Controlled Language through the Definitions of Coastal Terms in English*

Miriam Buendía Castro

University of Granada, Granada, Spain

Elsa Huertas Barros

University of Granada, Granada, Spain

Controlled languages are specially defined subsets of natural language which help to create clear and concise documents, and thus insure coherent communication. In this paper we approach controlled language through the definitions of specialized terms with a view to establishing a metalanguage for the basic format of definitions. Our research focuses on a corpus of specialized terms related to tides, extracted from the *Glossary of Coastal Terminology* elaborated by the NOAA Coastal Services Center.

The methodology used for elaborating controlled-language definitions is based on the Functional-Lexematic Model (FLM) (Martín Mingorance 1984, 1989, 1995; Faber and Mairal 1999). The conceptual structure of a domain is made explicit within definitions that are coherent in both their micro and macrostructure (Faber and Tercedor Sánchez 2001). The linguistic description of any specialized concepts should do the following: (1) make category membership explicit; (2) reflect its relations with other concepts within the same category; (3) specify its essential attributes and features (Faber *et al*., 2005).

Our results show that the use of controlled language in the definition of terms makes their conceptual description more coherent and systematic. Controlled language increases terminological consistency, facilitates standardization, simplifies syntax, and avoids semantic ambiguities.

1 INTRODUCTION

Controlled language is a specially simplified version of a language which is typically adopted by a company or by the documentation section of a company. Both vocabulary and syntax may be restricted. A controlled language attempts to reduce ambiguities, colloquialisms and synonyms (AECMA 1995). This is an approach which is particularly valid in certain environments when it comes to the preparation of technical documents for better translatability or easier computer processing. Nowadays, the best known controlled language is AECMA Simplified English of the aerospace industry (AECMA 1995). Texts written in controlled language are optimized for machine translation, CAT (Computer-Assisted Translation), or other forms of NLP (Natural Language Processing).

In this paper we approach controlled language through the definitions of specialized terms with a view to establishing a metalanguage for the basic format of definitions. Our research focuses on a corpus of specialized terms related to tides extracted from the Glossary of Coastal Terminology elaborated by the NOAA Coastal Services Center.

* This research is part of the Project MARCOCOSTA: Marcos de conocimiento multilingüe en la gestión integrada de zonas costeras (P06-HUM-01489), funded by the Andalusian Regional Government.

The initial section of this article gives a general overview of the advantages and disadvantages of using CL. Section two focuses on the methodology used for elaborating controlled-language definitions, which is based on the Functional-Lexematic Model (FLM) (Martín Mingorance 1984, 1989, 1995; Faber and Mairal 1999). We then explain what a linguistic description of specialized concepts should accomplish. To achieve our purpose, we analyze the nuclear part of each definition or *genus* and the adverbial modification of *differentiae* in order to establish the conceptual relationships that can be extracted from the textual information in the form of concordances. Section three gives some examples of controlled definitions. The last section includes a conclusion with the results of our study.

2 ADVANTAGES AND DISADVANTAGES OF CL

Controlled language (CL) has been created with the aim of enhancing the readability of texts and improving their comprehensibility and translatability. The main advantages of this type of languages are univocity, consistency, accuracy and readability. It also should be underlined that CL is not artificial, but rather a subsystem of natural languages that pursues a restriction of vocabulary and grammar (AECMA 1995).The international projection of CL makes a text easier to understand and also to translate into another language. It also avoids any ambiguity and simplifies the complexity of natural languages, which allows non-native speakers to understand texts clearly. Thanks to CL, the processing of natural language is also more efficient and reliable. Therefore, controlled language tools are a solution for developing and enhancing international communication since they provide an advanced solution for transferring an idea from one language to another. It should also be highlighted that the use of CL in the definition of terms makes their conceptual description more coherent and systematic. CL increases terminological consistency and standardization, simplifies syntax and avoids semantic ambiguities (Jiménez Hurtado and Seibel, 2004).

Nevertheless, we should also point out that there are some controversial aspects regarding CL. First of all, it is true that CLs cannot be applied to all fields, since they were mostly created for limited domains. There is also a fear that CL structures may at some point become standard patterns, and cause speakers to lose their command of emotional and subjective language. This possible lack of quality in CL is partly because many CLs target specialized language texts with solely an informative function, and those who produce these texts are experts in the field, who naturally have this same agenda.

