Semantic Representation of Context for Description of Named Rivers in a Terminological Knowledge Base

Juan Rojas-Garcia*

Department of Translation and Interpreting, University of Granada, Granada, Spain

The description of named entities in terminological knowledge bases has never been addressed in any depth in terminology. Firm preconceptions, rooted in philosophy, about the only referential function of proper names have presumably led to disparage their inclusion in terminology resources, despite the relevance of named entities having been highlighted by prominent figures in the discipline of terminology. Scholars from different branches of linguistics depart from the conservative stance on proper names and have foregrounded the need for a novel approach, more linguistic than philosophical, to describing proper names. Therefore, this paper proposed a linguistic and terminological approach to the study of named entities when used in scientific discourse, with the purpose of representing them in EcoLexicon, an environmental knowledge base designed according to the premises of Frame-based Terminology. We focused more specifically on named rivers (or potamonyms) mentioned in a coastal engineering corpus. Inclusion of named entities in terminological knowledge bases requires analyzing the context that surrounds them in specialized texts because these contexts convey specialized knowledge about named entities. For the semantic representation of context, this paper thus analyzed the local syntactic and semantic contexts that surrounded potamonyms in coastal engineering texts and described the semantic annotation of the predicate-argument structure of sentences where a potamonym was mentioned. The semantic variables annotated were the following: (1) semantic category of the arguments; (2) semantic role of the arguments; (3) semantic relation between the arguments; and (4) lexical domain of the verbs. This method yielded valuable insight into the different semantic roles that named rivers played, the entities and processes that participated in the events educed by potamonyms through verbs, and how they all interacted. Furthermore, since arguments are specialized terms and verbs are relational constructs, the analysis of argument structure led to the construction of semantic networks that depicted specialized knowledge about named rivers. These conceptual networks were then used to craft the thematic description of potamonyms. Accordingly, the semantic network and the thematic description not only constituted the representation of a potamonym in EcoLexicon, but also allowed the geographic contextualization of specialized concepts in the terminological resource.

Keywords: named river, frame-based terminology, terminological knowledge base, analysis of predicate-argument structure, semantic network, thematic description, specialized knowledge representation, geographic contextualization

INTRODUCTION

In linguistics, the convention has been established that a common noun, such as *river*, designates a category or class of individuals. Therefore, the meaning of *river* can be factorized to identify the multitude of individual rivers that are designated by this word. Nonetheless, a proper name, such as *Nile River*, is seen as the linguistic representation of an individual, namely, a unique entity in the world. Hence, *Nile River*, as a member of the class evoked by *river*, can be described but not defined. Based on such considerations, Sager (1990, p. 70) remarks that a proper name holds an individualizing value, whereas a common noun fulfills a classificatory function.

In the philosophy of language, a proper name is generally conceived, albeit with variations, as a linguistic expression that designates one and only one entity in the world because a fixed relation between the linguistic expression and that entity in the world can be established. From this basic conception, two schools of thought have emerged, namely, Referentialism and Predicativism. Referentialists defend that the only semantic function of a proper name is to designate an individual, and this referent constitutes its semantic content (Mill, 1843/2002, p. 21–22; Frege, 1892/1952, p. 54; Russell, 1905/1988; Wimmer, 1973/2011, p. 77; Kripke, 1980). In contrast, predicativists depart from this conservative stance, and start from the premise that a proper name is a type of common noun. They thus argue that the semantic function of a proper name is to designate properties of an individual, and that this set of properties comprises its semantic content (Quine, 1960; Burge, 1973; Elugardo, 2002; Matushansky, 2008; Fara, 2015).

The semantics of proper names thus remains a controversial issue, despite the fact that there is hardly any conceivable aspect which has not been exhaustively reconsidered. A comprehensive survey of the different philosophical and linguistic points of view can be found in Van Langendonck (2007, Ch. 1). However, in the following section "Proper Names in Linguistics," we hold a brief discussion on the topic to emphasize its complexity and contextualize the real objective of this paper.

Our aim is thus to propose a linguistic and terminological approach to the study of named entities when used in scientific discourse, with the purpose of representing them in EcoLexicon (Faber et al., 2016; San Martín et al., 2020). This is a digital terminological knowledge base (TKB) on environmental sciences, designed according to Frame-based Terminology (Faber, 2012, 2015), the theoretical framework of this research. Frame-based Terminology is a cognitive theory of terminology that contextualizes concepts in frames, also called semantic networks or knowledge structures in this paper, and is based on corpus analysis. It is worth clarifying that a TKB is a resource that describes "the concepts and terms of specialized knowledge domains for users with linguistic and/or cognitive needs" (Faber and León-Araúz, 2016, p. 2), namely, it represents specialized knowledge either in a relational database or in an ontology (Temmerman and Kerremans, 2003; Roussey et al., 2018, p. 228), and contains specialized concepts with their definitions, the semantic relations that link them, and terms that lexicalize concepts in different languages or language communities (Condamines, 2018, p. 338).

This study focuses on potamonyms (i.e., the proper name of rivers, according to Room, 1996, p. 84–85), and analyzes the predicate-argument structure of sentences that mention named rivers in a coastal engineering corpus. It should also be pointed out that the methods to analyze potamonyms could be applied to named entities in other specialized domains, such as planets in astronomy, named bays and beaches in coastal engineering, named volcanos in vulcanology, named lakes and wetlands in limnology, named islands in nisology, and named rivers in potamology.

It is our assertion that named rivers, in the coastal engineering domain, have meaning—not only a referential function—, which is encyclopedic in nature according to cognitive linguistics (Evans, 2019, Ch. 15). This meaning thus encompasses dictionary knowledge (i.e., the lexical meaning of the term *river*) and encyclopedic knowledge, which corresponds to the specialized knowledge that coastal engineering texts convey about named rivers. In fact, named rivers hold a large number of semantic relations (e.g., *causes*, *improves*, *takes_place_in*, or *has_function*) that link them to other knowledge units, or terms, in the coastal engineering domain, as shall be seen. These terms correspond to a wide range of features, such as processes (e.g., *sediment supply*, *salinity intrusion*, *siltation*, and *freshwater input*), entities (e.g., *salt march*, *soft mud*, *dam*, and *jetty*), and attributes (e.g., *discharge rate*, *evaporation*, *sediment load*, and *runoff*). All of these designate concepts that are directly related to a named river. These concepts highlight and reinforce the specific nature and behavior of a named river, and differentiate it from other named rivers. Examples of this are provided in sections "Named Entities in Terminology" and "Results."

In specialized discourse, verbs are means of such features to each named river. These verbs, which function as relational constructs, lead to the creation of a *semantic network*, or *semantic frame*, that represents specialized knowledge about a named river in the coastal engineering domain. This semantic network takes the form of a set of situational elements, namely, concepts and the semantic relations that link them.

For this reason, named rivers should have a *thematic description* in TKBs on the environment. The thematic description of a named entity is a textual explanation crafted from its semantic network, which depicts its relational behavior in a specialized domain. The thematic description pertains to the specialized domain in which the named entity has been analyzed because of its multidimensionality (Rogers, 2004), which involves that it is described depending on perspective and subject fields. Consequently, the same named river could have more than one description in a terminological resource, based on contextual constraints. One of the most important types of contextual constraint in terminology are the thematic constraints imposed by a knowledge domain, such as coastal engineering, hydrology, or potamology. As a result, the thematic description of a named entity is similar to the flexible approach to terminological definitions to represent thematic variation that proposed San Martín (2021). As an example of thematic description, that of the Salinas River (in California) in the coastal engineering domain, which is provided later, can be usefully summarized as follows: *Sea level rise is causing dune erosion of Monterey*

Bay beaches to progress at such a high rate that the sediments discharged by the Salinas River are not enough to alleviate the coastal erosion of the bay.

It goes without saying that the thematic description of a proper name is not to be confused with the *definite description* of a proper name, to which grammarians, such as Quirk et al. (1985, p. 294), and philosophers of language, such as Russell (1905/1988), allude. The definite description of a proper name is a noun phrase, or paraphrase, that also makes reference to the same unique entity in the world to which the proper name refers. For instance, for the proper name *Nile River*, a definite description could be *the most important river to Ancient Egypt*; for the proper name *Joe Biden*, a definite description could be *the President of the United States in June 2022*; and the noun phrase *this odd neighbor* acts as a definite description when singles out an individual in a situated context.

Inclusion of named entities, such as named landforms, in TKBs requires analyzing the context that surrounds them in specialized texts because these contexts transmit specialized knowledge about named entities. Applying the proposal by Faber and León-Araúz (2016) for the parameterization of context, this paper thus analyzes the local syntactic and semantic contexts that surround potamonyms in coastal engineering texts, and describes the semantic annotation of the predicate-argument structure of sentences where a potamonym is mentioned. The semantic variables annotated are: (1) Semantic category of the arguments; (2) semantic role of the arguments; (3) semantic relation between the arguments; and (4) lexical domain of the verbs. The findings prove that this linguistic and terminological approach to the study of named entities in scientific discourse facilitates their representation in a TKB designed according to the framework of Frame-based Terminology.

More specifically, the results, on the one hand, allow us to draw conclusions on how each lexical domain of the verbs employed in the context of potamonyms is configured, namely, the specific combination of semantic roles and categories, and the semantic relation encoded by their different patterns of combination. This method also provides valuable information on the different semantic roles named rivers play in the coastal engineering domain, the entities and processes that participate in the events educed by potamonyms through verbs, and how they all interact.

On the other hand, since arguments are specialized terms and verbs are relational constructs, the analysis of argument structure leads to the construction of semantic networks that depict specialized knowledge about potamonyms in the coastal engineering domain. These conceptual networks are then used to craft the thematic description of potamonyms. Accordingly, the semantic network and the thematic description not only constitute the representation of a potamonym in EcoLexicon, but also allow the geographic contextualization of specialized concepts of the coastal engineering in the terminological resource.

The geographic contextualization of a specialized concept should provide a context representation, in the form of a semantic network, that covers a background situation in which the concept is embedded. In this sense, the geographic contextualization we are referring to consists in viewing the specialized concept from a situation in which the concept is related to specific named geographic entities, such as rivers and bays, because it is involved in an environmental problem which affects those geographic entities. For instance, the geographic contextualization of the SEA LEVEL RISE concept in the coastal engineering domain, as shall be seen in the section "Results," would show a semantic network with situational elements (i.e., concepts and semantic relations) that would facilitate to represent and understand that sea level rise is causing dune erosion of Monterey Bay beaches to progress at such a high rate that the sediments discharged by the Salinas River are not enough to alleviate the coastal erosion of the bay.

The remainder of this paper is organized as follows. The section "Proper Names in Linguistics" summarizes different viewpoints in linguistics and the philosophy of language regarding the semantics of proper names. The section "Named Entities in Terminology" deals with the lack of named landforms in environmental terminology resources, gives reasons for this oversight, explains why named landforms should be included in terminological resources, and describes a set of principles to address the issues of categorization and inheritance, which arise upon representing named entities in a TKB. In section "Semantic Analysis From the Perspective of Frame-Based Terminology," the fundamental premises of Frame-based Terminology are explained, according to which the analysis of local syntactic and semantic contexts of potamonyms in coastal engineering texts is performed. The section "Materials" describes the coastal engineering corpus where the potamonyms are mentioned, the Geonames database for the automatic recognition of potamonyms in the corpus, and the INCEpTION tool for the semantic annotation of local contexts. The section "Methodology" details the semantic variables annotated in the corpus, their values, and the inter-annotator agreement. The section "Results" reports the findings of the analysis of local contexts with regard to the semantic roles played by potamonyms, the semantic relations they hold with other arguments in the sentences, the semantic configuration of predicate lexical domains, the construction of two river-evoked semantic frames, the thematic description of the named rivers in both semantic networks, and the geographic contextualization of two specialized concepts integrated in those frames. Critical reflection on the semantic behavior of named rivers in the coastal engineering domain is provided in the section "Discussion." Finally, the section "Conclusion" presents the conclusions derived from the semantic analysis of potamonyms, as well as plans for future research.

PROPER NAMES IN LINGUISTICS

Although lexicographers generally tend not to include proper names in conventional dictionaries, they do compile special dictionaries for them (e.g., *Placenames of the World* dictionary by Room, 2013). This indicates that most linguists believe proper names to be linguistic units that pertain to the lexicon of a language (Gardiner, 1940, p. 32–34; Quirk et al., 1985,

p. 288; Cruse, 2000, p. 315–318; Levinson, 2003; Bennett and Agarwal, 2007; Tenbrink, 2007; Stock et al., 2019). Nevertheless, they are "semantically different from so-called appellative words (roughly corresponding to common nouns), so that we need different techniques and kinds of description for the meanings of proper names versus appellatives." (Evans and Wimmer, 1990, p. 261). This need for a novel approach, which is more linguistic than philosophical, to the description of proper names is also recognized by other scholars.

The philosophers of language Strawson (1974/2004) and Searle (1969) came out in favor of a pragmatics and discourse view of proper names, and stressed that philosophers should consider the usage of proper names in natural language discourse, rather than only focusing on decontextualized short sentences or parts of sentences, for the elaboration of theories of proper names. Strawson's and Searle's claim possibly stemmed from the fact that context was systematically neglected in linguistic accounts for a long time since it was regarded as being too chaotic to be objectively described (Ervin-Tripp, 1996, p. 35).

It is thus hardly surprising that, as observed by the linguist Sjöblom (2006), the status of proper names in linguistics has not been satisfactorily addressed because the issue has been dominated by the philosophy of language. Hence, from the perspective of cognitive linguistics, she asserts that proper names are words that have meanings because they are inserted into the network of meanings that exists in our mind. In this way, Sjöblom's (2006) view is in line with the principles of Frame-based Terminology (Faber, 2012, 2015), the theoretical framework of this paper.

Similarly, the onomastician Van Langendonck (2007, p. 2–3) states that "theoretical linguists have often treated proper names as the poor cousin of other grammatical categories. […] Onomasticians, however, have sometimes forgotten that proper names are part of the system of natural languages. Both onomasticians and linguists should be aware of the fact that proper names are words which deserve linguistic attention in the first place."

NAMED ENTITIES IN TERMINOLOGY

Theoretical Principles for the Representation of Named Entities in Frame-Based Terminology

As previously mentioned, for the representation of a named entity in a TKB, we propose both the construction of a semantic network that reflects the relational behavior of the named entity with other concepts in a specialized domain, and a thematic description that is a textual explanation of the relational behavior of the named entity, elaborated from its semantic network. This is supported not only in the cognitive linguistics approach adopted by Sjöblom (2006) for proper names, but also in the theory of proper names formulated by Searle (1983).

Searle (1983, Ch. 9) points out that the reference made by a proper name not only includes some kind of necessary and sufficient knowledge about the referent (e.g., its semantic category), but also descriptive knowledge on peripheral aspects

related to the referent, provided that this conceptual content helps text senders to refer. Namely, everything that text senders semantically know about the referent helps them to accomplish their intention to refer. In this sense, Searle's view is in consonance with Michalski (1991), who states that the context of a concept (i.e., a named entity in our case) is the set of concepts that contribute significantly to describe its intended features.

Thus, the semantic network of a named entity, on the one hand, must represent the conceptual structure that underlies its usage in specialized discourse, according to Sjöblom (2006); and on the other hand, the network must be endowed with the explanatory adequacy that Searle (1983, Ch. 9) postulates. These two principles for the construction of the semantic network of a named entity substantiate Frame-based Terminology (see section "Semantic Analysis From the Perspective of Frame-Based Terminology"). Therefore, a terminological resource designed according to this framework enables users to understand the relevance of a named entity for a subject field by giving them access to the necessary information to activate the knowledge structure in which the named entity is integrated. In this way, users can acquire background knowledge about the named entity necessary in communicative situations, such as specialized translation (Faber, 2012).

Lack on Named Landforms in Environmental Terminology Resources

In terminology work, the description of named entities is a theoretically accepted activity, as evidenced by Sager (1990, p. 68–71), and the international standard for terminology work developed by the International Organization for Standardization (ISO) (2009, p. 36–37) (ISO 704: 2009). To ensure greater clarity, it is worth mentioning that this standard distinguishes between *proper name* and *appellation* for the designation of an *individual concept* (i.e., unique entity in the world, also referred to as individual entity in the standard, or named entity in this study). Accordingly, an appellation corresponds to a definite description (see section "Introduction"), used in a subject field, to designate an individual concept. For instance, the appellation (or definite description) *the United Nations Commissioner for Human Rights in June 2022* and the proper name *Michelle Bachelet* designate the same individual concept; the appellation *Il Duce* and the proper name *Benito Mussolini* also designate the same individual concept.

