# The compositional semantics of QuantML annotations

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

## Abstract

This paper discusses some issues in the semantic annotation of quantification phenomena in general, and in particular in the markup language QuantML, which has been proposed to form part of an ISO standard annotation scheme for quantification in natural language data. QuantML annotations have been claimed to have a compositional semantic interpretation, but the formal specification of QuantML in the official ISO documentation does not provide sufficient detail to judge this. This paper aims to fill this gap.

## 1 Introduction

The semantic annotation of quantification in natural language aims to enrich language data with information about the intended interpretation of the quantifications. The formulation of such annotations and their assignment to the data are challenging tasks, in view of the complexity of quantification phenomena in natural language. The many aspects of quantification, such as the distributivity, determinacy, countability, exhaustiveness, polarity and scope, make quantifications a major source of ambiguity and difficulty in computational semantics.

One of the challenges that quantifications pose for semantic annotation and representation is that, although much of the information about quantification is located in noun phrases, there may also be quantification information floating around in the form of adverbials, or encoded in language-specific morphosyntactic structure, or expressed by prosodic features (stress, pauses) (*"You heard a dog barking?" "TWO dogs barked."*) or typographical elements (use of capitals, underlining, punctuation). Semantic annotation faces the challenge of picking up these various pieces of information and assembling them in a useful form. The QuantML language is a proposal for such a form.

Another, fundamental challenge for quantification analysis concerns the choice of depth and detail, or 'granularity'. Studies of quantification phenomena in natural language have both benefited and suffered from studies of quantification in logic (Aristotle, 350 BC; Montague, 1971). The benefits are in the deep understanding of formal properties of and fine-grained distinctions between various types of quantification, which have contributed greatly to the emergence of the theory of generalized quantifiers (GQT, Barwise & Cooper, 1981). On the negative side, logic-based approaches tend to have weaknesses from an empirical linguistic point of view, since the fine-grained distinctions that can be expressed in formal logic tend to carry over to semantic representations of natural language expressions, while speakers and listeners are often unaware of these fine distinctions, thus creating a sort of artificial ambiguities. At this point semantic annotations may come in useful. A semantic annotation can be regarded as expressing constraints on the meaning of certain language data, without having the ambition of providing full-blown semantic representations, viz. to express 'the meaning' of the data. Annotations, by contrast, may express fewer or more, weaker or stronger constraints on interpretation. Still, in a context where high-precision interpretations are required, a semantic annotation scheme should allow detailed information to be captured by the annotations.

The present stage of defining an ISO standard annotation scheme for quantification, involving QuantML is documented in ISO CD 24617-12 (2023)(see Bunt et al. 2022). It follows the general principles for semantic annotation laid down in ISO 24617-5, Principles of Semantic Annotation (2015)(see also Bunt, 2016). One of these principles is that semantic annotations must have a well-defined semantics. In view of the challenges mentioned above, this means for QuantML the requirement to be flexible in the level of detail of its expressions, while having a semantics for its annotation structures, regardless the level of detail. The specification of QuantML in the ISO document goes a long way in this direction, outlining a compositional semantics of annotation structures, but this is not fully worked out for some of its structures. The present paper aims to remedy this.

The paper is organised as follows. Section 2 summarises the approach taken in developing QuantML. Section 3 discusses the annotation of scope relations, distinguishing wider, equal, and dual scoping. The semantics of these relations is considered, and their role in combining information from pairs of quantifications. Section 3 generalizes the semantic interpretation of annotations of multiple scoped quantifiers. Section 4 briefly discusses the instruments available in QuantML for varying the level of granularity in annotations. Section 5 wraps up and closes the paper.

## 2 ISO 24617-12 and QuantML

Following the ISO principles of semantic annotation (ISO 24617-5, 2015), QuantML has a triple-layered definition, based on a metamodel. The three layers are (a) an abstract syntax, using $n$-tuples of concepts; (b) a reference representation format based on XML, with encoding- and decoding mappings to the abstract syntax; and (c) a semantics, in the form of a function $I_Q$ which translates abstract annotation structures into DRSs in a compositional way. The fact that the semantics is defined for the *abstract* syntax makes it possible to accommodate alternative representation formats, while preserving the meaning. An example of the abstract and concrete syntax of a QuantML annotation is given in (1).

(1) At least three students called more than once.

a. Markables: $\mathrm{m}_1$ = “At least three”, $\mathrm{m}_2$ = “At least three students”, $\mathrm{m}_3$ = “students”, $\mathrm{m}_4$ = “called $\mathrm{m}_4$ = “more than once”, $\mathrm{m}_5$ = “more than once”

b. Abstract syntax:
$L_1$ = $\langle \epsilon_e, \epsilon_{P1}$,Agent, individual, narrow$\rangle$, where:
$\epsilon_e$ = $\langle$$\mathrm{m}_4$, $\langle$call, $\langle >, 1\rangle\rangle\rangle$, and
$\epsilon_{P1}$ = $\langle$$\mathrm{m}_2$, $\langle$student, indeterminate$\rangle$, count, $\langle \geq, 3\rangle\rangle$

c. Concrete syntax:
<entity xml:id=“x1 target=“#m2” domain=“#x2” involvement=“n1” definiteness=indet/>
<refDomain xml:id=“x2” target=“#m3” source=“#x3”/>
<cardinality xml:id=“n1” target=“m1” numRel=“greaterthan” num=“2”>
<sourceDomain xml:id=“x3” target=“#m3” individuation=“count” pred=“student”/>
<event xml:id=“e1” target=“#m4” pred=“call” rep=“#n2”>
<cardinality xml:id=“n2” target=“m1” numRel=“greaterthan” num=“1”>
<participation event=“#e1” participant=“#x1” semRole=“agent” distr=“individual” evScope =“narrow” />

