

# Derivational Morphology of Italian: Principles for Formalization

Francesca Carota

Dipartimento di Linguistica, Università degli Studi di Pisa, Pisa, Italy

## Abstract

The present paper investigates the major derivational strategies underlying the formation of suffixed words in Italian, with the purpose of tackling the issue of their formalization. After having specified the theoretical cognitive premises that orient the work, the interacting component modules of the suffixation process, i.e. morphonology, morphotactics and affixal semantics, are explored empirically, by drawing ample naturally occurring data on a Corpus of written Italian. A special attention is paid to the semantic mechanisms that are involved into suffixation. Some semantic nuclei are identified for the major suffixed word types of Italian, which are due to word formation rules active at the synchronic level, and a semantic configuration of productive suffixes is suggested. A general framework is then sketched, which combines classical finite-state methods with a feature unification-based word grammar. More specifically, the semantic information specified for the affixal material is internalised into the structures of the Lexical Functional Grammar (LFG). The formal model allows us to integrate the various modules of suffixation. In particular, it treats, on the one hand, the interface between morphonology/morphotactics and semantics and, on the other hand, the interface between suffixation and inflection. Furthermore, since LFG exploits a hierarchically organised lexicon in order to structure the information regarding the affixal material, affixal co-selectional restrictions are advantageously constrained, avoiding potential multiple spurious analysis/generations.

### Correspondence:

Francesca Carota,  
Institute for Cognitive  
Sciences,  
67, boulevard Pinel  
69675 BRON Cedex (Lyon),  
France.  
E-mail:  
Francesca.Carota@isc.cnrs.fr

## 1 Introduction

### 1.1 Motivation

The present paper investigates the major modalities that underly the formation of morphologically complex, suffixed words, on the basis of morphologically simple words in Italian.<sup>1</sup> Some studies have examined specific aspects of Italian Suffixation, which offers an essential derivational strategy for lexical renewal. However, what is still lacking is a comprehensive picture of the process, which represents an important pre-requisite for morphological

analysis and generation in computational applications. The *desideratum* is that the relevant derivational mechanisms involved into the suffixation process are systematically examined and then specified in formal terms in view of automatic analysis and generation of suffixed words. A general architecture is then proposed, in which the interactions between the various modules operating in morphological suffixation, i.e. morphotactics, morphonology and semantics are integrated in a unified approach. It is well established in the literature that the morphotactic and the morphonological

components can be modelled to a large extent using finite-state models (Jurafsky and Martin, 2000), whereas the semantics of suffixation is not included in the finite-state approach and requires a more specific treatment. The model adopted here declares explicitly and according to different degrees of semantic specificity the semantic information associated with derivational bases and suffixes, by means of the structures of Lexical Functional Grammar (LFG) (Kaplan and Bresnan, 1995). This allows us to constrain the suffix combinatory properties, by filtering potential incorrect morphotactic combinations.

The article is articulated as follows: after sketching a schematic overview of major approaches to derivational morphology and some related methodological concerns in Section 1, the basic mechanisms underlying suffixation are illustrated in Section 2, by examining naturally occurring data from Italian and modelling the corresponding morphonological and morphotactic rules in finite-state terms; Section 3 introduces the semantic mechanisms of suffixation, to be described within the framework of LFG; Section 4 reports some conclusive remarks.

## 1.2 Theoretical approaches to morphology and present orientation

Linguistic theories see morphology in terms of ‘word-syntax’ (ITEM AND ARRANGEMENT model) (Spencer, 1991), and, alternatively, as a set of lexical rules operating on affixal material and constraining affix combinations and modifications (ITEM AND PROCESS model) (Aronoff, 1976, 1994). On the other hand, the WORD AND PARADIGM approach, first, assumes that a morphological system consists of paradigms representing both inflectional classes and principal parts for individual lexemes, and, secondly, ascribes to the grammar an inventory of morphological ‘spell out’ rules that apply to the feature bundles specified for some selected grammatical features (Matthews, 1972; Anderson, 1977, 1992).

The perspective preferred here is the second one. Accordingly, following Schwarze (2001), morphological rules, first, express restrictions both on affixal pattern combination at the constituency level; secondly, constrain feature projections at the level

of functional information; thirdly, create predicates at the level of semantic information.

Such a formulation of the notion of morphological rules seems particularly felicitous with respect to the present desideratum, that is to describe the semantic aspects of derivation, besides the morphotactics and the morphonology, in order to gain a semantically motivated account of morphotactic combinations.

The ‘Item and Process’ approach is fully consistent with the theoretical model of lexicon assumed here: drawing on a cognitive perspective (Langacker, 1991; Schwarze, 2001), the lexicon is conceived as a repository of words related to a conceptual system, *situs* of cognitive experience, categorized objects, properties and events.

The theoretical lines just introduced orient the present formal and computational choices, which combine the finite-state approach of ‘Two-level’ morphology (Kaplan and Kay, 1981; Koskenniemi, 1983; Antworth, 1990) and a feature unification-based Grammar, the above-mentioned LFG (Bresnan, 1982; Ackermann, 1995; Darlymple *et al.*, 1995).

## 1.3 Methodology

We deal with the morphological complexity emerging from naturally occurring data collected into a corpus of written Italian, the ‘Le-Parole’ Corpus (Biagini *et al.*, 2000). Corpus data were systematically integrated with the dictionary definitions of suffixed words, extracted from machine-readable dictionaries, the Dizionario Macchina dell’Italiano (Calzolari *et al.*, 1983), the Dizionario Italiano Sabatini Coletti (Sabatini and Coletti, 1997) and the Zanichelli (Zingarelli, 2000) dictionaries. Some of the not yet represented, but representative suffixed words were picked up from specialistic *lexica* (Lurati, 1999; Schiavetti, 2000), and included into the list of the attested suffixed words.

