

DIRK SIEPMANN

**COLLOCATIONS AND EXAMPLES: THEIR RELATIONSHIP  
AND TREATMENT IN A NEW CORPUS-BASED LEARNER'S  
DICTIONARY**

**Abstract:** The production of learners' dictionaries raises a number of specific problems regarding the treatment of collocations and examples. It has been claimed, among other things, that example sentences should not be based exclusively on collocations. The present article argues that a clear distinction must be drawn between collocations and examples, and that, under certain conditions, collocations must also be exemplified. It discusses various approaches to the microstructural arrangement of collocations and establishes guiding principles for the choice and presentation of examples.

**1. Introduction**

In recent years great strides have been made in linguistic description – a development that was partly sparked by lexicographic projects (COBUILD, Herbst et. al.: *A Valency Dictionary of English* 2004). With unprecedented riches of data and powerful enquiry tools at their fingertips, lexicographers have produced dictionaries which give a highly accurate and very detailed picture of language in use. Ample testimony to this can be found in the five “big” learners' dictionaries and in many bilingual works of reference. One area that has received particularly close attention is collocation. Thus, the *Oxford Collocations Dictionary for Students of English* has succeeded in recording 150,000 collocations of 9,000 common nouns, verbs and adjectives.

That said, there is still room for improvement. One type of shortcoming, which I have enlarged on elsewhere (Siepmann 2005b; Siepmann 2006; Siepmann, forthcoming), is linguistic: particular types of collocations have so far been given short shrift or are difficult to access. Another type of shortcoming is pedagogic: the major monolingual learners' dictionaries are reference tools for occasional consultation, but they cannot be used for serious vocabulary expan-

sion. To assimilate the vocabulary of a foreign language, students therefore need to turn to works in which lexical items are arranged in topic areas.<sup>1</sup>

However, thematically arranged bilingual thesauri designed for vocabulary learning still lag badly behind their monolingual alphabetical counterparts in terms of both depth and breadth of coverage, and the production of such dictionaries raises a number of specific problems regarding selection and categorization as well as exemplification and translation. The present article provides an account of some of the solutions adopted in the Bilexicon project (see Siepmann 2006), whose purpose is to produce bilingual thesauri geared to the needs of the foreign learner. Section 1 gives a broad definition of collocation. Section 2 proceeds to discuss various solutions to the arrangement of collocations in the microstructure. Section 3 discusses exemplification, placing particular emphasis on the problems posed by a) the distinction between collocations and examples and b) the special characteristics of examples in bilingual learners' dictionaries.

## 2. Definitions of collocation

This is not the place for a detailed discussion of various definitions of collocation (see Herbst 1996; Hausmann 2003; Siepmann 2005b). Suffice it to say that the subject of collocation has been approached from three main angles.

Semantically-based approaches (e.g. Benson 1986; Mel'čuk 1998; González-Rey 2002; Hausmann 2003) assume that collocations are typically made up of two constituents which differ in their semantic status: a semantically autonomous base such as *compliment* combines with a semantically dependent collocate like *pay*, in such a way that the collocate takes on a specific meaning ('offer') contingent on the base. In many cases this difference in semantic status allows semanticists to make two important distinctions: a) between collocations and free combinations (semantically autonomous + semantically autonomous: *he likes money*) and b) between collocations and phraseology (i.e. semantically irregular items).

The frequency-oriented approach looks at statistically significant co-occurrences of two or more words. It is therefore alone in providing a workable heuristic for discovering the entire class of co-occurrences, but its exclusive reliance on automatic statistical analysis has sometimes led to the inclusion of chance co-occurrences such as *hotel at, either hotel* (Kjellmer: *A Dictionary of English Collocations* 1994) or *nature because* (Sinclair: *Collins Cobuild English Collocations on CD-ROM* 1995) and to an insufficient consideration of lexico-grammatical and semantic-pragmatic factors. As noted by Klotz (2000, 83), for example, a purely formal analysis of collocations such as *catch + ball* is insufficient to disambiguate the verb, which can be variously paraphrased as 'grab', 'record' or 'hit'.

---

<sup>1</sup> It goes without saying that such dictionary use must be supplemented by exposure to input and guided production of correct output.

A third, more recent approach to phrasemes and collocations (Feilke 1996; 2004) might be termed “pragmatic” since it claims that the structural irregularities and non-compositionality underlying such expressions are diachronically and functionally subordinate to pragmatic regularities which determine the relationship between the situational context and linguistic forms. The proponents of this approach take the line that collocation can best be explained via recourse to contextualisation theory (Fillmore 1976). However, they run into difficulties when it comes to explaining a large number of co-occurrences linked by common semantic features (cf. Siepmann 2005b, 424-430).

The problems attendant upon these three approaches can be resolved in all-encompassing approaches to language theory (Hoey 2005) and lexicography (Siepmann 2005b) which take collocation as their starting point. In keeping with neurological evidence on the structure of the brain (Lamb 1999), such approaches assume that, in speakers’ minds, lexical items become progressively loaded with all sorts of information about their typical grammatical, lexical, semantic and pragmatic contexts of use. In Hoey’s terms, words and other units may be “primed” for lexical and grammatical collocation as well as for semantic-pragmatic association. It is the task of the pedagogic dictionary maker to describe such collocational primings in a readily learnable form. To this we now turn.

### 3. Treatment of collocations

Just like individual words in traditional word-based dictionaries, collocations can be grouped and accessed either by purely formal criteria or by semantic criteria. By formal criteria are meant ordering by the alphabet, word class, case marking, clause function or “lexical function”, a term coined by Mel’čuk (see, for example, Mel’čuk 1998). These criteria may be combined in various ways. Thus, Kubczak and Costantino (1996, 22) organize the German section of their “syntagmatic” dictionary according to the form the noun takes within the collocation: collocations where the noun appears in the nominative case are given first, followed by genitive, dative and prepositional uses; likewise, the French section divides each entry into pre-verbal noun phrases, post-verbal noun phrases and prepositional phrases. Kubczak and Costantino (1996, 22-23) argue that this organizing principle greatly facilitates consultation, regardless of whether the dictionary is a book, a CD-ROM or an Internet reference work.

Entries in the *Dictionnaire Explicatif et Combinatoire du Français Contemporain* (DEC) are organized according to lexical functions. Thus, at *blesure*, lexical functions and the corresponding collocations have been grouped in the following way:

Epit	:	morale, psychologique; intime, secrète [...]
Magn	:	grande   antépos, grave   antépos ou postpos, atroce, cruelle, douloureuse < béante, profonde
AntiMagn	:	superficielle
AntiBon	:	envenimée
Oper1	:	avoir [ART ~], porter [ART ~ en Pron = X]; éprouver [ART ~ ] [...]

(DEC, s.v. *blessure*)

DEC is intended to be a “scientific” dictionary, and its symbols and terminology would put off all but the most sophisticated users. Nonetheless pedagogic lexicographers may use DEC as a quarry and its lexical functions as a guideline for structuring entries. They may, for example, juxtapose items that fulfil the same lexical function and/or may follow up specific collocations with their antonyms (e.g. Magn vs. AntiMagn).

*Langenscheidts Kontextwörterbuch Französisch-Deutsch* (LKF) arranges its material firstly by part of speech and secondly by the alphabet, with synonymic collocations added in random order after the first sub-entry:

**demande** f *Antrag, Gesuch*

**accéder, donner suite** à une ~ *einem Antrag stattgeben; einen Antrag genehmigen*: Malheureusement, nous ne pouvons pas accéder à votre ~. / **admettre, agréer, accorder** une ~ *einen Antrag billigen, gutheißen*: Le ministre a admis votre ~. [...]  
(LKF, s.v. *demande*)

The *Oxford Collocations Dictionary for Students of English* (OCD) adopts a similar approach in dividing each entry according to part of speech, but differs from LKF in structuring each division by intuitively plausible criteria. Thus, at *pollution*, the section devoted to verb-noun collocations proceeds from the strongest form of action (*avoid/prevent*) to the mildest (*monitor*). Probably unbeknownst to the editors of OCD, Reum followed the same principle with great ingenuity as far back as 1910 (cf. *Petit Dictionnaire de Style à l'Usage des Allemands* and *A Dictionary of English Style* [DES]). It is interesting to compare the ways in which OCD and DES handle the same content; a good example is provided by the entries for *river*.

