

On the elements activating the transmission of specialized knowledge in verbs

Anna Joan Casademont

In this paper we propose an analysis of phraseological verbs in specialized discourse in order to establish syntactic and semantic characteristics likely to indicate the transmission of specialized knowledge. By adopting and adapting theoretical and descriptive proposals existing for general language, we consider elements from the relation between the verb and its arguments (such as argument structure and thematic roles) and aspects from the arguments themselves (such as semantic features and termhood). Through the detection of combined features and behaviour patterns of occurrences, we are able to more specifically establish which characteristics allow us to recognise the terminological use of phraseological verbs and which kind of information is the more relevant to its conveyance.

Keywords: terminology; verb; argument structure; syntax; semantics; specialized discourse

1. Verbs and the expression of specialized knowledge¹

When considering parts of speech, it is well established that units transmitting terminological value are prototypically nouns (Rey 1979, 1993; Sager 1990; Cabré 1999), since their eminently referential features allow them to be used in a very direct way when representing the conceptual nuclei of specialized knowledge. So unsurprisingly, terminological dictionaries include nominal terms such as *chromosome*, *stock share*, *decree*, *molecule* or *ozone layer*.

Verbs, by their relational condition, would then seem to be less subject than nouns to present terminological values on their own. Moreover, except for some cases where a specialized use is morphologically visible in the formation of a verbal unit

(for example from the verbalization of a noun term, such as *crystallize* from *crystal* in Physics) (Lorente 2001, 2007), the possible specialized variational behaviours of a verb (i.e. *produce*) would seem more difficult to grasp unless its interaction with the surrounding text is considered.

The motivation behind this research is to provide some insights into the mentioned interaction of the verb with its surrounding text: Which context elements are the most important to fully express the verb meaning? Do they interact with each other or solely with the verb? Do they all have the same influence on the expressed meaning?

We start this paper by presenting our hypotheses and the main previous work existing for terminology and verbs, some of which being a starting point of this study. We then proceed to outline our methodology: the selection of the corpus, the structure of the created database and the analysis of data. In the latter, we focus on analyzing features of the relation between a verb and its arguments, features of the arguments themselves and features linking verb realizations and the meanings expressed by them. We then centre on the results obtained for the specialized uses of verbs and, in the conclusions, we go back to the hypotheses in order to treat them at the light of the data analysis results.

2. Discovering what activates specialized knowledge in predicative verbs

Experts seem to be able to detect whether there is transmission of specialized knowledge or not among everything that is being expressed in their discourse

exchanges. This ability of distinguishing between specialized and general discourse makes us wonder what elements point to the conveyance of specialized knowledge (in this case, of verbs), and if these elements are linguistic or not.

We aim at verifying the following hypotheses:

- There are elements in the syntactic projection of verbs and their arguments that show the activation of specialized value; these should be or tend to be the same ones in order to be understood in communication between experts.
- (Specialized) Knowledge is transmitted, above all, by features of the minimal construction of the verb with its most internal argument, the activation of a specialized meaning even being able to imply changes in the semantics of the verb itself (verb class).

As has been already pointed out, we think that even if a verb conveying specific knowledge, in most cases, it would be impossible for it to give concrete expression to its meaning unless it is developed with all its arguments projected into the syntax. Consequently, when aiming at analyzing the possible terminological use of verbs, it will be necessary to observe characteristics of the verb itself, properties of its arguments and elements linking the verb and its arguments (Lerat 1995; L'Homme 1998).

In this research we aim to identify the specific characteristics of these constructions that, once combined, may act to readers as implicit triggers indicating

that there is a terminological use in discourse and that specialized knowledge is thus being expressed. Concretely, as specified in the section devoted to the selection and the characteristics of the corpus, this study focuses on a specific group of verbs as defined by Lorente (2007) and that the author calls *phraseological verbs*.

3. Previous work

Though verbs transmitting specialized knowledge are comparatively less studied than nouns, we still find a broad spectrum of interesting approaches to the subject. First, it is worth mentioning the essential works that consider the subject from a conceptual point of view and link their thoughts to which words are documented in terminological dictionaries (Rey 1979; Sager 1990). We equally find proposals for classifying specialized verbs (Condamines 1993; Lerat 2002 in the domain of law; Lorente 2002a and 2002b) and work that seeks to characterize verbs in specialized discourse (L'Homme 1998; Lerat 2002; Lorente 2002a, 2002b, 2007; Pimentel 2012). Moreover, part of this work has led to descriptive models for verbs (taking into account their argument structure and lexical relations). One example of this is the research project DiCoInfo (*Dictionnaire fondamental de l'informatique et de l'Internet*), where verb terms are analyzed considering their actantial (i.e. argument) structure, the linguistic realizations of their actants (i.e. arguments) and lexical relations, based on the Meaning-Text Theory (Mel'čuk 2004a and 2004b).²

The present research follows the Communicative Theory of Terminology, proposed by Cabré (1999, 2002a, 2002b), and thus adopts a linguistic and

communicative approach, that considers terms to be part of language. Therefore, as we will see below, research theories and applied projects that already exist for general language should also allow us to observe syntactic and semantic features in the realizations expressing specialized communication and are, thus, considered and adapted to the goals of the database created for this research.

Our starting point is the verb analysis approach established by Lorente (2002) in which the author considers lexical linguistic references (as Levin 1993 and 2000, or Pustejovsky 1995) for the analysis of terminological units. Concretely, most of the main general language theoretical references analyzed and chosen deal with the interaction between syntax and semantics: verb types according to adicity and aspect, argument structure, argument alternations and lexical syntax, thematic roles and proto-roles, and semantic features of verbs. As for applied references, SIMPLE (Lenci et al. 2003, and the related project CLIPS, CNR/ILC 2004), as well as the lexical-based ontology EuroWordNet (LE-2 4003 & LE-4 8328 [online]; this one in the use of the YATE term extractor, Vivaldi 2006) are taken into consideration. The use of one or several of these features aims at obtaining the most possible descriptive database in order to identify general trends in the behaviour of verbs in real contexts.³

4. Methodology

This part is divided in three subsections, each one explaining one aspect of the research methodology: the criteria used in the selection of the corpus, the structure established for the database, and the analysis carried out at different levels of data.

4.1. Corpus

The domains studied in this research are Genomics and Economics. This is done in order to ensure that the possible observed generalizations are not just prototypical of an area of knowledge (such as Natural Sciences or Human Sciences), but more likely to describe a more widely spread behaviour of verbs.

Concretely, the corpus used in this research is composed of 40 verb lemmas⁴ in Catalan language and chosen according to the following criteria:

(a) Lemmas need units in their context use to be able to fully express one of their (specialized) meanings. We are then interested in possible “Phraseological Verbs” (denomination taken from the typology proposed by Lorente 2007), and that can be described based on the different following properties (Lorente 2007):

- Relation with a nominal term: They are part of a phrase (such as *deliver judgment*).
- Semantic class: They express an Action (such as *to say*), a Change (such as *to die*) or a Cause of Change (such as *to kill*).
- Syntactic-semantic information: They express an Event and have a dyadic or monadic structure (two or one arguments) and their accompanying Noun Group is a relevant term of the domain. According to fixation, they can be phraseological units (verbal idiomatic

expressions and support verb constructions such as *to make a diagnosis*) or collocations (such as *to pass sentence, to codify proteins*).

