

SEMANTIC PACKAGING AND SYNTACTIC PROJECTIONS IN WORD FORMATION PROCESSES: THE CASE OF AGENT NOMINALIZATIONS¹

RICARDO MAIRAL USÓN
UNED, Madrid

FRANCISCO CORTÉS RODRÍGUEZ
Universidad de La Laguna

ABSTRACT. *The aim of this paper is to present the methodological underpinnings of what we have termed the Affixal Lexical Model. This model constitutes an enrichment of the proposals stemming out of both Role and Reference Grammar and S.C. Dik's Functional Grammar. In contrast to the view that word formation processes are treated as either lexical rules or predicate formation rules, our major leading thesis is that affixal units should be regarded as lexical units in their own right. We advocate for the notion of lexical template as a means of representing the type of fusion processes which hold between the affixal unit and the base word.*

KEYWORDS. *Morphology, lexicology, syntax and semantics interface.*

RESUMEN. *El objetivo de este artículo es presentar las bases metodológicas que sustentan lo que denominamos el Modelo Léxico de Afijos. Este modelo constituye un enriquecimiento de las propuestas realizadas desde los modelos funcionales de la Gramática del Papel y la Referencia y la Gramática Funcional de S.C. Dik, que tratan los procesos de formación de palabras bien como reglas léxicas o como reglas de formación de predicados. En este sentido, sostenemos que las unidades afijales son unidades léxicas y postulamos la noción de plantilla léxica como sistema de representación que da cuenta de los procesos de fusión entre la unidad afijal y las bases.*

PALABRAS CLAVE. *Morfología, lexicología, interfaz sintaxis-semántica.*

1. INTRODUCTION

The primary goal of our analysis is explaining and describing the type of linguistic phenomena that occur in the affixal lexicon, why these do occur and in which way(s) these do affect grammar; thus, describing and explaining constitute the two basic methodological pillars.

The theoretical scope of this paper primarily falls upon two functionally-oriented models, Simon Dik's Functional Grammar (FG) (Dik 1997ab) and Van Valin and LaPolla's Role and Reference Grammar (RRG) (Van Valin and LaPolla 1997). Both theories assign the lexicon component a prominent place in the theory. However, none of the two models fully develop the rich set of regularities found within a lexicon organized in terms of coherent semantic classes. This latter assumption becomes more apparent when dealing with word formation processes, which are treated as either predicate formation or lexical rules, thus ignoring the fact that affixal units are in themselves lexical units, organized in lexical classes in much the same way as non-derived words are. This is in fact the first point of departure in this paper.

Furthermore, both theories seem to agree that it is necessary to dispense with redundant information in the lexicon, reducing lexical entries to a minimal amount of information, thus invoking some other mechanisms to derive the rest of the lexical material which a predicate contains. But, again none of the two theories explicitly design an explanatory device (e.g. lexical rules, inheritance mechanisms) which can account for the set of regularities found within lexical classes. This is our second point of departure, viz. the design of mechanisms in the lexicon that can predict information from the lexical entries, and more particularly, from derived words and their corresponding affixal units.

Finally, FG does not investigate the interface lexicon vs. grammar within an articulated theory. Surprisingly enough, this model reduces this complex theoretical issue to the expression rule component, which, by the way, has been relegated to an ancillary status². Conversely, RRG does present a very interesting and solid proposal on linking. Hence, it is no surprise that our approach adopts great part of the linguistic machinery of this model in developing a derivational system from the lexicon to the grammar. This is our third point of departure, viz. the design of a linking algorithm within the RRG framework. We claim that our approach is fully compatible with the RRG proposal on linking and is meant to enrich the RRG lexical representations in a number of substantial ways. With respect to FG, initially it is not that easy to accommodate such derivational process in the FG apparatus since this model lacks a fully developed linking algorithm. However, although in principle the FG and our approach are distant, we are firmly convinced that bridging both models is a feasible and challenging enterprise.

Our endeavor to develop these three initial theses has led us to postulate an alternative proposal for the affixal lexicon within both FG and RRG. The idea behind this proposal is to make explicit the fact that irregularities in the lexicon can to a certain extent be explained if new principles and rules are formulated. In this regard, the Affixal Lexical Model (ALM) contains an inventory of representations which interact in non-trivial ways with either lexical entries and an inventory of semantic constructions (see Appendix 1). This interaction is established *via* a set of fusion principles, which are regulated by means of a set of lexical mapping rules. Then, we introduce a conception of the lexicon as a core lexical grammar. The output of this process will be the initial input for the grammar.

Thus, the format of this paper goes as follows: Section 2 gives an account of the type of mechanisms for the affixal lexicon invoked within both FG and RRG. Section 3 primarily focuses on the fact that affixal units are lexical entries in their own right and consequently these should be represented in terms of what we have called an affixal template, an extension of RRG's logical structures (Van Valin and LaPolla 1997: chapter 3). Section 4 explores the methodological scope of this new system of representation for Agent nominalizations in English and Spanish. Section 5 shows the way the resulting representations fuse with their corresponding base words. Finally, the general layout of the Word-Formation model that we propose is shown in the Appendix.

2. WORD FORMATION WITHIN THE FRAMEWORK OF BOTH FG AND RRG

Before spelling out the details of the ALM, let us comment on the proposals made within the framework of both RRG and FG. Both theories coincide in postulating a lexical rule and a predicate formation rule respectively:

(1) RRG Lexical Rule for Word Formation (Van Valin & LaPolla 1997: 188-189):

verb + *-er* [_N *verb* + *-er*] 'x_i which *verbs*' ([LS ...(*x*₁, ...) ...]), where 'x' is the actor argument in the logical structure

(2) FG Predicate Formation Rule for Agent Noun Formation (Dik 1997b: 3)

input:	pred [V, +contr] (x ₁) _{Ag/Pos} ... (x _n) [n ≥ 1]
output:	{Ag pred} [N] (x ₁) ∅ ... (x _n)
meaning:	'a person who has the property of being (habitually) involved in the action of pred-ing'

These rules, in our view, miss important generalizations about the complex process of word formation. Apart from more specific problematic issues concerning the configuration of the FG model for *-er* nominals (detailed in length in Cortés and Pérez, in press) one common weakness to the two rules described above is their inefficiency to explain the following questions:

- (a) they do not include in their input the possibility of forming non-deverbal nominals, both in the case of *-er* formations (e.g. *astrologer*, *geographer*) and of other agent nouns, a common phenomenon in English (e.g. *cartoonist*, *technician*, *consultant*, *accountant*, *litigant*, etc.) and other languages, such as Spanish (*violinista*, *panadero*); in the case of FG Agent Noun Formation Rule it also excludes the possibility of forming nouns like *sufferer*, *lover*, *keeper*, *container* or *loser*, where the base is [-Con]³.
- (b) in the case of the FG rule, there is no consideration either in the input or the output of what morphological device will be triggered; that is, nothing refers to the final form that the {Ag pred} variable will take. It does not explain either

under what conditions the output form is monovalent (e.g. *writer* (x)) or bivalent (e.g. *writer* (x) (*poetry*)) and, in the second case, under what conditions the second argument will be a postmodifier (e.g. *the writer of this novel*) or a premodifier (e.g. *a book writer*).

-) the semantics of both rules is misleading and, furthermore, there is no interrelation between the meaning and the (internal and external) syntax of the derived word. For instance, how to distinguish between the Spanish formations *competidor* (“someone who competes in one (specific) contest”), *fumador* (“someone who smokes habitually”), *pescador* (“someone whose habitual profession is to fish) or *flotador* (“something to help you float in the water”, i.e. a float)? (cf. Cortés 1997). More complications arise with formations like *diner* (when it means “place where to dine”) or *villager*.

