Towards the Inference of Semantic Relations in Complex Nominals: a Pilot Study

Melania Cabezas-García, Pilar León-Araúz
University of Granada
C/ Buensuceso, 11 18002 Granada (Spain)
{melaniacabezas, pleon}@ugr.es

Abstract

Complex nominals (CNs) (e.g. *wind turbine*) are very common in English specialized texts (Nakov, 2013). However, all too frequently they show similar external forms but encode different semantic relations because of noun packing. This paper describes the use of paraphrases that convey the conceptual content of English two-term CNs (Nakov and Hearst, 2006) in the domain of environmental science. The semantic analysis of CNs was complemented by the use of knowledge patterns (KPs), which are lexico-syntactic patterns that usually convey semantic relations in real texts (Meyer, 2001; Marshman, 2006). Furthermore, the constituents of CNs were semantically annotated with conceptual categories (e.g. *beach* [LANDFORM] *erosion* [PROCESS]) with a view to disambiguating the semantic relation between the constituents of the CN and developing a procedure to infer the semantic relations in these multi-word terms. The results showed that the combination of KPs and paraphrases is a helpful approach to the semantics of CNs. Accordingly, the conceptual annotation of the constituents of CNs revealed similar patterns in the formation of these complex terms, which can lead to the inference of concealed semantic relations.

Keywords: Complex Nominal, Semantic Relation, Terminology

1. Introduction

Complex nominals (CNs) (e.g. *wind power*) are very frequent in English specialized texts (Nakov, 2013). They are distinguished by their syntactic-semantic complexity, since at least two concepts are juxtaposed with no clear indication of the link between them (Rosario et al., 2002). This means that in CNs such as *air pollution* and *oil pollution*, which have the same external form (the head *pollution* combines with a noun modifier), different semantic relations can be established between their constituents (*has_patient* vs. *caused_by*, respectively) (Maguire et al., 2010). The root of this issue is noun packing, which can be addressed by analyzing the formation processes of CNs, involving predicate deletion (e.g. *power system*, instead of *a system produces power*) and predicate nominalization (e.g. *energy transfer*, instead of *energy is transferred*) (Levi, 1978). This paper describes the use of paraphrases conveying the conceptual content of English two-term CNs (Nakov and Hearst, 2006; Butnariu and Veale, 2008; Cabezas-García and Faber, 2017) in the specialized domain of environmental science. Verb paraphrases were used to access the concealed propositions in two-term CNs formed by predicate nominalization and verb deletion. Some of these paraphrases were based on the lexico-syntactic patterns that generally convey semantic relations in real texts (Meyer, 2001; Marshman, 2006). Our goal was to access the semantics of these multi-word terms (MWTs) in order to (i) disambiguate the semantic relation between the constituents of the CN; and (ii) develop a procedure to infer the semantic relations in these MWTs.

2. Complex Nominals and Meaning Access

Complex nominals (CNs) are expressions with a head noun preceded by a modifying element (i.e. a noun or an adjective) (Levi, 1978), e.g. *wind turbine*. CNs can be endocentric, when one term is the head and the other is its modifier (Nakov, 2013) (e.g. *power generation*). Alternatively, they can be exocentric, when the MWT is not a hyponym of one of its elements, and thus appears to lack a head (Bauer, 2008) (e.g. *fire rainbow*).

The semantic relation linking the constituents of CNs is usually implicit because of noun packing. This results in CNs with parallel external forms but different semantic relations, which sometimes can lead to interpretation difficulties. For instance, according to our world knowledge, the semantic relations underlying *fine sand* and *carbonate sand* are interpreted as an attribute of sand (its grain size, which is fine) and the composition of that type of sand (*carbonate*). Therefore, different methods of meaning access are usually employed. On the one hand, verb paraphrases (Nakov and Hearst, 2006) can be used (e.g. *beach erosion* > *the beach was eroded*). These paraphrases, which take the form of a predicate, its mandatory arguments, and its adjuncts (optional complements) (Tesnière, 1976), make the concealed propositions explicit and further specify the semantic relation in CNs. In that case it could be the non-hierarchical semantic relations *has_patient* or that of *erodes*, depending on the granularity of the semantic relation inventory. On the other hand, knowledge patterns (KPs) can also be used to identify semantic relations. KPs are the lexico-syntactic patterns that usually convey semantic relations in real texts (Meyer, 2001; Marshman, 2006) (e.g. erosion *takes place at/occurs/affects/threatens/impacts (on)* the beach; erosion *of/along/on/across* the beach; beach *impacted by* erosion). This paper combines the use of paraphrases and KPs with the semantic annotation of the constituents of CNs with conceptual categories (e.g. *beach* [LANDFORM] *erosion* [PROCESS]) to disambiguate the semantic relation between the constituents of the CN and develop a procedure to infer the semantic relations in these MWTs.

