EXPLORING THE SEMANTICS OF MULTI-WORD TERMS BY MEANS OF PARAPHRASES

Melania Cabezas-García
Universidad de Granada
melaniacabezas@ugr.es

Pamela Faber
Universidad de Granada
pfaber@ugr.es

Abstract
Multi-word terms (MWTs) are the main way that concepts are linguistically expressed in specialized domains. Accessing the semantic content of these compressed propositions is the first step toward understanding and translating them. Until now, most studies have focused on two-term compounds (Kim & Baldwin, 2013). This paper, however, deals with three-term English and Spanish endocentric noun compounds in the specialized domain of Coastal Engineering. Our analysis involved parsing, bracketing, and the assignment of semantic relations. The meaning of the MWTs was then expanded through paraphrasing (Nakov, 2013). Our results showed that a predicate-based analysis facilitated the specification of the relations between the concepts in MWTs as well as the mapping of this content onto the corresponding term in the target language.

Resumen
Los conceptos especializados se expresan principalmente con términos compuestos. El primer paso para comprender y traducir estos términos yuxtapuestos, que representan proposiciones, es acceder a su contenido semántico. Hasta ahora, se habían estudiado fundamentalmente los compuestos formados por dos términos (Kim & Baldwin, 2013). Sin embargo, este artículo aborda los compuestos nominales endocéntricos en inglés y español formados por tres términos pertenecientes al campo de la Ingeniería Costera. Primero, realizamos un análisis sintáctico y gramatical y señalamos la estructura interna de los términos. A continuación, asignamos relaciones semánticas a los compuestos y expandimos su significado por medio de paráfrasis (Nakov, 2013). Nuestros resultados indicaron que este análisis facilitó la especificación de las relaciones conceptuales y la identificación de términos equivalentes en la lengua meta.

Keywords: multi-word term, noun compound, semantic relation, paraphrase, terminology
Palabras clave: término compuesto, compuesto nominal, relación semántica, paráfrasis, terminología

Professional Profiles
Melania Cabezas-García is a PhD student in Translation and Interpreting at the University of Granada (Spain). Her research interests are Terminology, Lexicography, Corpus Linguistics, Specialized Translation, and Verbal Lexicon.
Pamela Faber is a Full Professor at the Department of Translation and Interpreting of the University of Granada (Spain). Her research interests are Terminology, Specialized Translation, Cognitive Semantics, and Lexicography.

1. Introduction

Multi-word terms (MWTs) are the most frequently used units when conveying specialized knowledge (Horsella & Pérez, 1991; Daille et al., 2004; Hendrickx et al., 2013). In regards to grammatical category, 85% of MWTs are noun compounds (Nakagawa & Mori, 2003). These units represent juxtaposed concepts (Zelinsky-Wibbelt, 2012), which is why disambiguating their semantic content is a crucial step toward understanding them in context and, eventually, to translating them.

To this end, inventories of semantic relations have traditionally been the preferred option for specifying their meaning. Nevertheless, there are drawbacks to such relations, since they can only partially reflect the semantics of these terms. Furthermore, all too frequently, various relations are possible and a choice must be made among the many existing inventories (Nakov, 2013).

Natural language processing research on compound interpretation has a long history, mainly because the semantic relation between the compound's constituents cannot be inferred by its head and modifiers (Ó Séaghdha & Copestake, 2013). Doubtlessly, this is one of the disadvantages of using inventories of semantic relations. For this reason, we opted for another method proposed by Nakov & Hearst (2006), who argue that noun compound semantics is more easily accessed by means of paraphrases involving verbs and/or prepositions. Our results indicate that verb paraphrases better reflect the semantic universe of MWTs. For example, in MWTs designating processes, they specify the action performed and place the MWT within the context of a semantic field or domain.

Up until now, most studies have focused on two-term compounds (Kim & Baldwin, 2013), in particular, when devising methods for recognising and automatically extracting them from a corpus. Notwithstanding, we decided to study three-term noun compounds, since these terms can reflect the arguments of the predicate (Nakov & Hearst, 2013) when the noun compound is the nominalization of a process. In this sense, the more terms that an MWT has, the more specific it is.

Still another aspect to be considered is that every language has its own term formation patterns. This means that translators must first access the semantic content underlying an MWT in order to translate it properly. Along these lines, it is well known that advances in science and technology lead to the creation of new terms, especially in English, the *lingua franca* of communication (Humbley & García Palacios, 2012). However, in order to disseminate scientific results throughout the world, new terms need to be translated. Thus, research on MWTs, the most productive specialized units both in English and Spanish, is a priority.

This paper describes the use of verb paraphrases for accessing the semantic content of MWTs. For this purpose, the terms in our study were extracted from the EcoLexicon corpus (http://ecolexicon.ugr.es/), both in English and Spanish, although all of them are not necessarily equivalent terms. These terms designate specialized processes and represent compressed propositions whose implicit conceptual relationship must be

retrieved by text receivers. Our analysis involved parsing and bracketing, in order to specify the semantic structures and dependencies of the MWTs. Their meaning was then expanded through verb paraphrasing (Nakov, 2013) and, finally, the paraphrases elicited from Coastal Engineering experts were compared with those extracted from a web search engine.

This study involved the following:

i) specification of the relations between the concepts that make up the MWTs. This entailed accessing the underlying propositions (whose implicit conceptual load must be recovered by the receiver) and studying the role of micro-contexts in term formation (Hendrickx et al., 2013) and semantic interpretation;

ii) specification of the generic verbs of semantic fields and their hyponyms and classification of the verbs in EcoLexicon in semantic categories;

iii) establishment of mapping relations between terms in English and Spanish.

The objective was to disambiguate the MWTs and accurately access their conceptual load.

2. Multi-word Terms: an Approach to Noun Compounds and Verb Paraphrases

2.1 Noun Compounds

Noun compounds have been defined in various ways. However, the most commonly used definition is that proposed by Downing (1977), who defined them as a sequence of nouns which function as a single noun (for instance, *water quality management* or *propagación de un tren de ondas*).

Noun compounds can be endocentric (such as the terms in this study) or exocentric. In an endocentric compound, "one member functions as the head and the other as its modifier, attributing a property to the head" (Nakov, 2013: 299). In contrast, exocentric compounds lack a head and usually refer to pejorative properties of human beings (Nakov, 2013).