- Unit transmitting specialized knowledge: They convey specialized knowledge when combined with other units (and not on their own).

One should bear in mind that the fact of working with specialized discourse does not imply that the selected realizations are themselves specialized. In fact, we voluntarily discarded lemmas with morphological clues indicating their possible specialization uses (such as *clonar* [*to clone*] or *hipotecar* [*to mortgage*]), and instead choose lemmas that may appear in very different domain discourses. This is done to ensure that a specialized meaning will have to be detected based on the cooperative use of the context in which the verb appears.

(b) Lemmas should present a variety of lexicographic grammatical information, since we do not want this variable to be a distortion to possible generalizations.⁵ We are interested in working with prototypically transitive verbs (such as *alterar*, where “A alters B”), intransitive verbs (such as *morir*, where “A dies”), as well as pronominal verbs (such as *separar-se*, where “A and B separate from each other”),

(c) Lemmas should prototypically be able to express different semantic classes for the same reason as (b). We choose prototypical verbs from different classes following the classification in SIMPLE (Lenci et al. 2003).⁶ Some examples would be *respondre* (*to respond*, COMMUNICATIVE ACTION) or *augmentar* (*to increase*, (CAUSE OF) CHANGE OF VALUE).

(d) Lemmas should be rather productive in both domains studied herein (Genomics and Economy). Since the research aims at considering data statistically and finding tendencies affecting their behaviour, a minimal number of occurrences is needed. For that matter, the verbs are selected from the list of lemmas appearing at least 100 times in each corpus.

Thus, by analyzing those verbs and their contexts, we should be able to verify and detail the existence or absence of elements in their syntactic projection and their arguments that would be indicative of the activation of specialized value, those elements expected to be features of the minimal construction of the verb with its most internal argument.

The analyzed occurrences of the lemmas (50 occurrences in each domain for each of the 40 lemmas) are extracted from the IULA Technical Corpus (University Institute for Applied Linguistics, Barcelona).

4.2. Structure of the database

As for the Database skeleton, it must be kept in mind that we base our research on the idea that specialized knowledge is conveyed by different units building it in the discourse, so that the specialized nature of a unit is not a consequence of its features, in an individual or isolated way, but of different elements accompanying it in its immediate context (Cabr  1999).⁷

Hence, for each realization (verb in context), we observe syntactic, semantic and syntactic-semantic features. Variables are crossed by means of statistical methods. The relational database is composed of four general information tables (lemmas, meanings, occurrences, arguments) and several complementary information boards. Data entry is done using two different approaches: first, contexts are considered separately; second, each meaning of a lemma is documented separately and linked to the occurrences that correspond to it.

Specifically, these are the features that are considered for each occurrence of the corpus (see Figure 1 for an example of the data-entry interface of the Database):

- For each occurrence (relation between verb and arguments):
 - Voice: active, morphological passive or pronominal passive.
 - Argument structure: monadic, dyadic or triadic.
- For each argument:
 - Syntax: Noun Group, Prepositional Group, etc.
 - Thematic role: Agent, Patient, Location, Event, Measure, etc.
 - Semantic feature: Result, Physical Space, Natural Matter, Inanimate Entity, Animate Plant, Animate Organism, etc.
 - The realization of the argument itself.

- Automatically assigned level according to termhood⁸: T1 (high level), T2 (medium level) or T3 (low level).
- For each occurrence (verb classes and meaning): after considering the features above, we assign to the verbal occurrence:
 - A semantic class: STATE, ASPECT, ACTION, CHANGE, CHANGE OF, etc.
 - A syntactic-semantic pattern: IDENTIFICATIVE STATE, STATIVE PLACE, JUDGMENT, PERCEPTION, ACTION WITH A GOAL, COMMUNICATIVE ACTION, NATURAL TRANSITION, CAUSE OF CHANGE OF VALUE, CREATION, etc.
 - A meaning: from the selected reference dictionaries.
- For each lemma: the different meanings documented are compiled with simple definitions, as well as the (different) grammatical information of the verb: transitive, intransitive, pronominal, etc.

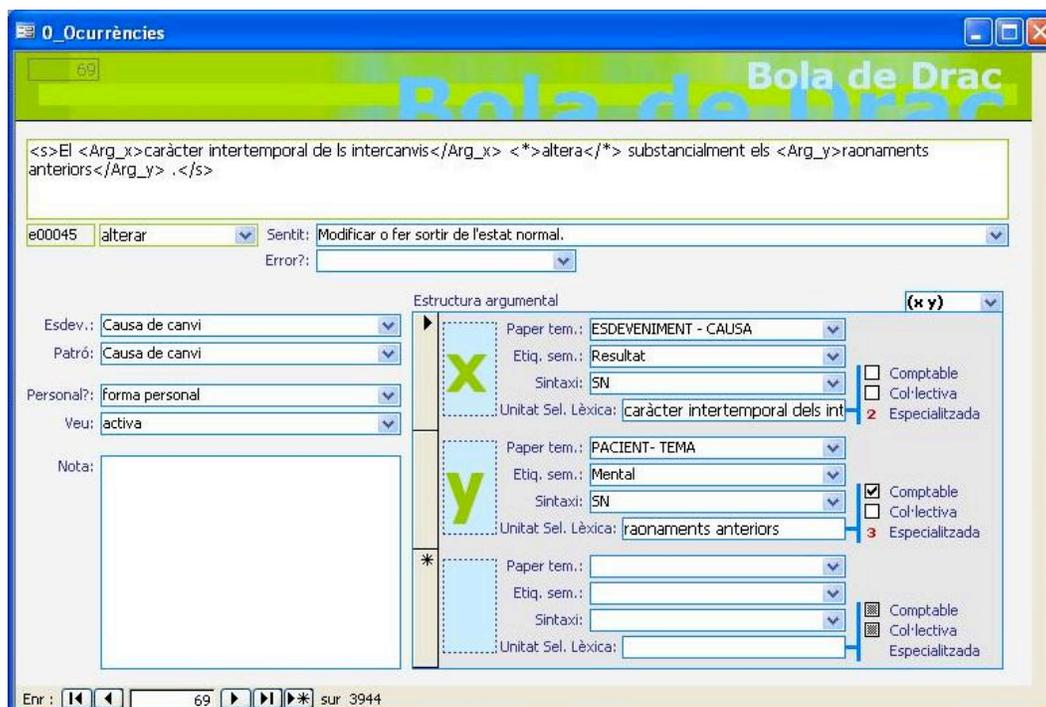


Figure 1. Data-Entry Interface of the Database (Occurrence and Arguments)

As a more concrete example, Table 1 offers an instance of the proposed analysis for the occurrence 1268 of the corpus.

