The Metaphor in Human Genome Discourse[1]

Míriam Suárez Gallo
Instituto de Ciencia Animal (ICA)
msuarez@ica.co.cu

Míriam Suárez Gallo treballa a l'Instituto de Ciencia Animal (ICA), centre de recerca de la República de Cuba, com a editora de la *Revista Cubana de Ciencia Agrícola = Cuban Journal of Agricultural Science*. L'any 2019 s'ha doctorat en lingüística a la Universitat Pompeu Fabra, amb la tesi *La metáfora en el discurso especializado del genoma humano*, dins del grup IULATERM de l'Institut de Lingüística Aplicada (IULA) i sota la direcció de Mercè Lorente.

Resum

La metàfora en el discurs del genoma humà

Les metàfores d'origen terminològic implicades en la creació de termes i les metàfores útils per explicar a experts i a no experts com funciona el sistema genètic s'han identificat i classificat segons els supòsits de la teoria de la metàfora conceptual de Lakoff i Johnson (1980) i des de la perspectiva de la teoria comunicativa de la terminologia (Cabré, 1999), d'acord amb els pressupòsits de la lingüística de corpus (Parodi, 2007 i 2008). Les unitats analitzades s'han extret d'un corpus de genoma humà, compost per diversos textos de diferents nivells d'especialització. S'ha aplicat una eina d'extracció automàtica de terminologia per seleccionar els termes més rellevants basats en metàfores. A més, també s'ha utilitzat un sistema d'anàlisi sintàctica amb l'objectiu de cercar contextos que subministressin metàfores explicatives. Es va concloure que les metàfores d'origen lèxic i les metàfores explicatives mostren la mateixa tendència cap a la selecció d'accions i estats (o atributs). Ambdós tipus coincideixen, bàsicament, en la metàfora militar (freqüència baixa) i en la metàfora de les ciències del llenguatge (la més abundant).

PARAULES CLAU: metàfora terminològica; metàfora explicativa; genoma humà

Abstract

Metaphors of terminological origin involved in term creation, and metaphors useful for explaining to experts and/or non-experts how the genetics system works, were identified and classified according to assumptions based on the conceptual metaphor theory of Lakoff and Johnson (1980) and on the communicative terminology theory (Cabré, 1999), and also considering corpus linguistics (Parodi, 2007, 2008). The metaphors were taken from a textual corpus composed of several texts of different specialization level. A terminology extraction tool was applied to select the most relevant metaphor-based terms. In addition, a syntax analysis system was applied with the goal of searching contexts that supply explicative metaphors. It was concluded that metaphors of lexical origin and explicative metaphors show the same trend towards the selection of actions and states (or attributes). Both agree, basically, on the military metaphor (low frequency) and on the metaphor of language sciences (the most common).

KEYWORDS: *terminological metaphor; explicative metaphor; human genome*

TERMINÀLIA 20 (2019): 30-38 · DOI: 10.2436/20.2503.01.139
Data de recepció: 1/10/2019. Data d'acceptació: 15/11/2019
ISSN: 2013-6692 (impressa); 2013-6706 (electrònica) · http://terminalia.iec.cat

1 Introduction

It is well known that the evolution that has taken place in the manner of appreciating and understanding the metaphor, from the merely ornamental to the conceptual, and from the artistic to the domain of daily life, began within the context of incipient cognitive linguistics, in the papers taking a cognitive approach by Lakoff and Johnson. Based on linguistic evidence, Lakoff and Johnson discovered that most of our ordinary conceptual system is of metaphorical nature. For them, metaphor is not only a matter of language. They understand that, conversely, human thought processes are mostly metaphorical and that, "as linguistic expressions, metaphors are possible because they are metaphors in the conceptual system of a person" (Lakoff and Johnson, 1980, p. 6).

Considering this background, this study starts with the idea that metaphor, due to its cognitive potentials, is vital to the shaping of thought, be it daily and trivial or strictly scientific. It is understood that metaphor participates in the formation of knowledge from the most varied specialties, that it is involved in its consolidation, that it is essential for its transmission, and that it thus "results in a mechanism of conceptualization of extreme importance: due to its epistemology potential to open new modes and pathways of thought, and because, by evoking daily experience domains, [it] constitutes an effective communicative resource for the explanation and exposition of scientific contents to different types of audiences" (Ciapuscio, 2011, p. 4).

In the genetics domain, for instance, the code metaphor was established by molecular biologists and, from this starting point, a terminology was established which has as its basis the concept of information: "el ADN se replica en la división celular, se transcribe de ADN a ARN, y se traduce de ARN a proteínas, el ADN porta información que **aparece** encriptada, y es partiendo de ella que se construye la vida. Esta molécula contiene un secreto, es el libro de la vida, el código de los códigos." ["DNA replicates in cell division, it is transcribed from DNA to RNA, and it is translated from RNA to proteins. DNA bears information that is encrypted, and it is on this basis that life is constructed. This molecule contains a secret; it is the book of life, the code of codes."] (Sentís, 2004, p. 199).

