From lexicological to lexicographical issues: Italian verbs with predicative complement

Francesca Strik Lievers
Istituto di Linguistica Computazionale, CNR

Intransitive (e.g., become) and transitive (e.g., consider) verbs obligatorily requiring a predicative complement are an interesting, and at the same time problematic issue both at a theoretical and at a lexicographical level. In this paper we focus on Italian verbs, and on the way two computational semantic lexica deal with them. Both in ItalWorNet and in SIMPLE the treatment of these verbs shows to be problematic, since the information appears to refer to the 'verb + predicative complement' complex rather than to the verb itself. Recognizing that verb and predicative complement contribute to the construction of a unitary event, we believe that it is nevertheless possible, and useful, to isolate the role of the two components. The description proposed here is based on the Generative Lexicon model (Pustejovsky 1995), and it is in line with the recent project of a lexical resource for (sub)event structure (Im and Pustejovsky 2009). Verb and predicative complement codify each a different part of the subevent structure. To give an example, 'diventare ('become') + predicative complement' is a transition, where diventare codifies the process subevent, and the predicative complement codifies the (result) state subevent. This kind of analysis can possibly be integrated into the SIMPLE lexicon, which is already built following the Generative Lexicon model.

1. Introduction

From a theoretical perspective, verbs with predicative complement (henceforth pred. compl.) are an interesting issue. In particular, intransitive verbs obligatorily requiring a (subject) complement (e.g., *seem*, *become*) have been largely studied in the literature, receiving many different descriptions and interpretations, as they show peculiarities both of 'ordinary' predicative verbs (e.g., *run*, *build*) and of the copula *be*. Also some transitive verbs (e.g., *consider*) differ from 'ordinary' predicative verbs, requiring the presence of an (object) complement to code a single event. The problematic issues emerging from theoretical reflections about both intransitive and transitive verbs with pred. compl. are reflected in their lexicographic treatment in computational lexica. This difficulty has already been recognized for the description of English verbs in lexica such as FrameNet (cf. Ruppenhofer et al. 2006: 56) and VerbNet (cf. Kipper Schuler 2005, Korhonen and Briscoe 2004). As Korhonen and Briscoe (2004: 36) point out, 'Verbs taking ADJP, ADVP, ADL, particle, predicative, control and sentential complements are still largely excluded [from VerbNet] […] As many of these verbs are highly frequent in language, NLP applications utilizing lexical-semantic classes would benefit greatly from a linguistic resource which provides adequate classification of their senses'.

In this paper we focus on Italian verbs, and on the way computational semantic lexica deal with them. We examine here two resources available for Italian: ItalWordNet (cf. Roventini et al. 2003), and SIMPLE (cf. Lenci et al. 2000). Since the treatment of verbs with pred. compl. shows to be highly problematic, the aim of the paper is to suggest possible directions for improving the description of the verbs under analysis.

2. ItalWordNet

ItalWordNet is structured in the same way as English WordNet (cf. Fellbaum 1993), i.e. the lexicon is organized in sets of synonyms (synsets). Each synset is accompanied by a gloss defining the sense of the word. Synsets are further linked with other synsets by semantic relations, mainly vertical relations of hyponymy and hyperonymy.

Verbs with pred. compl. are typically 'light' verbs, since most lexical information is contributed by the pred. compl.. It can thus be expected that a sense-enumerative lexicon (a lexicon consisting in a simple list of senses, cf. Pustejovsky 1995), as ItalWordnet, may have difficulties in giving a satisfactory analysis of these verbs. Examining some entries for verbs with pred. compl., it can moreover easily be observed that many difficulties are found in individuating the semantic relations with other synsets. To give an example, *diventare* ('become') has a synset that includes more than 800 hyponyms. Table 1 includes just a few examples:

