

SPECIALIZED KNOWLEDGE REPRESENTATION: FROM TERMS TO FRAMES*

PAMELA FABER

University of Granada, Spain
pfaber@ugr.es

MELANIA CABEZAS-GARCÍA

University of Granada, Spain
melaniacabezas@ugr.es

Abstract

Understanding specialized discourse requires the identification and activation of knowledge structures underlying the text. The expansion and enhancement of knowledge is thus an important part of the specialized translation process (Faber 2015). This paper explores how the analysis of terminological meaning can be addressed from the perspective of Frame-Based Terminology (FBT) (Faber 2012, 2015), a cognitive approach to domain-specific language, which directly links specialized knowledge representation to cognitive linguistics and cognitive semantics. In this study, context expansion was explored in a three-stage procedure: from single terms to multi-word terms, from multi-word terms to phrases, and from phrases to frames. Our results showed that this approach provides valuable insights into the identification of the knowledge structures underlying specialized texts.

Keywords: context expansion, frame, multi-word term, phrase, specialized discourse

1. Introduction

An important issue in translation is how to achieve sameness of meaning across languages and at all levels of the text. In the case of the translation of scientific and technical texts, a considerable percentage of translation quality depends on finding optimal correspondences for the specialized language units or terms used to convey the text message. These units, which may be single or multi-word terms, designate objects, events, processes, and attributes in the specialized field (Faber 2012).

Terms, semantic clusters of terms, and their configurations activate segments of the conceptual structure of a knowledge domain (Sager et al. 1980), which is present in the source and target language-cultures. If both language-cultures have

* This research was carried out as part of FFI2017-89127-P, *Translation-oriented Terminology Tools for Environmental Texts (TOTEM)*, research project funded by the Spanish Ministry of Economy and Competitiveness. Funding was also provided by an FPU grant given by the Spanish Ministry of Education to the second author.

terms for the entities designated, the assumption is that the text can be translated with a reasonable degree of accuracy. The translator must first be aware of what is happening in the text and the message that it conveys. Then he/she identifies term correspondences and finds the most accurate way to link them so as to highlight the semantic relations between concepts that are explicit in the text.

Understanding specialized discourse thus depends on the text receiver's capacity to grasp and then activate the knowledge structures underlying the text. When the text receiver is not an expert in the field (as often occurs in a specialized translation scenario), he/she must be able to rapidly acquire the necessary domain-specific knowledge (Faber 2012).

In the translation process, the specialized knowledge units in a text as well as their relations must be analyzed at various levels. Although the meaning of certain concepts and relations are evident in the surface structure of the text, this is merely the tip of the iceberg. There is a whole world of meaning lurking beneath the surface, which translators must be able to perceive. Relevant data from the source language text must be generalized or abstracted with a view to integrating new information into semantic memory. Understanding thus depends on the translators' ability to successfully construct a mental representation of a segment or segments of the specialized knowledge field. The expansion and enhancement of knowledge is thus an important part of the specialized translation process (Faber 2015).

This paper explores how the analysis of terminological meaning can be addressed from the perspective of Frame-Based Terminology (FBT) (Faber 2012, 2015), a cognitive approach to domain-specific language, which directly links specialized knowledge representation to cognitive linguistics and cognitive semantics. In FBT, knowledge acquisition involves a progressive expansion of meaning, which begins at the term-level, progresses to the phrase level, and finally results in the codification of an entire knowledge frame.

2. Theoretical background

To understand how knowledge is configured, and expanded, it is necessary to start with the brain. Neurological studies provide insights into how experts retrieve and activate stored knowledge (Quillian 1969; Anderson 1983; Gallese and Lakoff 2005; Patterson et al. 2007; Meteyard et al. 2012; Kiefer and Pulvermüller 2012). For this reason, Faber et al. (2014) conducted a pilot fMRI study in which brain activation images of expert geologists were compared to those of novices as they performed a series of different tasks, such as linking geological tools to their function and tools to images. The results showed that expert knowledge involves a supramodal conceptual representation, which transcends sensory input modalities such as vision or hearing. Conceptual representations thus have

multiple levels of input (Binder and Desai 2011), which do not only come from the senses.

At the top level, much research agrees that there is a non-modality-specific schematic representation, which is progressively fleshed out by sensory-motor-affective input when and as needed (Patterson et al. 2007). Faber et al. (2014) highlighted the key role played by contextualization and situation in specialized knowledge processing since the brain regions activated by experts (though not novices) were those strongly implicated in mental imagery, episodic memory, and context representation. Although more studies are necessary, Faber et al. (2014) further validated the need to explore how contextual information can be activated and thus facilitate frame creation in the non-expert.

