

NEW CHALLENGES FOR LEXICAL REPRESENTATION WITHIN THE LEXICAL-CONSTRUCTIONAL MODEL (LCM)*

Ricardo Mairal Usón
UNED, Madrid
Francisco José Ruiz de Mendoza Ibáñez
Universidad de La Rioja

ABSTRACT

Within the context of the Lexical Constructional Model, this paper focuses on the notion of lexical template and argues for a more enriched and compact version of this system of lexical representation by integrating Pustejovsky's *qualia* structures. After providing a sort of a historical context that situates the origins of the notion of lexical template from the pioneering work of Van Valin and Wilkins to the more recent notion of lexical template (cf. Mairal and Cortés; Mairal and Faber, "Lexical"; Ruiz de Mendoza and Mairal, "Challenging," "Constructing"), the new formalism is discussed within the framework of the following lexical classes: change of state verbs; contact-by-impact verbs; consumption verbs and cognition verbs.

KEY WORDS: Lexical templates, logical structures, lexical classes, operators, primitives, *qualia*.

RESUMEN

En el marco del Modelo Léxico-Construccional, este trabajo se centra en la noción de plantilla léxica y tiene como objetivo potenciar un sistema de representación léxica más rico y consistente a través de la incorporación de las estructuras de *qualia* ideadas por Pustejovsky. Tras introducir en la primera parte de este artículo el contexto en el que surge la noción de plantilla léxica, desde los primeros trabajos de Van Valin and Wilkins hasta la versión más reciente de esta noción (cf. Mairal y Cortés; Mairal y Faber, "Lexical"; Ruiz de Mendoza y Mairal, "Challenging," "Constructing"), se aplica el nuevo sistema de representación propuesto con respecto a cuatro clases léxicas verbales: verbos de cambio de estado, verbos de contacto-por-impacto, verbos de consumición y verbos de cognición.

PALABRAS CLAVE: plantillas léxicas, estructuras lógicas, clases léxicas, operadores, primitivos, *qualia*.

1. INTRODUCTION

As has been extensively described in Ruiz de Mendoza and Mairal ("Constructing," "Description," "Levels") and Mairal and Ruiz de Mendoza ("Internal," "Levels"), the Lexical Constructional Model (LCM) provides a comprehensive de-

scription of the full inventory of parameters involved in meaning construction, including those that go beyond so-called core grammar (e.g. traditional implicature, illocutionary force, and discourse coherence). One of the most attractive implications of developing a fully-fledged linguistic model based on a firm and sound semantic grounding is its potential application in the field of natural language ontologies and artificial intelligence systems (Mairal and Perrián).¹

If we want to build a lexicon that meets the requirements of an intelligent search engine, we will need to enrich lexical entries with very robust semantic and pragmatic information, an area where most linguistic models have but tiptoed. Creating such rich lexical entries is not an easy enterprise. Most models either formulate representations that—even if formally impeccable—have a very limited scope (by capturing only those aspects of the meaning of a word that are grammatically relevant) or else provide more ambitious representations that include encyclopedic information but lack a rigorous formal metalanguage. In this connection, the LCM, which aspires to cover all dimensions of meaning construction, aims to develop a lexical formalism that is formally elegant (and as a consequence, can be part of a meaning-syntax linking algorithm) and at the same time is sensitive to the sort of pragmatic, semantic and discourse information that is too pervasive to be captured in a formalism. Moreover, the resulting lexical representation should also serve as input for the elaboration of the syntactic apparatus of the model, an aspect of the model that is still in progress. In this regard, a compact and sound lexical formalism that combines the set of grammatically, semantically and pragmatically relevant features of a predicate into one single representation is in fact a major achievement for the specification of syntactic configurations.

In this context, the present paper focuses on one specific aspect of the LCM, the notion of lexical template and its more recent design in terms of Pustejovsky's *qualia*, a proposal that provides a nice format prior to the elaboration of the syntax. Section 2 briefly spells out the more relevant fundamentals of the LCM. Section 3

* Financial support for this research has been provided by the DGI, Spanish Ministry of Education and Science, grants HUM-2005-02870/FILO and HUM-2007-65755/FILO. Part of the research has been co-financed through FEDER funds. For further updated information on developments of the LCM, we refer the reader to the webpage <www.lexicom.es>.

¹ The underlying rationale for this kind of account is in full harmony with the most recent research in the area of the semantic web. The aim of providing web pages with meaning has resulted in the birth of what is called the semantic web and the development of a new computational language, i.e. *Ontology Web Language* (OWL), which is a more ambitious semantic description and analysis than found in the previous *Hypertext Markup Language* (HTML). Technical details aside, OWL is primarily based on semantic tags that can allow the machine to understand the meaning of a web page. Moreover, this type of language, if correctly executed, improves the type of web searches in both monolingual and multilingual environments in such a way that the user can retrieve the exact information that is being searched for, thereby avoiding the reception of massive amounts of information with only a marginal relation to the intended search (Aguado, Montiel Ponsoda, and Ramos; Montiel-Ponsoda, Aguado, and Gómez; Montiel-Ponsoda *et al.*; Perrián-Pascual and Arcas-Túnez, "Cognitive," "Microconceptual," "Modelling," "Reusing").

concentrates on lexical representation and provides the historical context that situates the origins of the notion of lexical template from the pioneering work of Van Valin and Wilkins to the more recent notion of lexical template (Mairal and Cortés; Mairal and Faber, “Lexical”; Ruiz de Mendoza and Mairal, “Challenging,” “Constructing”). Then, section 4 presents the new formalism and discusses the format of the following lexical classes: change of state verbs (4.1.), contact-by-impact verbs (4.2.), consumption verbs (4.3) and cognition verbs (4.4). Finally, section 5 includes some concluding remarks.

2. LEVELS OF DESCRIPTION IN MEANING CONSTRUCTION

As advanced above, the LCM is intended to be operational at all levels of linguistic description, including pragmatics and discourse. Hence, a four-level catalogue of construction types—including configurations that would be regarded by other theorists as a matter of pragmatics and discourse—is postulated as part of the semantic component of the model:

Level 1: constructions producing core grammar characterizations.

Level 2: constructions accounting for heavily conventionalized situation-based low-level meaning implications.

Level 3: constructions that account for conventionalized illocutionary meaning (situation-based high-level implications).

