Strategies for the Semi-Automatic Retrieval of Metaphorical Terms

José Manuel Ureña Gómez-Moreno and Pamela Faber
University of Granada

This article proposes a method for the semi-automatic extraction of resemblance metaphor terms from a manually annotated corpus of marine biology texts in English and Spanish. The corpus was first searched for target domain terms as well as for lexical markers indicative of metaphors. The combination of these search strategies for metaphor extraction resulted in a set of English-Spanish term pairs. After analysing and comparing these metaphor candidates, a quantitative analysis provided comparative statistical data regarding marine biology metaphor. Finally, the metaphorical nature of marine biology terms was verified in three ways. The first verification strategy entailed an adapted version of the Metaphor Identification Procedure (Pragglejaz Group, 2007). The second involved the analysis of contextual data extracted from the corpus, and the third involved the analysis of visual images from an online marine biology database and from the Google search engine.

There is a range of corpus-based studies underlying the role that metaphor plays in particular genres and discourses from a monolingual perspective (e.g., Charteris-Black, 2004 in politics, press reporting, and religion) as well as a cross-linguistic perspective (e.g., Charteris-Black & Musolff, 2003 and Semino, 2002 in financial reporting; Musolff, 2003 in politics). Although these studies deal with specific subject fields, none of them focuses on scientific domains.

Our study is innovative because it analyses terminological metaphor in a naturally-occurring corpus of texts from a cross-linguistic perspective, and applies a systematic and time-efficient method of identifying metaphorical terms in corpora. Earlier corpus-based studies of terminological metaphor (e.g., Alexiev, 2005; Caballero, 2006) do not propose a method of this nature, and some of them do not even analyse metaphorical terms in discourse, but rather isolated instances extracted from specialised dictionaries and glossaries (e.g., Alexiev, 2005).

Regardless of the theoretical approach, discourse type, objectives, and number of languages examined, searching a corpus for instances of metaphorical usage is not an easy task since metaphorical mappings are not straightforwardly associated with particular word forms (Koller, 2006, p. 241). For this reason, the vast majority of corpus-based research on metaphor cannot rely
on annotated corpora (Stefanowitsch, 2006, p. 2). This body of research draws on manual searching and on the four semi-automatic strategies suggested for detecting metaphorical expressions in a corpus: (i) searching for source domain vocabulary; (ii) searching for target domain vocabulary; (iii) searching for sentences containing lexical items from both the source domain and the target domain; and (iv) searching for markers of metaphor (Stefanowitsch, 2006, pp. 2–5). These four strategies yield interesting results, but they have seldom been combined in one research study. Consequently, new metaphorical patterns are often not detected.

As a way to solve this problem, this paper describes a method of finding metaphorical terms in a corpus of English and Spanish research articles on marine biology. In the first phase, the corpus was searched for target domain keywords. The great potential of this strategy has been documented by various researchers (cf. Tissari, 2003; Stefanowitsch, 2004, 2006; Koivisto-Alanko & Tissari, 2006), but never applied to scientific texts. Secondly, a search was made for a set of lexical markers that are typical of scientific discourse. Thanks to these markers, it was possible to find metaphorical terms that could not be retrieved by means of pre-defined target domain keywords. Besides being quicker and more effective than manual searching, the combination of both these search strategies was found to successfully retrieve metaphors, also providing interlinguistic information regarding terminological metaphor. A tagging system was also applied that further exploited such lexical markers, and helped to identify other metaphorical terms.

The advantages of manually tagging a corpus have been verified in many studies on a wide range of discourse types (e.g., Koller, 2006 in business media; Sardinha, 2008 in banking; Semino, 2006 in prose fiction, newspaper news reporting, and biography and autobiography; Wallington, Bardnen, Buchlovsky, Fellows & Glasbey, 2003, in medicine). Our tagging system was effective since it provided a quantitative account of metaphor in English and Spanish. This has valuable implications for metaphor theory because only a few studies have ever provided numerical data regarding the use of metaphor in language (Sardinha, 2008, p. 127). Such results are also conducive to establishing a standard of comparison for cross-linguistic research on terminological metaphor. This standard is relevant because cross-linguistic variation of metaphor is one of the areas in which knowledge is currently tentative or lacking (Stefanowitsch, 2006, p. 9).

Finally, we propose a method to verify the metaphorical nature of the candidate terms found in our corpus. This method, which complements the results of the qualitative and quantitative analyses, involves using an adapted version of the Metaphor Identification Procedure (Pragglejaz Group, 2007), the analysis of contextual data found in the marine biology corpus, and the analysis of visual images extracted from an online marine biology database and from the Google search engine.

TERMINOLOGICAL METAPHOR

Metaphor has been a focus of interest for scholars in a wide range of disciplines. It is a point of intersection for fields such as linguistics, literary studies, psychology, and philosophy of science, where the question of how to deal with metaphor is probably as old as the philosophy of science itself (Goschler, 2007, p. 7). Nevertheless, metaphor in science was long regarded as at best irrelevant and at worst detrimental (Semino, 2008, p. 131) since it supposedly diverged
from the empirical method. From a linguistic perspective, it was thought of as an obstacle to scientific communication, which in principle should be literal and free of synonymy and polysemy. It was not until the 1960s that experts in science, philosophy of science and epistemology began to acknowledge metaphor as a valuable instrument for specialised concept formation and denomination (Hesse, 1974; Kuhn, 1967).

This relatively late shift in perspective means that research on terminological metaphor is still relatively new. Boyd (1993, p. 486) presents a three-level typology of scientific metaphor: (i) *theory-constitutive metaphors*, which are “an irreplaceable part of the linguistic machinery of scientific theory”; (ii) *heuristic metaphors*, which support scientific knowledge by presenting new conceptual perspectives on ideas specific of each specialised discourse; and (iii) *exegetical metaphors*, which are contrived to play a pedagogical role and make science more understandable.

Boyd's claim has been verified in a variety of specialised domains, such as genetics (Knudsen, 2003), medicine (Baran, 2004), and science teaching (Bradford, Dana & Thomas 1996; Ciapuscio, 2004). The metaphorical terms analysed in this study are heuristic metaphors. In theory, they are not indispensable to the scientific field because, as shall be seen, the organisms that they designate can also be referred to by their Latin names. However, each metaphorical term foregrounds visual and/or behavioural information of the marine organism. Presumably, this has the effect of facilitating its mental retrieval more quickly than its Latin counterpart.

Over the last thirty years, metaphor has largely been studied within Cognitive Linguistics, which combines premises of cognitive sciences, psychology, and linguistics. Since Lakoff and Johnson (1980), research in general language has confirmed that metaphor is an integral part of everyday thought processes. The presence of conceptual metaphor in specialised language has also been studied. For instance, Temmerman (2000, 2006) affords insight into metaphorisation in genetics and biotechnology. She argues that this process critically influences category configuration in specialised domains. Caballero (2006) gives an account of the linguistic metaphors used by architects to assess design solutions in building reviews, and the conceptual mappings that underlie such linguistic metaphors. Alexiev (2005) contrastively analyses metaphorical architecture and civil engineering terms in Bulgarian, English, and Spanish with a view to filling a wide gap in the area of contrastive studies on metaphorical terms and their possible translation equivalents (Alexiev, 2005, p. 43).

The description of terminological metaphor traditionally involves the distinction between conventional (conceptual-structural) and resemblance metaphors (Alexiev, 2005, pp. 40). Resemblance metaphors arise because of comparison in physical appearance (i.e., image metaphors) or in behaviour (Grady, 1999), whereas conceptual metaphors emerge from richer and more abstract structures (Lakoff & Turner, 1989, p. 91). This typology has recently been modified and extended by studies in various specialised fields (cf. Caballero, 2006 in architecture; Ureña & Faber, 2010 in marine biology).

The first modification concerns resemblance metaphors. Evidence has been provided that image metaphors and behaviour-based metaphors should not be regarded as two clear-cut types of resemblance metaphor (Ureña & Faber, 2010). The dynamicity or stationariness of certain mental images highlights inherent similarities between both types of metaphor. Consequently, there is a group of metaphors that possess characteristics of both, and thus reside in a transition zone between the two. For instance, the species *boxer crab* (*Lybia tesselatta*) holds an anemone
in each pincer, and uses these anemones for protection (usually against octopuses) in the same way as a boxer uses his fists against his opponent. These tiny round-shaped anemones resemble boxing gloves (comparison in shape), while the action of attacking predators with the anemones is a type of behaviour that resembles that of a boxer.