3. Materials

For the purposes of this study, we used an English corpus of specialized environmental texts compiled for the terminological knowledge base EcoLexicon (http://ecolexicon.ugr.es). It was composed mainly of articles and books, and comprised 67 million words on different environmental subdomains, such as Coastal Engineering, Meteorology, Geology, etc. Part of this corpus is now available in Open Corpora (Sketch Engine)[1].

We selected CNs with a nominal or adjectival modifier that designated both entities (i.e. 56 hyponyms of *sand*) and processes (i.e. 57 hyponyms of *erosion*) in order to compare the semantic relations and conceptual categories that are characteristic of these different term types.

We identified relevant term candidates in TermoStat (http://termostat.ling.umontreal.ca/) (Drouin, 2003), namely two-term CNs that were hyponyms of *erosion* and *sand*. Then, Sketch Engine (https://www.sketchengine.co.uk/) (Kilgarriff et al., 2004) was used for term extraction, concordance analysis, paraphrase and KP search, and word sketch analysis. NooJ (http://www.nooj4nlp.net/pages/nooj.html) (Silberztein, 2003) was also employed for the semantic annotation of the constituents of CNs with conceptual categories, as explained below.

4. Corpus-based Semantic Analysis

The aim was to compare the semantic relations that are characteristic of CNs designating entities and processes. Not surprisingly, different semantic relations are established between the constituents of these types of terms. As will be shown, CNs that name a process (e.g. *sea erosion*) encode semantic relations such as *caused_by*, *has_patient*, *has_result*, etc., whereas CNs designating entities (e.g. *carbonate sand*) activate semantic relations such as *composed_of*, *has_origin*, *has_function*, etc. This is not surprising, since such constraints can be explained by the different semantic nature of entities and processes and their natural combinatorial potential (León-Araúz and Faber, 2010). However, what is not so obvious is the kind of constraints that may be inferred from the combination of each particular concept within the same CN. This suggests that conceptual categories play a role in the conceptualization of complex terms, which is directly linked to the notion of 'micro-context'. The head of a CN is considered to open slots that are filled by specific conceptual categories (Rosario et al., 2002; Maguire et al., 2010) that play a semantic role. Thus, the semantic category that is the head of the CN determines what can be done to it by means of the addition of modifiers that fill the slots opened by the head. Micro-contexts refer to this slot filling and are essential to the inference of semantic relations addressed in this study, because similar heads are considered to co-occur with related semantic categories and to evoke similar relations (Maguire et al., 2010).

Since the semantic nature of the head determines its combinatorial potential, CNs representing entities and processes must be differentiated as a first step in the development of a procedure of inference of the semantic relations in CNs. On the one hand, this pilot study focused on a set of CNs that were hyponyms of *erosion*, and thus represented processes, in order to ascertain the different semantic relations activated by the same head. These CNs were then classified according to the conceptual category of their modifiers, some of which are included in Table 1.

WATER BODY + PROCESS		
sea erosion marine erosion	fluvial erosion river erosion	stream erosion glacial erosion
PROCESS + PROCESS		
storm erosion subduction erosion	landslide erosion wind erosion	seepage erosion
LANDFORM + PROCESS		
dune erosion cliff erosion	bluff erosion slope erosion	mountain erosion delta erosion
PART OF LANDFORM + PROCESS		
toe erosion	bed erosion	berm erosion

Table 1: Example of the classification of the hyponyms of *erosion* based on the conceptual category of their modifiers.

