Combining semantic annotation schemes through interlinking

Harry Bunt
Department of Cognitive Science and Artificial Intelligence,
School of Humanties and Digital Sciences, Tilburg University
harry.bunt@tilburguniversity.edu

Abstract

This paper explores the possibilities of using combinations of different semantic annotation schemes. This is particularly interesting for annotation schemes developed under the umbrella of the ISO Semantic Annotation Framework (ISO 24617), since these schemes were intended to be complementary, providing ways of indicating different semantic information about the same entities. However, there are certain overlaps between the schemes of SemAF parts, due to overlaps of their semantic domains, which are a potential source of inconsistencies. The paper shows how issues relating to inconsistencies can be addressed at the levels of concrete representation, abstract syntax, and semantic interpretation.

Keywords: semantic annotation, ISO standards, combination of annotation schemes, interlinking

1. Introduction

Existing semantic annotation schemes are nearly always focussed on a specific type of semantic information, such as TimeML (Pustejovsky, 2003) on time and events, SpatialML (Mani et al., 2010) on spatial information, DAMSL (Allen & Core, 1997) on dialogue acts, PDTB (Prasad et al, 2008; 2019) on discourse relations, and RAF (Reference Annotation Framework, Salmon-Alt & Romary, 2005) on coreference. In a similar vein, the ISO Semantic Annotation Framework (ISO 24617, 'SemAF') was set up as a multi-part standard, with different parts focussing on different semantic domains. Table 1 lists the SemAF parts that have defined an annotation schema, with an indication of their semantic domain in the leftmost column. The second column specifies the SemAF part number, so for example the part that focuses on the annotation of time and events has defined the standard schema ISO 24617-1, the part for annotating dialogue acts the standard ISO 24617-2, and so on. The third column contains an unofficial name of the standard, which is often used for being mnemonically easier than the official ISO number. The rightmost column indicates some of the most important sources of each SemAF part.

Developing the SemAF standard as a set of separate sub-standards has proved useful, as it is more feasible to develop an annotation schema for a well-delineated semantic domain, and can benefit from the participation of different groups of experts for different domains. The first two parts of SemAF, informally known as 'ISO-TimeML' and 'DiAML', respectively, are successful examples of the application of this approach, as the annotation of time and events is clearly separable from the annotation of dialogue acts. However, some of the semantic domains are not entirely disjoint. The annotation schemes of the various SemAF parts are therefore not entirely complementary, and some semantic phenomena are covered in more than one sub-standard. More specifically, semantic phenomena that play central stage in one domain may play a peripheral role in another domain. For example, the temporal expression *"every Monday"* quantifies over mondays. Being a temporal expression, ISO-TimeML provides an annotation of this expression, including an indication of its quantifying character. ISO-TimeML has only a rudimentary treatment of quantification, however (Bunt & Pustejovsky, 2016), while it is the focus of SemAF part 12, QuantML.[1]

The marginal treatment of temporal quantification can be seen as a limitation of ISO-TimeML; on the other hand, ISO-TimeML offers a more detailed treatment of events and temporal entities than QuantML, which can be seen as a limitation of QuantML. Limitations of this kind are no problem when annotating language data with (a) information about events and time, or (b) about quantifications, but they present a problem for annotating data about *both* quantifications *and* time and events. In the latter case, one would like to combine the possibilities offered by the two annotation schemes. One way to do this is to define a new annotation schema that makes use of elements from the two schemes. In this paper we explore another idea: the combination of annotations provided by two (or more) annotation schemes without modifying them, but adding links between elements of the annotations in order to express that the two schemes annotate the same primary data with a different focus.

[1] At the time of writing, QuantML was the subject of a ballot for obtaining the status of an international ISO standard. See also Bunt (2024).

ISA-20 Proceedings @LREC-COLING-2024, pages 111–121
20 May, 2024. © 2024 ELRA Language Resource Association: CC BY-NC 4.0

Semantic domain	#	Name	Source
Time and Events	1	ISO-TimeML	TimeML (Pustejovsky, 2003)
Dialogue acts	2	DiAML	DIT++ (Bunt, 2007)
Semantic roles	4	ISO-SR	LIRICS and VerbNet, (Palmer & Bunt 2013, Bonial et al. 2011)
Spatial information	7	SpaceML	SpatialML (Mani et al., 2010; Pustejovsky & Lee, 2015)
Dscourse relations	8	DR-Core	PDTB (Prasad et al, 2008, 2019)
Coreference	9	ISO-RAF	RAF, Reference Annotation Framework (Salmon-Alt & Romary, 2005)
Measurable Quantitative Information	11	MQI	(Hao et al., 2019)
Quantification	12:	QuantML	(Bunt, 2019a) (under review)

Table 1: SemAF parts that have defined an annotation schema

The idea of this technique, *'interlinking'*, is very simple: given two annotation schemes A and B which represent different information about the same event or other kind of entity, interlinking adds to the A- and B-annotations an identity relation between the corresponding elements. This is illustrated in Figure 2, where a mini-discourse is annotated with TimeML, QuantML, and DR-Core, which all use XML-based representations, with <idLink>s indicating that the same three events are annotated in each of the three schemes.

