Combining annotations of SemAF parts through interlinking

Extended version of a paper presented at ISA-20@LREC-COLING2024

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 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 (see Bunt, 2024)).[1] [2] [3]

The marginal treatment of temporal quantification can be seen as a limitation of ISO-TimeML; by contrast, ISO-TimeML offers a more detailed treatment of events and temporal entities than QuantML,

[1] At the time of writing, QuantML was the subject of a ballot for becoming an ISO standard. As a result of this ballot, QuantML has obtained the status of Draft International Standard (ISO DIS 24617-12).

[2] This paper is a slightly revised version of the original article entitled "Combining semantic annotation schemes through interlinking" (Bunt, 2024), extended with an addendum in which some of the considerations in the original version are discussed in more detail. The revisions concern (1) correction of typos, (2) improved notations in formulas, (3) addition of missing attributes and values in the XML fragments in example (3) and in Figure 2, (4) corrections in the examples (6) and (10), and (5) replacing the example that originally constituted Annex A by a 6-page Addendum which includes a more detailed version of that example.

[3] This is version 2.2, October 2, 2024.

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

which can be seen as a limitation of QuantML. Such limitations are no problem when annotating 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.

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 a set of identity relations between the corresponding elements. This is illustrated in Figure 2 below, 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 (eF), decoding (dF), and interpretation functions (I); 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.

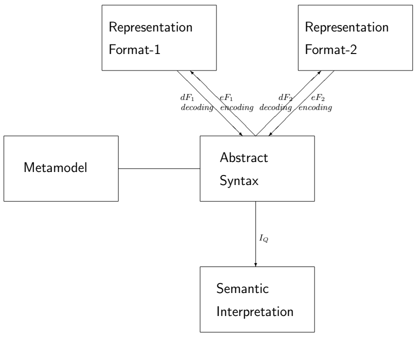

Figure 1: Architecture of SemAF parts.

(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="past" aspect="perfective" mood="none" polarity="positive"/>

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"/>

3.3. Overlaps of SemAF parts

3.3.1. Events

Events play central stage in ISO-TimeML, in which they have articulate annotations. Similarly, events that involve motion are 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 ISO-TimeML, DR-Core, and QuantML. The value '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) 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" type="date" pred="morning"
temporalFunction="true"/>
```

ISO-RAF :

```
<discourseEntity xml:id="e1" target="#m1" abstract-
ness="concrete"        referentialStatus-"discourseNew" ani-
macy="animate" naturalGender="male"/>
<discourseEntity xml:id="e1" target="#m3" animacy="inanimate"
abstractness="abstract"/>
```

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"/>
```

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.

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 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 at the price of often somewhat redundant semantic representations).

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 and other cases in SemAF that we have examined can all be treated in the same way as type (2) inconsistencies. Example (18) 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. Data segmentation: Markables

Each of the SemAF parts has its own way of choosing the segments of discourse that its annotations refer to, the markables

4.2. Concrete syntax

Figure 2 shows the concrete syntax of an interlinked annotation structure that combines elements from