However, on a practical level, named landforms, such as rivers, bays, and beaches, are not represented in terminological resources on the environment. In our opinion, reasons for this absence include the following:

1. Firm preconceptions, rooted in philosophy, as to what named entities are, have presumably led to named landforms (e.g., Salinas River, Monterey Bay, Sunset Beach) being regarded as mere instances, with only a referential function, of categories such as RIVER, BAY, or BEACH. Their relational behavior with other concepts in a specialized knowledge domain has never been semantically described in any depth. Therefore, terminologists have been inclined to believe that

the inclusion of the concepts of RIVER, BAY, or BEACH was sufficient. This belief doubtlessly applies to other named entities in specialized discourse as well.

2. Even though the inclusion of named landforms (and, in general, relevant named entities to a subject field) is justified, their semantic representation depends on knowing which concepts are semantically related to each of them, how those concepts are linked to each other, and which semantic relations should be included in the semantic network of each named landform. This is evidently a time-consuming task since terminologists rarely use natural language processing systems beyond corpus query tools such as Sketch Engine (Kilgarriff et al., 2004). On the other hand, although terminologists apply those natural language processing systems, the validation of the (semi-)automatically extracted information about a single named entity before storing it in a TKB is also a labor-intensive task.

3. The lack of clear guidelines for terminologists about how to deal with named entities has meant that their representation in TKBs is not a priority. In fact, although the relevance of named entities to certain specialized domains has been highlighted by prominent figures in the discipline of terminology (Sager, 1990, p. 68–71; Faber and León-Araúz, 2014; Faber, 2015, p. 26–27; L'Homme, 2020, p. 60–61), as far as we know, no research work has yet addressed in any depth how the description of a named entity, significant to a subject field, should be crafted in terminological resources.

Named landforms, among other named entities, are frequently found in specialized texts on environment. However, their representation in specialized knowledge resources has received little research attention. This is evident by the lack of named landforms in terminological resources for the environment such as EcoLexicon,[1] DiCoEnviro,[2] GEMET,[3] or FAO Term Portal.[4]

In contrast, AGROVOC[5] includes a list of named landforms with hyponymic information (only the semantic relation *type_of*), whereas ENVO[6] provides descriptions of named landforms with only geographic details (e.g., geographic coordinates, and rivers that discharge into a certain bay), and minimal semantic information consisting of the relations *located_in*, and *tributary_of* in the case of named rivers and bays. Although the ENVO resource includes named landforms with their descriptions, these correspond to general knowledge focused on geographic data. This type of information does not permit users to understand either the pertinence of a named landform to a certain domain of specialized knowledge such as coastal engineering, hydrology, or sedimentology, or what relation the named landform holds to specialized concepts of a subject field.

Why Named Landforms Should Be Included in Terminological Resources

So far, most TKBs have limited themselves to representing concepts such as RIVER, BAY, or BEACH, on the questionable assumption that the concepts linked to each of them are also related, respectively, to all named rivers, bays, and beaches in the real world. This issue is evident in the following explanation of forcing mechanisms acting on suspended sediment concentration (SSC) in rivers and bays.

According to Moskalski and Torres (2012), temporal variations in the SSC of bays and rivers are the result of a variety of forcing mechanisms. River discharge is a primary controlling factor, as well as tides, meteorological forcing (i.e., wind-wave resuspension, offshore winds, storm, and precipitation), and human activities. Various of these mechanisms tend to act simultaneously. However, the specific mix of active mechanisms is different in each bay and river. For example, SSC in the *San Francisco Bay* is controlled by spring-neap tidal variability, winds, freshwater runoff, and longitudinal salinity differences; whereas precipitation and river discharge are the mechanisms in the *Suisun Bay*. In the *Yangtze River*, SSC is controlled by tides and wind forcing; whereas river discharge, tides, circulation, and stratification are the active forcing mechanisms in the *York River*.

Consequently, in a specialized knowledge resource, a comprehensive list of forcing mechanism concepts semantically linked to the RIVER and BAY concepts, would not accurately represent the knowledge really transmitted in specialized texts because such a representation would inappropriately establish that all forcing mechanisms acting on SSC occur in all the rivers and bays in the world.

Indeed, as shall be seen in the section "Results," each named river in the coastal engineering domain educes a semantic network depicting a specific topic, associated with an environmental problem. Moreover, each of these river-evoked frames shows a different set of situational elements (i.e., concepts and semantic relations), a fact that proves the specific relational behavior of each named river. Therefore, given that each named river exhibits a specific relational behavior in specialized discourse, it is our assertion that TKBs should include the representation of named landforms and whatever named entity deemed to be relevant to a subject field.

Categorization and Inheritance

Upon representing named entities in a TKB, categorization and inheritance issues arise, which require basic principles to be laid down, similar to those applied to the representation of named rivers in EcoLexicon. These principles are based on research into the human categorization of spatial and non-spatial entities by Barsalou (1985), Davies (2009, 2020), and Rosch et al. (1976).

Rosch et al. (1976) provided evidence that categorization does not lead to clearly delimited categories of elements with shared properties based on necessary and sufficient conditions, but rather to categories with a graded structure and fuzzy boundaries, in which some members are deemed more prototypical than others.

---

[1]http://ecolexicon.ugr.es
[2]http://olst.ling.umontreal.ca/cgi-bin/dicoenviro/search_enviro.cgi
[3]https://www.eionet.europa.eu/gemet/en/themes/
[4]http://www.fao.org/faoterm/en/
[5]http://aims.fao.org/en/agrovoc
[6]http://www.environmentontology.org/Browse-EnvO

In turn, Barsalou (1985) established that, in addition to the resemblance to a prototype, there exist other three factors that determine the graded structure of categories: (1) goal-based ideals (i.e., features related to function or purpose of a concept in a context); (2) frequency of instantiation (i.e., how often a person has understood a concept to pertain to a category); and (3) personal familiarity with a concept. He found that the prototypicality of a concept depends on context and type of category, namely, taxonomic category (e.g., types of estuary, and means of transport), or goal-derived category (e.g., aesthetics of a place as well as its function). As such, prototypicality exerts more influence on taxonomic categories, whereas goal-based ideals and frequency of instantiation act on both types of category.

Goal-based ideals have been found to play a major role in determining the graded structure of even natural-kind categories despite being taxonomic, for instance, in the categorization of trees (Lynch et al., 2000) and birds (Burnett et al., 2005). Furthermore, in both research studies, knowledge domain experts employed goal-based ideals in contextualized categorization tasks (related to a utilitarian view and purposes rather than proximity to prototypes) much more frequently than novice students. At the highest level of expertise, personal familiarity had a greater impact than prototypicality. Similarly, in an fMRI study of expert-novice differences in the identification of geological field instruments, Faber et al. (2014) found that, unlike novices, experts activated, among other brain regions, those involved in the representation of context and the codification of meaningful contextual associations.

The environmental psychologist, Clare Davies, upon replicating experiments on human categorization by Barsalou (1985) for the case of geographic entities, found that named places are regarded as concepts (Davies, 2009) and may be treated cognitively as semantic categories of locations (Davies, 2020, p. 9–10), a fact that has been long considered intuitively reasonable in the discipline of Geography (Montello, 2003). In other words, in human mind, a place behaves as a semantic category, and each feature linked to the place is thus stored as a semantically related exemplar of it. Even physical items located within the place are cognitively processed as exemplars of the place, not just as contiguous points in space. In Davies's (2020) work, named places showed several of the same characteristics of categorization as the aforementioned studies of non-spatial objects. Namely, places are fuzzy categories, and influenced by context, expertise, goal-based ideals, prototypicality, and physical or spatial similarity.

Although Davies (2020) did not explore in her experiments aspects such as hierarchical structure and the presence of a basic level in place categories (Rosch, 1978; Murphy and Lassaline, 1997), she emphasized that work in Geographic Information Science has provided evidence of the presence of both characteristics (Hirtle and Jonides, 1985; Lloyd et al., 1996; Edwardes and Purves, 2007).

Regarding the categorization and inheritance issues that arise when including named rivers in EcoLexicon, the following principles are adopted, based on the previously discussed research.

For purposes of specialized knowledge representation, a named entity, such as a named river (e.g., the *Salinas River*, in California), should be considered to be a subordinate concept of the RIVER concept in virtue of its specific relational behavior with specialized concepts within a subject field. Accordingly, the named river inherits from the RIVER concept the properties that allow it to be identified as a member of the RIVER category within a cultural community. Therefore, other named rivers, such as the *Dee River* (in the United Kingdom), would be considered a cohyponym of the *Salinas River*.

Consequently, if each named river is related to a distinct set of specialized concepts within a subject field, this poses the challenging question as to what specialized concepts are then to be linked to the RIVER concept, as superordinate concept, within the same subject field in a knowledge resource. Although there is no simple answer since the problem can be approached from diverse points of view and disciplines, we offer three possible solutions.

Firstly, one could opt not to link any specialized concept to the RIVER concept, only to named rivers. As such, named rivers would only inherit from RIVER the properties that allow them to be identified as members of the RIVER category.

The second solution is based on Rosch et al. (1976), and Barsalou (1985). Namely, those specialized concepts (also referred to as features), which are related to the majority of named rivers in our sample, might also be linked to the RIVER concept. Hence, the named rivers, as subordinate concepts, would also inherit from RIVER the features that are common to the majority. The drawback is that some named rivers would inherit features that are not related to them. In this situation, the factors that determine the graded structure of the RIVER category become important. For this reason, named rivers could be regarded as more or less prototypical members of the RIVER category according to Rosch et al. (1976); or alternatively, their categorization would be based on the goal-based ideals of Barsalou (1985). In this way, these named rivers would be prevented from inheriting those features that are not related to them. Thus, the TKB should implement a mechanism to impose inheritance restrictions.

The third solution consists in linking to the RIVER concept all features related to the named rivers in our sample. However, the links would be numerically weighted according to the commonness of the features among the named rivers. In other words, the more named rivers associated with a feature, the greater the link weight for that feature. In this way, the RIVER concept would reflect all the potential features that could be activated depending on context. This is a solution in consonance with the major role of context in the selective activation of previously stored knowledge (Croft and Cruse, 2004, p. 75). However, this could also produce an excessive information load for users of a terminological resource, since *river* can rarely activate all those features at the same time in a specific context. For this reason, users could set a threshold for the weights of the links, so that only the features whose link weight is greater than the threshold would be shown in the semantic network of the RIVER concept. This could

be interpreted as a graded adhesion of the features to the RIVER concept.

In the three solutions proposed, the features linked to named rivers are those that corpus data ascribe to each named river. Obviously, if corpus data do not associate a feature with a named river, this does not mean that the feature is not actually related to it. However, as the terminological resource becomes populated with data from an increasingly large corpus, this issue will become less critical. In any case, whatever the implemented solution, the final representation should be validated by experts in the field in which the named entities are analyzed.

Cross-Cultural Conceptualizations of Landforms

Further explanations should also be provided with respect to cross-cultural differences in the conceptualization of landforms. The semantic content of words for parts of the physical world is determined by the cultures of the speakers (Sharifian, 2011). Consequently, landform terms, such as *river*, *mountain*, *bay*, and *wetland*, do not possess identical meanings in all languages (Smith and Mark, 2003). Culture and utility (i.e., affordance, understood as the resource that the environment offers people) also influence the categorization of landforms, not only intellectual interest (Smith and Mark, 2001). For instance, in the case of river, Bromhead (2018, Ch. 2) compares four languages and uncovers differences and common factors in the configuration of the RIVER concept.

This paper focuses on the usage of named rivers in a coastal engineering corpus in English language. Therefore, the analysis of cross-cultural differences in conceptualization of this landform, in the same subject field, in other languages is outside the scope of this study. Notwithstanding, in future work, with the purpose of making EcoLexicon an inclusive resource sensitive to cultural variation, it will thus integrate different cultural views on specialized concepts of the environment. The cultural parameters with which this project will begin are the following: geographical origin, variations from each environmental discipline, and degree of specialization. As such, the cultural adaptation of the conceptual module of EcoLexicon will allow to contextualize the semantic networks of named rivers according to the cultural parameter of geographical origin.

SEMANTIC ANALYSIS FROM THE PERSPECTIVE OF FRAME-BASED TERMINOLOGY

Frame-based Terminology (Faber, 2012, 2015), the approach applied in EcoLexicon and in this study, organizes knowledge in semantic frames, thereby creating non-language-specific representations. Such configurations are the conceptual meanings underlying specialized texts in different languages. This specification facilitates specialized knowledge acquisition because it relates entities and processes associated with a particular situation that is part of human experience (Barsalou, 2003).

According to Frame Semantics (Fillmore, 2006), in order to understand the meanings of words in a language, it is first necessary to have knowledge of the semantic frames, or conceptual structures, that underlie their usage.

Frames have the advantage of making explicit both the semantic and syntactic behavior of specialized language units. This necessarily includes a description of semantic relations as well as a term's combinatorial potential (Faber, 2009, p. 123). Frames conceptualize reality by means of a closed set of hierarchical relations, such as the hyponymic and meronymic relations *type_of*, and *part_of*; and non-hierarchical or associative relations, such as *causes*, *improves*, *results_of*, and *has_function*, which are domain-specific relations that make knowledge representation more meaningful and connected to reality since they show both multidimensionality and dynamism (Faber et al., 2009, p. 16; León-Araúz, 2009, p. 149, 176, and 184).

In summary, frames facilitate specialized knowledge acquisition and make knowledge representation more meaningful. These properties comprise what is called *explanatory adequacy* of a semantic network in Frame-based Terminology.

On the other hand, since a frame is activated by a linguistic item and the units in its cotext, its construction implies the semantic analysis of predicate-argument structures, which refer to the lexical representation of argument-taking lexical items (Levin, 2013/18). These are typically verbs and their nominalizations. The specification of the argument structure involves identifying the number of arguments that a lexical item can take, their syntactic expression, and their semantic relation, or semantic role, to the predicate.

Although syntactic expression is language-specific, semantic relations to the predicate are not. For that reason, what is important is not the syntactic realization of the predicate and its nominalization, but rather the combination of semantic roles and categories. In this way, the frame is generated by this combination of semantic roles and categories, and the relation between them (Faber and Cabezas-García, 2019, p. 202–204).

Consequently, this paper focused on the semantic analysis and annotation of sentences that mention potamonyms in coastal engineering texts. This permitted the subsequent construction of semantic frames that reflected the usage of named rivers in that domain for the purpose of representing them in EcoLexicon. These frames can function as interlingual representations, thereby facilitating their processing by computers (Boas, 2005; Segev and Gal, 2008; Baker, 2009; Pimentel, 2015), for instance, in machine-translation applications (Buendía-Castro and Faber, 2016), and computer-aided specialized translation (León-Araúz et al., 2020).

MATERIALS

Corpus Data

The sentences that cite named rivers were extracted from a subcorpus of English texts on coastal engineering, comprising roughly 7 million tokens. This subcorpus was composed of specialized texts (scientific articles, technical reports, and

Ph.D. dissertations), which amounted to 73.17% of the corpus size; and semi-specialized texts (textbooks and encyclopedias on coastal engineering), which constituted 26.83% of the corpus size. The total number of texts of the subcorpus was 2,249, whose publication data ranged from 1996 to 2018.

This subcorpus in costal engineering is part of the English EcoLexicon Corpus,[7] which currently contains over 100 million tokens in English and is focused on the environmental domain. It was manually compiled for the development of the EcoLexicon database. We refer the reader to León-Araúz et al. (2018) for a detailed description of its design and compilation criteria.

The coastal engineering domain was chosen for the semantic representation of named rivers because it is an interdisciplinary science that studies coastal processes, both natural and human-induced, for the design of maritime works and environmental recovery projects. Since one of the functions of coastal engineering is shore protection against erosion and flooding, coastal engineers design coastal defense structures such as breakwaters, dikes, and revetments. They also may envisage non-aggressive solutions such as dune restoration, artificial nourishment, and revegetation.

According to the Coastal Sediment Management Workgroup (2009, p. 1), although beach erosion is a natural process, human activities have reduced the natural supply of sediments to the coast and have thus modified alongshore sediment transport. For instance, dams block the transport of sediment through rivers, thus decreasing the downstream transport of sediments that reach bays; coastal structures, such as groins and breakwaters, alter the transport of sediment along the coast; harbors in bays trap sediment and also modify its transport patterns along the coast to naturally nourish beaches. Consequently, since the nature of coastal and environmental problems vary widely depending on the location, and the proper solution needs specific evaluation (Coastal Engineering Research Center, 1984, p. 1), the semantic analysis of individual named rivers (among other geographic entities such as named beaches and bays) is required for their accurate representation in TKBs on the environment.

