Conceptual and Representational Choices in Defining an ISO Standard for Semantic Role Annotation

Harry Bunt
TiCC, Tilburg Center for
Cognition and Communication
Tilburg University,
Tilburg, The Netherlands
harry.bunt@uvt.nl

Martha Palmer
Department of Linguistics
University of Colorado
Boulder, Co.
USA
martha.palmer@colorado.edu

Abstract

This paper presents two elements of the ISO standard for semantic role annotation which is under development (ISO CD 24617-4:2013), namely (a) the metamodel, which describes the types of concepts that may occur in semantic role annotation and their conceptual relations, and (b) an annotation language for expressing semantic role annotations, with its abstract syntax, XML-based dconcrete syntax, and semantics.

1 Introduction

ISO project 24617-4, Language resource management Semantic annotation framework Part 4: Semantic Roles, has the aim of defining an international standard for the annotation of semantic roles, including an inventory of core semantic roles defined as ISO data categories, and an annotation language with an XML-based representation format and a formal semantics.

Semantic roles are receiving increasing interest in the information processing community because they make explicit the key conceptual relations of participation between a verb and its arguments, i.e., they specify Who did what to whom, and when, where, why, and how. For English alone, there are already several different semantic role frameworks, including FrameNet, VerbNet, LIRICS, EngVallex and PropBank (see Fillmore & Baker, 2004; Kipper-Schuler, 2005; Schiffrin & Bunt, 2007; EngVallex, 2011; and Palmer et al., 2005, respectively). Although these have been developed independently, there are strong underlying compatibilities between these frameworks, and they share a central definition of what a semantic role is, and what its span is, within an individual sentence. In addition to defining key concepts, the ISO standard aims at clarifying and specifying these underlying compatibilities and providing where possible a mapping between similar semantic roles across different frameworks. This mapping illustrates how different semantic role definitions can be linked to each other across frameworks, and presupposes a specification of clearly defined criteria for distinguishing semantic roles.

The specification can be used in two different situations:

- in annotations where the semantic roles are recorded in annotated corpora;
- as a dynamic structure produced by automatic systems; a process typically called semantic role labelling (SRL)

The objectives of this specification are to provide:

- A reference set of data categories defining a structured collection of semantic roles with an explicit semantics.
- A pivot representation based on a framework for defining semantic roles that could facilitate mapping between different formalisms (alternative semantic role representations/syntactic theories/eventually different languages) promoting interoperability.
- Guidelines for creating new resources that would be immediately interoperable with pre-existing resources

The ISO semantic roles project follows a design strategy for semantic annotation projects that includes (a) the design of a conceptual model which contains the key concepts involved in the kind of semantic annotation and which describes how these concepts are related; such a model is called a 'metamodel' (see Bunt & Romary, 2004), and (b) the three-part definition of an annotation language, the parts being (1) an 'abstract syntax', specifying how the basic concepts defined by the metamodel may be combined into set-theoretic structures called annotation structures'; (2) a 'concrete syntax', defining a reference representation format, typically using XML, for representing the annotation structures defined by the abstract syntax, and (3) a formal semantics describing the meaning of annotation structures (see Bunt, 2010; 2013 for a description of this methodology, called the CASCADES methodology). This paper focuses primarily on the metamodel constructed in the project for semantic role annotation (section 2) and the definition of the annotation language (3). For a more detailed description of the frameworks discussed and of semantic roles in general see the ISO document ISO 24617-4:2013, Bonial et al. (2011) and Johnson et al. (2001). The paper concludes with a brief discussion of what has been achieved and what remains to be done.

2 A metamodel for semantic role annotation

2.1 Predicate-argument structures and eventualities

A predicative expression with its arguments can be viewed semantically as describing an actual or hypothetical eventuality with its participants. Associated with the predicate (most prototypically a verb) is a subcategorization frame, describing the participants that are expected in that particular type of eventuality. Each slot in the subcategorization frame can be given a semantic role label which can then be associated with any argument that fills that slot. In the most fine-grained view each individual lexical item can be seen as defining a unique eventuality type with a unique set of possible participants.

