Annotation of Quantification: The Current State of ISO 24617-12

Harry Bunt
Tilburg University
harry.bunt@uvt.nl

Abstract

This paper discusses the current state of developing an ISO standard annotation scheme for quantification phenomena in natural language, as part of the ISO Semantic Annotation Framework (ISO 24617). An approach that combines ideas from the theory of generalised quantifiers, from neo-Davidsonian event semantics, and from Discourse Representation Theory was proposed to the ISO organisation in 2019 as a starting point for developing such an annotation scheme. This scheme consists of (1) a conceptual 'metamodel' that visualises the types of entities, functions and relations that go into annotations of quantification; (2) an abstract syntax which defines 'annotation structures' as triples and other set-theoretic constructs; (3) an XML-based representation of annotation structures ('concrete syntax'); and (4) a compositional semantics of annotation structures. The latter three components together define the interpreted language QuantML. The focus in this paper is on the structuring of the semantic information needed to characterise quantification in natural language and the representation of these structures in QuantML.

Keywords: semantic annotation, quantification, ISO standards, QuantML

1. Introduction

The specification of an interoperable scheme for the semantic annotation of quantification phenomena in natural language has for some time been on the agenda for extending the ISO Semantic Annotation Framework. After preliminary studies, reported in Bunt (2017), Bunt et al. (2018), and Bunt (2019a), a concrete proposal for developing the specification of such a scheme, supported by a first 'working draft' (ISO/WD 24617-12), was adopted by the ISO organisation. This paper describes the current state of developing the proposed specification, elaborating the WD 24617-12 working draft. Although this work is still in a preliminary stage, the current state of the specification covers a fairly wide range of aspects and forms of quantification, including collective, cumulative, and group quantification, quantification over events, exhaustive quantification, negative-polarity quantification, quantified possessives, various forms of mass noun quantification, quantification involving parts of individuals, and quantification over complex domains including the use of quantifying modifiers with inverse linking. We refer to this annotation scheme under development by the name of its markup language, QuantML.

The interest in developing a semantic annotation scheme for quantification is twofold. First, there is the ubiquitous character of quantification in natural language. Quantification occurs when a predicate is applied to one or more sets of arguments. Since this happens in every clause when a verb is combined with its arguments (except perhaps in extremely simple sentences like *"John loves Mary"*, if proper names are regarded as referring expressions), quantification occurs in virtually every sentence. Quantification is moreover the most important source of structural ambiguity. Accurate question answering, information extraction, advice giving, negotiation, and other applications that rely on deep language understanding therefore struggle with the interpretation problems caused by quantification. Second, the ISO Semantic Annotation Framework (ISO 24617, 'SemAF') has parts for annotating temporal and spatial information, events, semantic roles, discourse relations, dialogue acts, and coreference relations, which together span a substantial range of semantic aspects of spoken and written language, but quantification forms a big gap in this range. Filling this gap would greatly enhance the coverage of SemAF.

Assuming that a semantic representation of a natural language (NL) expression is understood to be a formal expression that has a single well-defined interpretation corresponding to a possible meaning of the NL expression, a semantic *annotation* is somewhere in between a 'raw' NL expression and a semantic *representation.* A semantic annotation adds information to the annotated NL expression about its indented interpretation. In the simplest case, a semantic annotation identifies exactly one interpretation of the annotated NL expression, and thus corresponds to a single semantic representation, but in practice the situation is more complicated. First, semantic annotations are typically *constraints* on the possible interpretations, selecting a subset of its possible meanings rather than a single specific interpretation. Second, semantic representations do not necessarily carve out just one possible meaning. In fact, the pervasive ambiguity of quantifier scopes in NL expressions has prompted the definition of formalisms for underspecified semantic representations, thus blurring the distinction between semantic annotations and representations somewhat. Still, in practice the two are very different in two respects:

1. semantic annotations typically capture only certain aspects of natural language utterance meaning, such as properties of quantifications, or coreference relations, or spatiotemporal information;
2. semantic representations are typically designed as expressions in a formal logical language, while

> annotations are often designed to be a way of attaching certain labels to parts of NL expressions, such as semantic roles predicate-argument structures.