DES (s.v. <i>river</i> )	OCD (s.v. <i>river</i> )
E. great, broad, large, long : small, narrow, short, straitened between rocks; deep : shallow; <u>fordable</u> ; <u>rapid</u> , <u>abounding in rapids</u> , navigable for small boats, navigable; <u>rolling</u> , <u>rushing</u> , <u>sweeping by</u> , <u>gurgling</u> , <u>murmuring</u> , <u>purling</u> , <u>babbling</u>	<b>ADJ.</b> broad, great, large, long, <u>mighty</u> , <u>wide</u> ; <u>little</u> , narrow, short, small; deep; shallow; <u>high</u> ; <u>low</u> ; <u>fast-flowing</u> ; <u>slow-moving</u> , <u>sluggish</u> ; <u>winding</u> ; <u>swollen</u> ; navigable
<b>Va.</b> a river rises <i>or</i> <u>has its source</u> in the mountains; to <u>precipitate its course</u> ; to flow, to run between, through and over rocks; to <u>work</u> o.s. a <u>channel</u> , a <u>passage</u> ; to <u>wind its way through fields and meadows</u> ; to <u>form</u>	<b>VERB + RIVER</b> cross, ford, get across; bridge; dam; dredge; navigate

<p><u>islets; to form rapids, cascades and waterfalls; to turn to the right : left; to take another course or direction; to flow round a hill; to fork, to divide itself in several arms; to meander; to grow big; to increase, swell, become larger, mightier, more rapid and more unruly, to roll majestically along, to overflow, to inundate, overflow or deluge, submerge wide tracts of land; to roll muddy (miry, dirty) water; to freeze, to drift with ice; to fall, to re-enter its bed; to have a strong, mighty, rapid : feeble current; to receive several tributaries; to irrigate or water : drain a country; to flow or run through a town, a country; to pass by N.; to wash the walls of N.; to have or attain a length of a thousand kilometres, a breadth of a hundred metres, a depth of some ten metres; to flow towards the ocean; to fall, run, discharge itself, empty itself into the sea. –</u>  O.S. <u>to go, row, scull, paddle, steer, sail down : up the river (= downstream : upstream); to walk along, to skirt a river; to cross a river; to swim across, to ford a river; to bathe in a river; to skate on a river; to boat it on a river; to throw a bridge over a river; to cross a river on a bridge, by means of a pontoon-bridge; a river is crossed by several bridges; to change, to regulate the course of a river; a brook joins a river; to watch the glinting river with its drifting barges; to admire the trim, well-kept villas on both sides of the river</u></p>	<p><b>RIVER + VERB</b> flow, run, wind; rise; be in flood, burst its banks, flood (sth), overflow sth; dry up</p>
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------

Table 1. A comparison of the entries for *river* in DES and OCD

As far as adjectival collocations are concerned, it is evident that the two dictionaries cover roughly the same meaning spectrum, albeit with differing collocations (e.g. *rapid* vs. *fast-flowing river*). An interesting exception is the inclusion, in DES, of collocations denoting the sound of flowing water, such as *gurgling river* or *babbling river*. Although the participial adjective *swollen* is absent from the adjective section of DES, it has been entered in infinitival form in the verb section (*swell*).

Nowhere is the difference between the two dictionaries more noticeable than in the section given over to verb-noun collocations in which the noun functions as subject. DES builds up an almost complete panorama of actions and events that may occur as one follows the course of a river from its source to its mouth. These actions and events are clearly sequenced, and there is a wide variety of synonymous collocations expressing each idea. By contrast, the picture provided in OCD is rather sketchy, and the ordering of events is not as intuitive as claimed in the introduction. From a pedagogic perspective, then, the DES entry for *river* has clear advantages over its modern counterpart.

One of the reasons for this is the severe neglect, in OCD, of three-word collocations. Only one such item has been recorded in OCD (*the river bursts its banks*), whereas DES can boast more than a dozen. Most of these triples can be segmented into their component parts, which raises the question of whether to record them at the entry for the first or the second noun involved. Thus, at least theoretically, *have its source* could be recorded at *source*, and *re-enter its bed* could be recorded at a specific sense division of *bed*. This is how a strictly word-based dictionary would proceed. A pedagogic dictionary should strike a compromise by including three-word items at the entries for heavy-duty nouns as opposed to those for infrequent lexical items. For example, the three-word collocation *the river forms rapids*, which is fairly specific to rivers, should be entered at *river* and, space permitting, at *rapids*; it must be recorded at *river* if *rapids* is not given entry status.

At the risk of duplicating information, one should follow the example of DES by recording even seemingly unspecific collocations at *river*. A pertinent example is afforded by *river + take + course/direction* (e.g. *the river takes a southerly course, each river takes its own course*); it would be unreasonable to expect users to learn such combinations by consulting the entry for *river* as well as those for *course* or *direction*, especially since OCD makes no mention of the collocation *take + course* at *course*; the entry for *direction* does inform the user about its collocation with *take*, but the link with *river* is far from obvious, as the only example given is “Which direction do we take?”. Hardly any dictionary user would go to such lengths to look up a collocation, and users of a learner’s dictionary in particular would expect all the useful triples to be listed. Another reason for including three-word collocations is that these may sometimes correspond to two-word items in other languages (e.g. *régulariser un fleuve / einen Fluss begradigen* → *straighten / (dated) regulate [the course of] a river*).

A line should perhaps be drawn at combinations such as *river + attain a width of*, where the verb phrase applies to a vast range of objects, but even here the case can be made that students may gain some advantage from having such phrases ready-made at their disposal. Otherwise they would waste a great deal of time ferreting out such information from other entries. Similar considerations apply to word combinations such as *go/row/paddle down (the) river*. By contrast, *throw a bridge over a river*, recorded at *river* by DES, should be counted among the collocations of *bridge* rather than *river*, as *bridge* is the primary object; bridges can be thrown across ravines, ditches and rocks as well as rivers. Since *bridge* is a high-frequency noun, learners will have to learn its collocations anyway. By the same token, exemplificatory word groups such as *to admire the trim, well-kept villas on both sides of the river* (DES) should be kept apart from entries.

It is clear from this brief comparison of DES and OCD that three-word collocations should have a firm place in bilingual thesauri. It is only with the help of such items that the learner can form a broad picture of an item’s preferential uses and situational embedding.

The projected dictionary of French noun-based collocations (see Blumenthal 2005) goes one better than either LKF or OCD in being based on a more refined syntactic framework:

1.1	ADJ + MOT	( <i>un bon <u>sentiment</u></i> )
1.2	MOT + ADJ	( <i>un <u>sentiment</u> aigu</i> )
1.3	MOT + NOM	
2.1	MOT + de + NOM <sub>subjectif</sub>	( <i>le <u>sentiment</u> de notre peuple</i> )
2.2	MOT + PRP + NOM <sub>objectif</sub>	( <i>un <u>sentiment</u> d'insécurité</i> )
2.3	MOT + PRP + INF	( <i>le <u>sentiment</u> d'être trahi</i> )
2.4	MOT + COMPLÉTIVE	( <i>le <u>sentiment</u> que</i> )
2.5	MOT + PRP + NOM <sub>circonstanciel</sub>	( <i>un <u>sentiment</u> vis-à-vis de</i> )
2.6	MOT + RELATIVE	( <i>l'<u>hypothèse</u> selon laquelle</i> )
3.1	NOM <sub>quantifiant</sub> + MOT	
3.2	NOM + PRP + MOT	( <i>l'<u>ambivalence</u> de ses <u>sentiments</u>, le <u>chantage</u> au <u>sentiment</u>)</i> )
3.3	NOM + MOT	
4.	ADJ/ADV + PRP + MOT	( <i><u>animé</u> d'un <u>sentiment</u></i> )
5.1	MOT <sub>sujet</sub> + VERBE	( <i>un curieux <u>sentiment</u> l'<u>habitait</u></i> )
5.2	MOT + ATTRIBUT	( <i>ce <u>sentiment</u> de responsabilité <u>est</u> essentiel</i> )
6.1	VERBE + MOT <sub>objet</sub>	( <i><u>traduire</u> un <u>sentiment</u></i> )
6.2	VERBE <sub>attributif</sub> + MOT	( <i>la haine est un <u>sentiment</u> beaucoup trop lourd à porter</i> )
7.1	MOT <sub>circonstanciel</sub>	
7.2	PRP + MOT <sub>circonstanciel</sub>	( <i>ils emboîtent le pas à leurs partenaires <u>contre</u> leur meilleur <u>sentiment</u></i> )

Table 2. Entry for *sentiment* in the projected dictionary of French noun-based collocations

With only minor modifications, this framework can be transposed to other European languages such as English. The most important new category which Blumenthal (2005) sets up is category 7; no other dictionary has so far given sufficient consideration to the use of common nouns within adverbials. In line with Hoey's (2005) observation that semantically similar words are normally primed differently, Blumenthal (2005, 277-278) notes that near-synonyms such as French *euphorie* and *enthousiasme* can be distinguished on the basis of their adverbial uses; adverbials containing *enthousiasme* are closely linked to the verb and prefer postverbal positions, whereas adverbial phrases centred around *euphorie* occur more commonly in clause-initial position.

