Process-oriented terminology management in the domain of Coastal Engineering[1]

PAMELA FABER*
SILVIA MONTERO MARTÍNEZ**
MARÍA ROSA CASTRO PRIETO*
JOSÉ SENSO RUIZ*
JUAN ANTONIO PRIETO VELASCO*
PILAR LEÓN ARAÚZ*
CARLOS MÁRQUEZ LINARES*
MIGUEL VEGA EXPÓSITO*
**University of Granada*
***Universty of Valladolid*

Abstract
This article describes the theoretical premises and methodology presently being used in the development of the PuertoTerm database on Coastal Engineering. In our project there are three foci, which are highly relevant to the elaboration of lexicographic and terminological products: (1) the conceptual organization underlying any knowledge resource; (2) the multidimensional nature of conceptual representations; and (3) knowledge extraction through the use of multilingual corpora. In this sense we propose a frame-based organization of specialized fields in which a dynamic, process-oriented event frame provides the conceptual underpinnings for the location of sub-hierarchies of concepts within a specialized domain event. We explain how frames with semantic and syntactic information can be specified within this type of framework, and also discuss issues regarding concept denomination and terminological meaning, based on the use of definitional schemas for each conceptual category. We also offer a typology of images for the inclusion of graphic information in each entry, depending on the nature of the concept.

Keywords: terminology, terminography, dictionary, Coastal Engineering, Frame Semantics, conceptual representation, database

1. Introduction
PuertoTerm is a research project whose principal objective is the generation of terminological resources for the representation of specialized concepts in the domain of Coastal Engineering. The product envisioned is a suite of trilingual glossaries, dictionaries, and multimedia knowledge bases for this specialized field in Spanish, English and German.

In our project there are three foci, which are highly relevant to the elaboration of lexicographic and terminological products: (1) the conceptual organization underlying any knowledge resource; (2) the multidimensional nature of conceptual representations; and (3) knowledge extraction through the use of multilingual corpora.

The design and creation of a terminological database for a specialized domain is extremely complex, since this type of resource is the starting point and the basis for any type of dictionary, thesaurus or glossary. For this reason, it is useful to first evaluate the quality of available resources in order to have a realistic idea of how to develop an information system that will meet the needs of potential users. Such resources are now quite easy to find on the Internet, but they vary greatly in quality and reliability.

Since PuertoTerm targets Coastal Engineering, our interest naturally lay in the analysis of the specialized resources available in Earth Sciences (Edaphology, Hydrology, Meteorology, etc.). To this end, we obtained a sample of 113 terminological thesauri, encyclopaedias, glossaries and dictionaries that included at least one of the following languages: English, French, German and Spanish. We then identified the principal characteristics of each compendium of terms, and converted this information into quantifiable data.

The results obtained show that the number of online lexicographic resources in the field of Earth Sciences is very high, but that many do not meet even minimum standards of quality. In terms of their content, dictionaries and glossaries tend to be limited to one specific area (e.g., volcanoes, glacial geomorphology) whereas thesauri cover wider knowledge areas (e.g., Earth Sciences).

The size of the resources in the sample ranged from 18 to 2,000 term entries. Most organized terms alphabetically, though all of the thesauri in the sample naturally had a conceptual organization. Surprisingly, there were some resources that displayed terms randomly, with no sort of organization whatsoever. Some of these products were monolingual (English, French or German), whereas those that were multilingual included from two to twenty languages. However, not all languages received the same treatment. The majority of the multilingual resources were multilingual only to the extent that they indicated equivalents of the main term in other languages.

The information in each entry was usually very limited, and did not include any syntactic or usage data. Synonyms were not represented in a consistent way. Sometimes they were included in the main entry term, and other times they appeared in separate entries without any reference to the more frequently used term. We also found term definitions to be extremely concise, sometimes to the detriment of including necessary conceptual description. Whereas the descriptive information in entries was practically always relevant, it was often presented chaotically. Needless to say, definitions of conceptually related terms showed no overt interrelation, and there was a clear lack of systematicity in definition formats.

Of the lexicographic products studied, 31% included only a list of terms with no other information, while 57% gave descriptive information for each term in text form. Although 12% of the products offered visual information (images, photos, diagrams, etc.), these images did not always serve the purpose of enriching the conceptual description, since in 30% of the cases they were not relevant or useful to the user.

Only 27% of the resources analyzed made use of the possibility offered by the Internet of establishing hyperlinks to other terms within the same document. Similarly, only 13% offered links to other websites for additional information regarding the term.

The results of this mini-study point to the fact that there is a great need for terminological databases constituting authentic information systems that highlight conceptual interrelatedness and underline the multidimensionality of specialized concepts (Bowker and Meyer 1993; Bowker 1997). It goes without saying that in order to develop such a terminological product, important theoretical decisions must be made concerning the underlying structure of concepts.

2. Conceptual structure and terminology

The specification of the conceptual structure of specialized domains is a crucial aspect of terminology management. The way concepts are represented affects the configuration of information within individual terminological entries and the contents of each data field, especially in regards to individual conceptual description or definition.

For this reason, conceptual representation should be more than a hierarchical list of objects linguistically translated into either simple or compound nominal forms. Such conceptual hierarchies are static configurations unless conceived as part of a more dynamic structure. More specifically, we are referring to the representation of a prototypical domain event, which provides a frame for the basic processes that take place within the specialized field. Within this context, concepts are organized around an action-environment interface (Barsalou 2003: 513; Faber et al. 2005). The name that we give to this type of approach is *Process-Oriented Terminology Management*.

The description of specialized domains is based on the events that generally take place in them, and can be represented accordingly (Grinev and Klepalchenko 1999). In fact, each knowledge area can be said to have its own event template. As a result, part of the understanding of a specialized field can be described in terms of basic entities linked by different types of conceptual relations. Depending on the knowledge field, there is a Medical Event, Automotive Event, Engineering Event, etc. Repeated actions or events are easily apprehended by our conceptual system, and therefore form the basis for conceptual event structures that are in turn used to deal with new events and actions, which are categorized within the context of such idealized frames. The construction of an event template for a specialized knowledge domain helps the user to process its conceptual content more easily.

Basic concepts of object, event, and attribute are also the starter categories for any type of terminology management, which must first be carried out to facilitate understanding.

Terminology management means configuring terms with essential information and in meaningful arrangements or relations to each other so that the user can process them, absorb the knowledge they transmit, and remember them better (Faber and Tercedor Sánchez 2001). A key element in this approach is flexibility, since events are dynamic and real-world entities can play different roles in them.

3. Frames and specialized language

In the last decade, specialized language and general language studies have begun to show an approximation in perspective, as lexicographers and lexical semanticists now tend to take a more conceptual approach towards dictionary structure and terminographers have begun to pay more attention to conceptual description or to the definition of terminological units.