The various parts of the ISO Semantic Annotation Framework each deal with a certain type of semantic information, and thus with a certain type of constraints on semantic interpretation. Each of these parts defines an annotation scheme for the kind of information that it deals with, with the aim of specifying information that disambiguates an NL expression *in that respect*, such as which semantic role is played by an NP, or how is an anaphoric expression referentially related to which antecedent. Quantification is the most important source of structural ambiguity in natural language, and the annotation of the quantifications in an NL expression aims at disambiguating NL expressions in that respect.

The main challenge in developing an annotation scheme for quantification is to identify a limited number of categories of information that is sufficient for characterising aspects, forms and uses of quantification that are found in natural language, and to define the combinations of these categories that form meaningful building blocks in annotations. Annotations should be constructed in accordance with the methodological principles laid out in ISO standard 24617-6 (Principles of semantic annotation), which means that the annotations should have an abstract and a concrete syntax, related through an encoding function (from abstract to concrete) and an inverse decoding function, and a semantics defined for the abstract syntax (and inherited by any concrete encoding).

Annotations should moreover be in stand-off format. The use of stand-off formats is motivated primarily by the consideration that the integrity of the primary data should always be respected, and has the methodological advantage that the pointers from elements in an annotation to elements in the primary data ('markables'), formalise the relation between annotation structures and linguistic elements, making explicit that the semantic annotation of an NL expression is not a stand-alone object, but is formally attached to NL elements.

This paper is organised as follows. Section 2 outlines the analytical framework for quantification annotation that is proposed in ISO WD 24617-12 (2019). Section 3 discusses the categories of semantic information identified in the QuantML annotation scheme for characterising aspects and forms of quantification. First, a number of traditional categories are considered such as scope and distributivity. Second, a number of less well-established categories are introduced, and some novel uses of traditional categories. Section 4 closes the main paper with some concluding remarks and directions for further work. The appendix contains a summary specification of the QuantML markup language and its underlying metamodel.

2. Analytical Framework

2.1. Quantification: GQT

Quantification is linguistically, logically, and computationally extremely complex, and has been studied for centuries by logicians, linguists, formal semanticists, and computational linguists (e.g. Aristotle, 4th century B.C., Frege, 1879; Montague, 1974; Barwise and Cooper, 1981; Westerståhl, 1985; Keenan and Stavi, 1986; Hobbs and Shieber, 1987; Partee, 1988; Cooper, 1983; Kamp and Reyle, 1993; Bos, 1995; Peters and Westerståhl, 2006, Szabolcsi, 2010; Ruys and Winter, 2011, Champollion, 2015; Coppock and Beaver, 2015). Mostowski (1957) and Lindström (1966) noted that the universal and the existential quantifier, as used in predicate logic, can be viewed as expressing properties of sets of individual objects, involved in a predication: the universal quantifier expresses the property of containing all the elements of a given domain; the existential quantifier the property of containing at least one such element. This opened the way to generalise the notion of a quantifier to other properties of sets, such as the property of containing more than three elements, or of containing most of the elements of the quantification domain. The concepts in this broader class of quantifiers are called 'generalised quantifiers'.

The study of how generalised quantifiers are used and expressed in natural language has led to generalised quantifier theory (GQT, Barwise and Cooper, 1981). An important point in this theory is that there is a fundamental difference between quantification in natural language and quantification in logic in the following sense. Words like *"all"* and *"some"* in English, as well as their equivalents in other languages, may seem to be the counterparts of the universal (∀ 'for all') and existential (∃, 'for some' quantifiers of formal logic, and so-called 'cardinal quantifiers and 'proportional quantifiers' like *"three"*, and *"most"*, may seem to be the counterparts of certain generalised quantifiers, but they are not. In formal logic, if p is a formula that denotes a proposition then the expressions '$\forall x.p$' and '$\exists y.p$' are quantifications, saying that p is true of all individual objects and that p is true of at least one such object, respectively.

Such quantifications, which range over all individual objects in a universe of discourse, cannot be expressed in natural languages. Quantifying expressions in natural languages, instead, like *"all students", "quelques gens"*, and *"mais que cinco melodias"*, include the indication of a restricted domain. This is reflected in the view that quantifiers in natural language are not determiners like *"all"* and *"some"*, but noun phrases (Barwise and Cooper, 1981).

The QuantML annotation scheme takes an approach which combines generalized quantifier theory with the neo-Davidsonian event-based approach.