Theoretically, QuantML is inspired mainly by GQT, by event semantics (Davidson, 1967; Parsons, 1990), and by Discourse Representation Theory (DRT, Kamp & Reyle 1993). Quantifiers are thus interpreted as properties of sets of individuals, typically expressed by noun phrases, which play certain semantic roles as participants in sets of events. This is reflected in the metamodel in Fig. 1, where participant sets, event sets, and the relation between them play center stage. The use of DRT is primarily motivated by the consideration that other parts of the ISO Semantic Annotation Framework also make use of DRT; otherwise, second-order logic would be equally well suitable.

## 3 Annotation Semantics

### 3.1 Basic Concepts and Metamodel

The semantic information that is captured by QuantML annotations is concentrated primarily in the specification of participant sets and their relation to event sets through participation links. The annotation describing a participant set contains local semantic information about the entities that populate the set; a participation link structure specifies properties of the relation between a participant set and the events in which they are involved. This includes information about the relative scopes of the quantification over participants and the quantification over events (the ‘event scope’).

Semantic information which is less local in character concerns the relative scoping of quantifications over participant sets involved with different semantic roles. The main challenge of a compositional interpretation of annotation structures is to combine the local semantic information in the participant sets and the participation link structures into a single semantic representation. The relative scoping of quantifiers has been studied extensively in formal logic and in formal and computational semantics in terms of wide and narrow scope (e.g. Hobbs & Shieber, 1987; Montague, 1974; Kamp & Reyle, 1983: Szabolcsi, 2010), mostly for count nouns and for distributive readings. With these limitations, the semantics of quantification annotations would be fairly straightforward.

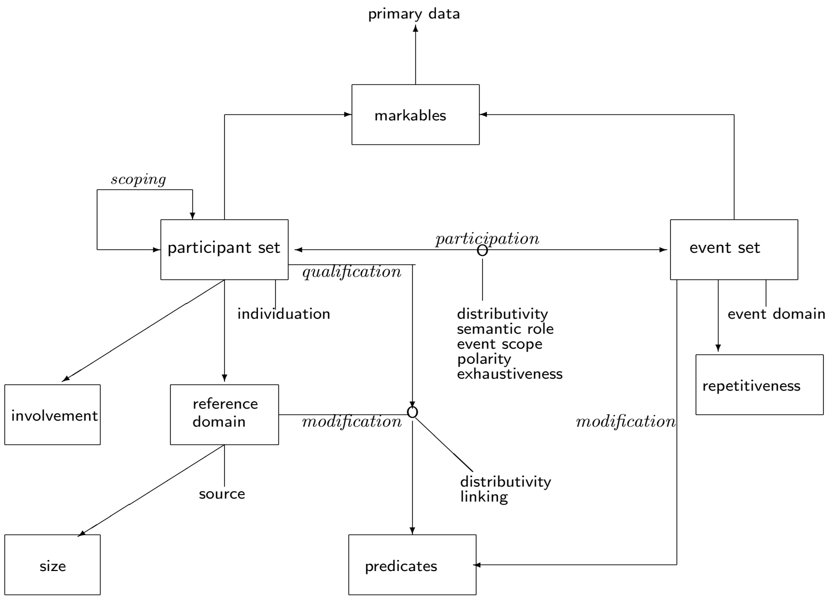


Figure 1: QuantML metamodel.

Besides distributive readings, also collective readings display scope ambiguities. A set of participants that is involved collectively in certain events might appear to be acting as a single entity, and the notion of scope would therefore not apply. However, consider the sentences in (2):

(2) a. The two men moved all the pianos.
All the pianos were moved by the two men.
b. All the pianos were moved by two men.
c. Two men moved all the pianos.

Both sentences in (a) can only be read as saying that the same two men moved all the pianos, giving the quantification over men wider scope than the one over pianos. Sentence (b), by contrast, has the preferred reading where the various pianos were moved by pairs of men, but not necessarily all by the same pair, and thus having the inverse scoping. Sentence (2c) has both readings. So clearly, issues of scope do apply also in the case of quantifications with collective distributivity.

### 3.2 Abstract syntax and semantics

Since the semantic contributions of the participant and event structures are included in the participation link structure, the semantic interpretation of the annotation of the sentence is just the interpretation of the participation link structures. More generally, the abstract syntax of the QuantML annotation for a sentence with two or more quantifiers (as expressed by NPs) is the collection of participation link structures plus the collection of scope relations between them, and the semantic interpretation of the annotation structure is obtained by combining the semantics of the individual participation structures in a way determined by the scope relations.

## 4 Scope relations in QuantML

### 4.1 Scoping and 'plint structures'

In QuantML three scope relations among participation link structures are distinguished:

1. Wider: one quantification outscopes the other. Example: "Every student speaks two languages" (but not necessarily the same two). The DRSs representing the semantics of the participation links are combined by means of an operation called *'scoped merge'*.

2. Dual: two quantifications mutually outscope each other (so-called 'cumulative' quantification). Example: "Three breweries supplied more than 5000 inns". The corresponding DRSs are combined by means of an operation called *'dual-scoped merge'*.