Level 4: constructions based on very schematic discourse structures.

The LCM has a central module, the *level 1* or *argument module*, consisting of elements of syntactically relevant semantic interpretation based on the principled interaction between lexical and constructional templates. As discussed in section 3, a *lexical template* is a low-level (i.e. non-generic) semantic representation of the syntactically relevant content of a predicate; a *constructional template* is a high-level (i.e. generic or abstract) semantic representation of syntactically relevant meaning elements derived from multiple lower-level representations. Constructional templates make partial use of the same metalanguage as lexical templates since constructions capture structure that is common to a number of lexical items, as is the case of the caused-motion construction, which contains structure from multiple caused-motion predicates:

[do' (x, y)] CAUSE [BECOME *NOT be-LOC' (y, z)]

Additionally, the LCM has other more peripheral analytical tiers that contain collections of conventionalized constructions or, alternatively, low or high-level situational cognitive models that can be accessed inferentially. Thus, the LCM features a *level 2* or *implicational module* that accounts for aspects of linguistic communication that have traditionally been handled in connection with implicature theory. There is a *level 3* or *illocutionary module* dealing with traditional illocutionary



force. Finally, a *level 4* or *discourse module* addresses the discourse aspects of the LCM, with particular emphasis on cohesion and coherence phenomena. Each level is either subsumed into a higher-level constructional configuration or acts as a cue for the activation of relevant conceptual structure that yields an implicit meaning derivation.

These four different layers are interrelated by two cognitive processes: *subsumption* and *cueing*. For example, at the argument-structure level of grammar constructional templates “coerce” lexical templates, a process that is called *lexical-constructional subsumption*, which is in turn regulated by two kinds of constraints on coercion: *internal* and *external*. The former arise from the semantic properties of the lexical and constructional templates, while the latter result from the possibility or impossibility of performing high-level metaphoric and metonymic operations on the lexical items involved in the *lexical-constructional subsumption* process. Internal constraints specify the conditions under which a lexical template may modify its internal configuration. For example, the *lexical class constraint* explains why ‘break’ verbs may take part in the causative/inchoative alternation (cf. *The child broke the window* and *The window broke*), while ‘destroy’ verbs may not. The reason is that ‘destroy’ verbs belong to the lexical class of ‘existence’ verbs, while ‘break’ verbs are verbs of ‘change of state’.

As an example of external constraint, consider the conversion of ‘laugh (at)’, an activity predicate, into a causative accomplishment predicate when taking part in the caused-motion construction: *They laughed him out of the room*. This is possible because of the correlation between two kinds of actor and two kinds of object. In the case of causative accomplishments, the actor and object are an *effector* and an *effectee* respectively. The effector is an actor whose action has a direct impact and subsequent effects on the object or effectee. With activities, the actor is a mere ‘doer’ of the action experienced by the object. This observation suggests an analysis of the subcategorical conversion process experienced by ‘laugh’ in terms of source and target domain correspondences (EXPERIENTIAL ACTION IS EFFECTUAL ACTION), of the kind proposed in Cognitive Linguistics (Lakoff).

At the pragmatic and discourse levels, *subsumption* takes the form of parametrization processes of the variable elements of non-argument constructions, which differ from level-1 constructions in that they are essentially *idiomatic* in nature, i.e. they consist of a combination of fixed and variable elements. A case in point is the level-2 *What’s X Doing Y?* configuration (first studied in detail by Kay and Fillmore), which conveys the idea that the state of affairs denoted by the non-interrogative content of the sentence is either incongruent or bothers the speaker (e.g. *What’s the child doing in the swimming pool?*). The construction has fixed elements that cannot be changed without altering its meaning implications (e.g. verb tense, cf. *What will the child do in the swimming pool?*) and variable elements that can be parametrized in a constrained way. For example, the X variable in the level-3 requestive *Can You X?* construction must contain a predicate that expresses the addressee’s control of the state of affairs (cf. *Can you close the window?* vs. *Can you see the window?*). In a similar way, the level-4 construction *Just Because X Doesn’t Mean Y* (e.g. *Just because we live in Berkeley doesn’t mean we’re left wing radicals*) is

used to indicate that the content of Y does not necessarily follow from X (cf. Holmes and Hudson).

Finally, *cueing* or *cued inferencing* is a form of constraining non-explicit meaning on the basis of lexical and constructional clues. It takes place at all levels of meaning construction as an alternative to subsumption. Thus, at the level of core grammar, it accounts for inferences obtained by making contextual adjustments on the meaning of some predicates (e.g. *He drinks [alcohol]*; *She's ready [for the party]*). At other levels it accounts for meaning implications based on potential conceptual connections between propositions (the case of discourse), or on metonymic activations or high-level (for illocution), and low-level (for implicature) situational models or scenarios. For example, the discourse connection between *It can't sound good; it's not digital*, which is one of conclusion-evidence, differs from the connection between *It doesn't sound good; it's not digital*, which is simply of cause-effect. The difference lies in the use of *can't* indicating (i.e. cueing) a deduced impossibility in the case of the conclusion-evidence pattern.

From this brief description two methodological issues are in order here: the question of the ubiquity of cognitive processes and the existence of continua between linguistic phenomena. In relation to the first issue, one of the relevant methodological features of the LCM is what has been termed the *equipollent hypothesis*, whereby all levels of linguistic description and explanation are postulated to make use of the same or at least comparable cognitive processes (Ruiz de Mendoza). For example, as commented above, cognitive processes such as generalization or parametrization as well as inferential activity, or *cued inferencing*, not only operate at a discourse and pragmatic level but are also influential in the argument structure level of grammar. The same can be said of idiomaticity which is an active process that not only refers to the lexicon but also functions constructionally at all levels of description. In previous work (Ruiz de Mendoza and Mairal, "High-level") metaphor and metonymy have been likewise found to be present not only at the lexical level of description, but also at the level of pragmatic implications, illocutionary meaning, and even syntactic alternations (Ruiz de Mendoza; Ruiz de Mendoza and Pérez).

