent prefix in the target language (see Figure 2, where only *former* is suggested as a possible equivalent of French *ex–*).

Figure 1. French *pré–* in RC

Figure 2. English *ex–* in HO

Second, the polysemy of a number of prefixes is not properly accounted for. For example, two meanings of *under–* are included in Figure 3 ('below' and 'junior'). However, the evaluative meaning 'not enough' (e.g. *undercook*), which is extremely productive in English, is not listed. In Figure 4, the prefixes *dé–*, *dis–*, *in–* and *mal–* are listed alphabetically as possible equivalents of English *un–*, without any indication as to the two meanings of the English prefix (negative and reversative), which call for different equivalents in French.
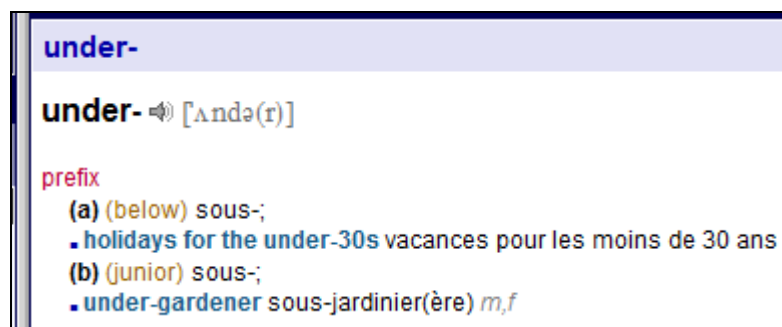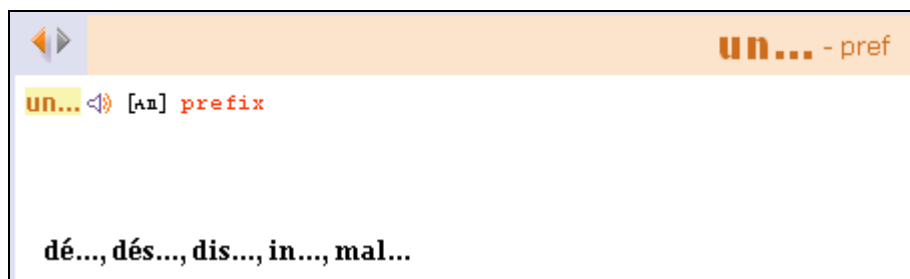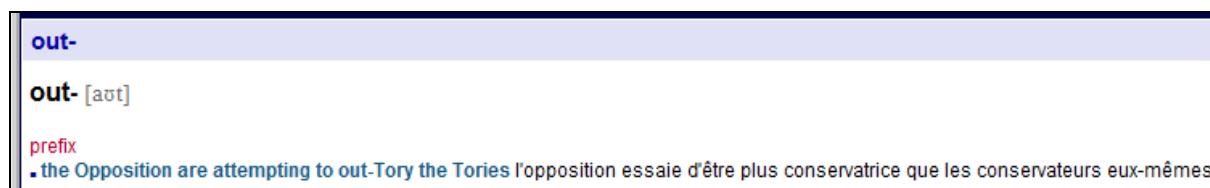

Figure 3. English *under–* in HU

Figure 4. English *un–* in RC

It is also important to note that in some cases, no equivalent is given and the entry only consists of a translated example, as illustrated in Figure 5.

Figure 5. English *out–* in HU

In addition, we can highlight the following shortcomings of prefix entries: there is usually no information about productivity or frequency (an extremely interesting, albeit ambitious, undertaking) and there is barely any information on usage (e.g. register, genre and domain). Even if the domain (e.g. medicine) is specified in some dictionaries, the labels are not systematically attributed.