3. Equal: two quantifications have equal scope (so-called 'cluster' or 'group' quantification). Example: "Seven boys played against eleven girls", in the sense of teams of seven boys playing against teams of eleven girls. The corresponding DRSs are combined by means of the standard DRS-merge.

The semantics of the scope links determines how the participation link interpretation structures, or '*plint structures*', should be combined. This is expressed in (3).

(3) For any scope relation $s$, if $L'_i =_D I_Q(L_i)$:
$I_Q(L_1, L_2, s) = I_Q(s)(L'_1, L'_2)$

The scope relation between two quantifications is semantically interpreted as an operation on two plint structures, using the standard DRS-merge and two scope-dependent forms of merge, called *scoped merge* ($\cup^*$) and *dual-scoped merge* ($\cup^\otimes$). These operations are defined below.

(4) $I_Q$(wider) = $\lambda x.\lambda y.x \cup^* y$;
$I_Q$(equal) = $\lambda x.\lambda y.x \cup y$;
$I_Q$(dual) = $\lambda x.\lambda y.x \cup^\otimes y$

The semantics of each of the scope relations is discussed in subsequent subsections.

## 4.2 Wider scope

The scoped merge operator $\cup^*$ takes two plint structures $L'_1$ and $L'_2$ as arguments and merges them into a single DRS. Since the $L_1$-quantification has scope over the $L_2$-quantification, the DRS that represents the latter quantification is moved into the DRS that represents the $L_1$-quantification, in such a way that it falls within the scope of that quantification. Moreover, since the two plint structures link participant sets to the same set of events, the two event quantifications are collapsed into one. In terms of DRS merging this means that a discourse referent is introduced which refers to the event set, in a position determined by the two event scopes, [1] and the nuclear content of $L'_1$ is added to the nucleus of the $L'_2$ - quantification. This expressed in (5).

(5) **Scoped merge.**
Given two plint structures $L'_1$ and $L'_2$, the scoped merge moves $L_2$' as a sub-DRS into the DRS $L'_1$, bringing the $L'_2$-quantification within the scope of the $L'_1$-quantification and merging the event quantifications.

The formal definition of the scoped merge is formulated in terms of pattern-matching based operations, since the structures that it applies to have certain specific structural properties. Every plint structure contains three parts:

(6)
1. the introduction of a participant set, i.e. a DRS of the form $[X|C_1, x \in X \rightarrow K_1(x)]$, where the discourse referent $X$ refers to the participant set, $C_1$ is a set of conditions, and the sub-DRS $K_1$ represents the quantifying predicate;
2. the introduction of an event set, a DRS of the form $[E|C_e, e \in E \rightarrow K_2(e)]$;
3. the nucleus, a sub-DRS of the form $R_i(e, x)$, where the semantic role $R_i$ relates events and participants.

A plint structure where the second part constitutes the $K_1$ subexpression of the first part represents a quantification with narrow event scope; one where the first part constitutes the $K_2$ subexpression of the second part represents wide event scope. Schematically, these two forms of a plint structure have the top-level structures shown in (7)

(7) a. $[X_i|C_i, x \in X_i \rightarrow K_i(x)]$, with
$K_1 = \lambda z.[E|C_e, e \in E \rightarrow R_i(e, z)]$

b. $[E|C_e, e \in E \rightarrow K_i(e)]$, with
$K_1 = \lambda u.[X_i|C_i, x \in X_i \rightarrow R_i(u, z)]$

Both forms come in two variants, depending on the distributivity of the quantification with individual or unspecific distributivity. In the individual case, the elements of the participant set are involved individually; in the unspecific case also as subsets. This leads to differences in $K_1$ and $K_2$ in (6). In the case of collective quantification we see a first part of the form $[X|C_1, K_1(X)]$ rather than $[X|C_1, x \in X \rightarrow K_1(x)]$, for narrow-scope interpretations and $[X, E|C_1, C_e, e \in E \rightarrow K_2(e, X)]$ in case of wide event scope,. The six possible forms of plint structures for all combinations of event scope and distributivity (and positive polarity and non-exhaustive, see below) are listed in (8).

(8) a. Narrow event scope, individual distributivity: $[X|\mathrm{C}_i,\ x \in X \rightarrow [E|\mathrm{C}_e, e \in E \rightarrow R(e, x)]]$ or, schematically, with $K$ as in (7a): $[X|C, x \in X \rightarrow K(x)]$

b. Wide event scope, individual distributivity: $[E|C_e, e \in E \rightarrow [XC_i, x \in X \rightarrow R(e, x)]]$ or, schematically, with $K$ as in (7b): $[E|C_e, e \in E \rightarrow K(e)]$

c. Narrow event scope, collective distributivity: $[E, X|C, C_e, e \in E \rightarrow R(e, X)]]$ or, schematically, with $K = \lambda z.R(e, z)$: $[E, X|C, C_e, e \in E \rightarrow K(e, X)]$

[1] More precisely, the quantification over events is in the position with most narrow scope which is consistent with the event scopes.

d. Wide event scope, collective distributivity:
$[E|C_e, e \in E \to [X|C, R(e, X)]]$
or, schematically:
$[E|C_e, e \in E \to K(e, X)]$,
with $K = \lambda u.[X|x \in X \to R(u, X)$

e. Narrow event scope, unspecific distributivity, where $X^* =_D X \cup \mathcal{P}(X)$:
$[X|C, x \in X \to [E|C_e, e \in E \to$
$[y \in X^*|x = y \vee x \in y, R_i(e, y)]]]$,
schematically: $[X|C, x \in X \to K(x]$,
$K = \lambda z.[E|C_e, e \in E \to$
$[y| \in X^*, ...R(e, y)$

f. Wide event scope, unspecific distributivity:
$[E|C_e, e \in E \to [X|C, x \in X \to$
$[y \in X^*|x = y \vee x \in y, R(e, y)]]]]$,
schematically, with $K$ similar to case e:
$[E|C_e, e \in E \to K(e)]$.