Different predicative expressions may share the same or a very similar set of possible participants. Obvious examples are nouns and adjectives that constitute derived forms of the same lexical item (*observe, observance, observer*). Other examples are *buy* and *sell*, and *give* and *receive*. Depending on the desired level of generalization, the grouping of lexical items into shared subcategorization frame classes may stop there (this is one view of the PropBank Frame Files) or may continue to include a small set of items with very closely related semantics (the FrameNet view) or may extend to include items that share specific patterns of argument types but may have a fairly tenuous semantic relation (the VerbNet view). These frameworks take the subcategorization frame as a whole into consideration when determining the choice of individual semantic roles; this is motivated by examples such as *replace*, which can have one participant as the old item being replaced and another participant as the new item replacing it, with an obvious dependency between these two roles.

LIRICS does not use subcategorization frames or any other a priori association of semantic roles, but uses a set of features, like intentionality of the involvement of a participant, to distinguish among individual semantic roles, in the spirit of Dowty (1991). For example, in (1a), the behaviour of 'Martin' is clearly intentional, and he would be assigned the Agent role. In (1b), there is no intentionality involved, and *The lightning* would be assigned the Cause role. Sentence (1c) is ambiguous as to whether Martin's behaviour caused the children to be frightened as an intended or as an unintended effect, and so the semantic role of *Martin's behaviour* is either Agent or Cause.

(1) a. Martin frightened the children by pulling faces at them.

b. The lightning frightened the children.

c. Martin's behaviour frightened the children.

Note that the same word can have multiple senses, each of which might be associated with a distinct event type, and therefore a distinct frame. In this case the word could be represented by several eventuality types, each one associated with a different frame or class. Therefore, for the approaches to semantic role labelling embodied in FrameNet, PropBank, EngVallex and VerbNet, there are three core elements that must be defined for semantic role labelling:

1. the word sense, or lexical unit, under consideration;
2. the frame associated with that word sense; and
3. specific semantic role labels associated with each slot in that frame that will be assigned to the participants filling the slot.

The more examples that can be provided to illustrate the degree of syntactic variation available to each sense, the better. These examples, or instances, are considered tokens that are each associated with the appropriate type definition.

An additional consideration in defining any semantic role labelling scheme is exactly which constituents are labeled as adjuncts and whether or not a set of general adjunct types is defined. It is notoriously hard to draw a clear line between arguments of a verb and adjuncts, and approaches to semantic role labelling differ in how they draw such a line, or finesse the question by giving individual labels to adjuncts associated with each eventuality type. Finally, frames may include information about likely semantic types of the semantic roles being specified.

The frames associated with a semantic role labelling scheme specify the roles associated with the eventuality types. (For FrameNet they would be the FrameNet Frames, for PropBank and for EngVallex they are the PropBank role sets or framesets, and for VerbNet they are defined in VerbNet classes.) The frames are typically consulted during annotation to guide the decisions and ensure consistency. This makes the specification of the frame a critical step in the path towards an annotated corpus. For each predicate in a language, a meta-level description of the predicate and its arguments needs to be created, with examples, which constitutes the definition of the eventuality type frame.

2.2 Eventualities, participants, types and tokens

Figure 1 visualizes the conceptual view that underlies semantic role annotation according to standard ISO 24617-4 under development. A predicative expression in natural language, in the sense in which it is understood in a given utterance, is viewed as denoting a certain type of eventuality, and the occurrence of the verb form in the utterance as denoting an instance (or 'token') of that type of eventuality. Each eventuality type has a semantic role set or 'frame' defined, which determines the possible choices of individual semantic roles for the participants in an instance of that eventuality type. Eventuality types may further be grouped into classes that have similar role sets, possibly defining hierarchies of event classes/types and the corresponding role sets/frames (not shown in Fig. 1).

Like eventualities, participants also have a semantic type, typically expressed by the lexical item that serves as the nominal head of a noun phrase or that forms the central element in a predicative expression. The metamodel in Fig. 1 indicates that in a given utterance, the semantic roles relate the participants that occurrences of nominal (or adverbial) lexical items refer to, to the eventualities corresponding to an occurrence of a verb (or noun, or other event-denoting predicative expression). Participants and eventualities are both tokens of certain types, which pertain to a semantic type system.