GeoNames Geographic Database

Automatic detection of the named rivers mentioned in the corpus was performed with a dataset obtained from the GeoNames geographic database (Ahlers, 2013). GeoNames[8] has over 10 million proper names for 645 different geographic categories, such as bays, beaches, rivers, deltas, estuaries, river basins, river valleys, mountains, bridges, and populated places. For each entity, information is stored regarding its normalized designations, alternate designations (including other languages than English), latitude, longitude, and location name. A daily GeoNames database dump is publicly available for download in the form of a large worldwide text file, which was used for the recognition of named rivers in the corpus.

---

[7]The English EcoLexicon Corpus is publicly available in Open Corpora in Sketch Engine: https://www.sketchengine.eu/ecolexicon-corpus/.
[8]http://www.geonames.org

INCEpTION Annotation Tool

The INCEpTION tool (Klie et al., 2018) is a state-of-the-art annotation platform for semantic annotation (e.g., semantic frame annotation, knowledge base population, and entity linking, *inter alia*), which integrates machine learning capabilities, knowledge management, an intuitive user interface, and the ability to manage multiple annotation projects with several users involved.

To improve the manual annotation process, INCEpTION system makes use of predictive machine learning algorithms, which continuously monitor the labels attached by the user to provide annotation suggestions that the user can accept or reject. In this way, the feedback provided, and the changes made during the manual annotation process flow to the algorithm, which retrains the predictive model to update the annotation suggestions.

For knowledge management, INCEpTION allows users to create and edit an internal RDF-based knowledge base by annotating facts (i.e., triplets formed by a subject, a predicate, and an object) in the predicate-argument structure of sentences in corpus texts. In doing so, a domain-specific knowledge base can be constructed and expanded as part of the annotation task. The population of this internal knowledge base can then be used for fact-linking. In addition, external knowledge bases, such as Wikidata, DBPedia, ENVO, or EcoLexicon, can be accessed *via* SPARQL. These external resources enable users to perform knowledge-driven annotations such as entity linking, which signifies linking terms, mentioned in texts, to the corresponding concepts, stored in a knowledge base, which are designated by those terms.

The INCEpTION annotation scheme organizes annotations into layers, which represent the features to be annotated in a project (e.g., semantic roles, semantic categories, or named entity types) and their labels (e.g., the labels AGENT, PATIENT, and THEME for the semantic role feature). Any number of layers can be defined, which can be spans or relations between spans. Each layer can also have any number of features, which can be strings, numbers, Booleans, concept references, or references to other annotations.

Figure 1 shows the annotation user interface of the INCEpTION tool and the semantic annotation of the predicate-argument structure of a sentence, mentioning the *Salinas River*, in the coastal engineering corpus. The annotation scheme designed for our semantic annotation task is explained in the following section.

METHODOLOGY

Recognition of Named Rivers

In the experiments by Stokes et al. (2008) which compared different natural language processing methods to detect toponyms in texts, it was found that a simple approach employing a gazetteer (i.e., a dictionary of proper names for geographic entities) to recognize the presence of toponyms outperformed other sophisticated methods. Hence, we applied the matching of named entities, using the GeoNames geographic database to identify the named rivers, deltas, estuaries, river basins, river valleys, and river mouths mentioned in the corpus.

The corpus texts were tokenized, tagged with parts of speech, lemmatized, and lowercased with the Stanford *CoreNLP* package
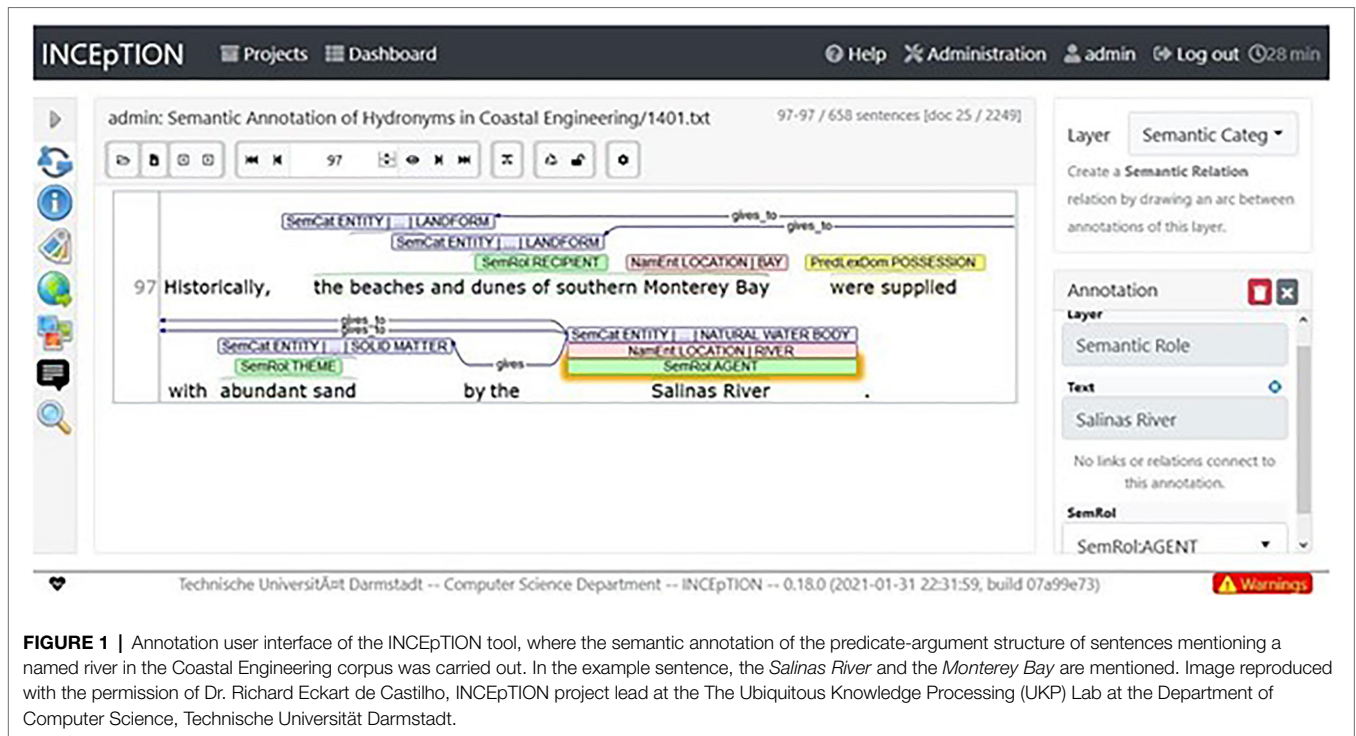
FIGURE 1 | Annotation user interface of the INCEpTION tool, where the semantic annotation of the predicate-argument structure of sentences mentioning a named river in the Coastal Engineering corpus was carried out. In the example sentence, the *Salinas River* and the *Monterey Bay* are mentioned. Image reproduced with the permission of Dr. Richard Eckart de Castilho, INCEpTION project lead at the The Ubiquitous Knowledge Processing (UKP) Lab at the Department of Computer Science, Technische Universität Darmstadt.

(Manning et al., 2014) for the R programing language (R Core Team, 2021). Then, both normalized and alternate names of rivers, deltas, estuaries, basins,[9] valleys, and mouths in the GeoNames database dump were searched in the lemmatized corpus. A total of 783 different designations were recognized and listed.

Most designations cited in the corpus were in GeoNames (97%), while others were identified by manual inspection (3%). Namely, with a view to estimating the capability of GeoNames for recognizing named landforms in the corpus, we first queried the corpus documents that contained the terms *river*, *delta*, *estuary*, *basin* (and the synonyms *catchment* and *watershed*), *valley*, and *mouth*. Then, we listed all the potamonyms that were manually identified in those corpus documents. In this way, we could ascertain that GeoNames matched 97% of the potamonyms in our list. This high performance of GeoNames allows us to trust that it will also be able to match a substantial percentage of other named landforms, relevant to coastal engineering, that are mentioned in the corpus, such as named bays, beaches, and coasts, which will be analyzed in future work.

Anaphoric elements that referred to a river, delta, estuary, basin, valley, or mouth were replaced by the corresponding full designations in the lemmatized corpus. For this task, the automatic anaphora resolution function of the *CoreNLP* package was used.

Since various designations can refer to the same river because of syntactic variation (e.g., *Nile River* and *River Nile*), and orthographic variation (e.g., *Yangtze* and *Yangtse River*), the variants were identified to give them a single designation in

the corpus. Once the variants were normalized in the lemmatized corpus and joined with underscores, the number of named rivers, deltas, estuaries, basins, valleys, and mouths was 676.

The mouths of the 360 rivers mentioned in the corpus were shown on a map with color-coded rectangles that represented their frequency in the corpus. Their latitudes and longitudes were retrieved from the GeoNames database dump. This reflected the representativeness of the corpus in reference to river locations and their number of mentions. The named rivers were in a large number of countries, but the most cited rivers were located in the United States.

A critical issue was to disambiguate to which river with the same name the text referred to. Namely, although latitudes and longitudes could be retrieved from the GeoNames database dump, the same designation occasionally referred to rivers in different countries. For instance, the corpus only located the *Yellow River* in China. However, GeoNames indicated that rivers with the same name also existed in the United States, Canada, Ireland, and Papua New Guinea. Such cases had to be resolved by corpus queries.

The occurrence frequency of the named rivers, deltas, estuaries, basins, valleys, and mouths ranged from 129 (*Scheldt River Estuary*) to only one mention (349 out of 676 designations). **Figure 2** shows a sample with the 35 most frequently cited designations, along with their number of mentions.

Semantic Annotation of Predicate-Argument Structures for Named Rivers

The 676 designations encompassed a total of 2,840 mentions of named rivers, deltas, estuaries, basins, valleys, and mouths

---

[9]In the case of river basins, the term variants *basin* (British English), *catchment* (Australian English), and *watershed* (American English) were also manually searched in the corpus to identify proper names of this landform. This terminological variation for river basins is pointed out by Renschler (2004).
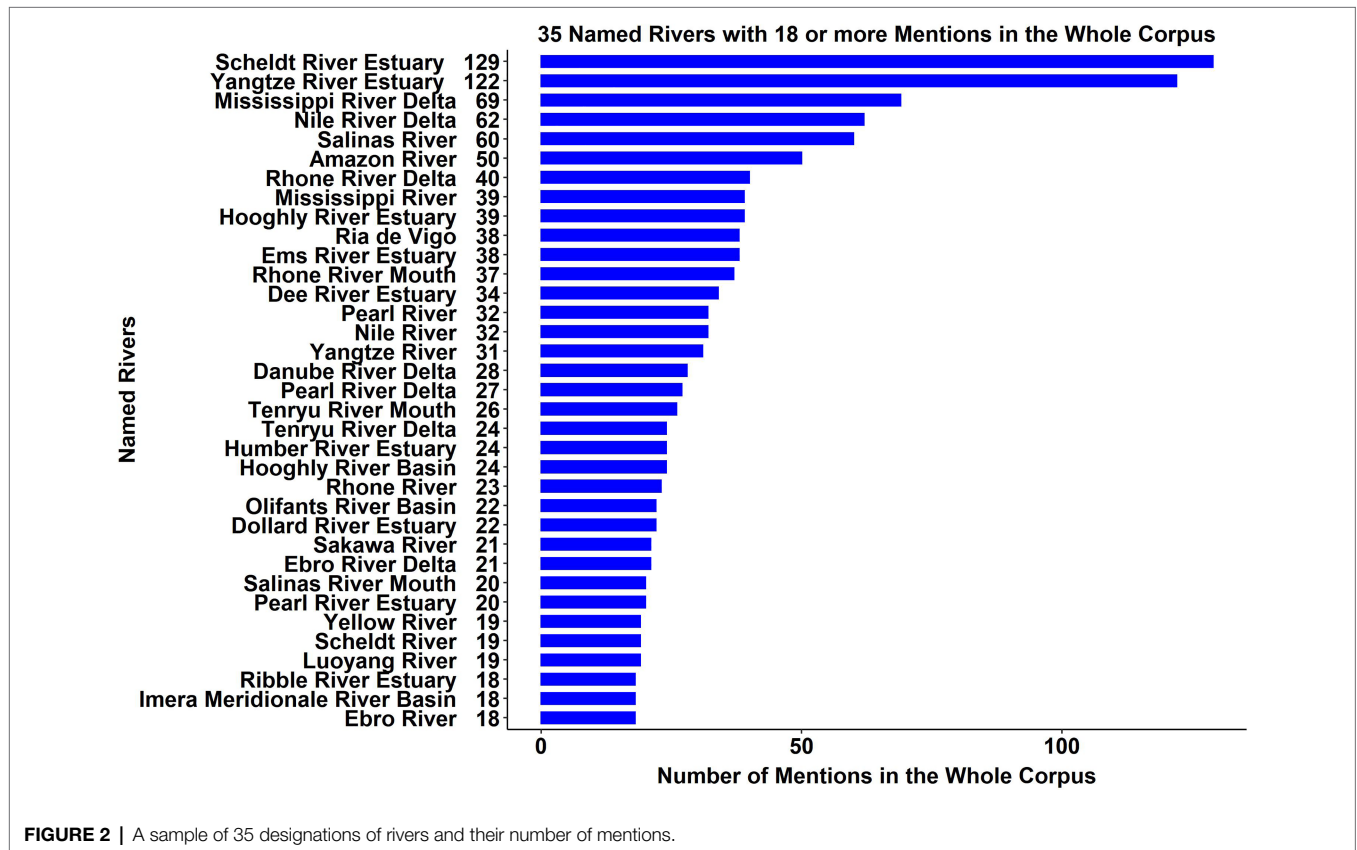
FIGURE 2 | A sample of 35 designations of rivers and their number of mentions.

in the corpus. We decided to analyze more than 50% of the mentions, which meant including in the sample the designations that appeared 8 or more times. This led to the semantic annotation of 1,694 mentions, which embraced rivers, deltas, estuaries, basins, valleys, and mouths. Accordingly, the sample of 1,694 annotated mentions represented 59.65% of the total number of mentions in the corpus.

For simplicity, named rivers, deltas, estuaries, basins, valleys, and mouths are henceforth referred to as named rivers or potamonyms. As such, the verbal and nominal predicates that occurred with named rivers in our corpus were first classified into the lexical domains defined by Faber and Mairal (1999). Then, the argument structure of the predicates was analyzed and annotated.

The set of 1,694 sentences from the corpus was annotated by three terminologists from the LexiCon research group at the University of Granada (Spain). They performed the semantic annotation of the predicate-argument structure of a sentence by assigning: (1) a lexical domain to the predicate; (2) a semantic role to the arguments of the predicate; (3) a semantic category to the arguments of the predicate; and (4) a semantic relation to the link between the potamonym and the other arguments in the sentence. In the following, these annotation categories are described, and examples of annotated sentences are provided later, in the section "Results."

Predicate Classification in Lexical Domains

According to Faber and Cabezas-García (2019, p. 205), the most frequent verbs in our corpus, as part of the English EcoLexicon Corpus, are general language verbs (e.g., *accumulate*, *pollute*, *increase*, *discharge*, *supply*, and *drain*), which are also used in specialized texts.[10] However, when they have terms for their arguments, this makes them domain-specific (L'Homme, 2003; Buendía-Castro, 2013). Even though verbs (especially general language verbs) have hardly been regarded as important in Terminology, they reflect how environmental entities interact (Buendía-Castro, 2013). In this sense, such verbs are also susceptible to classification into the lexical domains proposed by Faber and Mairal (1999), within the Functional-Lexematic Model (Martin, 1984/2017).

These authors propose, after analyzing over 10,000 verbs in the English language, a model for their lexical classification into domains based on the distinction between paradigmatic and syntagmatic relations. A lexical domain is thus formed by a hierarchy of verbs, all of which share the same nuclear meaning and syntax. The most prototypical verbs, or superordinate verbs, are those that have the largest combinatory potential from a semantic point of view.

---

[10]We focused on the fact that verbs, whether specialized or general language, are crucial to meaning because they are generally what relate concepts in specialized texts. This study thus did not address the phenomena of terminologization or determinologization of verbs, although we recognize that both aspects need to be further investigated.

In Frame-based Terminology, such verbs and their nominalizations provide the frames that characterize the actions and processes in the specialized field and link the semantic categories of the typical participants. In this regard, the majority of verbs in the English EcoLexicon Corpus were found to belong to the lexical domains shown in **Table 1** (Faber and Cabezas-García, 2019, p. 206), used to annotate the predicates of our set of sentences.

Semantic Roles

Specialized knowledge representation includes semantic properties that help to describe the nature of entities and processes. These semantic properties are reflected as the relations between a predicate and its arguments, which are typical semantic roles.

Although most linguists tend to believe that semantic roles exist, there is considerable disagreement as to their number,

nature, and function (Ureña et al., 2013, p. 180). The set of semantic roles in this study largely coincided with those specified by Kroeger (2005, p. 54–55), and Thompson et al. (2009). **Table 2** shows the list of the semantic roles used to annotate the arguments in our set of sentences.