Interestingly, while some entries suggest a formally congruent equivalent together with a possible non-morphological equivalent or paraphrase, which can greatly help users translate a given derivative into the target language (see Figures 6 and 7), this type of valuable information is regrettably absent in most entries.
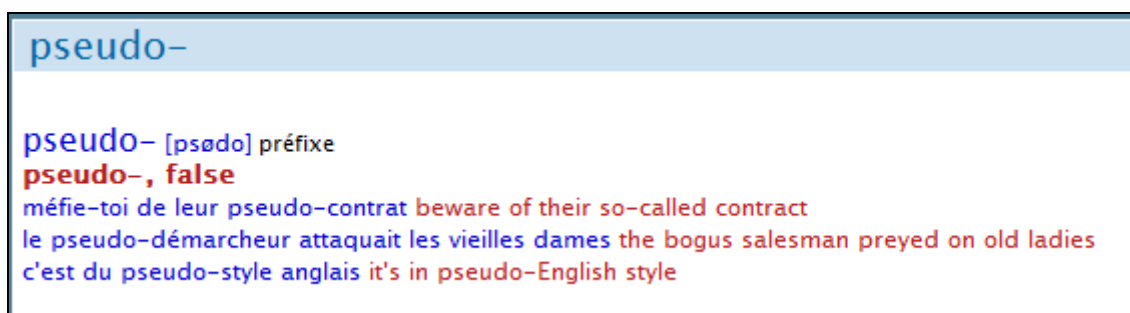
Figure 6. French *pseudo–* in LCh
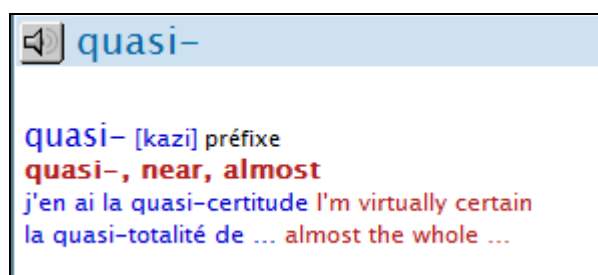
Figure 7. French *quasi–* in Lch

5. Corpus-based short list of productive and frequent prefixes

As appears clearly from Section 2, no systematic criteria seem to govern the inclusion of word-parts as headwords in English-French bilingual dictionaries. The choices made by lexicographers stand out as quite arbitrary and clearly lack empirical grounding. In this section I wish to argue that it is possible to systematise this selection by relying on data extracted from comparable corpora.

To illustrate this I have compiled a corpus-based short list of the top 25 most productive and frequent prefixes in English and French and compared this short list with the prefixes listed as headwords in the four dictionaries examined (see Table 2). The short list presented in Table 2 is based on a thorough analysis of data extracted from a comparable corpus of written English and French containing c. 3 million words per language. The corpus is made up of three components of c. 1 million words, each representing a written genre:

- novels from the Frantext corpus and the British National corpus;
- newspaper leading articles (also called editorials) from the Louvain Multilingual Corpus of Editorials[2];
- research articles in medicine, economics and linguistics from the KIAP corpus (see Fløttum *et al*., 2006)[3].

About 100 prefixes have been investigated in each language.

The prefixes included in the short list are all on the list of the 25 most productive prefixes (in terms of type frequency or 'realised productivity'; see Baayen, 2008) and/or the 25 most frequent prefixes in English and French writing, irrespective of the genre in which they occur.

---

[2] See http://cecl.fltr.ucl.ac.be/MULTED.html [access date: 13 Feb. 2010].

[3] I acknowledge the kind help of Prof. Kjersti Fløttum (University of Bergen), who granted me access to the KIAP corpus.

French					English				
prefix	*RC*	*HO*	*LCh*	*HU*	*prefix*	*RC*	*HO*	*LCh*	*HU*
a-					a-				
anti-	√			√	after-				
après-					anti-	√	√	√	√
auto-	√			√	back-				
bi-	√			√	bi-	√		√	√
bien-					co-	√		√	√
co-					counter-	√	√		
contre-	√				cross-				
dé-					de-	√			
demi-	√	√			dis-				
dis-					en-				
é-					ex-	√	√	√	√
en-					in-				
ex-	√	√	√	√	inter-	√			√
extra-				√	intra-	√			
hyper-	√		√	√	mid-	√	√		
in-					mis-				
inter-	√		√	√	multi-	√	√		
intra-					non-	√	√	√	√
mal-					over-	√	√	√	√
més-					out-				√
mi-	√	√	√	√	post-	√	√		
micro-	√			√	pre-	√			√
multi-	√				pro-	√	√	√	√
néo-	√	√	√	√	re-	√			
non-			√	√	self-			√	√
par-					semi-	√	√	√	√
pour-					sub-	√		√	√
post-	√	√			super-	√	√		
pré-	√	√			tele-	√	√		
pseudo-		√	√	√	un-	√			
quasi-	√	√	√	√	under-	√	√	√	√
re-					TOTAL	22	13	11	14
semi-	√		√	√					
sous-	√								
sur-	√								
télé-									
trans-	√								
ultra-	√								
TOTAL	20	8	9	14					

Table 2. Corpus-based short list of prefixes to be considered for inclusion in bilingual dictionaries

Surprisingly, French *dé–*, *in–* and *re–*, which are the top three most productive and frequent prefixes in writing, are not mentioned in any of the dictionaries reviewed. In the same vein, some English productive and/or frequent prefixes are not listed as headwords, including a

number of Germanic prefixes, which may be difficult to translate into French (e.g. *after–*, *back–*, *cross–*, *mis–*) and therefore deserve particular attention. In addition, English *un–* and *out–* are included in one dictionary only. Even though the short list of productive and frequent prefixes provided in Table 2 is only preliminary (it is based on a relatively small corpus), it constitutes a promising starting point for compiling definitive lists of affixes that should be incorporated as main entries. It clearly demonstrates that comparable corpora can make a significant contribution to bilingual lexicography, as they make it possible to establish empirically-based lists of the most fundamental derivational affixes.