As a working hypothesis, it is established that metaphor participates in the setting-up and spread of scientific knowledge as a resource of ***lexical creation***. A significant proportion of terms from diverse knowledge areas are denominations of metaphorical character. Furthermore, metaphor is also used to ***explain*** to semi-experts or non-experts what happens in certain fields of knowledge. Scientific knowledge is, in many instances, of great complexity and abstraction; thus, an expert from any given field makes use of metaphor quite regularly, since it is a resource that makes it possible to explain to semi-experts or non-experts abstract phenomena starting out from more concrete and better-known phenomena.

This research seeks to identify and classify, from a small human genome corpus, metaphors of terminological origin which are involved in term creation, and metaphors that serve to explain to experts and/or non-experts how the genetics system works. We seek to establish, *grosso modo*, trends in structures that lead to terminological and explicative metaphors.

In this study, metaphor of terminological creation is understood as metaphor which serves to create a term and which is internal to the terminological unit (TU). This type of metaphor may coincide with a simple term or may be marked in one of the components of the complex terms. A terminological metaphor is thus interpreted as a lexical unit derived from or being the product of term creation from phrases that, when used in a specialized context, undergo some modifications in their original meaning, reflecting the activation of certain specialized semantic traits. *Mapa*, *mapa cromosómico* [*map*, *chromosomal map*] are examples of terminological metaphors in the genome domain.

Explicative metaphor is metaphor that is used to explain to experts and/or non-experts how the genetic system works. It is expressed through sentence sequences that relate a term of the genome domain functioning as a subject with a verb in active voice or through sentence sequences in which the genome terms are subjects of copulative or pseudo-copulative sentences, or mid constructions.

A TU or term is a specialized unit of meaning (SUM), whose structure corresponds to a lexical unit derived from or which is the product of term creation from a phrase, which has a specific meaning in the domain to which it is related, being necessary in the conceptual structure of the domain of which it forms part (Cabré and Estopà, 2005, p. 10).

As to its morphology, a simple or mono-lexical TU is a lexical unit consisting of only one morpheme (*gen* [*gene*]).

A complex TU is a lexical unit that is formed by more than one morpheme (*cromosoma* [*chromosome*]). Complex terms can be composed of a combination of words that follow certain syntax structure (*brazo cromosómico* [*chromosome arm*]).

2 Theoretical framework

This research takes into account the significant contribution of Lakoff and Johnson (1980) with respect to the conception of metaphor as a cognitive mechanism. Metaphor is also addressed from the Communicative Terminology Theory (CTT) elaborated by Cabré (1999), who proposes, from linguistics, an approach to terminological units in their context of use, namely, in specialized communication.

By recommending the terminological units as part of the general language and sharing the same characteristics as any other lexical unit, the CTT states that terms and words are not different entities but lexical units created with the same resources of formation, which have the same performance at the syntactic level. Therefore, due to these units' interdisciplinary character, it holds that metaphor is a process of lexicon creation and, consequently, of specialized jargon. Moreover, it is understood as a linguistic cognitive device that contributes to the organization and building of specialized knowledge.

On referring to how the human mind understands objects and, by abstraction, builds concepts, Cabré (1993, p. 96) alludes to the three basic dimensions that are joined in one term: cognitive, linguistic, and communicative. As noted, a term has a cognitive component because it represents knowledge; it has at the same time a linguistic component because it is a unit of natural language; and it is of communicative character because it transmits specific information, hence intervening in the communicative act like any other lexical unit, conditions which mean that a term may also be, in no few instances, the result of a metaphor-based process. These assertions denote a more comprehensive account of metaphor in terminology which, according to the author, must be, describe and explain a cognitive, linguistic and sociocultural phenomenon.

In linguistics-oriented terminology, terms are addressed from their conditions of discourse combination, assuming the conceptual and denominative variation to which they are subject. On this point, the opinion of Cabré (2002, p. 89) is quite clarifying and, on the basis of CTT, it shares procedural cognitive models of textual linguistics:

> El enfoque de esta propuesta no tendría ningún sentido sin una concepción discursiva y textual de la terminología. Los términos constituyen en esta propuesta unidades insertas en el discurso, entendido el discurso en un sentido amplio: el texto y el contexto, o dicho de otro modo, el texto y sus condiciones de producción, transmisión y recepción.
>
> Los pilares en los que se fundamenta nuestra línea de trabajo son los siguientes: Por un lado, la lingüística cognitiva en sus postulados más generales [...]. Por otro lado, la lingüística textual, en una concepción discursiva y dinámica [...]. En tercer lugar nos situamos en una aproximación comunicativa a la terminología.
>
> ["This proposal's approach would be meaningless without a discursive textual conception of terminology. In this proposal, terms constitute units inserted into the discourse, understanding discourse in a broad sense: the text and the context, or in other words, the text and its conditions of production, transmission and reception.
>
> The pillars on which our line of work rests are as follows: On the one hand, cognitive linguistics in its most general postulates [...]. And on the other, textual linguistics, in a discursive dynamic conception [...]. In the third place we situate ourselves in a communicative approach to terminology."]

In accordance with Cabré (2002), the analysis of linguistic expressions based on metaphors with terminological value in the field of the human genome can only be possible from a discursive and textual conception of terminology. As reported by this author, in this proposal terms constitute units inserted into the discourse, understanding discourse in a broad sense: the text and the context, or in other words, the text and its conditions of production, transmission and reception.