Innovative as it is with respect to syntactic categorization, the dictionary of French collocations has no new policy to offer on semantics. In fact, it abandons semantic considerations in favour of a strictly alphabetical order. This befits its purpose as a reference tool, but, for reasons of learnability, such a policy would not be suitable in a learner's dictionary.

The Bilexicon therefore adopts the comprehensive syntactic categorization of the French dictionary of collocations, while making judicious use of semantic ordering criteria wherever possible. Unlike the procedure followed in OCD, where the reader is left to work out the organizing principles underlying entries, the Bilexicon lays open to inspection the lexicographer's semantic intuitions in the extra column (see the appendix for an example based on the noun *germ*).

#### 4. Treatment of examples

An example can be usefully defined as any stretch of text that contains an occurrence of the entry term or phrase (cf. Martin 1989, 600; Métrich 1993, 122). In collocational dictionaries this stretch of text should, of course, extend beyond collocational units, so that collocations are not deemed to be examples.

This section will investigate three main questions concerning exemplification in bilingual thesauri:

1. Is exemplification always necessary?
2. What are the functions of examples?
3. What should examples look like? Should examples be authentic or simulated instances of language in use?

Let us begin with the first question. As Hausmann (1977, 84) and Zöfgen (1994, 299) point out, sentence-length examples are not always necessary; nor does the informational value of the example necessarily increase as its length increases. This seems to suggest that an onomasiological collocational dictionary does not need to illustrate each entry with an example. Work on the Bilexicon has shown that several cases can be distinguished.

Many of the prototypical items discussed by the proponents of the semantically-oriented approach to collocation do not normally require exemplification. This is especially true when there are no restrictions on verbal or nominal inflection, when a noun-verb collocation has a human subject and when a collocation enters only one pattern, as with *suffer an injury*, *submit an application* or *eligible bachelor*. It is less so with items such as English *set off an explosion* or French *emporter la conviction*, which can take both animate and inanimate subjects, many of which are abstract. They therefore stand in need of illustration: *the advent of realism set off an explosion of categories for fiction*; *aucune de ces opinions extrêmes n'emporte la conviction*. Similarly, collocations subject to inflectional and other restrictions such as *pass judgment* (?*pass judgments*, ?*pass a ADJ judgment*) should be exemplified. Note that, in these and similar cases, onomasiological ordering confers certain advantages: it would record a polysemous collocation such as *set off an explosion* (= 1. cause a sudden loud bursting, 2. cause a sudden large increase) in two or three entirely different chapters of the dictionary (1. general and abstract terms; war 2. quantification); only in the second of these would it be of the essence to exemplify and to specify colligational patterning (*explosion + of + N*, *explosion + in + N*). Sometimes, however, a larger number of examples is needed, especially when a collocation occurs in two

or more patterns. With collocations such as *cause concern*, which admits of divalent and trivalent patterning (cf. Klotz 2000, 189-195), two examples are usually necessary to illustrate each pattern. The same is sometimes true of polysemous collocations, especially when the (usually) two senses of the collocation translate similarly or identically (cf. for example *leave the road – quitter la route*).

Pragmatically-based, situation-specific collocations, many of which are situated on the borderline between what is traditionally referred to as speech formulae and idioms, do not normally need to be exemplified if the lexicographer provides a one-to-one translation belonging to the same situation type. Thus, a chapter on emotions might include the following phrases in a sub-section on the notion of embarrassment: *I wanted the ground to open up (and swallow me [up]) – am liebsten wäre ich vor Scham in die Erde versunken* (synonyms: *am liebsten wäre ich in ein Mauseloch gekrochen / am liebsten wäre ich in den Boden versunken*); *I didn't know which way to look (where to look) – ich wusste nicht, wo ich hinschauen sollte*; *I was as red as a beetroot – ich bin ganz rot angelaufen* (synonym: *ich war rot wie eine Tomate*).

By contrast with collocations, which, if appropriately recorded (e.g. *s.th. causes [s.o.] concern*) automatically reveal their colligational possibilities, colligational patterns should always be exemplified in typical situational contexts. One reason for this is that traditional valency patterns should be regarded as selective abstractions from the reality of language use. Another is that collocation and colligation are tightly interlinked, such that particular patterns have preferred lexical realisations or occur in specific contexts. This makes it necessary to give the broadest possible portrayal of actual language use, especially with heavy-duty verbs such as *drive* or *run*.

It stands to reason that discourse markers, by virtue of encoding cohesion relations between longer stretches of text, require more ample exemplification (cf. Siepmann 2005a, 305); as with speech formulae or idioms, however, illustrative passages may be omitted in an onomasiological bilingual dictionary provided that there is perfect symmetry in usage between translation equivalents.

In sum, it can be said that onomasiological ordering diminishes the need for exemplification, as even pragmatically-based collocations will be immediately clear to the user provided that they are adequately translated and embedded within the relevant situation type. All this militates in favour of the policy of moderate exemplification suggested by the above theoretical considerations.

We now turn to our second question. If we consider the functions that have been posited for exemplification (cf. Siepmann 2005a; Zöfgen 1986; 1994, 185-186; Drysdale 1987; Métrich 1993), we find that most authors identify up to ten different such functions. On close inspection, only some of these functions prove to be relevant to bilingual pedagogic thesauri. Firstly, as Zöfgen (1994, 198) notes, it is not the role of the pedagogic dictionary to attest by example the mere existence of a word or usage. Secondly, and more importantly, the functions of meaning differentiation, definition, metacommunicative comment and guidance on register are all well served by translation. This is nowhere more

apparent than in relation to pragmatically-based collocations. Whereas monolingual learners' dictionaries experience great difficulty in providing suitable and succinct examples of such idioms as *I've put my foot in it* or *will I do (as I am)?*, a good translation easily gets around the problem by offering a perfect equivalent in the appropriate thematic or situational context (e.g. *ich bin in ein Fettnäpfchen getreten; kann ich so gehen?*). Since language learners will be able to retrieve typical contexts from experience of their source culture, there is no need to contrive a lengthy explanatory example. Exemplification will only be required if perfect translation equivalents have slightly different connotations or semantic prosodies (Louw 1993). Unlike *treffende Beschreibung*, for example, *apt description* is often used ironically.

Thirdly, although any example may be deemed to have a pedagogic function in teaching the reader about the uses of an entry word or phrase, the genuinely pedagogic functions of what we might call the "learnability" and "usability" of examples have been largely overlooked. This point will be enlarged upon below.