The second modification entails the relationship between resemblance metaphors and structural metaphors. Research has revealed that both types share features which until recently had been attributed exclusively to structural metaphors (cf. Caballero, 2006; Ureña & Faber, 2010). One of these features is conceptual structure, which is now shown to be inherent not only to conceptual-structural metaphors, but also to certain resemblance metaphors. Accordingly, some well-entrenched resemblance metaphors can be brought together under productive, encompassing metaphors resembling multiple-correspondence processes.

For example, terms such as surgeonfish (Acanthuridae), pilotfish (Naucrates ductor), anglerfish (Lophius), fiddler crab (Uca), harvestfish (Peprilus alepidotus), by-the wind sailor (Vellela spirans), boxer crab (Lybia tessellata), nurse shark (Ginglymostoma cirratum), innkeeper worm (Urechis), and rock cook (Centrolabrus exoletus) can be subsumed by the general metaphor MARINE ORGANISMS ARE WORKERS. In other words, resemblance metaphors may be conceptual as well as structural metaphors. For this reason, it has been proposed that conceptual-structural metaphors should be known as non-resemblance metaphors (Ureña & Faber, 2010). Precisely for this reason, the term conceptual metaphor is exclusively used in this study as opposed to linguistic metaphor, not to resemblance metaphor.

This study only focuses on resemblance metaphors (i.e., metaphors arising from physical and/or behavioural comparison). As shall be shown, our analysis supports the claim that the two types of resemblance metaphor, image metaphors and behaviour-based metaphors, are not clearly differentiated categories by providing empirical evidence of their motivation for metaphorical transfer. This article does not deal with non-resemblance metaphors because this would be beyond the scope of this study. Moreover, some of the term retrieval strategies in this study can only be applied to resemblance metaphors.

METHOD

Justification

This article provides evidence of the great potential of metaphor to model scientific thought and of the pivotal role that it plays in the terminologisation of general language lexical units. It explores the field of marine biology to show that this domain contains a rich repository of metaphorical terms, which are used by experts to refer to a wide variety of sea entities, and consequently, make and communicate science. Evidence is also provided of the existence of English-Spanish metaphorical term pairs, which are contrasted and compared. Finally, the analysis of this domain reveals cases of synonymy, lending support to studies carried out in other specialised fields (Tercedor & Méndez, 2000).

1See Ureña and Faber (2010, p. 137) for more shared characteristics of both types of metaphor.
Materials

This study proposes a systematic way of finding resemblance metaphor terms, i.e. well-entrenched linguistic units with specialised meaning. These terms were extracted from a corpus of specialised texts in order to ensure authentic, naturally occurring data. According to Charteris-Black (2004, p. 19), corpus evidence helps the user to detect cases of inactive conventional metaphors and compensate for the arbitrariness of dictionaries.

Corpora of small-scale studies of metaphor are generally compiled by the researchers themselves. This was also the case for this study though our corpus is somewhat larger than average (4,550,190 tokens) because we wished to obtain results as objective and representative as possible. Our corpus was composed of research articles in academic journals on marine biology. The Journal Citation Reports (JCR) website\(^2\) was found to be useful for the classification and quality evaluation of the journal articles. The following factors were considered:

- Citation index of the journal
- Subject of the journals
- Topic of the articles (only English-language articles)
- Availability of complete on-line articles
- Date of publication

The Web provides quick and easy access to electronic texts, which can subsequently be compiled and processed with corpus analysis software programmes. These articles were converted into plain text `.txt` for their processing in Wordsmith Tools\(^3\), a lexical analysis programme.

The JCR is an on-line service which provides a ranking of the most cited journals published by over 3,000 publishers worldwide. Journals are ranked according to a citation index defined by the website itself. Apart from the citation index criterion, the English-language journals used were singled out for the publishing companies that issue them. With the exception of *Ciencias Marinas*, all of the journals used are published either by Springer, Blackwell or Elsevier. *Ciencias Marinas* is a Mexican bilingual journal that publishes English and Spanish research papers in all areas of marine science. This was useful to identify interlinguistic term pairs though the number of articles chosen from this journal was limited to avoid computing possible literal translations of English terms into Spanish. Another criterion for selecting articles was their subject matter, which was as diverse as possible in order to guarantee a more varied set of metaphorical terms. The bibliographic references of the articles were compiled as well because they also contained terminological data. Finally, all articles selected were published between 1992 and 2010. Table 1 lists the English-language journals used in this study\(^3\).

Unfortunately, this classification could not be applied to the Spanish language journals because of their JCR ranking. Only three journals appeared on the list (see Table 2), and all had low rankings. Although the rest of the journals in the Spanish corpus were not on the JCR ranking list, they were considered to be quality publications. The journal *Boletín del Instituto Español de Oceanografía* is published by the Spanish Ministry of Science and Innovation. The other journals are published either on the *SciELO* (*Scientific Electronic Library Online*) or Redalyc (*Red de Revistas Científicas de América Latina, el Caribe, España y Portugal*) websites. These websites

\(^2\)http://www.scimagojr.com/index.php
\(^3\)JCR index consulted on 15 March 2010.
TABLE 1
English-Language Journals

<table>
<thead>
<tr>
<th>English-Language Journal</th>
<th>JCR Citation Index</th>
<th>Number of Articles</th>
<th>Number of Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fish and Shellfish Immunology</td>
<td>0.183</td>
<td>35</td>
<td>276,929</td>
</tr>
<tr>
<td>Coral Reefs</td>
<td>0.127</td>
<td>46</td>
<td>365,173</td>
</tr>
<tr>
<td>Marine Biology</td>
<td>0.104</td>
<td>45</td>
<td>356,736</td>
</tr>
<tr>
<td>Helgoland Marine Research</td>
<td>0.082</td>
<td>22</td>
<td>172,441</td>
</tr>
<tr>
<td>Journal of Fish Biology</td>
<td>0.074</td>
<td>25</td>
<td>196,570</td>
</tr>
<tr>
<td>Environmental Biology of Fishes</td>
<td>0.068</td>
<td>47</td>
<td>373,249</td>
</tr>
<tr>
<td>Marine Ecology</td>
<td>0.063</td>
<td>30</td>
<td>235,477</td>
</tr>
<tr>
<td>Fish Physiology and Biochemistry</td>
<td>0.054</td>
<td>25</td>
<td>188,960</td>
</tr>
<tr>
<td>Ciencias Marinas</td>
<td>0.041</td>
<td>10</td>
<td>74,792</td>
</tr>
</tbody>
</table>

Total: 280 Total: 2,240,327

TABLE 2
Spanish-Language Journals

<table>
<thead>
<tr>
<th>Spanish-Language Journal</th>
<th>JCR Citation Index</th>
<th>Number of Articles</th>
<th>Number of Tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ciencias Marinas</td>
<td>0.041</td>
<td>10</td>
<td>74,792</td>
</tr>
<tr>
<td>Revista de Biología Marina y Oceanografía</td>
<td>0.032</td>
<td>53</td>
<td>450,335</td>
</tr>
<tr>
<td>Boletín del Instituto Español de Oceanografía</td>
<td>0.028</td>
<td>68</td>
<td>609,998</td>
</tr>
<tr>
<td>Investigaciones Marinas</td>
<td>—</td>
<td>51</td>
<td>449,506</td>
</tr>
<tr>
<td>Revista de Biología Tropical</td>
<td>—</td>
<td>37</td>
<td>252,069</td>
</tr>
<tr>
<td>Boletín de Investigaciones Marinas y Costeras</td>
<td>—</td>
<td>56</td>
<td>473,163</td>
</tr>
</tbody>
</table>

Total: 275 Total: 2,309,863

follow strict norms, guidelines, and selection criteria that guarantee the quality of the scientific journal articles that they host.4

Procedure

After compiling the corpus, the next step was to devise a procedure to identify resemblance metaphors. Candidate metaphor units were extracted semi-automatically by the following strategies:

(i) Searching for target domain vocabulary;
(ii) Searching for lexical markers of metaphor. The lexical markers used in this study were the following: (a) taxonomic designations; (b) scientists’ names; (c) the phrases known as and conocido/a como.

5The taxonomic designation of a species is the Latin name in binomial nomenclature used by the scientific community to classify such a species into a specific taxon. The first and the second constituents of the binomial refer to the genus and the specific name, respectively. Both constituents must be written in italics (e.g. Dicentrarchus labrax).
Both strategies entailed identifying keywords in the corpus and concordancing them to discover metaphorical terms.