It should also be stressed that the use of a genre- and domain-stratified comparable corpus like the one mentioned above makes it possible to identify genre-(or domain-)sensitive affixes, and hence to attach empirically-based genre and domain labels to bilingual entries (see Table 3 for a sample).[4]

	English	French
Genre-sensitive prefixes		
Press editorials	*anti–, counter–, pro–*	*anti–, ex–, mini–, neo–, pro–, ultra–*
Research articles	*co–, inter–, micro–, multi–, post–, pre–, sub–*	*auto–, co–, multi–, pré–, pseudo–, sous–*
Domain-sensitive prefixes		
Medicine	*anti–, auto–, de–, hyper–, intra–*	*a–, anti–, dys–, extra–, hyper–, hypo–, intra–, micro–, péri–, poly–, post–, trans–*
Economics	*cross–*	*inter–*
Linguistics	*non–, self–*	*non–*

Table 3. Sample of English and French genre- and domain-sensitive prefixes (adapted from Lefer, 2009)

Section 4 has shown that corpus-based analyses yield results that can be extremely helpful when it comes to complementing or revising existing bilingual entries. More specifically, comparable corpus analyses can be used to identify lists of derivational affixes to include as main entries. A number of criteria can be taken into account, such as productivity, frequency, generality/specificity (i.e. are the prefixes equally productive or frequent in different registers, genres and domains or are they rather genre- or domain-promoted?). Comparable corpus analysis based on a semantic classification of prefixes can also be used to determine which prefixes are available in each language to convey given meanings. For example, the comparable study carried out in Lefer (2009) has shown that reflexivity can be conveyed by French *auto–* and by English *auto–* and *self–*. It thus emerges that Eng. *auto–* and *self–* ought to be included as equivalents of Fr. *auto–* (only *self–* is provided in *Robert & Collins* and *Harrap's Unabridged*, with the other dictionaries including no main entry for *auto–*).

6. Corpus-based bilingual entry: the case of French *dé–*

Translation corpora are a very useful complementary resource, offering a wealth of authentic translation options which, if recurrent, can be included in bilingual dictionaries. For example, the comparable corpus analysis carried out in Lefer (2009) has shown that a number of English compounding patterns make up for the lack of productive prefixation expressing numbered quantity, approximation and attenuation in English (e.g. *one–, two–, all–, near–*,

---

[4] The genre- and domain-sensitive prefixes listed in Table 3 display a statistically significantly higher productivity in one genre or one domain than in the others according to the chi-square test ($p < 0.05$).

half–). Examination of translation data would undoubtedly make it possible to identify recurrent and systematic correspondences between these compounds (and other patterns) and French *mono–, uni–, bi–, omni–, quasi–,* and *demi–,* thereby confirming (or contradicting) our intuitions. In addition, translation corpora can be used in a more exploratory way to uncover systematic translation patterns in cases where the comparable corpus analysis identified cross-linguistic productivity differences. These differences in productivity often lead to the use of indirect (non-morphological) translation strategies, which are difficult to detect without the help of translation data but which should nevertheless be included in bilingual dictionaries as they could help users cope with some translation difficulties.

Prefix entries should ideally be revised on the basis of solid comparable and translation data, according to the following guidelines:

- Allomorph(s) of the prefix (if any);
- Prefix used in solid words, hyphenated words, or both;
- Meaning(s) conveyed by the prefix;
- Word categories of the bases and categories of the resulting prefixed words;
- Prefix equivalents (if any) and other types of equivalents (where appropriate)[5];
- Productivity and frequency of the prefix in the source language;
- Register-, genre- or domain-sensitivity (where appropriate);
- Synchronically semantically transparent examples (avoid synchronically opaque derivatives, which are most probably processed as wholes by learners).