This leaves us with six main functions of examples in bilingual pedagogic thesauri. Firstly, examples should illustrate the syntactic, semantic and pragmatic patterns typically found in the discourse surrounding the entry word or phrase or, more simply put, to show it embedded in a natural context; the example should exhibit prototypical features of the entry word or phrase (Siepmann 2005a). As already mentioned, some collocations (e.g. *easy prey*), although they may be regarded as examples of individual words, should themselves be illustrated. If a full-sentence example is provided in an onomasiological translation dictionary which includes detailed information on collocation and colligation, this should ideally contain more than just a convenient illustration of collocational, colligational and situational features which are evident from the collocational entries and valency patterns recorded in the entry (cf. Zöfgen 1994, 186, 189). However, this clearly is an ideal that cannot always be achieved in real life. In incriminating the use of illustrative sentences that are exclusively based on collocations, such as *Il a affronté de grands dangers* (*Dictionnaire du Français Contemporain*), Zöfgen (1994, 189) fails to realise that it is almost impossible not to use collocations in examples since language is predominantly collocational (for an illustration of this point, see below); thus, it will be argued that *don't spread your germs among us* (*Dictionnaire de l'Anglais Contemporain* [DAC], s.v. **germ**) is a well-chosen prototypical example even though it makes use of a common collocation (*spread + germs*; illustrative function).

Secondly, examples may provide encyclopedic information. Zöfgen (1994, 187) views this function as being at variance with the aims of a language dictionary, but concedes that occasional forays into the realm of culture may be permissible. Given that cross-cultural approaches have recently come to the fore in language teaching, there are good grounds for systematically incorporating cultural information into examples or the extra column, especially in the case of cross-cultural differences. The aim may be to provide learners with the typical scenarios in which words occur (cf. Hausmann 2005; encyclopedic function).

One of the central functions of examples in vocabulary books is to offer instantiations of language use which can be easily memorized. In other words, examples should be as mnemonic as possible – this brings a new element into the equation, an element that has never been seriously discussed in the literature on monolingual learners' dictionaries. This function imposes a number of constraints on the length of examples (7 +/- 2 chunks), the choice of vocabulary, etc. (see below). It also means that the makers of bilingual vocabulary books cannot always simply copy examples from monolingual learners' dictionaries and translate them (learnability).

Another important function of examples is to provide a critique of the standard translation in cases where such a critique is needed, thus adducing evidence that specific contextual embeddings may result in different translations. Consider, for example, the collocation *spread + germs*, which translates literally into German as *Keime/Bazillen + verbreiten*. This translation will stand the learner in good stead in most situations (e.g. *birds may spread the virus – Vögel könnten das Virus verbreiten*); however, a phrase such as *don't spread your germs around* should be idiomatically rendered as *behalte deine Bazillen für dich* rather than *?verbreite deine Bazillen nicht*. This again shows – if the reader still needs proof – that collocations themselves may be contextually restricted, and this should be richly illustrated by examples. The entry for *butiner* in the projected BILEXICON provides another example of this “critique function”:

une abeille butine une plante	a bee visits a plant (to collect nectar)
<i>Les abeilles qui butinent les champs de tournesol traités au Gaucho sont peu à peu atteintes de paralysie et ne retrouvent plus le chemin de la ruche.</i>	<i>Bees which collect nectar from sunflower fields treated with Gaucho become paralysed and are unable to find their way back to the hive.</i>

Table 3. Entry for *butiner* in the projected BILEXICON

In an ideal world, examples should also entertain and provide food for thought. They should incite the reader to peruse and learn them. The *butiner* example just cited is a neat illustration of this, suggesting, as it does, that bees are an endangered species (motivational function).

Lastly, but equally importantly, examples have the pedagogic function of helping learners to encode meaning. Learners are wont to slur over information provided in pattern formulae and may prefer to model their production on examples. Although the Bilexicon differs from traditional monolingual learners' dictionaries in using both natural-language coding (*s.o. gets germs somewhere*) and symbolic complement frames (V + N + *on/onto/all over* N), learners are still likely to turn to the examples (*you might get the germs on your fingers*) for purposes of encoding and memorization. This is confirmed by suggestions made by experienced language teachers who have written on vocabulary learning (cf. Hill 2000; Hohmann 1999; Gallagher 2006, 79).

This brings us to our third question, namely what examples should look like. Related to this is the question as to whether examples should be authentic or invented. This has been variously answered in the literature. The Cobuild dictionaries have initiated a strict policy of authentication that has come under scathing attack from European meta-lexicographers. Zöfgen (1986; 1994, 183-200), for example, is adamantly opposed to exclusive reliance on corpora; his main arguments are as follows:

1. illustrative quotations may be just as cliché-ridden or stilted as invented examples (Zöfgen 1994, 193);
2. illustrative quotations may be more difficult to understand than made-up examples (Zöfgen, 1994, 194);
3. authentic examples may be strikingly unnatural or highly idiosyncratic (Zöfgen 1994, 194-195);
4. even attested examples may no longer be considered authentic when shorn of their context because their function within the dictionary differs from their original function (Zöfgen 1994, 198).

Although partially well-founded, Zöfgen's criticisms are mainly directed against manually collected examples culled from what must have been a fairly small corpus base (such as "er jagte drei irre blickende klapperdürre Rinder" in Mugdan 1985, 213; an example which clearly betrays its literary origin). Such criticisms are rendered ineffectual by the ready availability of electronic text from the Internet and other sources; there is such an abundance of corpus examples these days that lexicographers can choose the most natural and lively ones while discarding those that savour of contrivedness or affectation. Furthermore, comprehension problems encountered by users of monolingual dictionaries are less severe, if not inexistent, in the case of bilingual dictionaries because the translation provides an aid to understanding (although, for reasons of learnability, examples should not contain too much problem vocabulary). Zöfgen's last point of criticism, which concerns the inauthenticity of corpus examples, may be theoretically valid, but has no practical relevance; since collocations are inextricably connected in speakers' minds with specific contexts, lexicographers need not always reproduce these contexts. Here too the equivalence relation between source and target language items created by a perfect translation will help users to conjure up the right sort of context.

Elsewhere in his opus magnum, Zöfgen (1994, 277) argues that made-up examples may not always fill the bill. He considers illustrative sentences like the following to be empty, meaningless and unimaginative: *je bronze vite, il calcule vite, calmez-vous, la perdrix s'envola*. If we accept, however, that example sentences should exemplify meanings which are commonly encoded by native speakers, the above sentences may be considered well-chosen. This becomes abundantly clear if slight additions are made, even though such additions are by no means necessary: *vu que je suis originaire du sud, je bronze vite; plus la fréquence est élevée, et plus il (i.e. l'ordinateur) calcule vite*. If we also bear in mind that an ideal dictionary should distinguish between collocations and examples of

the headword, then the above examples are quite pertinent although they contain no collocations. The language learner has to acquire typical free combinations as well as typical collocations.

On balance, then, the arguments put forward by opponents of authentication turn out to be less compelling today than they may have been as little as ten years ago. They need to be set against clear evidence that made-up examples may lack some naturalness feature or other (cf. Sinclair 1984). Thus, Sinclair cites the textbook example “Prince Charles is now a husband”, whose colligational patterning is unusual. In similar vein, Fox (1987, 141) shows that isolated, artificial examples such as “the teacher used to cane me when I behaved badly” could not plausibly fit into an extended stretch of discourse. Quite unlike Zöfgen (1994), Sinclair (1991, 4) therefore comes to the valid conclusion that “however plausible an invented example might be, it cannot be offered as a genuine instance of language in use”. Of a piece with such reasoning is Landau’s view on the matter:

using invented examples is like fixing a horse race: the lexicographer invents an example to justify his definition instead of devising a definition to fit the examples. (Landau 2001, 210)

It seems that the main reasons why authentic samples of corpus material may not always be suitable as dictionary examples are in fact quite different from those advanced by Zöfgen. The following stretches of conversation from the BNC, which contain authentic occurrences of the English lexical unit *germ* (= bacillus), will serve to illustrate what is meant.

- (1) to take away all the nasty to kill the germs [...]
- (2) it'll mask the smell but it's also basically to kill the germs [...]
- (3) a germ, there's a germ that floats about in the air and it gets onto [...]
- (4) you don't want to tip it out on there cause you get germs, darling [...]

