

Semantic role annotation of instrument subjects

Rossella Varvara

Università di Trento

rossella.varvara@unitn.it

Elisabetta Ježek

Università di Pavia

jezek@unipv.it

Abstract

English. Semantic role annotation has become widely used in NLP and lexical resource implementation. Even if attempts of standardization are being developed, discordance points are still present. In this paper we consider a problematic semantic role, the Instrument role, which presents differences in definition and causes problems of attribution. Particularly, it is not clear whether to assign this role to inanimate entities occurring as subjects or not. This problem is especially relevant 1- because of its treatment in practical annotation and semantic role labeling, 2- because it affects the whole definition of semantic roles. We propose arguments to sustain that inanimate nouns denoting instruments in subject positions are not instantiations of Instrument role, but are Cause, Agent or Theme. Ambiguities in the annotation of these cases are due to confusion between semantic roles and ontological types.

Italiano. *L'annotazione dei ruoli semantici è ormai molto utilizzata nell'ambito del NLP e della creazione di risorse lessicali. Sebbene si stia cercando uno standard condiviso, vi sono ancora punti di disaccordo. Nel presente articolo si considera un problematico ruolo semantico, il ruolo di Strumento, il quale causa ambiguità nell'annotazione e nella sua definizione. In particolare, tra i ricercatori non vi è ancora accordo nell'assegnare questo ruolo a casi di entità inanimate in posizione soggetto. Tale questione è certamente significativa 1- in ragione dell'annotazione pratica di questi casi 2- in quanto interessa la definizione generale di ruoli semantici. Sosteniamo*

che nomi di entità strumentali in posizione soggetto non sono casi del ruolo Strumento, ma dei ruoli di Causa, Agente o Tema. Ambiguità nella loro annotazione sono dovute alla confusione tra ruoli semantici e tipi ontologici.

1 Background

Semantically annotated resources have become widely used and requested in the field of Natural Language Processing, growing as a productive research area. This trend can be confirmed by looking at the repeated attempts in the implementation of annotated resources (FrameNet, VerbNet, Propbank, SALSA, LIRICS, SensoComune) and in the task of automatic Semantic Role Labeling (Gildea and Jurafsky 2002, Surdeanu et al. 2007, Màrquez et al. 2008, Lang and Lapata 2010, Titov and Klementiev 2012 among others).

Since their first introduction by Fillmore (1967), semantic roles have been described and defined in many different ways, with different sets and different level of granularity - from macro-roles (Dowty 1991) to frame-specific ones (Fillmore et al. 2002). In order to reach a common standard of number and definition, the LIRICS (Linguistic Infrastructure for Interoperable Resources and Systems) project has recently evaluated several approaches for semantic role annotation and proposed an ISO (International Organization for Standardization) ratified standard that enables the exchange and reuse of (multilingual) language resources.

In this paper we examine some problematic issues in semantic role attribution. We will highlight a case, the Instrument role, whose definition and designation should be, in our opinion, reconsidered. The topic is particularly relevant since there is difference in its treatment in different lexical resources and since the theoretical debate is

still lively. Moreover, this matter highlights aspects of the nature of semantic roles, relevant both for their theoretical definition and for practical annotation, such as the difference between semantic roles and ontological types. The former refer to the role of participants in the particular event described by the linguistic utterance, the latter to the inherent properties of the entity. We argue that this is a main point in the annotation, because, even in the recent past, roles have been frequently tagged according to the internal properties of the entities involved, not, as it should be, because of their role in the particular event described.

This analysis arose from the first step of the implementation of the Senso Comune resource (Vetere et al. 2012). With the aim to provide it with semantic roles, a first annotation experiment was conducted to check the reliability of the set and the annotation procedure (Ježek et al. 2014). The dataset was composed of 66 examples without disambiguation, 3 for 22 target verbs, and it was annotated for semantic roles by 8 annotators. They were instructed with a guideline in which a set of 24 coarse-grained roles was defined, with examples and a taxonomy. During the evaluation process, the major cases of disagreement were highlighted. The present study is based on the evidence coming from these data: the Instrument role caused several misunderstandings (see also Varvara 2013). Nevertheless, our analysis will look primarily at examples from literature and other resources in order to rethink this role and to reach a standardization. We propose to consider what are called instrument subjects (Alexiadou and Schäfer 2006) as instances of three different roles (Cause, Agent and Theme) rather than as Instrument.