On the basis of these recommendations and on the basis of the results of comparable and translation corpus data analysed in Lefer (2009), I have devised a (provisional) model entry for French *dé–,* largely inspired by the *–ish* monolingual entry presented by Prćić (1999) (details are only provided for reversative *dé–* (sense 1) as the other two meanings of the prefix have not yet been thoroughly investigated). The entry is found in Figure 8. The translation comments made in the model entry are based on a detailed analysis of the English translation of French *dé–* in PLECI[6], a bidirectional translation corpus of fiction and journalese (c. 250 derivatives in *dé–* and their English translation were analysed; see Lefer, 2009).

dé– prefix (*dés–* if preceded by a vowel) [mainly found in solid derivatives]

1 added to *vs* to form *vs* | **REVERSAL OF AN ACTION** | Productivity/frequency ☺ ♦ *de–, dis–, un–* • *dénationaliser: denationalise* • *déconnecter: disconnect* • *déboucher (a bottle): uncork*
Notes the prefixal category 'reversal' is more productive in Fr. than in Eng. *dé*-verbs may therefore require a translation into a paraphrase, using verbs such as *remove, lose, clear* • *déculpabiliser: remove sb's guilt* • *désintéresser: lose interest* | *dé*-verbs are also sometimes translated into phrasal verbs, especially in speech and fiction • *dessoûler: sober up* • *décolérer: calm down* • *déplisser: smooth away*

2 added to *adjs/ns/vs* to form *adjs/ns/vs* | **NEGATIVE ('NOT')** | Productivity/frequency ☹ ♦ […]

3 added to *vs* to form *vs* | **INTENSIFYING** | Productivity/frequency ☹ ♦ […]

Figure 8: Model bilingual entry for French *dé–*

---

[5] Non-morphological equivalents, provided they are frequent and systematic in translation corpus data, may be especially useful when comparable corpus analysis has shown that a prefix is more productive in the source language than in the target language. Higher productivity often leads to the use of indirect equivalents and users could greatly benefit from this type of information.

[6] See http://cecl.fltr.ucl.ac.be/research.html [access date: 13 Feb. 2010].

As can be seen from Figure 8, French *dé–*, which is a productive and frequent prefix, can be translated into English *de–*, *dis–* and *un–*. Examples drawn from the PLECI corpus are provided. The 'notes' section in the entry emphasises that French *dé–* is more productive than its English counterpart and may therefore require a translation into a paraphrase using verbs such as *remove*, *lose* or *clear*. It is also indicated that some French verbs in *dé–* are commonly translated into phrasal verbs, especially in fiction and speech.

7. Concluding remarks

Word-forming elements such as derivational affixes are of utmost importance for the interpretation and translation of complex words. However, as shown in this paper, prefixation is currently not adequately described in English-French bilingual dictionaries. Three central issues seem to have been largely neglected: the criteria for the selection of affix entries, affix polysemy and cross-linguistic affix equivalences. Lexicographers could greatly benefit from relying on authentic data from comparable and translation corpora to consistently tackle these issues.

References

Alsina, V.; DeCesaris, J. (2002). 'Bilingual lexicography, overlapping polysemy, and corpus use.' In Altenberg, B.; Granger, S. (eds.). *Lexis in Contrast. Corpus-based approaches*. Amsterdam & Philadelphia: John Benjamins. 215-229.

Atkins, B.T.S. (2002). 'Then and now: competence and performance in 35 years of lexicography.' In Braasch, A.; Povlsen, C. (eds.). *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002*. Copenhagen: Centre for Sprogteknologi. 1-28.

Baayen, R.H. (2008). 'Corpus linguistics in morphology: morphological productivity.' In Lüdeling, A.; Kytö, M. (eds.). *Corpus Linguistics. An International Handbook*. Berlin: Mouton de Gruyter. 899-919.

Caluwe, J. de; Taeldeman, J. (2003). 'Morphology in dictionaries.' In Sterkenburg, P. van (ed.). A *Practical Guide to Lexicography*. Amsterdam & Philadelphia: John Benjamins. 114-126.

Cartoni, B. (2008). 'La place de la morphologie constructionnelle dans les dictionnaires bilingues: étude de cas.' In Bernal, E.; DeCesaris, J. (eds.). *Proceedings of the XIII EURALEX International Congress* (Barcelona, 15-19 July). Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. 813-820.

Corréard, M.-H. (2005). 'Bilingual lexicography.' In Brown, K. (ed.). *Encyclopedia of Language and Linguistics*. Oxford: Elsevier. 787-796.

Dardano, M.; Frenguelli, G.; Colella, G. (2006). 'What Lexicographers Do with Word Formation.' In Corino, E.; Marello, C.; Onesti, C. (eds.). *Proceedings XII Euralex International Congress*. Torino, Italia, September 6th-9th, 2006. Allessandria: Edizioni dell'Orso. 1115-1127.