In sum, plint structures can have the following schematic forms:

(9) a. $[X_i|C_i, x \in X_i \to K_i(x)]$
b. $[E|C_e, e \in E \to K_i(e)]$
c. $[E, X_i|C_i, C_e, e \in E \to K_i(e, X_i)]$

The scoped merge of two plint structures $L'_1$ and $L'_2$, where the first has wider scope than the second, combines the content of the two structures in a way that depends on their schematic forms. This is indicated in Table 1, where the '∪' indicator means that the scoped merge in this case is just the standard DRS-merge; the indicators 'A', 'B', and 'C' are defined in (10).

| | a, e | b, d, f | c |
|---|---|---|---|
| a, e | A | B | B |
| b, f | – | C | – |
| c | – | ∪ | ∪ |

Table 1: Scoped merge as depending on schematic argument structures

(10) Indicators used in Table 1:
∪: $L'_1 \cup L'_2$
A: $[X_1|C_1, x{\in}X_1 \to [X_2|C_2, y \in X_2 \to (K_1(x) \cup K_2(y))]]$
B: $[X_1|C_1, x{\in}X_1 \to (K_1(x) \cup L'_2)]$
C: $[E|C_e, e \;{\in}E \to [X_1|C_1, x \in X_1 \to [X_2|y \in X_2 \to N_1(e, x) \cup N_2(e, y)]]]$,
where $N_i$ is $\lambda z.\lambda u.R_i(u, z)$

Note that Table 1 indicates that the scoped merge is undefined for certain combinations of argument forms. This is because in those cases the relative scopes are inconsistent with the event scopes of the arguments. See Section 5.3. An example of applying the scoped merge is shown in (11).

(11) Some students read more than three papers.

a. Markables: $m_1$ = "Some students", $m_2$ = "students", $m_3$ = "read", $m_4$ = "more than three", $m_5$ = "more than three papers", $m_6$ = "papers"

b. QuantML annotation, XML-based concrete syntax:
<entity xml:id="x1" target="#m2" domain="#x2" involvement="some" definiteness=indet/>
<refDomain xml:id="x2" target="#m3" source="#x3"/>
<sourceDomain xml:id="x3" target="#m3" individuation="count" pred="student"/>
<event xml:id="e1" target="#m4" pred="read">
<participation event="#e1" participant="#x1" semRole="agent" distr="individual" evScope ="narrow" />
<entity xml:id="x4" target="#m5" domain="#x5" involvement="n1" definiteness=indet/>
<refDomain xml:id="x5" target="#m6" source="#x6"/>
<sourceDomain xml:id="x6" target="#m6" Rel="greaterthan" num="3">
<participation event="#e1" participant="#x4" semRole="theme" distr="individual" evScope ="narrow" />
<scoping arg1="#x1" arg2='#x4" scopeRel ="wider" />

c. QuantML annotation, abstract syntax:
$L_1$ = $\langle \epsilon_e, \epsilon_{P1}$,Agent, individual, narrow$\rangle$,
$L_2$ = $\langle \epsilon_e, \epsilon_{P2}$Theme, individual, narrow$\rangle$, where
$\epsilon_e$ = $\langle$m$_4$, $\langle$read$\rangle\rangle$,
$\epsilon_{P1}$=$\langle$m$_2$, $\langle$student, indeterminate$\rangle$, count, some $\rangle$,
$\epsilon_{P2}$ = $\langle$m$_2$, $\langle$paper, indeterminate$\rangle$, count, $\langle \geq, 4\rangle\rangle$
Scoping;
$sc_1$ = $\langle L'_1, L'_2$, wider $\rangle$

d. Semantics:
$L'_1 = [X_1|X_1 \subseteq \text{student}, x \in X_1 \to [E|E \subseteq \text{read}]$
$e \in E \to \text{agent}(e, x)]]$,
$L'_2 = [X_2|X_2 \subseteq \text{paper}, X_2| > 3, y \in X_2 \to$
$[E|E \subseteq \text{read}, e \in E \to \text{theme}(e, y)]]$,
$L'_1 \cup^* L'_2 = [X_1|X_1 \subseteq \text{student}, x \in X_1 \to$
$[X_2|X_2 \subseteq \text{paper}, y \in X_2 \to$
$[E|E \subseteq \text{read} \mid e \in E \to$
$[\ \text{agent}(e, x), \text{theme}(e, y)]]]$

In addition to the possible forms of the DRSs that interpret a participation structure with positive polarity, listed in (8), slightly different forms represent the semantics of negative-polarity quantifications. A participation link structure with wide-scope negative polarity corresponds to one the plint structures of (8) with an additional top-level negation; one with narrow-scope negative polarity and narrow event scope (cases (8a) and (8f)) have a negated sub-DRS that introduces the event set, which does not alter the schematic structure.

The scoped merge is defined only for two plint structures with the same polarity, with the following effects if both arguments have negative polarity.