Frame Semantics (Fillmore 1982, 1985; Fillmore and Atkins 1992) is an example of this change in lexicography as evidenced in the lexical tools and database developed within the FrameNet Project (Johnson et al. 2002; Fillmore et al. 2003). As is well known, Frame Semantics asserts that in order to truly understand the meanings of words in a language, one must first have knowledge of the semantic frames or conceptual structures that underlie their usage. Frames are a type of cognitive structuring device based on experience that provide the background knowledge and motivation for the existence of words in a language as well as the way those words are used in discourse.

Significant aspects of the FrameNet approach can also be applied to specialized language. We have found this to be true of biomedical domains such as Oncology as well as technical ones such as Coastal Engineering (Faber 2003).

3.1 The Coastal Engineering Frame (CEE)

The elaboration of frames in the Coastal Engineering Event (CEE) that we have created as part of our project is based on the evidence provided by the corpus devised for PuertoTerm. For this purpose we compiled a corpus of English, Spanish and German texts specific to Coastal Engineering and other related domains. These texts were extracted from reliable Internet sites, specialized electronic journals and extensive segments of course books.

Table 1 presents the number of tokens and types as well as their ratio in the English and Spanish corpus.[2]

Table 1. Corpus description[3]

	English	Spanish
Tokens	4,435,525	5,075,774
Types	68,685	115,558
Type/Token ratio	1.55	2.28

Although there are more texts in English than in Spanish, a great effort was made to build corpora of comparable size. Extensive use made of both corpora has shown that the relatively small difference in size poses no problems as regards the scope and usefulness of the information yielded in both languages.[4]

The text selection criteria for the PuertoTerm corpus were quantity, quality, simplicity and documentation, as recommended by EAGLES (1996) (Pérez Hernández 2002). Quantity refers not only to the size of the corpus itself but also to the fact that texts should be complete. Quality refers to the prestige of the author within the field as well as the value of the work itself. Simplicity encompasses the amount and the kind of information which is attached to each text. This is particularly usef Apart from these general criteria, the representativeness of texts was also taken into account. In this sense, an effort was made to include texts that represent communicative situations among experts or non-experts as well as between the two groups, thus accounting for different degrees of specialisation in the corpus. This criterion facilitates the identification of specialised versus non-specialised terms and maximum coverage of all possible occurrences in the field.

The initial phase in this approach was the development of a comprehensive list of terminological units in the corpus. Once this inventory was available, the second phase was the creation of the following frame (Figure 1), in which all concepts represented could fit in a coherent way.ul for the terminologist, since it provides very useful information on text typology and therefore specialisation level. Documentation is intrinsically linked to simplicity, for it refers to the availability of information on the origin and circumstances of production of the texts.

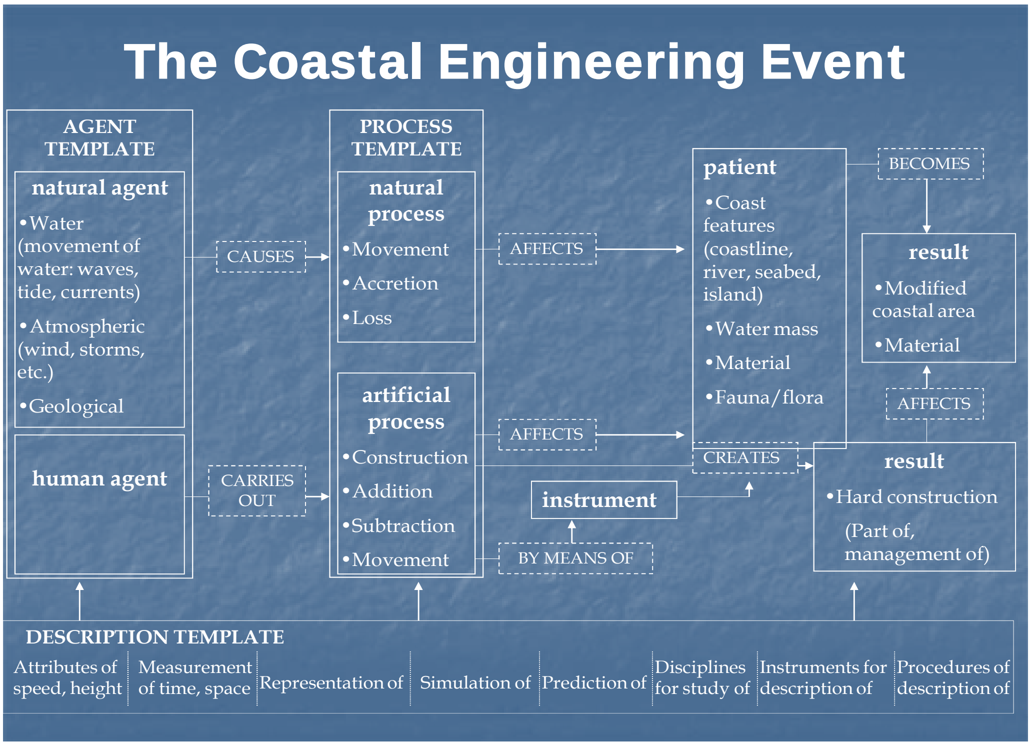

Figure 1. PuertoTerm Coastal Engineering Event

According to the corpus data, the Coastal Engineering Event (CEE) is conceptualized as a process that is initiated by an agent, and which affects a specific kind of patient and produces a result. These macrocategories (AGENT ➔ PROCESS ➔ PATIENT/RESULT) are the concept roles characteristic of the domain, and provide a model for representing their interrelationships. Additionally, there are peripheral categories that include INSTRUMENTS that are typically used during the CEE, as well as a category in which the concepts of measurement, analysis, and description of the processes in the main event are grouped together.

The AGENT, PROCESS and PATIENT/RESULT macrocategories present a parallel structure. The relationship established between entities that are NATURAL AGENTS and the NATURAL PROCESSES caused by them is parallel to the relationship between HUMAN AGENTS and the ARTIFICIAL PROCESSES triggered by the agent. This parallelism extends to their subcategories, which in both cases include processes that involve movement, accretion or loss of materials. Naturally, the CONSTRUCTION subcategory is a typically human process that has no equivalent in natural processes.

In the case of the PATIENT/RESULT category, coastal engineering deals with the highly dynamic processes that take place in the coastal environment. Beaches, cliffs, marshes and plants are affected by both natural and artificial processes, or can even be the result of a natural or artificial process. Thus, a beach is by definition a sediment deposit, and hence, the RESULT of erosion (a natural process of loss). However, at the same time it can be AFFECTED by erosion.