Table 1. Analysis of the occurrence 1268 of the corpus

DATABASE FEATURES	« oc.1268 Marx dividia el capital en dues parts [...] » (Marx divided capital in two parts)
For each occurrence (relation between verb and arguments)	
Voice	Active voice
Argument structure	Triadic structure <i>Marx (x) + capital (y) + two parts (z)</i>
For each argument	
Syntax	Noun Group + Noun Group + Prepositional Group (en)
Thematic Roles	Agent + Patient-Theme + Quantity
Semantic features	Animate Entity - Person + Artificial Inanimate Entity + Measure
Realization itself	<i>Marx + capital + two parts</i>
Termhood	T3 + T1 + T3
For each occurrence (verb classes and meaning)	
Semantic class	CAUSE OF CHANGE
Syntactic-semantic pattern	CAUSE OF CONSTITUTIVE CHANGE
Meaning	To classify.
For each lemma	
List of meanings	For the lemma <i>dividir</i> , a total of 5 definitions were compiled: “To separate from each other”, “To perform the mathematical process of division”, “To classify”, “To separate something or someone” and “(A number) to be able to perform the mathematical process of an exact division”.

4.3. Analysis of the data

In the following subsections we present the different levels of description by which occurrences of the corpus are analyzed.

4.3.1. Features of the relation between verb and arguments

Concerning the interrelation between syntax and semantics, we are interested in the adicity as a phenomenon that is projected into syntax and that has semantic foundations (Levin 1993, 2000; Levin and Rappaport 2003), because it can show us if: (a) some verbs always follow the same argument structure or have several ones; (b) whether or not the use of one structure or another is domain-dependent; and (c) if real sentences are the result either of canonical projections or of marked forms (focalization).⁹ Therefore, this information is useful to establish how context units unfold around a verb, and which of them seem to be more relevant than others at fully expressing its meaning.

When looking at our corpus, we see that the dyadic structure is the most common of all. As for the most common structure alternations, they are monadic – dyadic, and dyadic – triadic. The couple (x) and (x y) is usually related by inchoativization (with or without reflexive):

oc.198 <s> **augmentem** gradualment la quantitat utilitzada de l'altre _[y], [···] </s>

*[(we) gradually **increase** the quantity used by the other part]*

oc.196 <s>Les tarifes salarials pactades _[x] en els convenis col·lectius signats han

augmentat un 3,7%. </s>

*[The negotiated wage rates on the signed collective agreements have **increased***

a 3,7%]

The couple (x y) and (x y z) appears mainly by the agglutination of two of the three initial canonical arguments into one, and this is done either by reflexivization, by coordination or pluralization from two arguments to one:

oc.2061 <s>La que **lliga** les asseguradores [y] amb els subministradors de serveis [z]
[···] </s>

[The one that **binds** insurance companies with service suppliers]

oc.2069 <s>[...] pot arribar a detallar unes funcions [x] que **lliguin** les utilitats
individuals [y] [···] </s>

[... some functions that **bind** individual profits]

A gradation on impersonalization can also be drawn by means of the different interventions of focalization on a canonical structure. Not all levels of impersonalization exist for each verb of the language, but we think that they most probably arise depending on the needs of use in real discourse. Here are some examples of this gradation:

- Active Transitive: canonical form.

oc.1514 <s>Els caràcters externs del cos [x] **expressen** sobretot els resultats de la
selecció natural [y] </s>

[external features of the body mainly **express** the results of natural selection]

- Morphological Passive: the grammar allows the use of a by-agent complement, although it is not compulsory.

oc.1598 <s>aquesta realitat [y] ha estat **expressada** també pels grans comerciants [x]
</s>

[this reality has also been **expressed** by the big traders]

- Pronominal Passive or Pronominal Impersonal: grammar does not allow the use of a by-agent complement but the construction accepts an adjunct expressing cause or instrument of action. In some cases, a reflexive inchoative will overlap with a pronominal passive due to some ambiguity in the context.

oc.269 <s> en els últims 25 anys s'ha **avançat** en el terreny de la igualtat [...] </s>

[in the last 25 years, it has been **move forward** in the field of equality]

- Inchoative without reflexive pronoun: the subject corresponds to the subcategorized argument from the canonical structure, and all the possible formal clues on adicity do not appear, even if in some cases the realization might show an adjunct expressing a cause or an instrument of action.

oc.220 <s>l'aïllament de gens d'interès terapèutic [x], ha **avançat** considerablement en
els darrers anys.</s>

[the isolation of genes of therapeutic interests has been **moving forward** in lasts
years]

Therefore, we can see that, on the basis of a two-argument structure (one of the arguments being the Agent), the subcategorized argument tends to increase in importance (is focalized) and progressively takes the place of the Agent. At the same time, the intervention in the discourse of the agentive external argument of the canonical structure progressively decreases, the occurrences mainly moving from a CAUSE OF CHANGE to a CHANGE.

The way in which language uses its resources to express focalizations (and new uses and meanings of units) is also made in some cases from CHANGE towards CAUSE OF CHANGE. Some examples are found in the causative versions detected for *créixer* (CAUSE OF CHANGE OF NATURAL TRANSITION), replacing a type of expressivity that the old Catalan verb *acréixer* (equivalent to the transitive formulation of the English verb *to grow*, where “A grows B”, or “A causes B to grow”) would be able to fulfil:

oc.606 <s> [···] *les cèl·lules d'endoteli de cordó umbilical humà (HUVEC) [y] van ser*

crescudes. </s>

[HUVEC cells were **grown**]

oc.607 <s> *Les cèl·lules Jurkat i les cèl·lules HUVEC [y] es van créixer.* </s>

[Jurkat cells and HUVEC cells were **grown**]

It is also worth mentioning that the lemmas studied can also present alternation of their constituents. As seen in Lorente (1994), this switch of constituents appears

mainly between the noun group and the prepositional group, and between the noun group and propositions introduced by (non-)conjugated verbs.

To summarize, in most cases, the aim of focalization is to focus on the internal argument of the canonical structure and place it in a prominent position in the marked structure. It must be noted that focalization also works as an impersonalization resource, aiming at creating an objective discourse, typical of specialized texts.

4.3.2. Features of the arguments

The use of thematic roles as descriptive features also allows us to consider certain regularities that adicity or constituent alternation would not be able to show: are there some thematic roles more inclined than others to transmit (specialized) knowledge? In order to observe this, we take into account Dowty's (1991) proposal, especially in relation to the idea of two basic poles, Proto-Agent and Proto-Patient, as well as to the idea of a continuous list where thematic roles are ordered according to their prototypicality. We base our research on the idea of a continuum and non-compartmented classification among thematic roles and we propose a specific typology for our description (Figure 2).

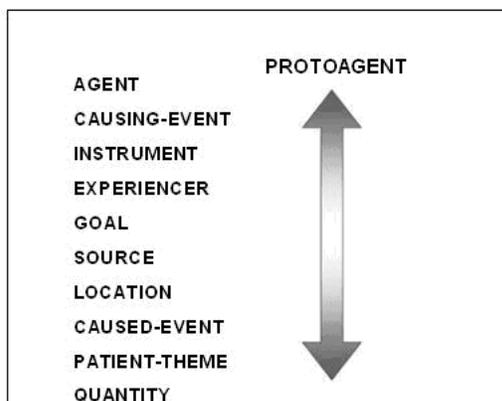


Figure 2. Thematic Role proposal for the Database

For both specialized domains, the most frequent combinations of thematic roles according to adicity would be Proto-Patient (oc. 2488), Agent + Patient (oc. 1314) and Agent + Patient + Goal (oc. 3809):

oc.2488 <s>L'home _[x] **mor.**</s>

*[men **die**]*

oc.1314 <s>[...] la capacitat de clonatge dels diferents vectors _[y] que hem **estudiat**.</s>

*[different vectors that we have **studied**]*

oc.3809 <s>Tots els teixits _[y] es van **tallar** en seccions de 14 xxx de gruix _[z] </s>

*[all tissues were **cut** into sections 14 xxx thick]*

The structures being monadic, dyadic or triadic, we see that the element appearing on all of them would be the (Proto)Patient argument (the internal one), placed either as a subject or as a subcategorized complement.

