

Antonietta Alonge
Istituto di Linguistica Computazionale, C.N.R., Pisa

**Motion Verbs:
Data on Meaning Components in Dictionaries and
Identification of Syntactic Properties ***

Abstract

Relying on the assumption that the syntactic properties of verbs are, to a large extent, determined by their meanings, we carried out an analysis of dictionary definitions of motion verb taxonomies, with the goal of acquiring information on verb meaning components. In keeping with Talmy (1985), we identified the typical “conflations” of meaning components found for motion verbs in Italian. The information extracted was then used to derive data, missing in our dictionaries, on syntactic properties of the verbs. The particular organisation of data within dictionaries allowed us to extend the observations made in connection with single verbs to large quantities of them.

1. Introduction

Recent research in lexical semantics showed that the syntactic behaviour of verbs is, to a large extent, determined by their meanings. Criticism of this hypothesis was rejected by emphasizing the importance of identifying the relevant meaning components of verbs when looking for the connection between semantics and syntax. Levin (1993), for instance, discusses Rosen’s (1984) position according to which meaning alone is not predictive of membership in the classes of unaccusative or unergative verbs, since there are verbs like, e. g., Italian bodily process verbs which do not show uniform syntactic behaviour: in fact, *russare* (to snore) is unergative while *arrossire* (to blush) is unaccusative. Levin points out that the notion “bodily process” is probably not relevant to the characterisation of a verb as unaccusative or unergative and claims that the two verbs can be semantically classified in different ways if one takes into consideration other, more basic, semantic features: while *russare* refers to an activity, *arrossire* refers to a change of state. As a matter of fact, relevant research (cf. Tenny (1988); Van Valin (1990); Levin & Rappaport (1991); etc.) already showed that these are the semantic notions involved in the determination of a verb status with respect to unaccusativity or unergativity.

Since lexicographers describe a word meaning by emphasizing the most salient characteristics of that word, dictionary definitions turn out to provide information exactly on the meaning components relevant to the determination of verb syntactic properties. If we take into consideration the

definitions given for the verbs *russare* and *arrossire* in *Il nuovo dizionario italiano Garzanti* (henceforth GRZ), we actually see that the former is described as referring to the process of “respirare rumorosamente nel sonno” (breathing noisily while sleeping), whereas the latter is defined as “divenir rosso, spec. in volto” (becoming red, especially in the face): that is, *russare* is connected with a genus term indicating a homogeneous activity (while the differentia of its definition refers to the MANNER in which the process is performed), whereas *arrossire* is connected with a change of state.

In the context of the Esprit project *Acquilex-II*, we are semi-automatically analysing machine-readable versions of Italian dictionaries¹, aiming at extracting information to be used in the development of a multilingual lexical knowledge base (LKB) for natural language processing, but also, more in general, in the design of high-quality multilingual lexical resources for both human and machine use. Indeed, acquiring and representing lexical knowledge by re-using existing lexical resources (in particular, dictionaries) have recently become fundamental concerns for computational linguists and lexicographers. Moreover, our research may help overcome the well-known limits which dictionaries present, notwithstanding their usefulness.

Thus, we are trying to re-use dictionaries for encoding data on the semantic and syntactic characteristics of large quantities of words. In the first stage of our work (cf. Alonge (1992a) and (1992b)) we built methodologies to extract data on the semantics of verb classes from definitions. Our research was then further developed with the main goal of obtaining information on syntactic properties of verbs, missing in our dictionaries, by carefully analysing the semantic data already extracted. The main goal of this paper, therefore, is to show how the semantic analyses carried out on definitions of Italian motion verbs were used for deriving information on various syntactic properties of large quantities of verbs.