Characteristic properties of noun compounds include the following (Nakov, 2013): (i) headedness (English endocentric noun compounds are mainly right-headed, Spanish endocentric noun compounds tend to be left-headed); (ii) transparency; (iii) syntactic ambiguity; and (iv) language-dependency.

Furthermore, as previously mentioned, there are propositions underlying the noun compounds. These propositions can be inferred by the formation processes of these MWTs, as pointed out by Levi (1978), who distinguished between predicate deletion and predicate nominalization. In MWTs formed by predicate deletion, the modifiers are usually the object of the predicate, which has been elided. On the other hand, in MWTs formed by predicate nominalization, the head is a nominalized verb, whose modifiers are the subject or object of the predicate. It is equally possible that the modifiers represent both the subject and the object. Most noun compounds in our study were formed by means of this process (e.g. *water level fluctuation, disipación de la energía por fricción*, etc.).

In summary, the terms in our study are MWTs or sequences of nouns that function as a single noun. They are mainly endocentric compounds and are characterized by their headedness, transparency, syntactic ambiguity, and language-dependency. There are propositions underlying these MWTs, as reflected in the two main formation processes: predicate deletion and predicate nominalization.

2.2 Accessing the Semantics of MWTs: Semantic Relations vs. Paraphrases

Linguists have traditionally used taxonomies of semantic relations to express the conceptual relation holding between the constituents of MWTs. To this end, a myriad of different inventories have been created, ranging from coarse-grained classifications, (e.g. Vanderwende's, 1994) to fine-grained groupings (e.g. Nastase & Szpakowicz, 2003) to domain-specific inventories (e.g. Rosario et al., 2002).

Although semantic relations have advantages, such as parsimony and generalization (Hendrickx et al., 2013), they also have many disadvantages. For example, it is necessary to decide which set will be used; the relations are abstract and limited; they are only a partial reflection of semantics; and several relations are often possible (Nakov, 2013).

Faced with these problems, Downing (1977) defended that noun compound semantics could not be expressed through any set of relations. Accordingly, Nakov & Hearst (2006), inspired by Finin (1980), proposed another solution and stated that the best way of expressing the semantic content of a noun compound is by means of multiple paraphrases. For instance, *malaria mosquito* can be paraphrased with the fine-grained verbs *carry*, *spread*, *cause*, *transmit*, etc. since this is the action performed by the mosquito who infects humans with the disease. This proposal is also closely related to the FrameNet project (Baker et al., 1998). In the words of Teubert (2005), the conceptual load of a unit of meaning can be formulated in a paraphrase or a set of paraphrases.

Not surprisingly, this idea of paraphrases for accessing the semantic content of noun compounds has become increasingly popular among natural language processing researchers (Butnariu & Veale, 2008; Nakov & Hearst, 2008; Nakov, 2008). Specifically, Task 9 at SemEval-2010 (Butnariu et al., 2010) focuses on this procedure. In fact, the annotators proposed lists of paraphrases for each noun compound to expand meaning at different levels of granularity and, in case of ambiguity, the different interpretations were reflected (Hendrickx et al., 2013).

The main focus of terminology work has always been on nouns (L'Homme, 1998). However, verbs are also important because they represent events and states, which make up a great deal of our knowledge (Faber, 1999). In this regard, increased attention is now being paid to predicates, which often uncover the path to the semantic content of the terms (Butnariu & Veale, 2008; Buendía, 2012; *inter alia*). In addition, Nakov & Hearst (2013) argue that verbs are one of the most frequent open-class parts of speech in English, and they can reflect fine-grained features of meaning.

In our opinion, the two approaches (inventories of semantic relations and paraphrases) are complementary. Although the sets of semantic relations have certain limitations, they are useful and can be particularized by means of verb paraphrases.

3. Materials and Methods

Our research was a mixed study in which a corpus and a questionnaire were employed. A basic quantitative analysis was also used to complement the more qualitative study.

3.1 Materials

For the purposes of our study, a parallel corpus was downloaded from EcoLexicon (http://ecolexicon.ugr.es/) that was composed of specialized texts on the

same subject. The corpus consisted of two subcorpora, one in English (9 million tokens) and the other in Spanish (2 million tokens). All of the texts were papers belonging to the domain of Coastal Engineering. The texts came from high-impact specialized journals, such as *Coastal Engineering*, *Journal of Hydrology*, *Ingeniería del Agua*, and *Ingeniería Hidráulica y Ambiental*, thus meeting the quality requirements.

The corpus was then uploaded to Sketch Engine (https://www.sketchengine.co.uk/), an online corpus-analysis tool that allowed us to extract the term candidates in both languages. Sketch Engine (Kilgarriff et al., 2004) is an open source tool that can process large amounts of text. It generates word sketches, a corpus-based thesaurus, sketch differences, word's grammatical and collocational behavior, etc.

In addition, we also recruited a small group of experts who agreed to participate in our research. The group was composed of five people (three men and two women) whose mean age was 30. They were mainly coastal engineers, researchers, and professors with 3-10 years of experience within their profession. All of the experts were Spanish native speakers with an excellent command of English.

These experts filled out a previously designed questionnaire composed of three sections with a view to eliciting different types of information. In the first section, they were asked to define terms. According to Saldanha & O'Brien (2013), it is best to start with the elicitation of factual or descriptive information, and the data obtained were very useful for our research. Secondly, they had to formulate verb paraphrases of the terms. Finally, they answered questions that elicited their perceptions and opinion regarding the questionnaire (see Appendix 1).

As part of our study, the *Web as Corpus* was used to extract more paraphrases and compare these results with those obtained from the experts. As its name indicates, the *Web as Corpus* is a new approach to corpus in which the texts on the web are searched as though they composed a huge corpus (Buendía, 2013). Initially, we had thought to use WebCorp (http://www.webcorp.org.uk/live/), another tool that also uses the web as corpus and shows the results in the form of concordances. However, after testing this application, it was found that this system makes it difficult to obtain valid data because it has excessive restrictions. For this reason, we finally decided to use the web search engine Google (www.google.es/) because of the specificity of our terms and the need to access a larger quantity of data.

Finally, we looked up equivalent terms and definitions in EcoLexicon (in addition to other specialized resources). EcoLexicon (http://ecolexicon.ugr.es/) is an environmental knowledge base that is the practical application of Frame-based Terminology (Faber, 2009; 2011; 2012). It represents the conceptual content of the domain of the environment in the form of a thesaurus with semantic networks. Moreover, it also provides conceptual, linguistic, administrative and phraseological information (Buendía & Faber, 2015).