As for another level of description, the specific and independent semantic features of lexicon units acting as arguments of an occurrence (based on Pustejovsky

1995 and the CLIPS project documents by CNR/ILC 2004) can also be important as per the positioning of the information and the characteristics of it in the expansion of the verb in real occurrences. These elements can show us whether there are correlations between the thematic roles of arguments and the semantic classes of nouns expressed in these arguments. In order to be able to observe the tendencies or regularities, we establish a hierarchy to help us group all twenty particular types into groups of labels expressing similar features (Figure 3).

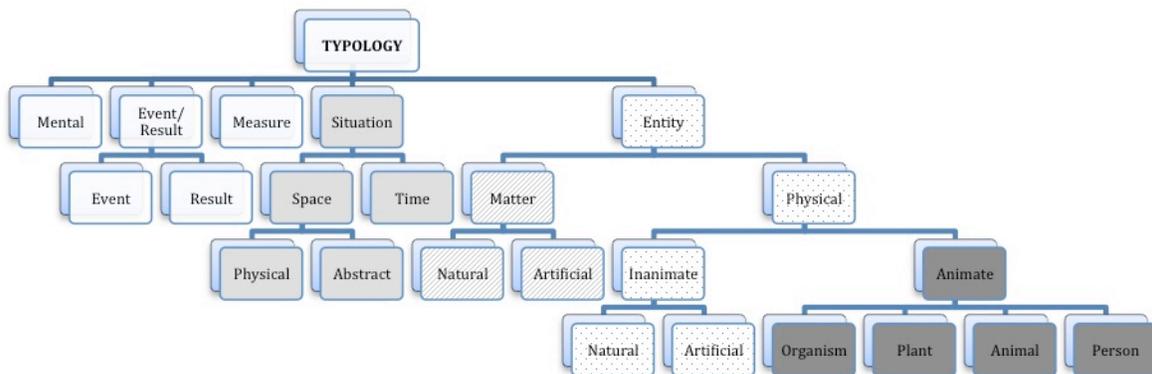


Figure 3. Hierarchical Semantic Feature Typology proposal for the Database

As shown in Figure 3, five meta-groups are created in order to observe tendencies related to the domain discourse (from left to right in the patterned typology):

- Event-related group, including arguments expressing Mental, Event, Result and Measure.
- Situation-related group, regrouping Space Location, either Physical or Abstract, and Time Location.
- Matter-related group, with arguments representing Matter Entities, either Natural or Artificial.
- Inanimate-related group, including Inanimate Physical Entities, either Natural or Artificial.
- Animate-related group, with the specific labels of Animate Physical Entities, specified as Organism, Plant, Animal or Person.

At this level, in the realizations of some specific verbs, though variation of the labels is very high, we can see some tendencies of use, which are mainly linked to each and every discursive domain. Most of the differences between domains appear in the subcategorized complement *y*, with a high variety of word selections in the Economy subcorpus, whereas the use of arguments expressing mainly Animates (people suffering diseases, patients or living organisms as cells), Natural Matters (*DNA, amino-acids* or *tissue*) and Events (referring to diseases, symptoms or physicochemical reactions such as *apoptosis, digestion, genetic alteration* or *miscarriage*) characterizes the Genomic subcorpus.

Apart from thematic roles and semantic labels, the possibility for an argument to convey specialized knowledge itself is also taken into account in the analysis. Our goal is to observe the termhood of the noun and prepositional groups acting as arguments of the studied verbs, since it can also be important in relation to the more global transmission of specialized knowledge of the argument structure and, hence, when considering the conveyance of specialized knowledge of verbs. In this sense, we establish termhood (from 0 to 1)¹⁰ for each argument and domain in order to obtain possible useful information which, in our Database skeleton, would be placed in between the descriptive granularity of the semantic labels seen above and the direct lexical selection.

The process used to establish the ranking for the information of termhood (T1, T2, T3) is the following: first, we use YATE (a terminology extractor using linguistic – EuroWordNet and reference documents from a domain – and statistical strategies; Vivaldi, 2006), then Mercedes (a term detection system using dictionaries from a specific domain; Vivaldi, 2004) and, finally, we consult specialists of both domains (in the Mercedes phase, and in order to reduce the weight of manual tasks, only arguments not having obtained a high level rank on previous instrumental methods are manually considered).

Information obtained from this description feature shows that the behaviour of most verbs varied considerably in relation to the termhood level assigned to each of their arguments. Some possible tendencies can be recorded for each domain, but no

relevant information on the specialized use of verbs is obtained from this descriptive feature alone.

Nevertheless, when combined with the semantic features of arguments, each discourse tends to make its own choices in relation to the termhood of the argument and its semantic features, mainly on the *y* argument, such as genetic organisms in Genomics and economic flow terms in the Economic discourse.

4.3.3. Features of the realizations of the verb and the meanings expressed

Once the argument structure and the different arguments are studied, we consider each realization as a whole and assign a semantic class to the verb. Following the classification by SIMPLE cited above, we detect verbs expressing STATES, ASPECTS, PSYCHOLOGICAL EVENTS, ACTIONS, CHANGES and CAUSES OF CHANGE. Uses in context show us some variation inside lemmas that is not expected in the classification variety used *a priori* to select the verbs. Similarly to the other individual features considered before, the variation among semantic classes is higher in economy texts than genomic ones.

Since we believe that each sense is tightly linked to realization and context, once all data from each context is gathered, we assign a meaning to each realization. In this research, *Meaning* is considered according to Ullmann (1962) and Lyons (1977):

A series of tests designed to study the influence of context has shown that there is usually in each word a hard core of meaning, which is relatively stable and can only be modified by the context within certain limits. At the same time

no one would deny the crucial importance of context in the determination of word-meanings. As far as the role of *verbal context* is concerned, this was already recognized as fundamental by some of the pioneers of modern semantics; Darmesteter, for example, spoke of the various elements of a sentence “conspiring”, by their distribution and their collocations, to modify the meanings of individual words. (Ullmann 1962, 49)

The sense of an expression [...] is a function of the senses of its component lexemes and of their occurrence in a particular grammatical construction.
(Lyons 1977, 206)

As for words transmitting specialized knowledge in one or several of their uses, we also take into account Adelstein (2007, for relational nouns) and the author’s following hypothesis:

[...] the specificity of specialized lexicon is found in the overlapping semantic configuration of its senses, being determined by different factors from the communicative situation, by phrasal factors and by local syntax-semantic factors. [our translation] (Adelstein 2007, 17)

Thus, we take into consideration different interlinked syntactic-semantic properties as a form by which specialized knowledge can be expressed.