A quantitative criterion was adopted to rank suffix classes<sup>2</sup> according to their degrees of transparency and lexicalization (Baayen, 1992; Baayen and Lieber, 1991; Baayen and Renouf, 1996). If quantitative measures of suffixed words are reliable indexes of morphological productivity is a matter of debate (Bauer, 2001). Notwithstanding this, it is assumed here that the frequency and

productivity values correlate with the degrees of suffixed word transparency, as reported in previous work (Carota and Prodanof, 2003). In this view, a quantitative analysis of the data is important in order to determine if a synchronic rule applies productively to a given derivational type. On the other hand, polymorphemic word types are ranked according to their relative degree of morphological transparency. This allows the identification of suffix semantic nuclei on which a same synchronic rule may operate with a variable scope. The scope of the rule is narrow if a suffix occurs more frequently in lexicalizations (i.e. suffixed words not yet analyzable on the base of the units of form-meaning that compose them), it is broad if the suffix occurs more frequently in transparent, compositionally analyzable words. These assumptions are congruent with our conception of lexicon (see next Section 2.1) and help us to establish close links between lexicalized and analyzable words, since semantic nuclei are compatible to both a lexicalized and a compositional reading.

## 2 The Formal Structure of Suffixation

### 2.1 Lexicon

The lexicon can be seen as being composed of 1) a set of verbalized concepts consisting of both primary words, which are associated to the cognitive concepts of the conceptual structure, and lexicalized words, idiosyncrasies, affixes and not autonomous words (i.e. particles, prepositions and so on), and 2) a dynamic module, where regular and productive rules take place at the level of inflection and derivation. Suffixes function as lexical operators, which are applied by rule to the primary words to create new words.

The lexicon is thus conceived not as a repository of word entries that are stored with their syntactic and semantic information (or the lemma level) but as a dynamic lexicon, where not only words (namely primary words and opaque words used both as derivational bases), but also suffixes are autonomously listed<sup>3</sup> with their relevant semantic and syntactic information. Consistently with this perspective, we introduce a further distinction pertaining *syntactic words* and *morphological words* (Di Sciullo and Williams, 1989; Schwarze, 2001).

The former consist of whole words, which are the complete atomic units available for building sentences; in Italian, they are marked with regard to syntactic category, valency, number, gender and person. The latter lack the syntactic configuration, are unmarked with regard to number and person and coincide with the stem (see Section 2.2). As we shall see in Section 2.3, this view of the lexicon can be formally tackled by combining a ‘two-level’ approach to morphology with a word grammar, where the lexical information is expressed by hierarchically organized features. Let us highlight some empirical remarks about the formal structure of suffixed words before.

### 2.2 Dealing with morphotactics

Within the view of lexicon just presented, derivational morphology represents a relatively general and systematic formal process, where by formal relations can be traced between pairs of word such as *forma* (form) – *formale* (formal), *comodo* (comfortable) – *comodità* (the quality of being comfortable), *suggerire* (to suggest) – *suggerimento* (suggestion). In these cases, the suffix determines the overall structure of a derived word: it works as an autonomous morphological exponent, which changes transpositionally the grammatical category of the word-base by adding a semantic meaning to it, according to a biunivocal relationship between form and function, in functionalistic terms.

Segmentation, however, is not always a trivial task. For instance, phonological phenomena alter along a diachronic line the transparency of word structure by generating allomorphic alternances. Allomorphy can be thus read as the synchronic projection of diachronic processes related either to phonological change, which produces weak allomorphy, like in the example (a), or to lexical change, which causes strong allomorphy, as in (b).

(a) Degno - Dignità

*Fuoco - Focoso*

(b) fuoco – pirico

*acqua – idrico*

On the other hand, morphologically-related forms do not always depend on fully syntagmatic principles. This happens if derivational chains,<sup>4</sup> as in (c),

lack any intermediate step \*B between A and C; a step which is instead attested for formations in (d).

- (c) A            \*B            C  
 mutuale - \*mutualista - mutualistico  
 pensione - \*pensionista - pensionistico  
 infermiere - \*infermierista - infermieristico
- (d) A            B            C  
 morale - moralista - moralistico  
 semplice - semplicitista - semplicitistico  
 storico - storicista - storicistico

Differently from (d), in (c), the suffix *-istico* seems to be not decomposable in a form *-ist-* and a form *-ic-o*, but it is rather an atomic whole.

As segmentation deals with heterogeneous types of suffixed words, often belonging to different diachronic stratifications of the lexicon, it requires some theoretical decisions on what exactly has to be assumed to form the derivational base and the related suffix. In the present study a stem-based derivational morphology is adopted for dealing with Italian data (Thornton, 1990; 1991; Gaeta, 2002). In the lack of general consensus on the point, it can be argued that such a choice avoids an analysis which would require the introduction of *ad hoc*<sup>5</sup> morphonological rules (Scalise, 1994), in favour of a neater one.

Besides the inherent issues just mentioned, segmentation is a basic pre-requisite for modelling the lexicon within the finite-state approach.

## 2.3 ‘Two-level’ morphology

### 2.3.1 Finite-State, ‘Two-level’ morphotactics and morphonology

Within the finite-state approach, the standard ‘two-level’ morphology describes morphotactics as a continuation class lexicon, which specifies the mutual ordering of the affixes in the suffixed word’s internal structure and models their morphotactic patterns (Koskeniemi, 1983; Karttunen, 1983; Antworth, 1990).

Table 1 reports an example of continuation classes based on the morphotactic patterns formed with the Italian suffixes *-ità*, *-ismo*, *-ista*. These suffixes are mutually exclusive from a paradigmatic point of view. In other words, they can occupy the same positional slot in the internal structure of

**Table 1** Potential right and left combinatory associations for the suffixes *-ità*, *-ismo*, *-ista*.

-ale	}	-ità	-ario		
-are					
-aneo					
-ano					
-bile					
-ico					
-ile					
-ino					
-ivo					
-oso					
-orio					
-ario					
				}	-ismo
	-ista	-ico			

suffixed words. These continuation classes are representative of the following suffixed words:

- *attuabilità* (feasibility) < *attua+bile+ità*,
- *formalismo* (formalism) < *forma+ale+ismo*,
- *inflazionistico* (inflationary)  
 < *inflazione+ista+ico*,
- *trasformazionalistico* (transformationalistic)  
 < *trasfoma+zione+ale+ista+ico*.

The continuation classes just sketched constrain recursive combinations that are unadmitted: for example, we have *contrarietà* (contrariety) ← *contro+ario+ità* (contrario), but not \**contrario+età+ario*.