3. THE FORMAT OF A LEXICAL TEMPLATE

In line with Mairal and Faber’s (2002) proposal for the representation of verbal units, we postulate the notion of lexical template (LT) as a system of lexical representation for affixal units. Basically, a LT is the maximal projection of the full repertoire of both semantic and syntactic parameters relevant to a lexical class. This presupposes an organization of the lexicon into coherent lexical classes. The purpose of using lexical classes as a system of semantic decomposition is to count with a formal device that would predict the morpho-syntactic behavior of lexical (in this case affixal) units. The inventory of lexical classes that have been used in this paper is adopted from RRG’s classification of Logical Structures (Van Valin and LaPolla 1997: 109), which basically codify the “modes of action” of predicates:

(3)

Verb Class	Logical Structure
State	predicate' (x) or (x,y)
Activity	do' (x, [predicate' (x) or (x,y)])
Achievement	INGR predicate' (x) or (x,y), <i>or</i> INGR do' (x, [predicate' (x) or (x,y)])
Accomplishment	BECOME predicate' (x) or (x,y), <i>or</i> BECOME do' (x, [predicate' (x) or (x,y)])
Active accomplishment	do' (x, [predicate1' (x, (y))]) & BECOME predicate2' (z,x) or (y)
Causative	α CAUSES β where , β are LS of any type

According to this classification, verbs are classified into *states*, *activities*, *achievements* and *accomplishments* together with their corresponding causatives. *States*

and *activities* are primitives, whereas *achievements* and *accomplishments* consist of either a state or activity predicate plus a BECOME and an INGR operator⁴.

Notice, however, that Logical Structures of this kind do not sufficiently explain the semantics of a lexical unit:

participation in one grammatical alternation does not sufficiently determine the semantic class of the verb. In fact, even once a complete cataloguing of participation in alternation classes is achieved, we must ask ourselves just what we have accomplished. (Pustejovsky 1995: 10)

We would still need further semantic decomposition, and this involves the utilization of other semantic parameters. By way of illustration, consider the following example from Mairal and Faber (2002):

(4)

[[**do'** (w, [**use.sharp-edged.tool**(α)**in**(β)**manner'** (w, x) & [**BECOME be-at'** (y, x)]] CAUSE [[**do'** (x, [**make.cut.on'** (x, y)]]] CAUSE [**BECOME pred'** (y, (z))]],
 $\alpha = x$.

This template encodes the syntactico-semantic scenario for *manner-of-cutting* verbs. This structure contains an effector (w) who carries out the cutting activity upon a patient (y) by means of a sharp-edged tool (x). As a result the affected entity acquires a new state, that of being cut. The representation can be more specifically described as follows: an effector (w) uses a sharp-edged tool (x) in such a way that the tool becomes in contact with a patient (y), causing an event such that x makes a cut on y and this, in turn, causes that y becomes cut. Furthermore, a new variable (z) is introduced to account for those cases where the final result is further specified (*into pieces, in strips, open, etc.*). The representation contains two types of variables: Roman letters represent external variables, and will have a syntactic realization; Greek letters mark internal variables and they are the means to represent the semantic parameters which characterize an entire lexical class. An important consideration about our conception of LTs is that they are not a receptacle of the particular idiosyncratic linguistic features of a predicate; but conversely a LT is a maximal projection of the conceptual substance found within a lexical class, each lexical class being represented by a canonical LT which accommodates into one compact representation the regularities found within the given lexical class. Individual lexical entries are the further expression of the unique selectional properties associated to the meaning of a specific lexical unit. These unique selectional properties are hyponymic expressions of the variables of the LT in ontological terms.

With regard to word-formation, the description of the lexical classes of the base words is one of the factors that explain the differences among competing processes. The following examples with the Spanish affixes *-ero* and *-ista* give us a clue as to the varied

lexical classes where each of these morphemes acts productively, showing quite a neat distribution of tasks:

These two suffixes are mostly operative on nominal lexemes, but it is significant that in the case of *-ero*, and despite its enormous productivity, there are almost no formations from abstract nouns, in opposition to *-ista* formations where a good number of bases are nouns like *exhibición, expresión, camorra, contrabando, segregación* that designate actions, or *progreso, paz, inflación, concepto, arte, estilo* that designate abstract attributes, or clearly evaluative/emotional concepts like *alarma, humor, terror*.

The affix *-ero*, on the other hand, needs base nouns that refer to physical, dimensional objects like *alfiler, alpargata, coraza, pan*, foods and drinks (*horchata, huevo, ensalada, cerveza*), animals and plants (*pájaro, gato, cabra, ballena, algodón, rosal, mimbre*), or places (*arrabal, olivar, mina, rincón*).

There are apparent overlapping cases: for instance, *-ista* is also very productive with bases that designate musical instruments, but formations in *-ero* with these bases are non-existent, but for very few cases (*guitarrero, gaitero, tamborilero*) and in the first of them there is also the formation *guitarrista* with a very different meaning (*guitarrero*, “someone who makes guitars”; *guitarrista* “someone who plays the guitar”). Similarly, *-ero* acts on some abstract bases that designate communication/speech processes (*embuste, recado, mensaje, chanza*) where those cases do not hold for the suffix *-ista*.

We propose that LTs are essentially the same for both the primary lexicon –words– and the affixal lexicon; hence derivational affixes are lexical units in their own right, thus forming a part of the lexicon of a language, viz. the affixal lexicon. Lexical representations form an inventory of structural canonical templates which are maximal representations of the semantic content in a lexical class. Then, morpho-syntactic alternations, which are explained by means of a set of linking rules, refer to both the set of clausal constructions where a lexeme participates and the (morpho-)syntagma underlying a derived lexical unit. In this sense, S-Syntax and W-syntax⁵ receive a unitary explanation.

4. A CASE STUDY: AGENT NOMINALIZATIONS

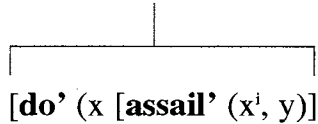
With regard to the affixal classes, the following Template corresponds to Actor nominalizations, which include a vast group of derived formations in *-er, -ant, -ist, -ician* for English, and *-ero, -nte, -or, -ista* in Spanish:

$$(5) _N^i: [\text{PRED} (x^i, [\text{LT}_{\text{BASE WORD}}])]$$

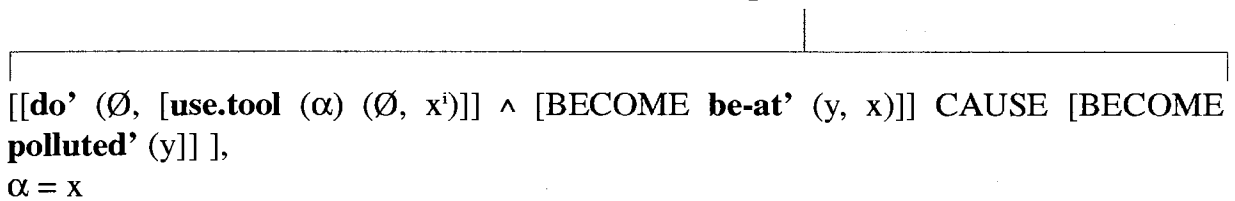
Note that the LT for Actor nominalizations is again the maximal projection of the conceptual substance codified in these affixes. In (5), we have the LT common to the Actor nominalizations; the label ‘Actor nominalizations’ that we use for this affixal class explains the wide scope of these types of derivational processes. The semantics of this template cannot be associated to a specific semantic function, as is proposed in Dik’s

description of the Predicate Formation rule that we have commented before. The term Actor explains the fact that all the formations are nominalizations of the macrorole Actor, as defined in RRG⁶. This in turn justifies the superscript *i* which co-indexes the lexical variable for the derived word ($_N$) with the participant that would receive that macrorole function. The following examples illustrate different instantiations of alternative Actor formations:

(6) $_N^i$: [**do'** (x^i [*assail*])]