These examples make the point that if real speech were incorporated into the dictionary with tape recorder accuracy, some of the resultant illustrative sentences, apart from being unreadable and difficult to memorise, could not be held up as specimens of good usage. Ordinary unscripted conversation is characterized by different kinds of dysfluency and error (cf. Biber et al. 1999, 1052-1066; Carter/McCarthy 1997): silent and filled pauses, repeats, reformulations, grammatically incomplete utterances, slurring, assimilation and syntactic blends. All of these will unnecessarily distract the learner from the essential features of the headword and should not be taken over into the learner’s repertoire. Apart from such performance phenomena, there are other features of authentic speech which may sometimes make it unsuitable for use in dictionary examples (cf. Quirk et al. 1985; Carter/McCarthy 1997; Gallagher 2006, 61-62): general words which are situationally rooted (*that thing's driving me mad; I don't want to get involved with that business*; Carter/McCarthy 1997, 16; emphases mine), context-specific deixis, ellipsis, non-standardisms (e.g. *criteria* used as a singular) and pleonasm (*and so therefore* instead of *and so* or *therefore*; *be in a position to be able to* instead of *be in a position to* or *be able to*, etc.).

That said, other features of authentic speech should be actively reproduced in learners' dictionaries so that learners can acquire some sense of the constructional regularities underlying this type of language. Most importantly perhaps, there should be abundant exemplification of prefaces (also known as "heads" or "topics"; cf. Carter/McCarthy 1997, 16), bodies and tags (cf. Biber et al. 1999, 1072-1082): *North and South London* (head), *they are two different worlds* (body), *aren't they* (tag)? Discourse markers (*actually, you know, I mean*, etc.) and vague language (*what with the weather and everything*) are also quite prevalent in spoken language, and their learnability may be increased through frequent use in examples.

From what has been said it should have become clear that a compromise must be struck between genuineness, usability and comprehensibility. Although only few concessions should be made on authenticity, the lexicographer must be allowed some freedom in manipulating corpus examples. Experienced lexicographers like Landau concur with this view:

there are times when for good practical reasons one has to modify a quotation for the particular readership of the dictionary in question, and sometimes even invent a quotation based on what the corpus evidence has told one about the likely context and syntax of the use of the word [...] To make sense of much real language, one needs to look at more than one sentence, but the dictionary editor hasn't the space available to give examples five or six lines. (Landau 2001, 208-209)

To this we might add that the editor of a bilingual thesaurus would not even wish to devote five or six lines to an example because this would severely reduce its learnability; indeed, learners might not even bother to read such lengthy examples. This implies three possible procedures:

- a) the lexicographer may pick authentic examples direct from the corpus – this should be the standard procedure whenever possible;
- b) the lexicographer may choose intuitively appealing corpus examples and modify them in some way;
- c) the lexicographer may construct an example on the basis of his linguistic intuitions and may proceed to consult the Internet to check whether the invented example corresponds to real usage. If, for instance, the lexicographer's corpus does not contain sufficiently simple examples, it may be advisable to make up a simple one and check its plausibility against the Internet.

With all three procedures, the lexicographer must have "quality control" criteria that will enable her to make reasoned choices. The following criteria might be adopted on the basis of our discussion and the research literature:

1. examples should be easy to understand and should therefore be carefully checked for rare or obsolete words;
2. examples should be prototypical, i.e. they should exemplify meanings typically encoded by native speakers;
3. examples should create the illusion of living speech or real writing;

4. examples of spoken language should be fully mnemonic (i.e. not longer than 7 [+/- 2] chunks);
5. examples should be “readable”, i.e. they should not be phrased in hermetic style nor should they contain disfluency phenomena; it may be preferable to use fictional conversations from radio or television drama rather than unscripted ordinary speech, for, as Svartvik (1988, 36) points out, “the very nature of speech is lost in an orthographic transcription”;
6. examples should be “reusable”, i.e. they should not normally be highly idiosyncratic or context-specific;
7. examples should illustrate typical differences between languages (which may not be apparent from patterns or standard translations, cf. for example the entry for *hijack* in the DAC; cf. *drive into the wall – gegen die Wand fahren*); learner corpora may be used to detect error-prone areas and examples should be selected accordingly;
8. examples that are clearly representative of a particular geographical variety could be marked accordingly.

With the exception of criteria 4 and 7, all these criteria apply to all learners’ dictionaries. 4 and 7 are peculiar to bilingual thesauri for learners. There is clearly a trade-off between criteria 4 and 3, and there may be a trade-off between 4 and 2; situationality may sometimes be achieved at the expense of brevity. It is worth recalling, though, that adequate translations make for quasi-automatic contextualization, where the monolingual learner’s dictionary often struggles to contrive appropriate embeddings for examples. It is also important to note that function words or phrases such as discourse markers usually require at least three exemplificatory sentences (cf. Siepmann 2005a).

Criterion 1 may be difficult to meet in cases where a word typically occurs in grammatically or lexically complex environments (cf. Fox 1987, 138). According to Landau (2001, 306-307), examples of such words include *residual*, *arbitrary*, *conciliatory*, *contingency* or *respite*. Here too, however, the bilingual thesaurus is superior to the monolingual learner’s dictionary in that the provision of translations is a great aid to decoding examples.

Criterion 2 also merits some further elaboration since it begs the question of how prototypicality can be identified. Of particular interest here are notions such as semantic or pragmatic prosody, isotopy, scenario and collocation. Hausmann (2005) ingeniously suggests that exemplification will take different forms depending on the extent to which the lexical item to be exemplified is collocationally restricted. According to Hausmann (2005, 283), three cases may be distinguished: a) collocation-based items (“mot-collocation”), b) scenario-based items (“mot-scénario”), c) isotopy-based items (“mot-isotopie”). Although these categories clearly have fuzzy edges, Hausmann sets out to show that each category requires a specific treatment; he claims that a scenario-based word such as *badaud* cannot be exemplified simply by citing collocations, whereas a collocationally specific word such as *délabrement* cannot merely be illustrated by means of isotopy.

Criterion 4 takes account of the fact that the bilingual thesaurus is intended to assist learners in memorizing typical chunks of the target language. It also raises the question of the ratio of spoken to written examples. This is, of course, dependent on the purpose and target group of the dictionary; a general-purpose bilingual thesaurus aimed at advanced learners and designed for vocabulary learning should give preference to examples illustrating spoken usage or to ones which are neutral with regard to the spoken/written distinction. This is because speech is the primary faculty, but foreign language learners are normally exposed to writing rather than speech, so that the provision of spoken or neutral examples may go some way towards redressing the balance.

To my knowledge, there are no empirical studies available of the ways in which lexicographers arrive at their examples. For this reason the application of the three aforementioned procedures will now be illustrated with a few concrete examples, which are also intended to show that the interplay of principled procedures and criteria tends to yield better illustrative sentences than the somewhat haphazard policy still being followed in the making of monolingual learners' dictionaries.

Let us first compare some example sentences designed to illustrate the use of different parts of speech: one noun (*germ*), three verbs (*grovel*, *hijack* and *pronounce*), one adjective (*constructive*) and one idiom (*put one's foot in it*). The sentences are from four well-known learner's dictionaries (OALD in two different editions). Collocations which occur in example sentences have been underlined (for the comparison see table 4 on the following pages).

As already mentioned, Zöfgen claims that example sentences should always do more than merely illustrate collocations. This idea is born of wishful thinking. Owing to the pervasiveness of collocation, it is only natural that most example sentences in the dictionaries under review should be based on high-frequency collocations and colligations. Indeed, collocation-based exemplification is required because learners' dictionaries do not normally provide examples or patterns illustrating collocational entries (say, *spread germs*), although such collocations often exhibit specific patterning: V (*spread*) + (poss) + N (*germs*) + N + *to/among/around*). It is true, though, that lexical items of very low occurrence (e.g. *traffic policeman*) or high specificity (*grovel*) generally have a very restricted collocational range so that they can be illustrated more readily without recourse to collocations.

