Encoding Context in Bilingual Specialized Dictionaries

Pamela Faber, Arianne Reimerink, Pilar León-Araúz
Department of Translation and Interpreting, University of Granada
e-mail: pfaber@ugr.es, arianne@ugr.es, pleon@ugr.es

Abstract

Specialized dictionaries and knowledge resources encode the meaning of specialized knowledge units but they rarely include contextual information. Context is crucial for language comprehension and production, since they both largely depend on users being able to activate the right frame. Moreover, context can be defined in many different ways. Therefore, it is timely and necessary to parametrize context with a view to more effectively facilitating knowledge representation in term entries. In this paper we propose a three-level parametrization of context based on scope (i.e. local vs. global) and according to the type of information conveyed (i.e. syntactic, semantic and pragmatic). The term 'absorption' is used to exemplify what kind of contextual information is normally included in bilingual specialized dictionaries and how this information could be expanded to provide more useful information for text comprehension and production.
Keywords: context parameters; bilingual dictionaries; specialized knowledge

1 Introduction

Although specialized dictionaries and knowledge resources encode the meaning of specialized knowledge units used in expert communication, they rarely include contextual information. Elman (2009: 572) highlights the importance of context in language comprehension and asserts that the meaning of a word is rooted in our knowledge of both the material and social world. He also highlights the major role of larger knowledge structures or events in the organization of experience. Accordingly, all specialized knowledge units need to be understood within the context of a larger event.

Contextual information is a crucial component in specialized knowledge resources, since understanding depends on users being able to activate the right frame (Faber 2012; Faber *et al.* 2014) in which the specialized knowledge unit should be processed. Frawley (1980) already mentioned the importance for a systematic relational representation of the specialized knowledge based on the premises of the Meaning-Text Theory (MTT) of Mel'cuk and Zolkovsky (Mel'cuk 1988). Of the lexical functions MTT proposed for general language dictionary construction, Frawley (1980: 24) proposed the following for the inclusion in specialized dictionary entries: taxonomy, synonymy, antonymy, gradation, cause, part/whole, source, result and etymology (apud cit. Montero-Martínez 2003). These lexical functions systematize usage information and explicitly link related entries, which would certainly help activate a larger frame for understanding. Nonetheless, such information is either not found in specialized resources or is inconsistently represented. This situation can be improved by designing entries that systematically include relevant contextual data. In the case of bilingual resources, this would signify the inclusion of both source language and target language contexts.

One of the obstacles to designing a 'contextualized' term entry lies in the fact that *context* has no

universally accepted definition. For this reason, the creation of a contextualized term entry first requires the categorization and parametrization of contextual information.

2 Context as Encoded in Specialized Dictionaries

Context is often defined in terms of its scope or extension (see Section 3). It may be a few words on either side of a term, the sentence or paragraph in which it is appears, or even a set of documents containing it. However, the contextual information in bilingual or multilingual specialized dictionaries, if included at all, is mostly restricted to the specification of subdomains and a list of terminological phrasemes in the source language and their translations in the target language.
All too frequently, the only information provided is the source language term and a list of possible equivalents in the target language. Table 1 shows the entry for 'absorption' in the *English/Spanish Dictionary of Environmental Science and Engineering* (Headworth & Steines 1997) in which only the English term and two Spanish correspondences are shown.

absorption *n.* absorción (*f*), empapamiento (*m*).

Table 1: Entry for 'absorption' in the English/Spanish Dictionary of Environmental Science and Engineering.