2.2. Neo-Davidsonian event semantics

Abzianidze & Bos (2019) note that neo-Davidsonian event semantics is adopted in most if not all semantically annotated corpora. Davidson (1989) introduced events as individual objects into semantic representations, notably as an extra argument of predicates that correspond to verbs, as in 'read(e, x, y)'. In a variation of this approach, known as 'neo-Davidsonian' (Dowty, 1989; Parsons, 1990) the number of arguments of verb-related predicates is not increased, but instead one-place predicates are applied to existentially quantified event variables, and thematic roles, a.k.a. semantic roles, are used to represent the roles of the participants in events, as in 'read(e), agent(e,x), theme(e,y)'.

QuantML combines GQT with neo-Davidsonian event semantics. This has two advantages: it allows a treatment of adverbial temporal quantifiers such as *"twice"*, *"more than three times"*, *"daily"*, and *"twice an hour"*, and it is convenient since this approach is also taken in other parts of SemAF.

Using a neo-Davidsonian approach implies the use of an inventory of semantic roles. For reasons of intra-SemAF compatibility and in line with the recommendation by Abzianidze & Bos (2019) to use an existing role inventory, QuantML uses the set of roles defined in ISO 24617-4, which is based on the LRICIS and VerbNet inventories (see Bunt & Palmer, 2013; Bonial et al. 2011; Petukhova & Bunt, 2008).

2.3. Annotation theory

The QuantML scheme is designed according to the ISO principles of semantic annotation (ISO standard 24617-6, 'SemAF Principles', see also Bunt (2015) and Pustejovsky et al. (2017). This means that the QuantML markup language has a three-part definition consisting of (1) an abstract syntax that specifies the possible *annotation structures* at a conceptual level as set-theoretical constructs; (2) a semantics that specifies the meaning of the annotation structures defined by the abstract syntax; (3) a concrete syntax, that specifies a representation format for annotation structures (for example using XML). Defining the semantics at the level of the abstract syntax puts the focus of an annotation standard at the conceptual level, rather than at the level of representation formats. Annotators (human or automatic) work with concrete representations only, but they can rely on the existence of an underlying abstract syntax layer and its semantics for the interoperability of their annotations.

The abstract syntax is a detailed formalization of the metamodel of the annotation scheme. It specifies a store of basic concepts, called the 'conceptual inventory', and it decribes how the elements of the inventory can be used to build well-formed annotation structures in set-theoretical terms, like pairs, triples, and more complex nested structures. Two types of structure are distinguished: *entity structures* and *link structures*. An entity structure contains semantic information about a segment of primary data and is formally a pair $\langle$m, s$\rangle$ consisting of a markable, which refers to a segment of primary data, and certain semantic information. A link structure contains information about the way two or more entity structures are semantically related. The most important entity structures in QuantML are those that describe events and their participants, corresponding to the elements <event> and <entity> in XML representations; the most important link structures are those that link participants to events and those that specify quantifier scopes. See for example Figure 1.

The annotation structures defined by the abstract syntax can be represented (or 'encoded') in a variety of ways; XML is the most popular representation format, but other formats, such as attribute-value matrices or annotation graphs would be equally possible (Ide and Bunt, 2010).

The concrete syntax specifies a vocabulary and a class of syntactic structures, such as XML elements, which together define a class of well-formed representations, and an encoding function that assigns such a representation to every well-formed annotation structure.

The QuantML semantics has the form of an interpretation-by-translation into semantic representations; the recursive interpretation function I_Q 'translates' annotation structures to Discourse Representation Structures (DRSs) in a compositional way, compositional in the sense that the interpretation of an annotation structure is obtained by combining the interpretations of its component entity structures and link structures. This particular form of the QuantML semantics is a choice of convenience rather than one of principle, inspired by the fact that DRSs have also been used as the semantic basis of several other (ISO) semantic annotation schemes. Other choices, such as the use of Minimal Recursion Semantics (Copestake et al., 2005) could work equally well, and although the compositionality of the semantics seems a desirable feature, not all existing proposals for interpreting quantifiers are compositional (e.g., Robaldo, 2011). The specification of the QuantML semantics most importantly shows exactly what QuantML annotations mean.