“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-TimeML:

```
<event xml:id="eT1" target="#m2" pred="move" class ="occurrence"
   type="transition" tense="past" aspect="perfective"/>
<event xml:id="eT2" ptarget="#m7" pred="drink" class="occurrence"
   type="transition" tense="past" aspect="perfective"/>
<event xml:id="eT3" ptarget="#m10"red="be-thirsty" class="occurrence"
   type="state" tense="past" aspect="perfective"/>
<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.02"/>
<event xml:id="eD2" target="#m7" pred="drink"/>
<event xml:id="eD3" target="#m10" pred="be-thirsty"/>
<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.

ISO-TimeML QuantML, QuantML, and DR-Core. In particular, <idLink>s connect elements from the three respective annotations, making explicit that they all three annotate the same events.

4.3. 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 entity structure describing an event; (2) a set of n entity structures describing the 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 is the same as (4) for a simple clause, except that the set of scope links is empty for schemes that 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 (i.e., set of XML expressions) of a clause, annotated according to two annotation schemes A and B, and X_{IL} the set of statements that interlink X_A and X_B. The decoding functions dF_A and dF_B assign quadruples of the form (4) to X_A and X_B:

(5) $dF_A(X_A) = \langle e_A, E_A, L_A, sc_A\rangle$,
$dF_B(X_B) = \langle e_B, E_B, L_B, sc_B\rangle$

The decoding function dF_{AB} assigns similar quadruples to interlinked annotations. For linked events and participants this function produces pairs that correspond to the arguments of <idLink> elements in the concrete syntax.

Using 'X'_{IL} to indicate the set of interlinks, and'+' to indicate concatenation of XML structures, the decoding function dF_{AB} is defined as delivering the quadruple of which:

1. the first element is the pair $\langle dF_A(X_A), dF_B(X_B)\rangle$ formed by the event structures of X_A and X_B;
2. the second element is the set of participant entity structures of X_A and X_B that are not connected by interlinking elements in X_{IL} plus the pairs of those structures that are connected through interlinking elements in X_{IL};
3. the third element is formed by the event - entity link structures that are encoded in the concrete syntax by <participation> links, <tLink> elements, etcetera;
4. the fourth element consists of the scope relations of X_A and X_B, if any, plus triples which express that the event - entity relations expressed by interlinked components have equal scope. Such triples have the form $\langle L_{Ai}, L_{Bj}$,equal$\rangle$.

This is expressed formally in (6), where the indices 1 - 4 indicate the four elements of a quadruple, and the predicate 'interlinked' is defined in (7).

(6) $dF_{AB}(X_A + X_B + X_{IL}) = \langle e_{AB}, E_{AB}, L_{AB}, sc_{AB}\rangle$,
with
a. $e_{AB} = \langle (dF_A(X_A))_1, (dF_B(X_B))_1\rangle = \langle e_A, e_B\rangle$
b. $E_{AB} = (dF_A(X_A))_2 \cup (dF_B(X_B))_2 \cup \{\langle x, y\rangle |$ interlinked(x,y)$\}$
c. $L_{AB} = F^A((X_A)_3) \cup F^B((X_B)_3) = L_A \cup L_B$
d. $sc_{AB} = (F^A((X_A))_4) \cup (F^B((X_B))_4) \cup \{\langle x, y,$equal$\rangle |$ interlinked(x,y)$\}$

(7) interlinked(x,y) $=_D \exists x_a, y_b . a \in X_A, b \in X_B$, $F_A(a) = x, F_B(b) = y$,
<idLink arg1="#a" arg2="#b"/> $\in X_{IL}$

4.4. 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 are 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.[4]

(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

[4]The 'merge' may take different forms, depending on the formalism used in semantic representtions. for example, if DRSs are used as semantic representations, as is the case in QuantML, the the 'merge' is the DRS-merge operation.

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 (18) shows in detail how this works out for the sentence *More than fifty thousand students graduated on 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_{QT}, is shown in (18b); its semantics as calculated by the interpretation function I_{QT} is shown in (18c) (where $\cup *$ is a scope-preserving merge operation on 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 more than 50.000 students for whom there is a Friday, for which the description "XXXX-WXX-5" applies, which had graduation events as its time of occurrence, and which includes the time of occurrence. This combines the information in the QuantML and ISO-TimeML annotations. There is some redundancy in the final result, but though not very elegant, this is semantically harmless.

(10) [X | |X| $>$ 50.000, x $\in$ X $\rightarrow$ [student(x),
[Y | y $\in$ Y $\rightarrow$ [friday(y), past(y),