Accordingly, FBT applies the notion of ‘frame’ (Minsky 1975; Fillmore 1985, 2006) defined as a schema or knowledge structure, which relates elements and entities associated with a particular scene, situation that is part of human experience. A frame is thus as an organized package of knowledge that humans retrieve from long-term memory to make sense of the world (Faber 2012). Given that concepts cannot exist in a vacuum, they are more meaningful when they are related to each other and integrated into progressively more complex knowledge configurations. Framing experience involves applying stored knowledge derived from similar contexts and situations with a view to understanding complex events and how to deal with them.

In Terminology, the usefulness of embedding concepts in situations has also been highlighted as a way of enriching conceptual representations (Dubuc and Lauriston 1997; Faber 2012; Temmerman 2013). Although context is often regarded as the segment/s which precede or follow a word or phrase (Lyons 1995), it can also be a related situation, events, or information that help users to understand something, and which reflect a specific knowledge profile (Kecskes 2014; Faber and León-Araúz 2016). The specification of contexts should thus take place at multiple levels that range from concept to frame.

3. MWTs and context expansion

As is well known, MWTs are terms composed of more than one word. In English, they can be of varying length: (i) two constituents (*transboundary pollution*); (ii) three constituents (*surface water pollution*); (iii) four constituents (*wood-burning-stove pollution*); and even (iv) five constituents (*volatile organic compounds pollution source*). Although in theory, MWTs can go on forever, it is extremely rare to find one longer than four or, at most, five words because of the cognitive demands made on text receivers.

In English, these complex terms resemble a type of expert shorthand, where there is no need for further explanation because of the level of knowledge presumably shared by the text sender and receivers. Users are thus obliged to unpack the meaning of MWTs and correctly access the relationship between the

constituents. To do this, they must mentally activate a specialized event or frame in which the relations between participants are specified. Although this is relatively easy for an expert in the field, it can be somewhat more difficult for a non-expert.

Consequently, the process of understanding terminological meaning begins with the concept designated by the term itself and is conceived as a progressive expansion of contexts. First, there is the term and its microcontext (Cabezas-García and Faber in press), which can be further expanded to a set of related multi-word terms (MWTs). As shall be seen, these MWTs can subsequently be unpacked by inserting implicit information and then by explicitly relating them to each other as well as to other specialized knowledge units. As we shall see, this gives rise to the specification of larger knowledge structures or frames.

3.1. From single terms to MWTs

Context expansion initially takes place when a single term undergoes further specification and becomes a multi-word term (MWT). In specialized language, most MWTs take the form of endocentric noun compounds (Nakov 2013), e.g. *climate change*.

Endocentric MWTs are informative because they point to relations between and within semantic categories. Generally speaking, an endocentric MWT is a specialization of the meaning of its head. This means that term structure can often be used as a way to automatically extract information regarding conceptual hierarchies as well as hyponymy subtypes (Sager et al. 1980). For example, *vessel-source marine pollution*, is a type of *marine pollution*, which is a type of *pollution*. For further semantic characterization, we can also say that pollution *affects* the sea and *is caused by* vessels.

In morphologically-poor languages, such as English, endocentric MWTs can take the form of sequences or stacks of nouns of varying length. In English, lengthy pre-modification in the form of a series of nouns, also modified by adjectives or even entire phrases, is a frequent method that is used to condense and concentrate domain-specific knowledge (Sager et al. 1980; Štekauer et al. 2012; Fernández-Domínguez 2016).

Concept specialization involves a slot-filling mechanism where the modifier is inserted into a slot in the head-noun schema, also known as its micro-context (Cabezas-García and Faber in press). In an MWT, the modifier is directly related to the base meaning of the head noun as (under)specified in the definition and is interpreted accordingly. In the second stage, world knowledge is used to expand the context of the headword and its interpretation.

For example in the case of *pollution*, this expansion of context starts with its definition:

(1)

pollution presence in the **environment** [Slot 1] of a **substance** [Slot 2], whose **nature** [Slot 3], **source** [Slot 4], **location** [Slot 5] or **quantity** [Slot 6] produces **undesirable effects** [Slot 7] for the environment or the health of living organisms.