On the other hand, an analysis of the conceptual categories combined to form the hyponyms of *sand* demonstrates that processes and entities usually combine with different conceptual categories as well as through different semantic relations. Table 2 shows the classification of some of the hyponyms of *sand* based on the conceptual category of their modifiers.

SIZE + MATERIAL	
fine sand medium sand	coarse sand
LANDFORM + MATERIAL	
beach sand dune sand	nearshore sand littoral sand
MATERIAL + MATERIAL	
carbonate sand silt sand	calcareous sand oil sand
PROCESS + MATERIAL	
nourishment sand construction sand	fill sand filter sand

Table 2: Example of the classification of the hyponyms of *sand* based on the conceptual category of their modifiers.

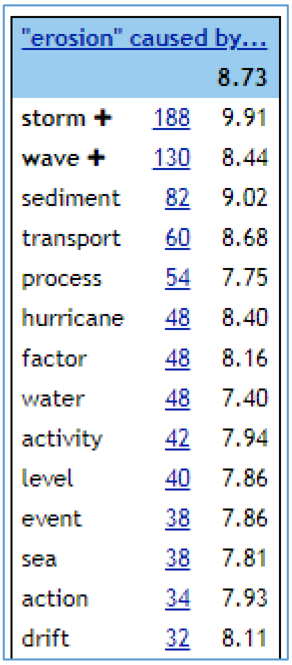
As stated earlier, although hyponymic CNs usually have the same external form, they can encode different semantic relations. At first sight, it might seem obvious that the relation held between landform-related concepts and material-related concepts, for instance, must be that of *located_at* (e.g. *beach sand* > sand *located_at* beach). However, with no world knowledge supporting our inferences, the elicitation of semantic relations may not be as straightforward. For instance, one should be knowledgeable in environmental issues in order to easily understand that sand can be used for filtering purposes in

[1] the.sketchengine.co.uk/open/

the domain of water treatment. Only then, the CN *filter sand* will thus evoke the *has_function* relation.

With a view to eliciting these relations, we performed different types of queries: (1) KP-based word sketches, (2) ws (word sketch) Corpus Query Language (CQL) queries, (3) free paraphrases, and (4) verb paraphrases.

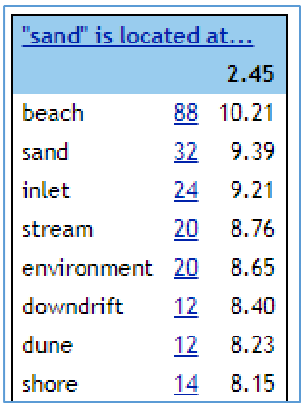
We first observed the KP-based word sketches generated by the head (*erosion* or *sand*) in Sketch Engine. These are automatic groupings of terms that specify the semantic relations between them, based on the application of KP-based sketch grammars (León-Araúz et al., 2016). Figure 1 shows an example of the main agents that cause erosion, which are usually the modifiers of its hyponymic CNs.



"erosion" caused by...		8.73
storm +	188	9.91
wave +	130	8.44
sediment	82	9.02
transport	60	8.68
process	54	7.75
hurricane	48	8.40
factor	48	8.16
water	48	7.40
activity	42	7.94
level	40	7.86
event	38	7.86
sea	38	7.81
action	34	7.93
drift	32	8.11

Figure 1: Sample of KP-based word sketches generated by *erosion*.

The automatic KP-based word sketches allowed the extraction of the internal semantic relations in 6 CNs out of the 57 hyponyms of *erosion*, namely the entities or processes being caused by erosion. As for the hyponyms of *sand*, the semantic relations in 6 out of the 56 CNs were elicited by means of these KP-based word sketches, namely the location of sand (see Figure 2), its origins, its composition, and its functions.



"sand" is located at...		2.45
beach	88	10.21
sand	32	9.39
inlet	24	9.21
stream	20	8.76
environment	20	8.65
downdrift	12	8.40
dune	12	8.23
shore	14	8.15

Figure 2: Sample of KP-based word sketches generated by *sand*.