The continuation classes specify suffixal combinatory properties for the transition networks of finite-state automata. Figure 1 gives an example of suffixed words formed with the word-base *centr-o* (center): *centrare* (to center), *centrale* (central), *centralità* (centrality), *centralizzato* (centralized), *centralizzazione* (centralization), *centralino* (switchboard), *centralinista* (switchboard operator).

In the ‘two-level’ approach, words are built from the finite set of affixes stored in the lexicon according to morphonological rules, which express the contexts where phonological alterations take place and map two levels of word representation. Basically, words are viewed as a correspondence between an underlying level, where affixes are represented as continuation classes and phonological

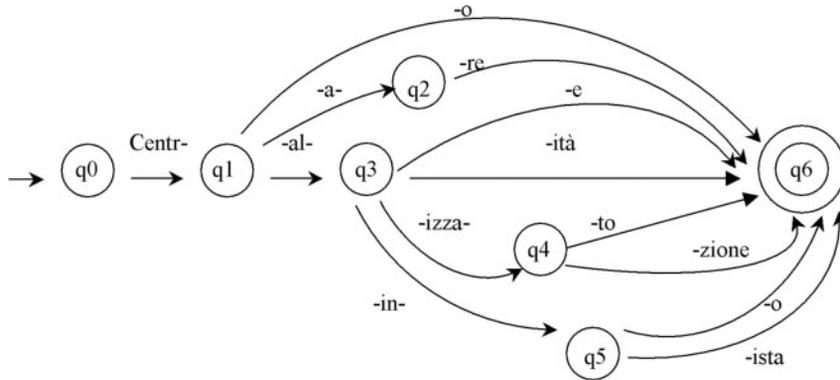


Fig. 1 Finite-state automata consuming suffixed words formed by the base centr-o (center): centrare (to center), centrale (central), centralità (centrality), centralizzato (centralized), centralizzazione (centralization), centralino (switchboard), centralinista (switchboard operator)

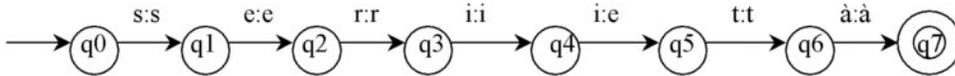


Fig. 2 Simple finite-state transducer for the morphonology underlying the word serietà

rules have not yet been applied, and a surface level corresponding to word spelling. The ‘two-level’ rules are encoded according to the ‘two-level’ formalism<sup>6</sup> and modelled as finite-state transducers (FST), i.e. two-tape automata defining a relation between symbols in sets of pairs of strings<sup>7</sup> and accepting these pairs of strings as input.

As for Italian morphology, some phenomena are rather fuzzy and unpredictable, while others can be depicted by means of rules conformable to the ‘two-level’ formalism, as we exemplify in Appendix I. An example of transducer for one of these rules is displayed in Fig. 2.

2.3.2 Efficiency and limits of the ‘Two-level’ model

The ‘Two-level’ Model has been successful in the field of automatic treatment of morphology, because it offers an efficient general tool. Notwithstanding this, the standard model is not adequate for the purposes of the current work. Specifically, the continuation classes representing the mutual ordering of the affixes in the word structure are not powerful enough to provide a motivated account of the co-selectional

restrictions constraining affixal combinations. In fact, affix co-selection is sensitive to semantic properties. Consequently, we need to capture affixal semantics in order to avoid spurious forms both in automatic analysis and in generation.

Furthermore, we intend to represent formally the distinction between syntactic and morphological words, which we left untreated until now, since the ‘two-level’ representation of the lexicon does not provide a neat separation between inflection and suffixation.

These considerations have led us to choose a framework interfacing, on the one hand, the morphotactic/morphonological modules and the semantic module of suffixation; and, on the other hand, inflection and suffixation. In Section 3.2 we shall see how these requirements can be met by means of the LFG apparatus, which makes explicit and organizes hierarchically in a feature-based lexicon both the syntactic and the semantic information.

3 Semantic Aspects of Suffixation

3.1 Approaching the data

Derivational semantics is largely noncompositional, the derivation of meanings not mirroring the formal

structure of words. Words belonging to a same formal class may exhibit a number of chaotic meanings (Scalise, 1997). Lexicalization and polysy-emy intervene in the organization of the lexicon to obscure word primary semantic motivation. Derived words acquire thus a fixed meaning which is independent of their formal structure to become “signs”-in the Saussurian sense-steadily associated with arbitrary meanings in the lexicon, a transition point towards semantic drifts.

Not only do different meanings correspond to the same formal class, but also distinct formal classes can share the same meaning (as it is illustrated by competing suffixes reported in Appendix II). Concerning what Beard (1995) calls the one(meaning)-to-many(forms)<sup>8</sup> mappings of morphological forms to function, recent studies deverbal nominalizations suggest, for instance, that the semantic behaviour of nominal suffixes is sensitively constrained by the *aktionsart* of the verbal base (Gaeta (2002)). The selection of different semantic features explains, on the one hand, the incompatibility between some suffixes and some bases, and, on the other hand, sheds light on the strategy whereby a particular suffix can be preferred among a group of competing suffixes. This is the case of the deverbal nominalizations in which several suffixes share the same semantic definition (see Appendix II, 3.2.1 (2)-(2.1)-a)). For example, the suffix *-ata* applies to verbal bases denoting a state and attributes to the derived word the features of individuality and countability, whereas the suffixes *-mento* and *-zione* select non-stative verbs and have a merely transpositional, i.e. base-category-changing function.

Furthermore, in contrast with the Aronoff’s Unique Base Hypothesis (UBH) (Aronoff, 1976), postulating that a suffix should select a class which is homogeneous with respect to grammatical category, the same suffix can co-occur with bases that are related to distinct grammatical categories.