The dynamic nature of natural processes can also be observed in artificial processes. Human processes RESULT in constructions and structures which become part of the coastal environment and, as such, can be AFFECTED by both natural processes and subsequent human processes. Therefore, the same entity which is categorised as the RESULT of a process in one context can be categorised as a PATIENT affected by a different process in another. This is resolved by the inclusion of both AFFECTED and RESULT categories in the same template.

INSTRUMENT refers to the specific machinery and devices that are used as aids to human processes carried out in the coastal environment. The importance of the DESCRIPTION template is underlined by the complexity and diversity of its subcategories, which range from the measurement and representation of the entities involved in the CEE to the instruments that are used for their study or the terms that are used to predict their evolution and interrelationships.

3.2 Specifying a frame: TIDE/MAREA

Once the Coastal Engineering Event is specified, it is necessary to establish the various sub-frames that make up this macro-event. To illustrate this process, the following section offers an explanation of how such a sub-frame is created and specified for the concept of TIDE as designated in Spanish and English.

3.2.1 Conceptual frame components

The first step is the identification of the specialized lexical units (SLUs) characteristic of TIDE. The information contained in the *Glossary of Coastal Terminology* (Voigt 1998) and the *Glosario de Marea y Corrientes* (Servicio Hidrográfico y Oceanográfico de la Armada de Chile 2002) gives us an overview of the process. The analysis of the information in the definitions of *tide* and *marea* points to the following: (i) vertical water movement; (ii) horizontal water movement; (iii) temporal regularity; and (iv) the role of gravitational attraction in this phenomenon. The English entry also specifies that *tide* is only used in relation to vertical water movement, whereas *tidal current* and *corriente de marea* are used to refer to horizontal movement.

To find the SLUs relating to this process, a word list was obtained from the corpus. The English list showed that *current* and units related to *tide*, such as *tide(s)*, *tidal*, *tidally* or *macro-tidal* were by far the most frequent. In the Spanish corpus, the units were *flujo(s)*, *corriente(s)* and different variants of *marea* (e.g., *marea(s)* or *mareal(es)*). All of these units point either to vertical or horizontal water movement in a general way. Yet the wordlists displayed other terms of lesser frequency such as *bore(s)*, *ebb(ing)*, *flooding* or *highwater. Highwater* can be regarded as a hyponym of *tide*, while the other three are hyponyms of *current.*

In Spanish, hyponyms of *corriente* (referring to horizontal movement) are *entrante(s)*, *macareo(s)*, *bores*, *reflujo(s)* and *vaciante(s).* Vertical movement (designated by *marea*) is lexicalized by the terms *bajamar(es)* and *pleamar(es)*. The concordances generated for this set of units give the most frequent collocates within the established horizon of the search word.

Results from the search string TIDE* and MAREA* listed grammatical and syntactic collocates such as *the tide* or *tides are* and the Spanish, *la marea* or *una marea*. There were also conceptual collocates reflecting dependencies between the events and entities and their specific attributes. For example, we found units such as *high* and *low*, possible attributes of TIDE, and lexemes such as *waves*, *sea*, or *currents*, conceptually related to *tide*. In Spanish, units such as *corriente* and *amplitud* appeared within a collocational span of ± 5 words from the search string MAREA*.

The next step is then the identification of the frame elements (FEs). The semantic or conceptual arguments that make up the categorization structure of TIDE can be identified through the analysis of relevant contexts and collocations such as (1) and (2):

> (1) Tidal bores form on rivers and estuaries near a coast where there is a large tidal range.
> (2) El término corriente de marea se aplica exclusivamente a las corrientes periódicas producidas por la marea.

Such sentences revealed units referring to the following entities and processes:

- the astronomical entities, *moon*, *sun* and *Earth*;
- a body of water, such as *sea waters* and *mar*;

- a rising and/or falling water movement, such as *high-water*, *ebb*, *pleamar* and *bajamar*;
- horizontal water movement, such as *ebb*, *current*, *flujos* and *reflujos*;
- coastal forms, such as *coast* and *playa*.

Other units formalize different types of relations among the entities and processes mentioned. A relevant example of this is the gravitational attraction between the moon, sun and Earth reflected in phrases such as *gravitational attraction of the moon and sun acting upon the rotating earth* and *atracción combinada del sol y la luna sobre la tierra*. They also lexicalize different attributes characteristic of the entities and processes mentioned. This can be seen in attributes relating to the frequency of both vertical and horizontal movements in units such as *diurnal*, *periodic*, *semimensualmente* or *período medio de 12,4 horas*.

Finally, within these contexts, units often appear that refer to a series of processes and objects used in the description and/or prediction of the different values of the attributes of the vertical and horizontal movement, such as *tide gage consisting of a vertical graduated staff from which the height of the tide can be read directly* and *la altura de la marea se mide con la ayuda de mareógrafos*. All of this information can be seen in the following TIDE sub-frame (Tables 2 and 3).

Table 2. Frame Elements

▪ SUN (SOL)
▪ MOON (LUNA)
▪ EARTH (TIERRA)
▪ WATER (AGUA)
▪ TIDE (MAREA)
▪ CURRENT (CORRIENTE)
▪ COASTLINE (LITORAL)
▪ ATTRACTION...
▪ FREQUENCY, ...
▪ MAREOGRAPHS, …
▪ PREDICTION, …

Table 3. Description of FEs

▪ AGENT_1
▪ AGENT_2
▪ PATIENT_1
▪ PATIENT_2
▪ MOVEMENT_1
▪ MOVEMENT_2
▪ REFERENT
▪ RELATIONS
▪ ATTRIBUTES
▪ INSTRUMENTS
▪ ARTIFICIAL-PROCESS

Within this frame, the sun and moon are AGENTS attracting a PATIENT, the Earth and its water-mass. This initiates a NATURAL PROCESS, i.e., a vertical displacement or tide, and a horizontal displacement, the current. Both are in relation to a reference point, the coastline. This process is also characterized by a series of relations and attributes that can be described by using different instruments. At this stage, there are no differences between languages, since frames characterize conceptual structures at a very basic level, at which potential cultural and linguistic divergences do not come into play.

3.2.2 Syntactic frame components

Dealing with interlinguistic differences is part of the next phase of the TIDE modelling process, in which the syntactic realizations of the different FEs, as well as their valence patterns or mappings between semantic and syntactic structures, are specified.