2 The case of instrument subjects

With “instrument subjects” we refer to examples in which a noun, denoting an inanimate entity frequently used as instrument by humans (and occurring in *with*-phrases), is the subject of the sentence, as in the examples below (Levin 1993:80, Schlesinger 1989:189): “**The hammer** broke the window”, “**The stick** hit the horse”. In the past, it has been frequently asserted that these subjects cover the role of Instrument (Fillmore 1967, Nilsen 1973, Dowty 1991), as much as the nouns preceded by the preposition *with*: “David broke the window **with a hammer**”, “Marvin hit the horse **with a stick**”. In Levin (1993)’s terms, these

are called “Instrument-Subject alternation”¹. On the other side, several authors have argued against this interpretation, suggesting other roles to these cases (Schlesinger 1989, DeLancey 1991, Van Valin and Wilkins 1996, Alexiadou and Schäfer 2006, Grimm 2013, among others). Although this interpretation is the most recent one and many scholars agree on that, in the implementation of lexical resources the trend is to consider instrument subjects as Instrument role. In Verbnet, instrument subjects are tagged with the role Instrument, as can be seen in the annotation of the verb *hit*: “**The stick** hit the fence”; “**The hammer** hit the window to pieces”; “**The stick** hit the door open”. In the LIRICS guidelines (Schiffrin and Bunt 2007:38) the Instrument-Subject alternation is used as exemplification of the role definition: “He opened the door [with the key (Instrument)]”; “[The brick (Instrument)] hit the window and shattered it.” The reason of the annotation of these last examples is not clear if we look at the role definition (as annotators usually do). It is said that the Instrument is the “participant in an event that is manipulated by an agent, and with which an intentional act is performed” (2007:38). Here, the agent and the intentionality of the act are explicitly mentioned, but while annotating the examples above a question arises: in order to tag a noun with the role Instrument, should the Agent be present in the context of the event in which the Instrument occurs, should it be inferable or could it be totally absent? We argue that, in order to assign the Instrument role, an Agent should be specified in the event representation and it should be linguistically expressed. From our data, it seems that, in presence of instrument subjects, there is not an Agent, neither expressed neither included inferentially in the scene. In the cases observed, it is clear that there are reasons for which speakers left the intentional Agent out of the scope of their utterance. Their intention could be to describe the instrument

¹The traditional examples of “instrument subjects” cover also other Levin’s alternations, such as Characteristic property alternation (1993:39) or Middle alternation (1993:26). Even the examples that will be a matter of discussion in the present study are ascribable to different alternations. We will then consider the term “instrument subject” in a broad way, taking into account every noun that can occur both in a *with*-phrase, both in subject position. Even if this term may bring confusion with the real semantic role Instrument, we will adopt it because of a lack of other appropriate terms. To avoid difficulties, we will use the capital initial letter for semantic roles and the lower initial for the words in their common sense (e.g. Agent vs agent).

noun as an autonomous entity, as the only known source of causation, not as an Instrument manipulated by an Agent, and as such its role in the event should be considered. In the next section, we will list and group in classes the occurrences of instrument subjects that we have encountered so far, according to our proposal.

3 Why instrument subjects do not perform the Instrument role

Nowadays it is a shared opinion that semantic roles are relational notions that express the role of participants in reference to the event expressed by the verb. As pointed by Pethukova and Bunt (2008), semantic roles should be defined not as primitives “but rather as relational notions that link participants to an event, and describe the way the participant is involved in an event, rather than by internal properties”(2008:40). From this statement, we argue that semantic roles should be considered as semantic qualities attributed to a participant not only in a particular event, but in the specific linguistic representation of that event. The same event can be the object of two different sentences that represent the event from different perspectives. In the words of DeLancey (1991:350): “case roles, like any other semantic categories, encode construals of events rather than objective facts”. This is the mistake that we make when we evaluate an instrument subject as Instrument role. Consider the examples “The janitor opened the lock with a key” and “The key opened the lock”. “The underlying argument is that since “*the key*” in 19 (the first example) is an Instrument, and since 19 and 20 could refer to the same scenario, “*the key*” must be Instrument in 20 (the second example) as well” (DeLancey 1991:348). Actually, examples like the second one are often not realistic, invented by linguists. We believe that, looking at corpus data, it appears clearly that subjects like “the key” are not usually represented as an instrument used by an human, but as a Cause that substitutes an unknown Agent in the causal chain (as in the previous example) or as an entity whose a characteristic is described (e.g. the property of opening a lock in an example such as “This key opens the lock”). As referenced in the Introduction, our idea is that instrument subjects usually cover the role of Cause, Theme or, metaphorically, Agent.