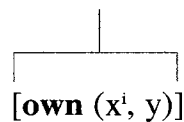
(7) $_N^i$: [**do'** (\emptyset , [**use.tool** (α) (\emptyset , x^i)])] CAUSE [*pollute*]]



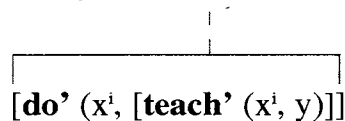
(8) $_N^i$: [**be-loc'** (x^i , [*abrevar*])]



(9) $_N^i$: [**pred'** (x^i , [*own*])]



(10) $_N^i$: [**do'** (x^i , [**manner.as.profession'** (β) (x^i , [*teach*])])]]



Example (6) illustrates the internal structure of the deverbal agentive noun *assailant*; example (7) corresponds to the instrumental formation *pollutant*; (8) corresponds to the

Spanish locative *abrevadero* (“watering place”), and (9) to the word *owner*, whose base word is a stative predicate; example (10) represents the structure for the formation of the deverbal professional agentive nominal *teacher*. These templates consist of the instantiation of the PRED variable plus the corresponding LT of the base word. In postulating a general label such as PRED we can capture the different derivational processes involved in this type of word-formation, namely agent, instrument, locative and state nominalizations. In addition, there is a fusion process between the affixal template and the predicate template which is regulated through a set of lexical mapping rules, which will be commented in the following section.

Finally, and in a parallel fashion to the primary lexicon, an ontology gives lexical expression to the set of corresponding concepts instantiated in both the internal and the external variables. So, the LT serves as input for the ontology which triggers the fully specified semantic representation corresponding to each lexical entry⁷. In this respect, Cortés and Pérez (in press) propose a description of the nominal bases of agentive formations in terms of the Qualia Structure proposed by Pustejovsky (1995); qualia structures describe about a concept the set of semantic constraints by which a word is understood when actually put into language; that is, it triggers semantic transformations by means of which concepts are related to other concepts, thus facilitating inferential mechanisms within a constellation of concepts, in much the same way as an ontology would work. The following example reproduces their proposal for the representation of the word *geographer*:

- (11) **geographer**^{i_N} : **geography**' {*Q_c* & *Q_f*, *Q_i* [<repeated' ^ profession'> do' (yⁱ, [**work**' (yⁱ)])], *Q_a* [**science**' (x), do' (yⁱ, [**work**' (yⁱ)])]}

Qualia structure classifies concepts in terms of the following *Qualia properties*:

- a. *Constitutive role (Q_c)*: the relation between an object and its constituents, or proper parts
 1. material
 2. weight
 3. parts and component elements
- b. *Formal role (Q_f)*: that which distinguishes the object within a larger domain
 1. orientation
 2. magnitude
 3. shape
 4. dimensionality
 5. color
 6. position

c. *Telic role (Q_t)*: purpose and function of the object

1. purpose that an agent has in performing an act
2. built-in function or aim that specifies certain activities

d. *Agentive role (Q_a)*: factors involved in the origin or 'bringing about' of an object

1. creator
2. artifact
3. natural kind
4. causal chain

The two first *Qualia* are useful for the semantic characterization of the nominal domains where one affix shows productivity, whereas the *telic* and *agentive Qualia* would be especially useful to account for the underlying structure of the derived words: as shall be seen later, these features provide the verbal nexus that binds the derived word to its semantic characterization in the Modeling Process. Thus, departing from the basic template, given in (5) it is possible to specify it by considering that the LT position would be occupied by the nominal base with its *Qualia* characterization and that (i) the semantics of PRED is compatible with either the *Q_t* or the *Q_a* characterization of the nominal base so that (ii) co-indexation takes place with one of the arguments expressed in either of these *Qualia*, as shown in example (11), where both the Telic and the Agentive roles in the noun's *Qualia* semantic properties make explicit the verbal nexus and agentive character of the formation, since they include the constant **do'** for activities.

5. THE INTERACTION OF THE PRIMARY AND THE AFFIXAL LEXICON: FUSION PRINCIPLE I

As commented in section 4, the LT in (5) is the lexical representation of all Actor nominalizations, and examples (6) to (10) on the one hand illustrate different realizations of this maximal LT, and on the other hand, presuppose a fusion between the derivational and the primary lexicon, as the base words (in italics in the examples) operative for a the specific derivational processes appear already inserted in the appropriate positions in the templates.

With regard to the first issue, I is necessary to device some sort of linking mechanism that accounts for the relation between the LT of a whole derivational class - as it is the case of Actor nominalizations- and the different morpho-semantic constructions that participate in the class (such as instrumental, agentive or locative nominalizations). Such linking algorithm is formulated in a number of lexical mapping rules which are encoded by a general Lexical Modeling Process (LTMP), already formulated by Mairal and Faber (2002) for the primary lexicon:

Lexical Template Modeling Process (LTMP)

Lexical templates can be modeled by suppressing external variables, instantiating internal variables, eliminating operators (e.g. CAUSE), or else, by introducing elements resulting from the fusion with other templates

The LTMP constitutes also a first phase of linking on the affixal lexicon. Recall that the structure of a word formation template (e.g. actor nominalizations, locative affixations, action nominals, undergoer nominalizations etc.) consists of two parts: the conceptual substance of the word formation process at hand and (ii) the LT of the base word. A further point of convergence with respect to the primary lexicon is that in the affixal lexicon fusion is also produced between a construction and a word, e.g. within the general class of actor nominalizations, one of the potential constructions resulting from the canonical word formation LT *via* the application of the LTMP is the instrument nominalization construction which underlies a vast number of formations such as the following:

(12) Instrument nominalizations in *-er*:

amplifier; burner; computer; cutter; disclaimer; eraser; freezer; pacifier; scanner; softener; sweetener; stapler; toaster; thinner; thickener; rubber; recorder; photocopier; etc.

(13) Instrument nominalizations in *-ant*:

contaminant; defoliant; deodorant; detergent; disinfectant; lubricant; repellent; stimulant; adherent; expectorant; pollutant; decongestant; determinant; astringent; irritant; solvent; deterrent; etc.

(14) Instrument nominalizations in the Spanish affix *-ero*:

mosquitero; yesquero; aliviadero; cafetera; coctelera; chocolatera; tetera; desenfadaderas; entendederas; explicaderas; posaderas; mantequera; lingotera; heladera; devanadera; freidera; plegadera; podadera; tapadera; regadera; etc.

(15) Instrument nominalizations in the Spanish affix *-or*:

acondicionador; apisonadora; cosechadora; esterilizador; excavadora; extensor; incumbadora; lavadora; ordeñadora; perforador(a); refrigerador; remolcador; abrillantador; acelerador; afilador; aspirador(a); borrador; bronceador; congelador; desmaquillador; destornillador; encendedor; escurridor; exprimidor; freidora; grabadora; rallador; secador; secadora; silenciador; sujetador; trilladora; etc.

(16) Instrument nominalizations construction

$_{-N}^i$: [**do'** (\emptyset , [**use.tool** (α) (\emptyset , x^i)])] CAUSE [LT_{BASE WORD}]] , where $\alpha = x^i$

In order to better explain this construction, let us bring back the LT of the class of Actor nominalizations:

$$(5) \quad _N^i: [\text{PRED} (x^i, [\text{LT}_{\text{BASE WORD}}])]$$

This structure describes the basic semantic configuration of a nominal predicate ($_N^i$) that designates one entity involved in some kind of event ($[\text{PRED} (x^i, [\text{LT}_{\text{BASE WORD}}])]$) such that its meaning involves a specific orientation of the semantics of the word formation processes towards that entity. Co-indexation with the superscrit i expresses such orientation or profiling of a participant (cf. Laca 1993; Cortés and Pérez 2002); that is, all formations from this class would depict one of the entities involved in a certain semantic scenario. The variable “PRED (x^i ...” expresses the fact that the event where this entity participates can be of any kind, a state (**pred'**) or an activity (**do'**) and, consequently, the semantic function of the entity nominalized is not necessarily that of Agent; as it is described by Cortés and Pérez (2002). This LT expresses in fact the whole range of formations that correspond with the Actor macrorole as described in RRG.