Corpus linguistics foundations are also considered (Parodi, 2007, 2008) as a methodological approach allowing the study of languages in use and, thus, making possible the description and analysis of all kinds of discourse. From this approach, original and complete linguistic information is analyzed from linguistic corpora supported in computational technology and *ad hoc* computer programs.

3 Methodology

3.1 Materials, method and procedure of analysis

Textual corpus from approximative results obtained through manual introspection into corpus concordances (Suárez, 2009), two texts of different levels of specialization about the human genome were selected from the Technical Corpus of IULA.

The Technical Corpus of IULA (University Institute of Applied Linguistics at Pompeu Fabra University) compiles texts written in five different languages (Catalan, Spanish, English, French, and German) from specialty areas such as economy, law, environment, medicine, and informatics. It also comprises parallel documents that facilitate translation studies. At the same time, this multilingual corpus includes a subcorpus of general language extracted from mass media and built as a contrastive corpus.

The purpose of the Technical Corpus of IULA is to facilitate linguistic data analysis with the aim of being able to establish rules of performance for each language in each area. Its users are researchers and anyone who needs to consult the specialty fields under consideration. By using this corpus, studies of terminology, discourse, morphology, syntax, neologism, or translation have been carried out. In order to make it easier to use the data, IULA has developed a series of exploration tools, such as an automated extractor of neologisms, an automated detector of terminology, a text aligner, and a dictionary feeder (Cabré and Bach, 2004, p. 173). The consultation of the IULA corpus

can be performed on Internet through BwanaNet, an interface developed at IULA. Consultations can be performed for almost all the corpus or certain subcorpora.

Since its conception, for thematic organization reasons, text collection on the human genome has been conceived as a subcorpus that is part of the corpus of medicine texts. This collection is composed of 276 documents written in Spanish, comprising a total of 1,649,844 words. These are texts produced by experts on medical sciences, especially in the genetics field, and other disciplines related to medicine, such as biotechnology, genetic engineering, pharmacology, and biochemistry. Texts written by experts on humanities are also included for disciplines such as anthropology and philosophy. The dates of publication of these materials correspond to the period 1990-2011.

3.2 Identification of terminological metaphors

In order to identify the terminological metaphors in the selected texts, the automated extraction of the terms was conducted with the YATE terminology extractor software.

YATE is a system of extraction of candidates of nominal terms in medicine texts which have been previously processed with the tools of the Technical Corpus of IULA. By means of this extractor software application, a list may be obtained and ordered with term candidates through a combination of outcomes from different extraction methods, namely, a series of modules (context analysis, educated learners, among others). They are the ones that analyze the candidates by using different techniques.[2]

YATE also uses semantic information in the extraction processes. The semantic information is obtained by means of the EuroWord Net (EWN) database. Although EWN includes, primarily, the vocabulary of general language, it also has a large number of entries from medicine and other specialized fields. This database shows basic semantic relations between words for different European languages (Dutch, Italian, Spanish). Each of the languages ("wordnets") is connected to WordNet 1.5 in American English and to an index of meanings (Interlingual Index or ILI). It also has a common ontology, so the specific traits of each language are kept in the various wordnets individually. This database is used in applications to retrieve multilingual information, with the aim of enhancing information retrieval by expanding the keywords from any user to a broader set of related variants and words in any of the interconnected languages.

The outcome YATE offers is, for each of the patterns considered (N, NA, NPN), a list of term candidates, ordered according to the degree in which a lexical unit represents a term that belongs to the domain under consideration.

When analyzing the contexts in which these simple or mono-lexical terms appear, it was proved that some of the metaphor-based mono-lexical terms, such as *código* [*code*] or *brazo* [*arm*], are variants of terminological units (TU) of syntactic character, in which the complement is elided (elision conditioned by the context) to simplify the expression. This observation helps to confirm that cases like these are reductions of syntactic units that appear also in the list of NA terms that YATE provides. It is just a matter of economizing the language, leading to a process of word formation of these units in their simplest form. Also, units were found to be performing like a single reduction in one context while not fulfilling this criterion in another, because the complement that comes with them tells them apart. An example of this is *map* which, in some contexts evidently refers to the genome map, while in other contexts it differs since in the genome domain there are different types of *map* (*mapa cromosómico*, *mapa molecular*, *mapa genético*, *mapa de ligamiento*, *mapa físico* [*chromosomal map*, *molecular map*, *genetic map*, *ligament map*, *physical map*]). Consequently, *map* cannot be determined to be a syntactic reduction of the phrase *mapa genético* [*genetic map*] every time we come across it.

Moreover, it was observed that, in the corpus exploration, some mono-lexical TUs of metaphorical origin, such as *brazo* [*arm*], constitute the basis for the formation of other combinations with a specialized sense, in which this unit from the field of anatomy is modified by units of specialized sense that belong to the genome domain.