Firstly, a preliminary, qualitative approach was carried out to find key inferences regarding metaphor in marine biology. A quantitative analysis was then performed with a view to reinforcing and refining these inferences statistically. For this purpose, the marine biology corpus was manually annotated using a tagset that complies with the standards of conciseness, perspicuity, and analysability (Leech, 1997, pp. 25–26). The annotated corpus was then processed with Wordsmith Tools to obtain comparative statistical data concerning the absolute frequencies of metaphor occurrences as well as the interlinguistic metaphorical term pairs that had been extracted. As shall be seen, this method was found to be more productive and time-efficient than manual searching.

The potentially metaphorical nature of the terms retrieved from the bilingual corpus was determined by combining a set of strategies. The first strategy involved an adapted version of the Metaphor Identification Procedure (Pragglejaz Group, 2007) to marine biology discourse. The second strategy concerned the use of contextual data extracted from the corpus, and the third strategy entailed obtaining visual evidence from an online marine biology database. This method was applied to the candidate terms as they were retrieved in the qualitative and quantitative analyses.

RESULTS AND DISCUSSION

Qualitative Analysis

This section demonstrates the effectiveness of the strategies used in this study by providing illustrative examples extracted from the marine biology corpus.

Extracting Resemblance Metaphors with Target Domain Keywords.

Given the size of our corpus, it was necessary to first search for keywords in the target domain. This strategy generally yields good results when a large body of representative and relatively monothematic texts is available (Stefanowitsch, 2006, p. 3). It is also effective because most domain-specific terms are compound nouns, in other words, uninterrupted collocations (Nakagawa & Mori, 2003, p. 201). Searching for target domain lexis involves searching for complex metaphorical expressions in which at least one source domain item is required for the whole lexical unit to be regarded as metaphorical. For instance, the target domain word fish needs to occur with clown to form the metaphorical term clownfish.

As a starting point, frequency was taken as the main factor defining salience (Geeraerts, 1997, p. 44; Koivisto-Alanko, 2000, pp. 47–48; Koivisto-Alanko, 2006, p. 194). In our view, the most frequent lexical words in a corpus are domain-relevant units indicative of large semantic classes that are a breeding ground for metaphorical thought. For this reason, the Wordlist function in Wordsmith Tools was used to obtain the frequency of the words in each sub-corpus. This is a crucial step because “no theory of metaphor can foresee which word forms will be used more
often metaphorically” (Sardinha, 2008, p. 128). Table 3 shows that some of the most frequent lexical words in the English corpus are nouns.

Table 4 shows the most frequent lexical words extracted from the Spanish corpus. Among them are agua [water], pez [fish], and mar [sea]. This is hardly surprising since all these words are the linguistic designations of basic concepts in the domain of marine biology.

Concordancing tools are the most usual way of studying a corpus for linguistic purposes (Deignan, 2005, p. 78). Accordingly, concordances were generated for the keyword pairs “water/agua”, “fish* / pez/peces”, and “sea/mar” by using the Concord function in Wordsmith Tools. The default option of this programme for the collocational horizons of the node (five words on the left and five words on the right of the node) was maintained. Once concordances were obtained, it was necessary to filter out obvious non-metaphorical occurrences since the frequency of the keywords in the corpus generated a lot of noise.

For enhanced data discrimination and candidate identification, the Show collocates function in Wordsmith Tools was used. This function yielded a list of collocates for each keyword in English and Spanish. This list includes the number of occurrences of collocates in each of the five positions on the left and right of the node. Figure 1 shows an English-language sample list of collocates, where harvest is highlighted. This collocate was regarded as a component of the potentially metaphorical term harvest fish because it occurred in the first position to the left of the node fish (L1), and does not originally belong to the field of marine biology, but rather to agriculture.

---

6Wordsmith Tools allows the user to append an asterisk (*) to (truncated) words to search for longer units.
Based on previous knowledge of the English terminology in the field, we also concordanced the form *fish, since complex terms can be written as separate words (multi-word items) or as compounds without spaces (see concordance pairs (1) and (2) below).

The Clusters option (with the default settings Words in cluster: 3 and Horizons: 5L-5R) was also useful for semi-automatic data discrimination and metaphor candidate identification. For example, whereas Spanish-language clusters such as desde el mar [from the sea] and mar filtrada a [sea filtered to] were discarded, the prepositional phrase de mar [from the sea] was found to be a frequent component of potentially metaphorical complex terms. A good example is pepinos de mar [sea cucumbers] in Figure 2. The Clusters function also indicates the number of occurrences of each cluster. This is very useful since frequency of occurrence usually provides evidence of the extent to which a linguistic metaphor has become conventional in a language. As can be seen, pepinos de mar occurred 16 times in the Spanish sub-corpus. The clusters de mar que [from the sea which] and de mar se [from the sea], also with 16 hits each, were also potential pointers to metaphor candidates.

Many English and Spanish resemblance metaphor terms were extracted from the corpus by these procedures. From a qualitative perspective, the concordance lines in both languages showed that the metaphorical terms in one sub-corpus had the same linguistic formalisation as metaphorical terms in the other sub-corpus (literal translation equivalents), which could well indicate that both languages share conceptual metaphors to designate the same specialised concepts. For this reason, they are regarded as exact pairs. Extended concordance pairs in (1) and (2) are given as examples:

(1) Significantly lower fungal growth rates on extracts of G. cf. mariae than G. ventalina support the hypothesis that G. cf. mariae is more chemically resistant to aspergillosis. When comparing sea fan disease across different regions of the Yucatan, we detected significantly
higher prevalence in *G. ventalina* near Akumal than further north near Cozumel and Puerto Morelos. (*Marine Biology, 149*(6), 2006, pp. 1355–1364)

Aunque en Colombia, se han llevado a cabo evaluaciones de la estructura de la comunidad de gorgonáceos en varias localidades (Botero, 1987; Sánchez, 1995, 1999; Sánchez et al., 1997), no existe información reciente sobre el estado y el desarrollo de las poblaciones del *abanico de mar* 15 años después de la mortalidad masiva. (*Boletín de Investigaciones Marinas y Costeras*, 35, 2006, pp. 77–90)

(2) Numerically and in biomass, the *lanternfish* is the dominant fish in the vast pelagic region of the Southern Ocean bounded on the north by the Antarctic Convergence and in the south by the Antarctic continental shelf. (*Marine Biology, 133*(1), 1999, pp. 145–158)

El método numérico determinó que la presa más representativa fue el *pez linterna* *Benthosema panamense* (38,0%; *n* = 305), seguido por el scianido *Larimus argenteus* (31, 3%; *n* = 252), *Sardinops sagax* (9, 2%; *n* = 74) y *Merluccius gayi* (9, 2%; *n* = 74; Table 1; *Revista de Biología Marina y Oceanografía*, 42(1), 2007, pp. 59–69)

The coral, *sea fan*, receives its names because its flat, wedge-like shape resembles that of a fan. The behaviour of the lanternfish is compared to the functioning of a lantern because of the photophore, a light-producing organ used by the lanternfish as both a defence and a mating strategy.

Even though target vocabulary retrieval shows great potential for metaphor candidate identification, it is not without its shortcomings. In fact, it is often criticised for capturing only subsets of metaphor occurrences since it only considers high-frequency lexical words. However, to avoid solely drawing on high-frequency words, we concordanced basic-level names, such as *crab*/cangrejo*, *shark*/tiburón*, and *dolphin*/delfín*. The advantages of this procedure have been highlighted in several studies (cf. Tissari, 2003; Stefanowitsch, 2004, 2006; Koivisto-Alanko & Tissari, 2006), but had not been previously documented in scientific discourse.
Stefanowitsch (2006) calls this strategy *metaphorical pattern analysis*. A metaphorical pattern is “a multi-word expression from a given source domain (SD) into which one more specific lexical item from a given target domain (TD) has been inserted” (Stefanowitsch, 2006, p. 66).

Many collocares of these entity names revealed potentially metaphorical terms designating types of crab, shark, and dolphin as shown in (3).

(3) Off Patagonian coasts, Argentina, the **dusky dolphin** is one of the most common small cetaceans. (*Marine Biology*, 152(1), 2007, pp. 165–177)

The dusky dolphin’s back is dark brown/black, a colour similar to that of dusk (the darker stage of twilight). Exact pairs were also found, as evidenced in the extended concordance shown in (4).

(4) Abstract Ninety four *scalloped hammerhead sharks*, *Sphyrna lewini* (53 females and 41 males) ranging in size from 121 to 321 cm total length (TL), were collected from surface gillnetters operating off northeastern Brazil and throughout the southwestern equatorial Atlantic Ocean between January and December 1996. (*Environmental Biology of Fishes*, 61, 2001, pp. 151–159)


The metaphorical motivation of *hammerhead shark*/*tiburón martillo* is evident since the head of this shark is has the same shape as a hammer.