Apart from the information regarding gender, no information is provided as to usage or degrees of equivalence. The entry is also flawed because along with the best all-purpose correspondence, *absorción* [absorption], it includes *empapamiento* [soaking], which is not used in scientific and technical writing. In addition, it is only applicable to liquid entities, and cannot be used in texts where the substance being absorbed is a non-liquid entity, such as air, light, radiation, etc. It is also not relevant to absorption processes that lead to chemical reactions.
In *Elsevier´s Dictionary of Soil Science* (Canarache *et al.* 2006), the entry for 'absorption' includes an English definition for Chemistry and for Physics. This is followed by a list of correspondences in French, German, and Spanish with no grammar or usage information.
In contrast, the entry for '*absorción*' [absorption] in the *Routledge Spanish Technical Dictionary/Diccionario Técnico Inglés* (Routledge 1997) includes contextual information regarding gender, specialized domain (ING MECÁ, FIS RAD, METEO, etc.), geographic variation (*AmE* or *BrE*), and multi-word combinations **(~atmosférica, ~ de microondas; ~máxima**, etc.) (see Table 2). This is important because it encodes: (i) the entity that absorbs, e.g. the atmosphere ('*absorpción atmosférica*' → 'atmospheric absorption'); (ii) the entity being absorbed, e.g. microwaves ('*absorción de microondas*' → **'**microwave absorption'); (iii) intensity of absorption, e.g. peak ('*absorción máxima*' → 'peak absorption'). Finally, in a few places, the entry includes the specific context of the absorption (*tintura* [dyeing], *tratamiento fisioquímico* [physical-chemical treatment], *KALC process*, etc.). However, relatively few items are provided with these markers, which seem to be randomly inserted.

absorción *f* GEN absorption, ING MECÁ indraft (*AmE*), indraft of air (*AmE*), indraught (*BrE*), indraught of air (*BrE*), TEXTIL *del tinte* uptake;**~de aceite** *f* P&C oil absorption;**~atmosférica** *f* FIS RAD, METEO atmospheric absorption; **~del baño** *f* TEXTIL *tintura* dip pickup; **~de calor** *f* TERMO heat absorption; ~**por carbon activado** *f* TERMO *tratamiento físicoquímico* active carbon absorption; ~**de carga** *f* TELECOM energy absorption; **~de criptón en anhídrido carbónico liquid** *f* NUCL krypton absorption in liquid carbon dioxide (*KALC process*); **~dieléctfica** *f* ELEC, ING ELÉC dielectric absorption; **~específica** *f* METEO specific absorption; **~de fondo** *f* FIS RAD background absorption; **~fotón-fotón** *f* FIS PART photon-photon absorption; **~fosférica** *f* FIS RAD photospheric absorption; **~de gas** *f* GAS gas absorption, QUÍMICA, *por metal* gassing, TEC PETR *refino* gas absorption; **~ de luz** *f* FIS RAD absorption of light; **~máxima** *f* ING ELÉC absorption peak; **~de microondas** *f* TELECOM microwave absorption; **~óptica** *f* FIS RAD, OPT optical absorption; **~ozónica** *f* FIS RAD ozone absorption; **~ de radiación** *f* FIS ONDAS, FIS RAD absorption of radiation; **~ de radiaciónionizante** *f* FIS ONDAS, FIS

RAD absorption of ionizing radiation; **~ de rayos X** *f* FIS RAD X-ray absorption; **~ de Sabine** *f* ACUST Sabine absorption; **~sonora***f* ACUST sound absorption; **~ del suelo** *f* AGUA absorption in the soil; **~ del terreno** *f* ACÚST ground absorption.

Table 2: Entry for 'absorción' in the Routledge Spanish Technical Dictionary/Diccionario Técnico Inglés.

Nevertheless, none of the entries provide information that facilitates text production in the target language. For this reason, it is necessary to analyze context in order to establish an interrelated network of context parameters.

3 Context Parameters for Specialized Knowledge Representation

As reflected in corpus analysis, when context is mentioned in a text, it is metaphorically conceived as a container or a bounded space, since an utterance can be "in context" or "out of context." As a relational construct in texts, context helps anchor linguistic designations to objective reality by providing background information, situating objects and processes, and explicitly relating them to one another as well as to the agents that manipulate and act on them (Faber & León-Araúz 2016: 2). It is thus a constraining factor that drives understanding. In other words, as stated by Leech (1981), the specification of context (whether linguistic or non-linguistic) has the effect of narrowing down the communicative possibilities of the message as it exists in abstraction from context.