As to the poly-lexical TUs of NA structure created by metaphor, in order to determine if a phrase has become a term or not, the adjective was analyzed, considering the criteria of Estopà, Lorente and Folguerà (2002). It is assumed that related adjectives (noun-derived), of the type *genético*, *cromosómico*, *proteico* [*genetic*, *chromosomal*, *protein*] facilitate term formation from the phrase, mainly if the nucleus of the phrase is a common noun, which becomes specialized thanks to the addition of the specialized relational adjective. Conversely, qualifying adjectives, which are gradual, tend to be an obstacle to word formation from a phrase. However, it is not assumed that there are no syntactic terms with AN structure that may be built with qualifying adjectives which form new terms, therefore creating TUs, but two conditions must be fulfilled for this purpose: 1) the base must be a TU, and 2) the qualifying adjective must take part in the construction of a TU of AN structure, being of a type or kind from another TU (N). In such cases, the terms of AN structure tend to have cohyponyms, for example, *virus*, *virus dócil*, *virus salvaje*, *virus-misil*, *virus-stop* [*virus*, *docile virus*, *wild virus*, *virus-missile*, *virus-stop*].

In this process of terminological metaphor identification through YATE, the consultation of papers on terminological fixation was also carried out to prove, in some instance, the word formation of specific units.

In order to establish the classification of these terminological metaphors into a certain metaphorical

field, attention was given to the extralinguistic reality that serves as the origin domain to set each of the identified units. For a better organization of the metaphorical fields, two groups have been established, taking into account the domains (specialized or non-specialized) with which each field is associated:

1) The fields that take different specialty domains as a reference.
2) The fields that present genome constituents as beings with specific qualities or that can perform diverse functions.

3.3 Identification of explicative metaphors

In order to conduct the automated detection of explicative metaphors, a search strategy was created from the design of another type of patterns that combine grammatical concepts with selected lexical units that may serve as tool for the expression of metaphors or that may accompany them, and the TUs of the genome domain.

Treebank was used to implement this consultation. It is a syntactic analyzer for carrying out searches on the part of the IULA corpus that is syntactically recorded.[3] In order to question the treebank, the most simple terms, the most frequent in the arrangements provided by YATE, were selected as the first search terms, whether they were metaphorical or not (*brazo*, *cromosoma*, *gen* [*arm*, *chromosome*, *gene*]) or the nucleus of syntactic terms, both choices depending on the subject. The adjective or the noun of the prepositional complement was chosen as a second search term.

The treebank questioning showed that there is a general metaphor of action which can be subclassified: cognitive actions (or of direction), communicative actions, and others. Moreover, a second general metaphor was confirmed, expressing state or quality. Out of these observations, two sentence scenarios were determined in which explicative metaphors are created:

1) Sentences with verbs of action in active voice, in which the genome terms perform actions proper to humans or not.
2) Sentences in which the genome terms are subjects of copulative, pseudo-copulative, or middle constructions.

The object of establishing these subdivisions was to have a more organized classification of metaphors in the sentences retrieved through the treebank.

4 Results and discussion

4.1 The metaphors of terminological creation

On reviewing the output of mono-lexical terms, NA terms, and NPN terms provided by YATE, it was proved that the metaphors of terminological creation, which are inherent to the terminological unit, may coincide with the simple term or with one of the components of the complex terms:

— Marked in the term (*brazo*, *código* [*arm*, *code*]).
— Marked in the nucleus of the terminological syntax (***brazo*** *cromosómico*, ***bloqueo*** *genético*, ***bloqueo*** *metabólico*, ***cartografía*** *genética*, ***código*** *genético*, ***lenguaje*** *químico*, ***mapa*** *bacteriano*, ***mapa*** *genético*, ***transcripción*** *génica*, ***hardware*** *genético*, ***alfabeto*** *genético*, ***diccionario*** *genético*, ***edición*** *genética*, ***expresión*** *genética*, ***mensaje*** *genético*, ***expresión*** *fenotípica*, ***cartografía*** *de cromosomas*, ***código*** *de triplete*, ***expresión*** *de gen*, ***mapa*** *de cromosoma*, ***mapa*** *de genes*, ***mapa*** *de baliza*, ***resumen*** *de proteína* [*chromosomal* ***arm***, *genetic* ***blockade***, *metabolic* ***blockade***, *genetic* ***cartography***, *genetic* ***code***, *chemical* ***language***, *bacterial* ***map***, *genetic* ***map***, *genetic* ***transcription***, *genetic* ***hardware***, *genetic* ***alphabet***, *genetic* ***dictionary***, *genetic* ***edition***, *genetic* ***expression***, *genetic* ***message***, *phenotype* ***expression***, ***cartography*** *of chromosomes*, ***code*** *of triplet*, *gene* ***expression***, ***map*** *of chromosome*, ***map*** *of genes*, ***map*** *of marker*, ***summary*** *of protein*]).
— Marked in the modifier of the terminological syntax (*células* ***frescas***, *fenotipo* ***silvestre***, *alelo* ***silvestre***, *parásito* ***invasor***, *genes* ***nómadas***, *virus* ***dócil***, *virus* ***salvaje***, *virus-***misil** [***fresh*** *cells*, ***wild*** *phenotype*, ***wild*** *allele*, ***invasive*** *parasite*, ***nomad*** *genes*, ***docile*** *virus*, ***wild*** *virus*, *virus-***missile**]).