Since annotations add linguistic information to stretches of primary data, the identification of relevant stretches in the data is essential. In standoff format, this realized through pointers to the primary data (the original text) or to elements at another layer of annotation, such as a syntactic parse, where the regions of primary data are identified. Following ISO practice, the term *'markable'* is used to refer to the entities that anchor an annotation directly or indirectly in the primary data. Note that the metamodel stipulates that participants and eventualities are expressed by markables in the original text ('source document'), but that semantic roles are not textually expressed.

3 SemRolesML

3.1 Abstract syntax

The abstract syntax of an annotation language consists of two parts (Bunt, 2010): (a) a specification of the elements from which annotation structures are built up, called a 'conceptual inventory', and (b) a specification of the possible ways of combining these elements in set-theoretical structures, called 'annotation structures'.

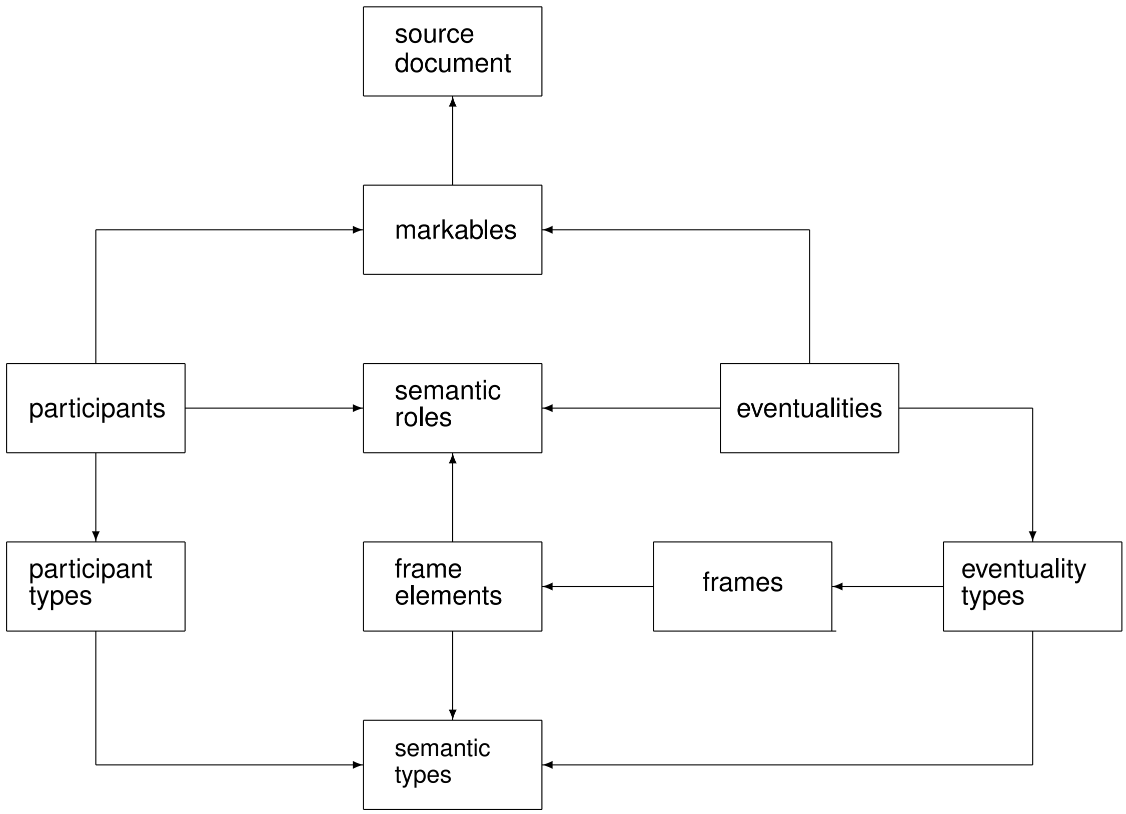

Figure 1: Metamodel for semantic role annotation.

a. Conceptual inventory

The conceptual inventory of the SemRoleML markup language, defined as part of ISO 24617-4, is derived from the metamodel shown in Fig. 1 by identifying among the categories of concepts in the metamodel those which are elementary and those which are composite, the latter being defined in terms of other concepts occurring in the metamodel. The listing of the basic concepts constitutes the conceptual inventory.