2. Semantic data within definitions

Within *Acquilex* we analysed a taxonomy of motion verbs aiming at identifying, in dictionary definitions, the meaning components “conflated” in verb roots, in keeping with Talmy’s (1985) study on these verbs. Talmy assumed that the process of lexicalisation is involved when a meaning component (or a set of meaning components) is found to be regularly associated with a morpheme, and provided a cross-linguistic study of lexicalisation patterns connected with the expression of motion. Moreover, Talmy dealt with the systematic relations between meaning and surface expressions, with a particular emphasis on the relation between meaning components conflated in verb roots and the syntactic configurations allowed by the verbs themselves. In our previous research we verified the possibility of identifying information on verb meaning components within dictionary definitions, by taking into consideration a taxonomy of motion verbs. In

Acquilex-II the issue of relating word meanings to varying lexicalisation patterns across the languages being investigated has become one of the major concerns because of the consequences that a better understanding of the principles of lexicalisation may have both from the theoretical point of view and for the further development of the Acquilex LKB. Our research was therefore further carried out, both on the same taxonomy and on other motion verb taxonomies, with the goals of extracting and formalizing the semantic information provided by definitions, and trying then to infer, from these data, information on the syntactic behaviour of the verbs under analysis.

By combining the semantic information contributed by the genus term with the information within the differentia, it was possible to obtain a description of the fundamental semantic characteristics of motion verbs. In particular, by the analysis of data contained in the differentia of motion verb definitions in GRZ and DMI, we identified recurrent patterns, referring to a restricted number of components of meaning. Thus, for instance, while the contribution provided by the genus term *muoversi* to the meaning of its hyponyms is the reference to MOTION on the part of the protagonist of the event, within the differentia of definitions we identified patterns connected with the following fundamental² meaning components (see Alonge (1992a; 1992b)):

- MANNER (of MOTION)
- DIRECTION:
 - GOAL
 - SOURCE
- PATH
- MEDIUM
- PURPOSE

The highly structured nature of the information coded within dictionaries allowed us to extend the results of the analysis of single words to whole classes of them. *Accasciarsi* (to collapse), for instance, is defined (GRZ, 1) as “cadere pesantemente al suolo” (to fall down on the ground heavily): the genus term *cadere* refers to change of position since there is the indication of MOTION to a GOAL in its meaning. In addition, *cadere* is a hyponym of *andare* (to go), which, in turn, is a hyponym of *muoversi* that indicates that the protagonist of the event referred to is also the “moving object”. Thus, *accasciarsi* will, first of all, inherit these semantic features from its superordinate verbs. Within the differentia of its definition we then find words/patterns which can be connected with two meaning components: *pesantemente* refers to the MANNER of motion, while *al suolo* again indicates a GOAL (which we may assume to be the “typical goal” of the event indicated by the verb). Therefore, we can state that *accasciarsi* refers

to change of position/manner of motion on the part of the protagonist of the event and has a “typical goal”.³

Thus, much information on typical conflations of meaning components found in Italian is available in dictionaries and can be extracted by means of semi-automatic procedures. Even the lack of specific data may be used advantageously (especially in a contrastive way): by looking at the set of verbs expressing motion in English, one can see that there are some “verbs of motion using a vehicle”, that are zero-related to nouns that are vehicle names (balloon, bicycle, bike, boat, bobsled, bus, cab, canoe...), therefore, within the meaning of these verbs there is a reference to the vehicle used to move. If we look for vehicle names within verb definitions in dictionaries of Italian, we very seldom find them, since we only have a small number of verbs which are derived from vehicle names, like, e.g., *pattinare* (to skate), *sciare* (to ski), *paracadutare/si* (to parachute).

3. From semantics to syntax

The data on semantic components obtained by the analysis of definitions were connected with various syntactic properties of verbs which are not described within our dictionaries.⁴ In addition, once stated the connection among specific meaning components and a syntactic property, such a connection could be generalized, using taxonomies, for a semantic class of verbs. The two superordinates *muoversi* and *muovere*, for instance, are respectively the inchoative and causative verb forms that express simply “the fact of motion”;⁵ if we take into consideration the hyponyms of both verbs, we can see that when a verb displays two senses, one defined by means of *muoversi* and the other by means of *muovere*, such senses will turn out to be connected with the inchoative and causative forms of the verb itself. A similar situation holds for (pairs of) senses of verbs defined by means of other superordinates (like, e. g., *andare* vs *far andare/mandare*), in such a way that, once stated, the relation between superordinates can be generalized for hyponym senses.