Contextual information can be specified in terms of scope (local vs. global) or according to the type of information conveyed (syntactic, semantic, and pragmatic variables). In reference to specialized knowledge units, the primary division of context is based on scope. Context may be a few words on either side of a specialized lexical unit (He *et al.* 2010), the sentence or paragraph in which it appears (Soricut & Marcu 2003), a set of documents containing it (Cilibrasi & Vitanyi 2007), a communicative act, or even a whole culture.

According to Akman and Bazzanella (2003: 325), an adequate multi-modal coding of context on both the global and local levels would be useful in delimiting inferences, disambiguating deictic expressions, and solving the problem of indeterminacy. Thus, the distinction of local vs. global can be found elsewhere in the literature though not with the same meaning. For instance, Bazzanella (1998), Akman and Bazzanella (2003), and Miecznikowski and Bazzanella (2007) refer to local context to denote a specific setting where the participants interact; and use global context for referring to the members of a community, their social norms, culture, beliefs, ideology, etc. However, Mihalcea (2007) uses the same distinction to refer to a different context span within textual excerpts (a pair of words vs. lexical chains), whereas Dash (2008) proposes a continuum of four contexts from local to global.

In our view, local contexts are usually limited to the words within the term itself, to a small number of words in the immediate vicinity of a term, or to words connected by syntactic dependencies to the term. In contrast, global contexts can encompass the whole text or go beyond the text: to the communicative situation (i.e. formal vs. informal); to the conceptual networks reflected in it; to the culture in which the text is interpreted, etc. This means that global contexts refer to items that are often quite a distance from the term or even outside of the text altogether though within the specialized domain (Faber & León-Araúz 2016: 5).

Local and global contexts can be further subdivided as syntactic, semantic, or pragmatic, depending on the nature of the information conveyed (see Table 4). It is true that it is extremely difficult to trace a clear boundary line between syntax, semantics, and pragmatics because there is a significant degree of overlap. However, this fuzzy three-level approach to context goes hand in hand with the

micro-theories proposed by FBT, which are related to the information encoded in term entries, the relations between specialized knowledge units and the concepts they designate (Faber 2015: 15).

3.1 Local Contexts

According to Agirre and Stevenson (2007: 225), the data that can be derived from local contexts are the following: part of speech, morphology, collocations, subcategorization, frequency of senses, syntagmatic and paradigmatic word association, selectional preferences, semantic roles, domain, topical word association, and pragmatics.

As mentioned before, in our parametrization of context, local contexts are usually limited to the words within the term itself or to words connected by syntactic dependencies to the term. In the case of 'absorption', this context is reflected in its pre- and post-modification. In English, pre-modifiers can be adjectives (*subcutaneous absorption*) or nouns (*photon absorption*), whereas post-modifiers are generally prepositional phrases (*absorption of electromagnetic radiation*).

More important, however, are the semantic categories of these words. Accordingly, adjective pre-modifiers most frequently refer to the absorbing entity or medium of absorption (e.g. *human body→skin→subcutaneous*) whereas noun pre-modifiers generally refer to what is being absorbed (e.g. *electromagnetic radiation→light→photons*). Alternatively, noun pre-modifiers can also be encoded as prepositional phrases. Thus, *electromagnetic radiation absorption* can be formulated as *absorption of electromagnetic radiation*. In this way, semantics generates selection restrictions that constrain the combinatorial capacity of a term in a certain specialized domain.

3.2 Global Contexts

Global contexts can encompass the whole text or extend to the communicative situation and conceptual categories reflected in it. In Frame-based Terminology (Faber 2012; Faber *et al.* 2014; Faber 2015), user understanding of an entity or group of entities in a specialized domain depends on having access to the information required to activate the right frame or knowledge structure in which the term should be processed. When conceptual categories are regarded as domains, these categories are constrained by the shared properties of category members. For example, scientific instruments and chemical processes (e.g. *absorption*) differ in the set of conceptual relations that reflect their interconnections with other entities.