Of the ten categories represented in Fig. 1, the ‘source document’ is present only as a source of the markables and a carrier of possibly relevant metadata. Of the other nine categories, ‘participants’ and ‘eventualities’ are tokens of the basic concepts ‘participant type’ and ‘eventuality type’, respectively, and are identified by the occurrences of predicates and argument NPs in certain markables; as such they are instances (or ‘tokens’) of basic concepts, rather than basic concepts themselves. (Technically, they correspond to so-called ‘entity structures’ in the abstract syntax, see below.)

Concepts from the three categories at the bottom of Fig. 1, ‘frames’, ‘frame elements’ and ‘semantic types’, do not necessarily show up in semantic role annotations (but they often do in FrameNet annotations); they are especially important in the lexical resources supporting semantic role annotation. With respect to our abstract syntax, frames are a composite concept, that include n-tuples of frame elements. Frame elements include pairs of semantic role labels and specifications of the most likely semantic type of a participant playing that role, and are thus also composite concepts. So the five categories of elemenary concepts that form the SemRoleML conceptual inventory are: *markables, semantic roles, participant types, semantic types*, and *eventuality types*.

The specification of the SemRoleML conceptual inventory is thus the following listing of elementary concepts:

1. EV, a finite set of eventuality types, typically corresponding to verbs, nouns and adjectives.

2. RL, a finite set of semantic roles, such as the LIRICS role set (Schiffrin and Bunt, 2007; Petukhova and Bunt, 2007). This set can have a hierarchical organization, such as the unified VerbNet-LIRICS hierarchy presented by Bonial et al. (2011), with lower tiers expressing more fine-grained meanings, however this is not part of the conceptual inventory as such, but follows from the definitions of these roles (cf. Miltsakaki et al., 2008).

3. MA, a finite set of markables to which semantic roles can be attached.

4. PT, a finite set of participant types.

5. ST, a finite set of semantic types. The set PT of participant types and the set EV of eventuality types are subsets of ST.

b. Annotation Structures

An annotation structure is a set of entity structures and link structures. An entity structure is a pair $\langle m, s\rangle$ consisting of a markable (element of MA) and a specification of semantic information about that markable. For semantic role annotation, entity structures describe the eventualities and participants (both at token level) that are related by semantic roles. There are two kinds of entity structures in SemRoleML, those where the component s characterizes an eventuality and those where it characterizes a participant.

A link structure in SemRoleML is a triplet $\langle \epsilon_e, \epsilon_p, \rho\rangle$ consisting of two entity structures ϵ_e and ϵ_p, corresponding to an eventuality and a participant, respectively, and a semantic role specification ρ, which is either simply a semantic role label R or a pair $\langle \phi, R\rangle$, where ϕ is a frame, i.e. a list of frame elements $\phi = \langle \phi_1, \phi_2, \phi_k\rangle$. A frame element is either just a specification of a semantic role, or a pair $\langle R_i, t_i\rangle$ consisting of the specification of a semantic role and a semantic type (expected to subsume the participant type of a participant filling that role).

For the example sentence (2) two entity structures are created, one for the markable *The soprano*, and another one for the markable *sang*, shown in (3):

(2) The soprano sang

(3) a. $\epsilon_1 = \langle \textit{the soprano}, \text{SOPRANO}\rangle$
b. $\epsilon_2 = \langle \textit{sang}, \text{SING}\rangle$

For easy of readability, the strings *the soprano* and *sang* are used here to indicate markables (i.e. an occurrence of a stretch of text in the source document), SOPRANO is a participant type (an element of PT), and SING is an eventuality type (an element of EV).