	OALD4	OALD7	CIDE1	Cobuild3	DAC
germ	disinfectant <u>kills germs</u> ; <u>germ warfare</u>	disinfectant <u>kills germs</u> ; dirty hands can be a <u>breeding ground for germs</u>	wash your hands so you don't <u>get germs on the food</u> ; germs can be <u>spread</u> by rats	chlorine is widely used to <u>kill germs</u> ; a germ that destroyed hundreds of millions of lives	Hop it! Don't <u>spread</u> your <u>germs</u> among us!; I was reading this article about <u>germ warfare</u> ; <u>appalling</u> , isn't it?; Brown bread contains the whole <u>wheat germ</u> .
grovel (show humility)	you will just have to <u>grovel</u> to the bank manager for a loan	–	I will <u>apologize</u> to him, but I won't <u>grovel</u> ; they felt obliged to write <u>grovelling</u> notes <u>apologising</u> for and <u>explaining</u> their <u>omissions</u>	I don't <u>grovel</u> to anybody; speakers have been <u>shouted</u> down, classes <u>disrupted</u> , teachers made to <u>grovel</u> ; a letter of <u>grovelling apology</u>	–
hijack	the <u>plane</u> was <u>hijacked</u> while on a flight to Delhi	the <u>plane</u> was <u>hijacked</u> by two armed men on a flight from London to Rome	A man armed with a <u>pistol</u> <u>hijacked</u> a <u>jet</u> that was travelling to Paris early today and demanded payment of \$125,000; 400 tons of food was stolen when a <u>convoy</u> of trucks was <u>hijacked</u> in the <u>famine-hit</u> town	two men tried to <u>hijack</u> a <u>plane</u> on a flight from Riga to Murmansk; a Chinese <u>airliner</u> was <u>hijacked</u> by two passengers and forced to fly to Taiwan; the <u>hijacked</u> <u>plane</u> <u>exploded</u> in a <u>ball of fire</u>	Report: "The number of <u>planes</u> that have been <u>hijacked</u> has decreased recently"; well, what would you do if someone <u>hijacked</u> your <u>plane</u> <u>into</u> Cuba?; <u>Hijack</u> a few <u>lorries</u> if you're that <u>hard up</u>

pronounce	people pronounce the word differently in this part of the country; how do you pronounce <i>phlegm</i> ?; The 'b' in 'debt' is not pronounced.	very few people can pronounce my name correctly; the 'b' in <i>lamb</i> is not pronounced	how do you pronounce your surname?; she pronounced his name so <u>badly</u> he didn't even recognise it; Sade, pronounced <i>shah-day</i> , is a singer	have I pronounced your name <u>correctly</u> ?; He pronounced it Per-sha, the way the English do	Look, don't pronounce <i>heat</i> and <i>bit</i> the same way!; the K in <i>know</i> isn't pronounced; do you understand Peter? – yes, he pronounces everything very clearly
constructive	criticism, proposals, remarks	criticism, suggestions, advice; his work helped hyperactive children to use their energy in a constructive way; the government is encouraging all parties to play a constructive role in the reform process	she criticised my writing, but in a way that was very constructive – I learned a lot from her; If you don't have anything constructive to say, I'd rather you kept quiet; he didn't offer any <u>constructive criticism</u> – just complained he didn't like it	she welcomes constructive criticism; after their meeting, both men described the talks as frank, friendly and constructive; the Prime Minister has promised that Israel will play a constructive role; at least I'm doing something constructive	I can take <u>constructive criticisms</u> , but I hate arguments.
put one's foot in it	–	I really put my foot in it <u>with</u> Ella; I didn't know she'd split up with Tom	I really put my foot in it by asking her when the baby was due and she wasn't even pregnant	our chairman has really put his foot in it, poor man, though he doesn't know it	–

Table 4. Comparison of example sentences in five different learner's dictionaries

With the exception of DAC, most dictionaries display a bias towards the written language (cf. *germ*, *hijack*, *constructive*). This is probably due to the nature of their corpus base. To compound matters further, sentences from different varieties and registers are usually placed side by side without any additional guidance for the learner. DAC alone stands out in explicitly marking one of its written examples by introducing it with a text type label (*report*), a practice that might well be adopted in future learners' dictionaries (see criterion 8 above). The bias towards writing in monolingual dictionaries is probably justifiable since they will typically be used for the purposes of writing and translation. What is perhaps more problematic is that the predominant source of illustrative material is newspaper language; it is doubtful whether learners will have frequent occasion to take over such material into their own text production. On the positive side, there are also a few examples of what has been described above as "neutral" style; such examples are eminently suitable for bilingual thesauri: *chlorine is widely used to kill germs; she welcomes constructive criticism*.

Some of the examples are a long way from reality because they state the obvious or carry too much meaning content. Thus, it is difficult to imagine a real-life context for such sentences as "disinfectant kills germs" (OALD 4 and 7). As Fox (1987, 143) notes, such examples work as isolated sentences, but cannot be fitted into extended texts. Yet minor alterations can produce a perfectly natural example: *this natural disinfectant kills germs (but doesn't leave behind a long-lasting residue)*.

One clear drawback of the monolingual dictionary is apparent from such entries as *constructive* and *put one's foot in it* in CIDE. The examples given at these entries might be referred to as "defining" or "explanatory"; the dictionary makers must have felt the definition to be lacking in concreteness and therefore made up examples that would clarify the meaning. Inevitably, this results in rather lengthy co-ordinate example sentences that, for reasons already explained, are usually unnecessary in bilingual learners' dictionaries: *she criticised my writing, but in a way that was very constructive – I learned a lot from her*.

Examples can provide invaluable guidance in cases where linguistic description is not sufficiently advanced. Thus, the predicative use of adjectives in copular constructions with human subjects (*je serai bref*) and in imperatives (*soyez constructif!*) seems to be more widespread in French than in English or German, but it is difficult to generalise in this area: ?*be constructive* -> (more commonly) *try to be constructive; je serai bref* → *I'll be brief / ich werde mich kurz fassen*. The best way to sharpen a learner's sense of such elusive differences is probably via exemplification coupled with translation.

Also noteworthy is the fact that "bigger" is not necessarily always "better". Thus, DAC, which has some 15,000 headwords, illustrates a colligational feature of the verb *hijack* (*hijack s.th. into [country, place]*) which goes unrecorded in the pattern information and in the examples section of the "big five" learners' dictionaries.

In what follows, the lexical unit *germ* (= bacillus) will be taken up to illustrate how the above procedures and criteria can be applied with a view to ferretting out “ideal” examples. The list below shows some of the corpus hits for *germ* found in my own archives of fictional and spoken English and in the spoken section of the BNC. It would seem to disprove Fox’s judgment (1987, 148) that real language does not contain “neat little sentences” that can be detached from context for purposes of exemplification; this judgment is probably due to the written bias of the initial Cobuild corpus.

---

#### Fiction and spoken English

---

we have to share a hotel room that probably has more germs than an eighteenth-century brothel  
 we should wash the stuff before we use it; germs and all  
 what do you get when you fall in love? You get enough germs to catch pneumonia.  
 The food in Ma Jo’s kitchen was cooked until it was limp, in case a few germs remained.  
 I don’t know why everyone’s afraid of a few germs.  
 School water. They put chlorine in there to kill germs.  
 Try this. It kills all known germs.  
 Do you know what kind of germs you can pick up on a beach?  
 I don’t want to catch your germs.  
 You’re not helping my chances by getting your germs all over my hand.  
 the troops contracted a germ indigenous to the Gulf  
 grease draws dirt, dirt draws germs and germs kill people  
 we wish her well, as long as she keeps her germs to herself  
 germs that float about in the air  
 it kills all known airborne germs sort of thing  
 to take away all the nasty to kill the germs  
 [...] mask the smell but it’s also basically to kill the germs  
 a germ, there’s a germ that floats about in the air and it gets onto [...]  
 you don’t want to tip it out on there cause you get germs, darling

(pragmatic prosody: people are concerned about germs and their effect on their health)

---

#### Academic English

---

the cause of typhoid fever is a germ known by the name of bacillus typhosus  
 this disease is caused by a germ, gonococcus  
 The poor had to deal with charity hospitals. If admitted to these institutions, they were housed and fed, but they also shared beds and germs with all the other diseased patients in their wards, and often received little medical help.  
 viruses and germs  
 poisons and germs  
 disease germs  
 the fire had destroyed much filth, many rats, fleas and germs; London had no further plagues

(pragmatic prosody: germs as disease-causing agents)

---

Table 5. Corpus hits for *germ* in the BNC and my own archives of fictional and spoken English

Rather than just select examples, the lexicographer should attempt to gain an overall picture of an item's semantic or pragmatic prosody. It appears from the available data that *germ* is imbued with two closely connected prosodies. In spoken English, *germ* typically occurs in environments where there is a preoccupation with the effect of germs on health: germs kill and should therefore themselves be destroyed. In academic English, the prevalent pragmatic prosody is to do with germs as disease-causing agents. Even if the lexicographer wishes to separate examples of one-word entries and multi-word entries, the typical pragmatic prosody should still be captured in the illustrative material; the following sentences are likely candidates:

- (5) The food in Ma Jo's kitchen was cooked until it was limp, in case a few germs remained.
- (6) We should wash the stuff before we use it, germs and all.