The analysis carried out allowed us to classify verbs according to their *Aktionsart* (or “lexical aspect”; cf. Alonge (1992b)). Since the *Aktionsart*-class of a verb determines the possibility of it occurring in certain tempo-aspectual forms, or as complement of other verbs (“aspectual verbs” and others), or with different temporal adverbs, etc., we derived such information for whole verb taxonomies.

Then, meaning components were also used to obtain data on possible arguments of verbs. For instance, we got evidence that Italian motion verbs may occur with goal or source phrases only in case there is a reference to a GOAL or SOURCE in their meanings:

1. Sono arrivata a casa alle cinque.
(I arrived home at five)

2. * Ho nuotato a quell'isola.
 (I swam to that island)

On the other hand, when the GOAL or SOURCE implied by the meaning of a verb is a *specific* GOAL or SOURCE (like, e.g., in the case of *rincasare*, which means “to go/come back home”), the verb cannot occur with a goal or source phrase, unless such a PP is in some way modified by an adjective or other modifier:

3. Sono rincasata dopo due giorni.
 (I went back home after two days)
4. *Sono rincasata a casa dopo due giorni.
 (I went back home (home) after two days)
5. Sono rincasata nella mia casa di campagna dopo due giorni
 (I went back (home) to my country-house after two days)

A similar behaviour was already noted in connection with some classes of verbs that are derived from nouns (cf. Jackendoff (1990); Levin (1993)), and it seems typical of all the verbs of this kind. In fact, verbs of motion using a vehicle, whose meaning implies the particular vehicle used, cannot occur with PPs indicating the vehicle unless such a PP contributes additional information about the vehicle itself:

6. Si è paracadutato subito.
 (He parachuted immediately)
7. * Si è paracadutato con il paracadute.
 (He parachuted with the parachute)
8. Si è paracadutato con il suo nuovo paracadute.
 (He parachuted with his new parachute)

An interesting case is that of the verb *guidare* (to drive): in GRZ it is defined as “manovrare un automezzo” (to operate a motor-vehicle), but it is also stated that when the specific motor-vehicle is not indicated, it is understood as being a car:

9. Guidò l'autobus per tre ore.
 (He drove the bus for three hours)
10. Guidò bene= Guidò la macchina bene.
 (He drove (the car) well)

Finally, the fact that vehicle names are very seldom found within definitions was used to state that complex phrases, usually containing *andare* (to go) plus PPs indicating the vehicles used, express in Italian what in English is expressed by means of a single verb.

Our research on meaning components in definitions and syntactic properties of verbs allowed us to reach interesting results also from a theoretical point of view. Sanfilippo (1992) reports on the work done in order to acquire and represent information on diathesis alternations involving English verbs. He distinguishes among alternations that generally hold also for Italian verbs; however, at least in one case, Italian motion verbs display a different behaviour and data from definitions were used to identify this difference. According to Sanfilippo some intransitive verbs taking a goal complement may display the “unbounded-path” (u-path) alternation, i.e. they can occur with or without a goal/source argument and when the goal/source is not expressed the path along which the movement occurs is not bounded:

11. Kim walked to school.
12. Kim walked.

This alternation is opposed to the “bounded-path” (b-path) alternation which concerns those intransitive verbs that always refer to motion along a bounded path, even if they are not followed by any goal/source complements:

13. Kim came away.
14. Kim came.

While the b-path alternation is displayed by all Italian change-of-position verbs, that is those verbs in whose definitions we find a GOAL or SOURCE meaning component, it is more difficult to identify a group of verbs in Italian displaying the u-path alternation. In fact, verbs referring to motion along an unbounded path, i.e. manner-of-motion verbs like *camminare* (to walk), *salizzare* (to skip), *zoppicare* (to limp), do not generally undergo this alternation, because they cannot occur with goal/source arguments:

● b-path alt.

15. Gianni è caduto a terra.
(Gianni fell down on the ground)
16. Gianni è caduto.
(Gianni fell down)

● u-path alt.