A link structure is moreover created consisting of the two entity structures ϵ_1 and ϵ_2 and the semantic role *Agent*. The link structure is thus the triplet:

(4) $L_1 = \langle \epsilon_1, \epsilon_2, Agent\rangle$

The annotation structure for sentence (2) is the pair consisting of these entity structures and link structure(s):

(5) $\alpha = \langle \{\epsilon_1, \epsilon_2\}, \{L_1\}\rangle$

Note that ST, the set of semantic types, can be used to distinguish semantic roles and help determine their applicability. These are specified as selectional preferences by VerbNet, and are often included in the textual descriptions in FrameNet. As with the semantic roles, inheritance relations can hold between semantic types; these can be based on an hierarchical classification such as the hypernyms in WordNet (Miller, 1990; Feelbaum, 1998). In the example *The soprano sang*, the verb *sing* will plausibly have a frame which specifies that the frame element for the Agent slot expects a participant with the semantic type ANIMATE (or maybe HUMAN $\cup$ BIRD, if we agree that only humans and birds sing); since sopranos are humans, the semantic type system should include the knowledge SOPRANO $\subset$ HUMAN, and therefore the participant type is indeed subsumed by the semantic type.

The frames discussed above specify for each eventuality type the associated set of semantic roles, and can be used to guide the annotation process. Each frame consists of an eventuality type, e (an element of EV), and a subset, S_e, of RL with at least one element, such that $e \in EV$, and $r_i \in RL$ for all $r_i \in S_e$. For example, the frame for *sing* as occurring in example (2) above would consist of the eventuality type, SING, and the possible roles, including *Agent* and *Theme*, both of which are members of RL.

3.2 Semantics

The CASCADES design methodology (Bunt, 2013), used in the development of ISO 246171-4, derives a formal semantics for a given abstract syntax through a translation of the components of annotation structures to discourse representation structures (DRSs, Kamp and Reyle, 1994), which are combined by unification operations into a DRS for the annotation structure as a whole.

An entity structure $\langle m, s\rangle$ is interpreted as a DRS which introduces a discourse marker paired with a name of the markable m,[1] and which contains for each component s_i of s a condition of the form $p_i(x, a_i)$, where a_i is the interpretation of the component s_i, p_i is a predicate that indicates the role of a_i, and x is the newly introduced discourse marker. So the entity structures ϵ_1 and ϵ_2 are interpreted as the following DRSs, where m_1 names the markable *the soprano* and m_2 the markable *sang*:

$$\text{(6) a. } \epsilon_1 \rightsquigarrow \begin{array}{|l|}\hline \langle m_1, x_1\rangle \\ \hline \text{PARTICIP\_TYPE}(x_1, soprano) \\ \hline \end{array}$$

$$\text{b. } \epsilon_2 \rightsquigarrow \begin{array}{|l|}\hline \langle m_2, e_1\rangle \\ \hline \text{EVENT\_TYPE}(e_1, sing) \\ \hline \end{array}$$

A link structure $\langle\langle m, s\rangle, \langle m', s'\rangle, \rho\rangle$ is interpreted as a DRS which introduces discourse markers z_1 and z_2, paired with the markables m and m', respectively, and which has a condition of the form $R'(z_1, z_2)$, where R' is the DRS-predicate interpreting the relation ρ.

So the link structure L_1 of (4) is interpreted as the following DRS:

$$\text{(7) } L_1 \rightsquigarrow \begin{array}{|l|}\hline \langle m_1, z_1\rangle, \langle m_2, z_2\rangle \\ \hline \text{AGENT}(z_1, z_2) \\ \hline \end{array}$$

Merging these interpretations of the entity and link structures results in the following interpretation of the annotation structure (5):

$$\text{(8) } \alpha \rightsquigarrow \begin{array}{|l|}\hline \langle m_1, x_1\rangle, \langle m_2, e_1\rangle \\ \hline \text{PARTICIP\_TYPE}(x_1, soprano) \\ \text{EVENT\_TYPE}(e_1, sing) \\ \text{AGENT}(e_1, x_1) \\ \hline \end{array}$$

Once the DRS-interpretations of the entity structures and link structure have been combined (see footnote 1), the markable names can be deleted, resulting in a DRS of the usual kind.