The general concept of *pollution* is thus defined in terms of seven meaning slots: (i) environment; (ii) substance; (iii) nature; (iv) source; (v) location; (vi) quantity; and (vii) undesirable effects. All of these slots are underspecified and thus susceptible to be filled by hyponyms of the terms in bold. When one or more of these slots are made more specific, this generates MWTs that are hyponyms of *pollution*. Table 1 shows examples of sets of MWTs corresponding to each slot.

Table 1. MWT hyponyms of pollution

Definition slots: <i>pollution</i>	Multi-Word Terms
ENVIRONMENT	air pollution, water pollution, soil pollution, marine pollution, ocean pollution
SUBSTANCE	oil pollution, particle pollution, solid waste pollution, nutrients pollution
NATURE (of substance)	volatile organic compounds pollution
SOURCE (of substance)	point-source pollution, non-point-source pollution, wood-burning-stove pollution, industrial pollution, traffic-related air pollution
LOCATION (of substance)	transboundary pollution, transfrontier pollution,
QUANTITY (of substance)	intensive air pollution
UNDESIRABLE EFFECTS	oxygen depletion pollution, thermal pollution

As can be observed in Table 1, this underspecified meaning of *pollution* is a rich source of possibilities since it predicts the subclasses of MWTs that can designate more specific types of pollution. This allows translators to grasp the different dimensions of pollution or perspectives from which the pollution process can be envisaged.

Knowledge of the types of entity that can fill those slots facilitates understanding of MWTs. This is important because in such cases, syntax cannot be used to clarify meaning. This is evident in compounds such as *water pollution* and *oil pollution*. Despite the fact that *water pollution* and *oil pollution* possess the same syntactic structure (N+N) and even combine the general semantic categories of LIQUID and PROCESS, they obviously differ in the semantic relation between modifier and head, as reflected in their definition slots. This means that

water is the affected entity or patient of pollution, whereas *oil* is the polluting agent (Cabezas-García and León-Araúz 2018).

Even though *water* and *oil* belong to the same semantic category of LIQUID, the accurate interpretation of *water pollution* and *oil pollution* depends, among other things, on the conceptual distinction between INGESTIBLE LIQUID and NON-INGESTIBLE LIQUID as well as the functions of both. *Water*, which is ingestible and necessary for life, is highly sensitive to pollution. In contrast, *oil*, which is non-ingestible and used as a fuel, can have a negative impact on water since oil is a polluting agent that destroys marine life.

This is a basic example of the general knowledge that users must be able to access and activate for an accurate interpretation of both terms. Having this information available at some level signifies that at least a partial representation of semantic structure must be encoded, and enriched by pragmatic information. Syntax and surface form is not sufficient (Štekauer et al. 2012; Buendía Castro and Faber 2016).

For example, in scientific and technical translation from English into another language, the translator does not generally possess the same level of expert knowledge as the source-language text receivers. When the translation is from English into morphologically-rich languages such as Spanish or French, where noun-stacking is not an option, it is necessary to make the relations between MWT components explicit, usually in the form of adjective or prepositional postmodification (Maniez 2009; Daille 2017). In the case of Spanish, the translation of *water pollution* would be *contaminación del agua* whereas *oil pollution* would be translated as *contaminación por hidrocarburos*. The prepositions *de* [of] and *por* [by] are used to encode the conceptual relations implicit in the English MWT.

3.2. Multi-word term level to phrase level

Multi-word terms (MWTs) are also characterized by concealed propositions that can be inferred in the term-formation processes (Levi 1978). This means that MWTs can also be further expanded, especially since many of these terms are the result of predicate deletion (*transboundary pollution* instead of ‘pollution crosses boundaries’) or predicate nominalization (*chemical water pollution* instead of ‘chemicals pollute water’). Both of these term-formation processes have predicate-argument structure.

As is well known, ‘predicate argument structure’ refers to the lexical representation of argument-taking lexical items (Levin 2013). These are typically verbs and their nominalizations. The specification of argument structure involves identifying the number of arguments that a lexical item can take, their syntactic expression, and their semantic relation to the predicate.

Although syntactic expression is language-specific, semantic relations are not. Semantic relations can be understood as semantic roles such as AGENT, PATIENT,

INSTRUMENT, EXPERIENCER, LOCATION, etc. Although most linguists tend to believe that semantic roles are a good idea, at least in some form, there is considerable disagreement as to their number, nature, and function. Currently, there are as many inventories of semantic roles as there are theories that use them (Van Valin and LaPolla 1997; Gildea and Jurafsky 2002; Fillmore et al. 2003; Palmer et al. 2005).