Table 2 shows an example of the sense assignment method used for our data of the lemma *créixer* [to grow]. The numbers appearing next to the different analyzed features refer to the number of occurrences in the corpus having this characteristic.

Table 2. Sense assignment for *créixer*

	<i>créixer</i> [to grow] = 37 [conjugated personal forms]			
Adicity	(x) = 29		(x y) = 8	
Subcategorization of the constituents	xNG = 29		x∅ = 2 y _{NG} = 2	xNG = 6 y _{PG} = 3 / _{NG} = 3
Thematic Role	xProto-Patient = 29		xProto-Agent = 2 yPatient-Theme = 2	xProto-Agent = 6 yQuantity = 6
Global Semantic Selection	xAnimate-rel = 14	xEvent-rel = 15	x∅ = 2 yAnimate-rel = 2	x∅=1 / Event-rel = 5 yEvent-rel= 6
Specific/Lexical Semantic Selection	xAnimate	xMeasure / Result / Mental	x∅ yOrganism	xVariation yMeasure
Termhood	xT1 / T2	xT2 / T3	x∅ yT1	xT2 yT2
Semantic Class	CHANGE		CAUSE OF CHANGE	
Syntactic-semantic Pattern	NATURAL TRANSITION	VALUE (STATE)	NATURAL TRANSITION	VALUE
Productivity according to domain	E = 1 G = 13	E = 14 G = 1	E = 0 G = 2	E = 6 G = 0
Sense	A	B	C	D

First, we divide between monadic and dyadic occurrences. Then, within the dyadic group, we distinguish two types of realizations according to the subcategorization of constituents (noun group and prepositional group). Concerning thematic roles, we do not find variation among monadic realizations, whereas it exists among dyadic ones.

This variation, appearing in the *y* argument, corresponds to the group already established in the subcategorization level.

Global semantic selection confirms the same previous subgroup, specifically with the presence of units expressing animate beings as arguments. As for the specific semantic selection, the subgroups of monadic realizations continue to express the growing process of living creatures, whereas the two subgroups of dyadic realizations add more specific distinctions between them in relation to the subcategorized complement: we find a group presenting Organisms, whereas the other usually presents Values.

According to termhood, the first monadic group presents the *x* argument with a higher termhood level than the second group. Similarly, in the case of dyadic realizations, the first group is also the one presenting an indisputably higher termhood in the *y* argument. Apart from adicity and thematic roles of the *y* argument, the combination between semantic selection, lexical selection, termhood and thematic domain is the most important role in the classification of occurrences among different meanings.

We thus have, in Table 2, two basic blocs (monadic and dyadic), either expressing CHANGE or CAUSE OF CHANGE (OF VALUE and OF NATURAL TRANSITION). As for the productivity of these groups, the expressions of natural transition appear in the genomic corpus, whereas the ones linked to Change of value come from the economic subcorpus.

The meanings established and marked with capital letters in Table 2 are assigned to definitions, taken from reference dictionaries, except in the case of [C], which is not documented in Catalan and that we create *ad hoc*. The English version of the definitions would be: (A) “to become bigger, taller etc. over a period of time in the process of becoming an adult”; (B) “to increase in amount, size, number, or strength” (monadic version); (C) “to make plants or crops develop and produce fruits or flowers” (as mentioned above, when the meaning of *créixer* is *raise*, though this definition already existed in English in the fields of Agriculture and Gardening –and it has later been transposed to Genomics– it is not the case for Catalan language); (D) “to increase in amount, size, number, or strength”, (dyadic version).¹¹

Finally, after revising the data according to meaning assignation, it is worth mentioning the following:

- In most cases, the meanings detected from occurrences correspond to those documented in dictionaries.
- Most cases of adicity difference (they are not many) do not imply a change of meaning and they usually refer to inchoative versions or to alternations between dyadic and triadic structures where some of the constituents are implicit.
- Most alternations do not imply a change of meaning, and they are usually variations of constituent subcategorizations not being explicit in the reference dictionaries but appearing in the uses of the corpus.

- The cases with transitivity/intransitivity switch are very rare and, in fact, when they occur, they express a specific meaning not documented in the reference works in Catalan, i.e. the meaning [C] of *créixer*.

5. Results on specialized uses of verbs

When considering the transmission of specialized value, 11 out of the 40 lemmas studied appear in contexts where specialized knowledge is being conveyed. Most of the specialized meanings of these lemmas are found in the genomic corpus (*créixer* [to grow], *expressar* [to express], *manifestar* [to manifest], *patir* [to suffer], *repetir* [to repeat]), except for *avançar* [to move forward], *disposar* [to dispose], *dividir* [to divide] and *lligar* [to bind], where specialized realizations appear in the economic corpus, and *respondre* [to respond] and *sotmetre* [to subject], presenting specialized senses both in genomic and economic contexts. For example, in the case of *respondre*, constructions expressing “to show a response to a drug” are found in genomics corpus whereas constructions expressing “to take responsibility of debts” are found in the economic corpus.

We can also distinguish between the behaviour of verbs expressing events and verbs expressing states. When expressing an event, the conveyance of specialized knowledge is done through the verb and the argument being its subcategorized constituent. This is the case for most of the occurrences observed, where the combination of the two elements allows the mentioned transmission of specialized knowledge.

For instance, with *expressar* [to express], in most of the genomic realizations, the event expressed is not a general COMMUNICATIVE ACTION but an action implying in the field of genetics the creation, fabrication, production or synthesis of concrete entities (the subcategorized element):

oc.1520 <s>Per comprovar si la proteïna xxx [y] s'**expressava** en cèl·lules de mamífer [...] </s>

[to check if xxx protein was **expressed** in mammal cells...]

On the other hand, when expressing a stative condition, the transmission of specialized knowledge takes place through the verb and all the arguments (being or not subcategorized). We find here cases where the combination of elements allows expressing a specific meaning of the stative genre, in which both arguments of the realization are rather in relation in a distinctive way.

For instance, with *manifestar* [to manifest; dyadic form], we can express a STATIVE POSSESSION between the two arguments of the construction: "Someone having the symptoms of a disease" or "A having symptoms of disease B":

oc.2116 <s>[...]els fills [x] **manifestin** caràcters que restaven amagats en els pares [y], en forma d'al·lels recessius .</s>

[children **manifest** features that were hidden in progenitors by means of recessive alleles]

Thus, the data analyzed allow us to confirm the main criteria established by Lorente (2007) for phraseological verbs mentioned in section 4.1 and display the following characteristics, concerning the behaviour of stative occurrences and triadic forms, when expressing specialized knowledge:

- When expressing events:
 - The distribution of marks of terminological value on their projections appears with the combination of the verb and the argument realized as the subcategorized constituent. The latter is either in the position of the *y* argument (in dyadic and triadic constructions) or in the position of the *x* argument (in monadic (pronominal) constructions with the argument being a ProtoPatient).
 - The *x* argument in dyadic and triadic forms does not seem to be so important in the conveyance of specialized knowledge, and we have seen this above in the fact that this argument is often implicit in the corpus. This is done by means of a passive / impersonal pronominal construction, and it is sometimes compensated by an instrumental adjunct. Moreover, the possibilities of lexical selection are less restricted in this kind of arguments when comparing them to the *y* argument.
 - In triadic constructions, the *z* argument is also important in the expression of specialized meaning. We even find some parallel pronominal dyadic constructions where the *y* argument moves into the subject position, and

the *z* argument into the typical position of the subcategorized constituent.