The instrument construction codifies a structure where the Actor formation is oriented towards the implement of an activity. This choice is not random but appears modulated by the RRG's scale of Actor Macrorole Assignment:

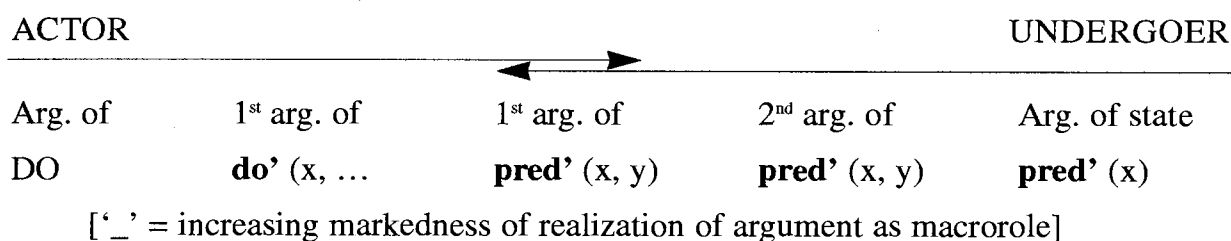


Figure 1. *The Actor-Undergoer Hierarchy*

The instrument construction, therefore, accounts for those formations from a causal chain where the first argument is left unspecified, thus yielding the INSTRUMENT entity as first candidate for the Actor function. Note that this construction is modulated by the LTMP as applied to the participants of the event: in these cases the Agent is left unelaborated (see the Lexical Mapping Rule 2 below) to the benefit of the Instrument, which is highly elaborated to the extent of having to make reference to an activity subevent in the first part of the causal chain ($[\text{use.tool} (\alpha) (\emptyset, x)]$)

In addition to the instrument nominalization construction, two other constructions that can fuse with the Actor nominalization LT are the following:

(17) Professional agent nominalizations:

$_N^i$: [**do'** (x^i , [**manner.as.profession** (β)' (x^i , [$LT_{BASE\ WORD}$])])]

e.g. *teacher, publisher, producer, anthropologist, economist, pianist, receptionist, enfermero, barrendero, pastelero, cocinero, investigador, enterrador, Labrador.*

(18) Locative nominalizations:

$_N^i$: [**be-loc'** (x^i , [$LT_{BASE\ WORD}$])]

e.g. *diner, dresser, bañera, papelera, criadero, gallinero.*

These two constructions reflect very different possibilities of realization of the Actor formations: the first one expresses the semantic content of the most prototypical nominalizations within the class: the derived words corresponding to this construction describe the Agent involved in the event described in the semantics of the base word. The internal variables (**[manner.as.profession'** (β) (x^i)] specify the non-episodic nature of the activity, and express the fact that agentivity in this case involves the habitual iteration of the activity expressed within the meaning of the base, as the Agent carries it out as its usual professional occupation.

Locative nominalizations, on the other hand, are less productive insofar as they express a more marked choice of Actor macrorole according to the Macrorole Scale. The construction expresses a locative state where the Location is profiled and the base word's event is the Theme of the locative relation.

Once the LTMP has been applied to the canonical template an underlying representation for the would-be derived word is obtained; this representation can be described as an "open underlying structure" since the LT variable corresponding to the base word needs still to be saturated. Such process –the Fusion Principle– is not unrestricted, but appears subject to certain Matching conditions, whose aim is to account for the semantic well formedness of word-formation underlying representations. In other words, the aim of the Fusion Principle is to yield a congruent combination of the base and the derived word's LTs. For the Fusion to be satisfactory one of the following Conditions must be met⁸:

Lexical Template Modeling Process in the affixal lexicon:

LTs for the affixal lexicon can be modeled iff the canonical LT of the affixal class and the construction-based LT of the base word meet the following two lexical mapping rules:

Lexical Mapping Rule 1'(LMR 1):

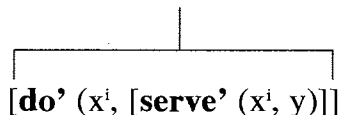
There must be a copying/identification of operators and variables between both LTs. (Matching Condition I: $LT_{BASE\ WORD} \equiv LT_{DERIVED\ WORD}$) or,

Lexical Mapping Rule 2' (LMR 2):

The LT of the base word must be a proper subevent of the derived word LT.
(Matching Condition II : $LT_{\text{BASE WORD}} \supset LT_{\text{DERIVED WORD}}$).

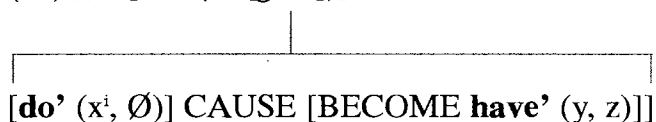
Let us consider different examples to illustrate both possibilities:

(19) $_N^i$: [**do'** (x^i , [**manner.as.profession'** (β) (x^i , *serve*))]]

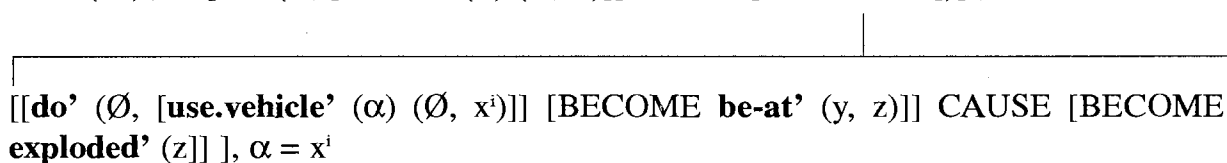


This structure codifies the underlying representation of the noun *servant*. As can be seen, the fusion between the constructional template of the derived word and the lexical entry of the base has taken place, and the result has become congruent since it is possible to identify the operators and variables of the base's lexical entry and those of the construction. In other words, fusion has satisfied LMR 1 ($LT_{\text{BASE WORD}} \cong LT_{\text{DERIVED WORD}}$). The same explanation applies to examples (20) (for episodic agentive *giver*) and (21) (for the Spanish instrumental *bombardero*):

(20) $_N^i$: [**do'** (x^i , [*give*])]]

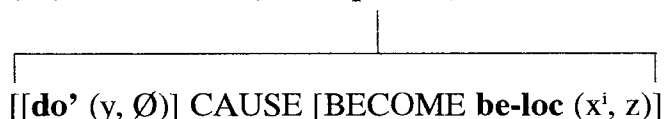


(21) $_N^i$: [**do'** (\emptyset , [**use.tool** (α) (\emptyset , x^i))] CAUSE[*bombardear*]] , where $\alpha = x^i$



The next representation, corresponding to the Spanish locative *despeñadero*, requires a different explanation: in this case fusion is accomplished by means of the LMR 2. The LT of the base expresses a causative accomplishment that becomes the Theme of the locative construction. Since dynamic events can be located with regard to spatial dimensions they can become a proper subevent of this structure: **be-loc'** (x^i , [+Dyn event]).

(22) $_N^i$: [**be-loc'** (x^i , [*despeñar*])]]



More interesting cases arise if other locative word-formation processes are considered: there are a great number of prefixal formations, both in Spanish and English that involve a locative, temporal or notional relation between the prefix and the base word. The locative predicate in the LT of this class is further instantiated in several constructions. Thus, there are –among others– anteriority (*antenatal*, *predate*, *forefinger*, etc), posteriority (*postscript*, *postgraduate*, *post-war*) and superiority (*foreman*, *proconsul*, etc) constructions. In these types of derivational processes the LMR 2 is highly operative since it filters out bases whose semantics involve an opposite locational, temporal or notional relation to that expressed by the construction (e.g. **ante-postdate*, **pre-postgraduate*, **fore-backgallows*, **post-precondition*, etc.⁹).