If we take a look at the argument structure of *pollute*, it would have the same number and semantic type of arguments as its correspondences in different languages (i.e. *polluer*, *verschmutzen*, *contaminar*, *inquinare*, *polua*, etc.). In all language-cultures, *pollute* is characterized by a polluting agent as well as a polluted (or affected) entity. The propositional representation of *pollute* is thus a type of *tertium comparationis* that can be used as the basis for semantic equivalence (Buendía Castro and Faber 2016). In fact, this type of representation and information is used, at least in some form, in various machine translation applications. One way of extracting these arguments, their semantic classes, and their combinations is by corpus analysis.

In our study of *pollution*, the corpus used for the extraction of linguistic information was the EcoLexicon English Corpus (over 54,000,000 words), which was subsequently validated by the English TenTen corpus (EnTenTen) of Internet texts, compiled by Lexical Computing. This English corpus contains over 19 billion words and is tagged with TreeTagger using the UTF-8 parameter file. The linguistic information was automatically extracted with the Sketch Engine application (www.sketchengine.eu). One of its most useful functionalities is the word sketch, which is an automatic corpus-derived summary of a word's grammatical and collocational behavior (Kilgarriff et al. 2014).

Based on the corpus information extracted from concordances of *pollute* and its different forms, Table 2 shows that the most frequent polluting agents or contaminants belong to the semantic categories of HUMAN ACTIVITY, INDUSTRY, WASTE, CHEMICAL, GAS EMISSION, VEHICLE, and MICROORGANISM.

In contrast, the second argument, which is the polluted entity, consists of different specifications of AIR, WATER, and SOIL.

Table 2. Semantic classes of the arguments of *pollute*

ARG 1 Polluting Agent	Contaminant
Human activity	[Activity] fracking, drilling, mining
Industrial location	[Location] factory, power plant, mine
Waste	[Solid] garbage, landfill, sludge [Liquid] effluent, wastewater, runoff
Chemical	[Element] mercury, carbon, nitrogen, phosphorus [Natural mixture] coal, oil, petroleum [Artificial mixture] pesticide, fertilizer
Gaseous emission	[Industrial source] gases, fumes

ARG 1 Polluting Agent	Contaminant
	[Vehicle source] exhaust
Vehicle	[Land vehicle] car, diesel vehicle [Water vehicle] container ship, oil tanker [Air vehicle] aircraft, jet
Microorganism	bacteria
POLLUTES	
ARG 2 Polluted Entity	Environmental element/location
Environment	environment
Water	[Water] water, groundwater, drinking water [Water body] aquifer, river, ocean, stream, creek, watershed, lake
Air	[Gas] air, airwaves, atmosphere
Soil	[Soil] land, soil, ground, Earth

What is important is not the syntactic realization of the predicate and its nominalization, but rather the combination of semantic roles and categories, which reflect the polluting activities of the human race (since the implicit agent is human) as well as the three main environmental spheres (air, water, and soil) where pollution occurs. Consequently, the frame is generated by this combination of semantic roles and categories, in this case, POLLUTING AGENT (CONTAMINANT) and POLLUTED ENTITY (ENVIRONMENTAL ELEMENT/LOCATION) and the relation between them.

3.3. From phrase level to frame level

The understanding of phrases in specialized language depends on the reader's ability to expand them so that they fit within a wider context or frame. The problem is that frames are slippery customers. Everyone talks about them but examples are rarely provided, except for the much-used example of the commercial transaction (Fillmore 1982). However, frames also exist in specialized language and can be specified for the knowledge domains, such as the environment (Faber 2012, 2015).

Generally speaking, a frame is a type of mental representation, involving the organization of knowledge about a concept or a set of related concepts. The elements within a frame are linked by different types of semantic relation (Minsky 1975; Fillmore 1985, 2006; Faber 2012, 2015).

The specification of a specialized language frame is the description of a space and the events that occur within it as well as the entities that participate in those events. Busse (2012) makes the useful distinction between concept frames and

predicative frames. Concept frames mostly refer to concepts designated by nouns and noun phrases. Concept frames represent the attributes and properties of an entity. As such, they provide a general format for the representation of categories and category structure (Barsalou 1992). In contrast, predicative frames describe actions and processes, which are designated by verbs and their nominalizations. They represent events and states of affairs in terms of their situation types and participants.