For example, for the FE TIDE/MAREA we used units such as *tide(s)*, *marea(s)* and *bajamar(es)* to find fixed recurrent patterns in the corpus. From a list of clusters ranging from two to five lexical units, with a minimum frequency of two, we ended up with clusters such as *high tides*, *high-tide*, *perigean spring tides*, *phase of the tide* or *components of the sea tide*. In Spanish, we identified units such as *marea baja*, *carrera de marea*, *profundidad en marea baja*, and *la amplitud de las mareas*. The annotation of these frequent terminological combinations according to the FEs listed in Table 2 and their grammatical constituents shows that TIDE/MAREA is linguistically represented by structures shown in Tables 4 and 5.

Table 4. Syntactic realizations of TIDE

(1) MOV_1
tide, tides, ebb, flood (**N**)
(2) AGENT (1_2) + MOV_1
lunar tide, solar tide
perigean tide, equatorial tide **(ADJ + N)**
moon tide **(N + N)**
(3) MOV_1 + ATTRIB.
highwater (**N**)
tidal period, tidal range (**ADJ + N**)
(4) ATTRIB. + MOV_1
range of tide, height of the tide (**N + of + (the) N)** diurnal tide, high water, high-tide (**ADJ + N**)

Table 5. Syntactic realizations of MAREA

(1) MOV_1
marea, mareas (**N**)
(2) MOV_1 + AGENT (1_2)
marea lunar, marea solar, marea lunisolar, marea luni-solar (**N + ADJ)**
marea de perigeo, marea de apogeo (**N + de + N**)
(3) MOV_1 + ATTRIB.
pleamar, bajamar (**N**)
marea alta, marea baja (**N + ADJ)**
(4) ATTRIB. + MOV_1
amplitud de marea, rango de la marea (**N + de + (la) N)** amplitud mareal, rango mareal (**ADJ+ N)**

In Table 4, (1) shows that TIDE is designated by simple lexical units referring to MOVEMENT_1. More concretely, it represents vertical movement through general terms such as *tide* or *tides*, and more specific hyponyms such as *ebb* or *flood* (a falling and rising tide, respectively).

Yet, *tide* most frequently co-occurs with other units as shown in (2), which activates elements such as the AGENT (1_2) + MOV_1. This is evidenced in *lunar tides* or *solar tides*, whose syntactic realization is a noun phrase (ADJ + N). This conceptual activation is also seen in units such as *perigean tides* or *apogean tide*, where *perigean* and *apogean* evoke the lunar agent through metonymic extension since both refer to lunar position. In the first case, the moon is at perigee, closest to the Earth; in the second, the moon is at apogee, farthest from the Earth.

Finally, this conceptual combination can be syntactically realized by a noun phrase (N + N), i.e., *moon tide*. Similar activations occur in Spanish; in Table 5, (1) shows that *marea* refers to MOV_1. However, due to different syntactic projections in these languages, differences arise when describing the order of activation of the FEs.

Again in Table 5, (2) shows that the combination of agent and vertical movement has the opposite order from English. In Spanish it is MOV_1 + AGENT (1_2), represented by noun phrases with the following syntactic structures: (i) (N + ADJ) such as *mareas lunares*, *marea solar*, *mareas lunisolares/marea luni-solar*; and (ii) (N + de + N) such as *marea de perigeo* or *marea de apogeo*.

The other examples in Tables 4 and 5 show that TIDE very often occurs in conjunction with its most relevant attributes. This syntactic information of the FE is further enriched with the specification of valence patterns, the semantic and syntactic templates found in a set of sentences where the element occurs.

For example, annotation of the most relevant lexical contexts of TIDE identify a set of sentences that relate the element with the frequency of the phenomenon, as shown in Tables 6 and 7 through the verbal elements *occur* and *tener lugar*, respectively. The underlying grammatical constituents are NP + VP + ADVP.

Table 6. Valence patterns of TIDE

(1a) TIDE *to occur* FREQUENCY
Spring tides occur every 14–15 days.
(NP) (V) (ADVP)
(2a) MOON *to raise* TIDE
The moon raises tides in the ocean.
(NP) (V) (NP)
(3a) TIDE *to create* CURRENT
Tides create currents.
(NP) (V) (NP)
(4a) TIDE *to flood* COASTLINE
The tide floods the beach.
(NP) (V) (NP)

Table 7. Valence patterns of MAREA

(1b) MAREA *tener lugar* FRECUENCIA
La marea de perigeo tiene lugar cada 27,5 días.
(NP) (VP) (ADVP)
(2b) LUNA *hacer crecer/decrecer* MAREA
La Luna hace crecer la marea.
(NP) (VP) (NP)
(3b) MAREA *generar* CORRIENTE
Las mareas generan un flujo periódico desde y hacia el puerto.
(NP) (V) (NP)
(4b) MAREA *inundar* LITORAL
Las pleamares inundan las marismas.
(NP) (V) (NP)

Contexts relating TIDE to astronomical bodies through verb phrases followed by a noun phrase are also quite frequent, as shown in (2a) and (2b) in Tables 6 and 7, where the causative relation between the agent MOON and TIDE is transmitted by predicates such as *raise* or the Spanish *hacer crecer/decrecer*. Sentences in which TIDE acts as an agent causing a process such as a CURRENT were also detected, for example, in the patterns found in (3a) and (3b) with *create* or *generar.* A similar situation is found in (4a) and (4b) where verbs such *flood* and *inundar* describe a process affecting the COASTLINE due to the TIDE.

3.2.3 Frame interrelations

The information obtained not only enriches the description of SLUs within the TIDE frame, but also helps to identify frame relatedness and specify frame description. Thus, the fourth stage of our process deals with the relations TIDE has with other sub-frames that make up the Coastal Engineering Event as well as the specification of the relevant parameters or attributes of the frame.

Through the hierarchical generic-specific relation, IS-A, TIDE is conceptualized as a NATURAL AGENT. More specifically, it is regarded as a type of WATER-MOVEMENT where TIDE appears as subordinate of the WAVE frame (Prieto Velasco 2005). This categorization is grounded in the lexical and conceptual analysis of specialized discourse which reflects Laplaces's *Dynamic Theory of Tides* (Tomczak 2000). According to this theory, tides are gigantic waves travelling around the earth's surface in shallow waters.

As a result, the TIDE frame inherits relations and attributes of the parent concept, such as the CAUSE-EFFECT relation with the agent producing the wave or attributes such as PERIOD, DURATION or LENGTH. Within the Coastal Engineering domain, tides are relevant at the three stages of the process: (i) the initial phase, since the analysis of the astronomical bodies' position is vital for the prediction of tides and tidal currents; (ii) the intermediate phase; and (iii) the final phase, since the rise and fall of tides are instrumental in the modelling of the coastline, and also affect the construction of artificial structures. Therefore, within the Coastal Engineering Event, the tidal phenomenon is the agent that initiates processes affecting other frames within the macro-event. Figure 2 shows the basic underlying structure of the TIDE frame.