These low figures can be explained by the fact that KP-based grammars for relations like *has_patient*, which are prevalent in the CNs in our study, have not yet been implemented in the sketch grammars. Furthermore, these word sketches do not include adjectives, which appear in 8 of the hyponyms of *erosion* and 26 of the hyponyms of *sand*. In addition, word sketches are a mere statistic summary of term combinations, i.e. they do not show all the terms actually annotated as a KP-based word sketch in the corpus.

For this reason, subsequent ws CQL queries were performed by combining the annotated word sketches with the specific components of each CN. This allowed us to find further knowledge-rich contexts, i.e. "a context indicating at least one item of domain knowledge that could be useful for conceptual analysis" (Meyer, 2001). Table 3 shows an example of a query that targets the sentences annotated as word sketches between *erosion* and *wind*, where ws means word sketch, "erosion-n" and "wind-n" are the terms that have been annotated as part of a word sketch in the corpus; and "\"%w\".*" means any relation defined in the KP-based sketch grammars. As can be observed, the semantic relation in this CN was found to be *caused_by*.

[ws("erosion-n","\"%w\" .*","wind-n ")]
Moreover, the **wind** *causes* more **erosion** when there are no plants and their roots to hold the soil in place.
It is not generally possible to do anything about the *causes* of the **erosion**, namely high tidal levels, **winds**, rain and wave action.
The land was devastated not only by acid deposition but also by the accumulation of toxic metals in the soil, the clearcutting of forested areas for fuel, and soil **erosion** *caused by* **wind**, water, and frost heave.

Table 3: Query for KPs between *erosion* and *wind*.

The same queries were performed to clarify the semantic relations in the CNs of *sand*. Table 4 illustrates a search for KPs between *sand* and *carbonate*, *oil* or *quartz*, which belong to the same category (MATERIAL). The semantic relation was found to be *made_of* since *sand* is usually composed of these materials.

[ws("sand-n","\"%w\" .*","carbonate-n|oil-n|quartz-n")]
The erosion of granitic mountains and the subsequent transport of the erosion products to the coastline by rivers have led to a very significant fraction (around 70 percent) of the beach **sand** being *composed of* **quartz**.
Everyone knew that, potentially, these **sands** *contained* almost unlimited amounts of **oil**, but they remained untouched until the price of oil rose above fifty dollars a barrel.
Most beach **sand** *contains* calcium **carbonate** (CaCO3) fragments from seashells.

Table 4: Query for KPs between *sand* and *carbonate*, *oil* or *quartz*.

The use of ws CQL queries allowed the extraction of the semantic relations in 14 more CNs of *erosion*, i.e. the relations in 20 CNs of *erosion* were accessed by means of KPs, either in the form of word sketches or ws CQL queries. As for the CNs of *sand*, the semantic relations in

6 more CNs were clarified by means of ws CQL queries. Therefore, the relations in 15 CNs of *sand* were elicited by means of KPs (word sketches or ws CQL queries).

CQL queries are more flexible and allowed the elicitation of the relations in CNs composed of adjectives by manually querying their nominal forms (e.g. *structural erosion* > *structure* + *erosion*). However, this is not always possible, because all adjectives do not have a nominal equivalence (e.g. *fine sand*). The relations extracted by means of CQL queries were *caused_by* and *has_patient*, for the hyponyms of *erosion*. As previously stated, although the *has_patient* relation is not yet implemented in the grammars, KRCs conveying this relation were elicited. This was due to the appearance of verbs such as *cause*, which do not only evoke the causative relation, but can also be part of other relations, as evidenced in the following concordance line: (…) *waves may attack the dune during storms, causing erosion and recession of the dune line*. As for the hyponyms of *sand*, the semantic relations elicited were *located_at*, *attribute_origin*, *made_of*, and *has_function*, which shows that the KP-based sketch grammars are so far more suited to the relations evoked by entities (e.g. location, composition, origin, etc.).

However, given that the KP-based sketch grammars need to be further developed in order to include more KPs and different relations (León-Araúz et al., 2016), these searches were complemented with the query for free paraphrases, i.e. specific CQL queries for co-occurrences of the CN constituents in a ±10 span, which allowed to obtain a general idea of the link between the constituents of the 57 CNs of *erosion* and the 56 CNs of *sand*. Even though the query for free paraphrases was entirely manual, offered more noise, and thus was more time-consuming than the previous steps, paraphrases were very useful for the elicitation of the semantic content concealed in CNs as a result of noun packing. Regarding the hyponyms of *erosion*, the semantic relations extracted by means of free paraphrases were *has_patient*, *has_result*, and *caused_by*, as shown in Table 5, which illustrates a query to extract words between *erosion* and *sea, river, stream* or *glacier*, and vice versa, in a span of 10 tokens.