In relation to the possible existence of continua between linguistic categories, which is a central claim of Cognitive Linguistics (Langacker, *Cognitive Foundations, Grammar*), the LCM takes no special stance on this issue. While it recognizes that such continua exist, the LCM regards them as epiphenomena arising from the intrinsic nature of the categories in question. The LCM focuses on the representational adequacy of each level in the model and on the principles that constrain interaction between representations from different levels. Thus, the model has lexical templates, which partially resemble constructional templates and interact with the latter in predictable ways. The output of this interaction is a level-1 representation that can be made part of higher-level representations by realizing their non-idiomatic (i.e. variable) components. For example, *Can you clean the kitchen, please?* has 'clean the kitchen' as a level-1 component that realizes the Y variable in the level-3 idiomatic configuration *Can You X, please?* This process is fully predictable on the basis of constraining factors such as coercion of the level-3 on the level-1 configuration (*Can You X, please?* is incompatible with level-1 out-



puts denoting states and non-active accomplishments, as in *Can you be tall, please?*, *Can you own the car, please?*). There is no special reason to give pride of place to the lexicon-grammar continuum in this meaning construction account: content-carrying lexical items are represented in the form of lexical templates that are related to other such items through various kinds of relations, among which lexical-class ascription figures prominently. Argument constructions, like lexical templates, are the result of abstracting conceptual material away from lexical items. Caused-motion, for example, is the result of finding structure that predicates like *push*, *pull*, *shove*, and others have in common: in all of them there is force causing an object to change from one location to another. Obviously, there is no such thing as a continuum from these predicates to the caused-motion construction, but simply an abstraction operation that allows us to create a higher-level construct that may be useful as a meaning construction factor. This is clearly evidenced by the coerced uses of the construction with verbs that do not match the construction in terms of their basic meaning structure (e.g. *laugh*, *listen*) so long as it is possible to find a licensing factor (in this case the high-level metaphor from effectual action to other forms of goal-oriented action).

In much the same way, it is unnecessary to postulate a pragmatics-semantics continuum. What we have is the possibility of constructing meaning representations that go beyond the argument level on the basis of inferential activity or on the basis of constructional interaction, or by combining both processes. Thus, we may have inferential activity based on the linguistic expression providing partial access to low-level situational models (traditional implicature), or to high-level situational models (traditional illocutionary force), or to discourse coherence patterns. Alternatively, we can often derive comparable meaning implications by grammatical means on the basis of levels 2, 3, and 4 constructional realization. A person can ask for a glass of water by saying *I'm thirsty* or *Can you give me some water, please?* The reasons to use one way or another are a matter of communication strategies, but what matters is that we have two alternative ways, with slightly different meaning effects, and there is no need to postulate a continuum from one to the other.

3. LEXICAL REPRESENTATION IN THE LCM: NEW CHALLENGES

The LCM uses lexical templates for the lexical representation of relational predicates. As discussed in Mairal and Ruiz de Mendoza ("Levels"), this notion is an alternative form of lexical representation that integrates relevant elements from both decompositional and frame-based proposals. Lexical templates are thus a development of the logical structures (LS) postulated in Role and Reference Grammar (RRG) (cf. Van Valin; Van Valin and LaPolla).

Let us firstly contextualize this proposal within the context of RRG and then spell out the specific details of the internal anatomy of a lexical template as well as the more recent proposal that suggests a reorientation of the notational device in terms of Pustejovsky's *qualia* structures.

3.1. TOWARDS A FINER SEMANTIC DECOMPOSITION IN THE PRIMARY LEXICON

RRG uses a decompositional system for representing the semantic and argument structure of verbs and other predicates (their *Logical Structure*, LS). The verb class adscription system is based on the *Aktionsart* distinctions proposed in Vendler, and the decompositional system is a variant of the one proposed in Dowty. Verb classes are divided into *states*, *activities*, *achievements*, *semelfactives*, and *accomplishments*, together with their corresponding causatives. Here is a representation of each verb class with their corresponding formalism (cf. Van Valin 45):

VERB CLASS	LOGICAL STRUCTURE	EXAMPLE	INSTANTIATION OF LS
State	predicate' (x) or (x,y)	see	see' (x,y)
Activity	do' (x, [predicate' (x) or (x,y)])	run	do' (x, [run' (x)])
Achievement	INGR predicate' (x) or (x,y), or INGR do' (x, [predicate' (x) or (x,y)])	pop (burst into tears)	INGR popped' (x)
Semelfactive	SEML predicate' (x) or (x,y) SEML do' (x, [predicate' (x) or (x,y)])	glimpse, cough	SEML see' (x,y)
Accomplishment	BECOME predicate' (x) or (x,y), or BECOME do' (x, [predicate' (x) or (x,y)])	receive	BECOME have' (x,y)
Active accomplishment	do' (x, [predicate' ₁ (x, (y))] and BECOME predicate' ₂ (z,x) or (y))	drink	do' (x, [drink' (x,y)]) and BECOME consumed' (y)
Causative accomplishment	α CAUSES β where α , β are LS of any type	kill	[do' (x, \emptyset)] CAUSE [BECOME [dead' (y)]]

RRG maintains that states and activities are primitives and thus form part of the logical representation of the rest of predicates; by way of example, an accomplishment is either a state or activity predicate modified by the telic operator BECOME. However, in Van Valin and Wilkins, and Van Valin and LaPolla we find the explicit claim that state and activity atomic predicates need further semantic decomposition and thus provide a first approach for the predicate *remember* and speech act verbs respectively. Here is the format of these two first representations:

remember (Van Valin and Wilkins 511)
BECOME think.again (x) **about something.be.in.mind.from.before** (y)

Speech act verbs (Van Valin and La Polla 117)
do' (x, [**express**(α).**to**(β).**in.language**.(γ)' (x,y)])

In these representations, event structures are enriched by the addition of a number of internal variables marked in Greek letters. These internal variables spell

out the exact semantic parameters that are operative within a lexical class and are bound to an *external or argument variable* pertaining to the eventive or logical structure of the item in question.

Further work on this area was extended to some other lexical classes: manner of cutting verbs, *break* verbs, consumption, *contact-by-impact*, cognition verbs, to name just a few (Mairal; Mairal and Faber, “Functional,” “Lexical”; Ruiz de Mendoza and Mairal, “Challenging,” “Levels”). Here is a sampled representation of some of these predicates:

Contact-by-impact verbs

[[do' (w, [use.tool.(α).in.(β).manner.for.(δ)'] (w, x)) CAUSE [do' (x, [move.toward' (x, y) and INGR be.in.contact.with' (y, x)], α = x.