Example (23) is the underlying representation of the Spanish agentive noun *novelista*; this structure illustrates how the LTMP works for denominal formations. In these cases, there is no direct matching between the activity part of the constructional template and the representation of the base, since nominal predicates are prototypically stative. However, matching occurs with other sections of the lexical entry of the base, namely the Telic and/or the Agentive Qualia¹⁰.

(23) $_i^N$: [**do'** (x^i , [**manner.as.profession'** (β) (x^i , [*novela*]))]]

[**novel'** (y), Q_c : [**narrative'** (y)], Q_f : [**book'** (y), **disk'** (y)], Q_i : [**do'** (z , [**read'** (z , y)]), Q_a : [**artifact'** (y), **do'** (x^i , [**write'** (x^i , y))] & BECOME **exist'** (y)]]

Once the fusion between both templates have been produced, it is necessary to check the fully specified underlying semantic representation against the entries for the affixes that participate in the class in order to retrieve the one that matches all the conditions imposed by the structure. In the case of *servant* (see representation no.19) this fusion process would retrieve the affix #ant_N## and insert it in the appropriate variable position:

(24) $_i^N$: [**do'** (#ant_N##, [**manner.as.profession'** (β) (#ant_N##, [*serve*]))]]

[**do'** (, [**serve'** ((#ant_N##, y))]]

Summarizing a bit, at this stage of the derivational process the actual output of this first phase of linking will be, for the primary lexicon, the semantic characterization of a given construction, and –for the affixal lexicon- the semantic description of the derived word.

6. SECOND PHASE OF LINKING

Once obtained the fully specified semantic representation of the derived word, it is necessary to postulate a second phase of linking, whose task is to derive –from the information codified in the underlying representation- the adequate morpho-syntagmatic structure of the derived word. This is made by means of two sets of rules:

- W-Syntactic rules, which will account for the internal grammatical configuration of the predicate, and
- S-Syntactic rules, which will explain its external syntagmatic structure.

W-Syntactic rules stipulate which elements of the underlying representation will be marked as Determinant (DT) and Determinatum (DM) components. Recall that the internal structure of a word can be described as a grammatical *syntagma* in which “*the determinatum is that element of the syntagma which is dominant in that it can stand for the whole syntagma in all positions*” (Marchand 1969: 11-12), being the Determinant a modifier of the Determinatum.

In the case of suffixal Actor derivations the affixal unit tends regularly to be the Determinatum as it is the element whose categorial features would determine the morpholexical category of the derived word. On most occasions, the base word will become the Determinant. The syntactic template associated to this structure is the following:

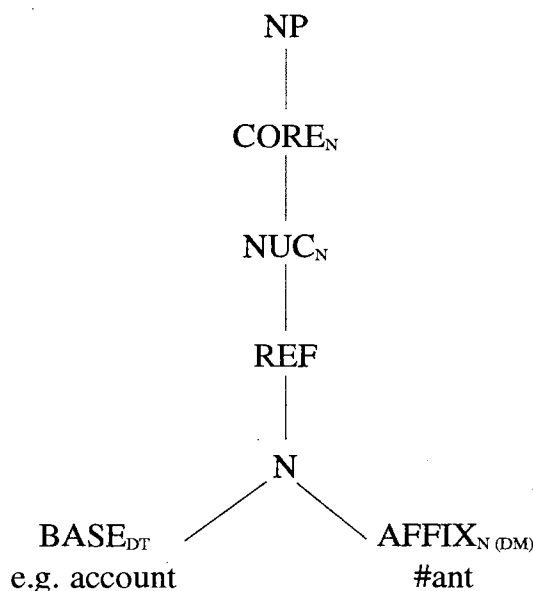


Figure 1. *Syntactic template for actor transitive constructions*

There is, however, one interesting case in which the Determinant is complex: this situation occurs when the subevent associated to the base in the underlying representation is an activity with an inherent second argument. In these cases, the

Determinant may be formed by the base verb and its second argument together; the corresponding syntactic template is the following:

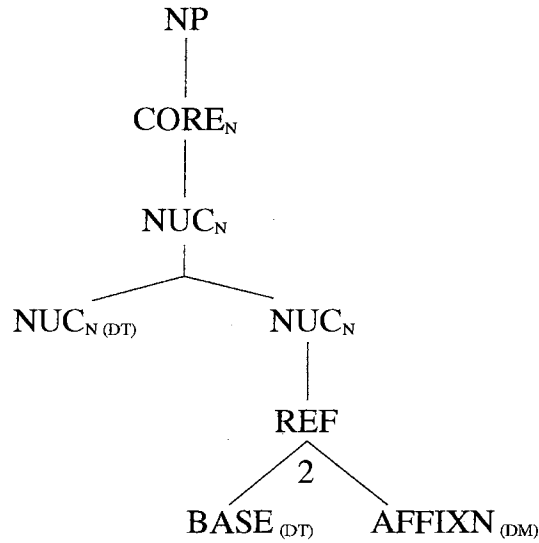
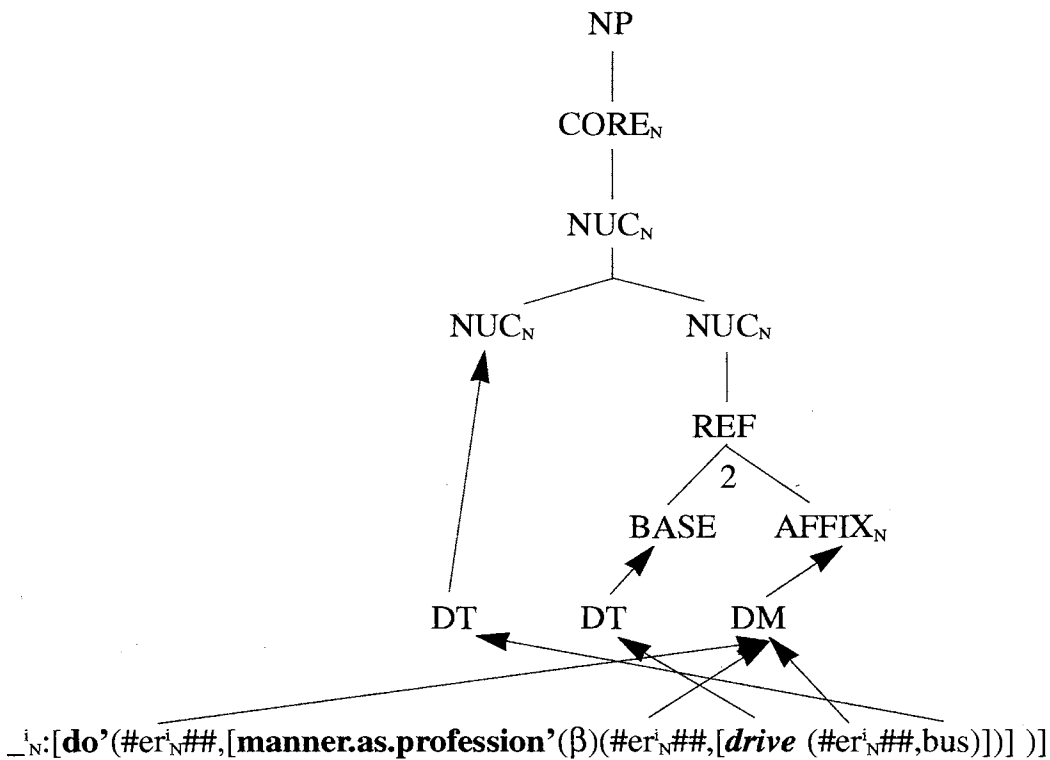