3.2 Methods

3.2.1 Corpus compilation and term extraction

As previously mentioned, the English and Spanish subcorpora were downloaded from EcoLexicon. However, the Spanish subcorpus was smaller than the English one because of the scarcity of specialized texts in Spanish. For this

reason, the size of the Spanish subcorpus was increased by means of the WebBootCat function of Sketch Engine. This tool allows the user to rapidly compile a corpus from the web, based on the seed terms entered. Furthermore, the user can choose the web sites to be included. This automatic compilation complemented the manual selection of corpus texts.

After uploading the corpus to Sketch Engine, the next step was term extraction. English term extraction was performed with the *Word List* function of Sketch Engine. The search attribute was established in *lemma*, and then 3 n-grams were used. A stop list eliminated irrelevant words. For Spanish term extraction, the *Word List* function was also used. However, the search attribute was set to *word*, and 5 or 6 n-grams were used. It was necessary to increase the number of n-grams in Spanish, because of the prepositions and articles in the MWTs of this language.

We preferred the *Word List* function over *Keywords & Terms* because the latter only offered a rather limited list of terms, many of which were not noun compounds or even three-term MWTs. They were thus not relevant to our research. As for the terms selected, we chose noun compounds designating Coastal Engineering processes, all of which were semantically related.

3.2.2 Parsing, bracketing and assignment of semantic relations

After the terms were selected, they were parsed and bracketed. As stated by Nakov (2013), parsing is a crucial step in semantic analysis because the syntactic structure indicates where semantic relations have to be assigned. Bracketing is also important because it disambiguates the syntactic interdependencies (see Table 1). The distinction between the head and the modifiers shows whether a noun compound is left or right-bracketed (Utsumi, 2014).

water quality management	
Bracketing	[water quality] management
Parsing	$[N_{modifier} + N_{head}]_{modifier} + N_{head}$

Table 1: *Bracketing and parsing for water quality management.*

The next step was to assign semantic relations to the noun compounds, in order to better understand their internal structure and semantic content. We initially used Nastase & Szpakowicz's (2003) taxonomy of 35 relations, and enhanced it with the semantic relations in EcoLexicon as well as other domain-specific relations that accounted for the semantic networks in the noun compounds of our research.

Table 2 shows the coarse-grained categories proposed by Nastase & Szpakowicz (2003) and the generic verbs in our study, all of which express the actions in the MWTs. Since our MWTs encoded processes, verbs were considered to be at the core of their meaning.

CAUSALITY: cause, manipulate PARTICIPANT: study, represent, create_model_of, see QUALITY: measure, change, maintain, decrease SPATIAL: move_outwards, move_upwards, move_over, change_movement TEMPORALITY: say_future

Table 2: *Generic verbs and verb phrases used to codify additional semantic relations.*

Regarding the participants in each process, a set of domain-specific categories were designed. When necessary, an attribute was also added. As can be observed, the semantic fields of water, negative situations, and movement have a high prevalence. The domain-specific categories are listed in Table 3:

WATER
WATER_WAVE
WATER_REPRESENTATION
WATER *MOVE_UPWARDS*
WATER *MOVE_UPWARDS* & *MOVE_OVER* LAND
NEG_SITUATION (MOVEMENT_WATER)
NEG_SITUATION (MET_DISTURBANCE *CAUSE* NEG_SITUATION)
MET_DISTURBANCE
MOVEMENT_ENERGY
MOVEMENT_AIR
MOVEMENT_VESSEL
MOVEMENT_SOLID_FRAGMENTS
FRICTION
SOIL_SURFACE
SEDIMENT
WIND
MEASUREMENT_LENGTH

Table 3: *Semantic categories designating the participants in the processes.*

The semantic relations in our research were the result of the combination of the generic verbs in Table 2 and the semantic categories in Table 3. Although these relations were domain-specific, we used paraphrase analysis to further specify them.

3.2.3 Expert paraphrases

A questionnaire was designed to elicit paraphrases of the MWTs from the group of experts (see Appendix 1). After an explanation of the activity, the experts wrote their paraphrases. The information was then organized in tables and the results were analyzed. In a few cases, the paraphrases were erroneous, and they were eliminated. Table 4 shows an example of paraphrase analysis.

Semantic relation	X			WATER (attribute: goodness)
	study			
Paraphrase	(X) estudio	sigue monitoriza mide controla realiza un seguimiento	((de) la)	calidad del agua
-Semantic relation -Paraphrase -Function words added to form a complete sentence -MWT: information that appears in the multi-word term				
				MWT

Table 4: *Color legend and paraphrase analysis for seguimiento de la calidad del agua*.

As can be observed, different colors were used to distinguish the following: (i) the information explicitly present in the MWT (grey); (ii) the semantic relation used (blue); (iii) the paraphrases that expanded the MWT and made its meaning more explicit (orange); (iv) function words added to form a complete sentence (green).

It was observed that semantic relations were more general than paraphrases and offered less information. In order to make semantic relations more specific, verb paraphrases were very useful, since they provided valuable information. For example, *study/estudiar*, a generic verb within the domain of MENTAL PERCEPTION, was further specified by means of its Spanish hyponyms (*sigue, monitoriza, mide, controla,* and *realiza un seguimiento*), thus providing a clearer view of the semantic universe of the MWT.

3.2.4 Web paraphrases

Although using a corpus can reduce noise, it has the problem of sparseness (Lapata & Keller, 2005). Given the specific nature of the domain, this was a problem even when using the web. As pointed out by Nakov & Hearst (2013), even if better processed linguistically, a corpus cannot compete with the vastness of the web. For this reason, we also retrieved information from Internet.

In the extraction of paraphrases from the web, the goal was to preserve the head-modifier relation by making the underlying propositions explicit. To this end, we first issued queries such as "flood risk management" "flood risk". We then retrieved the first five results in order to obtain the different participants in the semantic process (agent, location, etc.).

Secondly, when the object of the underlying proposition appeared in the MWT and the subject was not present, we searched queries such as "flood risk management" "to * the flood risk" and "flood risk management" "that * the flood risk". The * operator represents a wild card substitution. We then accessed the first five Google result pages for extracting different verbs in the semantic relation that could be used to find hyponyms of the generic verbs of our set of categories.