As to the classification of these terminological metaphors into a specific metaphorical field, regarding the extralinguistic reality that serves as a source domain for these denominations, units of the MILITARY domain were identified, whose frequency in the human genome domain derives from identifying some of the genome constituents with entities that take part in a military scenario. *Bloqueo genético*, *bloqueo metabólico*, *parásito invasor*, *virus-misil* [*genetic blockade*, *metabolic blockade*, *invasive parasite*, *virus-missile*] are poly-lexical units whose lexical base corresponds to a term coming from the military domain or to a unit specific to the genome.

The backgrounds of Lakoff and Johnson are revisited here. The identification of these terminological units permits the confirmation that, although in the genes area there is no battle with military weapons, there is indeed a fight from the verbal standpoint. Specific denominations, such as *bloqueo genético* [*genetic blockade*], though not making reference to real war, are thought and described making use of military or warlike terms. And this happens because these events, and the entities involved, have been internalized in the mind of those giving them their names as if there were a military conflict. *Brazo*, *brazo cromosómico* [*arm*, *chromosomal arm*] constitute terminological metaphors that are the result of conceptualizing chromosomes as humans. The term *brazo* [*arm*], which comes from the ANATOMY domain, and names one of the limbs of the human body, when inserted in the genome context, denotes by analogy each of the halves of the chromatids in the chromosomes.

Terminological metaphors like this one, that take the human body as a source domain, are found among

the so-called anthropomorphic metaphors, being noted by Boquera (2005) in her study on metaphor in civil engineering discourse, where this type of metaphor is more widespread, unlike what has been seen in the genome discourse, in which metaphors having specialty domains as their source, such as LANGUAGE SCIENCES, seem to be more productive.

When referring to anthropomorphic metaphors, Boquera (2005, p. 251) determines that there are several types: anatomic, formed by reference to human body parts (*hormigonado de riñones*, *machacadora de mandíbulas* [*concreted kidneys*, *jaw crusher*]); physiological, making reference to body functions or processes (*el curado del hormigón*, *el agotamiento de una estructura* [*hardened concrete*, *structure exhaustion*]); relational, formed by reference to human relations (*morteros bastardos* [*bastard mortars*]); sensorial, alluding to smell, sight, taste, hearing and touch (*pozo ciego* [*blind well*]); metaphors with reference to typically human actions (*abrazar*, *comer*, *descansar* [*hug*, *eat*, *rest*]) and metaphors expressing human qualities (*vigas esbeltas*, *roca sana* [*slender beams*, *healthy rock*]). As noted before, although anthropomorphic metaphors alluding to the human body do not seem to be the most productive among the terminological metaphors registered in the genome discourse, there were indeed, among the explicative metaphors, examples referring to human actions (*la proteína activa reconoce su diana*, *las células T examinan las moléculas*, *las proteínas dirigen todas las funciones celulares* ["the active protein recognizes its target, T cells examine the molecules, proteins direct all cell functions"]).

In the human genome area, the term *mapa* [*map*] designates the diagram of the positions of the genes, considering the genome as a large territory susceptible to be mapped.[4] *Map* has been the basis for the formation of multiple combinations, in which this unit is modified by units of specialized sense common in the genome domain or by prepositional complements that had as their nucleus a term from this field. *Mapa cromosómico*, *mapa genético*, *mapa bacteriano*, *mapa de baliza* [*chromosomal map*, *genetic map*, *bacterial map*, *map of marker*] constitute terminological metaphors that correspond to poly-lexical units that are part of the CARTOGRAPHY field. Besides, this field is composed of *cartografía cromosómica*, *cartografía genética*, *cartografía de cromosomas* [*chromosomal cartography*, *genetic cartography*, *cartography of chromosomes*].

Conceiving the genome constituents as participants in the processes of linguistic and cognitive character accounts for the metaphorical origin of the units *transcripción génica*, *lenguaje químico*, *alfabeto genético*, *diccionario genético*, *edición genética*, *expresión fenotípica*, *expresión genética*, *resumen de proteínas*, *expresión de gen* [*genetic transcription*, *chemical language*, *genetic alphabet*, *genetic dictionary*, *genetic editing*, *phenotypical expression*, *genetic expression*, *protein summary*, *gene expression*], which make up the metaphorical field of the LANGUAGE SCIENCES. These terminological metaphors of poly-lexical structure have at their base a term related to the language domain, modified by a unit of specialized sense from the genome area.

The DNA molecule, understood as a file containing information, has inside a *code* that is different from the Morse code, a penal code, or a bar code. This *code*, when deciphered, provides access to the written message containing the secret of life. As an expression of this conceptualization, terms characteristic to the information area (*código*, *mensaje* [*code*, *message*]) are taken as a reference to denominate DNA constituents. These units are also the basis of poly-lexical forms (*código genético*, *código de triplete*, *mensaje genético* [*genetic code*, *triplet code*, *genetic message*]) which have a specialized sense by combining with units of the genome domain. *Código*, *mensaje*, *código genético*, *código de triplete* [*code*, *message*, *genetic code*, *triplet code*] comprise the metaphorical field of INFORMATION.[5]

The terms created through metaphors identified in this study that are grouped in the metaphorical fields of CARTOGRAPHY and LANGUAGE SCIENCES confirm the criteria of Temmerman (2000) as to the formation of new terms from metaphorical cognitive models. The model of information, recognized by this author, encompasses the denominations that come from conceptualizing the genome as a space represented in a map containing information on the position of each of the components that form the human genome, and also from conceptualizing the genes as containers of messages that carry information.