Consumption verbs

do' (x, [CAUSE.BECOME.be-in'.([have.as.part'.(x, mouth)], α).in.(β). Manner' (x,y)) and BECOME consumed' (y) α = y

Causative change of state verbs

[[do' (x, [use' (x, y)) CAUSE [do' (y, Ø)] CAUSE [BECOME/INGR pred' (z)]]

These representations follow the same format such that the corresponding logical structure is enriched by a set of internal variables that express the relevant semantic parameters in a predicate meaning definition: for example, in the case of contact-by-impact verbs instrument (**use.tool.(α)**), manner (**in.(β).manner**) and purpose (**for.(δ)**) are the semantic parameters that permeate the lexical encoding of this class in English.

We still understand that these logical structures can be built on the basis of a universal semantic metalanguage or a set of indefinables. Doing so allows the analyst to avoid the problem of having to regard as undefinable predicates which can be further semantically decomposed, for example, defining the predicate *red* in terms of BECOME **red'**, or *popped* in terms of INGR **popped'**, or activity predicates like *sing* or *drink* in terms of **do'** (x,[**drink'**(x)]) or **do'** (x,[**sing'**(x)]). The innovation here with respect to the original RRG proposal resides in finding a systematic procedure to identify the correct prime together with a uniform framework for decomposing semantically every predicate until we arrive at the undefinable elements.

With this in mind, we introduced a new formalism that draws insights from Wierzbicka's Natural Semantic Metalanguage (NSM) (Goddard and Wierzbicka), Mel'cuk's Text-Meaning Theory (MTT) (Mel'cuk; Mel'cuk, Clas, and Polguère; Mel'cuk and Wanner), and the Functional-Lexematic Model (FLM) (Martín Mingorance, “Functional,” “Lexical”; Faber and Mairal, *Constructing*).²

² For a full description of the exact details of lexical templates based on a universal semantic metalanguage, we refer the reader to Mairal and Faber (“Functional,” “Lexical”) and the references and works posted on the LEXICOM webpage <www.lexicom.es>.

3.2. LEXICAL TEMPLATES

Despite the fact that the representations above involved more elaborate semantic decompositions, these first lexical templates were still not systematic enough in their use of activity and state primitives. Primitives such as *manner*, *tool* and *use* appear in these representations, but again no explanation is given of how they have been obtained. Moreover, we noted that the resulting representations turned out to be too unwieldy and lacked transparency and elegance in the expression.

Consequently, we decided to simplify the system by postulating two different modules both of which were based on a universal abstract semantic metalanguage. The resulting templates have two parts: (i) the semantic module, and (ii) the logical representation or *Aktionsart* module, each of which is encoded differently. Here is the basic representational format for a lexical template:

predicate: [SEMANTIC MODULE<lexical functions>] [AKTIONSART MODULE <semantic primes>]

The rightmost hand part of the representation includes the inventory of logical structures as developed in RRG with the proviso that the predicates used as part of the meaning definition are putatively candidates for semantic primes, or else, these cannot be further decomposed.

The semantic and pragmatic properties of the semantic module, as shown in the leftmost hand part of the representation, are formalized by making use of lexical functions such as those used in Mel'cuk's Explanatory and Combinatorial Lexicology (ELC) (Mel'cuk; Mel'cuk, Clas, and Polguère; Mel'cuk and Wanner; Alonso Ramos).³ These lexical functions have also been shown to have a universal status (Mel'cuk), something which is in keeping with our aim of providing typologically valid representations. Unlike what is the case in Mel'cuk's work and the complete literature on the *Explanatory Combinatorial Dictionary*, in our approach, lexical functions are essentially paradigmatic and capture those pragmatic and semantic parameters that are idiosyncratic to the meaning of a word, which allows us to distinguish one word off from others within the same lexical hierarchy. For example, if we want to account for the semantic differences between *mandar* ('command'), *ordenar* ('order'), *decretar* ('decree'),

³ According to Mel'cuk, Clas, and Polguère (126-127), a lexical function (LF) is written as: $f(x) = y$, where f represents the function, x , the argument, and y , the value expressed by the function when applied to a given argument. The meaning associated with an LF is abstract and general and can produce a relatively high number of values; e.g. **Magn** expresses intensification and can be applied to different lexical units thus yielding a high set of values:

Magn	(Engl. smoker)	=	heavy
Magn	(Engl. bachelor)	=	confirmed
Magn	(Sp. error)	=	craso
Magn	(Sp. llorar)	=	llorar como una Magdalena

preceptuar ('set up a precept'), *preinscribir* ('preregister'), from the lexical domain of speech acts, or *cautivar* ('captivate'), *arrebatar* ('seize'), *arrobar* ('entrance'), *embelesar* ('enrapture'), *extasiar* ('send into an ecstasy'), *hechizar* ('bewitch'), from the domain of feeling in Spanish, we would certainly need some mechanism that allows us to discriminate and encode those meaning elements that differentiate one predicate from others. Then, we have devised a semantic module that consists of a number of internal variables, i.e. world knowledge elements of semantic structure, which relate in very specific ways to the external variables that account for those arguments that have a grammatical impact. Now, let us consider the following examples:

fathom: [MAGNOBSTR and CULM_{1,2[all]}] **know'** (x, y)
 x = 1; y = 2

This predicate is a hyponym of *understand* and inherits all the properties from its superordinate, that is, it designates a state structure with a primitive predicate **know'** modified by two arguments (x,y). As an additional distinguishing parameter this predicate encodes two lexical functions that express the culmination of the process of knowing something [CULM_{1,2[all]}] and the great difficulty involved in this process [MAGNOBSTR]. If we move on to the domain of Speech Act verbs, a predicate like *command*, as a hyponym of *order*, inherits all its properties and adds its own specificity which lies in the political/military context:

command:

<MAGN_{1[PERM]23}, LOC^{soct}₍₁₎ (PLACE_TYPE: political/military)> [**do'** (x, [say' (x,y)])] CAUSE
 [**do'** (y,z)]

The subscripts (_{1, 2, 3}) codify the speaker, auditory percept, and the addressee, respectively. MAGN specifies that the action is intensified to a very high degree, thus making it more forceful, and PERM, applied to the first argument, indicates that the speaker has power over the addressee and is licensed to ask him/her to do things. As for the *aktionsart* module, this verb designates a causative accomplishment structure that is induced by an activity such that *x* says something to *y* and this causes *y* to do *z*.