DIVENTARE	
Gloss	cambiare stato, caratteristiche e sim., passare, spec. progressivamente, da una condizione a un'altra
Synset members	*divenire* [1], *diventare* [1], *farsi* [1], *trasformarsi* [1]
Examples	*'Diventare buono.' 'Diventare vecchio.' 'Il vino è diventato aceto.'*
Hyperonyms	accadere [1] avere luogo [1] avvenire [1] occorrere [2] succedere [1]
Hyponyms	*accendersi [1] accendersi [4] accentuarsi [1] acchetarsi [1] accigliarsi [2] acciocchire [1] accomodarsi [3] accontentarsi [1] accorciarsi [1] accorgersi [3] accostarsi [2] accrescere [2] accumularsi [1] acquiescere [1] acquietarsi [1] acquistare terreno [1] adattarsi [2] addensarsi [1] addolcire [2] addolcirsi [1] addolorarsi [1] addormentarsi [1] addormentarsi [2] addossarsi [1] adeguarsi [1] adempiersi [1] adempirsi [1] ...*

Table 1. *Diventare* in ItalWordNet

Why does ItalWordNet offer such a long list of hyponyms? Probably, the list has been compiled with regard to the '*diventare* + pred. compl.' complex (e.g., *abbronzarsi* roughly means 'diventare abbronzato, 'to become tanned'). Therefore, no significant information is provided about the semantics of *diventare* itself. Also many other verbs with pred. compl., whose entries can not be reported here, meet with the same problem.

The origin of these difficulties can be found in two connected matters: 1) the semantic 'lightness' of the verbs under analysis, and 2) the fact that it is not easy to isolate the contribution of the verb from that of the pred. compl. in the coding of the event.

The Generative Lexicon (Pustejovsky 1995) theory can provide useful means to describe these verbs, since it allows to give underspecified semantic representations of words. The meaning of words is then further specified and modulated in composition: in our case, mainly in composition with the pred. compl[1]. In the following paragraph we will therefore see how the verbs under analysis are treated in the SIMPLE lexicon, that is based on an extension of the Generative Lexicon model.

3. SIMPLE[2]

In the SIMPLE lexicon each lexical entry consists of one or more corresponding Semantic Units (USem in the Italian lexicon), describing the sense(s) of the lemma. For each USem, it is specified the relevant Template (i.e. the ontological information), and for event-denoting USems are also marked the Event Type (approximately corresponding to *Aktionsart*), the argument structure, and the selectional restrictions on the arguments. USems are also linked

[1] See Schmitt (2005), demonstrating the usefulness of the Generative Lexicon for describing the copulas of Portuguese (*ser, estar, ficar*).

[2] *Semantic Information for Multipurpose PLurilingual Lexica*. SIMPLE is available for 12 languages, among which Italian (developed by Istituto di Linguistica Computazionale, CNR, Pisa).

to other USems by semantic relations. The following tables show the most relevant information given in the entries for some representative verbs with pred. compl.:

DIVENTARE (USem61409) [3]	
Example	diventare grande / di ghiaccio
Gloss	trasformarsi in qlco di diverso, passare da una condizione ad un'altra
Template	Change
Event Type	Transition
Relations	Isa: evento; Synonym: divenire
Predicates	ARG0: Role_ProtoPatient; Entity
	ARG1: Role_Underspecified; Entity

Table 2. *Diventare* in SIMPLE

SEMBRARE (USem79107)	
Example	la questione sembra complicata; - essere risolta
Gloss	apparire; mostrarsi in una certa maniera
Template	State
Event Type	State
Relations	Synonym: parere, apparire; Isa: essere
Predicates	ARG0: Role_ProtoPatient; Entity
	ARG1: Role_Underspecified; Entity

Table 3. *Sembrare* in SIMPLE

RITENERE (USem60743)	
Example	Ritenere qlcu. intelligente, un amico/ ritenere un libro un capolavoro
Gloss	credere, giudicare
Template	Judgement
Event Type	State
Relations	Isa: giudicare; Synonym: reputare, considerare, trovare, stimare, pensare, giudicare
Predicates	ARG0: Role_ProtoAgent; Human
	ARG1: Role_ProtoPatient; Entity
	ARG2: Role_Underspecified; Entity