When the subject of the proposition appeared in the MWT, we searched for sequences of the type "wave energy conservation" "wave energy is *" or "wave energy can *". If no valid information was retrieved, we deleted the MWT within the quotation marks and searched queries such as "to * the wave energy". This way, a greater number of results was obtained, though much more noise was generated.

In the extraction of Spanish paraphrases, the same process was followed, except for verb extraction, when we issued queries such as "tren de ondas se *". This structure represents both a passive sentence and a reflexive passive sentence. In addition, we searched "que * un tren de ondas", to elicit both the subject and the verb. After extracting paraphrases from the web, we analyzed them and compared them with the ones proposed by the Coastal Engineering experts.

4. Results and Discussion

The analysis of these data produced the following results:

1. Regarding the formation of MWTs, the graph in Table 5 shows the syntactic structure of the propositions underlying the MWTs:

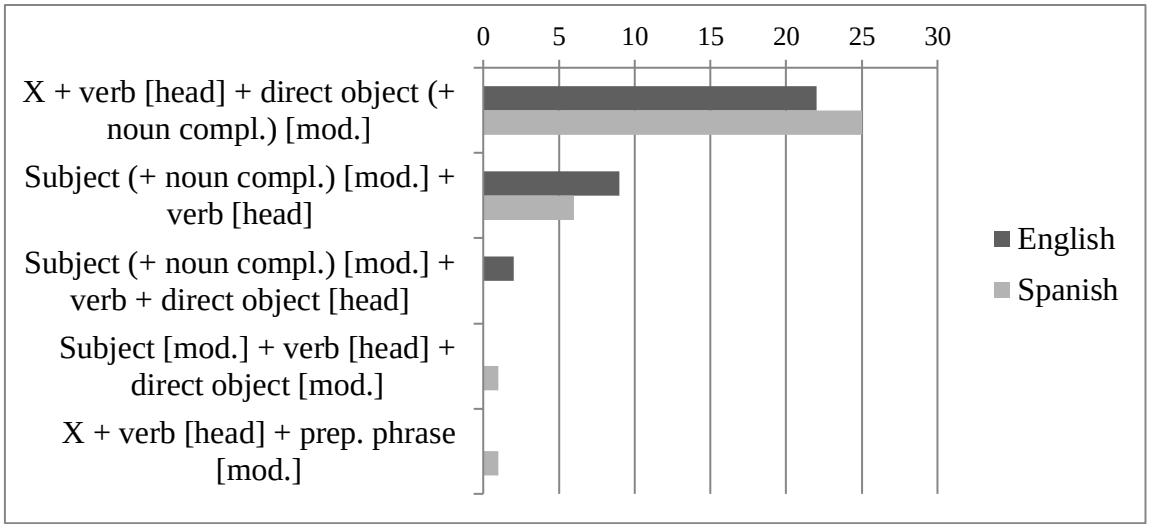

Table 5: *Syntactic structures of the propositions underlying the MWTs*.

As can be observed, there is a general pattern of slot filling in the argument structure of the MWTs. In most cases, the modifiers represent the direct object of the verb (which is composed of two nouns, one of which is a noun complement). The second most significant pattern occurs in MWTs in which the modifiers (a noun and a noun complement) are the subject of the underlying proposition. This way, the most common complements in our MWTs were the subject, direct object and noun complement. Additionally, it is worth mentioning that, since the MWTs in our study were composed of three terms, the semantic specificity was greater than in two-term compounds because more complements were involved.

2. The head of almost all of the MWTs was a verb, which was expressed as the nominalization of an environmental process (e. g. *wave energy conservation [conserve]*, *water level fluctuation [fluctuate]*, *propagación de las ondas de borde [propagar]*, *disipación de la energía del oleaje [disipar]*, etc.). This signals that all of the processes have a hidden predicate that also appears in the main position of the MWT, which highlights its importance. This is further evidence that predicates play a major role in the transmission of knowledge (in this case, in the transmission of specialized processes), as a vehicle for a domain-specific sequence in which several elements take part.

3. It was confirmed that MWTs composed of similar terms usually belong to the same semantic field and have similar combination patterns (Kim & Baldwin, 2013; Ó Séaghdha & Copestake, 2013). These combination preferences between similar terms are established both from the head to the modifiers and from the modifiers to the head. For example, the modifiers *riesgo de inundación* belong to the semantic field of NEG_SITUATION (MOVEMENT_WATER) and usually codify the meaning of *change*. On the other hand, the head *conservation* belong to the semantic field of POSSESSION and codifies the meaning of *maintain*, which can become a semantic relation within the domain of the environment.

In this sense, it would be more accurate to talk about lexical domains, as proposed by the Lexical Grammar Model (Faber & Mairal, 1999). In lexical domains paradigmatic and syntagmatic relations converge. Paradigmatic relations allude to the fact that terms share the same area of meaning, whereas the syntagmatic relations refer to the syntactic behavior of terms. Therefore, the semantic information shared by these areas of meaning can be used to predict the syntactic behavior of their terms (Faber & Mairal, 1999).

4. In this same line, it was shown that syntactic dependency is linked to semantic dependency. In other words, the combination of each modifier separately with the head represented by the verb, which constitutes the semantic core of the MWT, is only possible if a metonymic relation between the modifiers is established. In other words, one modifier is the attribute of the other modifier.

For example, in *water quality management*, the combinations *water management* and *quality management* are possible since there is a part-whole relation between *water* and *quality*. In contrast, if the semantic relation holding between the modifiers is not metonymic, the separate combination of each modifier with the head is not possible. For example, this occurs in the *result_of* and *located_at* relations, because the second modifier cannot represent both modifiers. This case is illustrated by MWTs such as *ocean wave propagation* or *propagación de las ondas de borde*.

This idea of semantic constraints that limit the argument structure was addressed by Pinker (1989), Gabrovšec (2007), and Sanz (2012) *inter alia*, who stated that words combine not only with chosen words, but also with chosen meanings. The lexical domains and paradigmatic and syntagmatic relations of the Lexical Grammar Model (Faber & Mairal, 1999) are also based on this assertion. Nevertheless, Zelinsky-Wibbelt (2012) noted that the semantic aspects enabling a modifier to form an MWT together with the head must still be studied in greater depth. For this reason, the results of our study are both timely and innovative, since they address semantic aspects that influence argument structure.