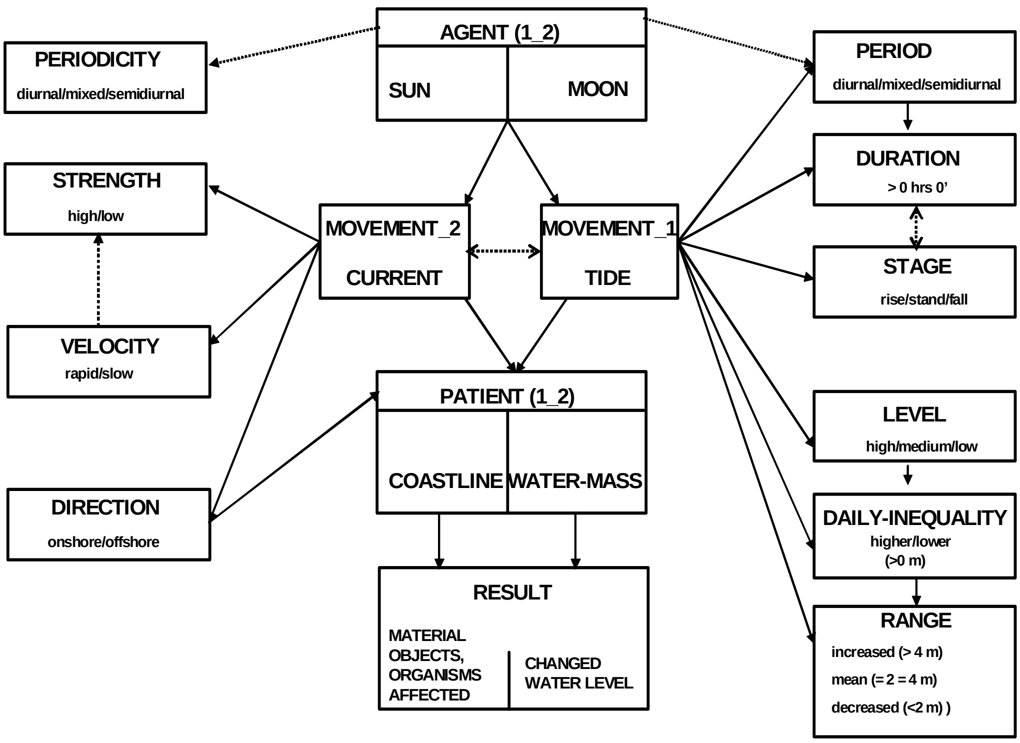

Figure 2. Specification of the TIDE frame

This diagram represents the elements of TIDE. Basically, there are two agents that simultaneously initiate two processes or movements that affect the patients, producing different results. Also specified are some of the attributes of two main elements of the event, CURRENT and TIDE.

The representation of this frame through a schema made up of slots and fillers allows the specification of the relevant information regarding conceptual entities as well as their properties (Tables 8 and 9):

Table 8. Relations of FE TIDE

TIDE SCHEMA: FE TIDE		
IS_A	:	**WAVE: WATER-MOVEMENT**
RELATIONS		
CAUSED-BY	:	ASTRONOMICAL-BODY
AFFECTED-BY	:	ASTRONOMICAL-BODY-POSITION
PRODUCES	:	CURRENT
CHANGES	:	WATER-LEVEL
TRANSPORTS	:	SEDIMENTS
MOVES-INTO	:	COASTLINE
MOVES-OUT-OF	:	COASTLINE
AFFECTS	:	LIVING-ORGANISMS
MEASURED-WITH	:	MEASURING-TEST
MEASURED-WITH-THE-USE-OF	:	INSTRUMENT
PREDICTED-WITH	:	PREDICTION-TEST
PREDICTED-WITH-THE-USE-OF	:	INSTRUMENT

Table 9. Attributes of FE TIDE

TIDE SCHEMA: FE TIDE		
ATTRIBUTES		
AGENT-POSITION	⇒	near (perigee, perihelion) far (apogee, aphelion)
TIDAL-PERIOD	⇒	diurnal (24 hrs 50') mixed semidiurnal (12 hrs 25') > 1 day
DAILY-TIDE-NUMBER	⇒	single/two
TIDAL-CYCLE STAGE	⇒	rise/stand/fall
TIDAL-LEVEL	⇒	high/medium/low
TIDAL- DAILY-INEQUALITY	⇒	higher/mean/lower (>0 m)
TIDAL-RANGE	⇒	increased (> 4 m) mean (≥ 2 ≤ 4 m) decreased (<2 m)
TIDAL-DURATION	⇒	> 0 hrs 0'

On the left side of Table 8, the small capitals indicate the conceptual relations through which TIDE is related to the concepts on the right. In addition to the hierarchical relation IS-A, some non-hierarchical relations, such as CAUSED-BY or PRODUCES, are found. Similarly, the small capitals on the left side of Table 9 refer to the attributes of the FE, such as AGENT-POSITION or TIDAL-PERIOD. Their values, which can be literal or scalar constants, are on the right side. This frame template is regarded as a model for all the concepts formalizing the FE TIDE and constitutes the framework for the terminographical definition of the terms realizing such elements.

The systematic application of the methodology described gives a series of frames, such as TIDE, WAVE or EROSION, which are part of the CEE. It also generates a list of SLUs that evoke these frames together with their lexical contexts and syntactic patterns. Also specified are the set of conceptual structures underlying the SLUs and, consequently, the set of conceptual relations between frames. Finally, it allows the establishment of definitional schemas for the SLUs belonging to each frame.

4. Terminographic definition

According to Bejoint (1997: 19–20), definitions have never been given due importance in Terminology. Traditionally, terminographic definitions have been associated with intensional definitions. In many termbases, definitions are simply inserted in a cut-and-paste fashion from other dictionaries, termbases or knowledge resources without taking into consideration their internal structure and coherence. However, definitions are mini-knowledge-representations, and accordingly, the organization of information encoded in definitions should be structured according to its perceptual salience. However, it is also necessary to take into account the definitions of other related concepts within the same category (Faber 2002). For this reason, terminographic definitions are presently receiving increased attention within Terminology theory.

The socio-cognitive approach (Temmerman 2000) proposes the creation of a definitional template for the description of *units of understanding*. Temmerman distinguishes traditional concepts from units of understanding. If a concept can be logically or ontologically classified, its description will be intensional. In contrast, if concepts show a prototypical structure, they are regarded as units of understanding on the basis of intracategorial and intercategorial information. Rather than restricting definitions to conceptual intension, templates activate broader conceptual structures in the form of frames.

