Working with Words: Research Approaches to Translation-Oriented Lexicographic Practice
Travailler avec les mots : vers une pratique lexicographique à l’usage de la traduction

Maribel Tercedor, Clara Inés López-Rodríguez and Pamela Faber

Dictionaries ideally should address the needs of particular types of users. Their micro and macrostructural design should be oriented towards what user groups need to know about words and the uses that will be made of such knowledge. One specific use for dictionaries is the activity of translation. From the perspective of professional translators, dictionaries should allow for creativity and dynamicity in text production, providing solutions for changing communication needs. Dictionary-making for such a purpose can and should benefit from insights into words and meanings from other fields. Interdisciplinarity in Lexicography is just one example of how other fields interact in Translation Studies. In this paper we analyse how working in an interdisciplinary way is crucial to developing useful lexicographic and terminographic tools for translators and how methodologies, such as corpus-based work and experimental methods should be combined to offer converging evidence of different aspects of use and processing. We illustrate such work methods with examples from real lexicographic projects.
Working with Words: Research Approaches to Translation-Oriented Lexicographic Practice

Maribel Tercedor
Clara Inés López-Rodríguez
Pamela Faber

Introduction

Interdisciplinarity is at the core of translation research and practice. The main reason for this is that translation is a multitask activity in which attention is focused on different cognitive subactivities related to transferring and mapping the content and nuances of the source onto the target. Translation research is generally divided into process-oriented studies and product-based investigations, and important empirical data can be obtained from both approaches (Alves et al., 2010). From the perspective of professional translation, dictionaries should allow for a wide range of user needs related to the source text (e.g., full understanding of word meanings and uses) as well as to the target text (i.e., creativity and dynamicity in the text production stage). It therefore follows that dictionary making as an area of Translation Studies can benefit from insights from other fields that are relevant to the study of the translation activity, the process of translating, and the creation of translation products.

1 This research is funded by a grant received from the Spanish Ministry of Science and Innovation for the projects Varimed (FFI2011-23120) and Record (FFI2011-22397) of the Lexicon Research group.
This paper describes how the research on lexical semantics and lexical representation from different fields can cross-fertilize lexicographic practice that targets translators as a user group. More specifically, we describe how process-oriented methodologies, such as experimental methods in psycholinguistics and cognitive science, can offer important insights into dictionary making. Furthermore, product-based research, such as corpus-based work, can provide crucial data for the elaboration of lexical and terminological resources. Moreover, empirical methods for studying translation as a product (i.e., corpus-based research) can provide authentic data on language use in context, and facilitate lexicographic tasks such as sense disambiguation, the selection of sentences to illustrate word use, the detection of neologisms, among other things. This article focuses on semantic prosody and collocations as a case in point.

The following sections highlight converging evidence from both approaches and stress certain research problems and issues through examples from lexicographic tools. Finally, world knowledge, frequency, and familiarity are analyzed as important factors to be taken into account in the making of lexicographic and terminographic resources.

1. Cognitive and Functional Uses of Dictionaries

It is generally agreed that an ideal dictionary for translators should provide information not only about the meaning of words (core meaning, and peripheral meaning), but about how words combine with other words and morphological elements (combinatory and derivational potential), and how they are used and activated in particular texts and contexts (use in particular genres, registers, dialects, links with other words in associative networks, and so on). However, when we look at what language in use means, the consensus is not so clear. It is necessary to consider both the users of dictionaries, and the use of dictionaries. Regarding the use of dictionaries by different groups, Jääskeläinen (1989) found that novice translators look up more units in the dictionary, whereas advanced students use more dictionaries per problematic unit.
When using dictionaries to translate, the translator must consider a wide variety of information regarding the words in the source and the target text, such as their core meaning, peripheral meaning and metaphorical extensions, position in associative networks, derivational and combinatorial potential, geographical uses, and use in particular genres and levels of expertise, among other things.

According to L’Homme and Leroyer (2009), functional approaches to lexicography should distinguish between cognitive functions, on the one hand, and communicative functions, on the other. In the first case, needs usually include the acquisition of encyclopaedic knowledge, specialized language units, cultural references, or general subject field knowledge. Communicative needs, on the other hand, are related to textual activities such as revision, reading, and translating. The authors, following Bergenholz (2005), argue that the consultation scheme is radically different, and they summarise it in the following way:

**COGNITIVE USE:** user → dictionary → user
**FUNCTIONAL USE:** user → text segment y → user → dictionary → user → text segment x (p. 20)

However, boundaries between both uses are rarely clear. In order for dictionaries to be suited to particular types of users, their micro and macrostructural design should be oriented towards the cognitive-functional uses that particular user groups make of dictionaries. Evidently, making a dictionary for translators involves describing the meaning of words, their use in context, and their possible correspondences in other languages. It also entails making their position explicit (at least at some level) in the configuration of the mental lexicon. This involves considering cognitive and functional criteria in a continuum, since the concepts of situation (as a set of knowledge acquisition needs) and linguistic context are intertwined.