5. The generic verbs of semantic fields, such as *maintain*, were specified by means of their hyponyms generated in the paraphrases. These verbs help to specify the relations in EcoLexicon (for example, *studies* or *affects*) and thus facilitate access to the semantic content of MWTs. The objective was to emphasize the relevance of micro-contexts in semantic characterization. By making explicit the relation between the predicate and its arguments, it was possible to retrieve the meaning of the MWTs and specify the abstract semantic relation, which ultimately favors the translation of MWTs. An example of how generic verbs are specified with verb paraphrases is shown in Table 6.

Semantic relation				SEDIMENT (attribute: volume)	*maintain*
Paraphrase	law	establishes	(that the)	sediment volume volume of sediments	is conserved is kept remains constant
					MWT

Table 6: *Paraphrase analysis for sediment volume conservation*.

As can be observed, the generic verb *maintain* was specified by means of its hyponyms, *conserve* (*is conserved*), *keep* (*is kept*) and *remain constant* (*remains constant*). These more specific verbs elicited by means of paraphrases allowed a better approach to the semantic content of the MWT since the verb is the core of its meaning.

6. In this sense, even when verbs are hyponyms of the generic verb of a lexical domain, their compatibility with all of the MWTs in this lexical domain depends on the predicate and argument structure of each MWT. In other words, if the hyponyms (e. g. *study, evaluate*) are semantically similar to the verb underlying the head of the MWT (*analyze*), they can be used in the paraphrases of the MWTs that have the same head (*analysis*). For example, *study* and *evaluate* can be used in the paraphrases of MWTs such as *wave height analysis, flood hydrograph analysis, surface temperature analysis*, and *flood risk analysis*.

On the contrary, in the case of more specific verbs (such as *mitigate* or *prevent*, which are hyponyms of *manipulate*, the generic verb underlying the head *management* in *flood risk management*), they may not be compatible with all of the MWTs whose head is *management*. For instance, *mitigate* or *prevent* cannot combine with the MWTs *water quality management* or *water resource management*.

In this same line, an important factor that restricts or permits certain term combinations is semantic prosody, or the negative/positive associations of a word. This is evident in the case of *flood risk management*, since the modifiers *flood* and *risk* have negative connotations and usually combine with positive heads (such as *management, analysis, assessment,* etc.) that palliate or diminish the negative consequences derived from the modifiers. This explains that verbs such as *mitigate* or *prevent* cannot combine with *water quality* or *water resource*, in spite of the fact that the head *management* appears both with *flood risk* and these modifiers.

As previously mentioned, this semantic richness allowed us to specify the general relations. Furthermore, it demonstrates the close relationship between the predicate and its arguments.

7. Additionally, there are interlinguistic correspondences between the languages under study. More specifically, the MWTs whose head or modifiers are the same both in English and Spanish usually combine with terms from the same lexical domains. In addition, the same semantic relation is established in both languages.

management: X *manipulate* WATER **gestión**: X *manipulate* WATER
propagation: X *move_outwards* WATER_WAVE / WATER_WAVE *move_outwards* **propagación**: X *move_outwards* WATER_WAVE / WATER_WAVE *move_outwards*

Table 7: *Semantic interlinguistic correspondences*.

As can be observed in Table 7, the head *management* (*gestión* in Spanish) combines mainly with terms from the lexical domain of WATER. English examples include *water quality management, water resource management,* and *catchment flood management*, whereas Spanish examples are *gestión de la demanda de agua, gestión de los servicios de agua,* and *gestión de la calidad del agua*.

Furthermore, the same semantic relation is usually established in MWTs that have the same head in both languages, as in the case of *propagation/propagación. Propagate/propagar* is a MOVEMENT verb that belongs to the dimension of OUTWARD MOVEMENT (*move_outwards*). As such, it tends to combine with WATER_WAVE, a specific type of movement of a water surface, which can be induced by an external agent.

However, term formation is different in English and Spanish. In English there are more MWTs composed only of noun modifiers, whereas Spanish MWTs tend to have adjective modifiers. For this reason, many equivalent MWTs were not included among the term candidates, since our object of study were three-term noun compounds. For translation purposes, correspondence can only be based on semantic content.

8. When parsing and bracketing, a uniform pattern was observed in each language. Modifiers are placed on the left in English (left bracketing), which is known as pre-modification, while they appear on the right in Spanish (right bracketing), i. e. post-modification, as stated by Sanz (2012). The modifier is composed of two terms, one of which is usually a noun complement (introduced by the preposition *de* in Spanish). This is valuable information that must be taken into account when translating MWTs. Table 8 shows the parsing and bracketing for *flood risk management* and *gestión del riesgo de inundación*, which are equivalent terms. The pre-modification pattern that usually occurs in English can be compared with the post-modification that is typical of Spanish.

flood risk management	
Bracketing	[flood risk] management
Parsing	$[N_{modifier} + N_{head}]_{modifier} + N_{head}$
gestión del riesgo de inundación	
Bracketing	gestión [del riesgo de inundación]
Parsing	N_{head} + [prep. + art. + N_{head} + prep. + $N_{modifier}]_{modifier}$

Table 8: *English pre-modification and Spanish post-modification.*

9. The paraphrases formulated by the experts as well as those extracted from the web provided additional information that was not present in the MWT, thus expanding its meaning. Along these lines, as pointed out by Faber & Mairal (1999: 89), "the description of a verb necessarily includes a specification of the number of arguments, their obligatoriness, and their semantic characteristics". This way, the additional information provided by the paraphrases was informative in regards to the context of the MWTs and it facilitated their translation. Table 9 shows the paraphrase analysis for *wave height analysis*, where additional information was offered.

Semantic relation	X	*study*		WATER_WAVE (attribute: height)		
Paraphrase	X	analyzes observes assesses	(the)	wave height	in a study area (LOCATION)	during a period of time (TIME)
			MWT			

Table 9: *Paraphrase analysis for wave height analysis*.

The information regarding the location and time of the process of *wave height analysis* was not present in the MWT. However, the paraphrases in Table 9, extracted from the Google search engine, provided these data, which are crucial to the description of verbs and micro-contexts (Faber & Mairal, 1999). The example in Table 9 shows that the wave height is analyzed/observed/assessed by X in a study area during a period of time, which makes the meaning of the MWT very explicit. These data are more informative than an MWT only described in terms of semantic roles/relations.