However, when a domain takes the form of a specialized knowledge field (e.g. Chemistry, Physics, or Biology), this also provides contextual constraints that lead to the recontextualization of versatile concepts (León-Araúz & Magaña 2010). For example, within the domain of Chemistry, the entities usually involved in 'absorption' (carbon dioxide, ammonia, etc.) are quite different from absorption-related entities within the domain of Physics (e.g. photons) or Biology (e.g. chlorophyll, nutrients).

Thus, specialized knowledge units need to be understood within the context of a larger event. Such conceptual information is not language-specific, and leads to the specification of mini-conceptual frames based on the combination of entity types in a role-set: ABSORPTION $(x)_{\text{absorbed entity}}$, $(y)_{\text{absorbing entity}}$, $(z)_{\text{absorption medium}}$. How this information combines to form terminological phrasemes depends on the language.

For example, in English, ABSORPTION $(\text{nutrients})_{\text{absorbed entity}}$, $(\text{plants})_{\text{absorbing entity}}$ is encoded as *plant nutrient absorption* or alternatively, as *nutrient absorption by plants*. Nevertheless, since Spanish does not accept this type of pre-modification, the user must first understand the world knowledge underlying the process. This means recognizing the action performed, which involves one substance

penetrating and being taken inside another. It also entails the identification of 'plant' as the absorbing entity and 'nutrient' as the absorbed substance since the conceptual links between term elements must be specified with prepositions.

Consequently, the absorbed entity, '*nutriente*' [nutrient], would be linked to '*absorción*' [absorption] by the preposition *de* [of] and the absorbing entity, '*planta*' [plant], by *en* or *por* [in/by] (see Table 3). Another language-specific difference that has to be taken into account in mapping rules involves the use of definite articles and singular and plural forms.

ABSORPTION $(\text{nutrient})_{\text{absorbed entity,}}$ $(\text{plant})_{\text{absorbing entity}}$	
plant nutrient absorption / nutrient absorption by plants	English
absorción de nutrientes en/por las plantas	Spanish

Table 3: Linguistic activation of the (plant-nutrient) absorption frame in English and Spanish.

Although this case is fairly straightforward, on other occasions, the relations between term components may be less transparent. For example, the terms 'water ammonia absorption' and 'amine acid gas absorption' are less transparent if the user has no knowledge of chemical processes or entities involved. Nevertheless, this is the knowledge required for text production in the target language.

4 A Context-Enhanced Knowledge Base Entry

Table 4 shows an integrated specification of contextual information for 'absorption' in English. Obviously, this is not the entry itself, but rather the information that the user needs to activate for text production. Moreover, not all terms would activate the same slots, because not all specialized units show meaningful contextual information at all levels. In this sense, the term 'absorption' itself does not activate a very elaborate pragmatic template, since there are no term variants associated with the same concept.

Info type		***Absorption***
SYNTACTIC CONTEXT		
part of speech	Noun	Absorption
morphology	action or result of a verb + *tion*	*absorb* + *tion* → *absorption*
syntagmatic word association	N + **N** **N** + PP (*of*) (+ PP[*by/into*]) (+PP [*through*]) Adj [+ Adj] + **N**	energy absorption absorption of energy (by electrons) atmospheric (gaseous) absorption
SEMANTIC CONTEXT:	**process** by which **one substance (x)** penetrates and is taken into the interior of **another substance (y).**	
frame	*x* is absorbed by/into *y* (through *z*)	
frame instantiation	• *x* [substance absorbed] absorption *by* [absorbing entity] • absorption of *x through/into* z [medium] • *y* [absorbing entity] absorption	• nutrient absorption by plants • carbon dioxide absorption by/into water • absorption of chemicals through the skin • gastrointestinal absorption
paradigmatic word association	[CHEMISTRY] *x* (gas/molecules/ions/atoms) *y* (liquid/solid/gas/chemical)	• carbondioxide/ solid particles/acid gases • water/sodium hydroxide/zinc oxide/ethanolamine/palladium
	[BIOLOGY] *x* (pharmaceutical/nutrient/toxin) *y* (living organism)	• botulinum/acetonitrile/chlorophyll • plant/skin/human body