The meaning of a word includes not only semantic information, but also information about its syntactic environment as well as the pragmatic parameters that influence its activation in different contexts. The meaning of a word is determined by its links with other words because the mental lexicon is a vast network,
whose organization accounts both for the existence of *signified* and *signifier*, and for the processes of word comprehension and word production, as shown by a wide range of psycholinguistic experiments (Aitchison, 1987).

Firstly, words with similar phonological or graphological representations (*signifier*) are stored together in order to facilitate their recognition both in oral and written communication. The principle underlying this type of organization of the lexicon has been partially reproduced in semasiological dictionaries, where words are arranged alphabetically. Psycholinguistic experiments have shown that words with similar beginnings, similar endings, and/or similar rhythm are likely to be tightly bonded and have labelled this phenomenon as the “bathtub effect.” For instance, it is very likely that a translator who is proficient in English stores the following terms in close proximity to each other so as to retrieve them quicker (Figure 1): **Commonwealth**—**common** law—**Commons** (The House of **Commons**).

Secondly, words are also stored according to their meaning (*signified*). Words with related meanings are stored together in the same lexical domain and are closely linked with others of a similar meaning while retaining looser links with words outside their domain. This organization of the lexicon is represented in onomasiological dictionaries, where the content is organized semantically, and therefore, users access words from concepts. Thesauri are good examples of onomasiological dictionaries.

However, regarding the mental lexicon, the boundary between language-dependent modules and knowledge of the world is far from clear cut. In fact, as cognitive science has proved, our interaction with the world is situated and embodied (Barsalou, 2003, 2005; Faber, 2011, 2012; Gibbs, 2006). As a result, situatedness models our language, pointing to a shared space between language and cognition that is a reflex of our dynamic representation of knowledge and language. In view of this dual organization of the mental lexicon and of the indivisibility of cognition and language, a dictionary for translators should be organized both onomasiologically and semasiologically.
Moreover, a dictionary for translators should reflect, to the extent possible, the same organization as the mental lexicon. This means that the presentation of information should include some sort of network representation, displaying common types of links such as co-ordination, collocation, and superordination, as well as more complex semantic relations with other words (such as location, has_function, made_of, etc.). The task is far from simple, since concepts are multidimensional in that their meaning can vary when the focus and perspective vary. Since this can even lead to different designations, information regarding different situated perspectives should be provided in the resource. This feature should guide the creation of dictionaries for translators, who frequently are obliged to search for a concept that they do not know how to lexicalise.

2. Studying Lexical Processes: What Experimental Methods Can Reveal About the Lexicon

The lexicon has been a research focus in both Cognitive Science and Linguistics, and thus interest in lexical meaning, organization, and representation is shared by areas such as Translation Studies,
Psycholinguistics, Second Language Acquisition, NLP, and Lexicography, to name a few. The lexicon is a problematic area in acquisition, and hence in translation. According to Laufer (1986), adults make four errors in the lexicon for each error in grammar.

Because translation involves analysing, processing, and producing texts, the notion of language in use, and the use of dictionaries should go hand in hand with these cognitive processes. Methods of empirical research from cognitive science and the social sciences, such as think-aloud protocols, semantic priming, eye tracking, ERPs (Event Related Brain Potentials) and fMRI (functional Magnetic Resonance Imaging) can shed light on the actual processing and production of lexical units, thus providing valuable data for the process of making lexical tools. By using such methods, it is possible to infer problems in dictionary use during the translation process. This provides valuable data related to cognitive processes such as the following: (1) how a word facilitates access to a semantically related word; (2) where attention is focused when we read a word in context, and; (3) which areas of the brain fire when words and multimodal stimuli are connected. The objective is to discover whether syntactic tags, definitions, and visual information actually help translators connect the information provided in dictionaries with their previous knowledge about the world.

The problem with these methods, as used traditionally in Psycholinguistics, is that they are not easy to apply to Translation Studies, due to their lack of ecological validity, which is always an issue in experimental studies (Göpferich and Jääskeläinen, 2009, p. 182). In fact, knowing how individuals behave in laboratories does not necessarily offer insights into how translators work when they are confronted with a text. Since ecological validity is evidently a prerequisite to the analysis of their lexical needs, the real challenge is to provide these methods with this type of validity so that they can replicate the environment of the translation practice as well as the actual texts that translators must work with in their professional environment.
2.1. Perception and Processing of Lexical Units

Reading aloud and reading silently in the context of lexical decision experiments determine the reader’s difficulty in processing a word. One of the psycholinguistic methods that Translation Studies is beginning to use is eye-tracking to gain insights into the fixation durations of words in particular contexts (Göpferich et al., 2008). The gaze duration (i.e., total fixation on a word before moving to another word) is taken to be a measure of difficulty in processing a word. However, other cognitive processes such as situated conceptualizations (Barsalou, 2005, p. 626) are crucial to processing and production tasks alike and are not accessible through eye-tracking monitoring. There are many cognitive processes not verbalized or included in semantic memory which are important for the lexicon. The meaning of a word includes its sensory aspects, its function and the embodied representation of the concept. In other words, word meaning reflects the situated way in which we make sense of the world around us. In order to access situated conceptualization processes through perceptual input as well as other salient features, the methods used are the elicitation of concept properties, imagery, as well as data generated by ERPs and fMRIs.