In addition, Strehlow (1993) affirms that the representation of concepts by means of definition statements alone is inadequate for many scientific terms. He highlights the fact that the representation of a definitional structure is comparable to a conceptual representation, and thus will show elements such as genus (referring to the domain or superordinate category to which the concept belongs), species (referring to the sub-domain), differentiae (expressing essential characteristics of the concept that makes it different from others in the same category) and accident (expressing non-prototypical characteristics of the concept).

Fillmore (2003) also refers to the use of frames in his double-decker definitions. Like Strehlow, he says that the user should have access to the conceptual structure that underlies the word, information about the position or orientation of its meaning within that conceptual structure, and a generous supply of examples of how the word interacts with other words and phrases in the utterances of language users.

Similarly the Functional-Lexematic Model (FLM) (Martín Mingorance 1984, 1989, 1995; Faber and Mairal Usón 1999) offers a linguistic way of organizing concepts, through the analysis and construction of definitions based on information provided by specialists as well as other lexicographic and terminological resources. The conceptual structure of a domain is made explicit in the form of terminographic definitions that are both coherent in both their micro and macrostructure (Faber and Tercedor Sánchez 2001).

Table 10. Coastal defense definitional hierarchy

COASTAL DEFENSE STRUCTURE

- **Hard Defense**: impermeable **coastal defense structure** of concrete, timber, steel, and masonry, which reflects a high proportion of incident wave energy.
  - **Bulkhead**: **hard coastal defense structure** or partition to retain or prevent sliding of the land, and also to protect the upland against damage from wave action.
  - **Groin**: **hard coastal defense structure** protecting a shore area, built perpendicular to the shoreline, made of timber, concrete or rubble mound to retard littoral drift and prevent beach erosion.
  - **Breakwater**: **hard coastal defense structure** protecting a shore area, harbor, anchorage, or basin from waves.
    - **Detached Breakwater**: **breakwater** without any coastal connection to the shore.
    - **Headland breakwater**: **breakwater** connected to the shore.

As shown in Table 10 the hierarchical relation IS-A can be derived from the genus of each definition. The genus of the definition indicates that BULKHEAD, GROIN and BREAKWATER are all types of HARD COASTAL DEFENSE STRUCTURE, whereas DETACHED BREAKWATER and HEADLAND BREAKWATER are defined according to the superordinate concept BREAKWATER. Apart from hyponymic information, definitions also offer valuable information in their differentiae, which are the characteristics that distinguish sub-ordinate and coordinate concepts.

The FLM establishes the frame elements (material, function, location, etc.) in a definitional schema that directly refers to the underlying event structure of the domain. In this sense, definitions are conceived as bridges between terms and concepts (Faber and Tercedor Sánchez 2001).

For example, the definition of *groin* shows the definitional schema that will hold for all the concepts within the subdomain (Figure 3):

GROIN hard coastal defence structure [**IS-A**], located perpendicular to the shoreline [**LOCATION-OF**], made of timber, concrete or rubble mound [**MADE-OF**] in order to retard littoral drift and prevent beach erosion [**HAS-FUNCTION**].

↓

GROIN ____________ [**IS-A**], located ________________[**LOCATION-OF**], made of ____________ [**MADE-OF**] in order to ______________________ [**HAS-FUNCTION**].

Figure 3. Definitional schema for GROIN and subtypes of groin

This schema can be considered a type of conceptual grammar or definitional frame that can be used as a format for the description of all types of coastal structures within the domain of Coastal Engineering. As a result, it is vitally important to apply frames at all levels of the elaboration of any sort of terminological product as a means of assuring the consistence of concept representation at both the macrostructural and microstructural levels.

5. Graphic information

Another important issue regarding concept representation in termbases is the need to include information in other formats to enhance textual comprehension. This is particularly important in knowledge areas such as Engineering. Consequently, one of the products presently being generated in our project is a database of images that will be complementary to the linguistic information provided in each term entry.

However, the inclusion of visual information is not a simple thing, since it also entails theoretical considerations regarding the nature of the concept represented as well as the general characteristics of images and their interrelation. Consequently, it was necessary to first explore the role of visual representation in the specification of conceptual structure in

Coastal Engineering. As a first step, we carried out a preliminary study to establish a descriptive model to be applied to images in scientific texts.

We compiled a corpus consisting of 60 multimedia texts on wave typology in English, containing a total of 22,682 words and 119 images. They were collected from educational and scientific websites on the basis of their electronic availability and the number of extratextual graphic elements present in the text. They were saved in plain text format.

An analysis of their salient characteristics led to the resulting typology of images, based on Monterde Rey (2002). The typology first establishes an initial difference between iconic and non-iconic images.

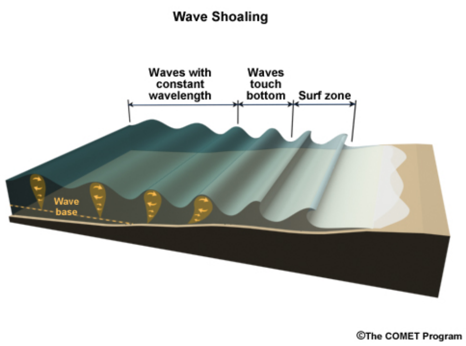

Figure 4. Example of an iconic image

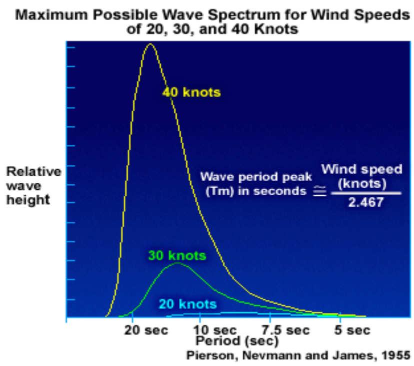

Figure 5. Example of a non-iconic image

As shown in Figures 4 and 5, an iconic image is similar to the entity that it represents, whereas a non-iconic image is not.

These categories were subsequently divided into static and dynamic images (depending on whether movement was represented). Since a textual description often accompanies such images, we used the presence or absence of descriptive text as another criterion for classification. Text may appear within the image, as a caption, or in its immediate context. Each of the twelve image types proposed was assigned an *ad hoc* tag (e.g., <IMG1Adesaout>), which was inserted in the corpus to replace the original image in the .txt file (Table 11).