The identification of the domains taken as source for setting these terminological metaphors shows that these are mostly, as noted by Humbley (2009), disciplinary fields that turn distant from the human genome area, as in the case of LANGUAGE SCIENCES, INFORMATION, and CARTOGRAPHY, although, in an exceptional manner, there were terminological metaphors not conceived from such distant sources, as those built by taking the ANATOMY domain as a source.

On this matter of recognition of the origin or source domain, if the classification of Lakoff and Núñez (2000, cit. by Humbley, 2009) were applied, it could be said that the terminological metaphors identified are the ones these authors denominate as "earth-connected metaphors", because they relate the genome field to other distant fields taken as a source for setting these denominations.

The terminological metaphors *alelo* ***silvestre***, *fenotipo* ***silvestre***, *células* ***frescas***, *virus dócil* [***wild*** *allele*, ***wild*** *phenotype*, ***fresh*** *cells*, ***docile*** *virus*] form the metaphorical field of QUALITIES. In these structures, which differ from the rest of those under analysis, the term of the genome domain constitutes the lexical base of the denomination, and receives the modification of a common unit, which grammatically corresponds to a qualifying adjective, in which the metaphor is marked.

Besides these metaphorical fields, there were examples of terminological metaphors that could belong to other thematic groups that have not been classi-

fied yet, until completing the analysis with the entire available corpus. The terminological metaphors from this unclassified group express ACTION (*enzima cortadora* [*cutting enzyme*]) and SHAPES (*filamento molecular*, *filamento proteico* [*molecular filament*, *protein filament*]).

It is noteworthy that by trying to identify these terminological metaphors through the automated extraction of terms, a quite recurrent scenario has emerged (term + adjective) which, while awaiting the completion of our analysis, must be provisionally explained as noise: *gen responsable*, *molécula responsable*, *enzima responsable*, *mutación responsable* [*responsible gene*, *responsible molecule*, *responsible enzyme*, *responsible mutation*]. This is so because, although this type of metaphor may always appear accompanied by a complement with the preposition *de* [*for*] (*el gen responsable de la mutación...* ["the gene responsible for the mutation..."]), lexical cases would not be involved and, hence, such complements would be, strictly speaking, terminological units.

4.2 Explicative metaphors formed through sentence sequences combining subject and verb

According to the methodology established, the questioning of the treebank allowed the retrieval of the following sentences:

1) Sentences with verbs of action in active voice, in which the genome terms do things which are characteristic of humans or non-predictable:
   Movement actions:
   *"En otras palabras, una **célula** tumoral invasora debe simultáneamente **perforar** un túnel, **agarrarse** a las paredes de ese túnel y **autopropulsarse** hacia delante."*
   ["In other words, an invasive tumoral **cell** shall simultaneously **perforate** a tunnel, **hang on** to the walls of that tunnel, and **self-propel** forward."]
   Cognitive actions:
   *"Las **células** eucarióticas han desarrollado sistemas complejos de señalización intercelular que les permiten **evaluar** las condiciones ambientales y responder en consonancia."*
   ["Eukaryote **cells** have developed complex intracellular signalings permitting them to **evaluate** environmental conditions, and respond to them in return."]
   *"La **proteína** activa **reconoce** su diana."*
   ["The active **protein recognizes** its target."]
   *"Al merodear por los tejidos, **las células T examinan** las moléculas del MHC y los antígenos que allí encuentran."*
   ["When around tissues, **T cells examine** the MHC molecules and the antigens found there."]
   *"Las **proteínas dirigen** todas las funciones celulares."*
   ["**Proteins direct** all cell functions."]
   Communicative actions or of 'dictum':
   *"Los **genes codifican** enzimas capaces de inactivar a determinados fármacos."*
   ["**Genes encode** enzymes capable of making inactive certain drugs."]
   *"El **gen** de la molécula de ácido desoxirribonucleico (ADN), que dirige la síntesis hormonal, **transcribe** una molécula de ácido ribonucleico mensajero (ARN)."*
   ["The **gene** of the molecule of deoxyribonucleic acid (DNA), which directs hormone synthesis, **transcribes** a molecule of messenger ribonucleic acid (RNA)."]
   *"Pero las **células** también **tienen que comunicarse** a través de distancias mayores que aquéllas que pueden facilitar las cadenas de contactos intercelulares."*
   ["But **cells** also **need to communicate** over distances longer than those that may facilitate the intercellular contact chains."]
   *"Las **células expresarán** entonces las proteínas deseadas."*
   ["**Cells will express** thus the desired proteins."]
   *"Quizá los **genes** nos lo **dirán**."*
   ["Maybe **genes** will **tell** us."]
   Military actions:
   *"Las **células** activadas, ya alertadas sobre la presencia de células tumorales, circularían por todo el organismo y **atacarían** a otros tumores".*
   ["The activated cells, being warned of the presence of tumoral cells, would circulate throughout the body and **would attack** other tumors."]
2) Sentences with copulative verbs (to be), whose attributes express certain properties characteristic of the terminological subject:
   *"Las **células están organizadas** en compartimentos."*
   ["**Cells are organized** in compartments."]
   *"Las **células están conectadas** para suicidarse."*
   ["**Cells are connected** to commit suicide."]
   *"Las **células** determinadas **están programadas** para seguir un proceso de desarrollo que acaba conduciendo a su diferenciación."*
   ["The determined **cells are programmed** to follow a development process that leads to their differentiation in the end."]
   *"Las **células** progenitoras y las proliferativas **son células en reciclaje**."*
   ["Progenitor and proliferating **cells** are **cells under recycling**."]
   *"Ambos **genes están reprimidos** por lexa, debido a la presencia de una caja SOS en la región intergénica."*
   ["Both **genes are repressed** by lexa, due to the presence of an SOS-box in the inter-gene region."]
3) Copulative sentences that are based on explicative metaphors through verbs that attribute qualities:
   *"El **código es** ordenado."*
   ["The **code** is ordered."]
   *"Las **mutaciones son** errores de copia que el sistema de verificación y reparación del ADN han pasado por alto."*
   ["**Mutations are** copying errors that the DNA verification and repair system have missed."]
   *"Las **células** del mono **son** permisivas para SV40, y las **células** de ratón **son** permisivas para polioma."*
   ["Monkey **cells are** permissive to SV40, and mouse **cells are** permissive to polyoma."]