A classical example of the heterogeneous range of meanings that correspond to the same suffix is displayed by the dictionary definitions found for the

suffix *-ata*, which also represents an infraction of the UBH (see Schwarze, 2001):

- (1) ‘act related to’ the event denoted by the verbal base, see:  
*andare* (to go) - *andata* (‘act of’ going)  
*trovare* (to find) - *trovata* (‘act of’ finding)
- (2) ‘period related to’ the time expressed by the nominal base:  
*anno* (year) - *annata* (‘period of’ a whole year)  
*mese* (month) - *mesata* (‘period of’ a whole month)
- (3) ‘quantity contained in’ the objects denoted by the nominal base:  
*forchetta* (fork) - *forchettata* (quantity contained in a fork)  
*secchio* (pail, bucket) - *secchiata* (quantity contained in a pail)
- (4) ‘action, behaviour specifically related to’ the person denoted by the nominal verb:  
*ragazzo* (boy) - *ragazzata* (boy’s typical behaviour)  
*pagliaccio* (clown) - *pagliacciata* (clown’s typical behaviour)

Regarding the definition in (1), Mayo *et al.* (1995), and Mayo (2000) point out how the nominalizations ending in *-ata* can be interpreted as individual or instantiated events. On the other hand, Gaeta (2002, pp. 154–155) establishes significant connections between these meanings and argues that the periphrasis ‘individual (quick) act, typical of’ represents the core semantic value underlying the senses listed above. In fact, a double motivation, raised from the copresence of both a noun and a verb sharing the same stem (e.g. *canto*, ‘song’/*cantare*, ‘to sing’ → *cantata*, ‘individual act of singing’, countable and pluralizable), could originate the mixed categorial scope of the rule involving *-ata*.

In conclusion, the denominal nouns formed by using the suffix *-ata* offer an example of under-specified derivational semantics (UDS), which is one of the major characteristics of the Italian derivation, as pointed out in Schwarze (2001). The UDS does not mark the morphological information that is relevant for the interpretation of derived

words, and can be defined as being the phenomenon whereby:

Given a word  $w = X + Y$  (where  $X =$  base,  $Y =$  suffix) having the meaning  $M$ , an attribute of  $M$  (i.e. the semantic role  $R$  in which the entity denoted by  $X$  is involved) has to be specified in order to interpret  $M$ .

The complexity of the interpretation depends thus on the need to accommodate both the variable  $X$  and the variable  $R$ , a condition met only if the context provides the source of information required for resolving  $R$ . In other words, a semantic interpretation becomes sensitive to pragmatic factors, especially in the case of newly occurring words (Dressler and Merlini Barbaresi, 1994).

To understand how the meaning of the suffixes constrains their possible combinations, the notion of thematic roles (i.e. the general roles that participants play in the event denoted by the verb), can also be invoked for the analysis of the suffixed words.

If we consider, for example, the verbal suffix *-izzare* in (1)-(1.1) and (3)-(3.2), its core semantic value consists of a change of status for the argument position occupied by the theme (i.e. object), which can be expressed by the following formula, by using lambda calculus (Church, 1941):

(5)  $\lambda X \lambda Y \epsilon S \text{ CAUSE } (X, \text{COME\_ABOUT } (S(Y)))$

where  $X$  is the patient/theme, and  $S$  is the entity denoted by the base of suffixed words in *-izzare*. Such core semantic value encompasses the instantiations of *-izzare*. The semantics of this suffix can contain also an inchoative meaning, which can be explicitated by a CAUSE predicate, in addition to the previous formula. For example, *modernizzare* (to make modern), which has an adjectival base, may have the following semantic representation:

(6)  $\lambda S \lambda X \text{ CAUSE } (X, \text{COME } (S, \text{ABOUT } (\text{moderno})))$

The verb *valorizzare* (to give value), which has a nominal base, is characterized by a slightly different semantics, although the meaning of status-changing just mentioned is not altered:

(7)  $\lambda S \lambda X \text{ CAUSE } (X, \text{COME } (S, \text{ABOUT } (\text{HAVE\_MORE: VALUE})))$

In order to make a further generalization, it can be claimed that the overall semantics of *-izzare* is given by the formula:

(8)  $\lambda S \lambda X \text{ CAUSE } (X, \text{COME } (S, \text{ABOUT } (Y)))$ .

In 3.2 this semantic information is represented into LFG word structures (i.e. m-structures), where the semantic roles fitting the argument structure of suffixed words are made explicit.

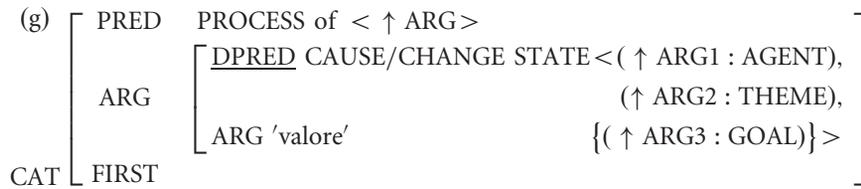
At this level of abstraction, specific or connotative shades of meaning may remain underspecified. For example, the suffix *-oso* shown in Appendix II loses in part some of its peculiar connotations, i.e. the meaning of 'rich in, full of' and of 'having the marked characteristic of', which are recoverable, respectively, from *pietoso* ('stony'), *roccioso* ('full of rocks, rocky'), *gommoso* ('gummy'), *pastoso* ('mellow' and also 'palatable'), and *morbidoso* ('softish'). Semantics modulates in a relevant way the constraints of co-selection between word bases and suffixes, as well as the combinatory patterns of suffix chains.

For the present purposes, it becomes thus of interest to make this type of information explicit, by specifying the subcategorization features more precisely into m-structures in LFG.

A preliminary specification of the suffixal semantics is sketched in Appendix II, which reports abstract generalizations on the lists of words extracted from the Corpus. The patterns of suffixed words have been systematically related to dictionary definitions and grouped according to the form and the meaning(s) of the suffix. Suffixed word types are represented in the form of schematic lexical rules, reflecting affix transcategorizations. The general semantic abstractions so defined intend to represent semantic stereotypic *nuclei* upon which the LFG m-structures can be built.

### 3.2 Describing derivational semantics in the LFG formalism

The semantics of the derived words is formalized within the unification-based framework of LFG (Kaplan and Bresnan, 1995). LFG offers a declarative formalism defining feature structures as sets of pairs of attributes i.e. atomic symbols, and of values, i.e. either atomic symbols or features structure.



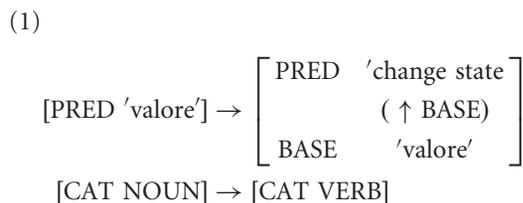
The basic mechanism in LFG, as well as in the other unification-based grammars, is unification, an operation of comparison and active recognition of the functional information internalized in the feature structures (Ferrari, 1991).