For example, it is possible to infer the types of sensory information that a word activates by identifying the brain areas that fire when the subject reads a word, perceives an image of what it designates, or even imagines it. These insights are relevant for lexical decision and definition strategies as well as for user perceptions of words in context. For instance, the meaning of a word includes its sensory aspects, its function, and the embodied representation of the concept that it lexicalizes. It goes without saying that such meaning is of a dynamic nature, and that it should reflect the situated way in which we make sense of the world around us. The dynamic multimodal information activated by a concept is illustrated in the following advertisement in Figure 2.
The fun never stops [EMOTION] with this giant 88” circumference [SIZE AND SHAPE] sprinkler that splashes water as you roll it around! [FUNCTION AND EMBODIED ACTION] Water sprays out of four fun spray heads [SHAPE]. This cool sprinkler [EMOTION] combines fast and fun beach ball action [EMOTION AND MOVEMENT] with wild water play [MEDIA]. Just inflate it, hook it up to a hose [USER INTERACTION] and watch the fun begin! [EMOTION]

Figure 2. Dynamic multimodal information in an advertisement

Although the persuasive nature of the advertisement leads the sender to foreground emotion-laden information, the explanation of the artefact is composed of sensory and functional information retrieved through the perceptual input and manipulation of the object. This information is not incidental, since meaning construction heavily relies on perceptually simulating the information that is presented to the comprehender, hence the importance of perception-based information for defining concepts in a situated manner.

In the definition of a specialized concept, the IS_A relation first situates it within a conceptual hierarchy, and then other information such as perceptual features, location and function are added:

bone marrow

soft and spongy [TOUCH_OF] substance [IS_A] in the centre of bones [LOCATION_OF] that produces blood cells, particularly red cells and platelets [FUNCTION].

2.2. Dynamicity and the Lexicon

As previously mentioned, lexicographic and terminographic tools should account for the dynamic way in which we conceptualize the world around us. Accordingly, the increasing emphasis
on dynamicity in Terminology has led to the proposal and elaboration of more flexible and dynamic specialized knowledge representation models that are better capable of managing and integrating information from different sources, and adapting terminological information to user needs (Faber, 2011). According to Wu and Barsalou (2009), a combination of word association and situated simulation is responsible for generating properties for concepts. In other words, dynamic representations and contextual uses of a word provide us with the necessary information about that word.

In Terminology, dynamicity is intimately related to multidimensionality (Bowker and Meyer, 1993). Since concepts can be viewed or represented from different perspectives or dimensions, lexicographic practice should thus involve the elaboration of more dynamic representations.

The dynamic representation of concepts in dictionaries involves providing information regarding terminological variation as a cognitive, situated phenomenon (Tercedor, 2011). Variation is related to the different ways of naming the same concept (Daille et al., 1996). For example, the Medline database provides the following terminological variants in Spanish for “malaria” under the heading “Nombres alternativos” [“Other names”]: fiebre cuartana; Paludismo o malaria por Plasmodium falciparum; Fiebre biduoterciana; Paludismo terciano; Plasmodio; Fiebre de las aguas negras o de los pantanos [Quartan malaria; Falciparum malaria; Biduotarian fever, Blackwater fever; Tertian malaria; Plasmodium]. Each of these variants has a motivation and the translator must be able to access information regarding the most appropriate use of each one in a particular context or situation. When lexicographic/terminographic tools offer such information, this makes the work of the translator considerably easier. In order to elicit terminological variants experimentally, crossmodal production tasks can be used in combination with other observational methods, such as extracting corpus-attested forms (Tercedor, 2011). For example, an image of a concept can be used to activate different terminological variants and then triangulate this data with corpus forms.
In Frame-based Terminology (Faber, Márquez and Vega, 2005; Faber et al., 2006; Faber, Leon, Prieto and Reimerink, 2007), conceptual networks are based on an underlying domain event as well as a closed inventory of hierarchical or vertical (i.e., IS_A, PART_OF, and so on) and non-hierarchical or horizontal (HAS_FUNCTION, LOCATION, IS_MEASURED_WITH) semantic relations. The EcoLexicon environmental knowledge base (http://manila.ugr.es/visual) was elaborated on the premises of dynamicity and multidimensionality. This terminological resource focuses on conceptual relations and the combinatorial potential of concepts, based on information extracted from corpus analysis. This prototypical domain event or action-environment interface (Barsalou, 2003) provides a template applicable to all levels of information structuring.