The identification of the explicative metaphors permits, in specific contexts, the detection of the conceptualization of the genome constituents as live entities, which are capable of moving and performing actions. The verbs *self-propel*, *recognize*, *communicate*, and *attack* mark this conceptualization. From this perspective, the explicative metaphors retrieved from the human genome texts were similar to those mentioned by Vandaele and Lubin (2009) when detecting metaphorical expressions in the descriptive anatomy discourse by the so-called verbal conceptualization indices.

In agreement with the proposal of Lorente (2007, p. 378) about the classification of verbs in specialized discourse, phraseological verbs were identified in the explicative metaphors retrieved from genome texts, specifically those referring to actions of movement-related, cognitive, communicative and military character. In the sentences analyzed, this type of verb creates explicative metaphors through the governing relation established with a nominal phrase that corresponds to a relevant term of the genome domain.

The copulative verbs (2) and the verbs that attribute qualities (3), which create metaphors connecting a nominal phrase functioning as a subject with attributes, are in correspondence with those called "verbs of logical relation" by Lorente.

The explicative metaphors, or at least those corresponding to the predication of terminological subjects that were extracted from the treebank, evidenced syntactic diversity (verbs in active voice, middle voice and attributes) and also thematic diversity (unclassified actions, movements, cognitive actions, communicative actions, states, and attribution of qualities, among others).

5 Conclusions

The metaphors of lexical origin and explicative metaphors showed the same trends in the selection of actions and states (or attributes).

With regard to thematic distribution, there was, basically, coincidence in both for the warlike metaphors (less frequent) and the linguistic and communicative metaphors (the most common).

On the basis of the available data for this analysis, it may be stated that neither the metaphors of lexical or terminological origin nor the explicative metaphors showed dissimilar solutions for the diverse levels of specialization.

The outcomes of this research on metaphor in texts about the human genome area are a contribution to the description of specialized discourses, specifically with respect to metaphor involvement in this type of discourse.

Although this study is based on a quite limited corpus, the outcomes presented help to determine certain trends and generalizations as to the presence of terminological metaphors in the creation of terms and metaphors that serve to explain to experts and/or non-experts how the genetic system works.

References

Boquera, María (2005). *Las metáforas en textos de ingeniería civil: estudio contrastivo español-inglés*. PhD thesis. Universidad de Valencia.

Cabré, M. Teresa (1993). *La terminología: Teoría, metodología y aplicaciones*. Barcelona: Antártida Empúries.

Cabré, M. Teresa (1999). *La terminología: Representación y comunicación*. Barcelona: Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada.

Cabré, M. Teresa (2002). "Análisis textual y terminología, factores de activación de la competencia cognitiva en la traducción". In: Alcina, A.; Gomero, S. (eds.). *La traducción científico-técnica y la terminología en la sociedad de la información*. Castelló: Publicacions de la Universitat Jaume I, p. 87-105.

Cabré, M. Teresa; Bach, Carme (2004). "El Corpus Tècnic del IULA: corpus textual especializado plurilingüe" ("The technical corpus of IULA: plurilingual specialized textual corpus"). *PANACE@*, 16, p. 173-176.

Cabré, M. Teresa; Estopà, Rosa (2005). "Unidades de conocimiento especializado, caracterización y tipología". In: Cabré, M. Teresa; Bach, Carme (eds.). *Coneixement, llenguatge i discurs especialitzat* ("Knowledge, language and specialized discourse"). Barcelona: Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada. (Sèrie Monografies; 7), p. 69-93.

Ciapuscio, G. Elena (2011). "De metáforas durmientes, endurecidas y nómades: un enfoque lingüístico de las metáforas en la comunicación de la ciencia". *ARBOR: Revista de Ciencia, Pensamiento y Cultura*, 747, p. 89-97.