The analysis of these concordances resulted in the extraction of units from many source domains. Since these units were indirectly signalled by target domain items, they were not used to find complex metaphorical candidates designating marine organisms. The terms from the land animal domain were found to be particularly salient in both languages, and exact pairs were even identified, such as (5) and (6):

(5) These species represent the most abundant scraping and excavating **parrotfish** species on inshore reefs. Grazing scars of each species were marked, measured and observed for seven consecutive days. (*Marine Biology*, 156(4), 2009, pp. 771–777)

**Dentro de la familia Scaridae** (*peces loro*), *Sparisoma radians* fue vista solamente en las estaciones SP. (*Boletín de Investigaciones Marinas y Costeras*, 37(1), 2008, pp. 111–127)


**M. gregaria** es una especie característica de las comunidades de fondos blandos, y junto al **cangrejo araña**, *Eurypodiodes lateillei*, son los decápodos mas importantes de la comunidad bentónica del golfo. (*Revista de Biología Marina y Oceanografía*, 42(3), 2007, pp. 221–229)

**Parrotfish/peces loro** are thus named for their dentition, which includes numerous teeth tightly arranged on the external surface of the jaw bones. The teeth form a parrot-like beak with which these fishes rasp algae from coral and other rocky substrates. **Spider crabs/cangrejos araña** resemble spiders because of their extremely long legs.
Extracting Resemblance Metaphors by Means of Lexical Markers

**Taxonomic designations.** The concordances and collocates showed taxonomic names (particularly species names) to be extremely productive lexical markers for metaphorical terms in English and Spanish. Taxonomic names were then collected from the co-texts of the previously extracted metaphorical terms. Taxonomic designations are used by experts to guarantee referential accuracy. However, as shall be seen, synonymy is quite common in specialised discourse.

Many English and Spanish metaphorical terms co-occurring with their corresponding taxonomic designations were found. These were generally names of sea organisms (especially fish and invertebrates). Apart from a desire for referential precision, another reason for this type of co-occurrence is that taxonomic designations strengthen the terminological status of their common names. The taxonomic designations associated with a metaphorical term in one language were concordanced in the sub-corpus of the other language to establish comparative interlinguistic patterns. This strategy permitted the identification of exact pairs, as shown by the extended concordances in (7):

(7) In order to study this enzyme in elasmobranchs, partial cDNA clones of 3β-HSD were isolated from steroidogenic tissue of the blacktip shark (*Carcharhinus limbatus*), **southern stingray** (*Dasyatis americana*), and the **spiny dogfish shark** (*Squalus acanthias*). The deduced amino acid sequences of the elasmobranchs were 40–44% identical to mammalian 3 β-HSD sequences, and approximately 50% identical to the **rainbow trout** (*Oncorhynchus mykiss*) form. (*Fish Physiology and Biochemistry, 19*(4), 2004, pp. 293–304)

Entre las especies de mayor valor económico de los Andes venezolanos, se encuentran la **trucha arcoiris** (*Oncorhynchus mykiss*), introducida en 1935 (Ginét al. 1952; *Revista de Biología Tropical, 46*(3), 1998, pp. 775–782)

The term **rainbow trout** was detected in the corpus by examining the co-text of **spiny dogfish shark**, which was in turn identified by retrieving expressions containing the basic-level keywords **fish** and **shark**. The metaphorical motivation of **rainbow trout** and its exact equivalent, **trucha arcoiris**, is colour since this fish has brightly colored longitudinal bands on each side of its body, similar to those in a rainbow.

As shown in the English text of concordance (7), the distributional behaviour of linguistic metaphors facilitated the identification of further metaphorical terms, which occurred in close proximity to the lexical marker *Oncorhynchus mykiss*. Metaphor distribution across texts has been assessed by various discourse researchers (e.g., Martin, 2006; Cameron, 2007). They found that linguistic metaphors are not evenly distributed, but rather occur in clusters in discourse (Cameron, 2007, p. 121). Also, the occurrence of a given metaphor increases the chances for this metaphor to be used again in the subsequent discourse (Martin, 2006). Still another example supporting this claim is given in (8).

(8) A small assemblage of reef fishes that included the **cherubfish**, *Centropyge argi*, **sunshine chromis**, *Chromis insolata*, **greenblotch parrotfish**, *Sparisoma atomarium*, **yellowcheek wrasse**, *Halichoeres cyanoccephalus*, **sargassum triggerfish**, *Xanthichthys ringens*, and the **longsnout butterflyfish**, *Chaetodon aculeatus* was most abundant or only present from stations deeper than 30 m, and thus appear to be indicator species of mesophotic habitats. (*Coral Reefs*, online first, 2010)
The repetition of one metaphor throughout a discourse unit can also be traced visually with the dispersion plot function in Wordsmith Tools, which graphically represents the distribution of a specific lexical item in a text. Figure 3 shows the spread of the term *hermit crab* (ascertained by concordancing the keyword *crab*) in each of the ten corpus articles where it occurred. This crab receives its name because it lives by itself inside its shell. This behaviour is compared to that of a hermit, who is a recluse and lives withdrawn from society in a solitary retreat.

The first article in the corpus (*Local, regional and global patterns of resource use in ecology: hermit crabs and gastropod shells as an example*) focused on this animal, which explains the 36 occurrences of *hermit crab*. The other articles had 1–4 occurrences.

However, a number of concordances also revealed a tendency to use metaphorical names only at the beginning of the articles, and rely on taxonomic names thereafter. This supports the existence of what Partington (2006, p. 268) calls *genre-specific metaphor*, which refers to the particular behaviour of metaphor in different genres or discourse types. Concordances 9 and 10 show the term *sea lettuce* (a type of alga that resembles a lettuce in shape and colour). In these texts, the authors either clearly state (concordance 10) or imply (concordance 9) that the name *Ulva lactuca* will be used in the remainder of their article.

(9) The influence of chemical deterrents on consumption does not appear to be a satisfactory explanation for the low consumption rate of the *sea lettuce*, Ulva, by Talitrus. Ulva was a minor biomass component (5%) on O Muiño beach and was poorly consumed by Talitrus.

(10) O Muiño beach receives inputs of the brown macroalgae, *Saccorhiza polyschides* (furbelow, hereafter Saccorhiza) and *Fucus spiralis* (wrack, hereafter Fucus) and the green alga, *Ulva lactuca* (*sea lettuce*, hereafter Ulva), as well as several species of red algae from nearby rocky intertidal and subtidal habitats. (*Marine Biology, 154*(1), 2008, pp. 163–174)

The presence of exact pairs in the corpus opened the door to the possible existence of two other types of interlinguistic term pairs: (i) those whose constituents were based on different metaphors; (ii) those in which only one constituent was metaphorical. Type (i) term pairs were regarded as *separate pairs*, and type (ii) as *unbalanced pairs*. These pairs were significant in
themselves because of the scarcity of studies addressing cross-linguistic variation of metaphor, especially in specialised language. However, it was impossible to attest the existence of separate pairs at this stage of our study since, as the quantitative analysis later revealed, this type did not occur as frequently as exact pairs. In contrast, we came across an unbalanced pair by concordancing taxonomic designations, as shown in (11).


Otros autores describieron la estructura de los folículos del ovario en *Urolophus jamaicensis* y *Dasyatis americana*, basada en la descripción de los tipos de células y tejidos que componen el trofonemata para estas especies, encontrando marcadas diferencias con otras como *Rhinoptera bonasus* y *Myliobatis* sp. Los hábitos reproductivos para *R. percellens* han sido mencionados sólo de manera descriptiva. (*Revista de Biología Marina y Oceanografía, 43*(3), 2008, pp. 469–481)

The cownose ray has a broad head with wide-set eyes, and a pair of creased lobes on its subrostral fin. It thus bears a definite resemblance to a cow’s nose and head.

**Scientists’ Surnames.** The analysis of taxonomic designations revealed that they usually co-occurred with the surnames of the scientists who first described the species. The surnames of two of the most prominent taxonomers, Carl Linnaeus and Georges Cuvier, were concordanced to search for further designations that did not occur in the co-text of the metaphorical terms previously found. Figure 4 highlights the metaphorical term *bib*, which occurred close to the lexical marker *Linnaeus*. The metaphorical motivation of this term is shape because it designates a fish whose lower jaw has a barbel, similar to a baby’s bib. Figure 6 also reveals the metaphorical term *boarfish* (concordances 53), thus named because of its projecting snout and bright red colouring.