	[PHYSICS] *x* (energy/radiation /light/ sound waves) *y* (surface/ atmosphere/ object/ material/ organism)	• radiant energy/x ray/photon/light/ electromagnetic radiation/sound • Earth's surface/building walls/ glass/body/plants
selectional preferences	*absorption* of (*x*) PHARMACEUTICALS by (*y*) HUMAN BODY *absorption* of (*x*) LIGHT by (*y*) PLANTS *absorption* of (x) GAS by (y) POROUS MASSES	
PRAGMATIC CONTEXT		
pragmatics	chemical absorption vs. physical absorption absorption vs. adsorption	

Table 4: Integrated specification of contextual information regarding 'absorption'.

A specialized knowledge resource should help the user create this integrated specification of the information related to the term 'absorption' and provide mapping rules for interlinguistic correspondences (see Figure 1).

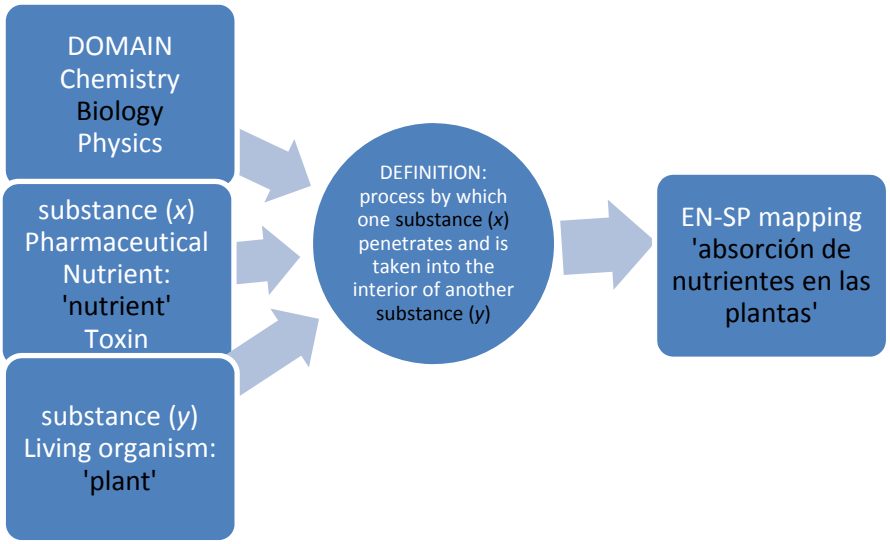

Figure 1: Contextualized entry design for 'plant nutrient absorption'.

The user does not need to know the information to fill in all the slots. The entry will provide the most plausible answer, based on the information that the user possesses. For example the term *absorption* is relevant to Chemistry, Biology or Physics. Accordingly, in the definition in the form of the micro-conceptual frame in Table 4 (i.e. process by which one substance (*x*) penetrates and is taken into the interior of another substance (*y*)), the slot fillers provide a set of semantic categories, depending on the domain. Conversely, if the user already knows (*x*) and/or (*y*), this will identify the specialized domain. When this information is provided in the source language, it will trigger the mapping relations, which will produce the target language correspondence.

5 Conclusion

Contextual information is a crucial component in specialized knowledge resources since understanding depends on users being able to activate the right frame in which the specialized knowledge unit should be processed. However, the creation of a contextualized term entry first requires the categorization and parametrization of contextual information. This paper analyzed

contextual information in the entries of 'absorption' in three specialized Spanish-English dictionaries and suggested how this information could be expanded so as to be more useful for text comprehension and production in the target language.

6 References

Agirre, E., Stevenson, M.(2007). Knowledge sources for word sense disambiguation. In E. Agirre, P. Edmonds (eds.) *Word Sense Disambiguation: Algorithms and Applications*. NewYork, NY: Springer.

Akman, V., Bazzanella, C. (2003). The complexity of context: guest editors' introduction. *J. Pragmat.* 35, pp. 321-329.doi:10.1016/S0378-2166(02) 00138-8.