This paper is organised as follows. Section 2 discusses related work. Section 3 summarises the ISO Semantic Annotation Framework as far as relevant for the present study, and explores overlaps and inconsistencies between SemAF parts. Section 4 specifies the mechanism of *interlinking*, with detailed examples. Section 5 summarises the present study, including its limitations, and an outlook of future work.

2. Related Work

The interest in combining annotation schemes has three main reasons.

First, specialised annotation schemes restricted to a specific semantic domain, like those of the SemAF parts, has the danger of designing schemes that have certain gaps, which may limit the coverage of individual annotation schemes in unwelcome ways for corpus annotation. Examples of such gaps are:

(1) anaphorically expressed participants in events cannot be annotated in QuantML, ISO-TimeML, and SpaceML (other than by simply assuming anaphora to have been resolved);

(2) temporal and spatial quantification have no adequate treatment in ISO-TimeML and SpaceML (Bunt & Pustejovsky, 2016);

(3) although semantic roles play a central role in QuantML annotations, they are undefined there - that is the subject matter of ISO-SR.

Some of these gaps could be resolved by combining SemAF annotation schemes, such as ISO-TimeML and QuantML, or SpaceML and ISO-RAF, or QuantML and ISO-SR.

Second, semantic annotation may play an important role in applications which require not just the annotation of one semantic domain, such as time and events, but also of other domains, such as coreference and discourse relations. This is for example the case in an application discussed by Silvano (2021) and Leal (2022), who used concepts from different SemAF annotation schemes to design a new, integrated schema to meet the requirements of the application. The design of integrated annotation schemes is also addressed in Malchanau et al. (2024).

Third, the markup language of an annotation schema may be used not only for the annotation of corpus data, but also as an internal interface language in an NLP system. For example, the dialogue act markup language DiAML has been used as an internal language in which the modules of an interactive language-based system communicate, in particular as an interface language for dialogue management (Malchanau, 2019). When used for this purpose, a notable limitation of DiaML is that, while it supports a rich annotation of dialogue acts, their communicative functions, and relations between them, it does not provide a way to indicate their semantic content. This limitation has been addressed by Bunt (2019), who proposed the use of *annotating schema plug-ins* for adding descriptive (and semantic) power to a host annotation schema.

Besides the definition of integrated schemes that combine elements from different schemes, which and the addition of plug-ins to a host annotation schema, another option is explored in this paper, in which existing annotation schemes are used in combination without altering them,.

3. The Semantic Annotation Framework

3.1. Architecture of SemAF Parts

All parts of SemAF follow the same architecture, described in ISO 24617-6: Principles of semantic annotation see also Bunt (2015) and Pustejovsky et al. 2017). QuantML thus has a triple-layered definition consisting of:

1. An abstract syntax, which specifies the class of well-defined *annotation structures* as pairs, triples, and other set-theoretical constructs containing quantification-related concepts. Annotation structures consist of *entity structures*, which contain information about a stretch of primary data, and *link structures*, which contain information relating two (or more) entity structures. The role of the abstract syntax is visualized in Figure 1.

2. A semantics, which specifies the meaning of the annotation structures defined by the abstract syntax. QuantML has an interpretation-by-translation semantics, which translates annotation structures to discourse representation structures (DRSs, Kamp & Reyle, 1993). The use of DRSs is mainly motivated by the fact that this formalism is also used in other SemAF parts.

3. A concrete syntax, that specifies a representation format for annotation structures The QuantML definition includes an XML-based reference format, again mainly motivated by the use of XML in other standards.

The three levels are interrelated by encoding (F_{ac}), decoding (F_{CA}), and interpretation functions; see Figure 1. Since the semantics is defined at the level of the abstract syntax, alternative representation formats may be used that share the same abstract syntax, as indicated in Figure 1 and are thus semantically equivalent. This adds to the interoperability of the schema.

3.2. Complementarity of SemAF parts

The various parts of SemAF are intended to be complementary, dealing with different semantic domains. However, as noted above, these domains often have overlaps, which is a potential source of inconsistencies. In particular, because of the common event-based semantic approach, events and their participants and the relations between them play a role in several SemAF parts. The following example highlights some of these overlaps, showing the information that six SemAF parts would annotate for the mini-discourse of (1a).

(1) a. After moving the pianos to the stage, the men had a beer. They were thirsty.

b. **ISO-TimeML:** a **move event** occurred, followed by a **beer-drinking event** which occurred **in the past**. A **be-thirsty event** occurred **in the past**.

ISO-SR: a **move event** occurred with pianos as *Theme*s and a stage as *Final Location*. A **drinking event** occurred with some men as *Agent*(s) and some beer as *Patient*. A **be-thirsty event** occurred, with certain individuals as *Experiencer*s.

SpaceML: a **move event** occurred with a stage as **end point**.

DR-Core: a **move event** occurred which *caused* a **be-thirsty event**, which *explains* the occurrence of a **beer-drinking event**.

ISO-RAF: the set of *discourse entities* that "they" refers to *is the same as* the it set referred to by "the men".

QuantML: some **move events** occurred in which certain *contextually determined men* participated *collectively* as an **Agent**. The *men* acted *individually* as the *Agent* in **drinking events** with *some beer* as **Patient**. A **be-thirsty event** occurred, with certain individuals as **Experiencers**.