Figure 3. Event types lexicalized as term entries in the EcoLexicon database
The visualization and representation of event types affords a more realistic picture of the types of objects that each action type allows as well as the type of event that the frame activates. For example, in Figure 4, through the definition of recharge (process by which new water is added to an aquifer or to the zone of saturation), the user learns that recharge takes an aquifer or the zone of saturation as the affected entities; that it is a type of action that entails a telic process through which the water level is affected (accomplishment); and that the action takes human or artificial agents as the doers of the event.

**Figure 4. Activation of the recharge frame in the EcoLexicon database**

This type of representation is coherent with how knowledge is dynamically processed in a situated context. According to Barsalou (2005), a given concept produces many different situated conceptualizations, each tailored to different
instances in different settings. Thus, context can be said to be a dynamic construct that triggers or restricts knowledge. For instance, the general event codifying an extremely general (and multidimensional) concept such as water can be recontextualized at any moment to centre on any of the more specific subevents. For example, when the water representation is recontextualized to focus on waste water (Figure 5), it takes the following form.

![Figure 5. Representation of waste water in EcoLexicon](image)

Such information is obtained from the combination of top-down processes of encyclopedic work with bottom-up methodologies using corpus-based collocation extraction. Conceptual relations, both horizontal and vertical, are the selection criteria for concordances in the EcoLexicon database (Faber et al., 2005; Tercedor and López, 2008).
3. Product-Based Research: Corpora, Collocational Meaning, Dialectal Variation and Semantic Prosodies

In WordSmith Tools, one of the most widely used lexical analysis programs, Mike Scott inserts short quotations to make users reflect upon the uses and limitations of corpus studies. One of these quotes states that much can be inferred from what is absent. This is undeniably true because corpora reveal both explicit and implicit information about the cognitive and functional uses of words. Studying a word in a corpus helps translators to learn the core definition of the word as well as its different senses. It also allows them to grasp conceptual information about the subject field in which the word is generally used. Moreover, corpora can help translators to learn the uses of the word in context. This includes acquiring knowledge about the following: (1) word combinations in collocations and syntactic structures; (2) lexical variation, depending on the speaker and the context of situation; (3) conventional and novel uses of the word, and; (4) intentionality motivating word use. Therefore, the study and analysis of language in use and in real translations is a valid methodology within Descriptive Translation Studies.

3.1. The Role of Corpora and Lexicographic Materials in Lexical Activation

Since the 1980s the use of corpora has become increasingly important in dictionary projects, beginning with Sinclair’s Birmingham Collection of English Text, which resulted in the English learner’s dictionary, Collins COBUILD. Most dictionaries are now compiled on the basis of instances of “authentic” language in use provided by corpora. The intuition of the lexicographer is essential in selecting examples and configuring the information. However, the final output available to dictionary users does not always include such corpus evidence. The entries in certain resources do not even include one sentence illustrating the use of the word in context. This is the case for one of the most widely used dictionaries in the Spanish language, the Diccionario de la Real Academia Española, (DRAE) [Dictionary of the Royal Academy of the Spanish Language]. Despite the fact that the Spanish Royal Academy provides free access to a corpus of
contemporary Spanish of more than 200 million words (CREA corpus), there is no direct link from the on-line dictionary to the on-line corpus, and the dictionary does not include any examples of actual word use. Moreover, corpus-based dictionaries, such as the Diccionario Clave, which does display examples, are not widely known amongst translators working with Spanish.

Figure 6 highlights the information provided by the previously mentioned resources about the verb amarrar [to moor], included in a text on oceanographic navigation. The use of examples clearly facilitates comprehension and the translation process by activating concepts and words functioning as prototypical subjects, objects, and adverbials.
Considering the fact that many translation students and translators worldwide usually search words in free on-line resources, which are generally abridged versions of dictionaries (full access must be paid for), they only have very limited (if any) access to information about the collocational meaning of words. In addition, translating under the pressure of tight deadlines is such a common practice among translators that they frequently do not have the time to search free on-line corpora, such the Collins Wordbanks Online English corpus, the Corpus of Contemporary American English or the British National Corpus. In fact, only a few are even aware of innovative research projects using Internet as a corpus in itself. On the brighter side, their pervasive use of Google provides them with both micro-contexts of language use and images that help them understand the core meaning of a word. Such textual, lexical, and visual activation helps them find candidates for translation equivalents. When translating, collocational meaning activated by the use of words in context, combines with the information and connotations evoked by images, together with presuppositions, knowledge, and mental schemata of the translator. All of these have a direct effect on the translation process, where creativity is essential (Kenny, 2001; Stewart, 2000).