We also find some pluralizations that allow grouping argument *y* and *z* in a single one.¹²

- When expressing stative situations:
 - The distribution of terminological value appears in the combination of the verb and the two arguments that are linked to the verb. We have seen an example of this phenomenon with *manifestar* [*to manifest*] above.

Moreover, the criterion proposed by Lorente (2007, 376) in relation to phraseological verbs (the nominal term linked to it is the argument that occupies the subcategorized constituent position) is confirmed by our data and reinforced by the results from the YATE extractor. In fact, in our analysis, all the cases where constructions transmit specialized value, the internal argument also automatically receives a high termhood value.

If we reconsider here the factors determining the activation of specialized value established by Adelstein (2007) for relational nouns, and we focus on the syntactic-semantic and the phrasal characteristics proposed by this author (overlapping in the case of verbs), our data establishes that a higher restriction on semantic selection of the internal argument is the main element to be taken into account. This sole element related to the activation is nevertheless subjugated under the two following aspects:

- The changing element: internal argument (one in event dyadic forms, and two in event triadic forms and stative realizations).
- The changing condition: tightly linked to the specific discourse (thematic global factor).

Finally, it is worth mentioning that this semantic selection may even affect the structure at other levels as well, for example in the number of arguments (such as the dyadic version of *dividir*; “(A number) to be able to perform a mathematical process of an exact division”) or in the restricted use of a concrete preposition when expressing a specific sense and not when expressing a more general one, where this preposition does not seem to be compulsory in the occurrence (*patir*; “to suffer”):

oc.1280 <s>[...] o bé xxx_[x] o bé xxx_[x] **divideix** xxx_[y]. </s>

[either xxx or xxx **divide** xxx]

oc.2601 <s>Els pacients_[x] **pateixen** una pèrdua de la visió macular_[y] </s>

[patients **suffer** a loss of macular vision]

oc.2749 <s>Els ports espanyols_[x] **pateixen** d’una multiplicitat d’empreses prestatàries de serveis_[y] </s>

[Spanish harbors **suffer** from a multiplicity of service borrower enterprises]

When observing the example of *dividir* above, the use of a dyadic structure instead of the prototypical triadic structure (person A dividing a number by another one) shows us a

specific meaning where the division is exact and without decimal point between two numbers, or “number X divides number Y” when X is a denominator of Y (or Y is a multiple of X). On the other hand, when considering the specialized Genomics discourse and the use of *patir* (to suffer) in the sense of having a disease, it is always expressed with a construction having a Noun Group as the subcategorized constituent (see the first example of *patir* above), whereas variation appears for this verb in the Economics corpus, randomly used with a subcategorized Noun Group or with a Prepositional Group.

6. Conclusions

To conclude, based on the data analyzed, we reconsider the main assumptions made at the beginning of this article.

The first hypothesis was: “There are elements in the syntactic projection of verbs and their arguments that show the activation of specialized value; these should be or tend to be the same ones in order to be understood in communication between experts”.

The data analyzed allow us to validate the first part of the hypothesis, since syntactic and/or semantic characteristics in the occurrences –adicity, subcategorization of constituents, thematic roles and semantic selection of arguments (the latter often combined to a high termhood of argument units)– show us the activation of terminological value in some realizations compared to others of the same verb. Nevertheless, for the second part of the hypothesis, we observe that not all the

elements considered convey the activation of terminological value in all cases. Moreover, they do not always have the same influence on the transmission of specialized knowledge. The most important element would be the semantic selection and sometimes even the lexical selection of a unit or a group of units. Moreover, this is not necessarily the only element in the activation of specialized value, since statistics show that the other properties mentioned above can play a complementary role reinforcing the expression of a meaning instead of another (Joan Casademont 2008).

The second hypothesis was: “(Specialized) Knowledge would be transmitted, above all, by features of the minimal construction of the verb with its most internal argument, the activation of a specialized meaning even being able to imply changes in the semantics of the verb itself (verb class)”.

The data allow us to validate one part of the hypothesis, since we see that the knowledge transmission of a verb could slightly vary according to its type. For the verbs expressing events, we see that in the monadic occurrences, the verb and the argument acting as the subject together form a minimal construction. In the case of dyadic occurrences, the relevant group would be the verb and the argument acting as the first complement of the verb, similarly to the triadic occurrences, which add also another internal argument to the relevant minimal construction. On the other hand, for verbs expressing stative situations, the marks of conveyance take place between the verb and the two complements being linked by it. The construction of these three elements would be then the minimal construction to convey specialized knowledge.

Finally, concerning the second part of the hypothesis, we can validate it in some cases, taking into account that the activation of specialized value of the minimal construction of a verb may include sense variation at two levels, since it is linked to two types of phenomena detected in the occurrences of our corpus. On the one hand, usually, when comparing specialized occurrences to other occurrences expressing general knowledge, only lexical or semantic restrictions on the arguments are found in the specialized cases, the general ones having a less restricted unit selection on the arguments. On the other hand, sometimes, combined to the restriction on the lexical or semantic selection (combined or not with other possible characteristics in the context indicating activation), the differences between general and specialized realizations become more complex. In those, we then find a change in the semantic class of the verb. This would be the case, for example, of the verb *créixer* [*to grow*] in Genomics, where the transmission of specialized knowledge in Catalan presents a verb of CAUSE OF CHANGE class instead of CHANGE (NATURAL TRANSITION), or *manifestar* [*to manifest*], where the verb expresses a genomic specialized meaning of STATE instead of ACTION, as in its related general occurrences.

All these results confirm that a phraseological verb transmits specialized knowledge by the cooperation of the units appearing in its context. Thus, it is not an item just appearing randomly in discourse, but highly interacting with the units expressed in its context (minimal constructions established above). It would hence be clearly useful to specialists in training that consult monolingual resources and to

translators when consulting bilingual dictionaries, if the information on this syntactic-semantic interaction were more systematically studied and documented.

Acknowledgements

I would like to thank Mercè Lorente Casafont, my PhD thesis supervisor who always found ways to motivate me. I cannot forget Jesús Carrasco Iglesias and Jorge Vivaldi Palatresi who have contributed in improving my research in a lot of different ways. To Albert Morales Moreno and Marie-Claude L'Homme, thank you for your valuable comments and for encouraging me to resume this research after some time on hold. Finally, I am very grateful to Marc Pomerleau for his notes on the English language.

References

Adelstein, A. 2007. *Unidad léxica y significado especializado: modelo de representación a partir del nombre relacional madre*. Barcelona: Institut Universitari de Lingüística Aplicada.

Cabré, M. T. 1999. *La terminología: representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Lingüística Aplicada.

Cabré, M. T. 2002a. "Teorías de la terminología: de la prescripción a la descripción." In *Innovazione lessicale e terminologie specialistiche*, 169-188. Florencia: Leo S. Olschki Editore.