A classical DRS is semantically equivalent to a formula in first-order logic; in this case the equivalent formula is (9), which says that there exist an eventuality, an eventuality type, a participant, and a participant type, such that the eventuality is a token of the eventuality type, the participant is a token of that participant type, and the participant is the agent of the event.

$$\text{(9) } \exists e_1.\exists et_1.\exists p_1.\exists pt_1.\ \text{EVENT-TYPE}(e_1, et_1) \land \text{PART-TYPE}(p_1, pt_1) \land \text{AGENT}(e_1, p_1)$$

In this semantic representation, AGENT is a first-order predicate constant that expresses the meaning of the semantic role Agent. The hardest part of the semantics of SemRoleML is in fact the formal definition of the logical predicates that express the meanings of the individual semantic roles. Defining these predicates comes down to formalizing the semantic role definitions in ISO CD 24617-4: 2013, Annex A. Figure 1 shows three examples of these definition. The Agent role, for example, is defined as one where a participant initiates and carries out an event intentionally or consciously, and who exists independently of the event. The condition of acting 'intentionally or consciously' distinguishes the Agent role from the Cause role; the existence independently of the event forms one of the distinctions between the Agent and Cause roles on the one hand and the Result role on the other hand (and, more significantly, also distinguishes the Result role from the Theme and Patient roles).

The formalization of such definitions can be used to complete the semantics of semantic role annotations; for example, the interpretation (9) of the

[1] The paring of discourse markers with markable names serves to ensure that, when an annotated text is interpreted which contains more than one occurrence of the same stretch of text, the right occurrences are combined in the semantics. See Bunt (2012) for details.

SemRoleML annotation of the sentence *The soprano sang* can be completed by replacing the predicate AGENT by (10a). Similarly, the semantics of CAUSE can be described by (10b).

(10) a. AGENT = $\lambda e.\lambda x.$ [Intent-Init(x,e) $\vee$ Consc-Init(x,e)] $\wedge$ [Intent-Do(x,e) $\vee$ Consc-Do(x,e)] $\wedge$ Indep-Exist(x,e)

b. CAUSE = $\lambda e.\lambda x.$ Init(e) $\wedge$ $\neg$Intent-Init(x,e) $\wedge$ $\neg$Consc-Init(x,e) $\wedge$ $\neg$Intent-Do(x,e) $\wedge$ Indep-Exist(x,e)

For some frameworks this approach to the semantics of semantic roles could be almost prohibitively burdensome. FrameNet has thousands of frame elements, and while VerbNet has less than 30, the definitions of each one can change subtly from class to class. On the other hand, this is perhaps the only way to semantically make sense of these elements with a formal rigour, required for automatic inferencing.

3.3 Concrete syntax

Following the CASCADES design methodology,a reference representation format for annotation structures, based on XML, can be defined as follows, given an abstract syntax specification.

1. For each element of the conceptual vocabulary define an XML name;
2. For each type of entity structure $\langle m, s\rangle$ define an XML element with the following attributes and values:
   (a) the special attribute `@xml:id`, whose value is an identifier of the entity structure representation;
   (b) the special attribute `@target`, whose value represents the markable m;
   (c) attributes whose values represent the components of s, and which themselves represent the significance of the components;
   (d) if s_i is an elementary concept then it is represented by its name.
3. For each type of link structure $\langle \epsilon_1, \epsilon_2, \rho\rangle$ define an XML element with three attributes, two which have values that refer to the representations of the entity structures ϵ_1 and ϵ_2, the value of the third denoting the semantic relation between them.
4. For each type of auxiliary structure (see below) specify an XML representation.