10. A comparison of the paraphrases formulated by experts with those extracted from the Google search engine shows that both of them were very useful, since they specified the abstract semantic relations and facilitated access to the semantic content of the MWTs. However, experts paraphrases were much more specific and more quickly obtained. The problem with the web was the noise, which complicated the extraction of useful information. Nevertheless, the data obtained with both procedures complemented each other and provided a more detailed view of the meaning of the MWTs.

11. The relevance of context for MWT interpretation (Meyer, 1993) became evident in our research. Although theoretical linguistics has attached great importance to noun compound interpretation in context, this aspect has been largely ignored by computational linguistics (Nakov, 2013). Thanks to the definitions of the MWTs offered by the experts, the paraphrases, and the context (accessed in the corpus concordances), our research confirmed that MWTs can be polysemous, which is why context is of crucial importance to the accurate interpretation of MWTs.

As an example, in some cases the lack of context prevented experts from understanding the MWTs. As a result, some proposed paraphrases did not correspond to the meaning of the MWT. Erroneous paraphrases also made it difficult to understand and semantically analyze the MWT.

The importance of context for MWT interpretation is evident in *control de la línea de flotación*. In this case, thanks to the definitions provided by the experts, it was observed that this MWT codify two meanings: *measure* and *manipulate*. However, because of the lack of context, some of the experts did not reflect the polysemy of the MWT in their paraphrases. Table 10 shows the definitions of *control de la línea de flotación* elicited from the Coastal Engineering experts, three of which allude to the meaning of *measure* whereas the other two refer to the meaning of *manipulate*:

***control de la línea de flotación*:**
-Establecer medidas para medir el nivel de agua y aplicar en el diseño de buques.
-Medida de la elevación del nivel del agua de cara a la operatividad portuaria.
-Medición del nivel del agua.
-La línea de flotación marca o separa la parte emergida de la sumergida en un cuerpo flotante. Sería controlar esa línea y que no se den unos límites.
-Mantenimiento de la flotabilidad de un objeto (ej. barco) entre unos valores o umbrales.

Table 10: *Definitions of control de la línea de flotación*.

12. Variation in MWT structures is more frequent in Spanish than in English (e.g., *dispersión de la energía del oleaje* and *dispersión de la energía de una ola*). In addition, the articles and prepositions in Spanish MWTs were not used homogeneously. In this sense, Sketch Engine was more effective in the extraction of English MWTs (3 n-grams), since there is a maximum of 6 n-grams to be extracted. Because of the articles and prepositions in Spanish MWTs, they can be between five and seven n-grams long. As a result, many times the full MWT was not extracted and it was necessary to look at the concordances.

5. Conclusion

This paper provided a perspective on the semantics of MWTs, specifically, three-term noun compounds. These are the most prolific units in domain-specific texts, which is why an accurate assessment of their meaning is a necessary step to understanding specialized texts. In particular, retrieval of the semantic content of these units is crucial for their translation into the target language, since these are units that rarely appear in terminographic resources.

Our perspective on the analysis of MWTs is innovative since specialized knowledge units designating processes were examined by means of verb paraphrases. These paraphrases encoded the propositions underlying the MWTs in our research. In this sense, the meaning of MWTs was retrieved through the analysis of micro-contexts that provide insights into MWT formation and their semantic interpretation in context. Following Nakov & Hearst (2013), we believe that the meaning of MWTs is best understood by specifying the semantic relations in their predicate-argument structure. This was performed by means of verb paraphrases elicited from experts as well as others retrieved from the Internet. The information from these paraphrases was used to further enhance our set of semantic relations.

Our results show that MWTs composed of similar terms usually belong to the same semantic domain and have similar combination patterns. The combination preferences are established from the head to the modifiers as well as in the opposite direction. In addition, syntactic dependency was found to be linked to semantic dependency. Similarly, the semantic aspects enabling a modifier to form an MWT with the head were also addressed.

Although term formation is different in English and Spanish, certain interlinguistic correspondences were observed. In this sense, the MWTs whose head is the same in both languages usually combine with terms from the same lexical domains.

Another important result of our research was the relevance of context for MWT interpretation. It was found that MWTs can be polysemous, and thus expanded context is crucial for understanding them. These results confirm that verbs play a major role in MWTs, as a vehicle for the transmission of specialized knowledge.

Plans for future research include a study of MWTs composed of nouns and adjectives, as well as the analysis of the verbs in a specific lexical domain in relation to MWTs and their structure, and the use of a parallel corpus in order to find equivalent MWTs that could be added to EcoLexicon. Our corpus of Coastal Engineering texts will also be used to design a method of paraphrase extraction and of formalizing semantic contexts for more accurate knowledge representation.

Acknowledgements

This research was carried out as part of project FF2014-52740-P, *Cognitive and Neurological Bases for Terminology-enhanced Translation* (CONTENT), 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 first author. Finally, we would like to thank the Coastal Engineering experts at the Andalusian Environmental Centre (*Centro Andaluz del Medio Ambiente*) for their participation.

References

Baker, C. F., Fillmore, C. J., & Lowe, J. B. 1998. "The Berkeley FrameNet project." *Proceedings of the 17th International Conference on Computational Linguistics*. *ICCL '98*, 86–90. Retrieved from http://www.aclweb.org/anthology/C98-1013

Buendía Castro, M. 2012. "Verb dynamics." *Terminology* 18(2), 149-166. doi: 10.1075/term.18.2.01bue

Buendía Castro, M. 2013. *Phraseology in Specialized Language and its Representation in Environmental Knowledge Resources*. PhD Thesis. Granada: University of Granada.