Semantic Categories

The concepts in EcoLexicon were classified into 153 semantic categories, hierarchically organized, and distributed in five categorization levels (Gil-Berrozpe et al., 2019). The most general level is composed of the three basic ontological categories, namely PROCESS (i.e., events extending over time and involving different participants), ENTITY (i.e., physical and mental objects), and ATTRIBUTE (i.e., properties of entities and processes).

However, depending on the ontological nature of concepts, they can be subclassified in up to five levels of specificity. For

TABLE 1 | The most frequent lexical domains of environmental verbs (Faber and Cabezas-García, 2019, p. 206).

Lexical domain	Prototypical verb	Verb examples from the corpus
CHANGE	[to become/change]	*change, become, decrease, increase improve, transform, worsen, consolidate, influence, vary*
MOVEMENT	[to move]	*carry, transport, transfer, introduce, go into, discharge, enter, drain, overflow*
EXISTENCE	[to be/exist]	*prevent, produce, originate, occur, arise, develop, form, grow, initiate, result in*
POSSESSION	[to have]	*absorb, drain, distribute, lack, obtain, offer, provide, receive, supply, discharge*
POSITION	[to be in a state/place/position]	*arrange, bound, confine, insert, place, remain, stay, deposit, accumulate, locate*
MANIPULATION	[to use]	*manipulate, influence, control, treat, deal, manage, monitor, recycle, use, utilize*
ACTION	[to do/make]	*perform, dam, respond, behave, make, operate, produce, work, construct, be under construction*
COGNITION	[to know/think]	*analyze, ascertain, assess, categorize, classify, compare, determine, estimate, underestimate, overestimate*
IMPACT	[to hit/break]	*thresh, strike, slam, shatter, rupture.*

TABLE 2 | List of semantic roles and their definitions, based on Kroeger (2005, p. 54–55) and Thompson et al. (2009).

Semantic roles	Definition
AGENT	Entity/process that causes an action, whether intentionally or unintentionally.
RESULT	Entity/process that has come about as a consequence of a voluntary/involuntary action.
PATIENT	Entity which is acted upon, affected, or created; or of which a state, or change of state, is predicated.
THEME	Entity which undergoes either a change of location or a change of possession. Also, entity whose location is being specified.
LOCATION	Spatial reference point of a process or an entity (the SOURCE, GOAL, and PATH roles are often considered to be subtypes of LOCATION).
RECIPIENT	Entity which receives or acquires something.
INSTRUMENT	Entity used by an AGENT to perform some action. INSTRUMENT refers to the tools, machinery, and devices that are used to carry out human process events in the environment. However, natural entities can also use natural instruments.
TIME	Phrase that situates an event in time or with respect to another event.
RATE	Phrase that describes changes in rate or level that occur as part of an event. In most cases, this role applies to the THEME of an event.
MANNER	Phrase that describes the method or way in which a particular event is carried out.
DESCRIPTION	Phrase that describes characteristics or behavior of the AGENT or THEME of the event.
CONDITION	Phrase describing the environmental conditions which must hold in order for the event to take place.
PURPOSE	Process that specifies why another process occurs, i.e., specifications of some sort of aim, purpose, goal, or reason for the process occurring.

instance, the semantic category of the BEACH SIZE SAND concept is MINERAL, placed on the fifth level of the category hierarchy ENTITY > MATTER > SOLID MATTER > MATERIAL > MINERAL.

From an ontological point of view, 15 categories were associated with attributes, 94 with entities, and 44 with processes. Accordingly, this hierarchically organized list of 153 semantic categories was used to annotate the arguments in our set of sentences. For a full list of the semantic category hierarchy and some examples of each category, see Gil-Berrozpe et al. (2019).

Semantic Relations

Conceptual description in EcoLexicon is based on the semantic category of concepts and their relational behavior. A fixed set of semantic relations, both hierarchical and non-hierarchical, was systematically defined by Faber et al. (2009) to make EcoLexicon a consistent resource at its different representational levels. These relations, with additional non-hierarchical relations specific to named rivers, are shown in **Table 3**, along with examples in the form of conceptual propositions (i.e., triplets

TABLE 3 | Semantic relations used during the annotation process.

Relation category	Relation	Example
Generic-specific	type_of	WIND EROSION type_of EROSION
Part-whole	part_of	RIGHT BANK part_of RIVER
	made_of	GROIN made_of WOOD
	delimited_by	MESOSPHERE delimited_by STRATOSPHERE
	located_at	SOFT MUD located_at MISSISSIPPI RIVER MOUTH
	takes_place_in	CONSOLIDATION OF THE LAND takes_place_in MISSISSIPPI RIVER MOUTH
	phase_of	PUMPING phase_of DREDGING
Non-hierarchical relations in EcoLexicon	affects	EROSION affects COASTAL DUNE
	causes	SACRAMENTO RIVER causes FRESHWATER INFLOW
	result_of	TENRYU RIVER DELTA result_of SEDIMENT SUPPLY
	attribute_of	DISCHARGE RATE attribute_of WESER RIVER
	has_function	SALINAS RIVER has_function SAND SUPPLY
	studies	POTAMOLOGY studies SURFACE CURRENT
	measures	OPTICAL BACKSCATTER SENSOR measures SUSPENDED SEDIMENT CONCENTRATION
	effected_by	DREDGING effected_by DREDGER
Additional non-hierarchical relations for the semantic frames evoked by Named Rivers	improves	SALINAS RIVER ESTUARY improves WATER QUALITY
	worsens	SEA LEVEL RISE worsens SALINAS RIVER ESTUARY
	creates	MISSISSIPPI RIVER creates NATURAL LEVEE
	becomes	SAINT BERNARD RIVER DELTA becomes CHANDELEURS ISLANDS
	gives	SALINAS RIVER gives BEACH-SIZE SAND
	gives_to	YANGTZE RIVER gives_to YELLOW SEA
	receives	DEE RIVER receives SEDIMENT
	receives_from	PO RIVER receives_from PO PLAIN
	drains	SANTA CLARA RIVER drains WATERSHED
has_path	moves_over	SALINAS RIVER has_path/moves_over RIVERBANK
	moves_into	BED SEDIMENT LOAD has_path/moves_into YANGTZE RIVER ESTUARY
	moves_across	MURRAY RIVER has_path/moves_across TERTIARY FORMATION
	transfers	WESER RIVER transfers SEDIMENT LOAD
	discharges_into	SALINAS RIVER discharges_into MONTEREY BAY
	places	MISSISSIPPI RIVER places SOFT MUD
	controls	VENTURA RIVER controls SEDIMENT SUPPLY
	applied_to	NOAH MODEL applied_to MISSISSIPPI RIVER BASIN

consisting of two concepts and the semantic relation that links them). They were all used to annotate the semantic relation between the arguments in our set of sentences, specifically, the link of a named river to another argument in the same sentence. The non-hierarchical relations specific to named rivers are explained in the section "Results," since they emerged from the semantic analysis of the predicate-argument structures.

Annotation Scheme in the INCEpTION Tool

The INCEpTION annotation scheme organizes annotations into layers, which represent the features to be annotated in a project and their labels. We set five layers for 10 annotation features: (1) the first layer for the *predicate lexical domain* feature; (2) the second layer for the *semantic role*; (3) the third layer for the *semantic relation*; (4) the fourth layer to describe the named entity, which included two features, namely the *named entity* feature (initially, with the three basic labels LOCATION, ORGANIZATION, and PERSON), and the *hydronym* feature to annotate whether the location was a bay, beach, coast, river, delta, estuary, river basin, river valley, or river mouth; and (5) the fifth layer for five features that stored, respectively, the five levels of the semantic category hierarchy.

Inter-Annotator Agreement

As previously stated, the annotation of the predicate-argument structures in the coastal engineering corpus was carried out by three terminologists. As for the inter-annotation agreement (Brezina, 2018, p. 87–92), *Cohen's kappa* coefficient (κ; Cohen, 1960) showed a very good agreement for all annotator pairs ($90\% < \kappa < 98\%$, values of $p < 0.05$) in the initial annotations of semantic roles, categories, and relations, according to Krippendorff's (2012) recommendations for text content analysis. A review of the differences between annotators showed no systematic pattern of disagreement. Given the nature of the judgment variables, the level of agreement was deemed acceptable. Notwithstanding, the disagreements in the original annotations were resolved based on discussion between the annotators to reach a consensus on the definitive annotations of semantic roles, categories, and relations.

For the initial annotation of predicates with lexical domains, the inter-annotation agreement was lower for all the annotator pairs ($84\% < \kappa < 88\%$, values of $p < 0.05$), indicating that this variable lent itself to alternative, though plausible, interpretations. A review of the differences between annotators showed that the lexical domains of MOVEMENT and POSSESSION were more prone to confusion. The issues fundamentally arose from verbs that could potentially belong to more than one lexical domain, as Faber and Mairal (1999) already proved. To arrive at a consensus on the definitive annotations of lexical domains, the factorization of meaning from the Functional-Lexematic Model framework was applied to verbs, such as *drain* and *discharge*. The meaning factorization of *drain* is described in the following, as an example of the process used when there was disagreement between the annotators.

Although *drain* is a general language verb, it becomes a specialized verb in domain-specific texts when its arguments

are filled with specialized knowledge units. As shall be seen, in some cases, the semantic content of its domain-specific arguments interacts with its base meaning to create a new sense that is appropriate for certain coastal engineering contexts (L'Homme, 2003; Faber and León-Araúz, 2016).

In the *Merriam-Webster Dictionary*, the intransitive use of *drain* can have four senses. However, only sense 2, *to discharge surface or surplus water*, was used in our sample of specialized contexts. Accordingly, *drain* belongs to the lexical domain of MOVEMENT, as shown in sentence (1) of the sample:

1. The [Salinas River]_{SemRol: THEME} *drains* into the [Monterey Bay]_{SemRol: LOCATION|SemCat: LANDFORM}

Similarly, the transitive use of *drain* can have nine senses. Of these senses, only 2b, *to carry away the surface water of*, was used in our sample of specialized contexts. This means that transitive *drain* also belongs to the lexical domain of MOVEMENT, as shown in sentence (2) of the sample:

2. Natural sediment supply within this region is defined by the [Ventura River]_{SemRol: AGENT} that *drains* large [watersheds]_{SemRol: PATIENT|SemCat: LANDFORM}

Therefore, sentences (1) and (2) maintain the base meaning of *drain* in general language. More specifically, they foreground the movement of water from one place to another, and thus convey the semantic role patterns THEME + LOCATION, and AGENT + PATIENT, respectively, where the semantic category of the LOCATION and PATIENT arguments is LANDFORM.

Nonetheless, sentences (3) and (4) show how the semantic content of the domain-specific arguments of *drain* interacts with its base meaning to create a new sense, which is not used in the general language. In other words, sentences (3) and (4) do not highlight the movement of water, but rather foreground the change of possession of sediments or water from one entity (e.g., Po plain) to another (i.e., named rivers). Consequently, in both sentences, the verb *drain* belongs to the lexical domain of POSSESSION because its argument structure reflects the semantic role pattern AGENT + THEME + RECIPIENT,[11] where the semantic category of the THEME argument is MATTER, and that of the RECIPIENT argument is LANDFORM, specifically, RIVER.

3. Normally, eutrophic conditions are caused by [waters]_{SemRol: THEME| SemCat: MATTER} *drained* by the [Po River]_{SemRol: RECIPIENT| SemCat: LANDFORM > RIVER} from the highly inhabited and cultivated [Po plain]_{SemRol: AGENT|SemCat: LANDFORM}

4. Not all the [sediments]_{SemRol: THEME| SemCat: MATTER} *drained* by the [Dee River]_{SemRol: RECIPIENT|SemCat: LANDFORM > RIVER} participate to coastal sediment transport

---

[11]The designation of the semantic roles for predicates could be different depending on the level of generality adopted (Van Valin, 2004). Another alternative would be POSSESSOR + THEME + BENEFICIARY.

In conclusion, the meaning of the verb *drain* in the coastal engineering domain integrates at least two aspects, namely, the movement of matter and the change of possession of matter. Since both aspects are closely interrelated, *drain* is difficult to categorize since it can belong to the lexical domain of either MOVEMENT or POSSESSION. However, it is true that specialized contexts foreground one of the aspects and relegate the other in the background. This type of fine-grained distinctions evidently required more careful analysis.

RESULTS

On the one hand, the percentages of annotated sentences classified into the predicate lexical domains were the following: The lexical domains of MOVEMENT (24.67%), POSSESSION (20.78%), and CHANGE (16.89%) covered 62.34% of sentences. Next in the ranking were the lexical domains of EXISTENCE (14.28%), ACTION (9.10%), POSITION (5.19%), MANIPULATION (5.19%), and COGNITION (3.90%). The lexical domain of IMPACT did not appear in our sample of sentences.

Coastal engineering is a process-oriented domain because it empirically describes and studies dynamic physical states (Faber et al., 2006). Therefore, dynamism is a phenomenon that pervades this domain, in which the interaction between different concepts is characterized by movement and change (León-Araúz, 2009, p. 24). For that reason, in our sample of annotated sentences, MOVEMENT predicates (24.67%) and CHANGE predicates (16.89%) are among the most frequent. These predicates describe the natural and artificial processes of agents in named rivers, bays, and beaches, and the consequences of phenomena such as erosion or sedimentation. Furthermore, named rivers participate in a change of possession when they receive matters from plains and valleys, and then provide them to other entities, such as beaches and dunes, as they discharge into bays, seas, and oceans. Consequently, POSSESSION predicates (20.78%) are also predominant.

On the other hand, named rivers were also found to have a variety of semantic roles, namely AGENT, LOCATION, THEME, PATIENT, and RECIPIENT. Among the 1,716 arguments, filled with a named river and annotated with a semantic role, the most frequent one was AGENT (52.55%), and not LOCATION (23.08%) as expected. The THEME and PATIENT roles both occupied the third position in the ranking (10.26%). RECIPIENT was the least frequent role for named rivers (3.85%), since there was less focus on the entity from which the rivers received matters (e.g., sediments, pollutants, or water). Instead, judging by the high percentage of POSSESSION predicates (20.78%), coastal engineering texts focused on the recipient entities that were provided with matter by the river (AGENT), primarily because they were directly affected by environmental problems such as erosion, pollution, or flooding, *inter alia*.

In the following, results are presented with regard to the semantic configuration of the eight lexical domains, and the semantic networks that arose from the semantic analysis.

Lexical Domain of MOVEMENT

Four combinations of semantic roles were found for the lexical domain of MOVEMENT (24.67% of sentences), shown in **Table 4**:

1. AGENT + PATIENT: This pattern could express two relations, namely *drains* (i.e., a river flows along a place, while taking matter from it, and transports such matter to another place), or *moves_over* (i.e., a river flows over its banks, and thus inundates a town, a building, land, or crops). However, the semantic relation could always be identified, because of the semantic category of the concept that took the PATIENT role. Accordingly, the pattern encoded the *drains* relation if the concept with the PATIENT role was a LANDFORM (e.g., *watershed*), whereas the pattern conveyed the *moves_over* relation if the concept was PART OF WATER BODY (e.g., *riverbank*), SPATIAL AREA (e.g., *town*, *land*, *crop*), or BUILDING (e.g., *temple*). Only the verb *drain* was found to transmit the *drains* relation, whereas several verbs could transmit the *moves_over* relation (e.g., *overflow*, *flow over*, *flood*, *inundate*, *drown*, and *submerge*).
2. AGENT + THEME: This pattern always expressed the *transfers* relation, linking a named river to the matter it transports.
3. THEME + LOCATION: This pattern conveyed either of two relations, namely *moves_into* (i.e., matter goes into a river, or more generally, into a landform), or *discharges_into* (i.e., a river meets the place of its mouth). Nonetheless, the semantic relation could always be differentiated, thanks to the semantic category of the concept with the THEME role. As such, the pattern encoded the *moves_into* relation if the THEME concept was MATTER (e.g., *bed sediment load*), whereas the pattern transmitted the *discharges_into* relation if the concept was a named river. Various verbs could express the *discharges_into* relation, namely *flow into*, *drain into*, *discharge into*, *debouche into*, *enter*, *reach*, and *meet*. The verb *drain*, followed by the preposition *into*, always referred to the place where a river mouth was located.
4. THEME + PATH: This pattern always conveyed the *moves_across* relation, linking a named river to the landform across which the river flows.

Depending on the ultimate application of the semantic annotations, the *moves_over*, *moves_into*, and *moves_across* relations could reasonably be collapsed into a single relation, namely the *has_path* relation.