The second example could be taken over as it stands. The first sentence could be shortened somewhat in accordance with criterion 4 to give the following "neutral" example:

- (5') The food was cooked until it was limp in case a few germs remained.

Some of the other examples can be adopted as illustrations of particular collocations; most of these illustrate spoken usage:

- (7) (collocation: *get germs*) you don't want to tip it on there cause you get germs, darling; you might get the germs on your fingers
- (8) (collocation: *catch germs*) I don't want to catch your germs.
- (9) (collocation: *kill germs*) School water. They put chlorine in there to kill the germs.

The lexicographer thus ends up with the article shown in the appendix.

## 5. Conclusion

This article has proposed a number of stringent principles which can be applied to the selection and treatment of examples in bilingual thesauri. Most of these principles apply equally to monolingual learner lexicography. It is to be hoped that their application will lead to a stricter distinction between collocations and examples, and that the wording and memorability of examples will be improved.

## Works Cited

### Dictionaries

- A Dictionary of English Collocations* (1994). Göran Kjellmer, ed. Oxford: Clarendon Press.
- A Dictionary of English Style* (<sup>2</sup>1955). Albrecht Reum, ed. Leverkusen: Gottschalksche Verlagsbuchhandlung. [DES]
- A Valency Dictionary of English* (2004). Thomas Herbst, David Heath, Ian F. Roe and Dieter Götz, eds. Berlin/New York: Mouton de Gruyter. [VDE]
- Cambridge International Dictionary of English* (1995). Paul Procter, ed. Cambridge/New York/Melbourne: Cambridge University Press.
- Collins Cobuild English Collocations on CD-ROM* (1995). John Sinclair, ed. London: Harper Collins.
- Collins Cobuild English Dictionary for Advanced Learners* (<sup>3</sup>2001). John Sinclair, ed. Glasgow: Harper Collins.
- Dictionnaire de l'Anglais Contemporain* (1980). Françoise Dubois-Charlier, ed. Paris: Larousse. [DAC]
- Dictionnaire du Français Contemporain* (1966). Jean Dubois, ed. Paris: Larousse. [DFC]
- Dictionnaire Explicatif et Combinatoire du Français Contemporain* (1984-1999). Igor Mel'čuk, ed. Montreal: Montreal University Press. [DEC]
- Langenscheidts Kontextwörterbuch Französisch-Deutsch. Ein neues Wörterbuch zum Schreiben, Lernen, Formulieren* (1989). Peter Ilgenfritz et al., eds. München: Langenscheidt. [LKF]
- Oxford Advanced Learner's Dictionary* (<sup>4</sup>1989). Albert S. Hornby; Anthony P. Cowie, ed. Oxford: Oxford University Press. [OALD4]
- Oxford Advanced Learner's Dictionary* (<sup>7</sup>2005). Albert S. Hornby; Sally Wehmeier, ed. Oxford: Oxford University Press. [OALD7]
- Oxford Collocations Dictionary for Students of English* (2002). Jonathan Crowther, Sheila Dignen and Diana Lea, eds. Oxford: Oxford University Press. [OCD]
- Petit Dictionnaire de Style à l'Usage des Allemands* (1941). Albrecht Reum, ed. Leipzig: J.J. Weber. [PDS]

### Other Sources

- Benson, Morton (1986). *Lexicographic Description of English*. Amsterdam: Benjamins.
- Biber, Douglas, Susan Conrad, Geoffrey Leech, Stig Johansson and Edward Finegan (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- Blumenthal, Peter (2005). "Le Dictionnaire des Collocations – un Simple Dictionnaire d'Exemples?" Michaela Heinz, ed. *L'Exemple Lexicographique dans les Dictionnaires Français Contemporains – Actes des «Premières Journées*

- Allemandes des Dictionnaires»* (Klingenberg am Main, 25-27 juin 2004), Tübingen: Niemeyer, 265-282.
- Carter, Ronald and Michael McCarthy (1997). *Exploring Spoken English*. Cambridge: Cambridge University Press.
- Drysdale, Patrick D. (1987). "The Role of Examples in a Learner's Dictionary." Anthony P. Cowie, ed. *The Dictionary and the Language Learner*. Tübingen: Niemeyer.
- Feilke, Helmuth (1996). *Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik*. Frankfurt, Main: Suhrkamp.
- (2004). "Kontext – Zeichen – Kompetenz. Wortverbindungen unter sprachtheoretischem Aspekt." Kathrin Steyer, ed. *Wortverbindungen – mehr oder weniger fest*. Berlin et al.: de Gruyter, 41-64.
- Fillmore, Charles J. (1976). "Pragmatics and the Description of Discourse." Siegfried J. Schmidt, ed. *Pragmatik/Pragmatics II. Grundlegung einer expliziten Pragmatik*. München: Fink, 83-104.
- Fox, Gwyneth (1987). "The Case for Examples." John Sinclair, ed. *Looking up. An Account of the COBUILD Project in Lexical Computing*. London: Collins ELT.
- Gallagher, John D. (2006). "Foreign Language Vocabulary Learning." Dirk Siepmann, ed. *Wortschatz und Fremdsprachenlernen*. Landau: Verlag Empirische Pädagogik, 11-96.
- González-Rey, Isabel (2002). *La Phraséologie du Français*. Toulouse: Presses Universitaires du Mirail.
- Hausmann, Franz J. (1977). *Einführung in die Benutzung der neufranzösischen Wörterbücher*. Tübingen: Niemeyer.
- (2003). "Was sind eigentlich Kollokationen?" Kathrin Steyer, ed. *Wortverbindungen – mehr oder weniger fest*. Berlin: Mouton de Gruyter, 309-334.
- (2005). "Isotopie, Scénario, Collocation et Exemple Lexicographique." Michaela Heinz, ed. *L'Exemple Lexicographique dans les Dictionnaires Français Contemporains. Actes des «Premières Journées Allemandes des Dictionnaires»*. Tübingen: Niemeyer, 283-292.
- Herbst, Thomas (1996). "What are Collocations – Sandy Beaches or False Teeth?" *English Studies* 77/4, 379-393.
- Hill, Jimmie (2000). "Revising Priorities: From Grammatical Failure to Collocational Success." Michael Lewis, ed. *Teaching Collocation. Further Developments in the Lexical Approach*. Hove, LTP, 47-69.
- Hoey, Michael (2005). *Lexical Priming: A New Theory of Words and Language*. London: Routledge.
- Hohmann, Heinz-Otto (1999). "Ausbau der Sprachkompetenz durch kontextuelle Wortschatzarbeit." *Fremdsprachenunterricht* 6, 409-412.
- Klotz, Michael (2000). *Grammatik und Lexik: Studien zur Syntagmatik englischer Verben*. Tübingen: Stauffenburg.
- Kubczak, Jacqueline and Sylvie Costantino (1998). "Exemplarische Untersuchungen für ein syntagmatisches Wörterbuch Deutsch-Französisch / Fran-