Table 11. Image typology

IMAGE TAG	DESCRIPTION	FREQ.	%
<IMG1Adesout>	iconic, static, with description outside the image	27	22.69
<IMG1Adesin>	iconic, static, with description inside the image	25	21.00
<IMG2Adesout>	non iconic, static, with description outside the image	23	19.33
<IMG1Asin>	iconic, static, without description	11	9.24
<IMG1Bdesin>	iconic, dynamic, with description inside the image	10	8.40

<IMG2Adesin>	non iconic, static, with description inside the image	9	7.56
<IMG1Bsin>	iconic, dynamic, without description	8	6.72
<IMG1Bdesin>	iconic, dynamic, with description outside the image	2	1.68
<IMG2Asin>	non iconic, static, without description	2	1.68
<IMG2Bdesin>	non iconic, dynamic, with description inside the image	2	1.68
<IMG2Bdesout>	non iconic, dynamic, with description outside the image	0	0.00
<IMG2Bsin>	non iconic, dynamic, without description	0	0.00
TOTAL	119	100	

The results obtained regarding the frequency of each image type are analyzed as follows (Figure 6):

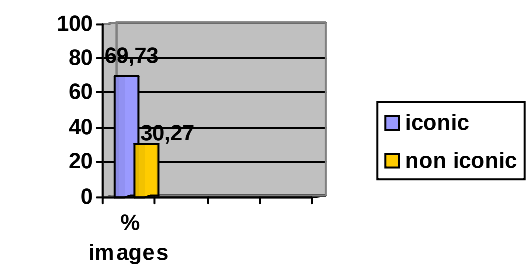

Figure 6. Iconic and non iconic images

In the sample studied, there were 69.73% iconic images and 30.27% non-iconic images. The higher frequency of iconic images is probably due to their greater resemblance to the real-world entity represented. In contrast, non-iconic images bear a greater resemblance to the concept rather than the real-world entity. Consequently, this higher level of abstraction means that the perceiver must possess more background knowledge to process them.

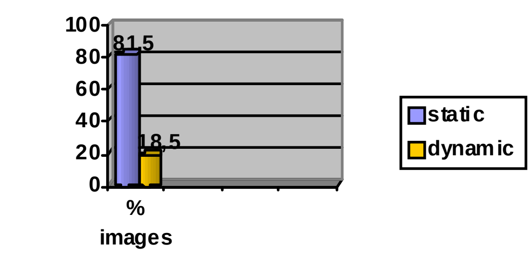

Figure 7. Static and dynamic images

Figure 7 shows that 81.50% of the images are static images, as compared to 18.50% that are dynamic. These numbers point to the fact that there is a technical constraint in multimedia texts, since very large files on the Internet would slow navigation and access to online multimedia resources.

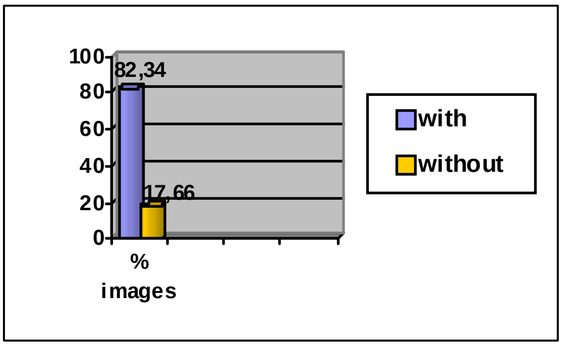

Figure 8. Images with/without text description

As shown in Figure 8, 82.34% of the images were accompanied by text description, whereas 17.66% were not. Such a description is one of the most common means employed to help the user achieve a better understanding of the concept.

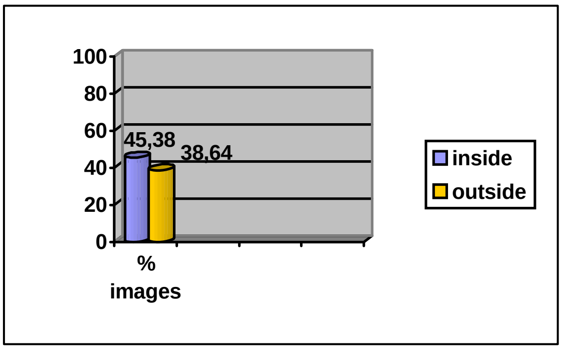

Figure 9. Images with description inside/outside the image

As shown in Figure 9, 45.38% of the images contained an embedded text description and 38.64% had the description within the immediate context, but not inside the image. There seems to be no significant difference in the use of a textual description either inside or outside the image. It appears to depend on other factors such as the size and layout of the image, as well as its colours and contents.

The results of our study indicate that scientific texts often employ images as a non-verbal resource to contribute to the transfer of expert knowledge, particularly to a lay audience. The general characteristics of these images are directly related to their degree of iconicity or similarity to the real-world entity represented and the level of abstraction necessary for a complete understanding. Generally, we found that more abstract concepts, such as *wavelength* or *wave spectrum*, were represented by non-iconic images, whereas more concrete concepts such as *steepness*, *crest* or *breaker* used iconic images.

Similarly, those images representing concepts referring to event-like processes appear to be more prone to show movement. Hence, dynamicity is a highly important criterion for process-oriented terminological research and for the configuration of the underlying conceptual event structure of a specific domain. In the short term, the number of dynamic images on the Internet is expected to increase considerably as a consequence of advances in technology. Textual explanation enhances the information conveyed by images when it is located in an unambiguous position in relation to the image. It is expected that these same characteristics will make up a continuum between the different levels of specialization of the scientific text.

In view of the results of our study, the images in our database are structured according to the typology presented. Furthermore, the inclusion of a given image type of image in terminological entries evidently depends on the nature of the concept.

6. Conclusion

The PuertoTerm database focuses on the conceptual structure, interrelations, and representation of specialized concepts in the domain of Coastal Engineering. Because of its dynamic nature, the representation of such a domain necessarily involves the elaboration of a complex event model in order to situate and contextualize the concepts within each sub-area of the event structure. For this reason, the FrameNet methodology is presently being used in our project for the global representation of the domain as well as for the inclusion of syntactic and semantic information within term entries. This is complementary to the terminographic definition construction used in the Functional-Lexematic Model, which establishes the frame elements in a definitional schema that directly refers to the underlying event structure of the domain. This linguistic representation of each concept is enhanced by graphic information that is chosen on the basis of concept type. It is our assertion that only through the principled analysis and design of the micro- and macrostructure of the concepts in a specialized domain is it possible to transform a list of terms into an authentic information system that will facilitate comprehension and enhance knowledge acquisition.

Notes

1. This research is part of the project PUERTOTERM: Knowledge representation and the generation of terminological resources within the domain of Coastal Engineering, BFF2003–04720, funded by the Spanish Ministry of Education.
2. Although the corpus also has a significant number of German texts, for the sake of simplicity we are limiting our analysis here to English and Spanish.
3. These statistics have been obtained with the Wordlist tool of the computer application Wordsmith Tools®.
4. The overall difference in the type/token ratio is probably due to the fact that the relative scarcity of coastal engineering texts in Spanish has led to the inclusion of texts belonging to closely related fields of knowledge, such as Environmental Science, Geology or Hydrology. This leads to a similar standardized type/token ratio (if analyzed in comparable chunks) but a differing overall type/token ratio, where the highly homogeneous English corpus is compared as a whole to the more heterogeneous Spanish corpus.