Lexical Domain of POSSESSION

Table 5 summarizes the findings for the lexical domain of POSSESSION (20.78% of sentences). Two combinations of semantic roles were found:

1. AGENT + THEME + RECIPIENT: This pattern expressed either of two relations, namely *gives/gives_to* (i.e., a named river supplies matter to a landform), or *receives/receives_from* (i.e., a named river takes matter from a landform while flowing along it). The specific semantic relation could always be specified, thanks to the semantic category of the concepts that took the AGENT and RECIPIENT roles. As such, if the

TABLE 4 | Results from the semantic annotations for the lexical domain of MOVEMENT.

Verb lexical domain: MOVEMENT

Arg1	Arg2	Arg3	Example	Term	Relation	Term
AGENT Named River	**PATIENT** ENTITY> […]>LANDFORM		Natural sediment supply within this region is defined by the [**Ventura River**]AGENT, that _drains_ large [**watersheds**]PATIENT.	Ventura River	_drains_	watershed
Named River	ENTITY > PART > PART OF WATER BODY		The [**Salinas River**]AGENT _overflows_ its [**banks**]PATIENT and deposits sediments in the flood plain.	Salinas River	_moves_over_	bank
	ENTITY > CREATION > STRUCTURE > BUILDING		The [**temples**]PATIENT, which date back to the thirteenth century b. c. e., were in danger of _being flooded_ by the [**Nile River**]AGENT during the construction of the Aswan High Dam.	Nile River		temple
	ENTITY > SPACE > AREA >ADMINISTRATIVE AREA		Rising 26 feet (10 m) above flood stage in some places, the [**Connecticut River**]AGENT _submerged_ at least four riverside [**towns**]PATIENT, blighting their corn and grain fields, and caused deep financial hardship for their weary inhabitants.	Connecticut River		town
	ENTITY > SPACE > AREA >LAND		When the [**Nile River**]AGENT _inundated_ the [**land**]PATIENT, the seepage naturally raised the water table.	Nile River		land
	ENTITY > SPACE > AREA >LAND		A crisscross network of earthen walls was formed in [**crops**]PATIENT that _would be flooded_ by the [**Nile River**]AGENT.	Nile River		crop
AGENT Named River	**THEME** ENTITY > MATTER > SOLID MATTER		The low concentration values are indicative for the [**sediment load**]THEME _carried_/_transported_ by the [**Weser River**]AGENT before it enters the estuarine zone.	Weser River	_transfers_	sediment load
THEME ENTITY > MATTER > SOLID MATTER	**LOCATION** ENTITY> […]>LANDFORM > NATURAL WATER BODY		Thus, several tens of millions of [**bed load**]THEME _goes into_ the [**North Passage of Yangtze estuary**]LOCATION along with the ebb currents.	bed load	_moves_into_	Yangtze River Estuary
Named River	ENTITY> […]>LANDFORM > NATURAL WATER BODY		The [**Salinas River**]THEME _flows into_/_drains into_/_discharges into_/_debouches into_/_enters_/_reaches_/_meets_ the [**Monterey Bay**]LOCATION.	Salinas River	_discharges_into_	Monterey Bay
THEME Named River	**PATH** ENTITY> […]>LANDFORM		The [**River Murray**]THEME _flows across_ [**Tertiary formations**]PATH to enter coastal lagoons behind the dune calcarenite barriers of Encounter Bay.	Murray River	_moves_across_	Tertiary formation

AGENT was mentioned in the sentence and was a named river, the predicate conveyed the _gives/gives_to_ relation. Nevertheless, if the RECIPIENT was mentioned in the sentence and was a named river, the semantic role pattern encoded the _receives/receives_from_ relation.

Whereas various verbs could transmit the _gives/gives_to_ relation (e.g., _provide, supply, contribute, deliver, discharge, bring,_ and _introduce_), only the transitive use of the verb _drain_ expressed the _receives/receives_from_ relation. In addition, the phrase that

took the RECIPIENT role had to be frequently inferred from the whole sentence since this argument did not belong to the target predicate (see the example in the second row of **Table 5**, where the inferred RECIPIENT, _Yellow Sea_, is not an argument of the predicate _discharge_).

2. RECIPIENT + THEME + LOCATION: This pattern always conveyed the _takes_place_in_ relation, linking a process to its spatial and temporal dimensions. It must be clarified that the argument with the LOCATION role did not pertain

TABLE 5 | Results from the semantic annotations for the lexical domain of POSSESSION.

Verb lexical domain: POSSESSION

Arg1	Arg2	Arg3	Example	Term	Relation	Term
AGENT Named River	**THEME** ENTITY > MATTER > **SOLID MATTER**	**RECIPIENT** ENTITY > PART OF LANDFORM	The [**Salinas River**]ₐGENT no longer _contributes_ substantial [**beach size sand**]THEME to the [**Littoral Cell**]RECIPIENT.	Salinas River Salinas River	_gives_ _gives_to_	beach size sand littoral cell
		ENTITY > GEOGRAPHIC FEATURE > NATURAL GEOGRAPHIC FEATURE > **LANDFORM** > NATURAL WATER BODY	The [**Yellow Sea**]RECIPIENT is influenced strongly by [**Yangtze River**]ₐGENT, which _discharges_ more than 1.6 billion tonnes of [**sediments**]THEME annually.	Yangtze River Yangtze River	_gives_ _gives_to_	sediment Yellow Sea
	ENTITY > MATTER > **FLUID MATTER** > WATER		The [**Changjiang River**]ₐGENT _provides/supplies/brings/introduces/delivers_ most of the [**fresh water**]THEME to the [**Changjiang Estuary**]RECIPIENT.	Changjiang River Changjiang River	_gives_ _gives_to_	fresh water Changjiang Estuary
ENTITY> […]> LANDFORM Ø	ENTITY > MATTER > **FLUID MATTER** > WATER ENTITY > MATTER > **SOLID MATTER**	Named River	Normally, eutrophic conditions are caused by [**waters**]THEME _drained_ by the [**Po River**]RECIPIENT from the highly inhabited and cultivated [**Po plain**]ₐGENT. Not all the [**sediments**]THEME _drained_ by the [**Dee River**]RECIPIENT participate to coastal sediment transport.	Po River Po River Dee River	_receives_ _receives_from_ _receives_	water Po Plain sediment
RECIPIENT ENTITY	**THEME** PROCESS > CHANGE > CHANGE IN INTENSITY > DECREASE	**LOCATION** Named River	However, in this instance, the [**anthropogenic effects**]RECIPIENT probably dominate and _include_ additional subsidence resulting from withdrawal of hydrocarbons, and the [**sediment supply reduction**]THEME in the [**Mississippi River**]LOCATION by the construction of upstream impoundments and jetties that direct the riverine sediment offshore to deepwater.	sediment supply decrease	_takes_place_in_	Mississippi River

to the same target predicate as the arguments with the RECIPIENT and THEME roles. In other words, the argument with the THEME role was filled with a deverbal noun designating a process, which acted as the predicate to which the argument with the LOCATION role belonged. A case in point is the sentence in the last row of **Table 5**, namely:

[Anthropogenic effects]RECIPIENT **include** [sediment supply reduction]THEME in the [Mississippi River]LOCATION

The predicate _include_ has _anthropogenic effects_ and _sediment supply reduction_ as its arguments. In turn, the nested predicate _reduction_ has _Mississippi River_ as an argument. The predicate _reduction_ means _to change by decreasing_, therefore, it could have been classified as belonging to the lexical domain of CHANGE. However, since _reduction_ is embedded in an argument structure that is steered by the predicate _include_, the decision was made to ascribe this type of sentence to the lexical domain of the higher-level predicate.

Lexical Domain of CHANGE

Table 6 summarizes the findings for the lexical domain of CHANGE (16.89% of sentences). Three combinations of semantic roles were found:

1. AGENT + PATIENT: This pattern expressed one of three relations, namely, _improves_ (i.e., an entity or a process changes an attribute or any other entity for the better), _worsens_ (i.e., a process changes a river for the worse), or _affects_ (i.e., an entity or a process causes a change in any other entity or process without producing a final result). However, the semantic relation was always evident, because of the semantic category of the concepts with the AGENT and PATIENT roles. Accordingly, the pattern encoded the _affects_ relation if the concept with the PATIENT role was a LANDFORM (e.g., _Quanzhou Bay_), whereas the pattern conveyed the _worsens_ relation if the concept with the AGENT role was a PROCESS INCREASING IN SIZE (e.g., _sea level rise_). In any other case,

TABLE 6 | Results from the semantic annotations for the lexical domain of CHANGE.

Verb lexical domain: CHANGE

Arg1	Arg2	Arg3	Example	Term	Relation	Term
AGENT ENTITY > HUMAN > INSTITUTION	**PATIENT** Named River		The [**United States Army Corps of Engineers**]AGENT adopted a modified version of his jetty plan for *improving* the [**St. Johns River entrance**]PATIENT.	US ACE	*improves*	Saint Johns River Mouth
Named River	ATTRIBUTE		[**Salinas River Estuary**]AGENT *helps improve* the [**quality of water**]PATIENT since it acts both as a filter […] and as a buffer between ocean storms and cities, protecting the inland areas from damage.	Salinas River Estuary	*improves*	water quality
PROCESS > CHANGE > TRANSFORMATION > RESTORATION	Named River		[**Vegetation removal effect**]AGENT over the entire study reach *changed* the [**Gila River**]PATIENT from a continually losing river for most years before clearing to a gaining stream during some months for most years following clearing.	vegetation removal effect	*improves*	Gila River
PROCESS > CHANGE > CHANGE IN SIZE > INCREASE	Named River		[**Sea level rise**]AGENT *changes*/*affects* [**Salinas River Estuary dynamics**]PATIENT and could thus potentially alter sediment supplies and process patterns.	sea level rise	*worsens*	Salinas River Estuary

Explanation: The above example allowed us to consider that the *sea level rise* process plays the AGENT role because it causes a change for the worse in the *Salinas River Estuary*. This process is really the result of other factors not mentioned in the example, such as the warming of the water and its subsequent expansion, the melting of glaciers, the sinking or rising of the land, the land movement, the pumping of freshwater from coastal aquifers resulting in sinking of the land, and the consolidation of the land (Dean and Dalrymple, 2004, p. 36–37). However, the *sea level rise* process was not annotated as taking the RESULT role because the sentence does not contain any clues which lead to deem the process as the result of other processes.

Named River	ENTITY> […]>LANDFORM		The top of the [**Quanzhou Bay**]AGENT is obviously *affected*/*influenced* by the [**Jinjiang River**]PATIENT.	Jinjiang River	*affects*	Quanzhou Bay
PATIENT ENTITY > MATTER	**LOCATION** Named River		[**Soft muds**]PATIENT *are consolidating* at the [**mouth of the Mississippi River**]LOCATION.	soft mud	*located_at*	Mississippi River Mouth
ATTRIBUTE	Named River	MANNER	The [**discharge rate**]PATIENT at the [**Weser River**]LOCATION *varies* [**greatly over a year**]MANNER, but a typical rate is in the vicinity of 200 m3/s.	discharge rate	*attribute_of*	Weser River
PATIENT Named River	**RESULT** ENTITY> […]>LANDFORM		The [**Chandeleurs Islands**]RESULT *are remnants of* the [**Saint Bernard River delta**]PATIENT, formed by the Mississippi River.	Saint Bernard River Delta	*becomes*	Chandeleurs Islands

the pattern expressed the *improves* relation. The verbs *improve*, *enhance*, and *change* encoded the *improves* relation, whereas the verbs *affect*, *influence*, and also *change* encoded both the *worsens* and *affects* relations.

2. PATIENT + LOCATION: This pattern could convey either of two relations, namely, *located_at* (i.e., an entity is located in a river), or *attribute_of* (i.e., a property that characterizes a river). Nevertheless, the semantic relation could always be differentiated, thanks to the semantic category of the concept with the PATIENT role. As such, the pattern transmitted the *located_at* relation if the concept with the PATIENT role was MATTER (e.g., *soft mud*), while the pattern expressed the *attribute_of* relation if the concept was an ATTRIBUTE (e.g., *discharge rate*).

3. PATIENT + RESULT: This pattern always encoded the *becomes* relation, linking a named delta or estuary that is transformed into any other landform.

Lexical Domain of EXISTENCE

Five different combinations of semantic roles were found for the lexical domain of EXISTENCE (14.28% of sentences), shown in **Table 7**:

1. AGENT + PATIENT: This pattern could convey one of three relations, namely, *result_of* (i.e., a process or an entity is derived from other process), *worsens*, or *creates* (i.e., an entity causes another entity to exist). Nonetheless, the semantic relation could always be distinguished, thanks to the semantic category of the concept with the AGENT role. Accordingly, the pattern transmitted the *result_of* relation if the concept was a MOVEMENT PROCESS (e.g., *sediment supply*). The *worsens* relation was conveyed if the concept was an ADDITION PROCESS (e.g., *salinity intrusion*), and the *creates* relation was conveyed if the concept was a named river. The verbs *form* and *build* expressed the *creates* relation, whereas *result in* and *be/represent/constitute/become a problem/issue/ challenge/trouble/matter* expressed the *worsens* relation.

2. AGENT + PATIENT + RESULT: This pattern always encoded the *worsens* relation.

3. AGENT + RESULT + LOCATION: This pattern transmitted the *takes_place_in* relation.

4. AGENT + THEME: This pattern conveyed the *causes* relation. In the corresponding row of **Table 7**, the verb *provide* in the example sentence has the sense *to cause something to happen,*

TABLE 7 | Results from the semantic annotations for the lexical domain of EXISTENCE.

Verb lexical domain: EXISTENCE

Arg1	Arg2	Arg3	Example	Term	Relation	Term
AGENT PROCESS > MOVEMENT > TRANSPORT	**PATIENT** Named River		The [**Tenryu River Delta**]ₚₐₜᵢₑₙₜ _has developed_ owing to the abundant [**sediment supply**]ₐ𝗀ₑₙₜ from the Tenryu R.	Tenryu River Delta	_result_of_	sediment supply
PROCESS > ADITTION	Named River	**CONDITION** PROCESS > ACTIVITY	For instance, [**salinity intrusion**]ₐ𝗀ₑₙₜ _constitutes a problem_ in the [**Yangtze Estuary**]ₚₐₜᵢₑₙₜ because of the [**vast area using water from the Yangtze**]𝒸ₒₙ𝒹ᵢₜᵢₒₙ.	salinity intrusion	_worsens_	Yangzte River Estuary
Named River	ENTITY > [...] > LANDFORM		The Chandeleurs Islands are remnants of the [**Saint Bernard River delta**]ₚₐₜᵢₑₙₜ, _formed_ by the [**Mississippi River**]ₐ𝗀ₑₙₜ.	Mississippi River	_creates_	Saint Bernard River Delta
AGENT PROCESS > MOVEMENT > TRANSPORT	**PATIENT** Named River	**RESULT** PROCESS > CHANGE > TRANSFORMATION > POLLUTION	Thus, [**several tens of millions of bed load goes into the North Passage of Yangtze estuary along with the ebb currents**]ₐ𝗀ₑₙₜ, and [**it**]ₐ𝗀ₑₙₜ _results in_ a continuous [**siltation**]ᵣₑₛᵤₗₜ in [**North Passage of Yangtze estuary**]ₚₐₜᵢₑₙₜ.	siltation	_worsens_	Yangtze River Estuary
AGENT PROCESS > CHANGE > CHANGE IN INTENSITY > DECREASE	**RESULT** PROCESS > CHANGE > CHANGE IN SIZE > DECREASE	**LOCATION** Named River	After the dam construction, the [**sediment supply decrease**]ₐ𝗀ₑₙₜ in the [**Tenryu River**]ₗₒ𝒸ₐₜᵢₒₙ _resulted in_ a [**delta coastline recession**]ᵣₑₛᵤₗₜ.	sediment supply decrease	_takes_place_in_	Tenryu River
PROCESS > CHANGE > CHANGE IN SIZE > DECREASE	PROCESS > MOVEMENT > SOIL MOVEMENT		The [**consolidation of the land**]ₐ𝗀ₑₙₜ at the [**mouth of the Mississippi River**]ₗₒ𝒸ₐₜᵢₒₙ _is causing_ the [**sinking of the land**]ᵣₑₛᵤₗₜ with respect to an absolute datum.	consolidation of the land		Mississippi River Mouth
AGENT Named River	**THEME** PROCESS > ADDITION		The [**Changjiang River**]ₐ𝗀ₑₙₜ _provides_ most of the [**fresh water input**]ₜₕₑₘₑ.	Changjiang River	_causes_	fresh water input
THEME Named River	**DESCRIPTION** ATTRIBUTE > PHYSICAL ATTRIBUTE > STATE	**TIME** Named Dam & PROCESS > ACTIVITY	[**Before construction of the High Aswan High Dam**]ₜᵢₘₑ, the [**Nile Delta shore**]ₜₕₑₘₑ _showed_ [**fluctuating equilibrium**]𝒹ₑₛ𝒸ᵣᵢₚₜᵢₒₙ between sediment supplied by the Nile River and the transport along the coast.	fluctuating equilibrium	_attribute_of_	Nile Delta
ENTITY > INFORMATION > REPRESENTATION > MODEL	ATTRIBUTE > MEASUREMENT	**LOCATION** Named River	Blackstone River draining into Narragansett Bay has been extensively dammed, and although not well quantified, [**models**]ₜₕₑₘₑ _show_ [**decreasing sediment load**]𝒹ₑₛ𝒸ᵣᵢₚₜᵢₒₙ in the [**Blackstone River**]ₗₒ𝒸ₐₜᵢₒₙ.	decreasing sediment load	_attribute_of_	Blackstone River
PROCESS > CHANGE > CHANGE IN INTENSITY > DECREASE		Named River	The dramatical [**sediment load variation**]ₜₕₑₘₑ in the [**Pearl River**]ₗₒ𝒸ₐₜᵢₒₙ, with the almost unchanged water discharge level, _represents_ an [**example of such effect that human activities can have on river deltas**]𝒹ₑₛ𝒸ᵣᵢₚₜᵢₒₙ.	sediment load variation	_takes_place_in_	Pearl River

making it possible, similar to _allow_, or _permit_ (Faber and Mairal, 1999, p. 279). In other words, the sentence foregrounds that the Changjiang River causes the entry of fresh water into the region (accordingly, in this case, _provide_ is ascribed to the lexical domain of EXISTENCE). In contrast, the fact that the river supplies water recedes into the background (and so, in this context, _provide_ is not regarded as POSSESSION verb).