This example clearly shows that each of the annotation schemes focuses on different information, but information concerning events with their participants and relations plays a role in nearly all of them. In the next subsection we consider the consequences of these overlapps.

3.3. Overlaps of SemAF parts

3.3.1. Events

Events play central stage in ISO-TimeML, in which they have articulate annotations as illustrated in example (3). Events that involve motion are equally important in SpaceML, and have a similar articulate annotation there. For annotating events expressed by verbs, ISO-SR makes use of 'eventuality frames', borrowed from VerbNet, which allows distinctions to be made between different verb senses. ISO-TimeML proposes articulate annotations both for events described by verbs and for events described by nouns. QuantML and DR-Core treat events, regardless of their lexical description, as predicate constants (in the spirit of DRT and other formal semantic approaches).

Example (2) shows annotations of a *call* event in the sentence *Peter called this morning* represented in each of these annotation schemes. The value

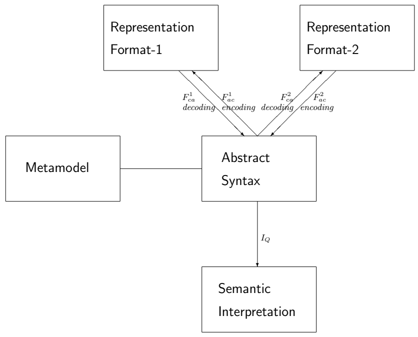

Figure 1: Architecture of SemAF parts.

(2) a. Peter called this morning.

b. **Representation of events** in various SemAF parts:

ISO-TimeML: `<event xml:id="e1" target="#w2" pred="call" class="occurrence" type="transition" pos="verb" tense="present" aspect="perfective" mood="none" polarity="positive" modality="certain"/>`

SpaceML: `<event> as in ISO-TimeML, with additional attributes (@latLong, @elevation,...)`

ISO-SR: `<eventuality xml:id="e1" target="#m2" eventFrame="#call.03"/>`

DR-Core, MQI: `<event xml:id="e1" target="#m2" type="call"/>`

QuantML: `<event xml:id="e1" target="#m2" pred="call" repetitiveness="1"/>`

'call.03' of the @eventFrame attribute in the ISO-SR annotation is assumed to identify the event frame for the intended sense of *call*, i.e. referring to an event that could also be described by the verb *to phone*.

To what extent are these alternative representations consistent? An important point to note is that all 6 annotations represent the *same* event, expressed in the primary data by the markable 'm2'. The ISO-TimeML representation just adds more information about the type of event and the way it is described in the primary data. A semantic difference between the ISO-TimeML and QuantML representations might seem to be that the latter is interpreted as a set of one or more events, whereas the ISO-TimeML representation refers to a single event. This is not quite the case, however, since the semantics of ISO-TimeML is defined by means of an existential quantifier, saying that *there has been a call-event such that...*, without ruling out that more than one event of the same type occurred. In this respect the two representations are therefore semantically equivalent. The additional @repetitiveness attribute in QuantML is used to accommodate expressions like *called twice*, indicating the cardinality of a set of events. If an annotation is intended to indicate the occurrence of a single event, this can be expressed in QuantML by the @repetitiveness attribute having the value '1'.

The fact that the various annotations represent the same concept, though possibly with more or less detail, will be essential for the interlinking mechanism described in the next section.

3.3.2. Participants

The entities that participate in events can be divided into (1) temporal and spatial entities, (2) events, (3) (measurable) quantities, and (4) objects of any other kind. Events participating in other events have the same articulate representation as the events in which they participate. Non-eventive entities have an articulate annotation in ISO-RAF, as shown in example (3). Entities of any kind (temporal, spatial, eventive, quantitative, other) occurring as participants in events all have articulate representations in QuantML; see example (3).

QuantML annotates the distinction between collective and individual (or 'distributive') quantification which is illustrated in example (1) if we assume that *the men* collectively moved *the pianos* and individually had a beer; therefore, participants in QuantML are represented by <entity> elements interpreted as sets.

(3) a. Peter called this morning.

b. **Representation of entities** as participants in events or inter-entity relations:

ISO-TimeML :

```
<timex3 xml:id="x1" target="#m3" pred="morning"...  />
```

ISO-RAF :

```
<discourseEntity xml:id="e1" target="#m1" abstractness="concrete"
         referentialStatus-"discourseNew" animacy="animate" ...  />
<discourseEntity xml:id="e1" target="#m3" animacy="inanimate" ...  />
```

QuantML :

```
<entity xml:id="x1" target="#m1" involvement="all"
            individuation="count" size="1"/>
  <refDomain xml:id="x2" target="#m1" pred="peter" determinacy="det"/>
<entity xml:id="x3 target="#m3" involvement="all"
            individuation="count" size="1"/>
  <refDomain xml:id="x4" target="#m3" pred="morning"
determinacy="det"/>
```

c. **Representation of relations** between events, participants, and time, as annnotated above and in (2):

ISO-TimeML :

```
<tLink eventID="#e1" relatedToTime="#x3" relType="isIncluded"/>
```

QuantML :

```
<participation event="#e1" participant="#x1" semRole="avent"/>
<participation event="#e1" participant="#x3" semRole="time"/>
```

Example (3) shows annotations of the participants in example (1). ISO-TimeML only provides a representation for the temporal expression *this morning*; ISO-RAF and QuantML provide a representation for both *Peter* and *this morning*. The QuantMbL representation indicates that both NPs are countable (as opposed to the mass NP *some beer* in example (1)), that both NPs quantify over a definite domain, consisting of only one individual in the case of the NP *Peter*, and that all the members of both domains participate in the event(s) under discussion.