value(y)="XXXX-WXX-5",
[E | e $\in$ E $\rightarrow$ [graduate(e),
class(e)=occurrence, type(e)=transition,
agent(e,x), time(e,y), includes(y,e)]]]]]]

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.

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. 2024a. Combining semantic annotation schemes through interlinking. In *Proceedings of the 20th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-20)*, pages 111–122, Turin, Italy.

H. Bunt. 2024b. 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*. Ph.D. Dssertation, Saarland University, Saarbruecken.

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: annotation 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 6th 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.

Addendum

A1. ISO-TimeML and QuantML

As mentioned above, the most fundamental and potentially problematic cases of inconsistency between annotation schemes are those that come from different conceptual views. Examples of such differences between ISO-TimeML and QuantML are:

- Quantification over sets of events
- Polarity (negation)
- Modality
- Use of attributes that represent syntactic or lexical semantic information

A.2 Concrete syntax

The attribute @target is generally used in SemAF parts to identify the region of primary data that an annotation refers to. ISO-TimeML uses IDs of the TEI-based representation of primary textual data as values of this attribute. QuantML uses the slightly more general concept of 'markables' as @target values, rather than pointing directly into the primary data. Markables can more generally be representational structures that include pointers to primary data. In QuantML markables they are treated simply as pointers to stretches of text or transcribed speech. As long as the primary data are strictly sequential, as in text, the two methods are equivalent.

A.3 Abstract syntax

Both ISO-TimeML and QuantML have a specification of their abstract syntax, though the one of QuantML has been specified in more detail than that of ISO-TimeML, due to the fact that the latter was the first part of SemAF, established as ISO 24617-1:2012, and the ISO principles of semantic annotation including the three-layer architecture of concrete syntax - abstract syntax - semantics (ISO 24617-6:2015) had not yet fully been developed at the time.

The encoding and decoding functions that relate the concrete and abstract representations (see Fig. 1) have not been defined explicitly in the formal specification of ISO-TimeML and QuantML; the only SemAF part where this has been done is ISO 24617-2 ('DiAML'), where these functions have been defined for three alternative concrete representation formats and have been implemented in the DialogBank (Bunt et al., 2019).

A.4 Semantics

The official document defining ISO-TimeML as international standard 24617-1 contains two alternative versions of a semantics of the annotations: (a) one based on interval temporal logic (ITL, Pratt-Hartman 2005), and (b) an event-based semantics in terms of second-order logic (SOL).

The ITL-based semantics is defined for the concrete XML representations of the annotations, which violates the ISO principles of semantic annotation according to which the semantic should be defined for the *abstract* syntax. This makes the semantics vulnerable for changes in the details of using XML. Moreover, it *"concerns only a subset of ISO-TimeML. This is [partly] because*

(11) a. some of the information recorded in ISO-TimeML is essentially syntactic rather than semantic in character, and

b. much of the semantic content of natural language texts, for one reason or another, no clear rendition within the usual apparatus of formal semantics"

(ISO document 24617-1:2012, section 8.4.1).

In the next subsection we will argue that both of the complicating factors mentioned in (11) can be dealt with by exploiting the level of abstract syntax as intermediate between concrete syntax and semantics.

Different from the ITL-based semantics, the event-based SOL semantics of ISO-TimeML is defined for the abstract syntax. It is, however, technically not entirely satisfactory.

QuantML has a fully specified semantics for its abstract syntax, documented succinctly in the official document that specifies the annotation standard, and described in more detail in Bunt (2023).

A.5 Exploiting abstract syntax

The two issues quoted in (11) can both be tackled by exploiting the distinction between concrete and abstract syntax. The <event> element used in ISO-TimeML to represent information about events has the following three kinds of attributes:

(12) a. Attributes whose values represent semantic information: @pred, @class, @type, @polarity. These come in two varieties:

1. attributes whose values may differ from one occurrence to another, such as @polarity and @pred (in case of lexical ambiguity);

2. lexical semantic attributes, whose values are the same for every occurrence;

b. Attributes whose values represent syntactic rather than semantic information: @tense, @aspect, @pos (part-of-speech), @mood;

c. Attributes whose values represent semantic information that escapes known formal treatment: @modality

Since attributes of type (12a.1) represent information of a syntactic rather than a semantic nature, they can only play a marginal role in semantic annotations. They are relevant for identifying syntactic phenomena with potential semantic relevance, but do not directly define constraints on the semantic interpretation of the annotated material. This can be implemented in the three-layer framework of the ISO principles by defining the decoding function in such a way that they are left out of consideration. So for example, the decoding function dF_T of ISO-TimeML is defined in such a way that:

(13) dF_T(<event xml:id=“eT1” target=“#mT1 pred=“graduate” class=“occurrence” type= “process” tense=“past” aspect=“none” pos=“verb” polarity=“positive” modality=“none” mood= “none”)> =
dF_T(<event xml:id=“eT1” target=“#mT1 pred=“graduate” class=“occurrence” type= “process” polarity=“positive” modality= “none”)>

This resolves the problem of (11a).

QuantML does not employ any syntactic attributes in its concrete syntax, so the problem of (11a) does not occur.

The problem of (11b) occurs in QuantML in dealing with generic quantification, which is an unresolved issue in formal semantics. In the concrete syntax an attribute @genericity is used, which is optional and has the possible values “generic" and “specific", the latter being the default value. The decoding function dF_Q of QuantML therefore skips such attribute-value pairs. Such a treatment can also be used in ISO-TimeML for the uninterpreted attribute @modality. The abstract syntax expressions that are calculated this way by the decoding functions contain only strictly semantic information. This allows a complete formal semantic interpretation for the abstract syntax.

An additional difficulty concerns the use of lexical semantic attributes, such as @class and @type in ISO-TimeML, and @animacy and @alienability in the reference annotation standard ISO-RAF. Since the values of lexical semantic attributes are the same for every occurrence (with the same sense), they don't need to be annotated. ISO-SR addresses this by using VerbNet entries rather than predicate constants in the annotation of events denoted by verbs, Dealing with lexical semantics and lexical ambiguity is an issue that should be addressed more generally in SemAF parts.

A.6 Combining semantic interpretations

The key issue in combining semantic annotation schemes is how the semantic information in the respective annotations can be combined into a single representation structure. More precisely, the idea behind the semantic side of interlinking is that each annotation scheme represents certain semantic information, and the semantics of their combination merges these pieces into a richer semantic representation. The QuantML semantics uses second-order DRSs to represent the meaning of its annotations. The ISO-TimeML semantics using second-order logic with lambda abstraction is similar in spirit, though different in its details. More specifically, the ISO-TimeML semantics views the meaning of a temporal event annotation as something that expects to be combined with more information about the event and its participants in order to obtain a full-fledged event representation. This is expressed by using a lambda expression over (complex) predicates, as shown in (14).

(14) a. I_T(⟨ E, C, T, positive⟩ = $\lambda P.$ ∃e. $I_T(E)$(e) ∧ $I_T(C)$(e) ∧ $I_T(T)$(e) ∧ P(e)

b. I_T(⟨ E, C, T, negative⟩ = $\lambda P.$ ¬∃e. $I_T(E)$(e) ∧ $I_T(C)$(e) ∧ $I_T(T)$(e) ∧ P(e)

For example, for the sentence *“Bob graduated on Friday"* the instantiation of (14a) would be applied to the complex predicate (15a). with the resulting formula (15b). This formula is logically equivalent to the DRS in (15c).

(15) a. λz. agent(z,bob) ∧ ∃t. friday(t) ∧ includes(t,z)

b. ∃e. graduate(e) ∧ $I_T(C)$(e) ∧ $I_T(T)$(e) ∧ agent(e,bob) ∧ ∃t. friday(t) ∧ include(t,e)

c. [e, t | graduate(e), friday(t), $I_T(C)$(e), $I_T(T)$(e), agent(e,bob), includes(t,e)]

The QuantML annotation of the same sentence would have the following semantic representation in DRS-form:[5]

(16) [E ⊆ graduate, T ⊆ friday | |T|=1, t ∈ T ↔ friday_0(t), t ∈ T → [e ∈ E | [agent(e,bob), time(e,t)]]]

[5] With a slight simplification of the treatment of the proper name *Bob*.