In sum, lexical templates provide enhanced semantic representation and consequently allow us to account for those properties which go beyond those aspects of the meaning of a word that are grammatically relevant. However, we believe that the formalism can be improved if we manage to find a system where both external and internal variables are easily integrated with a view to accounting for syntactic projection. In connection with this, we claim that lexical templates can be enriched by unifying internal and external variables within a system that allows the expression of both.

4. LEXICAL TEMPLATES REVISITED

One of the possible applications of the lexical architecture of the LCM is to build a knowledge base that allows us to retrieve contextual and pragmatic infor-



mation by means of a set of inferencing mechanisms. This goal requires some adaptations of the notational devices described in the previous section, which have a clear lexicological orientation, to make them compatible with computational needs. Simplifying a bit, this initial computational move involves reconverting our lexicon into a knowledge base linked to an ontology and to make the connections between the external and the internal variables more explicit. Recent research has evidenced the insufficient explanatory coverage shown by the inventory of lexical functions to account for the full gamut of semantic parameters that operate in the lexicon. In an attempt to simplify the formalism in order to avoid the sometimes *ad hoc* adscription of a lexical function to a semantic parameter, Mairal and Cortés have recently initiated a reconversion of the inventory of lexical functions by incorporating features from Pustejovsky's generative lexicon and more in particular to the set of *qualia*, which we reproduce here for convenience (Pustejovsky 76; 85-86):⁴

- CONSTITUTIVE (Q_C): the relation between an object and its constituent parts
 - i. material
 - ii. weight
 - iii. parts and component elements

- FORMAL (Q_F): that which distinguishes it within a larger domain
 - i. orientation
 - ii. magnitude
 - iii. shape
 - iv. dimensionality
 - v. color
 - vi. position

- TELIC (Q_T): its purpose and function
 - i. purpose that an agent has in performing an act
 - ii. built-in function or aim which specifies certain activities

- AGENTIVE (Q_A): factors involved in its origin or 'bringing it about'
 - i. creator
 - ii. artifact
 - iii. natural kind
 - iv. causal chain

⁴ Pustejovsky's (Ch. 5) generative lexicon includes four levels of representation: (i) argument structure; (ii) event structure; (iii) *qualia* structure and (iv) lexical inheritance structure, together with a complete set of generative devices (e.g. type coercion, selective binding, co-composition) that connect up the four levels. In the present paper, we focus on how *qualia* configurations serve the same purpose as the lexical functions in the semantic module. Unfortunately work on the notion of *qualia* structures has, to the best of our knowledge, been discontinued. We believe that the inventory of such configurations, as it stands, is far from exhaustive and a fined-grained description would certainly be welcome.



The following are examples of lexical representations based on this system (Pustejovsky, *Generative* 82, 101), although we have slightly changed some of Pustejovsky's notational devices and have adapted them to a system that is closer to ours:

book

ARGSTR = [ARG1 = **x: information**
[ARG2 = **y: phys_obj**]

QUALIA = **information-phys_obj_lcp**
FORMAL = **hold** (*y,x*)
TELIC = **read** (*e,w,x,y*)
AGENT = **write** (*e', v, x,y*)

This representation specifies that the nominal predicate *book* belongs to the lexical conceptual paradigm (*lc*) of physical objects and accounts for the telic and agentive interpretations that make reference to the dotted arguments (*x* and *y*), which are in turn featured as 'information' and 'physical object'. Now, consider a more complex representation:

build

EVENTSTR = [E₁ = **e₁: process**
E₂ = **e₂: state**
RESTR = <_α
HEAD = e₁

ARGSTR = [ARG1 = **x: animate_ind**
FORMAL = **phys_obj**]
[ARG2 = **y: artifact**
CONST = **z**
FORMAL = **phys_obj**]
[D-ARG = **z: material**
FORMAL = **mass**]

QUALIA = **create-lcp**
FORMAL = **exist** (*e₂, y*)
AGENTIVE = **build_act** (*e₁, x, z*)

This representation specifies the event, argument and *qualia* structures of the predicate *build*. The event structure indicates that the verb *build* is an accomplishment verb that involves a process and a result state ordered by the relation "exhaustive ordered part of," <_α. The initial event has been headed, which means that the action that brings about the state is fore-grounded. In relation to the argument structure, there are two true arguments (i.e. those that are syntactically realized) and a default argument (parameters that are relevant for the *qualia* but are not syntactically realized). In the *qualia* structure, the lexical conceptual paradigm is also noted, i.e. *build* is a creation verb, as well as the two processes involved: the agentive, involving both argument 1 and the default argument, which is related to

the object by the constitutive relation of argument 2. The formal role indicates the final result state (Pustejovsky 63; 71-73; 82).

Both representations include an event structure description —which, details aside, coincides to a large extent with the *Aktionsart* module— and a *qualia* structure, whose function is to specify the specific semantic properties of each of the arguments involved in the event. Interestingly enough, a brief mention to the lexical class is also included, which happens to be one of the hallmarks in our approach.

An added advantage in adopting this new formalism is that the two modules —the semantic and the eventive— are closely intertwined and the resulting lexical templates are an eloquent proof of it, as shown in the next section. This has interesting consequences in the semantics-to-syntax mapping possibilities of a predicate since, as pointed out in Pustejovsky (101-104), individual *qualia* compete for projection, and there are mechanisms such as foregrounding or ‘focalizing’ a single *quale* of the verbal semantic representation. To illustrate this, consider the lexical template for the causative change of state verb *break*:

break:

EVENTSTR: $\text{do}'(x, \emptyset)$ CAUSE [BECOME/ INGR *broken'*(y)]

QUALIASTR: { Q_f : *broken'*(y)}

Q_A : $\text{do}'(x, \textit{break_act})$ }

Change-of-state verbs typically involve an initial activity followed by a resulting state. These two phases in the causative structure map onto the agentive and the formal *qualia* respectively.⁵ Depending on which *quale* is fore-grounded (‘headed’ in Pustejovsky’s terminology) the verb can be constructed in a transitive (causative) or an intransitive (anticausative) structure. Foregrounding is in fact the effect of the cognitive operations that act as external constraints in our model.