Applied to the abstract syntax of SemRoleML, this results in the following concrete syntax:

1. The XML elements `<event>` and `<participant>` are defined for representing entity structures corresponding to eventualities and participants, respectively. Both of these elements have the attributes `@xml:id` and `@target`, and additionally they have the attributes `@eventType` and `@participantType`, respectively.
2. XML constants are chosen for the values of the attributes `@eventType` and `@participantType`.
3. The XML element `<srLink>` is defined for representing semantic role link structures; this element has the attributes `@event` and `@participant` whose values refer to the eventuality and the participant that are related by a semantic role, and the attribute `@semRole` whose value represents the semantic role of the participant in the eventuality.
4. For completeness, we mention that it is convenient to introduce auxiliary structures in the abstract syntax for frames and frame elements, which may occur within the relational component ρ of a link structure $\langle \epsilon_e, \epsilon_p, \rho\rangle$; see ISO CD 24617-4 (2013) for more details.

For the example sentence *The soprano sang* this gives us the following representation of the annotation structure (5):

(11)

```
<event xml:id="e1"
 target="#m2"
 eventType="sing"/>
<participant xml:id="x1"
 target="#m1"
 participantType="soprano"/>
<srLink event="#e1"
 participant="#x1"
 semRole="agent"/>
```

/agent/	
Definition	Participant in an event who initiates and carries out the event intentionally or consciously, and who exists independently of the event.
– Source	Adapted from Dowty [1989], EAGLES, SIL, Sowa [2000] and UNL
Explanation	An agent may be animate, or only seemingly, or perceived, as animate; this is so that cases of nonhuman agency such as a robot, or an institution will not be excluded from being able to initiate an event, e.g. "GM offers rebates on its new models".
Example	"John [agent e1] built e1 the house"

/cause/	
Definition	Participant in an event that initiates the event, but that does not act with any intentionality or consciousness; the participant exists independently of the event.
– Source	Adapted from: SIL (Causer) and Sowa [2000] (Effector)
Explanation	Except for the lack of intentionality of the participant, this semantic role is very similar to that of the agent and in fact shares all its other properties. The role of cause can often be identified with verbs of initiation, or causation, such as: to cause, to produce, to start, to originate, to occasion, to generate.
Example	"The wind [cause e1] broke e1 the window" "His talk [cause e1] produced e1 a violent reaction e2 from the crow

/result/	
Definition	Participant in an event that comes into existence through the event. It indicates a terminal point for the event: when it is reached, then the event does not continue.
– Source	Adapted from Sowa [2000]
Explanation	Result is the completed point of a process, and unlike goal is dependent upon the event for its existence.
Example	"(Within the past two months [duration e1]) (a bomb [cause e1]) exploded e1 (in the offices of El Espectador in Bogota [location e1]), (destroying e2 (a major part of its installations and equipment [patient e2]) [result e1])"

Figure 2: Examples of LIRICS semantic role definitions in the form of ISO data categories (from Schiffrin & Bunt, 2007)

4 Conclusion

In this paper we have described a number of fundamental decisions in the process of defining an international ISO standard for the annotation of semantic roles. Starting from the conceptual view of predication in natural language as referring to (actual or hypothetical) eventualities and their participants, and of semantic roles as ways in which a participant may be involved in an eventuality, we outlined a metamodel which specifies the categories of basic concepts involved in semantic role annotation, and which shows how these concepts are interrelated. We subsequently defined an annotation language, SemRoleML, which has an XML-based pivot representation format for semantic role annotations, and a semantics that is defined for an abstract syntax that underlies these representations. We showed how the formalization of semantic role definitions can in principle be the basis of a semantics of semantic role annotations.

Two advantages of defining the semantic role annotation language SemRoleML in this way, following the CASCADES methodology of defining semantic annotations, are

(1) that different representation formats, used to encode the same underlying abstract structures,

share the same semantics, and are thus semantically interoperable;

(2) that integration of the annotation of semantic roles with the annotation of other types of semantic information, such as information about time and events according to ISO 24617-1, or about spatial information (ISO 24617-7, under development) or about discourse relations (ISO 24617-8, under developent) is facilitated, since these all follow the same design methodology;

(3) that annotations of other linguistic phenomena, especially when following the ISO Linguistic Annotation Framework (ISO 24613:2012), such as annotations of syntactic, pragmatic and contextual information, can be combined with semantic role annotations; many of these are helpful and sometimes even necessary to determine word senses and resolve references for the automatic recognition of semantic roles.