(12) a. If both arguments have wide-scope negative polarity, then their scoped merge is as defined in (10), with the resulting DRS being negated.

b. If both arguments have narrow-scope negative polarity, then their scoped merge is exactly as defined in (10), since the negations are incorporated in sub-DRSs of the two arguments that represent the quantification over the event set.

Another complication for plint structures is due to the possible internal complexity of a participant set specification. As the metamodel in Fig. 1 indicates, a participant set may have 'qualifications', i.e. one or more specifications of non-restrictive modifications ('appositives'); moreover, the reference domain of which it is a subset may have an (absolute or relative) size specification, and may be co-determined by restrictive modifications (which come with their own distributivity and scope linking).

The plint structures listed in (8) all introduce a discourse referent used to refer to a participant set (indicated by '$X_i$') and include a set of conditions '$C_i$' that contains a restriction like $x \in X_i \rightarrow$ student$(x)$, stipulating that the participant set is a set of students. This is adequate for simple quantifiers like *"some students"* and *"five students"*, but it is not expressive enough for quantifiers like *"three of the four eggs"*, in example (13). The cardinal determiner *"three"* in this example indicates the size of the participant set ('involvement' in Fig. 1), while *"four"* indicates the size of the reference domain. To accommodate this, a second discourse referent is introduced that refers to the reference domain, indicated in (13) by$X'$, where the indexed predicate 'egg$_0$' is used to indicate the predicate 'egg' (denoting the source domain of all eggs) restricted to its contextually relevant subset.

(13) a. Three of the four eggs have hatched.

b. $[X, X'|C_1, x \in X \rightarrow [E|e \in E \rightarrow$
$[\,\text{hatch}(e), \text{theme}(e,x)\,]]]$, with
$C_1 = \{|X| = 3, |X'| = 4, X \subseteq X',$
$y \in X' \leftrightarrow \text{egg}_0(y)\}$

This addition does not alter the schematic form of the plint structure, apart from the introduction of a second discourse referent. This additional element does not play an active role in the scoped merge; it is merely dragged along when plint structures are combined. This possible complication is therefore disregarded in the rest of this paper.

### 4.3 Dual scope

The 'dual' scope relation is used in QuantML for the annotation of cases of cumulative quantification. Cumulative quantification may occur in sentences with two numerical determiners (Krifka, 1999) as in the most plausible reading of example sentence (14), due to Reyle (1983).

(14) Three breweries supplied twelvehundred inns.

In the cumulative interpretation, none of the two quantifiers has wider scope than the other; rather, it says that each one of a set of three breweries supplied some of 1200 inns, and vice versa. In QuantML, this is analysed as mutual outscoping: the quantification over breweries has wider scope than the one over inns, and vice versa.

The semantics of a dual-scope relation involves the use of an operation similar to the scoped merge operation, called *dual-scoped merge* and symbolised by $\cup^{\otimes}$. The operation is used for combining two plint structures for non-collective quantification with narrow event scope and positive polarity. Quantifications with collective distributivity, wide event scope, or negative polarity do not allow cumulative interpretations, hence only plint structures of the form (8) (a) or (e) are involved. The dual-scoped merge is defined as follows.

(15) **Dual-scoped merge.**
The dual-scoped merge combines two plint structures $L_1', L_2'$ into a DRS that inherits the discourse referents of both arguments, and branches out into two sub-DRSs, corresponding to the two sides of mutual outscoping, which both have the merge of the $L_1'$ and $L_2'$ nuclei as their nucleus.

To express this in formal terms, note that the operation is defined as applicable only to plint structures of the form of (8a) or (8e), which both have the schematic form $[X_i|C_i, x \in X_i \rightarrow K_i(x)]$ (see (9a)). Applying the dual-scoped merge to two arguments of this form is the following operation on plint structures:

(16) $L_1' \cup^{\otimes} L_2' = [X_1, X_2|C_1, C_2,$
$x_1 \in X_1 \rightarrow [x_2|x_2 \in X_2, K_1 \cup K_2],$
$x_2 \in X_2 \rightarrow [x_1|x_1 \in X_1, K_1 \cup K_2]]$

As an example, consider the sentence in (14). The abstract syntax of the QuantML annotation of this sentence would include two plint structures of the same form as those in (11). Application of the dual-scoped merge gives the following result:

(17) $[X_1, X_2 | X_1 \subseteq \text{brewery}, |X_1| = 3,$
$x_1 \in X_1 \rightarrow [x_2 \in X_2 \rightarrow E \subseteq \text{supply}|$
$[\text{agent}(e, x_1), \text{beneficiary}(e, x_2)]],$
$x_2 \in X_2 \rightarrow [x_1 \in X_1 \rightarrow E \subseteq \text{supply}|$
$[\text{agent}(e, x_1), \text{beneficiary}(e, x_2)]]]$

### 4.4 Equal scope

The equal scope relation is used specifically for cases of cluster quantification (or 'group quantification'), as mentioned in Section 4.1. The semantics of an "equal" scope annotation is defined through application of the standard DRS-merge. For example, the QuantML annotation of the sentence (18a) is as follows.

