

Semantic relationships within a set of verbal entries in the Italian Lexical Database

Donella Antelmi — Adriana Roventini

1. Introduction

This paper refers to a study carried out at the Institute for Computational Linguistics in Pisa as part of a project concerning the Italian Machine Dictionary (DMI) as a lexical database.

Within the field of computational linguistics there is an increasing need to represent the lexical entry in a consistent and exhaustive way, so that all the syntactic and semantic implications of the lemma are made explicit. In particular a lexical entry, for example of a verb, can be considered exhaustive if it is represented not only with its syntactic but also semantic features which must be given a unitary representation whenever possible. Over the last years a systematic study of the lexicon has shown that classes of verbs with homogeneous meanings often have a similar behaviour at the syntactic level. Furthermore, the morphological links which exist in a language are all strictly connected with syntactic and semantic attributes. They represent different structuring levels within the lexicon with many regular features which can be usefully correlated to particular semantic and syntactic aspects. In a previous study examining more than 20,000 nouns and adjectives derived by suffixation we connected morphologic and semantic features, and we created a new cluster in the database where these links were physically represented by means of pointers and semantic codes. For all these reasons the starting point of this work was also derivational morphology and the grammatical category considered was that of the verb.

In view of the electronic processing to be adopted, among the more than 6,000 derived verbs present in the database, we decided to choose the subset showing the largest number of regularities from both a semantic and a syntactic point of view.

The aim of our research was to make explicit some features which could not be immediately retrieved from the database but which have come out during our analysis.

The following were the most important phases of our work:

a) to explicit the link between the verb and its base in order to enable the connection with other derivatives from the same base and therefore to refer to the work already carried out for the adjectives and the nouns. E.g:

Incuriosire (to make curious)	→	Curioso (curious)
Curioso (curious)	→	Incuriosire (to make curious)
Curioso (curious)	→	Curiosità (curiosity)

b) to analyze the causative/inchoative alternation largely present in the subset;

c) to check whether the subset presented other significant regularities even from the point of view of the aktionsart;

d) to codify the regularities identified.

2. Description of the corpus

As already said, the subset is considerably homogeneous both from the morphological and the semantic/syntactic point of view. It mainly consists of a group of verbs derived from nouns and adjectives (mostly parasynthetic) which indicate processes of transformation and change. This semantic feature appears clearly, at the lexicographical level, from the repetition in all their definitions, of two identical models or formulae which alternate according to whether the verb is observed in its transitive or intransitive use (and relevant meaning). The two genus terms are 'to make' (Italian = 'rendere', in a few cases 'to reduce', Italian = 'ridurre') in its transitive use and 'to become' (Italian = 'diventare' or 'divenire') in its intransitive use, as we can see from the following examples:

Acidificare	= rendere acido (to make acid)
Acidificare	= diventare acido (to become acid)
Aggravare	= rendere più grave, pesante (to make worse, heavier)
Aggravare	= diventare più grave, pesante (to become worse, heavier)
Assordare	= rendere sordo (to deafen)
Assordare	= diventare sordo (to become deaf)
Imbellire	= rendere bello (to make beautiful)
Imbellire	= diventare bello (to become beautiful)
Ingrassare	= rendere grasso (to make fat)
Ingrassare	= diventare grasso (to become fat)
Invecchiare	= rendere vecchio (to make old)
Invecchiare	= diventare vecchio (to become old)

From a quantitative point of view the corpus is formed by approximately 1,700 definitions but the lexical entries of the database to which they refer are more than half that number, since there are (although not very many) cases of only transitive verbs such as: 'allergizzare', 'aromatizzare', and others of only intransitive verbs such as 'imbarbogire', 'imbolsire'.

In this respect, it is interesting to note that the verbs having only the transitive form are in most cases suffixed verbs (ending in -izzare, -ificare) or verbs created by transcategorization, while the verbs having only the intransitive form are generally the prefixed ones. From a morphological point of view the verbs are divided into:

- a) derivatives by simple transcategorization, such as 'fiorire', 'chiarire', 'scurire';
- b) derivatives by suffixes, such as 'acidificare', 'chiarificare', 'centralizzare';
- c) derivatives by prefixes and transcategorization or parasynthetic derivatives such as 'arrossire', 'imbalordire', 'indebolire', 'invecchiare'.