It is also true that syntactic rules differ substantially from language to language. Thus, it may be difficult to find CL structures which are candidates for interlinguistic correspondence. Evidently, not all structures can be easily controlled, due to grammatical restrictions.

Janowski (1998:1) mentions some of the risks and side effects of CL. According to this author, a compromise should be made between our objectives and the need for intelligibility, repeatability, low cost, balanced against potential risks, high error rate and probable rejection by target groups like private or professional product users- always being aware of the danger of excessive simplification and the deprivation of everyday colloquial language (1998:1).

3 THE FUNCTIONAL-LEXEMATIC MODEL

The elaboration of controlled-language definitions in this paper is based on the Functional-Lexematic Model (FLM) (Martín Mingorance 1984, 1989, 1995; Faber and Mairal 1999). This model facilitates the representation of conceptual relationships and collocations in general and specialized language (Faber, López Rodríguez and Tercedor Sánchez 2001). It

envisions a lexicon structured onomasiologically in terms of areas of meaning or lexical domains (Faber, Tercedor Sánchez 2001). The FLM proposes a lexical organization mainly based on the distinction between paradigmatic and syntagmatic relationships.

The paradigmatic axis of the FLM arranges concepts onomasiologically in a hierarchy of domains and subdomains. In other words, lexical subdomains are defined in terms of the nuclear part of conceptual descriptions or meaning definitions. This approach underlines conceptual areas, and is based on the assumption that each member of a certain lexical domain shares certain properties or meaning parameters with the other terms, but at the same time has other distinctive characteristics that differentiate it from other terms in the same domain. This type of onomasiological or conceptual organization has many advantages, not the least of which is its similarity to the organization of our mental lexicon (Faber and Mairal 1999).

As previously mentioned, the syntagmatic axis focuses on the combinatorial potential of lexical items, which is greater or lesser depending on the position of the item on the paradigmatic axis. Therefore, the combination of both axes represents the foundation of this conceptual structure. Both are complementary and inseparable in the description of meaning.

The FLM is a linguistic way of organizing concepts, using the information provided by specialists in the form of well-structured terminographic definitions (Faber and Tercedor, 2001).

4 DEFINING SPECIALIZED CONCEPTS

Defining a specialized language concept and all the terms associated with such a definition involves the following: (i) reproducing the basic information within such a concept as well as the information transmitted when this concept is activated within a text; (ii) establishing the difference between this concept and other concepts (Jiménez Hurtado and Seibel, 2005).

According to Faber et al., (2006), definitions can be regarded as mini-knowledge representations. Such a knowledge representation (KR) requires a metalanguage of definition for each category, which can be used to describe all of the concepts within that particular conceptual area. Such a metalanguage can be based on natural language and, when it is elaborated in principled way, can become a highly controlled language. (Jiménez Hurtado and Seibel, 2005).

The linguistic description of any concept should accomplish the following: (1) make category membership explicit; (2) reflect its relations with other concepts within the same category; (3) specify its essential attributes and features (Faber et al., 2005).