(18) Seven boys played against eleven girls.

a. Markables: $m_1$ = "Seven boys", $m_2$ = "boy", $m_3$ = "played against", $m_4$ = "eleven girls", $m_5$ = "girls"

b. QuantML annotation, XML-based concrete syntax:
<entity xml:id="x1" target="#m1" indiv="count", domain="#x2",
involvement="7" determinacy=indet/>
<refDomain xml:id="x2" target="#m3"
source="#x3"/>
<sourceDomain xml:id="x3" target="#m2" pred="boy"
<event xml:id="e1" target="#m4" pred="play">
<participation event="#e1" participant="#x1" semRole="agent" distr="individual"
evScope ="wide" />
<entity xml:id="x4" target="#m4" domain="#x5"
involvement="11" determinacy=indet/>
<refDomain xml:id="x5" target="#m6"
source="#x6"/>
<sourceDomain xml:id="x6" target="#m5"
<participation event="#e1" participant="#x4" semRole="theme" distr="individual"
evScope ="idew" />
<scoping arg1="#x1" arg2='#x4"
scopeRel ="equal" />

c. QuantML annotation, abstract syntax:
$L_1 = \langle \epsilon_e, \epsilon_{P1}$,Agent, individual, wide$\rangle$,
$L_2 = \langle \epsilon_e, \epsilon_{P2}$ Agent, individual, wide$\rangle$, where
$\epsilon_e = \langle m_4, \langle \text{play} \rangle \rangle$,
$\epsilon_{P1}$=$\langle m_2$, $\langle$boy, 7, indeterminate$\rangle$, count, some $\rangle$,
$\epsilon_{P2}$ = $\langle m_2$, $\langle$girl,11, indeterminate$\rangle$, count, $\langle \geq, 4 \rangle \rangle$
Scoping;
$sc_1 = \langle L'_1, L'_2$, equal $\rangle$

d. Semantics:
$L'_1 = [E|E \subseteq \text{play}, e \in E \rightarrow [X|X \subseteq \text{boy},$
$|X| = 7, \text{agent}(e, xX]]$
$L'_2 = [E|E \subseteq \text{play}, e \in E \rightarrow [Y|Y \subseteq \text{girl},$
$|Y| = 11, \text{agent}(e, Y)]],$
$L'_1 \cup L'_2 = [E \subseteq \text{play} \mid e \in E \rightarrow [X|X \subseteq \text{boy},$
$Y \subseteq \text{girl}, |X| = 7|, |Y| = 11,$
$[\text{agent}(e, X), \text{agent}(e, Y)]]]$

## 5 Clause-level annotation structures

### 5.1 Scoping multiple quantifiers

The semantics of the QuantML annotation of a clause with two scoped quantifications is defined by (3) and (4) plus the definitions of the scoped merge and the dual-scoped merge. For clauses with more than two scoped quantifications, the definitions of the scoped merge and the dual-scoped merge can be generalized so as to apply to more than two plint structures as arguments, or so as to apply to two arguments one of which is a plint structure and the other one a plint structure or the result of combining two ore more plint structures. In this section we take the latter approach, thus keeping all scope relations and merge operations binary.

The abstract syntax of a fully scoped clause annotation includes a number of binary scope relations of the form $\langle L_i, L_j, R \rangle$, where $R \in \{$wider, dual, equal$\}$. For example, if $L_1$, $L_2$, and $L_3$ are three participation links, of which $L_1$ has wider scope than $L_2$, while $L_2$ and $L_3$ have dual scope, then the semantics of their combination can be computed in two ways, shown in (19).

(19) a. $L'_1 \cup^* (L'_2 \cup^{\otimes} L'_3)$
b. $(L'_1 \cup^* L'_2) \cup^{\otimes} L'_3)$

More generally, for a clause annotation which contains $n$ participation links, $n-2$ of the links are involved in two scope relations, like $L_1 - L_2$ and $L_2 - L_3$ in the case of example (19). These links define a linked chain like $L_1 - L_2 - L_3$, of which the begin-and end points are the two links that are involved in only one scope relation. Following the approach of (19a), if $\sigma_{i,j}$ designates the scoping relation between $L_i$ and $L_j$, and $\sigma'_{i,j} = I_Q(\sigma_{i,j})$, the interpretation of such a chain is defined by (20).

(20) $I_Q([L_1, L_2, .., L_n] =$
$L'_1 \, \sigma'_{1,2} \, (L'_2 \, \sigma'_{2,3} \, ... \, (L_{n-1} \, \sigma'_{n-1,n} \, L'_n)..)$

### 5.2 Generalized scoped merge

To implement the semantic interpretation of linked chains of scoped participation links, we generalise the scoped merge and the dual-scoped merge operations to apply to two arguments, the first of which is a plint structure and the second either a plint structure or a DRS constructed by applying one of the merge operations defined above or the standard DRS-merge. This comes down to allowing the second argument to be a DRS which has a sub-DRS that expresses a quantification over the same set of events as in the first argument, since both arguments are concerned with participation in the same set of events. The two event quantifications are merged into one, in order to take that into account.

(21) **Generalized scoped merge.**
Given a plint structure $L_1'$ and a DRS $A_2$ which contains a sub-DRS expressing a quantification over the same events as in $L_1'$, the generalised scoped merge inserts the DRS $A_2$ into $L_1'$ immediately below the top level and merges the two event quantifications.