Concordance (44) in Figure 4 is significant because it provides evidence of the existence of single-word metaphorical terms in marine biology. Consequently, it also shows the relevance of taxonomic designations as good lexical markers of the different linguistic manifestations of resemblance metaphor.
**Known as/conocido/a como.** The English phrase *known as* and its Spanish equivalent *conocido/a como* were concordanced. As lexical markers, they signalled the presence of a number of metaphorical terms in both languages. They were particularly useful as indications of metaphorical terms co-occurring with taxonomic designations that were not found earlier, and those metaphorical terms that did not contain any of the basic-level keywords (fish, *pez*, etc.). However, these markers turned out to be less productive than taxonomic designations and basic-level keywords, as shown in (12–14).

(12) In Europe, *Ostrea edulis*, the European flat oyster, and *Ostreola stentina*, also known as the Provence oyster or *dwarf oyster*, are both present along the European and African, Atlantic and Mediterranean, coasts. (*Marine Biology, 150*(1), 2006, pp. 103–110)

(13) Los ejemplares de Pomacentridae capturados en diferentes oportunidades en la costa de Valdivia (ca. 39° 48′ S; Fig. 1), suman seis ejemplares, que al ser analizados taxonómica-mente con claves especializadas y comparándolos con otros de colecciones de peces del Instituto de Zoología de la Universidad Austral de Chile, resultaron ser inequívocamente pertenecientes a *C. crusma* (Valenciennes, 1833), comúnmente conocida como “castañeta” [*castanet*] (Fig. 2; *Investigaciones Marinas, 33*(1), 2005, pp. 101–107)


The dwarf oyster is compared to a dwarf because of its small size. *Castañeta* [*castanet*] is a round fish that resembles a castanet, a round percussion musical instrument that is very common in Spain. Round shape is also the metaphorical motivation for the chelicerate *cacerolita de mar* [little sea saucepan], which as its name indicates, resembles a small saucepan.

The lexical markers *known as* and *conocido/a como* provided resemblance metaphor forms that differed from the terms previously found (i.e., terms designating sea organisms). These new terminological forms were nouns designating body parts of sea organisms, not the organisms themselves. A much less numerous set of body part metaphors were identified by means of *known as* and *conocido/a como*. Since they were not associated with any taxonomic designations, their semi-automatic retrieval clearly entailed a strategy that did not exclusively draw on the previously lexical markers⁷. Concordances (15) and (16) show two occurrences of body part metaphor terms found in the English and Spanish corpora by concordancing *known as* and *conocido/a como*.

(15) *Hippolyte obliquimanus* showed sexual dimorphism in the third pereopod, absence of coupling hooks (also known as *cincinnuli*) in the first pleopod, and no reduction of the appendix masculina in the largest males. (*Marine Biology, 154*, 2008, pp. 127–135)


---

⁷We came across a considerable number of resemblance metaphors designating body parts of marine organisms while exploring the co-text of taxonomic designations and the basic-level keywords. However, this is obviously not a (semi-)automatic procedure.
Finally, known as and conocido/a como showed cases of synonymy involving metaphorical common names in English and Spanish. Concordances (17) and (18) are two examples:

(17) Modern stomatopods (also known as *mantis shrimps*) are placed in 7 superfamilies with over 450 species that occur mostly in coral reef environments. (*Marine Biology*, 150(2), 2006, pp. 213–220)


The term *mantis shrimp* is striking because it has a double metaphorical motivation. The term does not designate either a terrestrial praying mantis or a shrimp, but rather a type of sea crustacean that shares several features with both these animals. It resembles a mantis in behaviour (aggressive and solitary) and in physical appearance (big, round eyes). It is also similar to a shrimp because it has an elongated body with multiple short legs. As for the term *lapa ocho*, it refers to a limpet whose shape is more ellipsoid than normal, and which resembles the number eight. The term *lapa de sol* is not metaphorical. It refers to the fact that this limpet usually clings onto rock clefts where it has access to sunlight.

One of the advantages of relying on corpus evidence is that “it shows instances of performance rather than competence” (Koller, 2006, p. 239). A signal of the performative condition of a corpus of specialised data is precisely the presence of terminological synonyms and variants, phenomena which have been documented, for instance, in medical texts (cf. Tercedor & Méndez, 2000). Thus, these two instances of synonymy could well point to further cases of terminological variation, as was finally the case in our study.

**Corpus Annotation and Quantitative Analysis**

Corpus annotation has been shown to yield optimal results concerning different aspects of metaphor description. Manual tagging is the most frequent type of corpus formatting. For instance, Semino (2006) annotates her corpus to describe a specific kind of metaphorical speech act annotation, whereas Sardinha (2008) annotates his corpus to compute the probabilities of each candidate word form being a vehicle metaphor.

Recent research relies on artificial intelligence, which offers automatic semantic field annotation tools, such as *UCREL Semantic Annotation System* (Koller, Hardie, Rayson & Semino, 2008; Hardie, Koller, Rayson & Semino, 2007), and metaphor extraction systems that exploit the codification of pre-defined semantic relations between units in lexical databases, as is the case of *CorMet* (Mason, 2004).

Although convincing results have been obtained, the approaches of these studies are not suitable for our research purposes. First of all, the procedure used in projects such as *CorMet* is only valid for verbs, whereas this study focuses on nouns. Secondly, in our view, semantic domain

---

8Synonymy does not concern morphosyntactic variation of terms. Although cases of this variation were also identified in the English sub-corpus (e.g. *sea-fan*/sea fan, *clownfish*/clown fish), they were not of interest for this study.

9*CorMet* identifies metaphors by finding systematic differences in *selectional preferences* between domains. A selectional preference is “a verb’s predilection for a particular type of argument in a particular role” (Mason, 2004, p. 23).
tagging by the UCREL Semantic Annotation System (USAS) could not be used in our study for various reasons. As previously shown, many of the marine biology metaphorical terms are compounds with neither spaces nor hyphens. These terms are generally made up of one word belonging to the source domain LAND ORGANISMS and one belonging to the target domain SEA ORGANISMS (e.g., wolfish). The only semantic domains offered by USAS for this type of metaphorical pattern are much too general (i.e., Life and living things, Living creatures: animals, and Plants). Consequently, potential source domain units would be hard to identify since both the target and the source have the same semantic tag in this classification. The same applies to compound terms whose source and target units belong to the domain of marine biology (e.g., whale shark) and aquatic life in general (crocodile fish). Finally, it is not clear how the USAS could be systematically exploited to reflect cross-linguistic differences of resemblance metaphors, something which is guaranteed by our tag set.

Our corpus was also manually annotated by using a tagging system that yielded quantitative results, which were both reliable and revealing for resemblance metaphor in marine biology. This type of analysis supplements the qualitative analysis, and provides statistical data regarding the use of metaphor in language, something that is generally not done in metaphor studies (Sardinha, 2008, p. 127). As shall be seen, the annotation system involves a tagset that meets the three criteria proposed by Leech (1997, pp. 25–26) for tag labels, which are: (i) conciseness (short tags better than long ones); (ii) perspicuity (tags must be as readable as possible); (iii) and analysability (it should be possible to decompose tags into their logical parts). The corpus was tagged and processed with Wordsmith Tools in order to obtain the absolute frequencies of metaphor occurrences and the number of English–Spanish term pairs in the corpus.

**Tagging to obtain absolute frequencies.** As a starting point, the concordance lines of the basic-level keywords “fish/pez (peces)”, “sea/de mar”, “crab/cangrejo”, “shark/tiburón”, and “dolphin/delfín”, as well as the lexical markers known as and conocido/a como that contained no metaphorical terms were manually tagged. Clean lists were then obtained by means of the Zap function in Wordsmith Tools, which filtered out the marked concordance lines. The metaphorical terms of the remaining concordance lines were marked with the tag METAPH in their corresponding .txt files in each sub-corpus. Those co-occurring with their taxonomic designations were tagged METAPH TAXO, and were subsequently concordanced (e.g., Figure 6).

Given that many academic journal articles in both languages contained checklists of the species being described, these checklists helped to retrieve further resemblance metaphor terms in English and Spanish and statistically refined the cross-linguistic account. Not surprisingly, most checklists in research articles dispense with common names as a way to avoid ambiguity. A case in point is Figure 5, which shows a checklist extracted from an article of the Spanish-language journal Boletín de Investigaciones Marinas y Costeras.

More often than not, the checklists included all of the taxonomic designations quoted in the text body of their corresponding articles. Those taxonomic designations from the checklists that were not previously detected were concordanced and cleaned to search for metaphorical terms containing neither of the previously mentioned basic-level keywords (e.g. staghorn coral, camarón café). These concordances were tagged METAPH TAXO.