Therefore, Internet and freely available on-line resources such as reference corpora, websites offering not only dictionaries, but also feedback from language forums, and social networks have transformed the role of dictionaries in the translation process to the extent that it is now more appropriate to talk about lexicographic materials rather than dictionaries.

Accordingly, lexicographic practice in the making of dictionaries for translators should take into account the search patterns and habits of translators, and the benefits they obtain from conceptual and lexical activation through images and contexts

2 “Collocative meaning consists of the associations a word acquires on account of the meanings of words which tend to occur in its environment” (Leech, 1974, p. 20).

3 Projects such as WebCorp (http://www.webcorp.org.uk) or WebAsCorpus.org (http://webascorpus.org/searchwc.html).
(Tercedor, López and Robinson, 2005; Tercedor et al., 2009; Prieto and López, 2010). In the following example, the inclusion of a definition, an image, and conceptual and terminological information about the concept *amarrar* facilitates both word comprehension and production.

Figure 7. *Amarrar* in EcoLexicon
Working with Words

In conclusion, an interesting approach in translation-oriented lexicographic practice is producing on-line resources that integrate information from dictionaries, corpus data, images, and language forums.

3.2. Concordances, Meaning and Dialectal Variation

Concordances provide clues about conceptual information as well as lexical co-occurrence patterns of a keyword. In López and Tercedor (2008), and Tercedor and López (2008), we proposed four types of concordances, and described their usefulness in regards to the following:

(1) Extracting conceptual information (conceptual concordances) for the acquisition of knowledge about the subject field, its relevant concepts and their relationships.

(2) Knowing co-occurrence patterns in specialised discourse (structural concordances).

(3) Knowing the selection patterns of verbs and which verbs collocate with certain keywords (verbal structural concordances).

(4) Understanding the different senses of a word in regards to semantic prosody, metaphorical extensions, and word sense disambiguation.

Concordances can also show the different meaning or meanings a word may acquire as a result of geographical variation, an aspect that should be considered in lexicographic materials. For example, when sorting the concordances of the noun amarre (derived from amarrar, “to moor”) depending on the country (i.e., Spain vs. Cuba), it is interesting to note that in Spain, the word is mainly associated with navigation, whereas in Cuba, the word is primarily used in the context of witchcraft. In both cases, concordance lines point to the core meaning of the word: “to tie,” but different speech communities have extended this core meaning to different domains. Such geographical differences are reflected in the second sense of the word displayed in the DRAE, but not in other Spanish dictionaries such as Clave.
### Amarre in texts published in Spain

que conlleva mantener a una *flota en situación de amarre forzoso* “y sin perspectivas alentadoras.” Las **

Pesetas de la ayuda económica a cada *marinero en amarre forzoso* por la falta de acuerdo entre la UE y **

o de España. En los *puertos de Andalucía*, pese al amarre obligado de la *flota*, parece que empieza a flo **

LA UE Los *pescadores y armadores afectados por el amarre forzoso* provocado por la crisis con Marruecos **

70.000 pesetas por cada *pescador afectado por el amarre*. Estos mantuvieron ayer el bloqueo de las impo **

aluz señaló que cuando se cumplen quince días del *amarre forzoso de la flota comunitaria*, principalmente **

fin de semana rumbo a sus *respectivos puertos de amarre*. Los boniteros *españoles* tienen previsto celeb **

importantes daños en los cultivos y ha *forzado el amarre de la flota pesquera*. En España, esta situación **

os que ocasionó, según dicha central sindical, el *amarre posterior de los barcos de la flota pesquera d **

Rojaron al mar. El *incidente provocó entonces el amarre de la flota canaria*. El diputado Luis Mardones **

lar con Nueva Zelanda, país que ha *prohibido el amarre en sus puertos de barcos norteamericanos* que p *

ar una mosca. Después *liberó la embarcación de su amarre* y, clavando la pértiga en los *fondos someros*,

### Amarre in texts published in Cuba

anga significa en congo “amar rar” y por metáfora “amarre” o “hechizo.” Todo *embrujo* es, en rigor, el “a **

re” o “hechizo.” Todo embrujo es, en rigor, el “amarre” de una fuerza, de una *voluntad o de un destino* **

so, vana significaría que la acción del *conjuro, “amarre” o kanga* se proyectaría sobre el kanga o hechi **

Son de Choucoune. Aquí está el daño. Y esto es un *amarre de mujer celosa*. Busca a tu querida, ella fue **

ductores de su resistencia, *luchando contra mi amarre mientras me movía al borde del orgasmo* -pero e **

i del “medicamento magistral” romántico. Un *amarre de mujer siempre contiene gotas de sangre menstrual* **
Finally, concordances also show the different senses and nuances in the meaning of a word. They account for a collocational phenomenon called **semantic prosody**.