Example:

(22) Both candidates presented a view to the committee members.

a. Markables: $m_1$ = "All candidates", $m_2$ = "candidate", $m_3$ = "presented", $m_4$ = "a vision", $m_5$ = "a vision", $m_6$ = "vision", $m_7$ = "the committee members", $m_8$ = "committee members"

b. QuantML annotation, XML-based concrete syntax:
<entity xml:id="x1" target="#m2" domain="#x2" involvement="all" definiteness=det/>
<refDomain xml:id="x2" target="#m2" source="#x3"/>
<sourceDomain xml:id="x3" target="#m2" individuation="count" pred="candidate"/>
<event xml:id="e1" target="#m3" pred="present">
<participation event="#e1" participant="#x1" semRole="agent" distr="individual" evScope ="narrow" />
<entity xml:id="x4" target="#m4" domain="#x5" involvement="some" definiteness=indet/>
<refDomain xml:id="x5" target="#m6" source="#x6"/>
<sourceDomain xml:id="x6" target="#m6" individuation="count" pred="vision"/>
<participation event="#e1" participant="#x4" semRole="theme" distr="individual" evScope ="narrow" />
<entity xml:id="x7" target="#m7" domain="#x8" involvement="all" definiteness=det/>
<refDomain xml:id="x8" target="#m8" source="#x3"/>
<sourceDomain xml:id="x9" target="#m8" individuation="count" pred="committee-members"/>
<scoping arg1="#x1" arg2='#x4" scopeRel ="wider" />
<scoping arg1="#x4" arg2='#x7" scopeRel ="wider" />

c. QuantML annotation, abstract syntax:
$L_1$ = $\langle \epsilon_e, \epsilon_{P1}$,Agent, individual, narrow$\rangle$,
$L_2$ = $\langle \epsilon_e, \epsilon_{P2}$Theme, individual, narrow$\rangle$,
$L_1$ = $\langle \epsilon_e, \epsilon_{P1}$,Beneficiary, individual, narrow$\rangle$,
where
$\epsilon_e$ = $\langle$m$_4$, $\langle$present$\rangle\rangle$,
$\epsilon_{P1}$=$\langle$m$_2$, $\langle$candidate, determinate $\rangle$, count, all $\rangle$,
$\epsilon_{P2}$ = $\langle$m$_2$, $\langle$vision, indeterminate$\rangle$, count, some$\rangle$
$\epsilon_{P3}$ = $\langle$m$_2$, $\langle$commember, determinate$\rangle$, count, all $\rangle$,
Scoping;
$sc_1$ = $\langle L_1', L_2'$, wider $\rangle$, $sc_2$ = $\langle L_2', L_3'$, wider $\rangle$

d. Semantics:
$L_1' = [X_1 \subseteq \text{candidate}_0 \mid \text{candidate}_0 \subseteq X_1,$
$x \in X_1 \rightarrow [E \subseteq \text{present} \mid e \in E \rightarrow$
$\text{agent}(e, x)]],$
$L_2' = [X_2 \subseteq \text{vision} | y \in X_2 \rightarrow$
$[E \subseteq \text{present} \mid e \in E \rightarrow \text{theme}(e, y)]],$
$L_3' = [X_3 \subseteq \text{commember}_0 \mid \text{commember}_0 \subseteq X_3,$
$z \in X_3 \rightarrow [E \subseteq \text{present} \mid e \in E \rightarrow$
$\text{beneficiary}(e, z)]]$
$L_1' \cup^* (L_2' \cup^* L_3')$ =
$[X_1 \subseteq \text{candidate}_0 \mid \text{candidate}_0 \subseteq X_1,$
$x \in X_1 \rightarrow [X_2 \subseteq \text{vision}, y \in X_2 \rightarrow$
$[X_2 \subseteq \text{vision}, y \in X_2 \rightarrow$
$[X_3 \subseteq \text{commembers}, y \in X_2 \rightarrow$
$[E \subseteq \text{present} \mid e \in E \rightarrow$
$[\text{agent}(e, x), \text{theme}(e, y),$
$\text{beneficiary}(e, z)]]]]$

### 5.3 Generalized dual-scoped merge

The dual-scope merge can be generalized in a similar way. With the generalized scoped merge and dual-scoped merge (and the standard DRS-merge) we can compute the compositional semantic interpretation of any fully scoped collection of participation links, using (20) with $\sigma_{i,j} \in \{\cup^*, \cup^{\otimes}, U\}$. However, cumulative quantification, for which the dual-scoped merge is used, does not seem to make sense in combination with collective distributivity, wide event scope or negative polarity. The definition below therefore restricts its arguments to represent quantification annotations with individual or unspecific distribution, narrow event scope, and positive polariy.

(23) **Generalized dual-scoped merge.**
Given a plint structure $L_1'$ for non-collective distributivity and narrow event scope and a DRS $K$ that contains a sub-DRS expressing a quantification over the same events as in $L_1'$, a DRS is formed that inherits the discourse referents of both arguments and branches out just below the top level into two sub-DRSs, corresponding to either of the two argument scopings, and in both of which the two event quantifications are merged.

A representative example of the use of the generalized dual-scoped merge is shown in (24).

(24) Each of these breweries sold over six hundred thousand casks of beer to five hundred inns.

$L_1' \cup^* (L_2' \cup^{\otimes} L_3')$ =
$[X_1 \subseteq \text{brewery}_0, x \in X_1 \rightarrow$
$[X_2 \subseteq \text{cask}, X_3 \subseteq \text{inn} \mid y \in X_2 \rightarrow$
$[z \in X_3, E \subseteq \text{sell} \mid$

### 5.4 Event scope and participant scoping

Event scope, annotated in participation link structures, interacts with relative participant scoping; some combinations are inconsistent. Interestingly, such cases do not seem to occur in natural language. As an illustration, sentence (25b) does not seem to have a reading in which there was event in which all the inhabitants were killed (wide event scope), and for certain bomb fragments there were bombing events in which they caused inhabitants to die (narrow event scope).