Buendía Castro, M., & Faber, P. 2015. "EcoLexicon como asistente en la traducción." Corpas Pastor, G., Seghiri Domínguez, M., Gutiérrez Florido, R., & Urbano Medaña, M. (eds.) *VII Congreso Internacional de la Asociación Ibérica de Estudios de Traducción e Interpretación: Nuevos horizontes en los Estudios de Traducción e Interpretación (Comunicaciones completas) / New Horizons in Translation and Interpreting Studies (Full papers) / Novos horizontes dos Estudos da Tradução e Interpretação (Comunicações completas)*. Geneva: Tradulex. Retrieved from http://www.tradulex.com/varia/AIETI7.pdf

Butnariu, C., & Veale, T. 2008. "A concept-centered approach to noun-compound interpretation." *Proceedings of the 22nd International Conference on Computational Linguistics. COLING '08*, 81–88. Retrieved from http://www.aclweb.org/anthology/C08-1011

Butnariu, C., Kim, S. N., Nakov, P., Ó Séaghdha, D., Szpakowicz, S., & Veale, T. 2010. "SemEval-2010 Task 9 : The Interpretation of Noun Compounds Using Paraphrasing Verbs and Prepositions." *Proceedings of the 5th International Workshop on Semantic Evaluation: Recent Achievements and Future Directions*, 100–105. Retrieved from https://www.cl.cam.ac.uk/~do242/Papers/sew09_paraphrase.pdf

Daille, B., Dufour-Kowalski, S., & Morin, E. 2004. "French-English multi-word term alignment based on lexical context analysis." *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, 919-922. Retrieved from http://www.lrec-conf.org/proceedings/lrec2004/pdf/563.pdf

Downing, P. 1977. "On the creation and use of English compound nouns." *Language* 53, 810–842.

Faber, P., & Mairal Usón, R. 1999. *Constructing a Lexicon of English Verbs*. Berlin: Mouton de Gruyter.

Faber, P. 1999. "Conceptual analysis and knowledge acquisition in scientific translation." *Terminologie et Traduction* 2, 97-123. Retrieved from http://lexicon.ugr.es/pub/fab-con

Faber, P. 2009. "The Cognitive Shift in Terminology and Specialized Translation." *MonTI. Monografías de Traducción e Interpretación 1*, 107-134. http://dx.doi.org/10.6035/MonTI.2009.1.5

Faber, P. 2011. “The dynamics of specialized knowledge representation: Simulational reconstruction or the perception-action interface.” *Terminology 17*(1), 9-29. http://dx.doi.org/10.1075/term.17.1.02fab

Faber, P. 2012. *A cognitive linguistics view of terminology and specialized language*. Berlin, Boston: De Gruyter Mouton.

Finin, T. 1980. *The Semantic Interpretation of Compound Nominals*. PhD Thesis. Urbana, Illinois: University of Illinois.

Gabrovšek, D. 2007. “Connotation, semantic prosody, syntagmatic associative meaning: three levels of meaning?” *Elope 4*(1,2), 9-28. doi:10.4312/elope.4.1-2.9-28

Hendrickx, I., Kozareva, Z., Nakov, P., Ó Séaghdha, D., Szpakowicz, S., & Veale, T. 2013. “SemEval-2013 Task 4: Free Paraphrases of Noun Compounds.” *Second Joint Conference on Lexical and Computational Semantics (*SEM): Proceedings of the Seventh International Workshop on Semantic Evaluation* (SemEval 2013) *2*, 138–143. Retrieved from http://www.aclweb.org/anthology/S13-2025

Horsella, M., & Pérez, F. 1991. “Nominal compounds in chemical English literature: Towards an approach to text typology.” *English for Specific Purposes 10*(2), 125-138.

Humbley, J., & García Palacios, J. 2012. “Neology and terminological dependency.” *Terminology 18*(1), 59–85. doi:10.1075/term.18.1.04hum

Kilgarriff, A, Rychly, P., Smrz, P., & Tugwell, D. 2004. “The Sketch Engine.” *Proceedings of the 11th EURALEX International Congress*, 105–116.

Kim, S. N., & Baldwin, T. 2013. “A lexical semantic approach to interpreting and bracketing English noun compounds.” *Natural Language Engineering 19*(3), 385–407. doi:10.1017/S1351324913000107

Lapata, M., & Keller, F. 2005. “Web-based Models for Natural Language Processing.” *ACM Transactions on Speech and Language Processing 2*(2), 1–31. doi:10.1145/1075389.1075392

Levi, J. 1978. *The Syntax and Semantics of Complex Nominals*. New York: Academic Press.

L’Homme, M.-C. 1998. “Caractérisation des combinaisons lexicales spécialisées par rapport aux collocations de langue générale.” *Euralex ’98*, 513-522.

Meyer, R. 1993. *Compound Comprehension in Isolation and in Context: The Contribution of Conceptual and Discourse Knowledge to the Comprehension of German Novel Noun–Noun Compounds*. Linguistische Arbeiten 299. Tübingen: Niemeyer.

Nakagawa, H., & Mori, T. 2003. “Automatic term recognition based on statistics of compound nouns and their components.” *Terminology 9*(1), 201–219. doi:10.1075/term.9.2.04nak

Nakov, P., & Hearst, M. 2006. “Using Verbs to Characterize Noun-Noun Relations.” *Artificial Intelligence Methodology Systems and Applications 4183*, 233–244. doi:10.1007/11861461_25

Nakov, P., & Hearst, M. 2008. “Solving Relational Similarity Problems Using the Web as a Corpus.” *Proceedings of the ACL’08: HLT*, (January), 452–460. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.143.4811

Nakov, P., & Hearst, M. 2013. “Semantic Interpretation of Noun Compounds Using Verbal and Other Paraphrases.” *ACM Transactions on Speech and Language Processing 10*(3), 1-51. Retrieved from http://people.ischool.berkeley.edu/~hearst/papers/acm_tslp_2013.pdf

Nakov, P. 2008. “Paraphrasing Verbs for Noun Compound Interpretation.” *Proceedings of the LREC’08 Workshop: Towards a Shared Task for Multiword Expressions*, 2–

5. Retrieved from http://www.cs.berkeley.edu/~nakov/selected_papers_list/mwe2008.pdf

Nakov, P. 2013. “On the interpretation of noun compounds: Syntax, semantics, and entailment.” *Natural Language Engineering 19*(03), 291–330. doi:10.1017/S1351324913000065

Nastase, V., & Szpakowicz, S. 2003. “Exploring noun-modifier semantic relations.” *Fifth International Workshop on Computational Semantics (IWCS-5)*, 285–301.

Ó Séaghdha, D., & Copestake, A. 2013. “Interpreting compound nouns with kernel methods.” *Natural Language Engineering 19*, 1–26. doi:10.1017/S1351324912000368

Pinker, S. 1989. *Learnability and Cognition*. Cambridge: MIT Press.