The most frequent prefix is in-, which implies movement towards a state, a condition ('imbellire', 'invecchiare', 'incivilire'), followed by a-, in the sense of approach, direction, trend ('addomesticare', 'abbrutire', 'allentare'), ri- with an intensive or repetitive value or both ('rammorbire', 'rasserenare', 'riabilitare'), s- with an inten-

sive value as in ('scaldare', 'slargare') or a deprivative value as in ('sfoltire', 'spigrire'), and also di-, de-, dis-, with both deprivative ('depilare') and intensive value ('disseccare'). The most frequent suffix is -izzare followed by -ificare.

One interesting feature within the corpus is represented by the multiple formations from the same base. These are very frequent (as we can see in Table A) and they are due to either syntactic or semantic differentiations, or to diachronic stratifications caused by changes in the use of the language.

In many cases a form has become obsolete and has thus been substituted by another form derived from the same base but with a different prefix or different class of conjugation, as we can see in these examples:

Chiarare	(obsolete)	Chiarire	(new form)
Scurare	"	Scurire	"
Abbruttire	"	Imbruttire	"
Sterilire	"	Isterilire	"

In other cases, different prefixes give an opposite meaning as in:

Ingrassare	Sgrassare
Associare	Dissociare
Infiammare	Sfiammare
Gassare	Sgassare

or a different meaning, as in:

Allargare	Slargare
Appianare	Spianare
Affilare	Sfilare

or opposition literal sense/figurative sense:

Avvelenare	(both literal and figurative sense)
Invelenire	(only figurative sense)
Addolcire	(both literal and figurative sense)
Indolcire	(only literal sense)

Another different case is that in which the prefix or the change of conjugation (are/ire) distinguishes different meanings which are due to polysemic or homographic bases. Thus we have:

Ammollare	from	molle	= wet
Ammollire	from	molle	= tender
Ammezzare	from	mezzo	= half
Ammezzire	from	mezzo	= rotten
Allentare	from	lento	= loose
Rallentare	from	lento	= slow, etc.

3. Grammatical category and ergativity

Following the analyses proposed by a number of linguists (e.g. Perlmutter, Burzio) the most recent proposals of lexicographers are in favour of a grammatical description of the verbal entries which go beyond the distinction between the transitive and the intransitive form. In particular importance has been given to the concept of ergativity, by which it is possible to describe the regular behaviour of a wide class of verbs which present an alternation between the transitive and the intransitive form, often defined as causative/inchoative alternation, as for example in:

Mary broke the glass
The glass broke

This phenomenon is also present in Italian verbs, and the majority of the verbs of our corpus is involved in this type of alternation. E.g.:

- 1) il fabbro arroventa il ferro
- 2) il ferro si arroventa
- 3) troppo sole fa avvizzire /avvizzisce la pelle
- 4) la pelle si avvizzisce / avvizzisce al sole
- 5) Luisa si impigrì / impigrì e non fece più ginnastica
- 6) il caldo fece impigrire / impigrì Luisa
- 7) tutte le guerre del passato impallidiscono di fronte all'ultima
- 8) il terrore fece impallidire / impallidì il volto di Luisa
- 9) i commercianti ribassano i prezzi
- 10) quando la domanda cala i prezzi ribassano
- 11) i falegnami invecchiano il legno con prodotti speciali
- 12) si sa che il fumo fa invecchiare la pelle
- 13) lo scrittore, dimenticato, invecchio' in miseria
- 14) * Giorgio evapora l'acqua
- 15) Giorgio fa evaporare l'acqua
- 16) l'acqua evapora

The causative/inchoative alternation can be expressed in Italian in two ways which seem to depend on the specific verb, and for this reason it cannot be predicted.