3. Information Categories

3.1. Overview

For the annotation of quantification in the QuantML scheme, so far the following information categories have been identified:

1. domain specificity
2. definiteness and determinacy
3. distributivity
4. individuation (count/mass)
5. cardinality and size
6. absolute and proportional involvement
7. exhaustivity
8. participant quantification scope

9. event quantification scope
10. modification scope
11. polarity and scope of negation
12. repetitiveness and frequency

Several of these categories are very well known and/or have been discussed for their use in QuantML in one of the previous publications on the development of QuantML, mentioned in Section1. The use of these categories will be very briefly summarised in the next subsection. The rest of this section describes novel QuantML uses of the categories listed above, in particular relating to possessives, negation, exhaustiveness, quantification over masses and parts of individuals, and participation in repetitive events.

3.2. Traditional categories

3.2.1. Domain, definiteness and determinacy

Full-fledged noun phrases consist of two parts: (1) one or more determiners of various kinds of (*"all", "the", "a", "most", "all five", "two of his"*), and (2) a nominal head (bare noun or nominal complex). The latter part, called the *restrictor*, indicates a certain domain that is considered in the quantification. We use the term *'source domain'* to refer to the entities denoted by the restrictor. Quantifications are very often restricted to a contextually determined part of the source domain, the *'reference domain'*, also called 'context set' (Westerståhl, 1985; Partee et al., 1990). For example, the quantifier *"every student"* typically does not apply to literally every person who is a student, but only to the students in a particular class or school. The definiteness of an NP is often an indication that the reference domain of the quantification is a specific part of the source domain, rather than the entire source domain.

In English and in many other languages the determiner part of an NP is a prenominal sequence of determiners of different types. Grammars commonly distinguish different classes of determiners, with different possible sequencing and co-occurrence restrictions. In English grammar a distinction is made between predeterminers, central determiners, and postdeterminers (e.g. Quirk et al., 1972; Leech and Svartvik, 1975), having the following different functions:

- predeterminers express the quantitative involvement of the reference domain, and may, additionally, provide information about the distributivity of the quantification;
- central determiners determine the definiteness of the NP, and thus co-determine a reference domain;
- postdeterminers contain information about the cardinality of the reference domain.

This is illustrated by the NP *"All my nine grandchildren"* in (1), where *"all"* is a predeterminer, *"my"* a central determiner, and *"nine"* a postdeterminer.

(1) All my nine grandchildren are boys.

While being definite is often an indication that some particular, determinate entity or collection of entities is considered, the relation between the semantic property of determinacy and the morphological category of definiteness is not straightforward (Coppock & Beaver, 2015; Peters & Westerståhl, 2013). The semantic difference between definite and indefinite expressions has been discussed in terms of familiarity, novelty, salience, uniqueness, and existence presuppositions. In QuantML the view is taken that definiteness is an indication of determinacy, interpreted as restricting a quantification to a reference domain that is somehow constrained through considerations of familiarity and salience, but that this can be overruled by contextual information. Conversely, an NP being indefinite does not necessarily mean that the quantification applies to the NP's entire source domain; contextual considerations often carve out a more restricted reference domain.

3.2.2. Relative scope

Studies of quantifier scope have focused almost exclusively on the relative scopes of quantifications over sets of participants, as in the classical example *"Everybody in this room speaks two languages"*. Relative scopes of this kind are not a property of one of the quantifications involved, but are a semantic relation between them. This is annotated in QuantML as follows:

(2)
```
<entity xml:id="x1" target="#m1" involvement="all"
   definiteness="det" pred="person"/>
<entity xml:id="x2" target="#m3" involvement="2"
  definiteness="indet" pred="language">
<scoping arg1="#x1" arg2="#x2"
  scopeRel="wider"/>
```

(The reading with reverse scope order would be annotated with arg1="#x2" arg2="x1".) The relative scoping of participants and *events* is also a relevant issue. This is illustrated by the two possible readings of the sentence *"Everybody will die."* Besides the reading according to which everyone is mortal, there is also a reading which predicts an apocalyptic future event in which everyone will die. In the annotation in (3) the relative scope of events and participants is marked up by means of the attribute @eventScope, which has been added to the XML element <srLink> as defined in ISO 24617-4.[1]

(3)
```
<entity xml:id="x1" target="#m1" pred="person"/>
<event xml:id="e1" target="#m2" pred="die"
  time-"fut"/>
<srLink event="#e1" participant="#x1"
  semRole="theme" eventScope="wide"/>
```

---

[1] The @scopeRel attribute in <scoping> elements, which is used to annotate the relative scopes of two participant sets, has possible values that are not applicable to the relative scoping of events and participants.