### 3.3. Collocational Meaning and Semantic Prosody

Lexicographers who create resources for the general public use frequency of occurrence as the most relevant criterion for the selection of collocates. Although translators are also interested in frequent collocations, they must also have access to other collocations that are not relevant in frequency but reveal a particular connotation in meaning not readily accessible by introspection. Thus, the relevance of the concept of **semantic prosody** is self evident for translation. According to Louw (1993, p. 157), a semantic prosody is a “consistent aura of meaning with which a form is imbued by its collocates.” Semantic prosody has aroused considerable attention within corpus linguistics over the last twenty years, and recent studies have compared the semantic prosody of synonyms or near synonyms across languages. Stewart (2009, p. 30) points to two main issues related to semantic prosody: (1) “the role of semantic prosody in the translation process; (2) how corpus data is intuitively converted into evidence of semantic prosody.” He mentions two pitfalls for translators. The first pitfall occurs when translators select only the data that best suit or conform to their perception/preconceptions (Tymoczko, 1998 cited in Stewart, 2009, p. 44). The second occurs when translators are disheartened by corpus evidence of the semantic prosody of a word, which seems to lend weight to the notion of a supposed impossibility of translation.
The inclusion of semantic prosodies in dictionaries can provide valuable information for translators when it comes to identifying collocational uses in which a clash between meanings is either intentional or the result of a problem in rendering the most appropriate combinatorial structure. Moreover, translators, who evidently cannot have a native command of all their working languages, often find it difficult to identify the positive or negative attitudinal meaning associated with certain words. The availability of concordances and the statistical data provided by collocational analysis are powerful resources that make semantic prosody explicit (López, 2001; López and Tercedor, 2008; Tercedor and López, 2008, pp. 173-174).

López (2001) studied the semantic prosodies of general language words, such as *effect*, *cause*, *response* and *produce* in a specialised corpus of oncology texts. Concordances for “effect(s)” showed that this word has a negative semantic prosody in cancer texts. In particular, the word “effect” co-occurred with lexical units indicating negative entities, properties or events (highlighted in bold in Figure 9):

1 more drugs are needed for maximal effect. Mature results of prospecti
2 om pulmonary and hematologic toxic effects; proper administration requ
3 , those studies showing a positive effect for combined modality therap
4 0-200 cGy to avoid long-term toxic effects to normal brain tissue [13,
5 because of the high risk of toxic effects observed with such treatmen
6 acted out in destructive ways, the effects can be devastating. Likewis
7 actively on sleep patterns. Side effects of treatment that may affec
8 me anxiety, and serious withdrawal effects, such as seizures [3].The f
9 enzodiazepines have an anti-anxiety effect in low doses and a hypnotic
that melatonin exerts a hypnotic effect through thermoregulatory mecha
nisms. Barbiturates have a number of adverse effects, including the de
evelopment of clinical response and minimize adverse effects. Some pa
tients may be evaluated [34,45-50]. The major adverse effects include
headache (w
oress. This side effect increases the risk of i
associated with a wide variety of side effects and toxic effects. Pruritus has been a side effect associated with several
effective and does produce gastric side effects [17]. Aspirin seems to

Figure 9. Semantic prosody of effect in a specialized corpus of oncology

In order to verify if these particular uses of the word effect were also prevalent in general language, we analysed the semantic prosody of effect/effects in the Bank of English corpus. To ensure that the language of both corpora (our corpus and the Bank of English) was comparable in terms of dialect (American English) and medium (written text), we searched effect/effects in the section of the Bank of English including 9 million words of American English. The analysis of linguistic contexts was complemented with the generation of the most frequent collocates of the lemma effect according to t-score and mutual information. We concluded that in general language, this lemma has an “open” semantic prosody, allowing for neutral, positive (italics and bold) or negative (underlined) connotations of the word, as opposed to a more restricted semantic prosody in specialized texts (Figure 10).
Therefore, combining frequency as attested by corpora (both ad-hoc and reference corpora) and collocational analysis, with attitudinal or relevance of meaning as revealed by semantic
prosody seems to be a good way to complement the purely statistical—and sometimes incomplete—measurements obtained with bottom-up methodologies such as corpus analysis.