Let us see how we can reconvert the lexical templates for change of state, contact-by-impact, consumption and cognition verbs by using *qualia*, lexical functions and event structures (Mairal and Cortés).

4.1. CHANGE OF STATE VERBS

EVENTSTR: [$\text{do}'(x, e_1)$]_{E1} CAUSE [BECOME/INGR $\text{pred}'(y)$]_{E2}

QUALIASTR: { Q_f : MANNER $\text{pred}'(y)$ }

Q_A : e_1 : **Oper** x, z <Instr>}

As mentioned before, change-of-state verbs (e.g. *break*, *smash*) are causative telic predicates; their event structure involves an activity and a final resulting state

⁵ As can be seen, the information in the *quale* is often redundant as it tends to identify itself with the eventive description of logical structures, unless some specification is added.

modified by a telic operator (BECOME or INGR). Each of these subevents maps onto one *quale*: the state predicate is part of the formal *qualia* characterization of all change of state verbs. In fact, the semantic specificities of each predicate within the class will derive from the specific ontological values that the semantic function MANNER will receive. Thus, *smash*, *break* and *shatter* are semantically distinct as they encode different aspects of the ‘affectedness’ effect on the patient argument (y).⁶ The causing activity event maps onto the agentive *quale* as it expresses what is carried out by the effector argument (x) in order to make the resulting state come about. In this regard, the agent *quale* in the template includes a subevent (e₁) that depicts the use of an implement (z) by the effector (x); the formalized expression of the manipulation subevent is: **Oper** x, z <Instr>. The lexical function **Oper** is described by Alonso Ramos as a semantically empty verb that will have different values depending on its arguments. In other words, the specific nature of the object (z) that will be used as instrument will provide the exact content to the manipulation event; if a stone is used to break a glass, then **Oper** will stand for, say, *throw*. If (z) is to be a hammer, the value of **Oper** is most probably *hit*.

4.2. CONTACT-BY-IMPACT VERBS

We can now easily reformulate the lexical template proposed for this verb class in section 3.1 in the following terms:

EVENTSTR: [do' (x, e₁ < **Oper** e₂)]_{E1} CAUSE [INGR **touching'** (z, y)]_{E2}
 QUALIASTR: {Q_F: MANNER: MagnE1
 Q_A: e₁: **Oper** x, y <Instr>
 e₂: **move**^{adp} (z, y)
 Q_T: E₂}

The case of verbs of contact-by-impact is also very interesting. Because of the effects of hitting in the extralinguistic world, it is only natural to presuppose that ‘hit’ verbs lexicalize a change of state that affects the entity receiving the impact. In fact, their semantics encodes causative structures with a final locative component. However, ‘hit’ verbs are in essence verbs of contact and as such they integrate as a final subevent in their logical structure a stative contiguous—and not a result—location event. The relation of contiguous location that holds between a location argument (z) and a theme (y) is encoded by means of Wierzbicka’s prime **touching'**.

The *qualia* structure of these predicates is also very complex. The formal *quale* specifies the nature of the causing event; i.e. it is bound to the activity subevent and modifies it by one specific value of the MANNER operator. The intensifier lexical function ‘Magn’ (‘very’ ‘intense’) restricts the semantics of this class to those states

⁶ The other parameter that triggers semantic differences within the members of the class is duration: change of state verbs are either punctual (e.g. *shatter*) or durative (e.g. *break*).

of affairs where contact takes place by means of an impact. This semantic parameter sets the lexical class of ‘hit’ verbs apart from other semantically related predicates such as the class of ‘stroke’ verbs or contact-by-motion verbs like *join*, *link*, *unite*, etc. The Agent Quale combines to subevents e_1 and e_2 : in bringing about the contact event the effector (x) may use an implement (y) (the manipulation subevent e_1) so that it will displace itself towards the entity (z) that eventually will receive the impact. Such a motion subevent is encoded by another prime **move** modified by Melcuk’s lexical function **ad** (‘towards’). The temporal sequence between both subevents is encoded in the template by means of the relation $e_1 < oe_2$. It expresses the partial sequential overlap between the manipulation and the displacement of the instrument (y). Either of the two subevents may be foregrounded or ‘headed’ for its projection onto syntax. When e_1 is given more prominence (i.e. is headed) ‘hit’ verbs will allow its insertion in an instrument-as-subject construction. When, on the other hand, the event headed is e_2 the verbs will appear in a conative structure. The Telic Quale corresponds to the caused location subevent in the event structure characterization.

4.3. CONSUMPTION VERBS

Consumption verbs like *eat*, *drink*, *imbibe*, *gulp*, etc. are activities and therefore are not telic. Nevertheless, consumption verbs display an interesting behavior as regards telicity: they can become telic predicates if their second argument is referential; i.e. if it has a discourse referent, as in *Mario is drinking a can of beer*. Compare it with *Mario drinks beer daily* in which *beer* does not refer to a specific participant, but rather serves to characterize the nature of the action. The referential nature of the second argument causes a shift in the aspectual characterization of consumption verbs and renders them telic. RRG treats them as active accomplishments. Their semantic representation would be:

EVENTSTR: [do' (x, e_1)]_{E1} and [INGR NOT exist' (y)]_{E2}, E1 < E2
 QUALIASTR: {Q_F: MANNER E1
 Q_S: e: do' (x, [CAUSE.BECOME.LOCⁱⁿ '(part_of'x, y)])
 Q_T: E2}

The event structure encodes an activity and a subsequent achievement existential subevent. Again, the nature of the initial activity is specified in the agentive *quale*: consumption verbs involve a causal chain in terms of which the consumer-effector (x) places the affected entity (y) within a part of its body. Recall that there is also a formal *quale* that specifies the different manners of consuming something. The relation between this causal complex of events and the final telic state is encoded as an underspecified relation E1<E2. This relation involves firstly the existence of an ordered sequence between both events (in fact the symbol and is to be read as ‘and then’) and, secondly, either of the two events must be headed (i.e. foregrounded) for the semantics-to-syntax mapping. That is, underspecification involves



verbal polysemy (Pustejovsky 73-74). Headless event structures can have two possible interpretations. In the specific case of consumption verbs, if the activity event (i.e. the event encoded in the Agent *quale*) is headed (*E1 < E2), the non-telic interpretation will be selected for syntactic projection. If headedness falls upon the second subevent (the Telic *quale*, E1 < *E2) the verb must be interpreted as an active accomplishment and its syntactic behavior will vary in accordance with this feature.