(meet [lemma= "erosion"] [lemma="sea|river|stream|glacier"] -10 10) within <s/>
Glaciers *are powerful agents of* **erosion**, and are thought to have removed hundreds of feet (meters) from the continental surfaces during the last ice ages.
Guettard (1715), famous for his geological maps, believed the **sea** to *be the major agent in* land **erosion**, and that cliff coasts were the remnants of former extensive hill systems.
Sediment **erosion**, transport and deposition *by* **river**
Rivers *are the major agents of* water **erosion**.
Water, in the form of **streams** and **rivers**, changes mountains *by* **erosion**.
Streams and **rivers** *are significant agents of* **erosion**

Table 5: CQL query for paraphrases of *sea, river, stream* or *glacier* combined with *erosion*.

Alternatively, the relations elicited in the CNs of *sand* were *attribute_size*, *attribute_moisture*, *attribute_origin*, *located_at*, *patient_of*, *has_part*, and *has_function*, as shown in Table 6, which illustrates a query to extract words between *sand* and *nourishment*, *fill*, *filter* or *filtration*, and vice versa, in a span of 10 tokens.

(meet [lemma= "sand"] [lemma="nourishment|fill|filter|filtration"] -10 10) within <s/>
Characterize offshore sand sources to precisely identify the locations where suitable volumes of beach compatible **sand** exist so they may be *utilized for* beach **nourishment** purposes.
New sources of beach-quality **sand** need to be *readied for* beach **nourishment** following severe storm events and for long-term protection from rising sea level.
Adding sufficient **sand** *to* just **fill** the active profile of width at the rate of sea level rise.
It enables permeable **sands** to *function as* biocatalytic **filters**.
Subsequent **filtration** *using* either **sand**, anthracite, or GAC biofilters with EBCT up to 9 min.

Table 6: CQL query for paraphrases of *nourishment*, *fill*, *filter* or *filtration* combined with *sand*.

Thus, KPs and paraphrases turned out to be interdependent since paraphrases complemented the analysis of the semantic relations in CNs and, at the same time, they can be used to improve the grammars underlying the pattern-based word sketches.
Moreover, the word sketches of *erosion* and *sand* as subject or object of the proposition were analyzed in order to extract verb paraphrases (i.e. underlying verbs that further characterize the concealed semantic relation) of all of the CNs that complemented the previous queries. As can be seen in Tables 7 and 8, different verbs were extracted (i.e. *cause to erode, breach, destroy, cause damage, cause to degrade,* and *attack* for *dune erosion*; and *deliver, supply*, and *discharge* for *river sand*), all of which refine the semantic relation underlying the CN.

The same occurs in Estela (also Portugal), where dredging activities in the Cávado River and morphological changes in the river basin reduce littoral drift and *cause* the local **dunes** *to erode*.
When barrier beach **dunes** *are breached* by storm wave attack, the result may be the cutting of a new inlet.
Heavy storm events can *destroy* **dune** and beach nourishments by transporting the sediment to deeper water.
Nor is it intended to cover the well documented methods for the prevention or repair of *damage* to **dunes** *caused by* wind action, recreation or grazing.
This may *cause* the **dunes** *to degrade* resulting in loss of the protection provided by the natural dunes.
Waves approaching the beach during high water level may *attack* the **dune** during storms, *causing* erosion and recession of the dune line.

Table 7: Verb paraphrases of *dune erosion*.

Many of the major dune areas were originally formed several thousands of years ago from **sand** produced by the action of glaciers and *delivered to* the coast *by*

rivers in the last Ice Age.
Sand *supplied from* the **river** has mainly been transported by eastward longshore sand transport.
Fine **sand**, silt, and clay *transported by* the **river** and deposited on the floor of a sea or lake beyond the main body of a delta.
Large quantities of **sand** and especially silt and clay *are discharged by* the **river** and accumulate along the shore margins.