3.3.3. Relations

The following SemAF parts annotate relations among events, participants, time and place:

ISO-TimeML represents (1) information about the time of occurrence of events; (2) temporal relations between events, as expressed by conjunctions of clauses or by a main clause and a subordinate clause; (3) temporal relations between temporal objects. All these relations are represented using <tLink> elements.

SpaceML represents (1) spatial information about the occurrence of events, including locations of begin and end points, trajectories and paths of movements, (2) spatial relations between spatial objects, using a variety of links.

ISO-SR represents relations between events and participants in terms of semantic roles.

QuantML uses the semantic roles defined in ISO-SR as attribute values in <participation> links, and moreover represents (1) non-temporal semantic relations between events, as expressed by a main clause and a subordinate clause; (2) relations between any two kinds of entities as expressed by noun-noun modifiers, possessives, prepositional phrases, or relative clauses, using various links, such as <nnMod>, <ppMod>, and <possMod>.

DR-Core represents semantic relations such as *Cause, Contrast, Concession, Elaboration* between events as expressed in a discourse by clauses either within the same sentence or in different sentences.

Inspecting the information represented in these annotation schemes, we can again see a great deal of complementarity, but also some overlaps, and hence a danger of inconsistencies. We discuss these in the next subsection.

3.4. Levels of inconsistency

The various SemAF parts display inconsistencies in representing the same information in different ways, or as representing more detailed and different information about the same events, entities, or relations. To what extent do the inconsistencies

"After moving the pianos to the stage, the men had a beer. They were thirsty."

Markables: m1 = "After", m2 = "moving", m3 = "the piano", m4 = "to", m5 = "the stage", m6 = "the men", m7 = "had" m8 = "a beer", m9 = "They", m10 = "were", m11 = '"were thirsty" m12 = "thirsty"

QuantML:

```
<entity xml:id="xQ1" target="#m3" refDomain="#xQ2" individuation="count"
   involvement="all"/>
<refDomain xml:id="xQ2" target="#m3" pred="piano" determinacy="det"/>
<entity xml:id="xQ3" target="#m5" refDomain="#xQ4" individuation="count" size="1"
   involvement="all"/>
<refDomain xml:id="xQ4" target="#m5" pred="stage" determinacy="det"/>
<event xml:id="eQ1" target="#m2" pred="move"/>
<participation event="#eQ1" participant="#xQ1" semRole="theme"/>
<participation event="#eQ1" participant="#xQ3" semRole="finalLocation/>
<entity xml:id="xQ5" target="#m6 refDomain="#xQ6" individuation="count" .../>
<refDomain target="#m6" pred="man" determinacy="det"/ ...  />
<participation event="#eQ2" participant="#xQ5" semRole="agent"/>

<event xml:id="eQ2" target="#m7" pred="drink"/>
<entity xml:id="xQ7" target="#m8 refDomain="#xQ8" individuation="count"
   involvement="some"/>
<refDomain target="#m8" pred="beer" determinacy="indet"/>
<participation event="#eQ2" participant="#xQ5" semRole="patient"/>
<event xml:id="eQ3" target="#m10" pred="be"/>
<predication event="#eQ3" participant="#xQ1" predicate="thirsty" distr="individual"/>
```

ISO-TimetML:

```
<event xml:id="eT1" target="#m2" pred="move" .../>
<event xml:id="eT2" ptarget="#m7" pred="drink" ...  tense="past" />
<event xml:id="eT3" ptarget="#m10"red="be-thirsty" .../>
<signal xml:id="s1" target="#m1" pred="after"/>
<tLink arg1="#eT1" arg2="#eT1" relType="after"/>
```

DR-Core:

```
<event xml:id="eD1" target="#m2" pred="move" .../>
<event xml:id="eD2" target="#m7" pred="drink" ...  tense="past" ...  />
<event xml:id="eD3" target="#m10" pred="be-thirsty" ...  tense="past" ...  />
<drLink arg1="#eD2" arg2="#eD1" relType="succession"/>
<drLink arg1="#eD3" arg2="#eD2" relType="cause"/>
```

Interlinking ISO-TimeML to QuantML:

```
<idLink arg1="#eQ1" arg2="#eT1"/>
<idLink arg1="#eQ2 arg2="#eT2"/>
<idLink arg1="#eQ3" arg2="#eT3"/>
```

Interlinking DR-Core to ISO-TimeML:

```
<idLink arg1="#eD1" arg2="#eT1"/>
<idLink arg1="#eD2 arg2="#eT2"/>
<idLink arg1="#eD3" arg2="#eT3"/>
```

Figure 2: Example of interlinking at the level of concrete syntax.

noted above actually present a problem? So far, we discussed inconsistencies at the level of concrete (XML-based) representation; the addition of interlinking <idLink> elements (or a similar device in other representation formats) seems relatively straightforward, and the intuitive meaning of the interlinks is simple and clear, but they might cause inconsistencies at the deeper levels of abstract syntax and semantics. To remain in line with the ISO principles of semantic annotation (ISO 24617-6), the entire structure formed by the concatenation of the representations of interlinked schemes and the links between them should have a well-defined abstract syntax with a semantic interpretation.