The annotation process revealed cases in which taxonomic designations co-occurred with common names. However, they were not figurative in nature. These terms were tagged TAXO.
<table>
<thead>
<tr>
<th>Especie</th>
<th>Sin Protección</th>
<th>Especie</th>
<th>PNNT</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Acanthurus bahianus</strong></td>
<td>100</td>
<td><strong>Acanthurus bahianus</strong></td>
<td>100</td>
</tr>
<tr>
<td><strong>Bodianus rufus</strong></td>
<td>100</td>
<td><strong>Bodianus rufus</strong></td>
<td>100</td>
</tr>
<tr>
<td><strong>Chromis multilineata</strong></td>
<td>100</td>
<td><strong>Elacatinus illecebrosus</strong></td>
<td>100</td>
</tr>
<tr>
<td><strong>Haemulon flavolineatum</strong></td>
<td>100</td>
<td><strong>Haemulon flavolineatum</strong></td>
<td>100</td>
</tr>
<tr>
<td><strong>Myripristis jacobus</strong></td>
<td>100</td>
<td><strong>Haemulon plumierii</strong></td>
<td>100</td>
</tr>
<tr>
<td><strong>Sparisoma viride</strong></td>
<td>100</td>
<td><strong>Sparisoma viride</strong></td>
<td>100</td>
</tr>
<tr>
<td><strong>Cephalopholis cruentata</strong></td>
<td>98</td>
<td><strong>Canthigaster rostrata</strong></td>
<td>94</td>
</tr>
<tr>
<td><strong>Chaetodon sedentarius</strong></td>
<td>98</td>
<td><strong>Lutjanus mahogoni</strong></td>
<td>94</td>
</tr>
<tr>
<td><strong>Halichoerus garnoti</strong></td>
<td>98</td>
<td><strong>Myripristis jacobus</strong></td>
<td>94</td>
</tr>
<tr>
<td><strong>Pseudopleuron macrocephalus</strong></td>
<td>98</td>
<td><strong>Sparisoma aurofrenatum</strong></td>
<td>94</td>
</tr>
<tr>
<td><strong>Sparisoma aurofrenatum</strong></td>
<td>98</td>
<td><strong>Gnatholepis thompsoni</strong></td>
<td>92</td>
</tr>
<tr>
<td><strong>Chromis cyanca</strong></td>
<td>95</td>
<td><strong>Halichoeres garnoti</strong></td>
<td>92</td>
</tr>
<tr>
<td><strong>Elacatinus illecebrosus</strong></td>
<td>95</td>
<td><strong>Mullolidichthys martinicus</strong></td>
<td>92</td>
</tr>
<tr>
<td><strong>Thalassoma bifasciatum</strong></td>
<td>95</td>
<td><strong>Thalassoma bifasciatum</strong></td>
<td>92</td>
</tr>
<tr>
<td><strong>Acanthurus coeruleus</strong></td>
<td>93</td>
<td><strong>Acanthurus coeruleus</strong></td>
<td>89</td>
</tr>
<tr>
<td><strong>Gnatholepis thompsoni</strong></td>
<td>93</td>
<td><strong>Chaetodon capistratus</strong></td>
<td>89</td>
</tr>
<tr>
<td><strong>Mullolidichthys martinicus</strong></td>
<td>93</td>
<td><strong>Chromis cyanca</strong></td>
<td>89</td>
</tr>
<tr>
<td><strong>Stegastes partitus</strong></td>
<td>93</td>
<td><strong>Chromis multilineata</strong></td>
<td>89</td>
</tr>
<tr>
<td><strong>Canthigaster rostrata</strong></td>
<td>91</td>
<td><strong>Haemulon chrysargyreum</strong></td>
<td>89</td>
</tr>
<tr>
<td><strong>Holacanthus tricolor</strong></td>
<td>91</td>
<td><strong>Pseudopleuron macrocephalus</strong></td>
<td>89</td>
</tr>
</tbody>
</table>

**FIGURE 5** Checklist of taxonomic designations of marine organisms that includes no (metaphorical) common names.

**FIGURE 6** Extended concordance lines of METAPH from the English corpus.
(see examples further below). The concordances of Linnaeus and Cuvier were of help to find further taxonomic designations co-occurring with metaphorical terms. They were tagged METAPH TAXO.

Not only did the markers known as and conocido/a como help to find instances of terminological variation, but they also revealed other cases in the annotation process while checking the taxonomic designations against the different articles where they occurred (listed by Wordsmith Tools) and by examining their co-text. Those variants that were metaphorical were tagged METAPH TAXO, and those with a literal meaning, TAXO. Two cases of synonymy were found in the English corpus (apart from cloud sponges/sponge gardens and Provence oyster/dwarf oyster). The first case involves the taxonomic designation Galeorhinus galeus, which co-occurred in the corpus with the terms soupfin shark—tagged TAXO because of its non-figurative nature—tope shark, and school sharks—these latter tagged METAPH TAXO because of their metaphorical nature. The second case concerns the taxonomic designation Carcharias taurus, which co-occurred with the terms sand tiger shark and grey nurse—both tagged METAPH TAXO—as well as with raggedtooth shark, tagged TAXO.

Apart from lapa de sol/lapa ocho, we identified four cases of synonymy in the Spanish texts. The first entails the designation Anadara tuberculosa, which occurred in close proximity to the literal common names concha prieta [tight conch], concha negra [black conch], and piangua [no literal translation into English]. All three common names were tagged TAXO. Limulus polyphemus also co-occurred with the metaphorical name pata de mula [mule’s leg], which was tagged METAPH TAXO. The second case involves the designation Limulus polyphemus, which co-occurred with cangrejo cacerola [saucepan crab] and cacerolita de mar [little sea saucepan]. Both these common names were tagged METAPH TAXO. The common name variants of Dermochelys coriacea, tortuga baula [no literal translation into English] and tortuga laúd [lute turtle], were tagged TAXO and METAPH TAXO, respectively. Finally, the taxonomic designation Lepidochelys kempi was found in the Spanish corpus together with tortuga lora [female parrot turtle] and tortuga carpintera [carpenter turtle], which were tagged METAPH TAXO.

Once all occurrences of resemblance metaphor terms were tagged in English and Spanish (including both repeated units and hapaxes\(^\text{10}\)), they were computed. The absolute frequency of individual items was then calculated for each language and for both languages. The results are provided in Table 5.

<table>
<thead>
<tr>
<th></th>
<th>Number of Tokens</th>
<th>Number of Metaphoric Occurrences</th>
<th>% of Metaphoric Terms in Each Sub-Corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>English corpus</td>
<td>2,240,327</td>
<td>16,920</td>
<td>0.75</td>
</tr>
<tr>
<td>Spanish corpus</td>
<td>2,309,863</td>
<td>15,327</td>
<td>0.66</td>
</tr>
<tr>
<td>Overall</td>
<td>4,550,190</td>
<td>32,247</td>
<td>0.70</td>
</tr>
</tbody>
</table>

\(^\text{10}\)A hapax (legomenon) is a lexical item that occurs only once in a whole unit of analysis (in this case, the marine biology corpus).
As can be seen, the overall percentage of metaphorical terms is 0.70, which is not very high. However, it is sufficient to prove the incidence of metaphorical thought in scientific discourse. It should also be pointed out that this percentage only applies to one type of metaphor, resemblance metaphor. The analysis of the data also revealed a significant number of non-resemblance metaphors, which was hardly surprising since it has been shown that genre-specific texts typically feature complex metaphorical chains, which convey the underlying conceptual map of a particular discourse (Koller, 2003, p. 243). Nevertheless, most of the lexical markers used in this study could not be used to systematically detect these metaphors in the corpus. The same applies to body-part resemblance metaphors, the other type of resemblance metaphor found in our corpus.

The statistics showed that the Spanish corpus had more tokens than the English one, probably because the rhetoric style of Spanish is often wordier than English. However, the likelihood of encountering metaphorical terms was found to be higher in English because this language allows the repetition of the same lexical unit in close proximity in a text. In contrast, such repetition is generally avoided in Spanish. Moreover, metaphorical names for sea creatures are also more frequent in English than in Spanish. This fact was justified by the subsequent calculation and analysis of interlinguistic term pairs.