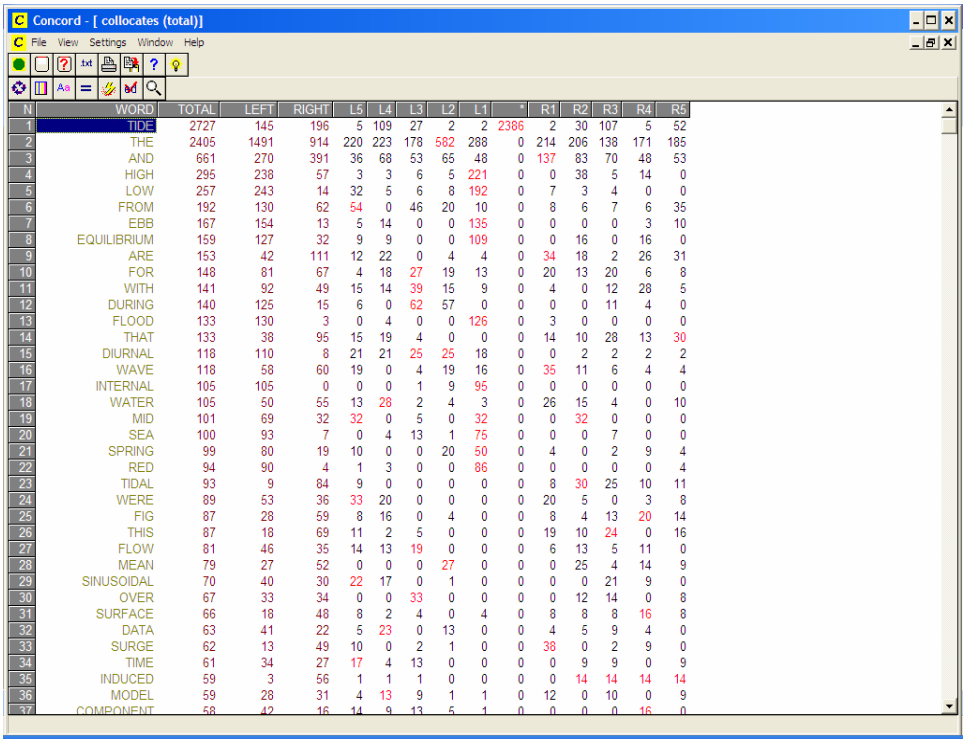
We have created a controlled language on the basis of information extracted from specialized dictionaries as well as a large electronic corpus of specialized texts on coastal engineering. Parameters of organization of this corpus of texts are: (i) level of specialization; (ii) group of target readers. The controlled language described in this article is of specialized texts written by experts and aimed at experts. The language used in the elaboration of terminographic definitions is natural language extracted from the corpus.

In a definition there are two major parts, the *genus* or nuclear part (which is indicative of the IS-A relationship) and the adverbial modification or *differentiae* that provides the characteristics that distinguish one concept from another within the same category (Faber et al., 2005).

4.1 Defining *tide*

Figure (1) shows the words which occur in the neighbourhood of our search word, namely *tide.* The study of these collocates helps us to find the meaning and usage of *tide.* As can be

seen, *high*, *low*, *ebb*, *equilibrium*, *flood*, *diurnal*, *wave*, *internal*, *mid*, *sea*, *spring* or *red* are the words which collocate most frequently with our search word. The occurrence of very common words like *the* and *and* does not offer significant information.



N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	TIDE	2727	145	196	5	109	27	2	2	2386	2	30	107	5	52
2	THE	2405	1491	914	220	223	178	582	288	0	214	206	138	171	185
3	AND	661	270	391	36	68	53	65	48	0	137	83	70	48	53
4	HIGH	295	238	57	3	3	6	5	221	0	0	38	5	14	0
5	LOW	257	243	14	32	5	6	8	192	0	7	3	4	0	0
6	FROM	192	130	62	54	0	46	20	10	0	8	6	7	6	35
7	EBB	167	154	13	5	14	0	0	135	0	0	0	0	3	10
8	EQUILIBRIUM	159	127	32	9	9	0	0	109	0	0	16	0	16	0
9	ARE	153	42	111	12	22	0	4	4	0	34	18	2	26	31
10	FOR	148	81	67	4	18	27	19	13	0	20	13	20	6	8
11	WITH	141	92	49	15	14	39	15	9	0	4	0	12	28	5
12	DURING	140	125	15	6	0	62	57	0	0	0	0	11	4	0
13	FLOOD	133	130	3	0	4	0	0	126	0	3	0	0	0	0
14	THAT	133	38	95	15	19	4	0	0	0	14	10	28	13	30
15	DIURNAL	118	110	8	21	21	25	25	18	0	0	2	2	2	2
16	WAVE	118	58	60	19	0	4	19	16	0	35	11	6	4	4
17	INTERNAL	105	105	0	0	0	1	9	95	0	0	0	0	0	0
18	WATER	105	50	55	13	28	2	4	3	0	26	15	4	0	10
19	MID	101	69	32	32	0	5	0	32	0	0	32	0	0	0
20	SEA	100	93	7	0	4	13	1	75	0	0	0	7	0	0
21	SPRING	99	80	19	10	0	0	20	50	0	4	0	2	9	4
22	RED	94	90	4	1	3	0	0	86	0	0	0	0	0	4
23	TIDAL	93	9	84	9	0	0	0	0	0	8	30	25	10	11
24	WERE	89	53	36	33	20	0	0	0	0	20	5	0	3	8
25	FIG	87	28	59	8	16	0	4	0	0	8	4	13	20	14
26	THIS	87	18	69	11	2	5	0	0	0	19	10	24	0	16
27	FLOW	81	46	35	14	13	19	0	0	0	6	13	5	11	0
28	MEAN	79	27	52	0	0	0	27	0	0	0	25	4	14	9
29	SINUSOIDAL	70	40	30	22	17	0	1	0	0	0	0	21	9	0
30	OVER	67	33	34	0	0	33	0	0	0	0	12	14	0	8
31	SURFACE	66	18	48	8	2	4	0	4	0	8	8	8	16	8
32	DATA	63	41	22	5	23	0	13	0	0	4	5	9	4	0
33	SURGE	62	13	49	10	0	2	1	0	0	38	0	2	9	0
34	TIME	61	34	27	17	4	13	0	0	0	0	9	9	0	9
35	INDUCED	59	3	56	1	1	1	0	0	0	0	14	14	14	14
36	MODEL	59	28	31	4	13	9	1	1	0	12	0	10	0	9
37	COMPONENT	58	42	16	14	9	13	5	1	0	0	0	0	16	0