Cabré, M. T. 2002b. "Terminología y Lingüística: La teoría de las puertas." *Estudios de Lingüística Española (ELIES)* 16. <http://elies.rediris.es/elies16/Cabre.html>. Accessed April 2005.

CNR/ILC. 2004. "CLIPS" [project documents]. [Pisa]: CNR-Consiglio Nazionale delle Ricerche, ILC-Istituto di Linguistica Computazionale. <http://www.ilc.cnr.it/clips>. Accessed October 2005.

Condamines, A. 1993. "Un exemple d'utilisation de connaissances de sémantique lexicale: acquisition semi-automatique d'un vocabulaire spécialisé." *Cahiers de lexicologie* 62: 25-65.

Dowty, D. 1991. "Thematic Proto-Roles and Argument Selection." *Language* 67 (3): 547-619.

Enciclopèdia Catalana. *Gran diccionari de la llengua catalana*.

<http://www.enciclopedia.cat/diccionaris#.Uqzr9WTuLdl>. Accessed 12 December 2013.

EuroWordNet. *Building a Multilingual Database with Wordnets for Several European Languages (EuroWordNet)*. LE-2 4003 & LE-4 8328. <http://www.ilc.uva.nl/EuroWordNet>. Accessed November 2005.

Ginebra, J., and A. Montserrat. 1999. *Diccionari d'ús dels verbs catalans*. Barcelona: Edicions 62.

Grimshaw, J. 1990. *Argument Structure*. Cambridge (Mass.): MIT Press.

Longman. *Longman Dictionary of Contemporary English*.

<http://www.ldoceonline.com/dictionary/grow>. Accessed 12 December 2013.

Joan Casademont, A. 2008. *Sintaxi i semàntica verbals en el discurs d'especialitat: elements per a l'activació del valor terminològic*. Barcelona: IULA.

Kageura, K., and B. Umino. 1996. "Methods of Automatic Term Recognition: A Review." *Terminology* 3 (2): 259-289.

L'Homme, M. C. 1998. "Le statut du verbe en langue de spécialité et sa description lexicographique." *Cahiers de lexicologie* 73 (2): 61-84.

L'Homme, M. C. 2012. "Le verbe terminologique : un portrait de travaux récents". In *Actes du 3e Congrès mondial de linguistique française*, ed. by F. Neveu et al. Lyon, France: EDP Sciences.
<http://dx.doi.org/10.1051/shsconf/20120100340>. Accessed February 2013.

Lenci, A., and N. Calzolari, N., and A. Zampolli. 2003. "SIMPLE: Plurilingual Semantic Lexicons for Natural Language Processing." In *Computational Linguistics in Pisa-Linguistica Computazionale a Pisa. Linguistica Computazionale, Special Issue XVI-XVII*. Vol. I, ed. by A. Zampolli, and N. Calzolari, and L. Cignoni, 323-352. Pisa-Roma: CNR- Consiglio Nazionale delle Ricerche, ILC-Istituto di Linguistica Computazionale.
<http://www.ilc.cnr.it/viewpage.php/sez=ricerca/id=58/vers=ita>. Accessed September 2005.

Lerat, P. 1995. *Les langues spécialisées*. Paris: PUF.

Lerat, P. 2002. "Qu'est-ce que le verbe spécialisé? Le cas du droit." *Cahiers de Lexicologie* 80: 201-211.

Levin, B. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

Levin, B. 2000. "Aspect, Lexical Semantic Representation, and Argument Expression." In *Proceedings of the 26th Annual Meeting of the Berkeley Linguistics Society*. Berkeley: 413-429.

Levin, B., and M. Rappaport. 2003. "The Semantic Determinants of Argument Expression. A view of the English Resultative Construction." In *The Syntax of time*, ed. by J. Guéron and J. Lecarme. Cambridge (Mass.): MIT Press: 477-494.

Lorente, M. 1994, 2003. *Aspectes de lexicografia: representació i interpretació gramaticals* [CD-ROM]. Barcelona: Universitat de Barcelona, Institut Universitari de Lingüística Aplicada.

- Lorente, M. 2001. "Tipología verbal y textos especializados." In *Cuestiones conceptuales y metodológicas de la lingüística*, ed. by M. González Pereira, and M. Souto Gómez, 143-153. Santiago de Compostela: Universidade de Santiago de Compostela, Servicio de Publicaciones..
- Lorente, M. 2002a. "Verbos y discurso especializado." *Estudios de Lingüística Española (ELIES)* 16. <http://elies.rediris.es/elies16/Lorente.html>. Accessed February 2004.
- Lorente, M. 2002b. "Terminología y fraseología especializada: del léxico a la sintaxis". In *Panorama actual de la terminología*, ed. by G. Guerrero, and M. F. Pérez, 159-179. Granada: Comares.
- Lorente, M. 2007. "Les unitats lèxiques verbals dels textos especialitzats. Redefinició d'una proposta de classificació." In *Estudis de lingüística i de lingüística aplicada en honor de M. Teresa Cabré Castellví. Volum II: De deixebles*, ed. by M. Lorente, and R. Estopà, and J. Freixa, and J. Martí, and C. Tebé: 365-380. Barcelona: Institut Universitari de Lingüística Aplicada.
- Lyons, J. 1977. *Semantics. Volume 1*. Cambridge, London, New York, Melbourne: Cambridge University Press.
- Mel'čuk, I. 2004a. *Actants in Semantics and Syntax*. I, *Linguistics* 42 (1): 1-66.
- Mel'čuk, I. 2004b. *Actants in Semantics and Syntax*. II, *Linguistics* 42 (2): 247-291.
- Observatoire de Linguistique Sens-Texte. *DiCoInfo Dictionnaire fondamental de l'informatique et de l'Internet*. <http://olst.ling.umontreal.ca/cgi-bin/dicoinfo/search.cgi>. Accessed February 2013.
- Pimentel, J. 2012. "Description de verbes juridiques au moyen de la sémantique des cadres." In *Terminologie / Ontologie: Théories et applications* (Toth 2011), Annecy: 145-166.

Pustejovsky, J. 1995. *The Generative Lexicon*. Cambridge (Mass.): MIT Press.

Rey, A. 1979, 1993. *La terminologie: noms et notions*. Paris : PUF.

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

Ullmann, S. 1962, 1992. *Semantics. An Introduction to the Science of Meaning*. New York: Barnes & Noble.

University Institute for Applied Linguistics (IULA). Technical Corpus

<http://www.iula.upf.edu/recurs01uk.htm>. Accessed April 2005.

Vivaldi, J. 2004. *Sistema de reconocimiento de términos Mercedes. Manual de utilización. Versión 3*. Barcelona: Institut Universitari de Lingüística Aplicada.

Vivaldi, J. 2006. *Sistema de extracción de candidatos a término YATE: manual de utilización*. Barcelona: Institut Universitari de Lingüística Aplicada.