Figure 2. Syntactic template for actor atransitive nuclear junctures

Example (25) shows the effect of W-Syntactic rules on the lexical representation for the nominal *driver*:

(25)



As can be seen in this example, the application of W-Syntactic rules in this phase of linking assigns the function of Determinatum to the argument that is co-indexed with the variable $_N^i$, since it refers to the entity designated by the derived formation process; this argument is lexically expressed by the suffix *-er*; given its status as determinatum it becomes the head for the nominal formation through linking, and its categorial features (expressed by the subscript N) percolate up to the higher node, thus indicating the nominal character of the word. The other elements that are expressed lexically become Determinants of the suffix; note that the second argument of the LT (*bus*) of the base word (*drive*) appears also as a nucleus, given its lexical character, but it does not dominate a REF node, given its non-referential status. That is, if the LT of an activity verb, such as *drive*, saturates its second argument slot with a non-referential or inherent participant, then it is incorporated as a Determinant by the W-syntactic Rules. The syntactic effect in these cases is the formation of a nuclear juncture that corresponds to a process of noun incorporation (*bus driver*). However, noun incorporation is blocked when the activity base is used with an active accomplishment interpretation (e.g. *the hunter of the deer* – **the deer hunter*), given the referential nature of the second argument.

After the application of W-Syntactic rules, S-Syntactic Rules are activated to finally determine the combinatorial behaviour of the word in its immediate syntagmatic context, the NP. The S-Syntactic behaviour of the word is essentially determined by Macrorole Assignment in terms of the Actor-Undergoer Hierarchy (see section 5). Actor formations are very often MR-Atransitive since Undergoer assignment does not take place¹¹. U-assignment takes place in those formations where the base word subevent is an activity with a referential second argument; in such cases the U-participant would appear in an *of*-postmodification structure:

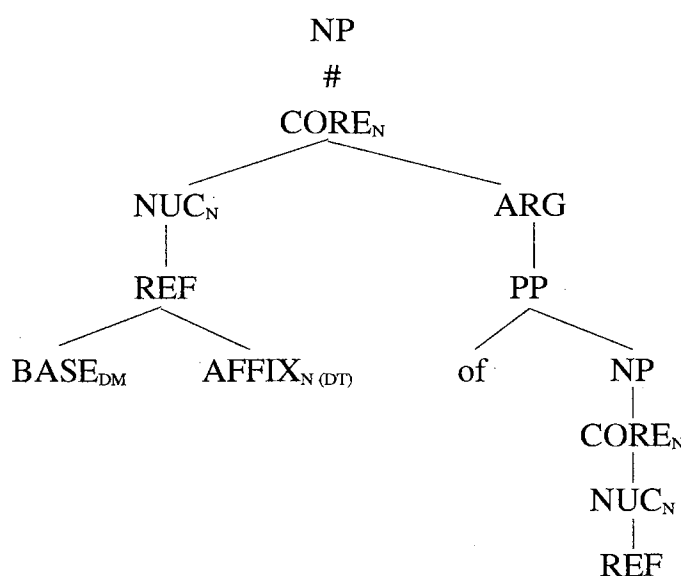
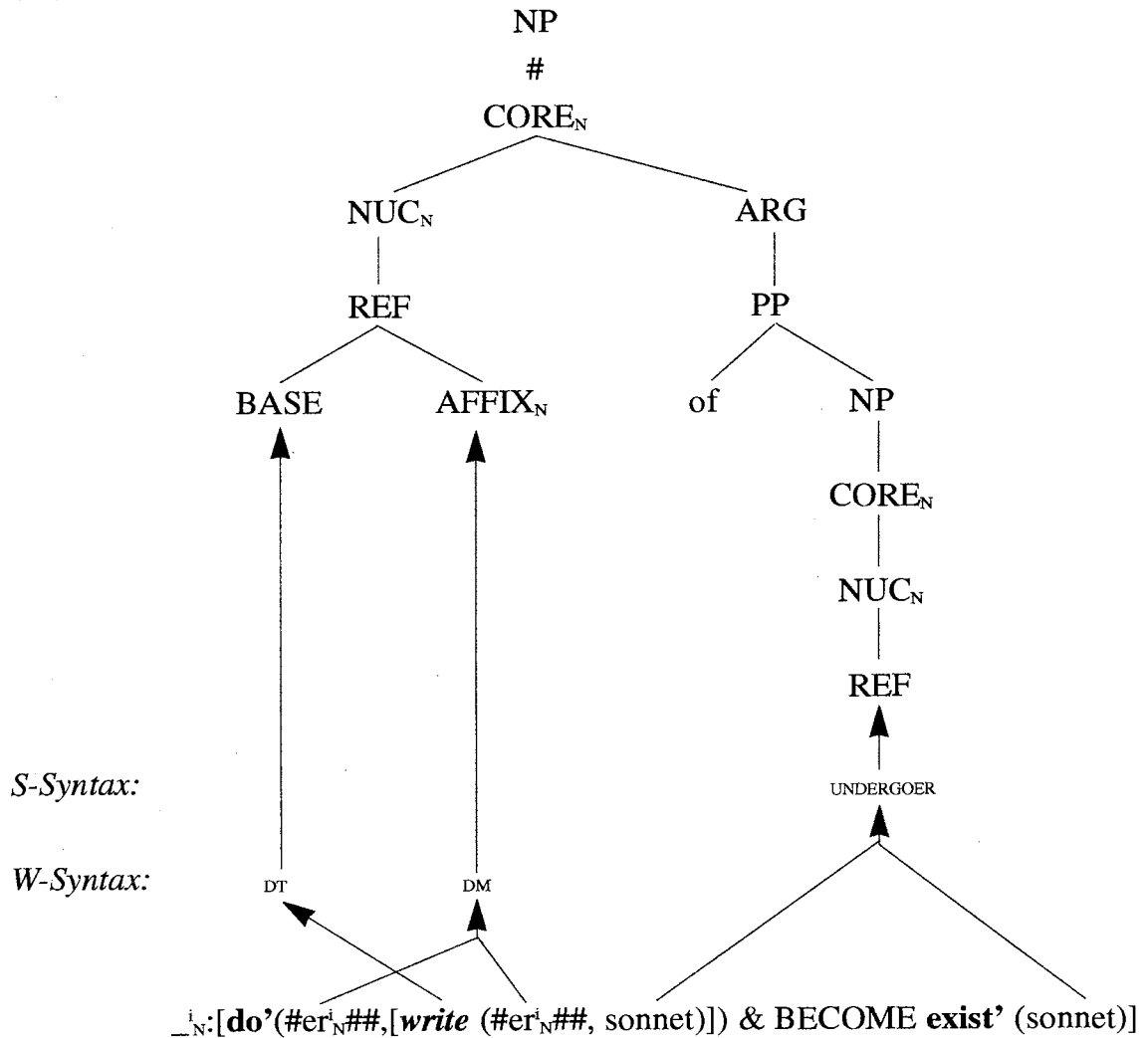


Figure 3. Syntactic template for actor transitive nominalizations

The following example shows the effect of U-assignment:

(26)



This diagram shows the underlying structure and the linking process for the formation *the writer of the sonnet*. The base verb *write* must be understood in its active accomplishment interpretation¹² and its second argument, *the sonnet*, is necessarily non-inherent, a feature that blocks its availability as a determinant in the W-syntactic linking process but at the same time makes it a suitable candidate for the assignment, in the S-.Syntactic phase, of the Undergoer macrorole in terms of the Actor-Macrorole Hierarchy.