Ferraresi, A.; Bernardini, S.; Picci, G.; Baroni, M. (2008). 'Web corpora for bilingual lexicography. A pilot study of English/French collocation extraction and translation.' In Xiao, R.; He, L.; Yue, M. (eds.). *Proceedings of The International Symposium on Using Corpora in Contrastive and Translation Studies*. Zhejiang University, Hangzhou, 25-27 September 2008.

Fløttum, K.; Dahl, T.; Kinn, T. (2006). *Academic Voices – across languages and disciplines*. Amsterdam & Philadelphia: John Benjamins.

Hacken, P. ten; Abel, A.; Knapp J. (2006). 'Word formation in an electronic learners' dictionary: ELDIT.' In *International Journal of Lexicography* 19 (3). 243-256.

Iacobini, C. (1998). 'Distinguishing Derivational Prefixes from Initial Combining Forms.' In Booij, G.; Ralli, A.; Scalise, S. (eds.). *Proceedings of the First Mediterranean Conference of Morphology*. Patras: University of Patras.

Jackson, H. (2000). 'Lexicography.' In Booij, G.; Lehmann, C.; Mugdan J. (eds.). *Morphologie / Morphology. Ein internationales Handbuch zur Flexion und Wortbildung. An International Handbook on Inflection and Word-Formation*. Berlin: Walter de Gruyter. 1882-1892.

Krishnamurthy, R. (2005). 'Corpus lexicography.' In Brown, K. (ed.). *Encyclopedia of Language and Linguistics*. Oxford: Elsevier. 250-254.

Lefer, M-A. (2009). *Exploring lexical morphology across languages: a corpus-based study of prefixation in English and French writing*. Unpublished PhD thesis. Louvain-la-Neuve: Université catholique de Louvain.

Prćić, T. (1999). 'The treatment of affixes in the 'big four' EFL dictionaries.' In *International Journal of Lexicography* 12 (4). 263-279.

Prćić, T. (2005). 'Prefixes vs initial combining forms in English: a lexicographic perspective.' In *International Journal of Lexicography* 18 (3). 313-334.

Prćić, T. (2008). 'Suffixes vs final combining forms in English: a lexicographic perspective.' In *International Journal of Lexicography* 21 (1). 1-22.

Roberts, R.P. (1996). 'Parallel text analysis and bilingual lexicography.' Paper presented at *AILA 1996*. http://www.dico.uottawa.ca/articles-fr.htm [Access date: 31 Jan. 2010].

Roberts, R.P.; Cormier, M. (1999). 'L'analyse des corpus pour l'élaboration du *Dictionnaire canadien bilingue*.' http://www.dico.uottawa.ca/articles-fr.htm [Access date: 31 Jan. 2010].

Salkie, R. (2002). 'Two types of translation equivalence.' In Altenberg, B.; Granger, S. (eds.). Lexis in *Contrast. Corpus-based Approaches*. Amsterdam & Philadelphia: John Benjamins. 51-71.

Salkie, R. (2008). 'How can lexicographers use a translation corpus?' In Xiao, R.; He, L.; Yue, M. (eds.). *Proceedings of The International Symposium on Using Corpora in Contrastive and Translation Studies*. Zhejiang University, Hangzhou, 25-27 September 2008.

Sharoff, S. (2004). 'Harnessing the lawless: using comparable corpora to find translation equivalents.' In *Journal of Applied Linguistics* 1 (3). 311-322.

Szende, T. (2004). 'Une ressource monolingue pour les dictionnaires bilingues. ' In Laurian, A.-M. (ed.). *Dictionnaires bilingues et interculturalité*. Bern: Peter Lang. 243-258.

Teubert, W. (1996). 'Comparable or parallel corpora?' In *International Journal of Lexicography* 9 (3). 238-264.

Teubert, W. (2002). 'The role of parallel corpora in translation and multilingual lexicography'. In Altenberg B.; Granger, S. (eds.). *Lexis in contrast. Corpus-based approaches*. Amsterdam & Philadelphia: John Benjamins. 189-214.

Williams, G. (2008). 'A Multilingual Matter: Sinclair and the Bilingual Dictionary.' In *International Journal of Lexicography* 21 (3). 255-266.

Dictionaries

Grand Dictionnaire Hachette Oxford Français-Anglais English-French on CD-ROM. (2004). Paris: Hachette Multimédia.

Robert & Collins électronique français-anglais / anglais-français - CD-ROM. (2003). Paris: Le Robert.

Harrap's French Unabridged PRO Dictionary on CD-ROM. (2004). Edinburgh: Chambers Harrap.

Larousse Chambers CD-ROM Advanced Dictionary English-French / French-English. (1999). Paris: Larousse-Havas Interactive.