(25) In the bombing, all the inhabitants were killed by bomb fragments.

Champollion (2015) claims that event scope is *always* narrow, which would mean that event scope does not need to be annotated at all and inconsistencies with relative scoping cannot occur. A sentence like *"Everybody died in the crash."* would seem to contradict this claim, however, as does (25).

## 6 Granularity in QuantML annotations

The preceding sections were inspired by the aim of allowing fine-grained annotation of quantification in a semantically well-defined way. As mentioned in Section 1, another important aim of semantic annotation is to allow representations which are *not* so fine-grained, since in many use contexts it is not relevant to make very fine-grained interpretations. This is especially true of quantifications, where issues of scope, distributivity, and exhaustiveness are not in all use cases of great interest. In this section we briefly consider the instruments that are available in QuantML for making annotations that are not maximally fine-grained.

First, QuantML annotations are modular. The abstract syntax of clause annotation structure contains a collection of entity structures and link structures When some of the components are missing, due to incomplete information, this and does not necessarily make the annotation structure uninterpretable, but allows it for example to be interpreted as an underspecified DRS (Reyle, 1993).

Second, some of the information in an annotation structure may be optional. Bunt et al. (2018) distinguish three types of optionality, which are all present in QuantML. *Semantic optionality* is that an annotation structure may have a certain component, according to its abstract syntax definition, but is also allowed without that component. Examples are the specification of the size of a reference domain and the specification of non-restrictive modifiers. Annotation structures with such components have a more specific semantics. *Syntactic optionality* is that a certain component does not need to be specified in annotation representations (using XML or some other format) but does have a default value in the encoded abstract syntax. Examples are the polarity and event scope of participation link structures. Finally, it may be convenient to allow certain components in concrete representations which do not encode anything in the abstract syntax, and thus have no semantic interpretation. Example are the marking up of a quantification as generic and, in ISO-TimeML (ISO 24617-1:2012) the encoding of parts of speech to distinguish verbal from nominal descriptions of events.

Third, some aspects of the information may be specified by more or less specific values. An example is the "unspecific" distributivity, which allows participant sets containing both individual objects and sets of individual objects. This is illustrated in plint structures of the form (8e) and(8f).

## 7 Concluding remarks

This paper presents certain details of the semantic definition of QuantML annotations that have so far been outlined only sketchily in the formal specification of QuantML (ISO CD 24617-12: 2023; see also Bunt, 2020). Various forms of merge operation on discourse representation structures, relying on pattern matching techniques, have been shown to allow for a compositional interpretation of annotation structures that describe quantifications in terms of sets of events and multiple sets of participants.

With the availability of the instruments mentioned in the previous section for avoiding being over-specific, QuantML aims to strike a balance between allowing fine-grained and more coarse-grained, empirically useful annotations of quantification phenomena, supported in all case by a compositional semantic interpretation.

## Acknowledgements

I would like to thank Jae-Woong Choe, Robin Cooper, Rainer Osswald and Stephen Pulman for their comments on an earlier version of this paper.

## References


Aristotle (350BC). *Analytika Protera. Translated into English as Prior Analytics by A.J. Jenkinson. Published by Kessinger Publishers (2010).*

Barwise, J. and R. Cooper (1981). Generalized Quantifiers and Natural Language. *Linguistics and Philosophy 4*, 159–219.

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

Bunt, H. (2020). *Semantic Annotation of Quantification in Natural Language, Ed. 2. TiCC Technical Report* 2019-12. Tilburg Center for Cognition and Communication, Tilburg University.

Bunt, H., M. Amblard, J. .Bos, K.Fort, P. de Groote, B. Guillaume, C. Li, P. Ludmann, M. Musiol, G. Perrier, S. Pavlova, and S. Pogadalla (2022). Quantification Annotation in ISO 24617-12, second edition. In *Proceedings 13th International Conference on Language Resources and Evaluation (LREC 2022)*, Marselle, France.

Champollion, L. (2015). The interaction of compositional semantics and event semantics. *Linguistics and Philosophy 38 (1)*, 31–66.

Davidson, D. (1967). The Logical Form of Action Sentences. In N. Resher (Ed.), *The Logic of Decision and Action*, pp. 81–95. Pittsburgh: The University of Pittsburgh Press.

Hobbs, J. and S. Shieber (1987). An algorithm for generating quantifier scopings. *Computational Linguistics 13*(1-2), 47–63.

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

ISO (2022). *ISO/CD 24617-12:2022, Language Resource Management: Semantic Annotation Framework (SemAF) - Part 12: Quantification*. Geneva.: International Standard. International Organisation for Standardisation ISO.

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

Krifka, M. (1999). At least some determiners aren't determiners. In K. Turner (Ed.), *The Semantics/Pragmatics Interface From Different Points of View*, pp. 257–291. Amsterdam: Elsevier.

Lee, K. (2023). *Annotation-based Semantics for Space and Time in Language*. Cambridge University Press.

Montague, R. (1971). The proper treatment of quantification in ordinary language. In R.Thomason (Ed.), *Formal Philosophy*. New Haven: Yale University Press.

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

Reyle, U. (1993). Dealing with ambiguities by underspecification: Construction, representation, and deduction. *Journal of Semantics*, 123–179.

Szabolcsi, A. (2010). *Quantification*. Cambridge (UK): Cambridge University Press.