Table 8: Verb paraphrases of *river sand*.

After combining KPs and free and verb paraphrases, we observed that the noise and ambiguities produced by KPs and paraphrases were counteracted by combining both procedures. This was found to lead to the disambiguation of CNs. Furthermore, although paraphrases are eminently manual, they can be very valuable for the improvement and refinement of KP-based word sketches (León-Araúz et al., 2016), which can lead to further automation of the process.

5. Extrapolation to Other Complex Nominals

To ascertain whether the semantic relations usually encoded between certain conceptual categories are also reproduced in other CNs composed of terms belonging to the same categories, we extrapolated the semantic analysis of the hyponyms of *sand* and *erosion* to other CNs of the same type. For that purpose, a dictionary was created in NooJ to lemmatize and annotate the terms according to a previously defined set of conceptual categories. This set still needs to be refined (as will be shown) and so far entities have been classified into more fine-grained categories than processes. Then we performed queries based on the combination of most prototypical categories found in the CNs of *sand* and *erosion*: 1) LANDFORM + MATERIAL, WATER BODY + MATERIAL, MATERIAL + MATERIAL, PROCESS + MATERIAL, as found in the hyponyms of *sand*; and 2) LANDFORM + PROCESS, WATER BODY + PROCESS, PROCESS + PROCESS, as found in the hyponyms of *erosion*.

As for the first set of conceptual combinations (based on those found in the CNs of *sand*), 100% of CNs formed by the semantic patterns LANDFORM + MATERIAL (Table 9) and WATER BODY + MATERIAL (Table 10) shared the relation *has_location*.

<LANDFORM> ***location_of***	<MATERIAL> ***has_location***
basin beach bluff coast dune reef shore seabed	sediment
beach channel delta dune floodplain plain shore seafloor	deposit
seafloor reef cliff	rock
shore	gravel
delta	silt

Table 9: The semantic relation *has_location* in CNs formed by the semantic pattern LANDFORM + MATERIAL.

<WATER BODY> ***location_of***	<MATERIAL> ***has_location***
bay lagoon lake marsh swamp	deposit
lake pond river lagoon bay	sediment
lake bay	mud
river	gravel

Table 10: The semantic relation *has_location* in CNs formed by the semantic pattern WATER BODY + MATERIAL.

In the CNs formed by the semantic pattern MATERIAL + MATERIAL (Table 11), nearly all of them shared the relation *made_of*. Only one exception was found out of 75 other CNs analyzed.

<MATERIAL> ***makes_up***	<MATERIAL> ***made_of***
cement clay lime mud	mortar
clay granite	rock
clay coal gravel iron loess mud peat salt sludge	deposit
coral peat quartz	sediment

Table 11: The semantic relation *made_of* in CNs formed by the semantic pattern MATERIAL + MATERIAL.

The exception was *rock salt*, which is not a type of salt made of rock but a type of salt that forms with the appearance of a rock.

In the CNs formed by the semantic pattern PROCESS + MATERIAL (Table 12), 70% of them shared the relation *result_of* and 30% the relation of *has_function*. As will be shown, the type of process (indicated in small caps under the concepts) clearly determines the type of relation held. Therefore, if processes are classified in more specific categories, such as natural processes and artificial processes, the semantic pattern MATERIAL + PROCESS can be disambiguated in 100% of the CNs analyzed.

<PROCESS> ***has_result***	**<MATERIAL>** ***result_of***
accretion erosion metamorphism <NATURAL PROCESS>	sediment
avalanche corrosion flood mudsflow tsunami <NATURAL PROCESS>	deposit
aeration digestion waste <NATURAL PROCESS>	sludge

<PROCESS> ***function_of***	**<MATERIAL>** ***has_function***
compacting <ARTIFICIAL PROCESS>	sediment
composting <ARTIFICIAL PROCESS>	sludge
construction <ARTIFICIAL PROCESS>	aggregate
construction <ARTIFICIAL PROCESS>	rock
construction <ARTIFICIAL PROCESS>	mortar
construction <ARTIFICIAL PROCESS>	stone
construction <ARTIFICIAL PROCESS>	steel

Table 12: The semantic relations *result_of* and *has_function* in CNs formed by the semantic pattern PROCESS + MATERIAL.