In any case, translators should be aware of the fact that texts are not neutral:

From the potential available in the language, [these] texts use particular actual selections, which attach particular connotations to the lemma. Meanings are conveyed not only by individual words and grammatical forms, but also by the frequency of collocations and by the distribution of forms across texts. There are general expectations in the language as a whole as to how words will be used.” (Stubbs, 1996, p. 86)

Finally, attitudinal meaning is not always available to the native speaker, as it depends on many factors and changes over time. For example, the analysis of the semantic prosody of the word *maquiladora* (an assembly plant in the Mexican side of the Mexico-U.S. border run by U.S. or other foreign interests) indicates different lexical selections and semantic prosody of the word *maquiladora*, depending on the identity of the author of the text: a Mexican worker (“explotación” [exploitation], “bajos salarios” [low salaries], “manifestar” [to protest], “largas jornadas de trabajo” [long work days], “mal pagado” [low paid], “mano de obra barata” [cheap work force], “sector afectado” [affected sector]) or a U.S. business man (“incremento” [increase], “dar trabajo” [hire], “desarrollo” [development], “generar empleo” [create jobs], “crecer” [grow]).

### 4. World Knowledge, Familiarity and Frequency

It has been shown that the frequency of a lemma facilitates its lexical processing (Morris, 2006, p. 379). However, an important criterion for lemmatization purposes in lexicographic and terminographic tools alike is judging the familiarity of a word, as assessed by frequency, age of acquisition, and subjective familiarity. In fact, whereas frequency might be opaque to a reader or speaker, familiarity and age of acquisition influence the reader’s processing time, and therefore, should be considered in the assessment of words and dictionaries in general, and translation-oriented lexicographic practices in particular. In this
respect, we argue that data from experimental methods for the identification of subjective frequency, combined with statistical correlations of attested frequency can offer converging evidence about words and their intrinsic difficulties in processing and production. This section illustrates how these factors interact and complement research approaches.

The text in Figure 11 was created by the British Ministry of Transportation, and deals with different types of pedestrian crossings.

The translation of this text requires the activation of a wide variety of world knowledge. For instance, from reading the text, it follows that pedestrian crossings is a hypernym of different types of crossing (which are thus co-hyponyms), and that their lexicalizations are metaphorical and vary with regard to familiarity or subjective frequency. World knowledge—and corpus use—should allow translators to grasp this quickly, as well
as the fact that “zebra crossing” is often used as a generic term in certain contexts. However, this cognitive information must be combined with communicative uses. For instance, most bilingual dictionaries make it clear that “zebra crossing” is a lexical unit used in British English. Interestingly enough, even though this concept is lexicalised in a similar way in Spanish, *paso de cebra*, the term does not appear in other varieties of English because the metaphoric motivation is not present (i.e., the resemblance of the stripes on the animal to the stripes painted on the street as a pedestrian walkway).

To verify this, we performed an experiment in which we showed a photograph of a zebra crossing to five eight-year-old Canadian children and asked them to tell us what it was. All of them replied that it was a street, revealing a lack of familiarity with this shape (and name) in Canadian culture. However, the question is if translators would be able to extract this information from a dictionary. In all likelihood, frequency would be the criterion used to include a compound unit such as *zebra crossing* in a general language dictionary. Nevertheless, this criterion rarely suffices when it is necessary to perform a more fine-grained analysis that must discriminate between geographical usages. Furthermore, familiarity does not always coincide with frequency, and is generally not taken into account in the dictionary-making process.

<table>
<thead>
<tr>
<th>Search pattern</th>
<th>Frequency in COCA (EN-US) (total tokens)</th>
<th>Frequency in BNC (EN-UK) (total tokens)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crosswalk</td>
<td>175</td>
<td>0</td>
</tr>
<tr>
<td>cross walk</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>pedestrian crossing</td>
<td>12</td>
<td>44</td>
</tr>
<tr>
<td>zebra crossing</td>
<td>3</td>
<td>39</td>
</tr>
</tbody>
</table>

Figure 12. Search patterns and frequencies of the different lexicalizations of the concept *pedestrian crossing* in the Corpus of Contemporary American English and the British National Corpus
To complicate things further, it can be observed that these forms are considered compounds in some cases, deserving a headword, whereas in other cases, they are part of the headword of the modifier of the compound. This problem illustrates the fuzzy boundaries of the syntactic constraints between collocations and compound structures.

Figure 13. Codification of compounds for “pedestrian” and “paso de peatones” in the Oxford Spanish Dictionary

In lexicographic practice, the criterion of frequency, along with restrictions of time and space, invariably excludes encyclopaedic or cultural information from dictionaries that would be extremely useful for translators. Such information is generally not explicitly included in texts because it constitutes the shared knowledge of the native speakers of a language. For example, the lack of encyclopaedic knowledge and cultural competence was responsible for an erroneous interpretation of the cartoon in a column published in The Economist about President Obama’s health care reform on 25 March 2010. The title of the column was “From Hope to Change. Barack Obama Has Made History. But He Can Still Make Mistakes.” Students who did not know that U.S. Democrats have a donkey as their symbol, and that Republicans have an elephant as theirs, did not understand that “social hero,” and “socialist” were two different perspectives for and against President Obama and Obamacare, corresponding to the Democrats and Republicans respectively. Some students even thought that the elephant was a dog or a pig.