Estopà, Rosa; Lorente, Mercè; Folguerà, Rosa-Anna (2002). "El rol de los adjetivos en los textos especializados". In: *VIII Simposio Iberoamericano de Terminología: La terminología: entre la globalización y la localización* (Cartagena, Colombia, 28-31 October). [CD-ROM ed.]

HUMBLEY, John (2009). "La nature de la métaphore dans une technologie émergente". In: DURY, Pascaline; MANIEZ, François (eds.). *La métaphore en langues de spécialité.* In: *Actes des Journées du CRTT* (Lyon, 26 September 2008). Grenoble: Presses Universitaires de Grenoble.

JARA, Carla (2013). "El *treebank* del español IPROCOLDI: componente anotado del corpus CODIMEP-CR". *Revista de Filología y Lingüística de la Universidad de Costa Rica*, 39 (2), p. 145-147.

LAKOFF, George; JOHNSON, Mark. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

LORENTE, Mercè (2007). "Les unitats lèxiques verbals dels textos especialitzats. Redefinició d'una proposta de classificació". In: LORENTE, Mercè; ESTOPÀ, Rosa; FREIXA, Judit; MARTÍ, Jaume; TEBÉ, Carles (eds.). *Estudis de lingüística i de lingüística aplicada en honor de M. Teresa Cabré Castellví*. Vol. II. Barcelona: Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada. (Sèrie Monografies; 11), p. 365-380.

NIVRE, Joakim (2008). Treebanks. In: LÜDELING, A.; KYTÖ, M. (eds.). *Corpus Linguistics: An International Handbook*. Vol. 1. Berlin: Germany: Walter de Gruyter, p. 225-241.

PARODI, Giovanni (ed.) (2007). *Lingüística de corpus y discursos especializados: puntos de mira*. Valparaíso: Ediciones Universitarias de Valparaíso.

PARODI, Giovanni (2008). "Lingüística de corpus: una introducción al ámbito". *Revista de Lingüística Teórica y Aplicada*, 46 (1), p. 93-119.

SENTÍS, Carlos (2004). "El ADN de las metáforas a los hechos y a la biotecnología". *Inguruak: Revista vasca de sociología y ciencia política*, 40, p. 199-220.

SUÁREZ, Míriam (2009). "La metáfora en la construcción del discurso especializado y en la divulgación científica. Análisis en el discurso del genoma humano". PhD thesis project. Barcelona: Universitat Pompeu Fabra.

TEMMERMAN, Rita (2000). *Towards new ways of terminology description: The socio-cognitive approach*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

VANDAELE, Sylvie; LUBIN, Leslie (2009). "Modes de conceptualisation et représentations fictives en anatomie". In: DURY, Pascaline; MANIEZ, François; ARLIN, Nathalie; ROUGEMONT, Claire (eds.). *La métaphore en langues de spécialité*. Grenoble: Presses Universitaires de Grenoble, p. 61-81.

VIVALDI, Jorge (2003). *Sistema de extracción de candidatos a término YATE: Manual de utilización*. Barcelona: Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada. (Papers de l'IULA. Sèrie Informes; 43)

Notas

1. This paper forms part of the PhD thesis of Míriam Suarez, under the tutorship of Prof. Lorente, focused on the study of metaphor and its importance in the processes of conceptualization, consolidation and dissemination of specialized knowledge in the human genome domain. The thesis deals with the presence of metaphor as a source of lexical formation in the knowledge creation process, and as a vehicle for explaining and transmitting scientific knowledge. Specifically, its aim is to determine whether metaphors of lexical creation and explicative metaphors are created by means of the same linguistic structures in the genome discourse, and whether the specialization level of discourse does or does not affect the selection of metaphors.

2. For more information, see Vivaldi (2003)

3. Treebanks are collections of examples of parsed sentences that serve to build statistical analyzers. Geoffrey Leech coined the term *treebank* to refer to a morpho-syntactically recorded corpus, in particular due to the fact that a very common manner of representing the syntactic structure is through tree diagrams (Jara, 2013, p. 145). However, at present, according to Nivre (2008), "a *treebank* is not necessarily represented by means of trees, since practices have been diversified in that respect, although the recording scheme will always be determined by the syntactic theory from which the analysis is performed." According to this author, it is common today to follow a model as general as possible or otherwise eclectic models are set taking aspects from dissimilar theories.
Treebanks are useful to study several lexical, morphological, and syntactic phenomena. When referring to their usefulness, Jara (2013, p. 145) notes that in computational linguistics they serve to develop, test or train automated or semi-automated analyzers. Applied to diachronic corpora, they allow a more productive study of linguistic change. This author adds that a treebank may account for the frequency of certain structures (how common they are in real use) and their spread (what new syntactic phenomena emerge and what communicative needs they cover).

4. Although *map* is usually seen as a non-specialized unit due to its so very common use, it is actually a basic TU in the cartography domain. It is understood in this way here.

5. The field of INFORMATION is established here to group the metaphorical TUs that make reference to some of the elements involved in the information theory (*message*, *receptor*, *code*, *program*, among others), with the aim of setting apart these units from those coming from the LANGUAGE SCIENCES field (*translation*, *transcription*, *reading*).