References

- Barsalou, L. 2003. "Situated simulation in the human conceptual system." *Language and Cognitive Processes* 18(5/6), 513–562.
- Béjoint, H. 1997. "Regards sur la définition en terminologie." *Cahiers de Lexicologie* 70(1), 19–26.
- Bowker, L. 1997. "Multidimensional classification of concepts and terms." In Wright, S.E. and G. Budin (eds.). *Handbook of Terminology Management*, vol. 1. 133–143. Amsterdam/Philadelphia: John Benjamins.
- Bowker, L. and I. Meyer. 1993. "Beyond 'textbook' concept systems: Handling multidimensionality in a new generation of term banks." In Schmitz, K.D. (ed.). *TKE'93: Proceedings of the Third International Congress on Terminology and Knowledge Engineering*. 123–137. Cologne, Germany. Frankfurt: Indeks Verlag.
- Calsamiglia, H., S. Bonilla, D. Cassany, C. López and M. Jaume. 1998. "Análisis discursivo de la divulgación científica." Paper read at the *I Simposio Internacional de Análisis del Discurso*. Universidad Complutense de Madrid, Madrid, Spain. http://www.upf.es/dtf/personal/danielcass/anali.htm. Accessed 27 February 2006.
- EAGLES. 1996. "Preliminary recommendations on corpus typology." Eagles Document (Expert Advisory Group on Language Engineering) EAG-TCWG-CTYP/P.
- Faber, P. 2002. "Terminographic definition and concept representation." In Maia, B. (ed.). *Training the Language Services Provider for the New Millennium*. 343–354. Oporto (Portugal): University of Oporto.
- Faber, P. 2003. "Terminological competence and enhanced knowledge acquisition." *Research in Language* 1, 95–116.
- Faber, P. and R. Mairal Usón. 1999. *Constructing a Lexicon of English Verbs.* Berlin: Mouton de Gruyter.
- Faber, P. and M.I. Tercedor Sánchez. 2001. "Codifying conceptual information in descriptive terminology management." *Meta* 46(1), 192–204.
- Faber, P., C. Márquez and M. Vega. 2005. "Framing Terminology: A process-oriented approach." *Meta* 50(4). CD-ROM.
- Fillmore, C.J. 1982. "Frame semantics." In Linguistic Society of Korea (ed.). *Linguistics in the Morning Calm*. 111–137. Seoul: Hanshin.
- Fillmore, C.J. 1985. "Frames and the semantics of understanding." In *Quaderni di Semantica* 6(2), 222–254.
- Fillmore, C.J. 2003. "Double-decker definitions: The role of frames in meaning explanations." *Sign Language Studies* 3(3), 263–295.
- Fillmore, C.J. and B.T.S. Atkins. 1992. "Towards a frame-based organization of the lexicon: The semantics of RISK and its neighbors." In Lehrer, A and E. Kittay (eds.). *Frames, Fields, and Contrast: New Essays in Semantics and Lexical Organization*. 75–102. Hillsdale: Lawrence Erlbaum Associates.
- Fillmore, C.J., C.R. Johnson and M. Petruck. 2003. "Background to Framenet." *International Journal of Lexicography* 16(3), 235–250.
- Grinev, S. and I.A. Klepalchenko. 1999. "Terminological approach to knowledge representation." In Sandrini, P. (ed.). *TKE '99: Proceedings of the 5th International*

Congress on Terminology and Knowledge Engineering. 147–151. Innsbruck, Austria. Vienna: TermNet.

– Johnson, C.R., C.J. Fillmore and M. Petruck. 2002. "FrameNet: Theory and practice." http://www.icsi.berkeley.edu/~framenet/book/book.html. Accessed 27 February 2006.

– Martín Mingorance, L. 1984. "Lexical fields and stepwise lexical decomposition in a contrastive English-Spanish verb valency dictionary." In Hartmann, R. (ed.). *LEXeter 83: Proceedings of the International Conference on Lexicography*. 226–236. Tübingen (Germany): Max Niemeyer.

– Martín Mingorance, L. 1989. "Functional grammar and lexematics." In Tomaszcyk, J. and B. Lewandowska (eds.). *Meaning and Lexicography*. 227–253. Amsterdam/Philadelphia: John Benjamins.

– Martín Mingorance, L. 1995. "Lexical logic and structural semantics: Methodological underpinnings in the structuring of a lexical database for a natural language processing." In Hoinkes, U. (ed.). *Panorama der Lexikalischen Semantik.* 461–474. Tübingen (Germany): Gunter Narr.

– Monterde Rey, A.M. 2004. "La importancia de la ilustración para la traducción técnica." In Gonzalo García, C. et al. (eds.). *Manual de Documentación, Terminología y Traducción especializada.* 259–274. Madrid: Arco/Libros.

– Pérez Hernández, C. 2002. "Terminología basada en hábeas: Principios teóricos y metodológicos." In Faber, P. and C. Jiménez Hurtado (eds.). *Investigar en Terminología*. 127–166. Granada: Comares.

– Prieto Velasco, J.A. 2005. "El papel de la información gráfica en la descripción de conceptos en la Ingeniería de puertos y costas." Paper presented at the *Jornadas Hispano-Rusas*. Granada, Spain.

– Servicio Hidrográfico y Oceanográfico de la Armada de Chile. 2002. *Glosario de Marea y Corrientes*, 3rd edition. Valparaiso (Chile): Servicio Hidrográfico y Oceanográfico de la Armada de Chile. http://www.shoa.cl/servicios/descargas/pdf/Pub3013.pdf#search='Glosario%20de%20Marea%20y%20Corrientes%2C'. Accessed 5 March 2006.

– Strehlow, R. 1993. "Terminological standardization in the physical sciences." In Sonneveld, H.B. and K.L. Loening (eds.). *Terminology: Applications in Interdisciplinary Communication*. 127–140. Amsterdam/Philadelphia: John Benjamins.

– Temmerman, R. 2000. *Towards New Ways of Terminology Description*. Amsterdam/Philadelphia: John Benjamins.

– Tomczak, M. 2000. "Tides in shallow seas." Shelf and Coast Lecture Notes, Chapter 5. http://gyre.umeoce.maine.edu/physicalocean/Tomczak/ShelfCoast/notes/chapter05.html. Accessed 27 February 2006.

– Voigt, B. 1998. Glossary of Coastal Terminology. Seattle: Washington Department of Ecology. http://www.ecy.wa.gov/programs/sea/swces/products/glossary.htm. Accessed 27 February 2006.