Figure 1

Collocations which appear with tide

Tide appears a total of 2727 times in our corpus. This figure allows us to study a great many concordances, with a view to establishing its definition. A concordance is an example of a given word or phrase, which shows its context and combinatory potential. Examples of concordances of *tide* are the following:

Representative concordances of *tide* extracted from the corpus[1]

(1) **Tides** are **caused by** slight **variations in gravitational attraction between the Earth and the moon and the sun** in geometric relationship with **locations on the Earth's surface. Tides** are **periodic** primarily **because of** the cyclical influence of the **Earth's rotation.**

(2) **Tides** are the **periodic rise and fall of the** ocean **waters**. They are **caused by** the **gravitational pulls of the Moon** and (to a lesser extent) **Sun**, as well as the **rotation of the Earth**.

(3) Maximum values were observed on the **rising and falling** with a minimum at high tide

(4) **The gravitational forces of the moon and sun** provide the driving forces of the ocean tides

(5) The **moon** produces two tidal bulges somewhere on the **Earth** through the **effects of gravitational attraction.**

We have highlighted in red the words or phrases which show a clear relation with *tide*, in order to extract the conceptual and terminological information which we would need for our definition. Once the conceptual and terminological information was extracted, the next step was to abstract a schema or template which, according to the concordances, encodes the definitional structure for all of the concepts belonging to this domain. The definitional template for *tide* the following:

[1] The corpus was elaborated as part of the PuertoTerm research project, funded by the Ministry of Education in Spain. It has a total of 4,435,525 words. The concordances have been studied with the Wordsmith Tools.

Tide	
IS_A	rise and fall of water
CAUSED_BY	gravitational attraction of the moon and sun
AFFECTS	rotating Earth
HAS_TIME	periodic

Table 1

Definitional structure of tide

Table 1 shows that *tide* can be characterized by four types of conceptual relations: IS_A, CAUSED_BY, AFFECTS, and HAS_TIME. *Tide* would this be defined by the phrases that designate these relations, and would be defined as follows: *periodic rise and fall of water caused by the gravitational attraction of the moon and sun, which affects the rotating Earth.*

The next set of concordances provides information relating to different types of *tides:*

Concordances of types of tide extracted from the corpus

(1) Maximum values were observed on the rising and **falling tides** with a minimum at high tide.
(2) Observations from the **ebb tide** of Aug. 26 provided the clearest examples of the initial steps of flow development.
(3) Pressure, permeability and sediment transport effects likely caused the trends in backwash friction values over the **high tide** cycle.
(4) Unsaturated flow is caused by gradients arising due to adjacent areas of **high and low water** content.
(5) Due to the geometry of the channel, the observed sill flow during **flood tide** is more complicated.
(6) The **rising tide** generates flood currents.
(7) A **semi-diurnal tide** is typical of the coastal waters.
(8) Cyclical tidal cycles associated with a **diurnal tide.**
(9) The height of the average **solar tide** is about 50% the average **lunar tide**.
(10) Caused by the nearly coincident new **moon tide**.