The first consists in the regular alternation as in English between transitive and non-transitive form, with no other morphosyntactic detail (examples 9 and 10).

The second way in which the alternation appears is with the pronominal form 'si' with the non-transitive form of the verb (examples 1 and 2).

This is a first important division regarding the morphosyntactic behaviour of these verbs.

The other examples reported show several variations either in the causative construction, where one can either use or not use the causative auxiliary 'fare' (see examples 3, 6), or in the inchoative construction where we can use or not the pronominal form 'si' (examples 4, 5). The causative form 'fare' can be used with most of the verbs, but in the particular subset we are considering, this type of construction makes it possible to re-establish a symmetry when the transitive version of the alternation is

missing (see examples 14, 15 and 16). In the other cases it is only a stylistic variation (but it is preferred when the transitive version of the verb is no longer topical as in 8).

Summing up, the verbs which present the causative/inchoative alternation in Italian are distributed according to the following schema:

	A	B	C
Causative	V	V	FA+V
Inchoative	V	SI + V	V

The type Ribassare belongs to column A, the type Arroventare to column B, the type Evaporare to column C, the types Avvizzire or Impigririre can behave according to any of the above models.

4. Aktionsart

As it is well known, the verbal action or aktionsart constitutes a semantic feature of the verb which is connected with aspect; however, while the aspect is strictly linked to the paradigm of tenses, the aktionsart is essentially connected with a lexical choice, and for this reason it is of considerable interest in a general analysis of the lexicon. In this study on the meaning of a particular subset of verbs, it was important to consider this characteristic and also useful in order to increase semantic information in the DMI. As regards the aktionsart we have considered the concepts of duration or «time» and goal or «telos» as elements pertinent to our analysis. With regard to time the verbs can be divided into 'durative' or 'non-durative'. The 'durative' verbs are opposed to the 'non-durative' since the latter are verbs in which the action can be considered as started and finished at the very same moment. In this respect, a verb like 'addormentarsi' = to go to sleep (non durative) is opposed to the verb 'dormire' = to sleep (durative). With regard to goal or «telos», the verbs can be seen as actions or processes directed towards a specific change, state, condition that can be achieved or not; in this respect the verbs can be divided into *Accomplishment and Achievement* (following Vendler's classic scheme '67) or in *Risultativi and Trasformativi* (using Italian terms, following Bertinetto's classification '86). Let us now consider the set of the verbs which we have chosen.

All these verbs are characterized by the movement towards a state or condition represented by the adjective or by the noun from which they are derived:

Arroventare	→ rovente	= red-hot
Ingrassare	→ grasso	= fat
Addolcire	↔ dolce	= sweet
Imbellire	→ bello	= beautiful
Depurare	→ puro	= clear
Rasserenare	→ sereno	= quiet
Allargare	→ largo	= wide

For this reason, by representing a process or a specific change, all these could be considered telic verbs and as a matter of fact for each of these verbs it is possible to

identify a context where the adverbial 'in x time', which is the syntactic test adopted by most of linguists to check this feature, is appropriate. What remains suspended is the actual completion of the process, which is nearly always likely to continue.

In this respect it must be noted that also the affixes give the sense of duration in time: while *in-* and *a-* imply not only the beginning but also an indefinite increase, this is not true for the suffixed verbs in *-ificare*, *-izzare* which give the idea of a process which takes place in various phases, with a beginning, a centre and a definite conclusion. It must also be noted that there is a difference between the transitive and the intransitive meaning of the verb with regard to the *aktionsart*. Let us consider for example the verb *Addolcire* in the two following sentences:

Lucia *addolcisce* il caffè' (sweetens the coffee)
 Lucia *addolcisce* con gli anni (softens with age)

We observe that the transitive use of the verb can be better placed among the *Accomplishment* or *Risultativi*, while the intransitive version is certainly to be placed among the *Achievement* or *Transformativi*.

This characteristic alternation appears with regularity in most of the verbs we are considering, so we decided to evidence this feature by means of codes.