These examples illustrate how the translator works by comparing lemmas in dictionaries and encyclopaedias (cognitive approach) as well as in texts so as to observe the behaviour of words in context (functional approach). A dictionary specifically for translators should thus target the specific needs of the user group and combine both cognitive and functional uses. Regarding entry selection and collocational choices, word frequency in the corpus should not be the only criterion used to include information in lexicographical resources. Finally, lexicographers should consider psychological and cognitive research insights into concept boundaries, conceptual relevance, dynamic representations, and familiarity when elaborating dictionaries.

Concluding Remarks

This paper has provided an overview of research approaches offering insights into the lexicon and the use that translators make of lexicographic resources. We have stressed the need to work
towards combining approaches in different knowledge fields. In usage-based linguistics—which includes lexicography—, interdisciplinarity and empirical methods can and should run in parallel. As Tummers, Heylen and Geeraerts state:

Overall, we will have to think of the empirical methodology of usage-based linguistics as having a helix-like structure, involving a gradual refinement of interpretations through a repeated confrontation with empirical data—all kinds of empirical data. An initial hypothesis, which may be derived introspectively, is confronted with the corpus data; interpreting the results leads to a more refined hypothesis and more questions, which may then be subjected to further experimental testing or a new confrontation with the corpus data—and so on. (2005, p. 233)

References


GÖPFERICH, Susanne and Riita JÄÄSKELÄINEN (2009). “Process Research into the Development of Translation Competence: Where are We, and Where Do We Need to Go?” Across Languages and Cultures, 10, 2, pp. 169-191.

GÖPFERICH, Susanne, A.L. JAKOBSEN and I.M. MEES (2008). Looking at Eyes: Eye-Tracking Studies of Reading and


LOUW, Bill (1993). “Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies.” In M.
Working with Words


**Corpora and Dictionaries**


*Diccionario Clave*. Available at: <http://clave.librosvivos.net/> [consulted 2 July 2010].

ABSTRACT: Working with Words: Research Approaches to Translation-Oriented Lexicographic Practice — Dictionaries ideally should address the needs of particular types of users. Their micro and macrostructural design should be oriented towards what user groups need to know about words and the uses that will be made of such knowledge. One specific use for dictionaries is the activity of translation. From the perspective of professional translators, dictionaries should allow for creativity and dynamicity in text production, providing solutions for changing communication needs. Dictionary-making for such a purpose can and should benefit from insights into words and meanings from other fields. Interdisciplinarity in Lexicography is just one example of how other fields interact in Translation Studies. In this paper we analyse how working in an interdisciplinary way is crucial to developing useful lexicographic and terminographic tools for translators and how methodologies, such as corpus-based work and experimental methods should be combined to offer converging evidence of different aspects of use and processing. We illustrate such work methods with examples from real lexicographic projects.

RÉSUMÉ : Travailler avec les mots : vers une pratique lexicographique à l’usage de la traduction — Un dictionnaire idéal devrait prendre en compte les usages particuliers de ceux qui s’en servent. La conception de leur micro- et macrostructure devrait être conçue en fonction du groupe auquel il est destiné, et notamment en fonction de ce que les usagers doivent savoir sur les mots et de ce qu’ils prévoient en faire. Nous utilisons les dictionnaires spécifiquement dans la cadre de notre activité de traduction. Du point de vue du traducteur professionnel, les
dictionnaires devraient stimuler la créativité et le dynamisme en proposant des solutions appropriées aux nouvelles exigences de communication. À cette fin, l’élaboration des dictionnaires peut et devrait se nourrir du regard que posent sur les mots les autres disciplines. L’interdisciplinarité dans la lexicographie est un exemple de l’interaction des différents champs dans les études de traduction. Dans cet article, nous analysons, tout d’abord, la raison pour laquelle il est crucial de travailler dans une perspective interdisciplinaire pour développer des outils lexicographiques et terminologiques destinés aux traducteurs. Ensuite, nous étudions la façon dont les méthodologies telles les analyses de corpus et les méthodes expérimentales devraient être combinées pour créer une synergie bénéfique. Nous illustrons ces méthodes de travail par des exemples extraits de projets lexicographiques réels.

**Keywords:** lexicographic tools for translators, process-oriented translation studies, product-oriented translation studies, experimental methods, collocational meaning

**Mots clés :** outils lexicographiques pour traducteurs, études de traduction orientées vers le processus, études de traduction orientées vers le produit, études expérimentales, sémantique des collocations

Maribel Tercedor  
Clara Inés López-Rodríguez  
Pamela Faber  
University of Granada  
Department of Translation and Interpretation  
C/ Buensuceso, 11  
Granada 18002 Spain  
itercedo@ugr.es  
clarailr@ugr.es  
pfaber@ugr.es