Since quantification over individual entities is logically equivalent to quantification over singleton sets, the two representations (15c) and (16) are semantically equivalent, apart from the additional information in the ISO-TimeML representation regarding event class and type, and the more fine-grained temporal relation 'is-included' than the 'time' relation in QuantML. This means that the two representations can be fused into a single representation which in DRT-form might look as follows:

(17) [E ⊆ graduate, T ⊆ friday | |T|=1, t ∈ T ↔ friday_0(t), t ∈ T → [e ∈ E | occurrence(e), transition(e), includes(t,e), agent(e,bob), time(e,t)]]

The condition 'time(e,t)' in this representation is redundant but harmless.

A.7 Concluding remarks

This paper, and in particular the detailed example (18) in this Addendum (below), shows that the combination of overlapping semantic annotation schemes through interlinking looks promising. However, not all the details have yet been considered in full, in particular the following:

- Besides the various types of attributes in <event> elements listed in (12), ISO-TimeML also has a number of 'special' attributes (with 'special' values) in <timex3> elements, such as @functionInDocument and @value. Their semantic impact on the possibilities of combining with non-temporal information needs further examination.

- Regarding the use of uninterpreted attributes (of types b and c in (12)), the solution of leaving such attributes out of consideration when decoding concrete annotations and generating abstract annotation structures seems to work perfectly well for optional attributes, such as @genericity in QuantML, as well as for obligatory attributes that have a 'not applicable' value, such as @tense in ISO-TimeML with possible value "none".

  Other obligatory non-semantic attributes are problematic for the *encoding* of abstract annotation structures, i.e., for the generation of XML-based or other concrete representations from underlying abstract annotation structures, since the value of a non-semantic attribute is simply not available at the levels of abstract syntax and semantics. The application of an encoding function therefore generates incomplete concrete representations. This means that conversions between representation formats can no longer be done via their common abstract syntax. Whether this should be regarded as a serious problem deserves to be investigated..

- Negation and temporal quantification are treated differently in QuantML and ISO-TimeML. In QuantML, a sentence with negative polarity is viewed as expressing that a certain event (with the specified participants and given time and place) does/did/will not occur. This is interpreted in the semantics by a negation operator applied to the entire structure that denotes that event. ISO-TimeML, by contrast, represents negative polarity by the attribute @polarity in an <event> element with value "negative". This comes down to allowing negated events which may be combined with information about participants, time and location. Whether these two views are technically compatible needs to be explored further.

  Temporal quantification is treated in QuantML in the same way as quantification over other domains, whereas ISO-TimeML treats temporal quantification as well as frequency as properties of temporal objects, using the @quant and @freq attributes in <timex3> elements. Bunt and Pustejovsky (2010) noted already that this treatment of these phenomena is not entirely satisfactory from a semantic point of view. Given the annotation schemes as they developed since then, another close look should be taken on the two approaches and how they might be combined.

A.8 References

Harry Bunt and James Pustejovsky (2010) Annotating temporal and event quantification. In *Proceedings* 5^{th} *Joint ACL - ISO Workshop on Interoperable Semantic Annotation (ISA-5)*t, Hong Kong, 15-22.

Harry Bunt, Volha Petukhova, Andrei Malchanau, Alex Fang and Kars Wijnhoven (2019) The DialogBank: Dialogues with Interoperable Annotations. *Language Resources and Evaluation* 52:213-249. Available online at DOI: 10.1007/s10579-018-9436-9.

Patrice Lopez and Laurent Romary (2010) A Framework for Multilevel Linguistic Annotations. Unpublished paper, available online at https://www.mpi.nl/ISLE/documents/papers/lopez/article.html

A.9 Details of ISO-TimeML - QuantML interlinking

The following detailed example shows the TEI-based data representation, the derivation of the abstract annotation structures by decoding functions, the internal structure of components of the abstract annotation structures, and the semantic interpretations in more detail.

(18) “More than three thousand students graduated on Friday"

XML REPRESENTATION:

Data representation, following TEI guidelines (`https://www.tei-org/release/doc/tei-5-doc/en/html/ref-w.html`)