From the study of the concordances of our corpus, we obtained the following types of *tides*. Those terms in the same square designate the same concept.

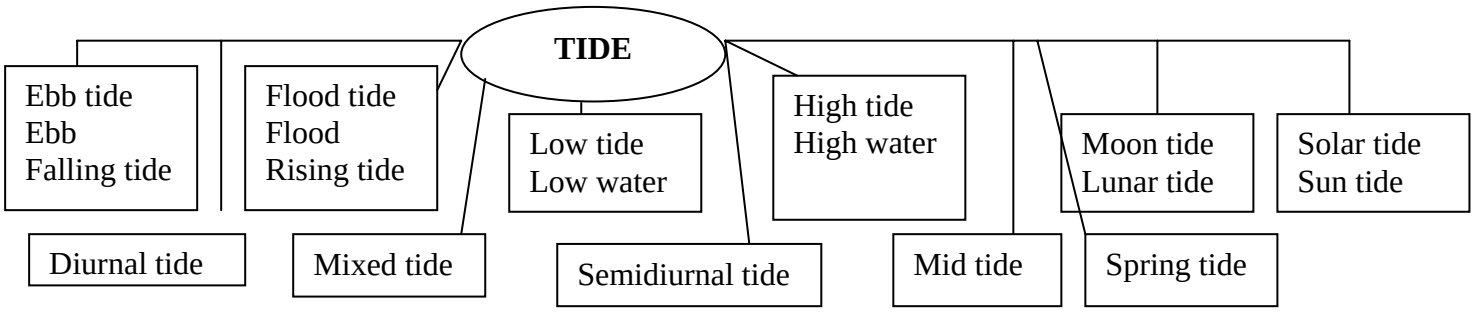


Figure 2

Types of tides

Table 2 shows how these terms can be categorized in terms of meaning parameters (water movement, water height, attractive force), and in terms of their respective degree of specialization.

WATER MOVEMENT		WATER HEIGHT		ATTRACTIVE FORCE	
Non-specialized receptor					
ebb tide	**flood tide**	**low tide**	**high tide**	**moon tide**	**solar tide**
Specialized receptor					
ebb **falling tide [syn.]**	**flood** **rising tide [syn.]**	**low water**	**high water**	**lunar tide**	**sun tide**

Table 2

Categorization of terms

Table 3 shows how our definitional template can be applied to the specific types of tide in order to establish their respective definitions. In those cases where one concept is designated by more than one term, the main term appears first in bold letters, followed by the synonyms in italics. We then apply the definitional template for *tide* to each concept. This means that the linguistic information for each of the more specific types of *tide* is proposed with a view to creating a uniform structure and a controlled definition for each concept. When necessary, within the conceptual relation, *HAS_TIME,* we have made the distinction between *phases* and *frequency. Default value* signifies that such a conceptual relation evokes the same concept activated in *tide.* We also point out when a specific conceptual relation is not activated within a concept. The analysis of the characteristics lexicalized in *differentiae* provides the defining features of the different types of *tides.*

TIDE

Rising tide (*flood tide, flood*)	
IS_A	tide
CAUSED_BY	[default value]
AFFECTS	[default value]
HAS_TIME	between low water and the following high water

Ebb Tide (*falling tide, ebb*)	
IS_A	tide
CAUSED_BY	[default value]
AFFECTS	[default value]
HAS_TIME	Between high water and the following low water

Moon tide (*lunar tide*)	
IS_A	tide
CAUSED_BY	direct attraction to the moon
AFFECTS	[default value]
HAS_TIME	[not activated]

Solar tide (*sun tide*)	
IS_A	tide
CAUSED_BY	direct attraction to the sun
AFFECTS	[default value]
HAS_TIME	[not activated]
Diurnal tide	
IS_A	tide
CAUSED_BY	[default value]
AFFECTS	[default value]
HAS_TIME	

	(i) phases	one flood and one ebb period
	(ii) frequency	each tidal day

Semidiurnal tide	
IS_A	tide
CAUSED_BY	[default value]
AFFECTS	[default value]
HAS_TIME	
(i) phases	two flood and two ebb periods
(ii) frequency	each tidal day