Cumulative quantification, a case of branching quantification (Barwise, 1978, Hintikka, 1973; Scha, 1981), as occurring in (4) (due to Reyle, 1993), is treated in QuantML as mutual outscoping of the quantifiers. That is, the reading where there is a set A of 3 breweries and a set B of 15 inns, such that the members of A supplied the members of B, and the members of B were supplied by the members of A, is annotated by the scope relation @scopeRel=“dual”.

(4) Three breweries supplied fifteen inns.

Group quantification is treated as a case of wide event scope in combination with collective distributivity; see Section 3.4. Other issues of scope concern the interaction between quantifiers and modifiers, and between quantifiers and negations; these are discussed below in the sections 3.5 and 3.6.
Scope underspecification is done in QuantML by simply omitting one or more <scoping> elements. The semantics of such a QuantML structure is an underspecified DRS (UDRS, Reyle 1993).

3.2.3. Distributivity

The distinction between distributive and collective quantification is well known, but other cases must be distinguished as well. Example (5) may describe a situation where the boys involved did not necessarily do all the carrying either collectively or individually, but where they carried some heavy boxes collectively and some other, less heavy boxes individually. More importantly, the question whether a set of participants is involved in certain events collectively or individually is not always relevant. So in some contexts it is inappropriate to make the collective/distributive distinction and consider the quantification as ambiguous.

(5) The boys carried all the boxes upstairs

The quantifications in this sentence have ‘unspecific’ distributivity (Bunt, 1985); the sentence just says that all the boxes were somehow carried upstairs by the boys, and all the boys somehow participated. This reading has also been called a ‘cover reading’ (Schwarzschild,1996), and can be seen as a cumulative reading with unspecific distributivity. (Ordinary cumulative readings have individual distributivity.) Cover readings are annotated in QuantML by both quantifiers having “unspecific” distributivity and “dual” relative scoping. Following Kamp & Reyle (1993), we use the notation X^* to designate the set consisting of the members of X and the subsets of X, and the predicate P^* to designate the characteristic function of the set X^*, where P is the characteristic function of X. Using moreover the notation R_0 to indicate the characteristic function of a reference domain that forms a subset of a source domain with characteristic function R, the ‘unspecific’ interpretation of (5) can be represented in second-order predicate logic as follows:[2]

(6) $$\forall \mathrm{x}.[\mathrm{box}_0(\mathrm{x}) \rightarrow \exists \mathrm{y}.\exists \mathrm{e}.[\mathrm{boy}_0{}^*(\mathrm{y}) \wedge \text{carry-up}(\mathrm{e}) \wedge \text{agent}(\mathrm{e,y}) \wedge \exists \mathrm{z}.[\mathrm{box}_0{}^*(\mathrm{z}) \wedge [\mathrm{x{=}z} \vee \mathrm{x} \in \mathrm{z}] \wedge \text{theme}(\mathrm{e,z})]]] \wedge$$
$$\forall \mathrm{y}.[\mathrm{boy}_0(\mathrm{y}) \rightarrow \exists \mathrm{u}.\exists \mathrm{e}.[\mathrm{box}_0{}^*(\mathrm{u}) \wedge \text{carry-up}(\mathrm{e}) \wedge \text{theme}(\mathrm{e,u}) \wedge \exists \mathrm{x}.[\mathrm{boy}_0{}^*(\mathrm{x}) \wedge [\mathrm{y{=}x} \vee \mathrm{y} \in \mathrm{x}] \wedge \text{agent}(\mathrm{e,x})]]]$$

The distributivity of a quantification is not a property of the set of participants in a set of events, but a property of the way of participating. This is illustrated by example (7), assuming that *“the men”* individually had a beer, and collectively carried the piano upstairs.

(7) The men had a beer before carrying the piano upstairs.

Distributivity should thus be marked up on the participation relation in the drinking and carrying events, as in the annotation fragment shown in (8), where the XML element <srLink> from ISO 24617-4 has been enriched with the attribute ‘@distr’:

(8)
```
<entity xml:id=“x1” target=“#m1”
  pred=“man”/>
<event xml:id=“e1” target=“#m2” pred=“drink”/>
<event xml:id=“e2” target=“#m3” pred=“carry”/>
<srLink event=“#e1” participant=“#x1”
  semRole=“agent” distr=“individual”/>
<srLink event=“#e2” participant=“#x1”
  semRole=“agent” distr=“collective”/>
```

3.2.4. Size and cardinality

Cardinal determiners indicate the size of a set; in (9), the central determiner “twenty-seven” indicates the cardinality of the reference domain, while the predeterminer “twenty-five” indicates the cardinality of the subset of the reference domain whose members were involved in vote-events.