Evidently, predicative frames are more relational since they are composed of various concepts. For this reason, they are the most useful for text understanding. They not only arise from single verbs but also from general configurations of verb meaning that converge in a single semantic space. In specialized language, this sounds strange because verbs are rarely regarded as terms, and thus usually not included in specialized knowledge resources (L'Homme 1998; Buendía Castro 2013). However, general language verbs are crucial to meaning because they are generally what relate concepts in specialized texts.

For example, of the 703 most frequent verbs in the EcoLexicon corpus of over 54 million words, only 10 verbs have no general language meaning (*denitrify*, *flocculate*, *hybridize*, *mineralize*, *nuclate*, *oxygenate*, *photosynthesize*, *solubilize*, *subduct*, and *supercool*). The other verbs are general language verbs (e.g. *accumulate*, *increase*, *develop*, *produce*, *vanish*, *pollute* etc.), which are also used in specialized texts with terms as their arguments. Their meaning underlies what happens in the environment and how we talk about it.

Even though verbs (especially general language verbs) have never been regarded as important in Terminology, they reflect how environmental entities interact. These verbs represent what in our opinion are conceptual invariants, which are present in the majority of documented language-cultures. The existence of such unique beginners or semantic near primitives that are lexicalized in most languages is a constant in the work of linguists such as Ana Wierzbicka, George Miller, and Juri Apresjan, *inter alia*. This culturally shared knowledge, stored in the lexicon, is composed of stable points of reference that comprise a cognitive map of our phenomenological universe.

In previous research within the framework of the Lexical Grammar Model, Faber and Mairal (1999) analyzed and categorized the semantic and syntactic structure of 12,000 general language verbs, first in English and subsequently in Spanish. This resulted in the following general lexical domains: EXISTENCE (*be*, *happen*), CHANGE (*become*, *change*), POSSESSION (*have*), SPEECH (*say*, *talk*), EMOTION (*feel*), ACTION (*do*, *make*), COGNITION (*know*, *think*), MOVEMENT (*move*, *go*, *come*), PHYSICAL PERCEPTION (*see*, *hear*, *taste*, *smell*, *touch*), MANIPULATION (*use*), CONTACT/IMPACT (*hit*, *break*) and POSITION (*to put*, *to be*). Other classes included LIGHT, SOUND, BODY FUNCTIONS, WEATHER, etc.

Faber and Mairal (1999) used this inventory of verb classes to classify the most general environment-related actions and processes in lexical domains derived from definition factorization, as described in the Lexical Grammar Model. This

highlighted the most prominent actions and processes within the environment as well as the semantic categories of the typical participants in these event frames.

For example, when the 703 most frequent verbs in the EcoLexicon corpus were analyzed, the majority were found to belong to the lexical domains, dimensions, and subdimensions of CHANGE, MOVEMENT, EXISTENCE, POSSESSION, POSITION, IMPACT, and MANIPULATION. Table 3 shows some of the verbs that belong to these lexical domains.

Table 3. Organization of verbs in lexical domains

Lexical domain	Verb examples
CHANGE [to become/change]	<i>abate, accrete, aggravate, ameliorate, clarify, decrease, deform, enrich, pollute, contaminate, etc.</i>
MOVEMENT [to move]	<i>whirl, vibrate, topple, thrust, submerge, spiral, stir, shake rotate, etc.</i>
EXISTENCE [to be/exist]	<i>prevent, produce, obliterate, originate, occur, interrupt, initiate, inhibit, inhabit, etc.</i>
POSSESSION [to have]	<i>absorb, catch, cling, collect, conserve, cumulate, drain, entrap, exchange, grab, grasp, harvest, etc.</i>
POSITION [to be in a state/place/position]	<i>block, cover, dump, embed, encase, encrust, envelop, juxtapose, lodge, plug, replace, etc.</i>
MANIPULATION [to use]	<i>burn, consume, exert, expend, exhaust, ignite, irradiate, recycle, tape, use, utilize, etc.</i>
IMPACT [to hit/break]	<i>thresh, strike, slam, shatter, rupture, etc.</i>

Notably absent was the (frequent) use of verbs belonging to the areas of FEELING, SENSORY PERCEPTION, and SPEECH. What is even more interesting is that the verbs in the same lexical domain tended to combine with specialized knowledge units in the same or similar semantic classes such as LIQUID SUBSTANCE, SOLID SUBSTANCE, CHEMICAL ELEMENT, WEATHER EVENT, LANDFORM, WATERBODY, etc.