**Tagging to obtain interlinguistic term pairs.** Comparable English–Spanish results from quantitative data were obtained by the following procedure. Concerning exact pairs, since the taxonomic designations are the same in both languages, a separate corpus was created that included the English and Spanish texts, and preserved the tags used up to that point. This was done with a view to finding exact pairs more easily. The tag **METAPH TAXO** was concordanced, focusing on those terms that contained neither of the basic-level keywords. Exact pairs were identified by searching for every taxonomic designation across the concordance lines with the search function, and comparing interlinguistic strings. Whenever an exact pair was identified, the tag **METAPH TAXO EXA** was added to both interlinguistic equivalents.

The same procedure was used to find term pairs containing one of the basic-level keywords (i.e., “fish”/*pez/*peces”, etc.). A search was made in the corpus of the other language for a literal translation for each basic-level metaphorical term with no taxonomic designation. This procedure had already been followed for certain terms at the beginning of the study to check for exact pairs. When there was a hit, the tag **METAPH EXA** was added to the English term. Subsequently, the concordance entries of **METAPH EXA** and **METAPH TAXO EXA** were added up. Since it was a question of computing term pairs rather than tokens, the figure obtained was divided by two to calculate the number of pairs (see Table 10).

The analysis of the co-text of the taxonomic designations not only generated exact pairs, but also separate and unbalanced pairs, which were tagged **METAPH TAXO SEP** and **(METAPH) TAXO UNB**, respectively. Each tag was then concordanced separately for computing. The tagging process thus confirmed the existence of separate pairs in the corpus. Figure 7 shows some of the separate term pairs found.

Table 6 shows the aligned pairs.

---

11The terms poblaciones [populations] and comunidad [community] found in the Spanish text of concordance (1) are metaphors that do not arise from resemblance but from different patterns of comparison.
The identification of unbalanced pairs in the corpus showed two sub-types of pairs:

(i) those in which only the English term was metaphorical, as previously shown in concordance (11);

(ii) those in which only the Spanish term is metaphorical.

For calculation purposes, the metaphorical terms of the pairs were tagged `METAPH TAXO UNB`, and the non-metaphorical ones, `TAXO UNB`. Furthermore, in both sub-types, the non-metaphorical constituent either had a common name or only the taxonomic designation was used. To retrieve all unbalanced pairs, the tag `TAXO UNB` was concordanced. Figure 8 shows
most of the interlinguistic pairs found in the corpus in which the non-metaphorical constituents have a common name associated to their taxonomic designations. Tables 7 and 8 contain the aligned pairs.

The pair *grunt-corocoro* is interesting. The term *grunt* is a case of behaviour-based metaphor whose metaphorical motivation is sound. It refers to a fish that is named so for its ability to produce sound by grinding their teeth as though it were grunting. This is thus a type of behaviour.

Figure 9 shows some of the 12 pairs found in the corpus in which a taxonomic designation co-occurred with a metaphorical term in the English texts, whereas the same taxonomic designation did not occur with any common name on all occasions that it appeared in the Spanish texts. No instances were found in which the inverse occurred. Nevertheless, the fact that a taxonomic designation does not occur with a common name in this corpus does not necessarily mean that experts use a common name for such taxonomic designation in other texts.

Table 9 contains the aligned interlinguistic pairs from Figure 9.

### TABLE 7

<table>
<thead>
<tr>
<th>Concordance Line</th>
<th>English Metaphorical Term</th>
<th>Spanish Non-Metaphorical Name [Literal Translation into English]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 and 17</td>
<td>School shark</td>
<td>Cazón [-]</td>
</tr>
<tr>
<td>4 and 18</td>
<td>Milkfish</td>
<td>Sabalote [-]</td>
</tr>
<tr>
<td>5 and 7</td>
<td>King crab</td>
<td>Centolla [-]</td>
</tr>
<tr>
<td>8 and 16</td>
<td>Rose shrimp</td>
<td>Gamba blanca [white shrimp]</td>
</tr>
<tr>
<td>9 and 14</td>
<td>Olive ridley turtle</td>
<td>Tortuga golfina [gulf turtle]</td>
</tr>
<tr>
<td>11 and 20</td>
<td>Grunt</td>
<td>Corococo [-]</td>
</tr>
<tr>
<td>13 and 15</td>
<td>Queen conch</td>
<td>Caracol rosado [pink snail]</td>
</tr>
</tbody>
</table>
TABLE 8
Aligned Unbalanced Pairs Including Spanish Metaphorical Names and English Literal Common Names

<table>
<thead>
<tr>
<th>Concordance Line</th>
<th>Spanish Metaphorical Name [LITERAL TRANSLATION INTO ENGLISH]</th>
<th>English Non-metaphorical Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 and 22</td>
<td>Camarón café [coffee shrimp]</td>
<td>Brown shrimp</td>
</tr>
<tr>
<td>3 and 6</td>
<td>Lobo fino [slender wolf]</td>
<td>Fur seal</td>
</tr>
<tr>
<td>10 and 21</td>
<td>Tiburón peregrino [pilgrim shark]</td>
<td>Basking shark</td>
</tr>
<tr>
<td>12 and 19</td>
<td>Pez luna [moon fish]</td>
<td>Sunfish</td>
</tr>
</tbody>
</table>

FIGURE 9  Concordance lines of unbalanced pairs including metaphorical terms in English and only the taxonomic designations in Spanish.

TABLE 9
Aligned Unbalanced Pairs Including Metaphorical Terms in English and Only the Taxonomic Designations in Spanish

<table>
<thead>
<tr>
<th>Concordance Line</th>
<th>English Metaphorical Term</th>
<th>Taxonomic Designation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 and 8</td>
<td>Bullseye stingray</td>
<td>Dasyatis brevis</td>
</tr>
<tr>
<td>2 and 11</td>
<td>Cownose ray</td>
<td>Rhinoptera steindachneri</td>
</tr>
<tr>
<td>3 and 13</td>
<td>Ribbed mussel</td>
<td>Aulacomya ater</td>
</tr>
<tr>
<td>4 and 15</td>
<td>Sea gooseberry</td>
<td>Pleurobranchia bachei</td>
</tr>
<tr>
<td>5 and 7</td>
<td>Fan mussel</td>
<td>Pinna nobilis</td>
</tr>
<tr>
<td>6 and 10</td>
<td>Moon jelly</td>
<td>Aurelia aurita</td>
</tr>
<tr>
<td>9 and 17</td>
<td>Dwarf sperm whale</td>
<td>Kogia simus</td>
</tr>
<tr>
<td>12 and 16</td>
<td>Spanish hogfish</td>
<td>Bodianus rufus</td>
</tr>
<tr>
<td>14 and 18</td>
<td>Spanish dogfish</td>
<td>Squalus acanthias</td>
</tr>
</tbody>
</table>
Table 10 shows the figures of the interlinguistic pairs extracted from the corpus. The figures show that English and Spanish often conceptualise marine organisms in the same way. However, considering the influence that English as a *lingua franca* exerts on the rest of languages in international scientific contexts, many Spanish metaphorical names might be the literal translations of their English counterparts. As far as unbalanced pairs are concerned, the figures suggest that while English language experts tend to give marine organisms a metaphorical name, Spanish language experts often either use a literal common name or simply use the taxonomic name.

The *METAPH TAXO* concordances revealed that a considerable number of taxonomic designations and their corresponding metaphorical terms were only quoted in one language, something which stands to reason, since we were dealing with a comparative rather than a parallel corpus. Consequently, this study provides a tentative, though illustrative, picture of the percentage of exact, separate, and unbalanced resemblance metaphor pairs of the designations for marine organisms.

The specification of the motivations for metaphorical transfer is essential for a clear classification of resemblance metaphors. From a cross-linguistic perspective, this specification should also be quantified. For this reason, the distribution of the five types of metaphorical motivation found in this study across the term pairs was numerically accounted for. The number of pair terms according to their metaphorical motivation is given in Table 11.

The data showed that both English-language and Spanish-language experts most frequently use shape to conceptualise sea organisms. Combinations of behaviour and shape are the least frequent in both languages because they feature complex patterns, which cannot always be combined in the metaphorical conceptualisation. The data also showed that the metaphorical motivations are evenly distributed across the two languages, with a higher number of terms in English because of the scarcity of Spanish metaphorical terms in the unbalanced pairs.

Determining the Metaphorical Nature of Terms

Metaphor identification is often understated in current research, and this applies to specialised language as well (Caballero, 2006, p. 65). This can lead to somewhat unreliable results since
metaphor identification is far from easy. This identification process can be approached from a symbolic or behavioural perspective (Steen, 2007). In this regard, our analysis is limited to the symbolic aspect of metaphor, specifically to its linguistic dimension. We have not dealt with the behavioural side because this would involve addressing the psychological processes of metaphorical thought, which is contingent on the type of user.