```
<w xml:id="w0">more/>
<w xml:id="w1">than/>
<w xml:id="w2">three/>
<w xml:id="w3">thousand/>
<w xml:id="w4">students/>
<w xml:id="w5">graduated/>
<w xml:id="w6">on/>
<w xml:id="w7">friday/>
```

QuantML part (‘X_Q’)

Markables: mQ1=“#w0 #w1 #w2 #w3", mQ2=“#w0 #w1 #w2 #w3 #w4", mQ3=“#w4", mQ4=“#w5", mQ5=“#w7"

```
<entity xml:id="xQ1" target="#mQ2" refDomain="#xQ1a" involvement="#n1"
      individuation="count"/>
   <refDomain xml:id="xQ1a" target="#mQ3" pred="student" determinacy="indet"/>
   <cardinality xml:id="n1" target="#mQ1" numRel= "greater" num="3000"/>
<event xml:id="eQ1" target="#mQ4" pred="graduate"/>
<participation event="#eQ1" participant="#xQ1" distr=“individual" semRole="agent" evScope="narrow"/>
<entity xml:id="xQ2" target="#mQ5" refDomain="#xQ2a" individuation="count"
      involvement="some"/>
   <refDomain xml:id="xQ2a" target="#m8" pred="friday" determinacy="indet"/>
<participation event="#eQ1" participant="#xQ2" distr=“individual" semRole="time" evScope="wide"/>
<scoping arg1="#xQ1" arg2="#xQ2" scopeRel="wider"/>
```

ISO-TimeML part (‘X_T’)

Markables: mT1=“#w5 ", mT2=“#w6", mT3=“#w7"