Bazzanella, C. (1998). On context and dialogue. In S. Cmejrková, J. Hoffmannová, O. Müllerová, and J. Svetla (eds.) *Dialogue in the Heart of Europe, Proceedings IADA (Prague1996)*. Tübingen: Niemeyer, pp. 407-417.

Canarache, A., Vintila, I. & Munteanu, I. (2006). *Elsevier's Dictionary of Soil Science*. Burlington MA: Elsevier Science.

Cilibrasi, R. L., Vitanyi, P. M. B. (2007). The google similarity distance. *IEEE Trans.Knowl.DataEng.* 19, pp. 370-383.doi:10.1109/TKDE.2007.48

Dash, N. S. (2008). Context and contextual word meaning. *SKASEJ. Theor. Linguist.* 5, pp. 21-31.

Elman, J. (2009). On the meaning of words and dinosaur bones: lexical knowledge without a lexicon. *Cognitive Science* 33, pp. 1-36.

Faber, P. (2015). Frames as a framework for terminology. In H. J. Kockaert, F. Steurs (eds.) *Handbook of Terminology*. Philadelphia: John Benjamins, pp. 14-33.

Faber, P. (ed.) (2012). *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin, Boston: De Gruyter Mouton.

Faber, P., León-Araúz, P. (2016. Specialized knowledge representation and the parameterization of context. *Frontiers in Psychology*, 7(00196). doi:10.3389/fpsyg.2016.00196.

Faber, P., León-Araúz, P. and Reimerink, A. (2014). Representing environmental knowledge in EcoLexicon. In E. Bárcena, T. Read & J. Arhus (eds.) *Languages for Specific Purposes in the Digital Era*, Educational Linguistics 19. Berlin, Heidelberg: Springer.

Frawley, W. (1980). *Lexicography and the Philosophy of Science*. Dictionaries, 3, pp. 18-27.

He, Q., Pei, J., Kifer, D., Mitra, P., Giles, C. L. (2010). Context-aware citation recommendation. In *Proceedings of WWW 10th Proceedings of the 19th International Conferenceon World Wide Web.* New York, NY: Association for Computing Machinery, pp. 421-430.

Headworth, H., Steines, S. (1997). *English/Spanish Dictionary of Environmental Science and Engineering*. Chicester, West Sussex: Wiley and Sons.

Leech, G. (1981). *Semantics: The Study of Meaning*. Harmondsworth: Penguin.

León Araúz, P., Magaña, P. (2010). EcoLexicon: contextualizing an environmental ontology. In *Proceedings of the Terminology and Knowledge Engineering (TKE) Conference 2010*. Dublin: Dublin City University.

Mel'cuk, I. A. (1988). Semantic description or lexical units in an explanatory combinatorial dictionary: Basic principles and heuristic criteria. *International Journal of Lexicography* 1(3), pp. 165-188.

Miecznikowski, J., Bazzanella, C. (2007). The attenuating conditional: context, appropriateness and interaction. In A. Fetzer (ed.) *Context and Appropriateness. Micro meets Macro*. Amsterdam:

John Benjamins, pp. 203-233.

Mihalcea, R. (2007). Using Wikipedia for automatic word sense disambiguation. In *Proceedings of NAACLHLT*. Rochester, NY: Association for Computational Linguistics, pp. 196-203.

Montero-Martínez, S. (2003). *Estructuración conceptual y formalización terminográfica de frasemas en el subdominio de la oncología*. PhD Thesis. Universidad de Valladolid.

Routledge (1997). *Routledge Spanish Technical Dictionary/Diccionario Técnico Inglés*. New York: Routledge.

Soricut, R., Marcu, D. (2003). Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of the Human Language Technology and North American Association for Computational Linguistics Conference (HLT/NAACL)*. Edmonton, AB.

Acknowledgements

This research was carried out within the framework of the project CONTENT [Cognitive and Neurological Bases for Terminology-enhanced Translation], funded by the Spanish Ministry of Economy and Competitiveness