17. * Gianni ha camminato a casa.
(Gianni walked home)
18. Gianni ha camminato.
(Gianni walked)

There is only a small subset of manner-of-motion verbs which shows a different behaviour in Italian. These are verbs which have both an unergative and an unaccusative use:

19. Gianni ha corso per ore/* a casa.
(Gianni (has run) ran for hours/home) (unergative)
20. Gianni è corso a casa.
(Gianni (is run) ran home) (unaccusative)

Since the meaning component DIRECTION characterizes verbs which allow goal/source arguments, we should hypothesize that the unaccusative variant of the verbs considered implies a DIRECTION (GOAL/SOURCE) meaning component besides the MANNER one. Dictionaries sometimes refer to the presence of such semantic indication to explain the unaccusative uses of these verbs: in DMI, for example, it is pointed out that when *correre* is used with the auxiliary *essere* (and is, therefore, unaccusative) it implies a GOAL. However, we do not find information of this kind for all the verbs behaving like *correre*. In any case, this component of meaning has already been described as being connected with unaccusativity in theoretical works (cf., e. g., Levin & Rappaport (1991)), since it would correspond, on the spatial dimension, to telicity (that is, to the existence of a temporal goal).

Some more reflections can be added regarding this and related subjects. Levin & Rappaport (1991), aiming at explaining the fact that intransitive motion verbs, although they apparently belong to the same semantic class, display a different syntactic behaviour (some are unergative and some unaccusative), identify three semantic sub-classes within the larger class. In the first group there are verbs indicating change of position (that is, connected with a meaning component DIRECTION) and behaving as unaccusatives; then there are two groups of manner of motion verbs with a different behaviour with respect to unaccusativity. The authors do not seem to reach a satisfying solution to the problem of finding a semantic "cause" for this different behaviour: they hypothesize that the unergative manner of motion verbs presuppose protagonist control, while the unaccusatives lack such a control, but involve a DIRECT EXTERNAL CAUSE. However, there are verbs which do not fit this description: *sgattaiolare* (to go far away stealthily), *fuggire* (to run away), *scappare* (to escape), *inginocchiarsi* (to kneel down), etc., are all unaccusative manner of motion verbs which instead of requiring a DIRECT EXTERNAL CAUSE, imply protagonist control. On the other hand, manner of motion verbs which present both unergative and unaccusative uses do not require a DIRECT EXTERNAL CAUSE in their unaccusative uses. *Volare* (to fly), e.g., is a manner of motion verb which can be either unergative or unaccusative:

21. Il gabbiano ha volato per ore.
(The seagull (has flown) flew for hours) (unergative)
22. Il gabbiano è volato via poco dopo.
(The seagull (is flown) flew away a short time later)
(unaccusative)

It seems to us that in both cases we should speak of the existence of control on the part of the protagonist of the event and something similar holds for other examples of this kind which we find in Italian. In our opinion, what differentiates the two instances of the verb *volare*, and what, in general, differentiates the two groups of manner of motion verbs is the component DIRECTION. Actually, even if as far as the small group of unaccusative/unergative verbs is concerned we do not always find useful information on components of meaning within definitions, relevant data are found in connection with the other unaccusative manner of motion verbs, which are always connected also with GOAL or SOURCE components, so that this seems to be the relevant combination of meaning components (cf. Alonge (in preparation) for a more detailed discussion of this issue). Thus, motion verbs can be split into three sub-classes, as Levin and Rappaport (1991) claim, but while one is of change of position verbs, of the remaining two, one is of manner of motion verbs and the other of manner of motion/change of position ones.

This solution seems to provide a better explanation as far as Italian is concerned, but it is necessary to reconsider English motion verbs. Italian manner of motion verbs cannot occur with goal/source phrases, if they are not unaccusative and do not refer to a GOAL/SOURCE component. English manner of motion verbs may always occur with PPs indicating goal/source, but, when the goal/source phrase is present, they behave as unaccusatives. We agree with Levin & Rappaport (1991) who claim that the behaviour of these verbs can be explained "with the process of lexical subordination which allows verbs to acquire certain types of extended meanings". Thus, such manner of motion verbs will present two meanings, one of which implying a DIRECTION component.

4. Conclusion

The goal of this paper was to show how the semantic information extracted from dictionaries may be used to infer information on syntactic properties of verbs. To be able to draw such a connection is first of all important because it allows one to overcome, at least in part, the limits which our dictionaries present and that have often been emphasized. Furthermore, it may have considerable consequences because it both provides evidence for theories which describe syntactic properties of verbs as determined by semantics and offers the researcher fundamental information on large quantities of data.