From a linguistic perspective, studies of terminological metaphor are mostly based on intuition and random inference to determine the metaphorical nature of terms. However, there are strategies that can be used to avoid subjective choices. We propose three strategies that aim to provide objective evidence of metaphorical usage. The first strategy is the use of a three-step method similar to the Metaphor Identification Procedure (MIP; Pragglejaz Group, 2007), which tests the potentially metaphorical meaning of a lexical unit in a particular context against its non-metaphorical meaning in other contexts. In our study, we identified the motivation for metaphorical transfer of the marine biology terms by determining whether the metaphor candidate had a more basic contemporary meaning in contexts other than the domain-specific one. The basic meaning of the metaphor candidate was then compared with its meaning in the specialised context. If the comparison of these meanings gave rise to semantic tension and such semantic tension led to referential incongruity, then this was regarded as indicative of a case of terminological metaphor.

As previously mentioned, this study relied on authentic data, and searched standardised observational patterns in a corpus. However, when it came to determining the figurative meaning of specialised language units, a dictionary was used as a norm of reference in order to have an independent reflection of what counted as the meanings of words for a particular group of users (Steen, 2007, p. 97). The expressions in our study were evidently included in dictionaries and other types of reference book. The reference sources used for the English terms were the general language dictionary American Heritage® Dictionary of the English Language, the field guide Sharks of the World, and the on-line marine biology database SeaLifeBase13. The resources consulted for the Spanish terms were the general language dictionary, Diccionario de la Real Academia Española®, and the specialised field catalogues, Guía de Identificación de los Peces Marinos de Europa y del Mediterráneo and Guía Submarina de Invertebrados no Artrópodos. The metaphorical term escorpión marino [sea scorpion] is given as an example.

---

13http://www.sealifebase.org
First of all, we determined the basic meaning (description) of escorpión, which, according to the Diccionario de la Real Academia Española, is the following:

Arácnido con tráqueas en forma de bolsas y abdomen que se prolonga en una cola formada por seis segmentos y terminada en un aguijón curvo y venenoso [a tail composed of six segments and tipped by a curve and venomous sting].

We then compared the basic meaning of escorpión with the meaning of the multiword unit escorpión marino, which is provided in the Guía de Identificación de los Peces Marinos de Europa y del Mediterráneo. The meaning of escorpión marino is:

(Miembro de la familia) Cottidae. Cabeza con espinas y aristas óseas bien visibles en la nuca [...] Las espinas de la cabeza pueden ser venenosas durante el período de reproducción [the spines on its head may be venomous during the reproduction period].

The comparison of both definitions reflects their semantic tension, which gives rise to the referential incongruity of the term escorpión marino, and thus underlines its metaphorical motivation. Sea scorpion was classified as a behaviour-based metaphor since the sea scorpion’s dorsal, anal, and pelvic spines have a venom gland, and thus cause the same effect as a scorpion’s sting.

In those cases in which the definition of a term or the description of the marine organism included no clear indicators of the possible motivation for metaphorical transfer, it was analysed by either exploring the linguistic context of the term in the academic article(s) in search of an explicit or implicit explanation or by looking at pictures of the marine organism.

Context analysis was performed by using the View Text option in Concord of Wordsmith Tools. This function displays a co-text of 400 words for single concordance lines, as shown in (19).

(19) The bioluminescent pelagic squaloid shark, Isistius brasiliensis (Figure 1), has acquired the common name of ‘cookie-cutter shark’ because of its very unusual feeding style, that apparently involves extracting cookie-shaped plugs of flesh from fast swimming, larger predators. (Environmental Biology of Fishes, 53, 1998, pp. 267–273)

Identifying the metaphorical motivation of this metaphor clearly illustrates the claim that some resemblance metaphors are based on both physical comparison and behavioural comparison. This shark resembles a cookie-cutter because it cleanly bites mouthfuls of tissue from its victims in the same way as a cookie cutter separates pieces of dough from a larger mass. Thus, there is a complex mental image involving shape (image metaphor) plus the fact that the shark acts in the same was as a kitchen utensil (behaviour-based metaphor).

Example (20) gives an implicit explanation:

(20) Urechis caupo (the fat innkeeper worm) is an echiuran which inhabits a U-shaped burrow in both intertidal and subtidal mudflats at three major population sites on the northern coast of California. (Marine Biology, 113, 1992, pp. 613–623)

Of course, corpus data also helped to discover the non-metaphorical nature of other terms. For instance, the term mosquito fish may mislead the reader into thinking that this animal looks and/or behaves like a mosquito. Example (21) implicitly specifies that the fish receives this name simply because mosquitoes are part of its diet:

(21) Eastern mosquito fish Gambusia holbrooki (Girard) and western mosquito fish Gambusia Gambia affinis (Baird & Girard) have been artificially propagated throughout the world to
FIGURE 10  Shamefaced crab (Calappa granulata) (SeaLifeBase).

When the contextual analysis was not sufficient, it was also necessary to examine images of the marine organisms. Many pictures are available in the electronic database consulted as well as on the Google image search engine. Figure 10 shows a shamefaced crab (Calappa granulata). The dark reddish patches on this crab’s shell, which resemble the red face of a person who feels ashamed, reveal the reason why this animal receives its name.

By applying these strategies, we were able to test the metaphorical nature of marine biology terms against empirical data. What the MIP calls the contextual meaning and the basic meaning of a lexical unit was determined in this study by using dictionary definitions and encyclopaedias, contextual data extracted from the corpus, and visual evidence from an online marine biology database.

CONCLUSION

Of the four strategies traditionally used to semi-automatically retrieve metaphorical units in corpora, we searched for target domain vocabulary as well as for lexical markers to detect terminological metaphors in a specialised corpus of marine biology texts in English and Spanish. The research study described in this article shows this method to be extremely productive and reliable. It detected resemblance metaphor terms in both languages as well as interlinguistic term pairs.

This qualitative analysis was afterwards reinforced by a quantitative analysis, which was performed after the corpus was manually annotated. The quantitative results showed that our annotation system is more time-efficient than manual searching, and confirmed the potential of combining the abovementioned strategies to discover metaphorical terms. The quantitative analysis complemented the qualitative results since it revealed interlinguistic term pairs that had not been detected previously. It also provided valuable statistical information regarding cross-linguistic terminological metaphor, an area in which such data are virtually non-existent.

Nouns, the grammatical form analysed, turned out to be well represented in the corpus, which corroborates Goatly’s (1997, p. 84) observation that nouns, referring directly to things [physical entities], can more directly evoke images than other parts of speech since the meaning of nouns...
is conceptualized as bundles of semantic features. The corpus analysis led to the following conclusions:

- Although the percentage of resemblance metaphors in the corpus is not high, it is sufficiently indicative of the great potential of metaphor to model scientific thought both in English and Spanish.
- The statistics showed that English is more prone to conceptualising marine organisms metaphorically than Spanish. This fact was confirmed by the numerical data regarding the interlinguistic term pairs retrieved.
- Three types of interlinguistic term pairs were extracted: (i) exact pairs (both terms based on the same conceptual metaphor); (ii) separate pairs (the two terms based on different conceptual metaphors); and (iii) unbalanced pairs (only one of the terms is metaphorical). Exact pairs were the most frequent type, which signifies that scientists in both languages tend to conceptualise marine biology entities in the same way.
- Empirical evidence was found of the existence of intralinguistic synonyms both in English and Spanish. This disproves the frequent assumption that scientific discourse is exempt of synonymy and terminological variation.

The advantages of the two strategies used for semi-automatic term retrieval not only lie in their potential to yield reliable results, but also in the fact that the size of the corpus can be increased without these strategies losing their effectiveness.

Finally, a method based on three strategies was proposed to verify the metaphorical nature of term candidates, detected by the previous analyses. The first strategy was an adaptation of the MIP (Pragglejaz Group, 2007) to the domain of marine biology. It involved the consultation of dictionaries and other resources containing encyclopedic information. The second strategy entailed the analysis of contextual data extracted from the corpus, and the third involved the analysis of visual images from an online marine biology database and from the Google search engine.

This methodology revealed five types of motivation for metaphorical transfer: shape, colour, size, behaviour, and combinations of behaviour and shape. This latter type involves a group of resemblance metaphors that reside in a transition zone between image metaphors and behaviour-based metaphors. The quantitative analysis showed that the metaphorical motivations are evenly distributed across the two languages, with a higher number of terms in English because of the scarcity of Spanish metaphorical terms in the unbalanced pairs.

REFERENCES