5. THEME + DESCRIPTION + LOCATION: This pattern expressed one of two relations, namely, _attribute_of,_ or _takes_place_in_. Nevertheless, the semantic relation could always be discriminated, thanks to the semantic category of the concept that took the THEME role. Accordingly, the pattern encoded the _attribute_of_ relation if the concept was an ENTITY (e.g., a mathematical model or a named river),

TABLE 8 | Results from the semantic annotations for the lexical domain of ACTION.

Verb lexical domain: ACTION

Arg1	Arg2	Arg3	Example	Term	Relation	Term
PATIENT ENTITY > CREATION > STRUCTURE > DEFENCE STRUCTURE	**LOCATION** Named River		At least 90 dams over 60 m *are under construction*, including the [**Three Gorges dam**]PATIENT on the [**Yangtze**]LOCATION (175 m).	Three Gorges Dam	*located_at*	Yangtze River
AGENT Named River	**DESCRIPTION** ATTRIBUTE > ABILITY		[**Salinas River Estuary**]AGENT helps improve the quality of water since **it** *acts* both *as a* [**filter**]DESCRIPTION that catches nutrients, sediments, and even pollution, preventing them from moving further into the environment, and *as a* [**buffer**]DESCRIPTION between ocean storms and cities, protecting the inland areas from damage.	Salinas River Estuary Salinas River Estuary	*has_function* or *type_of (function)* *has_function* or *type_of (function)*	filter buffer
PATIENT Named River	**PURPOSE** PROCESS > ADDITION		In the year 1950, the [**Camboriú River**]PATIENT *was dammed* for [**water supply**]PURPOSE, reducing the sediment input to the beach, thus the **river damming** caused coastal erosion.	Camboriú River damming damming damming	*has_function* *causes* *worsens* *has_function*	water supply coastal erosion beach water supply

whereas the pattern transmitted the *takes_place_in* relation if the concept was a PROCESS (e.g., *sediment load variation*). In the corresponding rows of **Table 7**, the verb *show*, in both example sentences, has the sense *to cause something to exist in the perception of others* (Faber and Mairal, 1999, p. 279). For that reason, *show* fits into the lexical domain of EXISTENCE, and not of PERCEPTION.

Lexical Domain of ACTION

Three combinations of semantic roles were found for the lexical domain of ACTION (9.10% of sentences), shown in **Table 8**:

1. PATIENT + LOCATION: This pattern conveyed the *located_at* relation.
2. AGENT + DESCRIPTION: This pattern transmitted the *has_function* relation, namely, a process or an entity (whether natural or artificial) that is linked to its specific function. In this case, a named river, delta, or estuary, despite being a natural entity, which is not goal-directed, is used for human profit (e.g., SALINAS RIVER ESTUARY *has_function* FILTER). Although this relation could also have been regarded as functional hyponymy (e.g., SALINAS RIVER ESTUARY *type_of (function)* FILTER), the annotators agreed to assign the *has_function* relation to this semantic role combination within the lexical domain of ACTION.
3. PATIENT + PURPOSE: This pattern also expressed the *has_function* relation (e.g., CAMBORIÚ RIVER *has_function* WATER SUPPLY). Like most sentences in which a named river is mentioned, the corresponding sentence in **Table 8** with this combination of semantic roles is a high-density knowledge-rich context, namely, a context containing several terms of interest in a particular knowledge domain that are linked to other terms through different semantic relations (Meyer, 2001; León-Araúz and Reimerink, 2019). For instance, the sentence also conveys

the following conceptual propositions associated with RIVER DAMMING, which reflect its multidimensionality. More specifically, although river damming is a beneficial activity because it assures the population of a water supply, it can also be regarded as a damaging activity because of its potentially adverse impact on the environment: DAMMING *causes* COASTAL EROSION; DAMMING *worsens* BEACH; and DAMMING *has_function* WATER SUPPLY.

Lexical Domain of POSITION

Table 9 summarizes the findings for the lexical domain of POSITION (5.19% of sentences). Two combinations of semantic roles were found:

1. AGENT + THEME + LOCATION: This pattern expressed the *places* relation, linking a named river to the matter that it deposits at a particular location. The argument with the THEME role was always matter, whether solid (e.g., *sediment*) or fluid (e.g., *soft mud*). The verbs that frequently encoded this relation were *deposit* and *accumulate*.
2. THEME + LOCATION: This pattern conveyed the *located_at* relation. In contrast to the pattern above, the argument with the THEME role was either a landform (e.g., *salt marsh*) or a defense structure (e.g., *jetty*).

Lexical Domain of MANIPULATION

Table 10 summarizes the findings for the lexical domain of MANIPULATION (5.19% of sentences). Only one combination of semantic roles was found, namely AGENT + PATIENT. This pattern transmitted the *controls* relation, linking a named river to the PROCESS (e.g., *natural sediment supply*) or ATTRIBUTE (e.g., *shoreline characteristic*) that the river

TABLE 9 | Results from the semantic annotations for the lexical domain of POSITION.

Verb lexical domain: POSITION

Arg1	Arg2	Arg3	Example	Term	Relation	Term
AGENT Named River	**THEME** ENTITY > MATTER	**LOCATION** ENTITY>[…]> LANDFORM	The [**Salinas River**]AGENT _deposits/accumulates_ [**sediments**]THEME in the [**flood plain**]LOCATION.	Salinas River	_places_	sediment
THEME ENTITY> […] > LANDFORM ENTITY > CREATION > STRUCTURE > DEFENCE STRUCTURE	**LOCATION** Named River		The field site for this study is the [**Zuidgors salt marsh**]THEME, _located in_ the [**Western Scheldt estuary**]LOCATION in The Netherlands. However, the construction of the [**jetty**]THEME, _located on_ the [**Camboriú River mouth**]LOCATION, changed the beach planform classification to natural beach reshaping or self-reshaping.	Zuidgors salt marsh jetty	_located_at_	Scheldt River Estuary Camboriú River Mouth

TABLE 10 | Results from the semantic annotations for the lexical domain of MANIPULATION.

Verb lexical domain: MANIPULATION

Arg1	Arg2	Arg3	Example	Term	Relation	Term
AGENT Named River	**PATIENT** PROCESS > MOVEMENT > TRANSPORT ATTRIBUTE		[**Natural sediment supply**]PATIENT within this region _is defined_ by the [**Ventura River**]AGENT that drains large watersheds. [**Shoreline characteristics**]PATIENT _are defined_ by [**Santa Clara River**]AGENT.	Ventura River Santa Clara River	_controls_	natural sediment supply shoreline characteristic

manipulates. The most frequent verb employed to express the relation was _define_.

Lexical Domain of COGNITION

Table 11 summarizes the findings for the lexical domain of COGNITION (3.90% of sentences). Two combinations of semantic roles were found:

1. INSTRUMENT + THEME + LOCATION: This pattern encoded the _applied_to_ relation, linking a mathematical model to the named river, whose magnitude was estimated by the model. The argument with the INSTRUMENT role was always a mathematical model (e.g., _NOAH model_), and that of the THEME role was a magnitude (e.g., _evaporation_).
2. THEME + LOCATION + RATE: This pattern conveyed the _attribute_of_ relation. As in the pattern above, the argument with the THEME role was a magnitude, but its value was specified by adding the RATE role.

From the Semantic Annotation to Semantic Frames

The semantic analysis of the predicate-argument structure of the sentences offered a comprehensive set of conceptual propositions. These propositions reflected the entities and processes that participated in the events educed by named rivers through predicates, and their interaction.

Furthermore, the sentences that preceded and followed our set of sentences were also semantically analyzed. This permitted us to construct the frames elicited by the rivers with a much

broader explanatory adequacy. The semantic networks were then validated by a coastal engineering expert from the University of Granada (Spain).

In the following, the frame evoked by the Salinas River is depicted, and also the frame educed by the Dee, Mersey, Ribble, and Solway Firth estuaries.

Frame Evoked by the Salinas River

The Salinas River (California, the United States) evoked the frame shown in **Figure 3**. In this semantic network, Monterey Bay and Salinas River are two named entities, from different categories, associated with the same environmental problem depicted in the frame. Therefore, the semantic network fulfills Tobler's (1970, p. 236) First Law of Geography, which states that "everything is related to everything else, but near things are more related than distant things." In other words, since the Salinas River discharges into the Monterey Bay, both entities are spatially close and thus associated with the negative effects of sea level rise on shoreline erosion.

For the thematic description of the Salinas River, a textual explanation of its relational behavior was crafted from the semantic network as follows.

> **Thematic description of the Salinas River**: Sediment is a resource essential both to the economic and environmental vitality of _Monterey Bay_ beaches and to the mitigation of _shoreline erosion_. The sources of _sand_ to the southern _Monterey Bay_ are from the discharge of the _Salinas River_ and from the erosion of the _beaches_ and _coastal dunes_. However, human

TABLE 11 | Results from the semantic annotations for the lexical domain of COGNITION.

Verb lexical domain: COGNITION

Arg1	Arg2	Arg3	Example	Term	Relation	Term
INSTRUMENT ENTITY > INFORMATION > REPRESENTATION > MODEL	**THEME** ATTRIBUTE > MEASUREMENT > MAGNITUDE	**LOCATION** Named River	Feng and Houser (2008) also find that the [**NOAH model**]INSTRUMENT _underestimates_ [**evaporation**]THEME in the [**Mississippi River basin**]LOCATION.	NOAH model	_applied_to_	Mississippi River Basin
THEME ATTRIBUTE > MEASUREMENT > MAGNITUDE > LEVEL > MEAN	**LOCATION** Named River	**RATE**	The [**average discharge rate of beach size sand**]THEME in the [**Salinas River**]LOCATION _is estimated_ at [**approximately 65,000 cubic yards per year**]RATE.	average discharge rate of beach size sand	_attribute_of_	Salinas River

activities and natural processes are changing sand availability. Namely, _dams_ constructed along the _Salinas River_ have decreased its _sand supply_. Hence, most sediment from the river is driven north and potentially lost into the _Monterey Submarine Canyon_, and _beach sand mining_ and _sea level rise_ cause _dune erosion_ to progress at a higher rate.

On the other hand, conceptual representations in TKBs can be enhanced when specialized concepts are embedded in situations (Meyer et al., 1992; Faber, 2011), for instance, situations geographically contextualized. As such, the representation of named rivers from coastal engineering in EcoLexicon enables the geographic contextualization of specialized concepts from that subject field in semantic networks. In this work, geographic contextualization consists in viewing a specialized concept from a situation in which the concept is related to specific named geographic entities, such as rivers and bays, because it is involved in an environmental problem which affects those geographic entities. Therefore, for the geographic contextualization of the SEA LEVEL RISE concept in the coastal engineering domain, EcoLexicon would show the semantic network in **Figure 3**. The situational elements in such a frame (i.e., concepts and semantic relations) would facilitate to represent and understand that sea level rise is causing dune erosion of Monterey Bay beaches to progress at such a high rate that the sediments discharged by the Salinas River are not enough to alleviate the coastal erosion of the bay. The frame in **Figure 3** would also be valid for the geographic contextualization of any of the specialized concepts that are integrated into that network.

Frame Evoked by the Dee, Mersey, Ribble, and Solway Firth Estuaries

The Dee, Mersey, Ribble, and Solway Firth estuaries (in the United Kingdom) educed the frame in **Figure 4**. The four estuaries are spatially close and associated with the same environmental problem, according to Tobler's (1970) First Law of Geography.

The thematic description of the estuaries was elaborated from the semantic network as follows.

Thematic description of the Dee, Mersey, Ribble, and Solway Firth estuaries: In Great Britain, the _Department for Environment, Food and Rural Affairs_, and the _Welsh Assembly Government_ have required to produce _shoreline management plans_ (SMPs) for the length of coastline which stretches from _Great Orme's Head_ in Wales to the _Scottish Border_ on the _Solway Firth Estuary_, including the _Dee_, _Mersey_, and _Ribble_ estuaries.

The overall aim of SMP is the _flood_ and _erosion risk management_ along the coast. Hence, SMP sets out policies for managing the coastline to reduce those risks to _urban areas_, _industrial_ and _commercial activities_, and natural environments such as _marine protection areas_. One of those policies is the _managed realignment_, namely, removing _coastal defenses_ or building new ones further inland to allow an area to become flooded by the sea. _Managed realignment_, usually pursued in estuarine areas, permits: The restoration of _accommodation space_ containing _sediment sinks_ for sediments mobilized by erosion; _habitat creation_, such as salt marshes and mud flats; and long-term _coastal defense resilience_. However, in areas where there are benefits in reverting to natural processes through _managed realignment_, there may be an increase in _tidal flooding_ or _erosion risk_ with associated negative impacts on _historic assets_.

Other plans, incorporated into the SMP, have been developed to coordinate works for flood and erosion risk management, such as _catchment flood management plans_, which predominantly consider _fluvial flood risks_. SMP also includes a _monitoring program_ to check _shoreline features_ and _wetland bird surveys_, among others, and _strategic studies_, for instance, for the _extreme water level prediction_ in the _Dee River Estuary_.