The inconsistencies between SemAF parts, due to overlapping semantic domains, can be divided into three categories:

1. Different terms used for the same concept, e.g. the attribute @pred in some of the schemes is called @type in others.
2. Different sets of attributes and values used to describe the same events or other entities, reflecting the focus of different schemes.
3. Different views on how events and other entities are conceptually related.

Inconsistencies of type (1) arise purely at the level of concrete syntax, have no semantic consequences, and may be considered trivial. The decoding function that computes the abstract syntax of interlinked annotations can simply map equivalent terms to the same concepts in the abstract syntax.
Inconsistencies of type (2) are potentially more serious, but not necessarily so. They are not problematic if the differences in sets of attributes correspond to semantically complementary information, or if one set of attributes and values is semantically more specific than another. An interesting case is the difference between ISO-TimeML and SpaceML on the one hand, and ISO-SR and QuantML on the other, regarding the annotation of relations between events and their time and place of occurrence. ISO-SR includes 4 temporal relations: *Time, Initial-Time, Final-Time*, and *Duration* and 5 spatial relations: *Location, Initial-Location, Final-Location, Distance, Path*, ISO-TimeML, by contrast, makes use of 7 relations: *Simultaneous, Includes, IsIncluded, Before, I-Before, After, I-After* (where I-Before and I-After mean immediately before and immediately after, respectively, and SpaceML has a large set of spatial relations. These differences reflect that ISO-TimeML and SpaceML have the domains of time and space as their respective focus, and these are semantically not problematic, since the relations of ISO-SR are less specific than those of ISO-TimeML and SpaceML, so the former entail the latter. This makes the 'inconsistency' semantically harmless (although somewhat redundant).
Inconsistencies of type (3) are the most fundamental, and are often the cause of a type (2) inconsistency. This is for example the case for temporal relations among events and for relations between events and time of occurrence. These cases, and all other cases in SemAF that we have examined, can all be treated in the same way as type (2) inconsistencies. Example (7) shows that interlinking can be used to accommodate different conceptual views at the level of concrete representations while providing a consistent semantic interpretation.

4. Interlinking

4.1. Concrete syntax

The example in Figure 3.3.3. illustrates the use of interlinking for the annotation structure that combines elements from ISO-TimeML and QuantML, where a mini-discourse is annotated with TimeML, QuantML, and DR-Core, with <idLink>s indicating that the same events are annotated in all three schemes.

4.2. Abstract Syntax

The decoding function of an annotation schema, which computes the abstract syntax of the concrete representation (see Fig. 2) uses the interlinking specifications to merge the semantic information about the same events and the same entities that occur in the respective annotations.
In QuantML, the unit of annotation is a clause. At the abstract syntax level, a clause annotation structure is a quadruple of the form (4), consisting of specifications of (1) an event; (2) a set of n participants ($n > 0$) (3) a set of n participation links; and (4) a set of $n-1$ scope links.

(4) $A_Q = \langle \epsilon_e, \{\epsilon_1, ...\epsilon_n\}, \{L_1, ...L_n\}, \{s_1, ...s_{n-1}\}\rangle$.

The abstract syntax of the annotations of other SemAF-parts that annotate events and participating entities are the same as (4) for a simple clause, except that the set of scope links is empty, as they do not annotate scope relations. Moreover, ISO-TimeML and SpaceML consider only temporal and spatial entities, and hence use specific time-and space-related relations rather than general participation relations. The interlinking of two or more of these annotation schemes has the effect of creating another annotation structure in the general quadruple form of (4), as follows.

Let X_A and X_B be the XML-representations of a clause, annotated according to the annotation schemes A and B, and X_{IL} the set of statements that interlink X_A and X_B. Application of the decoding functions F^A_{ca} and F^B_{ca} can be represented schematically as follows:

(5) $F^A_{ca}(X_A) = \langle \epsilon_A, E_A, L_A, sc_A\rangle$,
$F^B_{ca}(X_B) = \langle \epsilon_B, E_B, L_B, sc_B\rangle$

Let $\mathcal{R}_{IL}$ be the function that replaces in a given set of expression sll occurrences of an identifier x_i which occurs either as first or as second item in the set of pairs $F^{AB}_{ca}(X_{IL})$ by the corresponding pair $\langle \epsilon_{xAi}, \epsilon_{xBj}\rangle$ (where $\epsilon_{xAi} \in E_A$ and $\epsilon_{xBj} \in E_B$).
The decoding function F^{AB}_{ca} of the interlinked schemes constructs pairs of elements that correspond to the arguments of an <idLink> in the concrete syntax applied to the set X_{IL} of interlinks, a set of corresponding pairs $\langle \epsilon_{xAi}, \epsilon_{xBj}\rangle$ is constructed, where $\epsilon_{xAi} \in E_A$ and $\epsilon_{xBj} \in E_B$.
Using '+' to indicate concatenation, the decoding function applied to the entire XML representation $X_A + X_B + X_{IL}$, is defined as:

(6) $F_{ca}^{AB}(X_A + X_B + X_{IL}) = \langle \epsilon_{AB}, E_{AB}, L_{AB}, sc_{AB} \rangle$, with

a. $\epsilon_{AB} = \langle \epsilon_A, \epsilon_B \rangle$,

b. $E_{AB} = \mathcal{R}_{IL}(E_A) \cup \mathcal{R}_{IL}(E_B)$,

c. $L_{AB} = \mathcal{R}_{IL}(L_A) \cup \mathcal{R}_{IL}(L_B)$,

d. $sc_{AB} = \mathcal{R}_{IL}(sc_A) \cup \mathcal{R}_{IL}(sc_B)$

The set L_{AB} of event - entity links and the set of scope links sc_{AB} are computed in the same way as the set of entities E_{AB}, by merging the corresponding components of the linked schemes after replacing single entities by pairs in case they are interlinked.

4.3. Semantics

The semantic interpretation of interlinked A- and B-annotations is computed by the interpretation function I_{AB}, defined in terms of the interpretation functions I_A and I_B. Central in the definition of I_{AB} is the interpretation of pairs of events or pairs of participants which were linked by <idLink>s in the XML representation and which occur as participant pairs in the abstract syntax, simply as the merge of the two interpretations.[2]

(8) $I_{AB}(\epsilon_A, \epsilon_B) = I_A(\epsilon_A) \cup I_B(\epsilon_B)$

The semantic interpretation of a fully connected annotation schema, in which the relative scopes of all participants are specified, can be computed by combining the interpretations of all the event - entity link structures, since these structures embed the event structures and entity structures that describe the events and participants. This can be done in a compositional manner, using the semantics of scope links to determine how the interpretations of event and entity structures are combined; this has been worked out in detail for the semantics of QuantML (Bunt, 2023). The upshot of this is expressed in (9), where the set L_{AB} of link structures is ordered by their relative scopes; σ_{ij} is the composition function that is computed by applying I_{AB} to the corresponding scope relation in the abstract syntax.

(9) $I_{AB}(\epsilon_{AB}, E_{AB}, L_{AB}, sc_{AB}) = I_{AB}(L_{AB}) =$
$= I_{AB}(L_1, L_2, ...L_n)$
$= \sigma_{12}(I_{AB}(L_1, \sigma_{23}(I_{AB}(L_2, ...$
$I_{AB}(\sigma_{n-!,n}(I_{AB}(L_N)...)$

Example (7) shows in detail how this works out for the sentence *Ninety-five students graduated on a Friday*, instantiating the 'A' and 'B' in (5), (8), and (9) by 'Q' (for QuantML) and 'T' (for ISO-TimeML). The abstract syntax of the XML representation, computed by the decoding function F_{ca}^{QT}, is shown in (7b); its semantics as calculated by the interpretation function I_{QT} is shown in (7c) (where $\cup*$ is a scope-preserving merge operation oon DRSs; see Bunt, 2023). The XML representations, are slightly simplified to save space.

The final semantic interpretation, formulated as the DRS in (10), effectively says that there is a set ('X') of 95 students for whom there is a set of 1 friday, for which the description "XXXX-WXX-5" applies, which have graduation events as their time of occurrence, and include the time of occurrence. This combines the information in the QuantML and ISO-TimeML annotations. There is some redundancy in the final result, but such semantic redundancy is perhaps not very elegant, but formally harmless.

(10) [X | |X|=95, x ∈ X → [student(x),
[Y | |Y|=1, y ∈ Y → [friday(y),
value(y)="XXXX-WXX-5",
[E | e ∈ E → [graduate(e),
class(e)=occurrence, type(e)=transition,
agent(e,x), time(e,y), is_included(e,y)]]]]]]

5. Conclusion and Further Work

In this paper we have presented an exploration of the possibilities of using combinations of semantic annotation schemes. This seems particularly interesting for the use of annotation schemes developed under the umbrella of the ISO Semantic Annotation Framework, since these schemes were intended to be complementary, serving to express information in different semantic domains. The schemes developed as SemAF parts have certain unavoidable overlaps, however, due to unavoidable overlaps of semantic domains, which are a source of potentially problematic inconsistencies and which may be harmful for their interoperability.

For truly complementary schemes, like DiAML, QuantML, and DR-Core, the interlinking technique seems perfectly suitable. For interlinking annotations of overlapping schemes, such as ISO-TimeML and QuantML, we have shown promising possibilities for constructing semantically consistent interlinked annotations, but a more elaborate exploration of all the overlaps in SemAF parts is needed to fully evaluate this proposal.

[2]This notation assumes interpretations to have the form of DRSs. ISO-TimeML has a semantics defined in different terms, which is however readily converted to DRS form.