All this helps to make these annotation schemes mutually interoperable and combinable.

Important work that remains to be done is the formalization of all the semantic role definitions which are included in ISO CD 24617-4, including the specification of meaning postulates for the predicates used in their interpretation, in order to fully specify the inferences that may be drawn from the semantic roles used in an annotated corpus.

References

Bonial, Claire, William Corvey, Volha Petukhova, Martha Palmer, and Harry Bunt (2011) A Hierarchical Unification of LIRICS and VerbNet Thematic Roles, in *Proceedings ICSC Workshop on Semantic Annotation for Computational Linguistic Resources (SACL-ISCS 2011)*, September 21, 2011, Stanford, CA.

Bunt, Harry (2010) A Methodology for designing semantic annotation languages exploiting syntactic-semantic iso-morphisms. In: Alex Fang, Nancy Ide, and Jonathan Webber (eds.) *Proceedings ICGL 2010, the 2nd International Conference on Global Interoperability for Language Resources*, Hong Kong, pp. 29–45.

Bunt, Harry (2012) Annotations that effectively contribute to semantic interpretation. Forthcoming in Harry Bunt, Johan Bos and Stephen Pulman (eds) *Computing Meaning, Vol. 4*. Springer, Berlin.

Bunt, Harry (2013) A methodology for designing semantic annotations. Forthcoming in *Language Resources and Evaluation*.

Bunt, Harry and Laurent Romary (2004) Standardization in Multimodal Content Representation: Some Methodological Issues. In *Proceedings LREC 2004*, Lisbon, pp. 2219-2222.

Dowty. David (1991) Thematic Proto-Roles and Argument Selection. *Language*, 67:547-619.

EngVallex 2011 Charles University in Prague. Available at (Accessed 9/10/2012): http://ufal.mff.cuni.cz/lindat/EngVallex.html

Fellbaum, Christiane (1998) *WordNet: An Electronic Lexical Data-base. Language, Speech and Communications.* MIT Press, Cambridge, MA.

Fillmore, Charles; Collin Baker, and Hiroaki Sato (2004). FrameNet as a "Net". In *Proceedings 4th International Conference on Language Resources and Evaluation (LREC 2004)*, pp. 1091-1094

ISO 24612:2012 Language Resource Management - Linguistic Annotation Framework. Spring 2012. ISO, Geneva.

ISO 24617-1:2012 Language Resource Management - Semantic Annotation Framework, Part 1: Time and events. ISO International Standard, Spring 2012. ISO, Geneva.

ISO CD 24617-4:2013 Language Resource Management - Semantic Annotation Framework, Part 4: Semantic roles. ISO Committee Draft, March 2013. ISO, Geneva.

Johnson, Christopher R., Charles J. Fillmore, Esther J.Wood, Josef Ruppenhofer, Margaret Urban, Miriam R. L. Petruck, and Collin F. Baker, 2001. The FrameNet Project: Tools for Lexicon Building. Unpublished Report, University of Berkeley.

Kamp, Hans and Uwe Reyle, (1993) From discourse to logic. Kluwer, Dordrecht.

Kipper-Schuler, Karin. 2005. VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon. Ph.D. Thesis,University of Pennsylvania.

Miller, George A., 1990, WordNet: An on-line lexical database. *International Journal of Lexicography* 3(4) pp. 235312

Miltsakaki, Eleni, Livio Robaldi, Alan Lee and Aravind Joshi (2008) Sense Annotation in the Penn Discourse Treebank. In *Computational Linguistics and Intelligent Text Processing*, Lecture Notes in Computer Science Vol. 4919. Springer, Berlin, pp. 275-286.

Palmer, Martha, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank:An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71106.

Petukhova, Volha, Harry Bunt and Amanda Schiffrin (2007) LIRICS semantic role annotation: Design and evaluation of a set of data categories. In *Proceedings 6th International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech.

Schiffrin, Amanda and Harry Bunt. 2007. LIRICS Deliverable D4.3. Documented compilation of semantic data categories. `http://lirics.loria.fr`.