4.4. COGNITION VERBS

We are going to focus on the lexical subdomain that expresses knowledge acquisition. If we look back at the representation for the predicate *fathom* above, we could rewrite the semantic module in the template as follows:

fathom:
 EVENTSTR: **know'** (x, y)
 QUALIASTR: {Q_F: MANNER : MagnObstr **think'** (x, y)
 Q_T: Culm **know'** (x,y <ALL>)}

This new format is expressed in terms of two *qualia*: the *formal* and the *telic*. The formal *quale* describes the great difficulty involved in carrying out the process of thinking, i.e. it includes the semantic attributes by means of which *fathom* is semantically distinguished within the larger set of cognition predicates in English. The telic, as encoded in Q_T: Culm **know'** (x,y), specifies the culmination of the process of acquisition of knowledge, that is, the final process of understanding something.

Another interesting example from the cognition domain is the template for *realize*:

realize:
 EVENTSTR: **know'** (x, y)
 QUALIASTR: {Q_A: LOCin (**body_part**: **mind**, **see'** (x, y))
 Q_T: Culm **know'** (x,y <ALL>)}

Realize is also a verb that involves 'understanding' (as encoded in the formal *quale* {Culm **know'** (x,y)}). Furthermore, the cognizer achieves understanding by seeing the mental percept in its mind conceptualized as a location (Mairal and Faber, "Lexical"). The mind is represented as an abstract body-part, which means it is in a partitive relation to *body*. The whole perception subevent is encoded as an agentive *quale* as it is the kind of action carried out in order to obtain knowledge.

5. CONCLUSIONS

Within the context of the LCM, this paper argues for a more enriched and compact version of the notion of lexical templates by integrating Pustejovsky's *qualia* structures. It has been noticed that it is not always easy to find a lexical function



that gives expression to some of the semantic and pragmatic parameters that are operative within a lexical class. Moreover, within a lexical template, internal and external variables do not always communicate with one another as they should. Then, in our attempt at developing the syntax of the model, together with a first computational version of it, we have noted that it would be desirable to develop a system of lexical representation such that the two modules in the new formalism — the semantic and the eventive — are closely interrelated and therefore the relation between internal and external variables is easily accounted for. Several cases from different lexical domains have been discussed in this respect. We finally sustain that a complementary advantage is the greater power of this new lexical formalism for syntactic projection.

WORKS CITED

- AGUADO DE CEA, Guadalupe, Elena MONTIEL PONSODA and José Ángel RAMOS GARGANTILLA. "Multilingualidad en una aplicación basada en el conocimiento." *Revista de procesamiento de lenguaje natural* 38 (Abril 2007): 77-97.
- ALONSO RAMOS, Margarita. "Colocaciones y contorno en la definición lexicográfica." *Lingüística Española Actual* 24.1 (2002): 63-96.
- CORTÉS RODRÍGUEZ, Francisco José. "The English Constructicon." Diss. Universidad de La Laguna, 2007.
- DOWTY, David R. *Word Meaning and Montague Grammar*. Dordrecht: Foris, 1979.
- FABER, Pamela and Ricardo MAIRAL. *Constructing a Lexicon of English Verbs*. Berlin: Mouton de Gruyter, 1999.
- GODDARD, Cliff and Anna WIERZBICKA, eds. *Meaning and Universal Grammar*. Amsterdam: John Benjamins, 2002.
- GOLDBERG, Adele. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford UP, 2005.
- GONZÁLEZ-GARCÍA, Francisco and Christopher S. BUTLER. "Mapping Functional-Cognitive Space." *Annual Review of Cognitive Linguistics* 5 (2006): 39-96.
- HOLMES, Jasper and Richard HUDSON. "Just Because X Doesn't Mean Y." Paper delivered at the Linguistics Association of Great Britain, Spring Meeting. University College London, 2000.
- KAY, Paul and Charles J. FILLMORE. "Grammatical Constructions and Linguistic Generalizations: The 'What's X doing Y' Construction." *Language* 75 (1999): 1-33.
- LAKOFF, George. "The Contemporary Theory of Metaphor." *Metaphor and Thought*. Ed. Andrew Ortony. Cambridge: Cambridge UP, 1993. 202-251.
- LANGACKER, Ronald W. *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford UP, 2008.
- . *Foundations of Cognitive Grammar. Volume 1: Theoretical Prerequisites*. Stanford: Stanford UP, 1987.
- . *Grammar and Conceptualization*. Berlin: Mouton de Gruyter, 1999.
- LEVIN, Beth, and Malka RAPPAPORT HOVAV. *Argument Realization, Research Surveys in Linguistics Series*. Cambridge: Cambridge UP, 2005.