- zösisch-Deutsch." Daniel Bresson and Jacqueline Kubczak, eds. *Abstrakte Nomina. Vorarbeiten zu ihrer Erfassung in einem zweisprachigen syntagmatischen Wörterbuch*. Tübingen: Narr, 11-120.
- Lamb, Sidney M. (1999). *Pathways of the Brain: The Neurocognitive Basis of Language*. Amsterdam: Benjamins.
- Landau, Sidney I. (2001). *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- Louw, Bill (1993). "Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies." Mona Baker, Gill Francis and Elena Tognini-Bonelli, eds. *Text and Technology: In Honour of John Sinclair*. Amsterdam: Benjamins, 157-176.
- Martin, Robert (1989). "L'Exemple Lexicographique dans le Dictionnaire Monolingue." Franz Josef Hausmann, Oskar Reichmann, Herbert E. Wiegand and Ladislav Zgusta, eds. *Wörterbücher: Ein internationales Handbuch zur Lexikographie*. Vol. 1. Berlin: Mouton de Gruyter.
- Mel'čuk, Igor (1998). "Collocations and Lexical Functions." Anthony Cowie, ed. *Phraseology. Theory, Analysis and Applications*. Oxford: Clarendon Press, 23-53.
- Métrich, René (1993). *Lexicographie Bilingue des Particules Illocutoires de l'Allemand*. Göppingen: Kümmerle.
- Mugdan, Joachim (1985). "Pläne für ein grammatisches Wörterbuch. Ein Werkstattbericht." Henning Bergenholtz and Joachim Mugdan, eds. *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28.-30.6.1984*. Tübingen: Niemeyer, 187-224.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
- Siepmann, Dirk (2005a). *Discourse Markers across Languages. A Contrastive Study of Second-level Discourse Markers in Native and Non-native Text with Implications for General and Pedagogic Lexicography*. New York: Routledge.
- (2005b). "Collocation, Colligation and Encoding Dictionaries. Part I: Lexicological Aspects." *International Journal of Lexicography* 18/4, 409-444.
- (2006). "Collocation, Colligation and Encoding Dictionaries. Part II: Lexicographical Aspects." *International Journal of Lexicography* 19/1, 1-39.
- (forthcoming). "Routine Formulae in Learners' Dictionaries. What, Where and How?" Christelle Cosme, Céline Gouverneur and Fanny Meunier, eds. *Phraseology 2005. The Many Faces of Phraseology. An Interdisciplinary Conference*. Amsterdam: Benjamins.
- Sinclair, John (1984). "Naturalness in Language." Jan Aarts and Willem Meijs, eds. *Corpus Linguistics – Recent Developments in the Use of Computer Corpora in English Language Research*. Amsterdam: Rodopi, 203-210.
- (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Svartvik, Jan (1988). "On the Study of Spoken English". Wolfgang Kühlwein and Bernd Spillner, eds. *Sprache und Individuum*. Tübingen: Gunter Narr Verlag, 30-37.

- Zöfgen, Ekkehard (1986) "Kollokation – Kontextualisierung – (Beleg)Satz. Anmerkungen zur Theorie und Praxis des lexikographischen Beispiels." Albert Barrera-Vidal, Hartmut Kleineidam and Manfred Raupach, eds. *Französische Sprachlehre und bon usage. Festschrift für Hans-Wilhelm Klein zum 75. Geburtstag*. München: Hueber, 219–238.
- (1994). *Lernerwörterbücher in Theorie und Praxis. Ein Beitrag zur Metalexikographie mit besonderer Berücksichtigung des Französischen*. Tübingen: Niemeyer.

## Appendix: Sample article

	<b>the germ</b>	<b>der Krankheitskeim (-e) / der Keim (-e) / der Bazillus (-en) / die Bakterie (-n) / der Erreger (-)</b>	
N	<i>We should wash the stuff before we use it, germs and all.</i>	<i>Wir sollten das abwaschen, bevor wir es benutzen, wegen der Bakterien und so.</i>	
	<i>The food was cooked until it was limp in case a few germs remained.</i>	<i>Das Essen wurde so lange gekocht, bis es ganz weich war, aus Angst, es könnten noch ein paar Krankheitserreger drin sein.</i>	
ADJ + N	<b>an airborne germ</b>	<b>ein in der Luft befindlicher Keim</b>	syn: ein durch die Luft übertragbarer Bazillus
syn: <i>a disease-causing agent, a pathogene</i>	<b>a disease-causing germ</b>	<b>ein krankheitserregender Keim</b>	syn: ein Krankheitserreger, ein krankmachender Keim, ein gesundheitsschädlicher Keim, ein gesundheitsgefährdender Keim
syn: <i>a dangerous germ</i>	<b>a nasty germ</b>	<b>ein gefährlicher Keim</b>	syn: ein übler Bazillus
	<b>a resistant germ</b>	<b>ein resistenter Keim</b>	syn: ein widerstandsfähiger Keim
	<b>a virulent germ</b>	<b>ein bösartiger Keim</b>	
N + N: types of germs	<b>a cholera / flu (etc.) germ</b>	<b>ein Cholera- / Grippe- / (usw.) bazillus</b>	
N + V	<b>germs multiply</b>	<b>Keime vermehren sich</b>	
N + V	<b>germs spread</b>	<b>Keime verbreiten sich</b>	
N + V (+ ADV); syn: <i>germs spread (somewhere)</i>	<b>germs accumulate (somewhere)</b>	<b>Keime siedeln sich (irgendwo) an</b>	
N + V + into/on/onto (etc.) N syn: <i>enter s.th. / be introduced into s.th.</i>	<b>germs get somewhere</b>	<b>Keime gelangen irgendwo hin</b>	syn: Keime dringen irgendwo ein
	<i>The germs will get into the food and give you an illness.</i>	<i>Die Keime gelangen in die Nahrung und machen uns krank.</i>	
N + V + (often poss) N; syn: <i>pick up germs</i>	<b>s.o. catches germs</b>	<b>j-m fängt sich einen Bazillus ein</b>	syn: j-m steckt sich mit einem Bazillus an
	<i>Hop it. I don't want to catch your germs.</i>	<i>Hau ab. Ich will mir nicht deine Bazillen einfangen.</i>	

N + V + N + on/onto/ all over (etc.) N	<b>s.o. gets germs somewhere</b>	<b>Keime gelangen irgendwo hin</b>	syn: Keime geraten irgendwo hin
	<i>You might get the germs on your fingers.</i>	<i>Die Bazillen könnten auf Ihre Hände gelangen.</i>	syn: Die Bazillen könnten auf Ihre Hände kommen.
N + V + N: only if clear from context	<i>You don't want to tip it on there cause you get germs, darling.</i>	<i>Da solltest Du es besser nicht hinschütten, Schatz, sonst entstehen Bakterien.</i>	
N + V + (often poss) N + around; to s.o. / among people	<b>s.o. spreads germs (somewhere / among people)</b>	<b>j-m verbreitet Keime (irgendwo / unter Leuten)</b>	V + (oft poss) N + unter Leuten
syn: <i>keep your germs to yourself</i>	<i>Don't spread your germs around!</i>	<i>Behalte deine Bazillen für dich!</i>	
	<i>Do you want me to spread my germs among the rest of the staff?</i>	<i>möchten Sie, dass ich meine Bazillen an die anderen Mitarbeiter weitergebe?</i>	
N + V + N; syn: <i>destroy germs</i>	<b>s.th. kills germs</b>	<b>etw. tötet Keime ab</b>	syn: etw. erstickt Keime
	<i>School water. They put chlorine in there to kill the germs.</i>	<i>Trinkwasser für Schulen. Man setzt ihm Chlor zu, um die Keime zu töten.</i>	
	<i>It kills all known germs.</i>	<i>Es tötet alle bekannten Keime ab.</i>	
	<i>The Great Fire had destroyed much filth, many rats, fleas and germs; London had no further plagues.</i>	<i>Durch den großen Brand waren viel Schmutz sowie viele Ratten, Flöhe und Bazillen beseitigt worden. Danach trat die Pest in London nicht mehr auf.</i>	
N + V + N; syn: <i>carry germs</i>	<b>s.o. / s.th. harbours germs</b>	<b>etw. enthält Keime</b>	
-> a breeding ground for germs	<i>Dirt harbours germs.</i>	<i>Schmutz ist eine Brutstätte für Keime.</i>	
	<i>Puppies and kittens may harbour the germ.</i>	<i>In Welpen und Kätzchen kann der Keim überleben.</i>	syn: Welpen und Kätzchen können den Keim in sich tragen
	<b>s.o. / s.th. passes on germs</b>		
	<i>It's quite easy for simians to pass on germs and viruses to humans.</i>	<i>Affen können sehr leicht Keime und Viren auf Menschen übertragen</i>	
phrases; -> s.o. / s.th. harbours germs	<b>a breeding ground for germs</b>	<b>eine Brutstätte für Bakterien</b>	