(7) “Ninety-five students graduated on a Friday”

Markables: m1 = “Ninety-five”, m2 = “Ninety-five students”, m3= “students”, m4= “graduated ”, m5 = “on”, m6 = “on a Friday”, m7 = “a Friday”, m8 = “Frday”

a. XML REPRESENTATION:

X_QuantML

<entity xml:id=”xQ1” target=”#m2” refDomain=””#x1” involvement=”95” individuation=”count”/>
<refDomain xml:id=”x1” target=”#m3” pred=”student” determinacy=”indet”/>
<event xml:id=”eQ1” target=”#m3” pred=”graduate”...>
<participation event=”#eQ1” participant=”#xQ1” semRole=”agent” >
<entity xml:id=”xQ2” target=”#m7” refDomain=””#x2” individuation=”count” involvement=”some”/>
<refDomain xml:id=”x2” target=”#m3” pred=”friday” determinacy=”indet”/>
<participation event=”#eQ1” participant=”#xQ2” semRole=”time”/>
<scoping arg1=”#xQ1” arg2=”#xQ2” scopeRel=”wider”/>

X_ISO-TimetML

<event xml:id=”eT1” target=”#m23 .... pred=”graduate” ...>
<signal xml:id=”s1” target=”#m5” pred=”on”/>
<timex3 xml:id=”xT1” target=”#m8” pred=”friday” type=”date” value=”XXXX-WXX-5”/>
<tLink signalID=”#s1” eventID=”#eT1” relatedToTime=”#xT1” relType=”isIncluded”/>

X_Interlinking:

<idLink arg1=”#eQ1” arg2=”#eT1”/>
<idLink arg1=”#xQ2” arg2=”#eT1”/>

b. ABSTRACT SYNTAX:

QuantML:

$A_Q = \langle eQ1, \{xQ1, xQ2\}, \{pL1, pL2\}, \{pL1, pL2,\text{wider}\rangle\}\rangle$

ISO-TimetML:

$A_T = \langle eT1, \{xT1\}, \{tL1\}, \{\}\rangle$

Interlinked structure :

$A_{QT} = \langle\langle eQ1, eT1\rangle, \{xQ1, \langle xQ2, xT1\rangle\}, \{pL1, pL2, tL1\}, \{\langle pL1, pL2,\text{wider}\rangle\}\}\rangle$

c. SEMANTICS:

QuantML:

$I_Q(A_Q) = I_Q(\langle eQ1, \langle xQ1, xQ2\rangle, \langle pL1, pL2\rangle, \langle pL1, pL2,\text{wider}\rangle\rangle$
$\cup^*(I_Q(xQ1), I_Q(eQ1), \langle I_Q(\text{agent, individual})\rangle)$
$\cup^*([X \subseteq \text{student} \mid |X|=95\ ], I_Q(\ [Y \subseteq \text{friday}| \ |Y|=1), I_Q(\text{agent, individual})\rangle)$
$= [\ X\subseteq \text{student} \mid |X|=95,\ x \in X \rightarrow$
$[Y \subseteq \text{friday}| \ |Y|=1,\ y \in Y \rightarrow$
$[\ E \subseteq \text{graduate} \mid e \in E \rightarrow [\ \text{agent(e,x), time(e,y)}]]]]$

ISO-TimetML:

$I_T(A_T) = [\ Y\subseteq \text{friday} \mid y \in Y \rightarrow [\ \text{value(y)=”XXXX-WXX-5”},$
$E \subseteq \text{graduate} \mid e \in E \rightarrow [\ \text{class(e)=occurrence},$
$\text{type(e)=transition, is\_included(e,y)}]]]$

Interlinked interpretation:

$I(A_{QT}) = I_{QT}(\langle\langle pL1, pL2,\text{wider}\rangle, \langle pL2, tL1, equal\rangle\rangle)$
$= \cup^*(I_{QT}(pL1), I_{QT}(pL2, tL1, \text{equal})$
$= \cup^*(I_{QT}(pL1),\ (I_{QT}(pL2) \cup I_{QT}(tL1)))$
$= \cup^*(I_Q(pL1),\ (I_Q(pL2) \cup I_T(tL1)))$
$= [X \mid |X|=95,\ x \in X \rightarrow [\text{student}(x), [Y \mid |Y|=1,\ y \in Y \rightarrow [\text{friday}(y), \text{y)=”XXXX-WXX-5”},$
$[E \mid e \in E \rightarrow [\ \text{graduate(e), class(e)=occurrence, type(e)=transition},$
$\text{agent(e,x), time(e,y), is\_included(e,y)}]]]]]]$

6. Bibliographical References

References

J. Allen and M. Core. 1997. *DAMSL: Dialogue Act Markup in Several Layers (Draft 2.1). Technical Report*. University of Rochester, Rochester, NY.

C. Bonial, W. Corvey, M. Palmer, V. Petukhova, and H. Bunt. 2011. A hierarchical unification of LIRICS and VerbNet semantic roles. In *Proceedings IEEE-ICSC 2011 Workshop on Semantic Annotation for Computational Linguistic Resources*, Stanford, CA.

H. Bunt. 2009. The DIT++ taxonomy for functional dialogue markup. In *Proceedings of AAMAS 2009 Workshop 'Towards a Standard Markup Language for Embodied Dialogue Acts'*, pages 13–24, Budapest.