In the cases in which the formation involves some inherent argument that has not been incorporated in the nominal nucleus, it can optionally appear also under the format of a postmodifier (e.g. *a composer of melodies*); there is a restriction in this case which affects the inherent argument if it is a countable noun, and which indicates its non-U status: it must appear in plural form. Thus,

- Truncations: they delete morphological material that lies in an internal position after affix adjoining. Then, the general format of a truncation rule (cf. Aronoff 1976: 82 and ff.) is:

$$(23) \begin{array}{cccc} [[[\text{root} + \text{A}]_x + \text{B}]_y & & & \\ 1 & 2 & 3 & \rightarrow 1 \ \emptyset \ 3 \end{array}$$

where x, y = lexical categories

One case of the effect of these rules is that which affects *-ant* nominalizations from verbal bases ending in *-ate*: *tolerate* → *toler(ate)ant*; *stimulate* → *stimul(ate)ant*.

Allomorphy rules: they produce some morphologically motivated phonological changes. These rules affect certain morphemes when they get into contact; for example, the suffix *-ation* has two allomorphic variants for the cases in which it is a adjoined to latinate roots: when the root ends in a [+Coronal] sound the morpheme will be realized by the allomorph *-ion* (*rebellion, decision*); the variant for [-Coronal] ending bases is *-tion* (*consumption, absorption*).

After the morphological operations, the phonological rules come into play. Their task is to yield the final phonological contour of the derived word, in terms of two different operations, namely syllable distribution and stress assignment. Both operations are sensitive to the characterization of the affix boundary: affixes are classified phonologically in terms of their boundary. Class I affixes (represented with a + boundary; e.g. *+ician##*, and all Spanish suffixes) are heterosyllabic, since they fuse themselves with the syllabic structure of the base word and cause a redistribution of the stress configuration in the derived word with regard to its base; on the other hand, Class II affixes (with a # boundary, as are *#er##* and *#ist##*) are tautosyllabic, i.e. they neither alter the initial syllabic structure of the base nor its stress configuration¹⁴.

The output of the expression rules is the actual linguistic expression, that is, the derived word, in the case of a word formation process. This word encapsulates all the morphological, phonological, syntactic and semantic information that has been constituted along its derivational trajectory, and it is with all this information –which ultimately conforms its lexical entry– that will be incorporated into the primary lexicon.

8. CONCLUSIONS

In this paper we have shown the rudiments of an alternative approach (the ALM) to both FG and RRG within the domain of word formation. Our conception of word-formation processes is drastically different in the sense that it is essentially lexical. That is, a derived word is the result of two simultaneous processes that can be labelled as a grammaticalization of the lexicon and as a lexicalization of grammatical structures. The study of word formation has made manifest that derivational products are related to other

analytical structures, namely clauses and phrases, to the extent that the grammatical behaviour of synthetic word-formation products is similar, if not identical, to some of the features of other grammatical units. Suffice to mention, by way of example, the close correlation between the complement structure of action nominals and their corresponding clausal constructions. This has led many researchers to study word-formation as a process of “syntactic recategorization”. However, such a “grammaticalized” or “syntacticist” conception oftentimes entails the sacrifice of the semantico-lexical aspects involved in the generation of a new word. Word-formation is also a lexicological phenomenon by means of which new lexical units are coined to convey new meanings in a given communicative context, and the material used is also lexical: words and affixes, each of them imposing a certain number of combinatorial conditions at all levels: semantics, syntax, morphology, phonology and even pragmatics. In essence, word-formation is a grammaticalization of the lexicon and, as natural entailment, must be part of what we have termed an ALM.

From the point of view of the ALM affixes are lexical units endowed with a lexical entry, and, accordingly, the derived lexicon is also organized in terms of lexical derivational classes that would capture semantic, syntactic and, in this case, morphological regularities. Briefly put, the ALM has the following structure: (i) a set of word-formation lexical templates that correspond to the various types of word formation processes in a language (causative formation, action nominalizations, etc); (ii) an affix lexicon, where all affixes of a language have a lexical entry which encodes their morphophonological and lexico-semantic structure, together with their restrictions and combinatorial properties; (iii) two sets of rules with the following tasks: (a) the first set links a basic schema with the fully developed underlying structure of the derived word; (b) the second set of rules links the fully specified underlying representation of the derived word with its final morphosyntactic configurations; (iv) finally, the expression rules will account for the phonological contour of the derived word.

A further line of research is the implementation of this type of analysis in the primary lexicon.

NOTAS

1. This paper is part of the research project no. PI 1999/136 funded by the Gobierno Autónomo de Canarias.
2. In this regard, Bakker (1999, 2001) constitute an exception to this general state of affairs.
3. [-Con] predicates denote the so-called non-controlled States of Affairs; i.e. they are lexical units whose meaning involves processes or states which are not caused or instigated volitionally by any participating entity.
4. For a detailed discussion of this type of lexical representation within the larger framework of lexical semantics, we refer the reader to Mairal (2001).
5. These terms S(entence)-Syntax and W(ord)-Syntax are borrowed from Bresnan (1982) though the theoretical implications are substantially different, since Bresnan introduces these concepts to account for the word-internal syntactic structure (word-syntax) within the framework of X-Bar Theory.
6. “Macroroles are generalizations across the argument-types found with particular verbs which have significant grammatical consequences; it is they, rather than specific arguments in logical structure, that grammatical rules refer primarily” (Van Valin and LaPolla 1997: 139). RRG distinguishes two macroroles: the Actor, or generalized agent-type role, and the Undergoer or generalized patient-type role; it is

important to emphasize that the term actor is compatible with non-volitional things such as in *The key opened the door* where *key* is the actor (Van Valin and LaPolla 1997: 141). See section 5 for the mechanism of macrorole assignment to the arguments of any Logical Structure.

7. For further details of the way the ontology would interact with lexical representations, we refer the reader to Moreno and Pérez (2002).
8. We use * to differentiate those lexical mapping rules operating on the primary lexicon from the ones in the affixal lexicon. Both the content and function of these rules illustrate the mirror image of the halves of the lexical component, the primary and the affixal lexicon. In this regard, let us comment that Mairal (fc) presents a full research program on the primary lexicon.
9. For a detailed analysis of the prefixal lexicon of English see Sosa (2001).
10. As was mentioned in section 4, we follow Pustejovsky's (1995) Qualia Theory for the description of the lexical entry for nouns, which is also adopted in by RRG. A matter for further research is whether a more profound analysis, such as that provided by an ontology gives a more substantive backup for the description of lexical units.
11. We refer the reader to Cortés and Pérez (in press) for a detailed account of the effect of MR assignment in the case of Actor nominalizations. Interestingly enough, these authors explain that Actor nominalizations follow an ergative pattern, since the essential function of this class of formations is to profile the Actor entity, which consequently becomes incorporated in the W-Syntax.
12. One example of the use of this verb as an activity is: *John writes poetry*; in this case, the logical structure is: **do'** (John,[*write* (John, poetry)])
13. For the application of the expression rules within the domain of the primary lexicon, we refer the reader to Dik (1997a: chapters 14-18).
14. The distinction between these two types of affixes stems from the works of Aronoff (1976), Strauss (1982) and Kiparsky (1982ab, 1983), among others. However, they are not utilized in the same manner as expressed in this paper: the distinction presupposes for these authors a cyclic ordering of word-formation processes: Cycle I affixes are prephonological, whereas Cycle II morphemes are postphonological. Readjustments are also previous to Cycle II affixation, being thus forbidden for formations with affixes whose boundary is #. Both the data analyzed (Actor nominalizations) and works of several authors have shown that this distinction is problematic. This is the reason why our proposal restricts the explanatory potential of boundaries and leaves readjustment rules beyond their scope of influence.