3.1 Instrument subjects as Cause

Most frequently instrument subjects cover the role of Cause. It is usually the case when: 1- it is not possible to find another Agent or general causer other than the instrument inanimate subject; 2- it is possible to imagine an Agent that has “activated” the inanimate entity, but it is no longer present in the scene or it is not known. This could be a choice of the speaker that does not want to include or talk about the Agent or it could be the case with generic events with non specific agents. Consider the example “The clock was ticking so loudly that it woke the baby”(DeLancey 1991: 347): it is not possible to find another Agent other than the clock. The same can be seen in this sentence taken from the corpus ItTenTen (Jakubček et al. 2013): “Un masso caduto da una galleria ha messo fuori uso la metro. Il sasso ha rotto il pantografo, l’antenna che trasmette l’energia al treno, e ha interrotto la tensione per 600 metri di linea aerea” (“A stone falling down from a tunnel put out of order the metro. The stone has broken the pantograph, the spar that transmits the energy to the train, and it has interrupted the tension for 600 meters”). The stone is a Cause² because nobody has thrown it, but it has taken its own energy by its falling³. The same interpretation could be applicable to the sentence cited before from the LIRICS guidelines “The brick hit the window and shattered it”: from this context we do not know if there is an agent that has thrown the brick; if we do not have evidence about that, we cannot consider “the brick” an Instrument in this sentence. There are cases in which our real-world knowledge enables us to understand that the instrument subject has been manipulated by somebody, but it has been focused in the sentence as the principal or the only known element of the causal chain⁴: “The poison killed its victim”, “The camomile

²The definition of the role Cause in SensoComune is “participant (animate or inanimate) in an event that starts the event, but does not act intentionally; it exists independently from the event”.

³A reviewer pointed out that the real Cause is the event of falling, not the stone. Although this is a true inference, we argue that the stone is metonymically reinterpreted as the falling of the stone and for this reason the cause of the event. This interesting matter deserves a deeper analysis that will be subject of further work.

⁴Alexiadou and Schäfer note: “They are Causers by virtue of their being involved in an event without being (permanently) controlled by a human Agent. The fact that this involvement in an event might be the result of a human agent having introduced these Causers is a fact about the real world, not about the linguistic structure” (2006: 42-43).

cured the patient”. There is a case of this sort in the dataset of the SensoComune’s annotation experiment. The subject of the sentence “leggi che colpiscono il contrabbando” (“Laws that hit the smuggling”) has been tagged by 2 annotators upon 8 as Instrument role instantiation: it is possible that they have thought that there was an inferred Agent (the legislator) that was using the laws as an instrument. Putting instruments as subjects can be seen as a stylistic means adopted by the speaker to “defocus” the Agent: “ricorda teste sbattute contro il muro, saluto romano, ustioni con sigaretta e accendino. Un’altra le minacce mentre **le forbici** tagliavano ciocche di capelli” (‘she remembers heads hit against the wall, cigarette and lighter burns. Another the threats while the scissors cut locks of hair’). Lastly, instrument subjects can be Cause if the sentence expresses a generic event with a non-specific agent: ‘The piano addressed this by a mechanism where the way the key is struck changes the speed with which the **hammer** hits the string’.

3.2 Instrument subjects as Agent

We argue that the cases in which an instrument subject covers the role of Agent are sporadic and involve metaphorical or metonymical interpretations (Jezek et al. 2014). It should be kept in mind that it is widely assumed that the Agent role implies animacy and intentionality; as such an inanimate entity like an instrument noun cannot be Agent. This view contrasts with what has been claimed by some linguists (Schlesinger 1989, Alexiadou and Schäfer 2006) that were arguing anyway against the Instrument role attribution to instrument subjects. The Agent role can be fulfilled by instrument subjects in case of personification or metaphorical extension of the meaning of the lexeme: “Un giorno una forbice gigante tagliò della carta a forma di burattino. Un altro giorno ha ritagliato due palle giganti che erano il sole giallo e la Terra” (“Once upon a time a giant scissor cut a paper into a puppet. Later, it cut two giant balls, the yellow sun and the Earth”); “Tante penne scrivono su Napoli, usano Napoli per vendere copie” (“A lot of pens (writers) write about Naples, they use Naples to sell”); “Tutto l’ufficio ha lavorato bene” (“All the office has worked well”).

3.3 Instrument subjects as Theme

Analyzing the SensoComune dataset, a case has been found that has not been previously discussed in the literature on semantic roles. The examples to which we refer are: “**La penna** scrive nero” (“The pen writes black”), “**Forbici** che tagliano bene” (“Scissors that cut well”). These subjects have been tagged as Instrument by respectively 3/8 and 4/8 annotators. As previously claimed, the ambiguity is caused by the possibility of these nouns to occur as real Instrument with the preposition “with” (ex. “I have written the letter with this pen”). We suggest that in these cases the instrument subjects are neither Instrument, nor Cause, because they are not presented as causing an event or as being used by an Agent. The verb predicates a property of the subject and as such the Theme role is fulfilled. The Theme is defined in SensoComune as “participant in an event or state, which, if in an event, it is essential for it to take place, but it does not determine the way in which the event happens (it doesn’t have control) and it is not structurally modified by it; if in a state, it is characterized by being in a certain condition or position throughout the state and it is essential to its occurring”. In other resources, these examples could be referred to roles similar to our Theme, such as the role Pivot in LIRICS.