Author's address

UER Sciences humaines, Lettres et Communications (Télé-université)
5800, rue Saint-Denis
Montréal (Québec) H2S 3L5
anna.joan.casademont@teluq.ca

About the author

Anna Joan Casademont obtained a PhD in Language Sciences and Applied Linguistics in 2008 (Universitat Pompeu Fabra) and a Master's Degree in Museology in 2011 (Université du Québec à Montréal). She is a collaborator of the IULATERM Research Group (Lexicon,

terminology, specialized discourse and linguistic engineering, Universitat Pompeu Fabra) and an associate member of the OLST (Observatoire de linguistique Sens-Texte, Université de Montréal). Her research interests include different fields of (applied) linguistics, such as lexical semantics, terminology, information retrieval, translation and sociolinguistics, as well as museology and Japanese language and culture. She is currently in charge of workshops on Catalan language (Université de Montréal), in addition to courses on terminology and translation at Tél-Université (Université du Québec), where she works as a tutor and an external course developer. She is also a translator and an interpreter.

Notes

¹ This article focuses on one of the topics covered in my PhD thesis, supervised by Mercè Lorente Casafont. This research would not have been possible without the fellowships received from IULA-UPF, from the IULATERM group and from the research projects TEXTERM2 (BFF2003-02111), RICOTERM2 (HUM2004-05658-C02-01/FILO) and RICOTERM3 (HUM2007-65966-C02-01/FILO).

² For a more detailed portrait on research on terminological verbs, see L'Homme (2012).

³ We are not going to see in detail the use of statistics as a tool in this specific article, since we decided to use these pages to focus on other aspects of the research. Nevertheless, though not treated in detail here, it is worth mentioning it because it allows us to confirm some of the linguistic intuitions that we have when looking at raw data. In the research, we want to establish a basis for systematic methodologies in order to describe linguistic phenomena or make generalizations on analyzed linguistic data. In some cases, a more important amount of data would be necessary to completely confirm those intuitions and we are able to only establish behaviour tendencies. Either way, based on the contrast in data, statistics helps us (a) to compare among uses from different thematic corpus, (b) to try to establish the sense corresponding to each occurrence in the most possible systematic way, and (c) to detect specially prominent features in the uses transmitting specialized knowledge comparing to the ones in the uses not conveying this kind of information. For this, we use correspondence analysis, discriminant analysis and ANOVA (analysis of variance).

⁴ The corpus compiles 3,944 occurrences (2,049 being in personal form). Concretely, the lemmas studied are the following: *alterar* [to alter], *augmentar* [to increase], *avançar* [to move forward], *canviar* [to change], *contenir* [to contain], *contribuir* [to contribute], *créixer* [to grow], *creure* [to believe], *dependre* [to depend], *descobrir* [to discover], *detectar* [to detect], *disposar* [to dispose], *dividir* [to divide], *estudiar* [to study], *evitar* [to avoid], *expressar* [to express], *funcionar* [to work], *generar* [to generate], *incorporar* [to incorporate], *iniciar* [to start], *lligar* [to bind], *manifestar* [to manifest], *matar* [to kill], *mesurar* [to measure], *morir* [to die], *necessitar* [to need], *patir* [to suffer],

perdre [to loose], *produir* [to produce], *proporcionar* [to provide], *proposar* [to propose], *quedar* [to stay], *repetir* [to repeat], *representar* [to represent], *respondre* [to respond], *separar* [to separate], *sotmetre* [to subject], *tallar* [to cut], *unir* [to unite], *vincular* [to link]. The number of lemmas is chosen in the context of a doctoral thesis where only one person analyzes all occurrences.

⁵ Two dictionaries have been used to ensure that the chosen verbs display a variety of grammatical characteristics: *Diccionari d'ús de verbs catalans* by Ginebra and Montserrat (1999) [The Dictionary of Catalan verb uses] and *Gran diccionari de la llengua catalana* online by Enciclopèdia Catalana [Great dictionary of the Catalan language].

⁶ The list is the following: STATE (CONSTITUTIVE STATE, RELATIONAL STATE, STATIVE POSSESSION, EXISTENCE, IDENTIFICATIVE STATE, STATIVE PLACE), ASPECT (ASPECTUAL EVENT, ASPECTUAL CAUSE), PSYCHOLOGICAL EVENT (JUDGMENT, PERCEPTION, COGNITIVE EVENT, MODAL EVENT, EXPERIENTIAL EVENT), ACTION (COOPERATIVE ACTION, CAUSE OF ACTION, ACTION WITH A GOAL and COMMUNICATIVE ACTION, being the latter either EXPRESSIVE, SUGGESTIVE or COOPERATIVE), CHANGE (CHANGE OF STATE, NATURAL TRANSITION) and (CAUSE OF) CHANGE (OF CHANGE, OF VALUE, RELATIONAL, CONSTITUTIVE, OF POSSESSION, OF PLACE, OF STATE, OF NATURAL TRANSITION, CREATION). In this research, we have excluded COPULATIVE Verbs by excellence (with a high productivity on both studied domains) and METEOROLOGICAL Verbs (with high restrained realizations).

⁷ We follow the Communicative Theory of Terminology, proposed by Cabré (1999 and 2002a and 2002b), on the idea that the difference between words and terms is simply found in the use that we make of them in context, and that a linguistic unit carrying specialized knowledge works as it, according to the communicative situation in which it appears. In the line of Cabré, then, we believe in the existence of a situational context of a semantic or pragmatic type that includes factors implying the activation of the terminological value. The semantic, syntactic and linguistic context would reflect this possible activation. In this research, we assume the existence of factors directly related to the communicative situation among specialists, but we will specifically focus on the linguistic context in which occurrences appear. We only use here the direct observation of the elements indicating the activation of the terminological value in texts. Nevertheless, this does not mean that we deny the existence of this situational context where factors activating the specialized value in the communication take place.

⁸ Term proposed by Kageura and Umino (1996) and used by many authors afterwards, including Vivaldi and his research on term extraction.

⁹ In this research, we use the forms *x*, *y* and *z* when we discuss arguments individually. On the other hand, as for argument structures, we use forms such as (*x*), (*x y*) or (*x y z*), where the parentheses do not reflect externality or internality of arguments. Moreover, the selection of each symbol is made in relation to the positional/functional characteristic of each argument and not strictly considering the external and internal argument hierarchy proposed by Grimshaw (1990). We choose this option because the distinction is not always clear in several stative meanings of the studied verbs, and especially in the case of passives or impersonal passives. The externality or internality feature can be detected from the information on the thematic role of each argument in all realizations.

¹⁰ As a matter of readability, we use T1, T2 and T3 to label the different levels of termhood in arguments assigned in the Database, and according to the probability of some arguments to be acting as specialized ones. Thus, T1 [from 0.65 to 1] refers to the units being indisputably used as specialized in their context, T2 [from 0.4 to 0.64] represents the units probably working as specialized in context, and T3 [from 0 to 0.39] refers to the units not considered as transmitting specialized knowledge in context. The frontiers between T1, T2 and T3 are established after some previous tests with data from the corpus (see Joan Casademont 2008 for more information).

¹¹ English definitions are taken from the *Longman Dictionary of Contemporary English*.

¹² We could have, for example, sentences like “scientists bind a DNA chain to some adaptors” (*x y z*), “a DNA chain is bound to some adaptors” passive (*x y*), “scientists bind a DNA chain and some adaptors” active (*x y*) with pluralization or “a DNA chain and some adaptors were bound” passive (*x y*) with pluralization.