Mixed tide	
IS_A	tide
CAUSED_BY	[default value]
AFFECTS	[default value]
HAS_TIME	
(i) phases	Two flood and two ebb periods per day but of different ranges.
(ii) frequency	each tidal day

Mid tide	
IS_A	tide-moment
CAUSED_BY	[default value]
AFFECTS	[default value]
HAS_TIME	midway between the high water and the low water

High water	
IS_A	tide-moment
CAUSED_BY	flood
AFFECTS	[default value]
HAS_TIME	when water is at highest level

Low water	
IS_A	tide-moment
CAUSED_BY	[default value]
AFFECTS	[default value]
HAS_TIME	when water is at lowest level

Spring tide	
IS_A	higher high tide/ lower low tide
CAUSED_BY	combined gravitational attraction of both the moon and sun when aligned with the Earth
AFFECTS	[default value]
HAS_TIME	
(i) frequency	every two weeks when there is a full moon

Table 3

Definitional structure of the different types of tide

The definitions make the intercategorial relations very clear. Since this conceptual category is largely based on oppositions between its members: (1) rising tide/ebbtide; (2) moon tide/solar tide; (3) high water/low water. However all can be defined by the same set of conceptual relations, which activate basically the same or a similar set of lexical items. This template can be considered a conceptual grammar for the description of all types of tides,

which ensures a high degree of systematisation at the micro-structural level (Faber et al., 2005).

5 CONCLUSIONS

In this paper we have briefly introduced the concept of CL and have explained the advantages and disadvantages of using it. We propose a type of metalanguage for the basic format of controlled-language definitions which are based on Martín Mingorance's Functional-Lexematic Model (FLM).

Our research corroborates the notion that the linguistic description of any specialized concepts should make category membership explicit, reflect its relations with other concepts within the same category and specify its essential attributes and features (Faber *et al*., 2005).

Our results indicate that the use of controlled language in the definition of terms makes their conceptual description more coherent and systematic. Controlled language increases terminological consistency and standardization, simplifies the syntax and avoids any semantic ambiguities.

This is the preliminary research for the creation of controlled language, which we plan to do in the subsequent phase of our research.

REFERENCES

AECMA (1995). *AECMA Simplified English: A Guide for the Preparation of Aircraft Maintenance Documentation in the International Aerospace Maintenance Language.* AECMA, Brussels

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. et al. (2005). "Framing Terminology: A process-oriented approach." *Meta* 50(4). CD-ROM.

Faber, P. et al. (2006). "Processing of Terms in Specialized Dictionaries" *Terminology*. 12 (2), 2006, 189-213.

Jiménez Hurtado, C. and C. Seibel. (2004). "El lenguaje controlado para una definición variable: no ambiguity through homonyms, no redundancy through synonyms". Faber, Pamela, Jiménez Hurtado, Catalina & Gerd Wotjak (eds.). *Léxico especializado y comunicación interlingüística.* 117-130. Granada: Granada Lingüística.

Jiménez Hurtado, C. and C. Seibel. (2005). "La definición variable como base semántica del lenguaje controlado: Un ejemplo en textos sobre procesos marinos (alemán-español". Campos Plaza, Nicolás A., Miguel A. García Peinado, Emilio Ortega Arjonilla & Cécile Vilvandre de Sousa (eds.) *El español, lengua de cultura, lengua de traducción: aspectos teóricos, metodológicos y profesionales.* Granada: Editorial Atrio. 465-477.

Martín Mingorance, L. (1984). "Lexical fields and stepwise lexical decomposition in a contrastive English-Spanish verb valence 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.

Voigt, B. Glossary of Coastal Terminology. Seattle: Washington Department of Ecology. http://www.ecy.wa.gov/programs/sea/swces/products/glossary.htm Accessed: 20 November 2006.

Miriam Buendía Castro

Departamento de Trad. e Interpretación
C/ Buensuceso, 11
Universidad de Granada
Granada (Spain)
18002
Spain

mbuendia@ugr.es

Elsa Huertas Barros

Departamento de Trad. e Interpretación
C/ Buensuceso, 11
Universidad de Granada
Granada (Spain)
18002
Spain

ehuertas@ugr.es