REFERENCES

- Aronoff, M. 1976. *Word Formation in Generative Grammar*. Cambridge, Mass.-London: MIT Press.
- Bakker, D. 1999. "FG Expression Rules: from templates to constituent structure". *Working Papers in Functional Grammar* 67.
- Bakker, D. 2001. "The FG Expression Rules: A Dynamic Model". *Revista Canaria de Estudios Ingleses* 42: 15-54.
- Cortés Rodríguez, F.J. 1997. *La creación léxica: Una aproximación funcional*. Servicio de Publicaciones. Universidad de La Laguna.
- Cortés, F. and M. J. Pérez. 2002. "On the syntax-semantics interface in word formation: the case of English *-er* nominalizations". *New Perspectives on Argument Structure in Functional Grammar*. Eds. R. Mairal and M.J. Pérez Quintero. Berlin and New York: Mouton.
- Dik, S. 1997a. *The Theory of Functional Grammar: The Structure of the Clause*. Berlin: Mouton de Gruyter.

- Dik, S. 1997b. *The Theory of Functional Grammar: Complex and Derived Constructions*. Berlin: Mouton de Gruyter.
- Faber, P. and R. Mairal. 1999. *Constructing a Lexicon of English Verbs*. Berlin: Mouton de Gruyter.
- Kiparsky, P. 1982a. *Explanation in Phonology*. Dordrecht: Foris.
- Kiparsky, P. 1982b. "From Cyclic Phonology to Lexical Phonology". *The Structure of Phonological Representations (Part I)*. Eds. H. Van der Hulst and N. Smith. Dordrecht: Foris. 131-175.
- Kiparsky, P. 1983. "Word-Formation and the Lexicon". *Proceedings of the 1982 Mid-America Linguistics Conference*. Ed. F. Ingemann. Lawrence: University of Kansas. 3-22
- Laca, B. 1993. "Las nominalizaciones orientadas y los derivados españoles en *-dor* y *-nte*". *Fundamentos de morfología*. Ed. S. Varela Ortega. Madrid: Taurus Universitaria. 180-204.
- Mairal, R. 2001. "En torno a la interficie léxico-semántica en los modelos gramaticales". *La investigación en lenguas aplicadas: enfoque multidisciplinar*. Eds. P. Durán and G. Aguado. Madrid: Fundación Gómez Pardo. 115-151.
- Mairal, R. Forthcoming. "Lexical structures and syntactic projections: Towards a design of a lexical grammar model" (Unpublished manuscript).
- Mairal, R. and P. Faber. 2002. "Reconsidering Lexical Representations in Functional Grammar". *New Perspectives on Argument Structure in Functional Grammar*. Eds. R. Mairal and M.J. Pérez Quintero. Berlin and New York: Mouton.
- Mairal, R. and M. J. Pérez Quintero (eds.) (in press). *New Perspectives on Argument Structure in Functional Grammar*. Berlin and New York: Mouton.
- Marchand, H. 1969. *The Categories and Types of Present-Day English Word-Formation. A Synchronic-Diachronic Approach*. 2nd edition. München: Beck'sche Verlagsbuchhandlung.
- Moreno Ortiz, A. and C. Pérez Hernández. 2002. "Ontological Semantics and Lexical Templates: mowing the grass from the other side of the fence". *New Perspectives on Argument Structure in Functional Grammar*. Eds. R. Mairal and M. J. Pérez Quintero. Berlin and New York: Mouton.
- Pustejovsky, J. 1995. *The Generative Lexicon*. Cambridge MA: MIT Press.
- Rappaport, H. M. and B. Levin. 1998. "Building verb meanings". *The Projection of Arguments: Lexical and compositional factors*. Eds. M. Butt and W. Geuder. Stanford: CSLI. 97-134.
- Selkirk, E. 1982. *The Syntax of Words*. Cambridge, Mass.: MIT Press.
- Sosa, E. 2001. *Análisis funcional cognitivo de los procedimientos de prefijación locativa en inglés*. Unpublished Ph. D. Universidad de La Laguna.
- Strauss, S. L. 1982. *Lexicalist Phonology of English and German*. Dordrecht: Foris.
- Van Valin, R. D. and R. LaPolla. 1997. *Syntax: structure, meaning and function*. Cambridge: Cambridge University Press.

APPENDIX: THE WORD-FORMATION MODEL

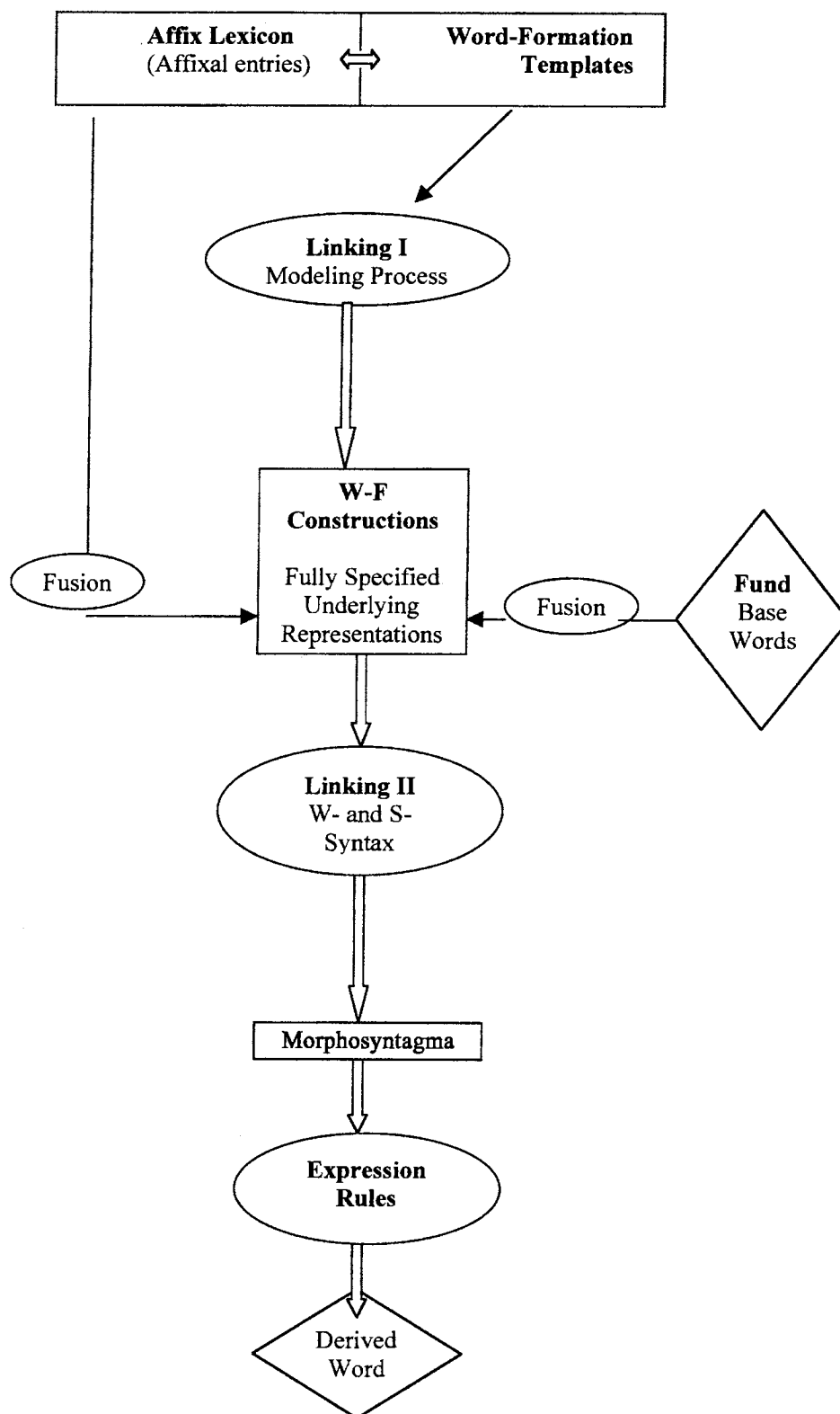


Figure 1. *Word Formation Component*