For the geographic contextualization of a specialized concept such as MANAGED REALIGNMENT, EcoLexicon would show the semantic network in **Figure 4**.
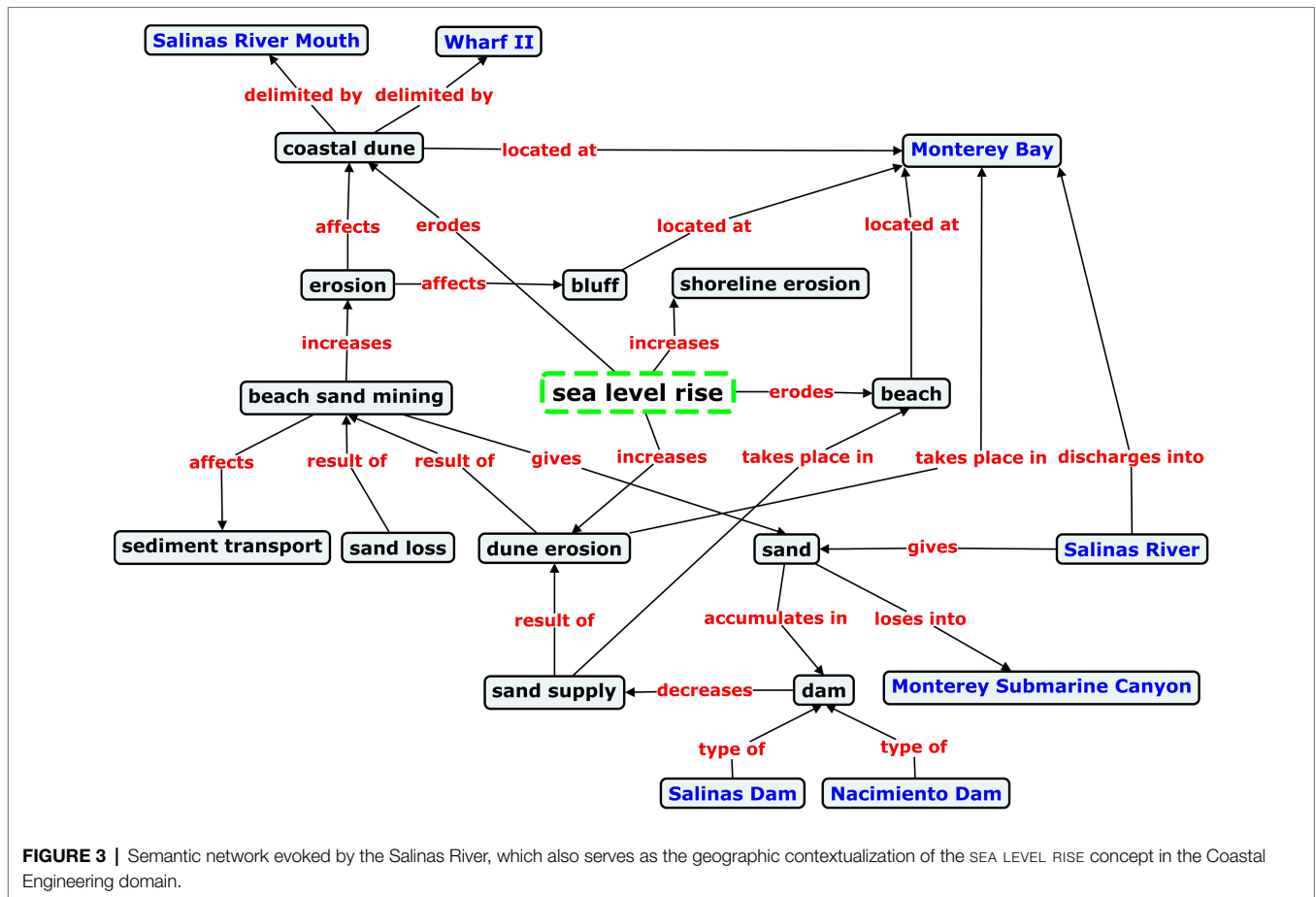
FIGURE 3 | Semantic network evoked by the Salinas River, which also serves as the geographic contextualization of the SEA LEVEL RISE concept in the Coastal Engineering domain.

DISCUSSION

It seems appropriate to embark on a discussion of the semantic behavior of named rivers in the coastal engineering domain. They are not solely conceptualized as the backdrop and scenario where human activities and environmental processes occur, as evidenced by the diversity of semantic roles that named rivers could play. The corpus data revealed that named rivers are generally conceived as *agents*. As such, they initiate natural processes, which in turn have an effect on or produce a result in another entity. Not surprisingly, coastal engineering texts attach major significance to the study of the processes that each named river triggers. For example, rivers deposit sediments, function as a filter to ameliorate pollution, and control shoreline characteristics. They are also deeply involved in the prevention of coastal erosion by supplying sand, and there exists a close relationship between rivers and bays in sediment concentration and transport.

Obviously, for the supply of sand, it is necessary for rivers first to act as *recipients* of sediments and water when draining valleys and plains.

As *patients*, named rivers undergo a change of condition for the better, when defense structures such as jetties, or processes such as vegetation removal, maximize the affordances

offered by rivers. However, their conditions can also worsen when river damming and sea level rise cause the supply of fluvial sediment to decrease, or siltation pollutes rivers.

In the role of *theme*, rivers are subject to change. For instance, they can undergo a change of location, since, when flowing along their course, they cross other entities. They also participate in a change of possession, when they provide sediments to other entities, such as bays, beaches, and dunes, upon discharging into bays, seas, and oceans. Their nature may even change, when they become another type of landform (e.g., a river delta becomes an island). According to Faber and Mairal (1999), one of the most important environmental processes is change. In fact, the results showed that, in the context of named rivers, the change in sediment possession is predominant.

When rivers function as *locations*, the corpus examples specify the following: (1) the entities located on them, whether natural (e.g., *salt march*, *soft mud*) or artificial (e.g., *jetty*, *dam*); (2) the properties of the rivers (e.g., *discharge rate*, *evaporation*, *sediment load*, *runoff*); and (3) the mathematical models applied to predict the values of those properties (e.g., *NOAH model* for evaporation, *Grid-to-Grid model* for river runoff).

Consequently, from this discussion, one can infer that named rivers, at least in situations of specialized
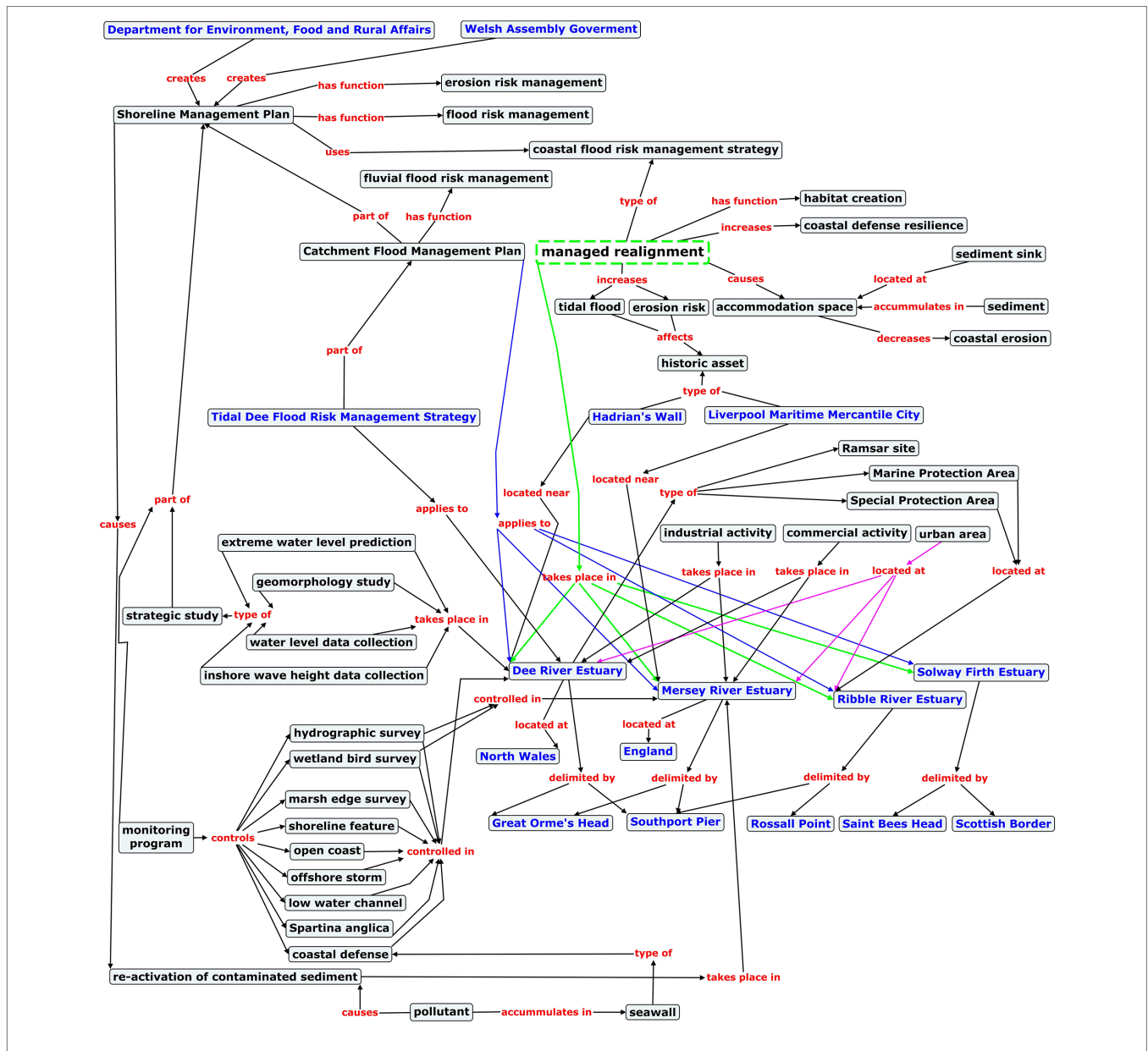
FIGURE 4 | Semantic network evoked by the Dee, Mersey, Ribble, and Solway Firth estuaries, which also serves as the geographic contextualization of the MANAGED REALIGNMENT concept in the Coastal Engineering domain.

communication, are used in ways otherwise than to perform acts of reference. These findings point toward Evans and Wimmer's (1990, p. 274), Searle's (1969, and 1983, Ch. 9), Sjöblom's (2006), and Strawson's (1974/2004) claim as to context cannot be omitted in linguistic accounts of proper names, nor in terminological accounts of named entities relevant to a subject field.

CONCLUSION

A set of 1,694 sentences, in which a potamonym was an argument of the predicate of the sentences, from a coastal

engineering corpus were semantically analyzed and annotated with the lexical domain of the predicates, the semantic role and category of the arguments, and the semantic relation between the arguments. The aim was to propose a linguistic and terminological approach to the study of named entities in scientific discourse to represent them in a TKB within the framework of Frame-based Terminology, more specifically in EcoLexicon.

The semantic analysis and annotation of argument structures were powerful tools that effectively extracted usage information regarding named rivers in coastal engineering texts. The combination of lexical domains, semantic roles, categories, and relations generated frames that reflected the entities and processes

that participated in the events educed by named rivers, and how they all interacted. Knowledge acquisition about named rivers could be conceived as a progressive expansion of meaning, which began at the phrase level, and resulted in the codification of entire semantic frames for named rivers to be represented in TKBs since those frames underlay the usage of named rivers in the corpus.

These propositional representations, derived from the analysis of predicate-argument structure, are a type of *tertium comparationis* that can be used as the basis for semantic equivalence in machine-translation applications (Buendía-Castro and Faber, 2016). In fact, the analysis evidenced that the predicates in the same lexical domain tended to combine with terms in the same or similar semantic categories such as MATTER, LANDFORM, PROCESS OF RESTORATION, PROCESS OF CHANGE IN INCREASE, and MAGNITUDE ATTRIBUTE.

For the conceptualization of the behavior of named rivers in the coastal engineering domain, the representation of a large number of non-hierarchical relations was essential (e.g., *drains*, *discharges_into*, *gives*, *has_path*, *becomes*, *worsens*, and *creates*). These domain-specific relations make knowledge representation more meaningful and connected to reality because they are both multidimensional and dynamic (Faber et al., 2009, p. 16; León-Araúz, 2009, p. 149, 176, and 184). Coastal engineering is a process-oriented domain that studies dynamic physical states (Faber et al., 2006). Therefore, dynamism is a phenomenon that pervades this domain, whose representation requires the use of non-hierarchical relations.

Semantic networks facilitated for EcoLexicon to geographically contextualize those specialized concepts integrated into the river-evoked frames. In other words, the analysis of the local contexts of potamonyms (i.e., the analysis of predicate-argument structure of sentences that mention named rivers) allowed the transition to global contexts (i.e., semantic frames that depicted environmental problems) that encompassed the conceptual networks reflected in the texts as background situations for specialized concepts. Therefore, since context, knowledge, and reasoning are closely intertwined (Brézillon, 2005), it will be examined how the river-evoked frames can be applied to enhance the geospatial modeling of rivers in geographic information systems, as envisaged by Feng et al. (2004), Garrido and Requena (2011), and Lindenschmidt and Carr (2018).

In future research, the statistical analysis of the annotations will be carried out applying machine-learning techniques, specifically decision tree and random forest, to construct a predictive model. It is expected that the results will reveal which rules permit the prediction of the *semantic relation* between two arguments in a sentence from the predictor variables *verb lexical domain*, *semantic role*, and *semantic category*. This is a framework that has not been explored in terminology, and could be beneficial to the implementation of automatic systems that perform semantic annotation, and construction of semantic networks and thematic description of named entities in specialized discourse.

Furthermore, this study focused on the usage of named rivers in a coastal engineering corpus in English language. Therefore, the analysis of cross-cultural differences in conceptualization of this landform, in the same subject field, in other languages such as Spanish, German, and modern Greek is also deferred for further investigation. It is also planned the semantic analysis of colponyms (i.e., named bays), litonyms (i.e., named beaches and coasts), helonyms (i.e., named wetlands), and named protected areas in the coastal engineering.

Finally, another question that will be investigated in future work is the possibility to conceptualize a subject field in EcoLexicon considering the named entities, relevant to that specialized domain, as starting points for knowledge extraction from corpora and for conceptual analysis. Two of the phases of the workflow in terminology work is term extraction and term selection (Chiocchetti et al., 2013). Both phases could be performed in a specialized corpus taking into consideration the terms associated with relevant named entities, in a similar way to the procedure followed in this study for named rivers in a coastal engineering corpus.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because they are subject to ongoing research. Requests to access the datasets should be directed to juanrojas@ugr.es.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Ahlers, D. (2013). "Assessment of the accuracy of GeoNames gazetteer data." in *Proceedings of the 7th Workshop on Geographic Information Retrieval (GIR'13)*, eds. C. Jones and R. Purves. November 5, 2013; Orlando, Florida (New York, NY: ACM), 74–81.

Baker, C. (2009). La sémantique des cadres et le projet FrameNet: une approche différente de la notion de "valence". *Langages* 176, 32–49. doi: 10.3917/lang.176.0032

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *J. Exp. Psychol. Learn. Mem. Cogn.* 11, 629–654. doi: 10.1037/0278-7393.11.1-4.629

Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Lang. Cogn. Process.* 18, 513–562. doi: 10.1080/01690960344000026

Bennett, B., and Agarwal, P. (2007). "Semantic categories underlying the meaning of 'place'," in *Spatial Information Theory. Vol. 4736*. eds. S. Winter, M. Duckham, L. Kulik and B. Kuipers (Berlin/Heidelberg: Springer), 78–95.

Boas, H. C. (2005). Semantic frames as interlingual representations for multilingual lexical databases. *Int. J. Lexicogr.* 18, 445–478. doi: 10.1093/ijl/eci043

Brézillon, P. (2005). "Task-realization models in contextual graphs," in *Modeling and Using Context. Vol. 3554*. eds. A. Dey, B. Kokinov, D. Leake and R. Turner (Berlin: Springer), 55–68.

Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge, United Kingdom: Cambridge University Press.

Bromhead, H. (2018). *Landscape and Culture: Cross-linguistic Perspectives*. Amsterdam/Philadelphia: John Benjamins.

Buendía-Castro, M. (2013). Phraseology in Specialized Language and Its Representation in Environmental Knowledge Resources. dissertation/thesis. Granada: University of Granada.

Buendía-Castro, M., and Faber, P. (2016). "Phraseological correspondence in English and Spanish specialized texts," in *Computerised and Corpus-Based Approaches to Phraseology: Monolingual and Multilingual Perspectives*. ed. G. Corpas (Geneva: Tradulex), 391–398.

Burge, T. (1973). Reference and proper names. *J. Philos.* 70, 425–439. doi: 10.2307/2025107

Burnett, R. C., Medin, D. L., Ross, N. O., and Blok, S. V. (2005). Ideal is typical. *Can. J. Exp. Psychol.* 59, 3–10. doi: 10.1037/h0087453

Chiocchetti, E., Heinisch-Obermoser, B., Löckinger, G., Lušicky, V., Ralli, N., Stanizzi, I., et al. (2013). *Guidelines for Collaborative Legal/Administrative Terminology Work*. Bolzano, Italy: Eurac Research.

Coastal Engineering Research Center (1984). *Shore Protection Manual*. Vicksburg, MS: U.S. Army Corp of Engineers. *Vol. 1, 4th Edn.*

Coastal Sediment Management Workgroup (2009). The California Coastal Sediment Master Plan. Status Report. Santa Rosa, CA: California Geological Survey.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46. doi: 10.1177%2F001316446002000104

Condamines, A. (2018). "Terminological knowledge bases," in *The Routledge Handbook of Lexicography*. ed. P. A. Fuertes-Olivera (London: Routledge), 335–349.

Croft, W., and Cruse, D.A. (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Cruse, A. (2000). *Meaning in Language. An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.

Davies, C. (2009). "Are places concepts? Familiarity and expertise effects in neighborhood cognition," in *Spatial Information Theory. COSIT 2009. Lecture Notes in Computer Science. Vol. 5756*. eds. K. S. Hornsby, C. Claramunt, M. Denis and G. Ligozat (Berlin/Heidelberg: Springer), 36–50.

Davies, C. (2020). Places as fuzzy locational categories. *Acta Psychol.* 202:102937. doi: 10.1016/j.actpsy.2019.102937

Dean, R. G., and Dalrymple, R. A. (2004). *Coastal Processes with Engineering Applications*. Cambridge, United Kingdom: Cambridge University Press.