- MAIRAL USÓN, Ricardo. "Reconsidering Lexical Representations in Role and Reference Grammar." *Proceedings of the XXVII International AEDEAN Conference*. CD-ROM. Salamanca: Universidad de Salamanca, 2004.
- MAIRAL USÓN, Ricardo, and Francisco José CORTÉS RODRÍGUEZ "Revising Semantic Representation in Role and Reference Grammar." [Forthcoming].
- MAIRAL USÓN, Ricardo, and Pamela FABER. "Functional Grammar and Lexical Templates." *New Perspectives on Argument Structure in Functional Grammar*. Ed. Ricardo Mairal Usón and María J. Pérez Quintero. Berlin: Mouton de Gruyter, 2002. 41-98.
- . "Lexical Templates within a Functional Cognitive Theory of Meaning." *Annual Review of Cognitive Linguistics* 5 (2007): 137-172.
- MAIRAL USÓN, Ricardo, and Carlos PERIÑÁN-PASCUAL. "First Steps towards the Computational Implementation of the LCM." [Forthcoming].
- MAIRAL USÓN, Ricardo, and Francisco José RUIZ DE MENDOZA IBÁÑEZ. "Internal and External Constraints in Meaning Construction: The Lexicon-Grammar Continuum." *Estudios de Filología Inglesa: Homenaje a la Dra. Asunción Alba Pelayo*. Madrid: UNED, 2008. [Forthcoming].
- . "Levels of Description and Explanation in Meaning Construction." *Deconstructing Constructions*. Ed. Christopher S. Butler and Javier Martín Arista. Amsterdam: John Benjamins. [Forthcoming].
- MARTÍN MINGORANCE, Leocadio. "Functional Grammar and Lexematics." *Meaning and Lexicography*. Ed. Jerzy Tomaszczyk and Barbara Lewandowska-Tomaszczyk. Amsterdam: John Benjamins, 1990. 227-253.
- . "Lexical Logic and Structural semantics: Methodological Underpinnings in the Structuring of a Lexical Database for Natural Language Processing." *Panorama der Lexikalischen Semantik*. Ed. Ulrich Hoinkes. Tübingen: Gunter Narr, 1995. 461-474.
- MEL'CUK, Igor. "Semantic Primitives from the Viewpoint of the Meaning-Text Linguistic Theory." *Quaderni di Semantica* 10-1 (1989): 65-102
- MEL'CUK, Igor, André CLAS, and Alain POLGUERE. *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot, 1995.
- MEL'CUK, Igor, and Leo WANNER. "Lexical Functions and Lexical Inheritance for Emotion Lexemes in German." *Recent Trends in Meaning-Text Theory*. Ed. Leo Wanner. Amsterdam: John Benjamins, 1996. 209-227.
- MICHAELIS, Laura. "Word Meaning, Sentence Meaning, and Syntactic Meaning." *Cognitive Approaches to Lexical Semantics*. Ed. Hubert Cuykens, René Dirven and John R. Taylor. Berlin: Mouton de Gruyter, 2003. 93-122.
- MONTIEL PONSODA, Elena, Guadalupe AGUADO DE CEA, and Asunción GÓMEZ PÉREZ. "Localizing Ontologies in OWL." *Workshop at the 6th International Web Conference: From Text to Knowledge*. November, 2007. Busan Korea, 13-22.
- MONTIEL PONSODA, Elena, Guadalupe AGUADO DE CEA, M. Carmen SUÁREZ FIGUEROA, Raúl PALMA, Asunción GÓMEZ PÉREZ, and Win PETERS "LexOMV: An OMV Extension to Capture Multilinguality." *Workshop at the 6th International Web Conference. From Text to knowledge*. November, 2007. Busan Korea. 118-127.
- PERIÑÁN-PASCUAL, Carlos, and Francisco ARCAS-TÚNEZ. "Cognitive Modules of an NLP Knowledge Base for Language Understanding," *Procesamiento del Lenguaje Natural* 39 (2007): 197-204.



- . “Microconceptual-knowledge Spreading in FunGramKB.” *Proceedings on the 9th IASTED International Conference on Artificial Intelligence and Soft Computing*. Anaheim: ACTA, 2005. 239-244.
- . “Modelling OLIF Frame with EAGLES/ISLE Specifications: An Interlingual Approach.” *Procesamiento del Lenguaje Natural* 40 (2008): 9-16.
- . “Reusing Computer-Oriented Lexica as Foreign-Language Electronic Dictionaries.” *Anglogermánica Online* 4 (2006): 69-93.
- PUSTEJOVSKY, James. *The Generative Lexicon*. Cambridge: MIT P, 1995.
- RUIZ DE MENDOZA IBÁÑEZ, Francisco José. “High Level Cognitive Models: In Search of a Unified Framework for Inferential and Grammatical Behavior.” *Perspectives on Metonymy*. Ed. Krzysztof Kosecki. Frankfurt Main: Peter Lang, 2007. 11-30.
- RUIZ DE MENDOZA IBÁÑEZ, Francisco José, and Ricardo MAIRAL USÓN. “Challenging Systems of Lexical Representation.” *Journal of English Studies* 4 (2008). [Forthcoming].
- . “Constructing Meaning: A Brief Overview of the Lexical Constructional Model.” *Converging and Diverging Tendencies in Cognitive Linguistics*. Ed. Mario Brdar Amsterdam: John Benjamins, 2009. Forthcoming.
- . “High-Level Metaphor and Metonymy in Meaning Construction.” *Aspects of Meaning Construction*. Eds. Günter Radden, Klaus-Michael Köpcke, Thomas Berg, and Peter Siemund. Amsterdam/Philadelphia: John Benjamins, 2007. 33-51.
- . “Levels of Description and Constraining Factors in Meaning Construction: An Introduction to the Lexical Constructional Model.” *Folia Linguistica* 42 (2008). [Forthcoming].
- . “Levels of Semantic Representation: Where Lexicon and Grammar Meet.” *Interlingüística* 17 (2007): 26-47.
- RUIZ DE MENDOZA IBÁÑEZ, Francisco José, and Sandra PEÑA CERVEL, eds. *Cognitive Linguistics: Internal Dynamics and Interdisciplinary Interaction*. Berlin: Mouton de Gruyter, 2005.
- RUIZ DE MENDOZA IBÁÑEZ, Francisco José and Lorena PÉREZ HERNÁNDEZ. “Metonymy and the Grammar: Motivation, Constraints, and Interaction.” *Language and Communication* 21 (2001): 321-357.
- VAN VALIN, Robert D. Jr. “Lexical Representation, Co-composition, and Linking Syntax and Semantics.” SUNY at Buffalo, 2004. <<http://linguistics.buffalo.edu/people/faculty/vanvalin/rrg.html>>.
- . *The Syntax-Semantics-Pragmatics Interface: An Introduction to Role and Reference Grammar*. Cambridge: Cambridge UP, 2005.
- VAN VALIN, Robert D. Jr., and Randy LAPOLLA. *Syntax, Structure, Meaning and Function*. Cambridge: Cambridge UP, 1997.
- VAN VALIN, Robert D. Jr., and David P. WILKINS. “Predicting Syntactic Structure from Semantic Representations: *Remember* in English and Its Equivalents in Mparntwe Arrernte.” *Advances in Role in Reference Grammar*. Ed. Robert D. Van Valin Jr. Amsterdam: John Benjamins, 1993. 499-534.
- VENDLER, Zeno. *Linguistics in Philosophy*. Ithaca: Cornell UP, 1967.