Table 4. *Ritenere* in SIMPLE

RENDERE	
(USem7657)	
Example	Rendere ridicolo qlcu. Rendersi ridicolo
Gloss	far diventare
Template	Cause_Change
Event Type	Transition
Relations	Isa: causare
Predicates	ARG0: Role_ProtoAgent; Entity
	ARG1: Role_ProtoPatient; Entity
	ARG2: Role_Underspecified; Entity
(USem76544)	
Example	Rendere la vita un inferno/ difficile a qlcu.
Gloss	far diventare
Template	Cause_Change (Temp97)
Event Type	Transition
Relations	Isa: cambiare
Predicates	ARG0: Role_ProtoAgent; Entity
	ARG1: Role_ProtoPatient; Entity
	ARG2: Role_Beneficiary; Human
	ARG3: Role_Underspecified; Entity

Table 5. *Rendere* in SIMPLE

[3] In the tables it is specified which is the Usem described, since some verbs are associated to more than one Usem (see *rendere*).

This kind of description seems to be more appropriate for verbs with pred. compl., as compared to that provided in ItalWordNet. First of all, in the 'Predicate' slot it is highlighted that the presence of the pred. compl. is semantically (as well as syntactically) required[4]. Secondly, the extreme underspecification of the description also seems to be appropriate (the role of the pred. compl. is 'Underspecified', and the selectional restriction is Entity, i.e., the highest level of the ontology). On the one hand, this properly reflects the semantic 'lightness' of the verbs, and is coherent with the vast range of semantically different elements that can fill the pred. compl. position. On the other hand, the risk is that of having entries that are scarcely informative.

However, uninformativity is avoided thanks to the information about the Event Type. The contribution given by *diventare*, *rendere* and similar verbs is mainly aspectual in nature. Therefore, the most relevant information is that concerning the Event. In SIMPLE, three Event Types are distinguished: state, process, and transition[5]. Both *diventare* and *rendere* are classified as transitions. However, do these verbs really denote a transition, or do these verbs *together with the pred. compl.* denote a transition?

We believe that the second possibility is the right one: the event of transition is coded by the verb and by the pred. compl. together. SIMPLE apparently attributes to the verb only the Event Type that should be attributed to the 'verb + pred. compl.' complex. In what follows, some preliminary ideas are suggested to improve the description of the event.

4. A tentative proposal for representing verbs with predicative complement

Following the description of the class of (transitive and intransitive) verbs with pred. compl. proposed in Strik Lievers (2009), two main semantic subclasses are identified: a) Verbs building – together with the pred. compl. - simple events (e.g., *sembrare* 'seem', *ritenere* 'consider'), and b) Verbs building complex events, i.e., events with subevent structure (e.g., *diventare* 'become', *rendere* 'make'). The idea is that verb and pred. compl. lexicalize each a different part of the subevent structure. The following scheme shows, as an example, the representation proposed for '*diventare* + pred. compl.', where the verb lexicalizes the first subevent (e_1, a process), that precedes ($<_\alpha$) the second subevent (e_2, a – result - state), lexicalized by the pred. compl.. The process and the state subevent together form a transition event:

diventare P

EVENTSTR = [$E_1 = e_1$:process; $E_2 = e_2$:state; RESTR= $<_\alpha$]

ARGSTR = [ARG1= x]

QUALIA = [FORMAL= *P_state* (e_2, x)]

Figure 1. Representation proposed for *diventare*.

[4] Arguments (ARG) are here intended as semantic arguments, that may be syntactically realized in a different way, or even not realized at all (cf. Lenci et al. 2000: 48). It has to be observed that the label 'ARG' is not really appropriate for predicative complements, or at least it is misleading, and it should perhaps be replaced by a more specific label.

[5] States consist of a unique event, processes of a temporally ordered sequence of identical events, and transitions of two different subevents (cf. Pustejovsky 1995).