Faber and Mairal (1999) highlighted the fact that one of the most important environmental processes is CHANGE. CHANGE is a lexical domain with a number of dimensions, which are specific to variation in parameters of time, space, and evaluation (e.g. *to become better, to become worse, to become larger, to become smaller, etc.*). *Pollute*, for example, belongs to this lexical domain. A segment of the lexical domain of CHANGE (*To cause something to become worse*) is shown in Table 4.

Table 4. Segment of the lexical dimension *To cause sth to become worse* in the lexical domain of CHANGE

<p>LEXICAL DOMAIN OF CHANGE TO CAUSE STH TO BECOME WORSE (IMPURE/DANGEROUS/UNCLEAN)</p> <p><i>contaminate</i> to cause sb/sth to become worse by making it less pure.</p> <p><i>pollute</i> to contaminate sth (esp. water/air/soil), by adding a harmful substance to it so that it is dangerous to the health of living organisms.</p> <p><i>adulterate</i> to pollute sth (esp. food products) by adding sth that lowers quality esp. to defraud the user.</p> <p><i>taint</i> to pollute (sth) so that it is spoiled or damaged.</p> <p><i>poison</i> to contaminate sb/sth by adding a harmful substance to it so that it will die or make others die.</p> <p><i>infect</i> to contaminate sb/sth with disease-producing organisms.</p>
--

As can be observed, *pollute*, *poison*, and *infect* are hyponyms of *contaminate*, which is the most general term in this subdimension. The difference between *pollute*, *poison*, and *infect*, lies in the polluting substance or what is polluted. When the semantic (and syntactic) characteristics of the verbs are also specified, this type of lexical organization codifies the range of choices available to each speaker in the lexicalization of a given area of meaning.

The assumption here is that verbs within the same lexical subdimension have a similar syntax and, even more important, combine with the same semantic types of argument. In the case of specialized language, the polysemy of these general language verbs is limited because the scope of their meaning is restricted to the field of Environmental Science. However, verb meaning is not restricted by syntax, but rather the nature of their arguments, which belong to a set of specific conceptual categories such as LANDFORM, CHEMICAL ELEMENT, ATMOSPHERIC PHENOMENON, WATER BODY, PLANT, etc.

The POLLUTION frame can also be further extended to include verbs that codify the remedy for pollution, in this case, a cleaning action in the form of the polysemic general language verb *flush*. Depending on whether there is a focus on liquid movement (*flowing*) or the result (*cleansing*), it is a member of the lexical domain of MOVEMENT or CHANGE. Although *flush* is polysemic, it only has one meaning in Environmental Science. The semantic nature of its arguments is what restricts its meaning to movement in a liquid medium. Its definition is the following:

(2)

flush to cause a **liquid** to flow into/through [MOVEMENT] a **place**, cleaning it [CHANGE] of **something**.

It thus activates a frame with three arguments or participants: (i) a liquid; (ii) a place; and (iii) an (undesirable) substance. In the EcoLexicon corpus, these argument slots are filled by the following terms in the following semantic classes, as illustrated in Table 5.

Table 5. Terms and semantic classes that can fill the argument slots of *flush*

<i>flush</i> to cause a liquid (usu. water) to flow into/through a place , cleaning it of something		
Argument 1: Liquid	Argument 2: Place	Argument 3: Substance
WEATHER EVENT → storms/rainfall	WATER BODY → ENCLOSED → lagoon, pond, lake SEMI-ENCLOSED → estuary, harbor, basin, embayment, river, bay OPEN → beach, channel, slope, reef	SOLID SUBSTANCE → sediment, sand, littoral material
WATER → water COMPOSITION → freshwater, saltwater, salt brine, seawater VELOCITY → water cascades QUANTITY → flood		LIQUID SUBSTANCE → soil water, acid, dissolved metals
SEA/OCEAN MOVEMENT → tides, tidal currents, tidal action		CHEMICAL ELEMENT → magnesium, sodium
		HARMFUL SUBSTANCE → pollutant, contaminant, organic matter, harmful salts, acid, dissolved metals

As can be observed, each definitional slot is potentially filled by a specific set of semantic types and subtypes. In this sense, each argument generates a mini-ontology. The frame activated pertains to water movement into a water body, resulting in the cleansing of that place of a usually harmful substance. In this sense, *flush* is a predicate that is related to pollution, and which provides a subframe that relates clusters of semantic categories that represent entities in the environment.