To apply LFG to the morphological domain, word structure needs to be determined, and a feature-based representation of the lexicon has to be assumed in which the features are hierarchically organized.

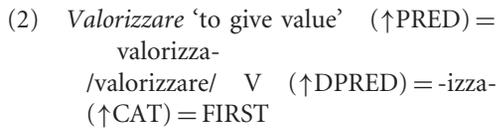
In LFG, two parallel levels of morphosyntactic representation are specified. The first level consists of constituent structures (c-structures), which have the form of context-free structure trees for words; at this level of representation, grammatical functions are associated with constituents, by means of c-structure rules. The second level consists of feature structures (f-structures, m-structure for morphological semantics), and sets of pairs of attributes and values; attributes may be features or functions.

Terminal symbols are marked with a projection equation (PE), indicated by the notation ( $\uparrow = \downarrow$ ), which means that the functional information attached to the symbol is unified with the functional information attached to the constituent node identified by the head of the rule.

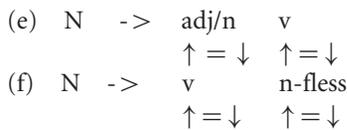
For instance, the derivation of the verb *valorizzare* considered in Section 3.1 and formed by a nominal base, is schematically described in (1) by using operations on sets of equations  $\langle \text{Attribute} \rangle \langle \text{Value} \rangle$ , where the Attributes PREDicate and CATegory are changed during the deverbal derivation.



The scheme corresponds to the general transcategorization  $N \rightarrow V$ . The lexical form (2) for *valorizzare* contains the word citation form and functional equations which project the attributes and the relative values of the affixes into the whole word structure.

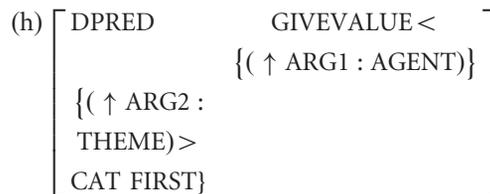


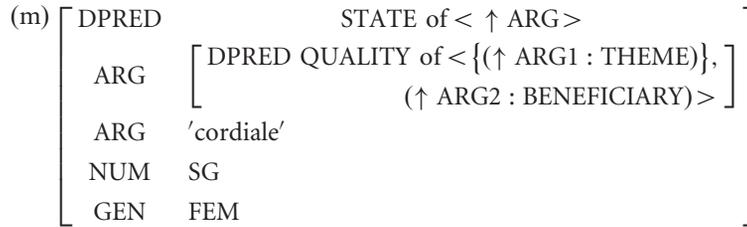
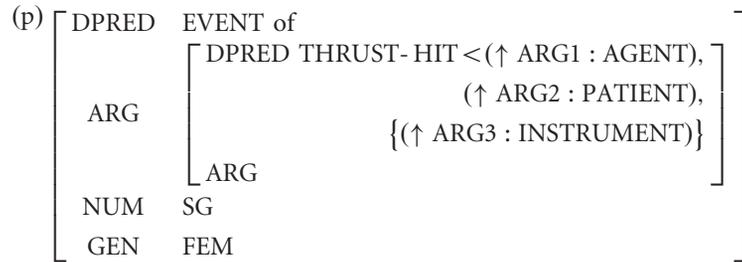
This word structure is determined by two rules, the rule (e), which prevents both the base-word genre and inflectional class from being transferred to the derived word; and the rule (f), which concerns in particular the inflectional information:



Equations in the lexical form (2) are represented as a directed acyclic graph in (g). In the graph, the terms of the pairs are chained from left to right to yield the corresponding path, and the thematic roles, which are part of the Lexical Form, are visualized (see (g) above).

In the graph, the bound roles of the lambda-expression (7) given in Section 2, are mapped onto the argument structure. The DPRED outlined in (g) specifies an action concept, where the GOAL corresponds to 'value', which has to be related to an object concept, as in (h).





In what follows, some schematizations are provided for other representative derivational types:

- (3) Adj -> N  
 Cordialità ‘the quality of being hearty, kindness’  
 /cordialità/, Noun ( $\uparrow$ PRED) = cordial-  
 ( $\uparrow$ DPRED) = -ità’  
 ( $\uparrow$ NUM) = SG  
 ( $\uparrow$ GEN) = FEM  
 ( $\uparrow$ NUM) = SG

Rules (i), for genre and inflectional class and (l) for inflection:

- (i) Adj -> n      adj/n  
 $\uparrow = \downarrow$      $\uparrow = \downarrow$   
 (l) Adj -> n      n-fless  
 $\uparrow = \downarrow$      $\uparrow = \downarrow$

See Acyclic graph (m) above.

- (4) N -> N  
 Manata “hit with the hand”  
 /manata/, Noun ( $\uparrow$ PRED) = man-  
 ( $\uparrow$ DPRED) = -ata  
 ( $\uparrow$ NUM) = SG  
 ( $\uparrow$ GEN) = FEM  
 ( $\uparrow$ NUM) = SG

Rules for gender and category blocking and for inflection:

- (n) Adj -> n      adj/n  
 $\uparrow = \downarrow$      $\uparrow = \downarrow$   
 (o) Adj -> n      n-fless  
 $\uparrow = \downarrow$      $\uparrow = \downarrow$

See Acyclic graph (p) above.

As additional remark, we wish to highlight that, together with semantic information, we have described inflection consistently with the distinction between morphological and syntactic words previously introduced. In a lexicon hierarchically conceived within the LFG framework, words are appropriately provided with inflectional endings after having undergone suffixation category-changing operations.

## 4 Concluding Remarks

The attempt of formalization presented in the article meets the requirements of a future system, based on the integration of the finite-state method and of a feature unification-based grammar for encoding structural and semantic properties associated with

suffixed words. The word grammar appropriately defines coselectional restrictions by ‘context-free’ lexical rules and restrictions on semantic features. Admitting a distinct module for morphotactic and morphological rules, the feature unification-based grammar supports the overall model sketched in the paper, which extends the standard ‘two-level’ morphology and recalls the architecture of a system such as KIMMO (Antworth, 1990), which employs, however, a word grammar based on a different formalism, the PATRII (Shieber, 1986).