As for the second set of conceptual combinations (based on those found in the CNs of *erosion*), approximately 90% of the CNs formed by the semantic pattern LANDFORM + PROCESS shared the relation *has_patient*. Once again, this figure was 100% when the PROCESS category was further refined to LOSS PROCESS and ADDITION PROCESS (see Table 13). However, there were some exceptions, such as *volcanic eruption*, which did not activate the relation *has_patient*, since volcano cannot be considered the patient of eruption but its agent.

<LANDFORM> ***patient_of***	**<PROCESS>** ***has_patient***
soil-aquifer	treatment <CHANGE>
alluvium	transport <MOVEMENT>
aquiclude beach	recharge <ADDITION>
beach	accretion <ADDITION>
beach dune shore cliff	recession <LOSS>
beach	replenishment <ADDITION>
beach dune inlet	stabilization <CHANGE>
canyon channel	flood <ADDITION>
channel	dredging <LOSS>
channel inlet	shoaling <CHANGE>
dune	construction <ACTIVITY>
dune	deflation <LOSS>
land	desertification <CHANGE>
land	subsidence <MOVEMENT>

Table 13: The semantic relation *has_patient* in CNs formed by the semantic pattern LANDFORM + PROCESS.

In the same manner, CNs formed by WATER BODY + PROCESS establish the pattern WATER BODY *causes*

PROCESS when the process entails an addition or movement (e.g. *lake flooding, river flooding, river deposition, runoff infiltration*, etc.). In contrast, when the process involves a change, the WATER BODY is *the_patient_of* the PROCESS (e.g. *lake acidification, lake eutrophication, river pollution*, etc.) (Table 14).

<WATER BODY> ***causes***	**<PROCESS>** ***caused_by***
lake	flooding <ADDITION>
river	flooding <ADDITION>
river	deposition <ADDITION>
river	transport <MOVEMENT>
runoff	infiltration <MOVEMENT>
<WATER BODY> ***patient_of***	**<PROCESS>** ***has_patient***
lake	acidification <CHANGE>
lake	eutrophication <CHANGE>
river	pollution <CHANGE>

Table 14: The semantic relations *caused_by* and *has_patient* in CNs formed by the semantic pattern WATER BODY + PROCESS.

Finally, in most cases the PROCESS + PROCESS pattern was found to encode the relation *causes* in weather- or water-related processes (e.g. *cyclone storm, hurricane storm, hurricane flooding, storm flooding, seepage erosion*, etc.). Nevertheless, exceptions were also present, mainly when there was human intervention in the process. Then, a PROCESS is usually *represented_by* another PROCESS (e.g. *flood simulation* or *erosion simulation*), or a PROCESS is the *patient_of* another PROCESS (e.g. *flood control*) (Table 15).

<PROCESS> ***causes***	**<PROCESS>** ***caused_by***
cyclone <WEATHER PHENOMENON>	storm <WEATHER PHENOMENON>
hurricane <WEATHER PHENOMENON>	storm <WEATHER PHENOMENON>
hurricane <WEATHER PHENOMENON>	flooding <WATER PHENOMENON>
storm <WEATHER PHENOMENON>	flooding <WATER PHENOMENON>
seepage <WATER PHENOMENON>	erosion <LOSS>
<PROCESS> ***represented_by***	**<PROCESS>** ***represents***
flood <WATER MOVEMENT>	simulation <HUMAN INTERVENTION>
erosion <LOSS>	simulation <HUMAN INTERVENTION>
<PROCESS> ***patient_of***	**<PROCESS>** ***has_patient***
flood <WATER MOVEMENT>	control <HUMAN INTERVENTION>

Table 15: The semantic relations *caused_by*, *represents*, and *has_patient* in CNs formed by the semantic pattern PROCESS + PROCESS.

The different semantic patterns of complex term formation highlight the fact that a fine-grained set of conceptual categories is essential in the inference of the semantic relations in CNs. Indeed, all examples in this section call for the recategorization of processes in deeper semantic levels.