```
<event xml:id="eT1" target="#mT1 pred="graduate" type="transition" class="ocurrence" pos="verb"
     tense="past" aspect="none"/>
<signal xml:id="s1" target="#mT2" pred="on"/>
<timex3 xml:id="xT1" target="#mT3" pred="friday" type="set" value="XXXX-WXX-5"/>
<tLink signalID="#s1" eventID="#eT1" relatedToTime="#xT1" relType="isIncluded"/>
```

Interlinking part (‘X_{IL}’)

```
<idLink arg1="#eQ1" arg2="#eT1"/>
<idLink arg1="#xQ2" arg2="#xT1"/>
```

ABSTRACT SYNTAX:

QuantML:

A$_Q$ = dF_Q(X$_Q$) = $\langle \epsilon_{Q1}, \{\xi_{Q1}, \xi_{Q2}\}, \{pL1, pL2\}, \{\langle pL1, pL2,$wider$\rangle\}\rangle$, where
ϵ_{Q1} = dF_Q(<event xml:id="eQ1" target="#mQ4" pred="graduate"/>) = ⟨mQ4, graduate⟩
ξ_{Q1} = dF_Q(<entity xml:id="xQ1" target="#mQ2" refDomain="#xQ1a">
involvement="#n1"individuation="count"/>
<refDomain xml:id="xQ1a" target="#mQ3" pred="student"
determinacy="indet"/>
<cardinality xml:id="n1" target="#mQ1" numRel= "greater"
num="3000"/>) =
= ⟨mQ2, ⟨⟨mQ3, ⟨student, indeterminate⟩⟩, count ⟨mQ1, ⟨greater, 3000⟩⟩⟩⟩
ξ_{Q2} is derived similarly;
$pL1$ = dF_{QT}(<participation event="#eQ1" participant="#xQ1" distr="individual"
semRole="agent"/>) =
= $\langle \epsilon Q1, \xi Q2,$time, individual, narrow$\rangle$. $pL2$ = $\langle \epsilon Q1, \xi Q2,$time, individual, narrow$\rangle$.

ISO-TimeML:

A$_T$ = dF_T(X$_T$) = $\langle \epsilon T1, \{\xi T1\}, \{tL1\}, \{\}\rangle$, where
ϵ_{T1} = dF_T(<event xml:id="eQ1" target="#mQ4" pred="graduate" class= 'occurrence"
type="transition" pos="verb" tense="past" aspect="none" polarity="positive"/>)
= ⟨mT1, ⟨ graduate, ocurrence, transition, past, positive⟩⟩
ξ_{T1} = dF_T(<timex3 xml:id="xT1" target="#mT3" pred="friday" type="set"
value="XXXX-WXX-5"/>)=
= ⟨mT3, ⟨friday, set, XXXX-WXX-5⟩⟩
$tL1$ = $\langle \epsilon_{T1}, \xi_{T1},$ is-included ⟩

Interlinked structure :

A_{QT} = $dF_{QT}(X_Q + X_T + X_{IL})$ = $\langle \epsilon_{QT}, E_{QT}, L_{QT}, sc_{QT}\rangle$,
a quadruple consisting of an interlinked event structure, a set of interlinked entity structures, a set of event-entity link structures, and a set of scope relations.
These components are computed as follows (see (6)):

ϵ_{QT} = $\langle (dF_Q(X_Q))_1, (dF_T(X_T))_1\rangle$ = $\langle e_Q, e_T\rangle$
E_{QT} = $(dF_Q(X_Q))_2 \cup (dF_T(X_T))_2 \cup \{\langle x, y\rangle|$ interlinked(x,y)} = {$\xi_Q 1, \langle \xi Q2, \xi T1\rangle$}
L_{QT} = $\{pL1, \langle pL2, tL1\rangle\}$
sc_{QT} = {$\langle pL1, pL2, wider\rangle, \langle pL2, tL1,$equal$\rangle$}

SEMANTICS:

QuantML:

Note that the noun “Friday", although looking like a proper name, in the reading that more than three thousand graduated on *a Friday* is a common noun, denotng the set of all those days which are called *Friday*, rather than referring to a single entity with that name.
Moreover, the NP “Friday" is a generalized (existential) quantifier.

$I_Q(A_Q)$ = $I_Q(\langle eQ1, \{xQ1, xQ2\}, \{pL1, pL2\}, \{\langle pL1, pL2$,wider$\rangle\}\rangle)$ = $I_Q(pL_1) \cup^* I_Q(pL_2)$
= $I_Q(eQ1), I_Q(xQ1), \langle I_Q$(agent, individual$\rangle) \cup^*$
$I_Q(eQ1), I_Q(xQ2, \langle I_Q$(time, individual$\rangle)$
= [X $\subseteq$ student | |X| $>$ 3000, $x \in$ X $\rightarrow$ [E $\subseteq$ graduate | e $\in$ E $\rightarrow$ agent(e,x)]] $\cup^*$
[Y $\subseteq$ friday| y $\in$ Y $\rightarrow$ [E $\subseteq$ graduate | e $\in$ E $\rightarrow$ time(e,y)]]
= [X$\subseteq$ student | |X| $>$3000, $x \in$ X $\rightarrow$ [Y $\subseteq$ friday| y $\in$ Y $\rightarrow$
[E $\subseteq$ graduate | e $\in$ E $\rightarrow$ [agent(e,x), time(e,y)]]]]

ISO-TimeML:

$I_T(\xi_{T1})$ = [Y | y $\in$ Y $\rightarrow$ [friday(y), code(y) = XXXX-WXX-5]
$I_T(A_T)$ = $I_T(tL1)$ = $I_T(\epsilon_{T1}, \xi_{T1}$ is-included$\rangle$
= [Y $\subseteq$ friday | y $\in$ Y $\rightarrow$
[E $\subseteq$ graduate [e $\in$ E $\rightarrow$
[past(e), class(e)=occurrence, type(e)=transition,
code(y)="XXXX-WXX-5", includes(y,x)]]]

Interlinked interpretation:

Generalizing the QuantML interpretation rules (B34) and (B35), specified in ISO DIS 24617-12, the link structures $pL1, pL2$, and $tL1$ are semantically combined as determined by the scope relations $\langle pL1, pL2$,wider$\rangle$ and $\langle pL2, tL1, equal\rangle$ using the ordinary merge ($\cup$) and scoped merge ($\cup^*$) operations. (See Bunt, H., (2023). The compositional semantics of QuantML annotations. *Proceedings 19th Joint ACL - ISO Workshop on Semantic Annotation (ISA-19)*, Nancy, France, pp. 3-13.)

$I_{QT}(A_{QT})$ = $I_{QT}(\langle\langle pL1, pL2$,wider$\rangle, \langle pL2, tL1, equal\rangle\rangle)$
= $I_{QT}(pL1)) \cup^* I_{QT}(pL2, tL1,$ equal)
= $I_{QT}(pL1) \cup^* (I_{QT}(pL2) \cup I_{QT}(tL1))$
= [X $\subseteq$ student | |X| > 3000, x $\in$ X $\rightarrow$
[Y $\subseteq$ friday | y $\in$ Y $\rightarrow$
[E $\subseteq$ graduate | e $\in$ E $\rightarrow$
[past(e), class(e)=occurrence, type(e)=transition,
code(y)=XXXX-WXX-5", includes(y,x)]]]]