Much more information could be added, such as that concerning event headedness, and that concerning opposition structure[6]. However, as a first step, at least subevent structure should be made explicit. This kind of description could be well integrated into the SIMPLE lexicon, since it is already built following the Generative Lexicon model. Of course, subevent structure should be marked for all event-denoting entries of the lexicon. This is in line with Im and Pustejovsky (2009), where an ongoing project has been presented, SUBEVITA[7], which follows exactly that direction. In Im and Pustejovsky (2009), marking the subevent structure is presented as a means for recognizing textual inferences. For example, the semantic relation between *kill* and *be dead* in a text is identifyable because it is already included in the event structure of *kill*, formed by a pre-state (*be_alive*), a process (*dying*), and a result state (*be_dead*). Verbs with pred. compl. provide further evidence to support the necessity of making the subevent structure explicit. As we saw, this appears to be in fact the only way to isolate and describe the meaning of these verbs.

5. Concluding remarks

In this paper, we have examined the lexicographic treatment of Italian verbs with pred. compl. in two computational lexica. It emerged that both ItalWordNet and SIMPLE provide descriptions that are not entirely adequate, since they apparently refer to the meaning of the 'verb + pred. compl.' complex, rather than to the meaning of the verb itself. The proposal made here is to represent in explicit form the subevent structure. This can be useful for all event-denoting lexical entries, and in particular for the verbs analysed here, where two subevents of one (complex) event are lexicalized by two different lexemes (the verb and the pred. compl.). This is, in our opinion, a theoretically sound way to isolate the contribution of the verb to the coding of the event, and to represent it. The next step will be to try to integrate this kind of information into the SIMPLE lexicon.

[6] The importance of giving information about opposition structure (Pustejovsky 2000) is particularly evident for verbs like *diventare*, for illustrating the fact that after the event has occurred, the subject is in a state P (codified by the predicative complement), in which he was not before the event.

[7] SUBEVents In Text Analyzer.

References

Fellbaum, C. (1993). 'A Semantic Network of English Verbs'. In Miller, G. A., Fellbaum, C., Miller, K. J. (eds.). *Five papers on WordNet*. 40-61.

Im, S., Pustejovsky, J. (2009). 'Annotating Event Implicatures for Textual Inference Tasks'. In *GL2009 Proceedings*. 113-119.

Kipper Schuler, K. (2005). *Verbnet: A broad-coverage, comprehensive verb lexicon*. PhD Dissertation (http://repository.upenn.edu/dissertations/AAI3179808).

Korhonen, A.; Briscoe, T. (2004). 'Extended Lexical-Semantic Classification of English Verbs'. In *Proceedings of the HLT/NAACL Workshop on Computational Lexical Semantics*. Boston, MA. 38-45.

Lenci, A. et al. (2000). *SIMPLE. Linguistic Specifications, Deliverable D2.1*

Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.

Pustejovky, J. (2000). 'Events and the Semantics of Opposition'. In Tenny, C., Pustejovsky, J. (eds.). *Events as Grammatical Objects*. Stanford: CSLI Publications. 445–482.

Roventini, A., et al. (2003). 'ItalWordNet: Building a large semantic database for automatic treatment of Italian'. In Zampolli, A., et al. (eds.). *Computational linguistics in Pisa : special issue II. Pisa-Roma: Istituti editoriali e poligrafici internazionali*. 745-791.

Ruppenhofer, J. et al. (eds.) (2006). *FrameNet II: Extended Theory and Practice* (http://framenet.icsi.berkeley.edu/book/book.pdf).

Schmitt, C. (2005). 'Semi-Copulas: Event and Aspectual Composition'. In Kempchinsky, P., Slabakova, R. (eds.). *Aspectual Inquiries*. Dordrecht: Springer. 121-45.

Strik Lievers, F. (2009). *I verbi con complemento predicativo. Uno studio semantico-lessicale sull'italiano*. Unpublished PhD Dissertation, Università di Pisa.

SIMPLE [on line]. http://www.clarin-it.it/Simple/ [Acces date: 23 Feb. 2010].

ITALWORDNET [on line]. http://www.ilc.cnr.it/iwndb_php/ [Acces date: 23 Feb. 2010].