Evidently, this type of context specification enhances understanding since it identifies the types of entity that participate in events. The focus here is on the actions and processes designated by verbs. As previously mentioned, when specialized knowledge units fill their respective argument slots, the meaning of these general language verbs is constrained by the semantic categories of their arguments. This highlights the relational potential of predicative frames and their usefulness for specialized knowledge acquisition.

4. Conclusion

In this paper we have described how knowledge acquisition can be conceived as a progressive expansion of meaning, which begins at the term-level, progresses to the phrase level, and finally results in the codification of an entire knowledge

frame. In this sense, the definition of a single term can predict how its meaning can be potentially specified in MWTs (Cabezas-García and Faber in press). MWTs that designate processes can be represented in terms of their predicate-argument structure. The importance of exploring semantic types and their combinations cannot be overstressed because semantics, rather than syntax, is what can disambiguate MWTs and phrases in specialized texts (Buendía Castro and Faber 2016). This was also evident in the analysis at the phrase and frame levels, where semantic categories and roles were found to be the basis for knowledge activation.

As reflected in our analysis of *pollution*, it is also necessary to take a closer look at the semantics of general language verbs in specialized texts. They show how specialized knowledge units are combined and encode the basic activities, processes and events in a specialized domain. The specification of context is a way of clarifying the meaning of the terms in a text. The examples given highlight the usefulness of using language as a conceptual mirror that reflects how specialized knowledge is structured and configured.

References

- Anderson, John R. 1983. *The Architecture of Cognition*. Hillsdale: Lawrence Erlbaum.
- Binder, Jeffrey R. and Rutvik H. Desai. 2011. The neurobiology of semantic memory. *Trends in Cognitive Sciences* 15 (11). 527-536.
[Online]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3350748/>.
[Accessed: 22 April 2018].
- Buendía Castro, Miriam. 2013. *Phraseology in Specialized Language and its Representation in Environmental Knowledge Resources*. PhD Thesis. Granada: University of Granada.
- Buendía Castro, Miriam and Pamela Faber. 2016. Phraseological Correspondence in English and Spanish Specialized Texts. In Gloria Corpas Pastor (ed.), *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives = Fraseología computacional y basada en corpus: perspectivas monolingües y multilingües*, 391-398. Geneva: Tradulex.
- Busse, Dietrich. 2012. *Frame-Semantik: Ein Kompendium*. Berlin and Boston: De Gruyter.
- Cabezas-García, Melania and Pamela Faber. In press. Phraseology in specialized resources: An approach to complex nominals. *Lexicography*.
- Cabezas-García, Melania and Pilar León-Araúz. 2018. Towards the Inference of Semantic Relations in Complex Nominals: a Pilot Study. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, 2511-2518. Miyazaki: ELRA.
- Daille, Béatrice. 2017. *Term Variation in Specialized Corpora: Characterisation, automatic discovery and applications*. Terminology and Lexicography Research and Practice, vol. 19. Amsterdam and Philadelphia: John Benjamins.
- Dubuc, Robert and Andy Lauriston. 1997. Terms and contexts. In Sue Ellen Wright and Gerhard Budin (eds.), *Handbook of Terminology Management: Basic Aspects of Terminology Management*, vol. 1, 80-87. Amsterdam and Philadelphia: John Benjamins.
- Faber, Pamela. (ed.) 2012. *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin and New York: Mouton de Gruyter.