The choice of LFG turns out to be consistent with our overall theoretical premises. Allowing the equation of estensional and intensional concepts, the LFG apparatus is congruent with the interplay between the cognitive and the verbalized objects assumed here: semantic interpretation, as it is encoded in the m-structures, seems to provide the mutual link between these two types of objects.

Moreover, the general view of the lexicon as forming a hierarchical organization of features and, in particular, of semantic features enables us to reconcile lexicalized and analyzable words, by means of semantic nucleus identified for each word type and potentially compatible with both a lexicalized and a compositional reading, according to different levels of generalization.

As a conclusive remark, we wish to point out that the study of derivational semantics is far from being exhaustive. Finer-grained analysis and a consequent deeper understanding of the semantic paths followed by the affixal combinations would lead to an advantageous improvement of the flexibility regarding both formal descriptions and computational applications.

## Acknowledgements

I am grateful to Prof Jeff Rydberg-Cox for contributing suggestions to the article. I also owe an heartfelt thanks to Prof Christoph Schwarze for providing comments about some aspects of this article. I am solely responsible for the content of the final version of the paper.

## References

- Ackermann F.** (1995). Lexical functional grammar. In Verschueren, J., Óstman, J.-O., Blommaert, T., and Bulcaen, C. (eds), *Handbook of Pragmatics*. Amsterdam: John Benjamins, pp. 346–350.
- Anderson, S. R.** (1977). On the formal description of inflection. Papers from the 13th annual regional meeting of the Chicago Linguistic Society, pp. 15–44.
- Anderson, S. R.** (1992). *A-Morphous Morphology*. Cambridge: Cambridge University Press.
- Antworth, E. L.** (1990). PC-KIMMO: a two-level processor for morphological analysis, Occasional Publication in *Academic Computing*, 16. Dallas: Summer Institute of Linguistics.
- Aronoff, M.** (1976). *Word Formation in Generative Grammar*. Cambridge, MA: The MIT Press.
- Aronoff, M.** (1994). *Morphology by itself. Stems and Inflectional Classes*. Cambridge, MA: The MIT Press.
- Baayen, H.** (1992). Quantitative aspects of morphological productivity. In Booj G. and van Marle J. (eds), *Yearbook of Morphology 1991*. Dordrecht: Kluwer, pp. 109–49.
- Baayen, H and Lieber, R.** (1991). Productivity and English Word-formation: a Corpus-based Study. *Linguistics*, 29: 801–43.
- Baayen, H. and Renouf, A.** (1996). Chronicing the times: productive lexical innovations in an English newspaper. *Language*, 72: 69–96.
- Bauer, L.** (2001). *Morphological Productivity*. Cambridge: Cambridge University Press.
- Beard, R.** (1995). *Lexeme-Morpheme Base Morphology: a general theory of Inflection and Word Formation*. Albany, NY: SUNY Press.
- Biagini, L., Bindi, R., Goggi, S., Marinelli, R., Picchi, E., and Rossi, S.** (2000). *LE-PAROLE Italian Corpus Documentation WP2.1*. Pisa: ILC-Technical report.
- Bresnan, J.** (1982). *On the mental representation of Grammatical Relations*. Cambridge, MA: The MIT Press.
- Calzolari, N., Ceccotti, L. M., and Roventini, A.** (1983). *Documentazione sui tre nastri contenenti il DMI*. Pisa: ILC-Technical Report.
- Carota, F. and Prodanof, I.** (2003). A corpus-based account of Italian suffix productivity. *Linguistica Computazionale*, XVI–XVII: 201–20.
- Church, A.** (1941). *The Calculi of Lambda Conversion*. Princeton, Princeton University Press: NJ.

- Darlymple, M., Kaplan, R. M., John III Maxwell, T., and Zaenen, A.** (eds) (1995). *Formal Issues in Lexical-Functional Grammar*. CSLI Lecture Notes 47. Stanford: CSLI Publications.
- Di Sciullo, A. and Williams, E.** (1989). *On the Definition of Word*. Cambridge, MA: The MIT Press.
- Dressler, W. U. and Merlini Barbaresi, L.** (1994). *Morphopragmatics*. Berlin-New York: Mouton de Gruyter.
- Gaeta, L.** (2002). *Quando i verbi compaiono come nomi*. Materiali Linguistici Università di Pavia. Milano: Franco Angeli.
- Ferrari, G.** (1991). *Introduzione al Natural Language Processing*. Bologna: Calderini.
- Jurafsky, D. and Martin, H. J.** (2000). *Speech and Language Processing: an Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. New Jersey: Prentice Hall.
- Kaplan, R. and Kay, M.** (1981). Phonological rules and finite-state transducers, Paper presented at the annual meeting of the Linguistic Society of America, New York.
- Kaplan, R. and Bresnan, J.** (1995). Lexical functional grammar: a formal system for grammatical representation. In Darlymple *et al.* (eds), *Formal Issues in Lexical-Functional Grammar*. CSLI Lecture Notes 47, Stanford: CSLI Publications, pp. 173–281.
- Karttunen, L.** (1983). KIMMO: a general morphological processor. *Texas Linguistic Forum*, 22: 163–86.
- Koskenniemi, K.** (1983). *Two-level morphology: a general computational Model for Word-Form Recognition and Production*, Publ 11. Helsinki University.
- Langacker, R. W.** (1991). *Concept, Image, and Symbol. The Cognitive Basis of Grammar*. Berlin: Mouton de Gruyter.
- Lurati, O.** (1999). *3000 parole nuove: la neologia negli anni 1980-1990*. Bologna: Zanichelli.
- Matthews, P. H.** (1972). *Morphology*. Cambridge: Cambridge University Press.
- Mayo, B.** (2000). *A Computational Model of Derivational Morphology*. Ph.D. thesis, University of Hamburg.
- Mayo, B., Schepping, Th. M., Schwarze, C., and Zaffanella, A.** (1995). Semantics in the derivational morphology of Italian: implications for the structure of the lexicon. *Linguistics*, 33: 883–938.
- Rainer, F.** (1999). I meccanismi di diffusione di allomorfi: il caso della rianalisi del tipo latino «man-u-alis» in italiano. *AttiSLI*, 41: 423–52.
- Sabatini e Coletti** (1997). *Dizionario Italiano Sabatini-Coletti in CD-Rom*. Firenze: Giunti.
- Scalise, S.** (1994). *Morfologia*. Bologna: Il Mulino.
- Scalise, S.** (1997). Aspetti problematici della semantica in morfologia derivazionale. In Bernini, G., Cuzzolin, P. L., and Molinelli, P. (eds), *Ars Linguistica*. Roma: Bulzoni, pp. 491–505.
- Schiavetti, V.** (2000). *Le parole de Il Sole 24 ORE*. II ed. Milano: Il Sole 24 ORE.
- Schwarze, C.** (2001). *Aspetti semantici della formazione delle parole*, Arbeitspapier 107, Fachbereich Sprachwissenschaft der Universität Konstanz.
- Shieber, S. M.** (1986). *An Introduction to Unification-Based Approaches to Grammar*. Stanford: CSLI Publications.
- Spencer, A.** (1991). *Morphological Theory. An Introduction to Word Structure in Generative Grammar*. Oxford: Basil Blackwell.
- Sproat, R.** (1992). *Morphology and Computation*. Cambridge, MA: The MIT Press.
- Thornton, A. M.** (1990). Sui deverbali italiani in -mento e -zione (I). *Archivio glottologico italiano*, 75: 169–207.
- Thornton, A. M.** (1991). Sui deverbali italiani in -mento e -zione (II). *Archivio glottologico italiano*, 76: 79–102.
- Thornton, A. M.** (1998). Quali suffissi nel ‘vocabolario di base’? In Leoni F. A., Gambarara D, Gensini S, Lo Piparo F., and Simone R. (eds), *Ai limiti del linguaggio*. Bari: Laterza, pp. 385–397.
- Zingarelli** (2000). *Vocabolario della lingua italiana di Nicola Zingarelli*. Bologna: Zanichelli.