Finally, it might allow encoding of only that information which is really necessary within the LKB which is being developed, with obvious advantages for the whole system.

Notes

- * This research is part of the work being carried out within the Esprit project *Acquilex-II* (BRA 7315), a continuation and extension of *Acquilex* (*Acquisition of Lexical Knowledge for Natural Language Processing Systems*). Research teams from different European countries are collaborating within the project, together with dictionary publishing partners.
- 1 The computerized version of the already mentioned GRZ and the DMI (*Dizionario-macchina dell'Italiano*) which is a machine-readable dictionary developed at ILC, in Pisa, using mainly the Zingarelli *Vocabolario della Lingua Italiana*. While within Acquilex we used only dictionaries as sources of data, within *Acquilex-II* also textual corpora are being exploited. In any case, the present paper will only deal with the work being carried out on dictionaries.
- 2 Actually, other meaning components were also individuated, but they are infrequent and not typically found in motion verb definitions.
- 3 The conflation of meaning components found (MANNER + GOAL) is not the most typical for a Romance language, according to Talmy (1985). Indeed, there are a number of verbs, within the taxonomies we have analysed, that present this kind of conflation, but they are generally verbs at the lowest level within taxonomies; that is, they have very specific meanings and are not the most commonly used words.
- 4 The syntactic information available in our dictionaries of Italian only concerns transitivity, intransitivity and reflexivity (and with many limits, since, for example, obligatoriness or optionality of direct objects are not indicated; different kinds of reflexivity are not clearly distinguished; etc.).
- 5 As a matter of fact, *muovere* can also be intransitive-inchoative, but this use of it is very rare and never found when the verb is utilized as a superordinate.

References

- Alonge, A. 1992a. "Analysing Dictionary Definitions of Motion Verbs", in *Proceedings of the 14th International Conference on Computational Linguistics*, Nantes, France.
- Alonge, A. 1992b. "Machine-Readable Dictionaries and Lexical Information on Verbs", in *Proceedings of the 5th Euralex International Congress*, Tampere, Finland.
- Alonge, A. in prep. "On the Syntax and Semantics of Motion Verbs: An Analysis of Data Extracted from Dictionaries and a Textual Corpus".
- Jackendoff, R. 1990. *Semantic Structures*, Cambridge, Mass.: The MIT Press.
- Levin, B. 1991. *The Lexical Organization of English Verbs*, Evanston, Ill.: Northwestern University.
- Levin, B. & Rappaport M. 1991. "The Lexical Semantics of Verbs of Motion: the Perspective from Unaccusativity", ms.
- Rosen, C. 1984. "The Interface between Semantic Roles and Initial Grammatical Relations", in Perlmutter, D. M. & Rosen, C. (eds.), *Studies in Relational Grammar 2*, Chicago, Ill.: University of Chicago Press.
- Sanfilippo, A. 1992. "Verbal Diatheses: Knowledge Acquisition, Lexicon Construction and Dictionary Compilation", Acquilex WP.
- Sanfilippo, A., T. Briscoe, A. Copestate, M. A. Martí, M. Taulé, A. Alonge 1992. "Translation Equivalence and Lexicalization in the Acquilex LKB", in *Proceedings of the 4th International Conference on Theoretical and Methodological Issues in MT*, Montreal, Canada.
- Talmy, L. 1985. "Lexicalization Patterns: Semantic Structure in Lexical Form", in Shopen, T. (ed.), *Language Typology and Syntactic Description: Grammatical Categories and the Lexicon*, Cambridge: CUP.
- Tenny, C. 1988. "The Aspectual Interface Hypothesis: The Connection between Syntax and Lexical Semantics", in Tenny, C. (ed.), *Studies in Generative Approaches to Aspect*, Lexicon Project Working Papers, 24, Cambridge, Mass.: The MIT Press.
- van Valin, R. D. 1990. "Semantic Parameters of Split Intransitivity", *Language*, 66, 2.
- Vendler, Z. 1967. *Linguistics in Philosophy*, Ithaca: Cornell University Press.