5. Codification of the corpus

In order to assign to each meaning of the verb its relevant representation, so that the lexical entry of each verb appears in a more detailed and complete form, we started a first – completely automatic – processing of our set of data.

The first nucleus of data was obtained using a query program which allowed us to extract from the database three sets of lemmas, the definition of which contained the genus term 'rendere' (make), 'diventare' (become), 'divenire' (become). These were brought together and sorted both alphabetically by lemma (which evidenced the sequence of the formative prefixes) and by grammatical category (which grouped them according to their syntactical features). By using another program, the lemma number and the definition number were added to this data for each entry and the file containing this information was used to connect each verb to the base, noun or adjective, from which it derived. Finally we added to each entry the codes relative to the features analyzed. Table B is a sample from our set of entries as they appear after codification: the code *ERG* (which stands for ergative) in column 8 indicates that the verb under consideration shows the semantic/syntactic features typical of the ergative verbs, in particular the causative/inchoative alternation: codes *CAUS/INCO* in column 7. The transitive tag *VT* has been referred to the meaning 'rendere' (make) and the non-transitive tag (pronominal, intransitive, pronominal-intransitive: *VP, VI, VE*) to the meaning 'diventare' (become), and the relative auxiliary has been added (column 6: *A* = to have, *E* = to be).

As regards the *aktionsart* all the verbs of the set were given, by default, the code of 'telos' (goal in English), and the value *Accomplishment* (in table *TT*, column 9, which stands for the Italian term *Risultativo*) were assigned to the transitive use, while the value *Achievement* (in the table *TR* for the Italian term *Transformativo*) were assigned to the non-transitive use. A further codification was also given to the

affixes indicated in the last column, and next to them, in column 10, we indicated the value they have in the corpus.

We shall obviously have to intervene later, and this time manually, to correct those values which (assigned by default) are not always correct. For example, while for the prefixes *in-* the intensive value is perfectly regular and thus correct, when the deprivative code was assigned to the prefixes *de-* *dis-* *s-* the value is not always exact, because, as already said, in some cases they have an intensive or a neutral value.

Proposing this codification, compared to the previous organization of the entries, the following values are made explicit:

- 1) the two genus terms 'rendere' and 'diventare' have been linked to the transitive and non transitive use respectively;
- 2) the auxiliary selected by each of the two forms is indicated;
- 3) the ergative nature of the verbs and the causative/inchoative alternation are evidenced;
- 4) a value has been assigned to the aktionsart although in its most basic and regular meaning;
- 5) the affix and the semantic value: intensive, deprivative, repetitive, of which the affix is the carrier are made explicit.

References

- ANTELMI, D., ROVENTINI, A.: *Derivatives: a new relation in the Italian Lexical Database*, forthcoming.
- ATKINS, B., KEGL, J., LEVIN, B. *Explicit and implicit information in Dictionaries*, Report 5, Cognitive Science Laboratory, Princeton University, November 1986.
- ATKINS, B., KEGL, J., LEVIN, B. «Anatomy of a verb entry: from Linguistic Theory to Lexicographic Practice», *International Journal of Lexicography*, Vol. 1, No. 2, 1988, Oxford University Press.
- BERTINETTO, P. M., *Tempo aspetto e azione nel verbo italiano. Il sistema dell'indicativo*. Study di Lessicografia Italiana, Firenze: Accademia della Crusca, 1986.
- BOGURAEV, B., BRISCOE, T., *Computational Lexicography for Natural Language Processing*, 1989, Harlow, Longman Group Ltd.
- BURZIO, L., *Italian Syntax*, Reidel, Dordrecht, 1986.
- CALZOLARI, N., «Structure and access in an automated lexicon and related issues», in: Walker, D., Zampolli, A., Calzolari N. (eds.), *Automating the Lexicon: Research and Practice in a Multilingual Environment*, 1988. Cambridge University Press.
- CALZOLARI, N., PETERS, C., ROVENTINI, A., *Computational Model of the Dictionary Entry*, Esprit Basic Research Action No. 3030, Preliminary Report, Pisa, April 1990.
- LEVIN, B., «The Representation of Semantic Information in the Lexicon», in Walker, D., Zampolli, A., Calzolari, N. (eds.), *Automating the Lexicon: Research and Practice in a Multilingual Environment*, 1988. Cambridge University Press.
- VENDLER, Z., «Verbs and Times», in *Linguistics in Philosophy*, Cornell University Press, Ithaca, N.Y., 1967.