H. Bunt. 2015. On the principles of semantic annotation. In *Proceedings 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-11)*, pages 1–13, London.

H. Bunt. 2019a. An annotation scheme for quantification. In *Proceedings 14th International Conference on Computational Semantics (IWCS 2019)*, pages 31–42, Gothenburg, Sweden.

H. Bunt. 2019b. Plug-ins for content annotation of dialogue acts. In *Proceedings 15th Joint ISO-ACL Workshop on Interoperable Semantic Annotation (ISA-15)*, pages 34–45, Gothenburg, Sweden.

H. Bunt. 2023. The Compositional Semantics of QuantML. In *Proceedings of the 19th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-19)*, pages 56 – 65, Nancy, France.

H. Bunt. 2024. QuantML: A proposed new standard for semantic annotation. In *Proceedings LREC-COLING 2024, Turin, Italy*, Paris. ELRA.

D. Davidson. 1967. The Logical Form of Action Sentences. In N. Resher, editor, *The Logic of Decision and Action*, pages 81–95. The University of Pittsburgh Press, Pittsburgh.

T. Hao, S. Liu, H. Wang, C. Xinyu, and K. Lee. 20219. The semantic annotation of measurable quantitative information. In *Proceedings of the 15th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-15)*.

ISO. 2012. *ISO 24617-1: 2012, Language Resource Management - Semantic Annotation Framework (SemAF) - Part 1: Time and events*. International Organisation for Standardisation ISO, Geneva.

ISO. 2014. *ISO 24617-4: 2014, Language Resource Management - Semantic Annotation Framework (SemAF) - Part 4: Semantic roles*. Geneva: International Organisation for Standardisation ISO.

ISO. 2015. *ISO 24617-6:2015, Language Resource Management - Semantic Annotation Framework (SemAF) - Part 6: Principles of semantic annotation*. International Organisation for Standardisation ISO, Geneva.

ISO. 2019. *ISO/WD 24617-12:2019, Language Resource Management: Semantic Annotation Framework (SemAF) - Part 12: Quantification*. International Standard. International Organisation for Standardisation ISO, Geneva.

ISO. 2024. *ISO/WD 24617-12: 2021, Language Resource Management - Semantic Annotation Framework (SemAF) - Part 12: Quantification*. International Organisation for Standardisation ISO, Geneva.

H. Kamp and U. Reyle. 1993. *From Discourse to Logic*. Kluwer Academic Publishers, Dordrecht.

A. Leal, P. Silvano, E. Amorim, I. Cantante snd F. Silva, A. Jorg, and R. Campos. 2022. The place of ISO-space in txet2story multilayer annotation scheme. In *Proceedings of the 18th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-17)*, pages 57–66.

A. Malchanau. 2019. *Cognitive architecture of multimodal multidimensional dialogue management*. Saarbruecken, publisher=Ph.D. dissertation, Saarland University.

A. Malchanau, V. Petukhova, and H. Bunt. 2024. On the principles of semantic annotation. In *Proceedings 20th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-20)*, Turin, Italy.

I. Mani, C. Doran, D. Harris, J. Hizeman aand R. Quimby, and J. Richer. 2010. SpatialML: annotatopn scheme, resources, and evaluation. *Language Resources and Evaluation*, 44 (3):263–280.

T. Parsons. 1990. *Events in the Semantics of English: A Study in Subatomic Semantics*. MIT Press, Cambridge, MA.

V. Petukhova and H. Bunt. 2008. LIRICS semantic role annotation: design and evaluation of a set of data categories. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco. ELRA, Paris.

R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber. 2008. The Penn Discourse Treebank 2.0. In *Proceedings 6^{th} International Conference on Language Resources and Evaluation (LREC 2008)*, Marrakech, Morocco.

R. Prasad, B. Webber, A. Lee, and A. Joshi. 20019. *Penn Discourse Teebank Version 3.0.* Linguistic Data Consortium.

J. Pustejovsky, H. Bunt, and K. Lee. 2010. ISO-TimeML. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*, Malta. ELDA, Paris.

J. Pustejovsky, H. Bunt, and A. Zaenen. 2017. Designing annotation schemes: From theory to model. In Nancy Ide and James Pustejovsky, editors, *Handbook of Linguistic Annotation*, pages 21–72. Springer, Berlin.

J. Pustejovsky, J. Castano, R. Ingria, R. Gaizauskas, G. Katz, R. Saurí, and A. Setzer. 2003. TimeML: Robust Specification of Event and Temporal Expressions in Text. In *Proceedings of the Fifth International Workshop on Computational Semantics (IWCS-5)*, pages 337–353, Tilburg, Netherlands.

S. Salmon-Alt and L. Romary. 2005. The Reference Annotation Framework: A Case for Semantic Content Representation. In *Proceedings of the Sixth International Workshop on Computational Semantics (IWCS-6)*, pages 259–270, Tilburg. Tilburg University Computational Linguistics and AI Group.

P. Silvano, A. Leal, F. Silva, I. Cantante, F. Oliveira, and A. Jorge. 2021. Developing a multilayer semantic annotation scheme based on ISO standards f or the visualization of a newswire corpus. In *Proceedings of the 17th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-17)*, pages 1–13.