## Notes

- 1 Neither prefixed words or adjectival suffixes will be discussed here.
- 2 The ‘LE-PAROLE’ Corpus (Biagini *et al.*, 2000) consists of 20,948,736 words. This number refers to a composition of books (3,752,643 words), newspapers (14,596,694 words), periodicals (959,255 words), miscellaneous works (1,640,189 words). The material is referred to the years from 1970 to 1996. The newspapers, which represent the most relevant portion, cover the years 1992–1996.
- 3 This criterion is especially suitable with highly inflectional or highly agglutinative languages like Finnish (Koskenniemi, 1983, Sproat, 1992).
- 4 By derivational chain I refer to (partially) recursive and step by step suffix application.
- 5 Nominalizations ending in *-azione*, *-ezione*, *-uzione*, *-izione* represent a vexata crux: the solution here

preferred is in favour of a stem-based analysis of the productive forms. Hence, the suffix form assumed in this case is *-zione*.

- 6 The 'two-level' formalism differs from the rewriting rules of the generative framework (Kaplan and Kay, 1981). A detailed specification of the 'two-level' formalism is provided by Antworth (1990) and Sproat (1992).
- 7 Sets of strings are defined by the formal regular language which corresponds to a finite-state automaton (FSA), in terms of bi-univocal relation. A FSA accepts as its input a single string. Sets of pairs of strings are an extension of regular languages which defines regular expressions and corresponds to a FST.
- 8 Notations in brackets are mine.

\*Sporadic and contextually independent consonantic epenthesis affects instead words like *Perù - Peruviano*, *Congo- Congolese*, *Caffè - caffettiera*, *Amalfi - Amalfitano*.

## Appendix I

Main Italian morphonological rules in the 'two-level' formalism:

- (1)  $V[-\text{stressed}]_0 \leftrightarrow \_ + V$

The rule states that the final vowel of the base is always deleted in front of another vowel. This happens when the final vowel of the base is unstressed, otherwise (i.e. if stressed) it remains unaltered. Examples: *rosso* (red) - *rossastro* (reddish), *virtù* (virtue) - *virtuoso* (virtuous).

- (2)  $i:e \leftrightarrow \_ i$  (tà)

A dissimilation rule affects the sequence *i+i*, occurring in words which are formed, for example, by the suffix *-ità*, for example, *serio* (serious) - *serietà* (seriousness), *bonario* (meek) - *bonarietà* (good nature).

- (3)  $C:C \leftrightarrow V[-\text{voiced}] [+ \text{stressed}]C \_ + V$

Epenthesis has a restricted domain, which often involves borrowed words ending with a consonant: *gas* (gas) - *gassoso* (gassy)\*.

- (4)  $0: u/i \leftarrow (t:t, s:s) \_ (-\text{ale}, -\text{are}, -\text{ario}, -\text{oso})$

This rule can be introduced to account for a particular case of allomorphy (Thornton, 1998), although the phenomenon is not predictable a

priori, because it is influenced by the analogical and coderived models. Insertion of a not etymological epenthetic glide *j/w* in new formation with *-ale*, *-ario*, *-are*, *-oso*, e.g. *Internet-internettiano* (related to the Internet), *salto* (jump)- *saltuario* (discontinuous) can be explained as a case of reanalysis of the lexicon belonging to Latinate heritage (especially latin forms belonging to the third and fourth declinations) (Rainer, 1999). Therefore, the 'two-level' (4) says that the glide does not always appear in the context specified at the right of the arrow.

## Appendix II

### (1) Denominal Derived Words

General Formula:  $X + Y = Z$ , where  $X = N(\text{oun})$ ,  
 $n/v/\text{adj}$   $Y = \text{Nominal/Verbal/}$   
 $\text{adjectival Suffix}$ ,  
 $Z = \text{Suffixed Word}$