- Faber, Pamela. 2015. Frames as a framework for terminology. In Hendrik J. Kockaert and Frieda Steurs (eds.), *Handbook of Terminology*, vol. 1, 14-33. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Faber, Pamela and Pilar León-Araúz. 2016. Specialized knowledge representation and the parameterization of context. *Frontiers in Psychology* 7 (196). 1-20. [Online] Available from: <https://www.frontiersin.org/articles/10.3389/fpsyg.2016.00196/full>. [Accessed: 19 April 2018].
- Faber, Pamela and Ricardo Mairal Usón. 1999. *Constructing a Lexicon of English Verbs*. Berlin: Mouton de Gruyter.
- Faber, Pamela, et al. 2014. Neural Substrates of Specialized Knowledge Representation: An fMRI study. *Revue française de linguistique appliquée* 19 (1). 15-32. [Online] Available from: <https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2014-1-page-15.htm>. [Accessed: 20 April 2018].
- Fernández-Domínguez, Jesús. 2016. A morphosemantic investigation of term formation processes in English and Spanish. *Languages in Contrast* 16 (1). 54-83.
- Fillmore, Charles J. 1982. Frame Semantics. In The Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm*, 111-137. Seoul: Hanshin.
- Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 6 (2). 222-254.
- Fillmore, Charles J. 2006. Frame Semantics. In Dirk Geeraerts (ed.), *Cognitive Linguistics. Basic readings*, 373-400. Berlin and Boston: De Gruyter.
- Fillmore, Charles J., Johnson, Christopher R. and Miriam R. L. Petruck. 2003. Background to FrameNet. *International Journal of Lexicography* 16 (3). 235-250.
- Gallese, Vittorio and George Lakoff. 2005. The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology* 22 (3-4). 455-479.
- Gildea, Daniel and Daniel Jurafsky. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics* 28 (3). 245-288. [Online] Available from: <https://dl.acm.org/citation.cfm?id=643093>. [Accessed: 22 April 2018].
- Keckes, Istvan. 2014. *Intercultural Pragmatics*. Oxford and New York: Oxford University Press.
- Kiefer, Markus and Friedemann Pulvermüller. 2012. Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex* 48 (7). 805-825.
- Kilgarriff, Adam et al. 2014. The Sketch Engine: ten years on. *Lexicography* 1 (1). 7-36.
- Levi, Judith. 1978. *The Syntax and Semantics of Complex Nominals*. New York: Academic Press.
- Levin, Beth. 2013. Argument Structure. In Mark Aronoff (ed.), *Oxford Bibliographies in Linguistics*. New York: Oxford University Press.
- L'Homme, Marie-Claude. 1998. Le statut du verbe en langue de spécialité et sa description lexicographique. *Cahiers de lexicologie* 73 (2). 61-84.
- Lyons, John. 1995. *Linguistic semantics: an introduction*. Cambridge and New York: Cambridge University Press.
- Maniez, François. 2009. L'adjectif dénominal en langue de spécialité: étude du domaine de la médecine. *Revue française de linguistique appliquée* 14 (2). 117-130. [Online] Available from: <https://www.cairn.info/revue-francaise-de-linguistique-appliquee-2009-2-page-117.htm>. [Accessed: 19 April 2018].
- Meteyard, Lotte et al. 2012. Coming of age: a review of embodiment and the neuroscience of semantics. *Cortex* 48 (7). 788-804.
- Minsky, Marvin. 1975. A framework for representing knowledge. In Patrick H. Winston (ed.), *The Psychology of Computer Vision*, 211-277. New York: McGraw-Hill.
- Nakov, Preslav. 2013. On the interpretation of noun compounds: Syntax, semantics, and entailment. *Natural Language Engineering* 19 (3). 291-330. [Online] Available from: <https://pdfs.semanticscholar.org/08af/6f1481ca99a07f01b5dfde042ea00d1bfb8b5.pdf>. [Accessed: 20 April 2018].

- Palmer, Martha, Gildea, Daniel and Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics* 31 (1). 71-106. [Online] Available from: <https://dl.acm.org/citation.cfm?id=1122628>. [Accessed: 20 April 2018].
- Patterson, Karalyn, Nestor, Peter J. and Timothy T. Rogers. 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience* 8. 976-987.
- Quillian, Ross. 1969. The teachable language comprehender. *Communications of the ACM* 12 (8). 459-476.
- Sager, Juan C., Dungworth, David and Peter F. McDonald. 1980. *English Special Languages. Principles and Practice in Science and Technology*. Wiesbaden: Brandstetter Verlag.
- Štekauer, Pavol, Valera, Salvador and Livia Körtvélyessy. 2012. *Word-formation in the world's languages: a typological survey*. Cambridge: Cambridge University Press.
- Temmerman, Rita. 2013. Primary and secondary term creation and the process of understanding. Presentation given at the international seminar *Application of Cognitive Terminology Theories in Terminology Management*, Zagreb, 27-28 September 2013. [Online]. Available at: http://struna.ihj.hr/uploads/act/Temmerman_Zagreb_2013.pdf. [Accessed: 17 April 2018].
- Van Valin, Robert D. and Randy J. LaPolla. 1997. *Syntax: Structure, Meaning, and Function*. Cambridge: Cambridge University Press.