FILE: VERBI	SORTI	F	VM/SP CONVERSATIONAL MONITOR SYSTEM
0494090VTE	INCARBONIRE		DIV CARBONE
0494050VHP	INCARARE		DIV CARO
0494460VE	INCAROGNIRE		DIV CAROGNA
0894340VTP	SCAROGNIRE		DIV CAROGNA
0494520VE	INCARTAPECORIRE		DIV CARTAPECORA
04944950VE	INCATARRARE		DIV CATARPOSO
0495100VIE	INCATTIVIRE		DIV CATTIVO
0495100VTE	INCATTIVIRE		REN CATTIVO
0895550VTI	SCATTOLICIZZARE		REN CATTOLICO
0366470VI	ESCAVARE		REN CAVO
0495220VTI	INCAVARE		REN CAVO
0895670VI	SCAVARE		REN CAVO
0191710VT	CENTRALIZZARE		REN CENTRALE
0496180VI	INCERCONIRE		DIV CERCONE
0198200VT	CHIARARE		REN CHIARO
0198230VT	CHIAREGGIARE		REN CHIARO
0198350VT	CHIARIFICARE		REN CHIARO
0198440VIE	CHIARIRE		DIV CHIARO
0198440VTE	CHIARIRE		REN CHIARO
0282960VTR	DECLARARE		REN CHIARO
0299380VTR	DICHIARARE		REN CHIARO
0853730VTE	RISCHIARARE		REN CHIARO
0898210VTE	SCHIARARE		DIV CHIARO
0898240VTE	SCHIARIRE		DIV CHIARO
0898260VTE	SCHIARIRE		REN CHIARO
0007770VTI	ACCECARE		DIV CIECO
0009680VTI	ACCIECARE		DIV CIECO
0497600VTE	INCITRULLIRE		DIV CITRULLO
0497660VTI	INCIUCCHIRE		DIV CIUCCO
0497660VII	INCIUCCHIRE		REN CIUCCO
0497690VI	INCIVETTIRE		DIV CIVETTA
0497730VTP	INCIVILIRE		DIV CIVILE
0497730VTP	INCIVILIRE		REN CIVILE
0498160VI	INCODARDIRE		DIV CODARDO
0283340VT	DECOLONIZZARE		REN COLONIA
0498580VE	INCOLORARE		DIV COLORE
0225260VT	COMICIZZARE		REN COMICO
0500300VP	INCONTADINARE		DIV CONTADINO
0500980VTP	INCORDARE		DIV CORDA
0502060VTE	INCRETINIRE		REN CRETINO
0269020VT	CRISTALLARE		REN CRISTALLO
0270900VT	CROMATIZZARE		REN CROMATICO
0502490VTI	INCRUDELIARE		REN CRUDELE
0502490VTI	INCRUDELIARE		REN CRUDELE
0502540VTE	INCRUDIRE		DIV CRUDO
0502540VTE	INCRUDIRE		REN CRUDO
0843220VTE	RINCRUDIRE		DIV CRUDO
0502590VTP	INCRUSCARE		DIV CRUSCA
0498290VTE	INCUDIARE		DIV CUDOIO
0502340VTE	INCUDIARE		DIV CUDOIO
0502790VTE	INCUDIARE		DIV CUGLIO
0502820VTE	INCUPIRE		DIV CUPPO
0502820VTE	INCUPIRE		REN CUPPO
0502920VTP	INCURIOSIRE		DIV CURIOSO
0276480VTI	CURVARE		DIV CURVO

TABLE A