- (1.1)  $X + Y_v \rightarrow Z$ , where  $Z = V$ ,  $X = N$   
 a. Core semantic value  $Y_v = \text{'EVENT/PROCESS of | X'}$ ,  
 where  $Y_v: \{\text{IZZARE}, \text{IFICARE}\}$
- (1.2)  $X + Y_n \rightarrow Z$ , where  $Z = N$ ,  $X = N$   
 b. Core semantic value  $Y_n = \text{'|PERTAINING to/FOLLOWING| X'}$   
 where  $Y_n: \{\text{ISMO}, \text{ESIMO}\}$   
 c. Core semantic value  $Y_n = \text{'|ACTIVELY INVOLVED in|X'}$ ,  
 where  $X = N/\text{ProperN}$ ;  $Y_n: \{\text{ISTA}\}$   
 d. Core semantic value  $Y_n = \text{'|QUANTITY of| X'}$ ,  
 where  $X = N$ ;  $Y_n: \{\text{AGLIA}, \text{AME}, \text{UME}, \text{ATA}\}$   
 e. Core semantic value  $Y_n = \text{'|HIT with/of to| X'}$ ,  
 where  $X = N$ ;  $Y_n: \{\text{ATA}\}$   
 f. Core semantic value  $Y_n = \text{'|PLACE with| X' >}$ ,  
 where  $X = N$ ;  $Y_n: \{\text{ETA}, \text{ETO}\}$   
 g. Core semantic value  $Y_n = \text{'|PLACE for| X'}$ ,  
 where  $X = N$ ;  $Y_n: \{\text{ILE}\}$   
 h. Core semantic value  $Y_n = \text{'|CONTAINING| X'}$ ,  
 where  $X = N$ ;  $Y_n: \{\text{ATO}\}$

- i. Core semantic value  $Y_n = \text{'[FEMININ of] X'}$ ,  
where  $X = N$  masculine;  $Y_n: \{\text{[ESSA]}\}$

(1.3)  $X + Y \rightarrow Z$ , where  $Z = \text{Adj}$

- j. Core semantic value  $Y_{adj} = \text{'[RELATED to] X'}$ ,  
where  $X = N$ ;  $Y_{adj}: \{\text{[ALE], [TORIO], [INO], [EO], [ARE], [ANO], [ESCO], [ASCO], [ICO], [ISTICO], [IVO], [IZIO], [ILE], [OSO]}\}$
- k. Core semantic value  $Y_{adj} = \text{'[SIMILAR to] X'}$ ,  
where  $X = N$ ;  $Y_{adj}: \{\text{[ACEO], [INO], [OIDE], [IGNO], [ARE]}\}$
- l. Core semantic value  $Y_{adj} = \text{'[BELONGING/RELATED to] X'}$ ,  
where  $X = N$ ;  $Y_{adj}: \{\text{[ARDO], [ASCO], [ICO], [ANO]}\}$
- m. Core semantic value  $Y_{adj} = \text{'[INHABITANT of/FROM] X'}$ ,  
where  $X = N$ ;  $Y_{adj}: \{\text{[ANO], [INO], [ETANO], [ITANO], [ANEO], [EO], [ESE], [ATE], [OTA], [ARE], [ESCO], [ICO], [ENO]}\}$

(2) Deverbal Derived Words

General Formula:  $X + Y$ , where  $X = V$ ,  $Y = A$  (unspecified Affix)

(2.1)  $X + Y_n \rightarrow Z$ , where  $Z = N$  (Nominalizations)

- n. Core semantic value for  $Y_n = \text{'[ACT/EFFECT/MODUS] OF] X'}$ ,  
where  $X = V$ ; Competing  $Y_n: \{\text{[ATA- (ITA1)-UTA-], [IONE], [MENTO], [AGGIO], [ITO], [ITA2], [URA], [URA], [NZA], [IO], [ERIA]}\}$
- o. Core semantic value for  $Y_n = \text{'[STATE of] X'}$ ,  
where  $X = V$ ;  $Y_n: \{\text{[ATO], [NZA], [IA], [IO], [ORE]}\}$
- p. Core semantic value  $Y_n = \text{'[SOMETHING for/RELATED to] X'}$ ,  
where  $X = V$ ;  $Y_n: \{\text{[TOIO], [TOIA], [TORIO]}\}$

- q. Core semantic value  $Y_n = \text{'[ON THE POINT of] X'}$ ,  
where  $X = V$ ;  $Y_n: \{\text{[ANDO]}\}$

- r. Core semantic value  $Y_n = \text{'[AGENT of] X'}$ ,  
where  $X = V$ ;  $Y_n: \{\text{[TORE], [TRICE], [IERE], [IERO], [ANTE], [ARO]}\}$

(2.2)  $X + Y_n \rightarrow Z$ , where  $Z = \text{Adj}$

- s. Core semantic value  $Y_{adj} = \text{'[REFERRING to] X'}$ ,  
where  $X = V$ ;  $Y_{adj}: \{\text{[EVOLE], [IVO], [TORIO], [ARIO]}\}$

(2.3)  $X + Y_v \rightarrow Z$ , where  $Z = V$

- t. Core semantic value =  $\text{'[GRADE of] X'}$ ,  
where  $X = V$ ;  $Y_v: \{\text{[ELLARE], [ARELLARE], [ERELLARE], [ICCHIARE], [ACCHIARE], [ETTARE]}\}$

(3) Deadjectival Derived Words

General Formula:  $X + Y$ , where  $X = \text{Adj}$ ,  $Y = A$  (unspecified Affix)

(3.1)  $X + Y \rightarrow Z$ , where  $Z = N$

- u. Core semantic value =  $\text{'[QUALITY of] X'}$ ,  
where  $X = \text{Adj}$ ;  $Y_{adj}: \{\text{[ITA]}, \text{[(IT)UDINE], [EZZA], [IZIA], [IGIA], [AGGINE], [IA]}\}$

(3.2)  $X + Y \rightarrow Z$ , where  $Z = V$

- v. Core semantic value  $Y_v = \text{'[CHANGE STATUS of/MAKE] X'}$ ,  
where  $X = \text{Adj}$ ;  $Y_v: \{\text{[izzare], [ificare]}\}$

(3.3)  $X + Y_n \rightarrow Z$ , where  $Z = N$

- w. Core semantic value  $Y_n = \text{'[PHENOMENON related to] X'}$ ,  
where  $X = \text{Adj}$ ;  $Y_n: \{\text{[ISMO], [ESIMO]}\}$
- x. Core semantic value  $Y_n = \text{'[ACTIVELY INVOLVED in] X'}$ ,  
where  $X = \text{Adj}$ ;  $Y_n: \{\text{[ISTA]}\}$

(3.4)  $X + Y \rightarrow Z$ , where  $Z = \text{Adv}$

- y. Core semantic value  $Y_{adv} = \text{'[WAY/MANNER of] X'}$ ,  
where  $X = \text{Adj}$