6. Conclusions

CNs usually have similar external forms, although they encode different semantic relations. Our goal was to access the semantic relation linking the constituents of the CNs with a view to developing a procedure to infer the semantic relations in other similar CNs. For that purpose, we studied a set of 57 CN hyponyms of *erosion*, which represented environmental processes, and 56 CN hyponyms of *sand*, which designated environmental entities. After classifying the CNs based on the conceptual categories of the modifiers (e.g. LANDFORM, WATER BODY, etc.), we used KPs and paraphrases to semantically analyze the CNs. This analysis based on conceptual categories and semantic relations was extrapolated to other CNs formed by the same categories. It was observed that with a few exceptions, the semantic relations codified in the hyponyms of *erosion* and *sand* were parallel to those of the CNs formed by the same conceptual categories. The semantic analysis of CNs by means of KPs, paraphrases and conceptual categories was found to be a valuable starting point towards the inference of the concealed semantic relation in CNs.

This research can be applied to the inference of semantic relations in different domains and languages, as well as to the translation of CNs since there is a need to render terms into languages other than English. In future work, CNs formed by more than two constituents will be studied, since these longer structures give an insight into the combinatorial potential of semantic categories and their concealed semantics.

7. 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.

8. Bibliographical References

Bauer, L. (2008). Les composés exocentriques de l'anglais. In D. Amiot (Ed.), *La composition dans une perspective typologique*. Arras: Artois Presses Université, pp. 35–47.

Butnariu, C., and Veale, T. (2008). A concept-centered approach to noun-compound interpretation. In Proceedings of the 22nd International Conference on Computational Linguistics. COLING '08, pages 81–88.

Cabezas-García, M. and Faber, P. (2017). Exploring the Semantics of Multi-word Terms by Means of Paraphrases. In M. A. Candel-Mora & C. Vargas Sierra (Eds.), *Temas actuales en terminología y estudios sobre el léxico*. Granada: Comares, pp. 193-217.

Drouin, P. (2003). Term extraction using non-technical corpora as a point of leverage, *Terminology*, 9(1): 99-117.

Kilgarriff, A., Rychly, P., Smrz, P., and Tugwell, D. (2004). The Sketch Engine. In G. Williams & S. Vessier (Eds.), *Proceedings of the Eleventh EURALEX International Congress.* Lorient: EURALEX, pp. 105–116.

León-Araúz, P., San Martín, A., and Faber, P. (2016). Pattern-based Word Sketches for the Extraction of Semantic Relations. In *Proceedings of the 5th International Workshop on Computational Terminology* (Computerm2016), pages 73-82. Osaka, Japan: COLING 2016.

León-Araúz, P. and Faber, P. (2010). Natural and contextual constraints for domain-specific relations. In V. Barbu Mititelu & V. Pekar (Eds.), *The Workshop Semantic Relations, Theory and Applications*, pp. 12-17.

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

Maguire, P., Wisniewski, E. J., and Storms, G. (2010). A corpus study of semantic patterns in compounding. *Corpus Linguistics and Linguistic Theory*, 6(1): 49–73.

Marshman, E. (2006). Lexical Knowledge Patterns for Semi-automatic Extraction of Cause-effect and Association Relations from Medical Texts: A Comparative Study of English and French. PhD Thesis, Université de Montréal, Montréal.

Meyer, I. (2001). Extracting Knowledge-Rich Contexts for Terminography: A Conceptual and Methodological Framework. In D. Bourigault, C. Jacquemin, & M.-C. L'Homme (Eds.), *Recent Advances in Computational Terminology*. Amsterdam: John Benjamins, pp. 279-302.

Nakov, P. and Hearst, M. (2006). Using Verbs to Characterize Noun-Noun Relations. *Artificial Intelligence Methodology Systems and Applications*, 4183: 233-244.

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

Rosario, B., Berkeley, U. C., Hearst, M. A., and Fillmore, C. (2002). The Descent of Hierarchy, and Selection in Relational Semantics. *ACL*: 247–254.

Silberztein M. (2003). NooJ Manual. Available at: www.nooj4nlp.net.

Tesnière, L. (1976). Eléments de syntaxe structurale. Klincksieck, Paris.

9. Language Resource References

LexiCon Research Group (2017). EcoLexicon.