Edwardes, A. J., and Purves, R. S. (2007). "A theoretical grounding for semantic descriptions of place," in *Web and Wireless Geographical Information Systems. Vol. 4857*. eds. J. M. Ware and G. E. Taylor (Berlin/Heidelberg: Springer), 106–120.

Elugardo, R. (2002). "The predicate view of proper names," in *Logical Form and Language*. eds. G. Preyer and G. Peter (Oxford: Oxford University Press), 467–503.

Ervin-Tripp, S. M. (1996). "Context in language," in *Social Interaction, Social Context and Language: Essays in Honor of Susan Ervin-Tripp*. eds. D. I. Slobin, J. Gerhardt, A. Kyratzis and J. Guo (Mahwah, NJ: Lawrence Erlbaum), 21–36.

Evans, V. (2019). *Cognitive Linguistics. A Complete Guide*. Edinburgh, United Kingdom: Edinburgh University Press, *2nd ed.*

Evans, M. J., and Wimmer, R. (1990). "Searle's theory of proper names, from a linguistic point of view," in *Speech Acts, Meaning and Intentions. Critical Approaches to the Philosophy of John R. Searle*. ed. A. Burkhardt (Berlin: de Gruyter), 259–278.

Faber, P. (2009). The cognitive shift in terminology and specialized translation. *MonTI* 1, 107–134. doi: 10.6035/MonTI.2009.1.5

Faber, P. (2011). The dynamics of specialized knowledge representation: simulational reconstruction or the perception–action interface. *Terminology* 17, 9–29. doi: 10.1075/term.17.1.02fab

Faber, P. (ed.). (2012). *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin: de Gruyter.

Faber, P. (2015). "Frames as a framework for terminology," in *A Handbook of Terminology. Vol. 1*. eds. H. Kockaert and F. Steurs (Amsterdam/Philadelphia: John Benjamins), 14–33.

Faber, P., and Cabezas-García, M. (2019). Specialized knowledge representation: from terms to frames. *RiL* 17, 197–211. doi: 10.18778/1731-7533.17.2.06

Faber, P., and León-Araúz, P. (2014). Specialized knowledge dynamics: From cognition to culture-bound terminology, in *Dynamics and Terminology: An Interdisciplinary Perspective on Monolingual and Multilingual Culture-bound Communication*, eds. R. Temmerman and CampenhoudtM. Van (Amsterdam/Philadelphia: John Benjamins), 135–158.

Faber, P., and León-Araúz, P. (2016). Specialized knowledge representation and the parameterization of context. *Front. Psychol.* 7:196. doi: 10.3389/fpsyg.2016.00196

Faber, P., León-Araúz, P., and Prieto, J. A. (2009). Semantic relations, dynamicity, and terminological knowledge bases. *Curr. Issues Lang. Stud.* 1, 1–23.

Faber, P., León-Araúz, P., and Reimerink, A. (2016). "EcoLexicon: new features and challenges." in *Proceeding of GLOBALEX 2016: Lexicographic Resources for Human Language Technology*. eds. I. Kernerman, I. Kosem, S. Krek, and L. Trap-Jensen. May 24, 2016; Portorož, Slovenia (Ljubljana: Euralex), 73–80.

Faber, P., and Mairal, R. (1999). *Constructing a Lexicon of English Verbs*. Berlin: Mouton de Gruyter.

Faber, P., Montero, S., Castro, M. R., Senso, J., Prieto, J. A., León-Arauz, P., et al. (2006). Process-oriented terminology management in the domain of coastal engineering. *Terminology* 12, 189–213. doi: 10.1075/term.12.2.03fab

Faber, P., Verdejo, J., León-Araúz, P., Reimerink, A., and Guzman, G. (2014). Neural substrates of specialized knowledge representation: an fMRI study. *Revue Française de Linguistique Appliquée* XIX, 15–32. doi: 10.3917/rfla.191.0015

Fara, D. G. (2015). Names are predicates. *Philos. Rev.* 124, 59–117. doi: 10.1215/00318108-2812660

Feng, C. C., Bittner, T., and Flewelling, D. M. (2004). "Modeling surface hydrology concepts with endurance and perdurance," in *Geographic Information Science*. eds. M. J. Egenhofer, C. Freksa and H. J. Miller (Berlin: Springer), 67–80.

Feng, X., and Houser, P. (2008). An investigation of GSWP-2 Mississippi River basin surface water and energy budgets. *J. Geophys. Res.* 113:D15118. doi: 10.1029/2007JD009144

Fillmore, C. J. (2006). "Frame semantics," in *Cognitive Linguistics: Basic Readings*. ed. D. Geeraerts (Berlin: Mouton de Gruyter), 373–400.

Frege, G. (1892/1952). "On Sense and Reference," in *Translations From the Philosophical Writings of Gottlob Frege*. eds. P. Geach and M. Black (Oxford: Blackwell), 56–79.

Gardiner, A. (1940). *The Theory of Proper Names: A Controversial Essay*. London: Oxford University Press.

Garrido, J., and Requena, I. (2011). Proposal of ontology for environmental impact assessment: an application with knowledge mobilization. *Expert Syst. Appl.* 38, 2462–2472. doi: 10.1016/j.eswa.2010.08.035

Gil-Berrozpe, J.C, León-Araúz, P., and Faber, P. (2019). "Ontological knowledge enhancement in EcoLexicon, in *Electronic Lexicography in the 21st Century.*" in *Proceedings of the eLex 2019 Conference*. October 1–3, 2019; Sintra, Portugal (Brno, Czech Republic: Lexical Computing CZ), 177–197.

Hirtle, S. C., and Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Mem. Cogn.* 13, 208–217. doi: 10.3758/BF03197683

International Organization for Standardization (ISO) (2009). *Terminology Work—Principles and Methods*. ISO 704:2009. Geneva: ISO.

Kilgarriff, A., Rychly, P., Smrz, P., and Tugwell, D. (2004). "The sketch engine." in *Proceedings of the 11th EURALEX International Congress*, eds. G. Williams and S. Vessier. July 6–10, 2004; Lorient, France (Lorient, France: Université de Bretagne-Sud), 105–116.

Klie, J.C., Bugert, M., Boullosa, B., Eckart de Castilho, R., and Gurevych, I. (2018). "The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation.' in *Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018)*. August 20–26, 2018; Santa Fe, New Mexico (Stroudsburg, PA: ACL), 5–9.

Kripke, S.A. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Krippendorff, K. (2012). *Content Analysis: An Introduction to Its Methodology*. London: Sage.

Kroeger, P.R. (2005). *Analyzing Grammar: An Introduction*. Cambridge, United Kingdom: Cambridge University Press.

León-Araúz, P. (2009). Representación Multidimensional del Conocimiento Especializado: el Uso de Marcos desde la Macroestructura hasta la Microestructura. dissertation/thesis. Granada: University of Granada.

León-Araúz, P., and Reimerink, A. (2019). High-density knowledge rich contexts. *AJAL* 7, 109–130.

León-Araúz, P., Reimerink, A., and Faber, P. (2020). "Translating Environmental Texts With EcoLexiCAT," in *Translating and Communicating Environmental Cultures*. ed. M. Ji (London: Routledge), 3–42.

León-Araúz, P., San Martín, A., and Reimerink, A. (2018). "The EcoLexicon English corpus as an open corpus in sketch engine." in *Proceedings of the 18th EURALEX International Congress*. July 17–21, 2018; Ljubljana, Slovenia (Ljubljana: Euralex), 893–901.

Levin, B. (2013/18). "Argument structure," in *Oxford Bibliographies in Linguistics*. ed. M. Aronoff (New York: Oxford University Press)

Levinson, S.C. (2003). *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge, United Kingdom: Cambridge University Press.

L'Homme, M. C. (2003). Capturing the lexical structure in special subject fields with verbs and verbal derivatives: a model for specialised lexicography. *Int. J. Lexicogr.* 16, 403–422. doi: 10.1093/ijl/16.4.403

L'Homme, M.C. (2020). *Lexical Semantics for Terminology*. Amsterdam/Philadelphia: John Benjamins.

Lindenschmidt, K. E., and Carr, M. K. (2018). Geospatial modeling of river systems. *Water* 10:282. doi: 10.3390/w10030282

Lloyd, R., Patton, D., and Cammack, R. (1996). Basic-level geographic categories. *Prof. Geogr.* 48, 181–194. doi: 10.1111/j.0033-0124.1996.00181.x

Lynch, E. B., Coley, J. D., and Medin, D. L. (2000). Tall is typical: central tendency, ideal dimensions and graded category structure among tree experts and novices. *Mem. Cogn.* 28, 41–50. doi: 10.3758/bf03211575

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., and McClosky, D. (2014). "The stanford CoreNLP natural language processing toolkit." in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. June 23–24, 2014; Baltimore, Maryland (Stroudsburg, PA: ACL), 55–60.

Martin, L. (1984/2017). "Lexical fields and stepwise lexical decomposition in a contrastive English-Spanish verb valency dictionary." in *LEXeter'83: Proceedings: Papers from the International Conference on Lexicography*, ed. R. Hartmann. September 9–12, 1983; Exeter, United Kingdom (Berlin/Boston: Max Niemeyer Verlag), 226-236.

Matushansky, O. (2008). On the linguistic complexity of proper names. *Linguist. Philos.* 31, 573–627. doi: 10.1007/s10988-008-9050-1

Meyer, I. (2001). "Extracting knowledge-rich contexts for terminography: a conceptual and methodological framework," in *Recent Advances in Computational Terminology*. eds. D. Bourigault, C. Jacquemin and M. C. L'Homme (Amsterdam/Philadelphia: John Benjamins), 279–302.

Meyer, I., Skuce, D., Bowker, L., and Eck, K. (1992). "Towards a new generation of terminological resources: an experiment in building a terminological knowledge base." in *Proceedings of the 14th International Conference on Computational Linguistics*, *Vol. 3*. August 23–28, 1992; Nantes, France (Nantes: ICCL), 956–960.

Michalski, R.S. (1991). *Concepts as Flexible and Context-Dependent Sets: The Two-Tiered View*. Fairfax: George Mason University.

Mill, J.S. (1843/2002). *A System of Logic: Ratiocinative and Inductive*. Honolulu, HI: University Press of the Pacific.

Montello, D. R. (2003). "Regions in geography: process and content," in *Foundations of Geographic Information Science*. eds. M. Duckham, M. F. Goodchild and M. Worboys (London: Taylor & Francis), 173–189.

Moskalski, S., and Torres, R. (2012). Influences of tides, weather, and discharge on suspended sediment concentration. *Cont. Shelf Res.* 37, 36–45. doi: 10.1016/j.csr.2012.01.015

Murphy, G. L., and Lassaline, M. E. (1997). "Hierarchical structure in concepts and the basic level of categorization," in *Knowledge, Concepts and Categories*. eds. K. Lamberts and D. R. Shanks (Cambridge, MA: The MIT Press), 93–131.

Pimentel, J. (2015). "Using frame semantics to build a bilingual lexical resource on legal terminology," in *Handbook of Terminology*. *Vol. 1*. eds. H. J. Kockaert and F. Steurs (Amsterdam/Philadelphia: John Benjamins), 425–450.

Quine, W.V.O. (1960). *Word and Object*. Cambridge, MA: MIT Press.

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. New York, NY: Longman Publishers.

R Core Team (2021). R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. R version 4.1.1 (August 2021). Available at: https://www.R-project.org/

Renschler, C. S. (2004). "Watershed," in *Encyclopedia of Geomorphology*. *Vol. 1* ed. A. S. Goudie (London/New York: Routledge), 1101.

Rogers, M. (2004). Multidimensionality in concepts systems: a bilingual textual perspective. *Terminology* 10, 215–240. doi: 10.1075/term.10.2.04rog

Room, A. (1996). *An Alphabetical Guide to the Language of Name Studies*. Lanham, MD: The Scarecrow Press.

Room, A. (2013). *Placenames of the World: Origins and Meaning of the Names for 6,600 Countries, Cities, Territories, Natural Features and Historic Sites*. Jefferson, NC/London: McFarland, *2nd Edn.*

Rosch, E. (1978). "Principles of categorization," in *Cognition and Categorization*. eds. E. Rosch and B. L. Lloyd (Hillsdale, NJ: Lawrence Erlbaum), 27–47.

Rosch, E., Simpson, C., and Miller, R. S. (1976). Structural bases of typicality effects. *J. Exp. Psychol. Hum. Percept. Perform.* 2, 491–502. doi: 10.1037/0096-1523.2.4.491

Roussey, C., Hernandez, N., and Zargayouna, H. (2018). "Domain ontologies," in *The Routledge Handbook of Lexicography*. ed. P. A. Fuertes-Olivera (London: Routledge), 217–324.

Russell, B. (1905/1988). "On denoting," in *Bertrand Russell: Logic and Knowledge. Essays 1901-1950*. ed. R. C. Marsh (London: Routledge), 41–56.

Sager, J. C. (1990). *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: John Benjamins.

San Martín, A. (2021). A flexible approach to terminological definitions: representing thematic variation. *Int. J. Lexicogr.* 2021:ecab013. doi: 10.1093/ijl/ecab013

San Martín, A., Cabezas-García, M., Buendía-Castro, M., Sánchez-Cárdenas, B., León-Araúz, P., Reimerink, A., et al. (2020). Presente y futuro de la base de conocimiento terminológica EcoLexicon [Present and future of the EcoLexicon terminological knowledge base]. *Onomázein* 49, 174–202. doi: 10.7764/onomazein.49.09

Searle, J. R. (1969). *Speech Acts. An Essay in the Philosophy of Language*. Cambridge, United Kingdom: Cambridge University Press.

Searle, J. R. (1983). *Intentionality. An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.

Segev, A., and Gal, A. (2008). Enhancing portability with multilingual ontology-based knowledge management. *Decis. Support. Syst.* 45, 567–584. doi: 10.1016/j.dss.2007.07.011

Sharifian, F. (2011). *Cultural Conceptualisations and Language: Theoretical Framework and Applications*. Amsterdam/Philadelphia: John Benjamins.

Sjöblom, P. (2006). A cognitive approach to the semantics of proper nouns. *Onoma* 41, 63–82. doi: 10.2143/ONO.41.0.2119611

Smith, B., and Mark, D. M. (2001). Geographical categories: an ontological investigation. *Int. J. Geogr. Inf. Sci.* 15, 591–612. doi: 10.1080/13658810110061199

Smith, B., and Mark, D. M. (2003). Do mountains exist? Towards an ontology of landforms. *Environ. Plann. B Plann. Des.* 30, 411–427. doi: 10.1068/b12821

Stock, K., Jones, C. B., and Tenbrink, T. (2019). "Speaking of location: Communicating about space with geospatial natural language." in *Proceedings of the Workshop on Speaking of Location 2019: Communicating about Space co-located with 14th International Conference on Spatial Information Theory (COSIT 2019)*, eds. K. Stock, C. B. Jones, and T. Tenbrink. September 10, 2019; Regensburg, Germany (Aachen: CEUR Workshop Proceedings), 1–7.

Stokes, N., Li, Y., Moffat, A., and Rong, J. (2008). An empirical study of the effects of NLP components on geographic IR performance. *Int. J. Geogr. Inf. Sci.* 22, 247–264. doi: 10.1080/13658810701626210

Strawson, P.F. (1974/2004). *Subject and Predicate in Logic and Grammar*. London: Routledge.

Temmerman, R., and Kerremans, K. (2003). "Termontography: Ontology building and the sociocognitive approach to terminology description." in *Proceedings of XVII International Congress of Linguists (CIL 17)*, eds. E. Hajicová, A. Kotesovcová, and J. Mírovský. July 24–29, 2003; Prague, Czech Republic (Prague: Matfyzpress), 1–10.

Tenbrink, T. (2007). *Space, Time, and the Use of Language: An Investigation of Relationships*. Berlin: Mouton de Gruyter.

Thompson, P., Iqbal, S. A., McNaught, J., and Ananiadou, S. (2009). Construction of an annotated corpus to support biomedical information extraction. *BMC Bioinform.* 10:349. doi: 10.1186/1471-2105-10-349

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* 46, 234–240. doi: 10.2307/143141

Ureña, J. M., Faber, P., and Buendía-Castro, M. (2013). Frame blending in specialized language. *Terminology* 19, 175–201. doi: 10.1075/term.19.2.02gom

Van Langendonck, W. (2007). *Theory and Typology of Proper Names*. Berlin/ New York: Mouton de Gruyter.

Van Valin, R. Jr. (2004). "Semantic macroroles in role and reference grammar," in *Semantische Rollen*. eds. R. Kailuweit and M. Hummel (Tübingen, Germany: Narr), 62–82.

Wimmer, R. (1973/2011). Der Eigenname im Deutschen. Ein Beitrag zu seiner linguistischen Beschreibung [The proper name in German. A contribution to its linguistic description]. Berlin: de Gruyter.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.