Rosario, B., Hearst, M. A., & Fillmore, C. 2002. “The Descent of Hierarchy, and Selection in Relational Semantics.” *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, ACL ’02, (July), 247–254. doi:10.3115/1073083.1073125

Saldanha, G., & O’Brien, S. 2013. *Research Methodologies in Translation Studies*. Manchester: St Jerome Publishing. Retrieved from https://books.google.com/books?id=RuxQAwAAQBAJ&pgis=1

Sanz Vicente, L. 2012. “Approaching secondary term formation through the analysis of multiword units: An English–Spanish contrastive study.” *Terminology 18* (1), 105–127. doi:10.1075/term.18.1.06san

Teubert, W. 2005. “My version of corpus linguistics.” *International Journal of Corpus Linguistics 10*(1), 1-13. doi:10.1075/ijcl.10.1.01teu

Utsumi, A. 2014. “A semantic space approach to the computational semantics of noun compounds.” *Natural Language Engineering 20*, 185-234. doi:10.1017/S135132491200037X

Vanderwende, L. 1994. “Algorithm for automatic interpretation of noun sequences.” *Proceedings of the 15th Conference on Computational Linguistics 2*. *COLING ’94*, 782–788. doi:10.3115/991250.991272

Zelinsky-Wibbelt, C. 2012. “Identifying term candidates through adjective-noun constructions in English.” *Terminology 18*(2012), 226–243. doi:10.1075/term.18.2.04zel

Appendix 1
ANÁLISIS DE TÉRMINOS DE INGENIERÍA COSTERA/ANALYSIS OF COASTAL ENGINEERING TERMS

Nombre/Name:

Sexo/Gender:

Edad/Age:

Estado civil/Marital status:

Nivel educativo/Level of education:

Profesión/Occupation:

Años de experiencia en la profesión/Years of experience within the profession:

Lugar de procedencia/Place of origin:

Idiomas/Languages:

1. Defina brevemente los siguientes términos./Define succintly the following terms.

- **Términos en inglés/English terms**

flood risk management:

water quality management:

water resource management:

catchment flood management:

wave height analysis:

flood hydrograph analysis:

surface temperature analysis:

flood risk analysis:

wave height reduction:

storm damage reduction:

length scale reduction:

surface roughness reduction:

wave energy conservation:

sediment volume conservation:

energy flux conservation:

water level fluctuation:

wind stress fluctuation:

water table fluctuation:

water wave propagation:

shock wave propagation:

ocean wave propagation:

wave energy propagation:

storm surge elevation:

storm surge prediction:

storm surge modeling:

storm surge simulation:

storm surge inundation:

flood risk assessment:

flood hydrograph reconstitution:

- **Términos en español/Spanish terms**

control del flujo de aire:

control de la calidad del agua:

control de la línea de flotación:

control de la contaminación del agua:

control de las oscilaciones del mar:

gestión del riesgo de inundación:

gestión de la demanda de agua:

gestión de los servicios de agua:

gestión de la calidad del agua:

gestión de canales de navegación:

propagación de un tren de ondas:

propagación de la energía del oleaje:

propagación de las ondas de borde:

propagación de las ondas de marea:

disipación de la energía del oleaje:

disipación de la energía por fricción:

disipación de la energía de las olas:

percepción del riesgo de inundación:

adaptación al riesgo de inundación:

cambio del riesgo de inundación:

reducción del riesgo de inundación:

prevención del riesgo de inundación:

modelación de la calidad del agua:

análisis de la calidad del agua:

seguimiento de la calidad del agua:

distribución del transporte de sedimentos:

cálculo del transporte de sedimentos:

estimación del transporte de sedimentos:

control del transporte de sedimentos:

análisis del transporte de sedimentos:

estudio de transporte de sedimentos:

medición del transporte de sedimentos:

2. Parafrasee los siguientes términos empleando un verbo. Por ejemplo, una paráfrasis del término *soil polluting element* sería *element that pollutes the soil*. Si conoce más verbos implicados en el proceso designado por el término, puede indicar más paráfrasis./Paraphrase the following terms using a verb. For example, a paraphrase for the term *soil polluting element* is *element that pollutes the soil*. If you know more verbs that are implied in the process denoted by the term, you can propose more paraphrases.

- **Términos en inglés/English terms**

flood risk management:

water quality management:

water resource management:

catchment flood management:

wave height analysis:

flood hydrograph analysis:

surface temperature analysis:

flood risk analysis:

wave height reduction:

storm damage reduction:

length scale reduction:

surface roughness reduction:

wave energy conservation:

sediment volume conservation:

energy flux conservation:

water level fluctuation:

wind stress fluctuation:

water table fluctuation:

water wave propagation:

shock wave propagation:

ocean wave propagation:

wave energy propagation:

storm surge elevation:

storm surge prediction:

storm surge modeling:

storm surge simulation:

storm surge inundation:

flood risk assessment:

flood hydrograph reconstitution:

- **Términos en español/Spanish terms**

control del flujo de aire:

control de la calidad del agua:

control de la línea de flotación:

control de la contaminación del agua:

control de las oscilaciones del mar:

gestión del riesgo de inundación:

gestión de la demanda de agua:

gestión de los servicios de agua:

gestión de la calidad del agua:

gestión de canales de navegación:

propagación de un tren de ondas:

propagación de la energía del oleaje:

propagación de las ondas de borde:

propagación de las ondas de marea:

disipación de la energía del oleaje:

disipación de la energía por fricción:

disipación de la energía de las olas:

percepción del riesgo de inundación:

adaptación al riesgo de inundación:

cambio del riesgo de inundación:

reducción del riesgo de inundación:

prevención del riesgo de inundación:

modelación de la calidad del agua:

análisis de la calidad del agua:

seguimiento de la calidad del agua:

distribución del transporte de sedimentos:

cálculo del transporte de sedimentos:

estimación del transporte de sedimentos:

control del transporte de sedimentos:

análisis del transporte de sedimentos:

estudio de transporte de sedimentos:

medición del transporte de sedimentos:

3. Responda brevemente a las siguientes preguntas:/Answer succintly the following questions:

-En una escala del 1 al 10, ¿qué nivel de dificultad considera que ha tenido la prueba?/Please rate the difficulty level of the test on a scale from 1 to 10.

-¿Qué entiende por término? ¿Eran términos las unidades sobre las que se le ha preguntado en la prueba?/What do you understand by term? Were terms the units that you have been asked about in the test?

-¿Con qué finalidad cree que se realiza la prueba?/In your opinion, which is the purpose of the test?

-Exprese brevemente su opinión sobre la prueba./Express succintly your opinion on the test.