4 Conclusion

In this paper we have shown how theoretical and data analysis can be mutually improved by each other. Literature has offered critical discussion about the Instrument role and the case of instrument subjects, a discussion that can be useful for the definition and annotation of semantic roles in the implementation of lexical resources. Moreover, analysis of annotated data can reveal fallacies in the reliability of the set, coming back from application to theoretical topics. At last, our study highlights the importance of distinguishing between semantic roles - relational notions belonging to the level of linguistic representation - and ontological types, which refer to internal qualities of real-world entities. We believe that this topic, because of its importance, should be taken into consideration for a more complete treatment in future work.

Acknowledgments

Thanks to Guido Vetere, Laure Vieu, Fabio Zanzotto, Alessandro Oltramari and Aldo Gangemi for the ongoing collaborative work on semantic role annotation within the Senso Comune initiative, and to the participants of the 10th Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation at LREC 2014 for feedback on early results of this work. We also acknowledge two anonymous reviewers for their very useful comments and suggestions.

References

- Alexiadou, A. and F. Schäfer. 2006. Instrument subjects are agents or causers. *Proceedings of West Coast Conference on Formal Linguistics*, vol.25
- Burchardt, A., E. Katrin, A. Frank, A. Kowalski, S.Padó, M. Pinkal. 2006. The salsa corpus: a german corpus resource for lexical semantics. *Proceedings of LREC 2006*
- DeLancey, S. 1991. Event construal and case role assignment. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, vol.17
- Dowty, D. 1991. Thematic proto-roles and argument selection. *Language*, 126: 547-619
- Fillmore, C.J. 1967. The case for case. *Universals in Linguistic Theory*, Bach and Harms (eds). New York, Holt, Rinehart and Winston edition.
- Fillmore, C. J., C. F. Baker and H. Sato. 2002. The framenet database and software tools. *Proceedings of the Third International Conference on Language Resources and Evaluation*, vol.4
- Gildea, D. and D. Jurafsky 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245288.
- Grimm, S. 2013. The Bounds of Subjecthood: Evidence from Instruments. *Proceedings of the 33rd Meeting of the Berkeley Linguistic Society*, Berkeley Linguistic Society.
- Jakubiček, M., Kilgarriff, A., Kovář, V., Rychlý, P., Suchomel, V. 2013. The TenTen Corpus Family *Proceedings of the International Conference on Corpus Linguistics*.
- Ježek, E., Vieu L., Zanzotto F.M., Vetere G., Oltramari A., Gangemi A., Varvara R. 2014. Extending ‘Senso Comune’ with Semantic Role Sets. *Proceedings 10th Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation*, LREC 2014.
- Lang, J., and Lapata, M. 2010. Unsupervised induction of semantic roles. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 939-947
- Levin, B. 1993. *English verb classes and alternations: A preliminary investigation*. University of Chicago press Chicago, IL.
- Màrquez, L., Carreras, X., Litkowski, K. C., and Stevenson, S. 2008. Semantic role labeling: an introduction to the special issue. *Computational linguistics*, 34(2), 145-159.
- Nilsen, Don L. F. 1973. *The instrumental case in english: syntactic and semantic considerations..* The Hague; Paris: Mouton.
- Petukhova, V. and Bunt, H.C. 2008. LIRICS semantic role annotation: Design and evaluation of a set of data categories. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco, 2830.
- Schiffirin, A. and Bunt, H.C. 2007. LIRICS Deliverable D4.3. Documented compilation of semantic data categories. <http://lirics.loria.fr>.
- Schlesinger, I.M. 1989. Instruments as agents: on the nature of semantic relations. *Journal of Linguistics*, 25(01). 189210.
- Surdeanu, M., L. Màrquez, X. Carreras, and P. R. Comas. 2007. Combination strategies for semantic role labeling. *Journal of Artificial Intelligence Research (JAIR)*, 29:105151.
- Titov, I., and Klementiev, A. 2012. A Bayesian approach to unsupervised semantic role induction. *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp.12-22. Association for Computational Linguistics.
- Van Valin, R. D. and D. P. Wilkins. 1996. The case for “effector”: case roles, agents, and agency revisited. *Grammatical constructions: Their form and meaning.*, eds. Shibatani and Thompson, 289322. Oxford: Oxford University Press.
- Varvara, R. 2013. *I ruoli tematici: Agente, Strumento e la nozione di causa*. Master thesis. University of Pavia
- Vetere, G., A. Oltramari, I. Chiari, E. Jezek, L. Vieu, F.M. Zanzotto. 2012. ‘Senso Comune’: An Open Knowledge Base for Italian. *Revue TAL (Traitement Automatique des Langues), Journal Special Issue on Free Language Resources*, 52.3, 217-43.