(9) Twenty-five of the twenty-seven states voted in favour.

At least the following quantitative aspects of a quantification must be taken into account: (1) the cardinality of the reference domain; (2) the number of elements in the reference domain involved in the predication; and (3) the size of sets, groups, or sums of individuals that are involved in a collective predication. See also Section 3.4 on group quantification.

3.3. Involvement and exhaustivity

The meaning of a cardinal determiner may depend on the speaker’s intention, as expressed by the stress pattern of an utterance in which it is used. Used with focal stress, “two” may give rise to a partitive interpretation; for example, in (10a) “two salesmen” means “two of the salesmen”, different from (10b) where the stress is on “salesmen”.

---

[2] Plural entities involved in quantifications can be viewed as mereological objects rather than as sets (Bunt1983; Champollion (2019); for the annotation as proposed in QuantML this makes little difference – see Bunt, 2019c).

(10) a. TWO salesmen came in.
b. Two SALESmen came in.

The occurrence of a cardinal determiner in focus relates also to the much debated issue whether a determiner (or a numeral) like "two" should be interpreted as "exactly two", as "two or more", or as "at most two". Consider the following examples:

(11) a. Two dogs are growling.
b. Do you have two AA batteries?
c. How many children does Mary have?
Mary has two children.
d. Mary has at most two children.

The standard GQT interpretation of quantifiers of the form *"two N"* is the property of being a set that contains two Ns. So for example, in DRT (Kamp and Reyle, 1993) sentence (11a) is interpreted as claiming the existence of a set X containing two growling dogs. Now suppose there are in fact three growling dogs - in that case it is also true that there are two growling dogs. So *"two"* in (11a) is in fact interpreted as 'two or more'. This seems reasonable for sentence (11a). For sentence (11b), uttered in a context where the speaker is examining a remote control with two apparently flat batteries, this is the only reasonable interpretation. But in (11c) the answer to the question licences the inference that Mary does not have more than two children or less than two children, so in this case *"two"* means 'exactly two'.

It is widely assumed (e.g. Partee, 1988; Kamp and Reyle, 1993; Krifka, 1999) that the numeral *"two"* indicates that the cardinality of the set (or individual sum) denoted by the NP that it modifies is exactly 2, but that the generalized quantifier *"two N"* is interpreted in some contexts as "at least two N' and in others as "exactly two N', due to context-specific (Gricean) pragmatic inferences - see Kadmon (2001). Quantifier readings of the type "exactly two N" are called 'exhaustive', and can be thought of as generated by a covert operator, an 'exhaustivizer', that could be lexicalized as *"only"* (see Szabolcsi, 2010). In (11), replacing *"two"* by *"only two"* in case a and case c enforces or reinforces the 'exactly two' reading, whereas in case b the replacement would be distinctly odd. Similar issues arise when *"two"* forms part of a monotone-decreasing quantifier, as in (11d), which is inherently exhaustive. The exhaustiveness of a quantifier relates to focus placement, as illustrated by (10a).

Exhaustive linking occurs when the set of individuals involved in a quantified predication contains all the participants of which the predication is said to hold, as in *"(Only) Two people attended the wedding"* and in *"(Only) Two colleagues did not attend the wedding"*.

(12) (Only) TWO dogs barked.
Markables: m1=Two dogs, m2=dogs, m3=barked

QuantML annotation:

```
<entity xml:id="x1" target="#m1" domain="#x2"
  involvement="2" exhaustiveness="exhaustive"
   definiteness="indet"/>
<sourceDomain xml:id="x2" target="#m2"
  individuation="count" pred="dog"/>
<event xml:id="e1" target="#m3" pred="bark"/>
<participation event="#e1" participant="#x1"
  semRole="agent" distr="individual"
  eventScope="narrow"/>
```

3.4. Group quantification

Quantifications with wide event scope and collective distributivity allow readings with so-called 'group quantification', as illustrated by the quantification over "*parts*" in example (13).

(13) Each of these machines assembles more than fifty parts.

Upon the 'group' reading, in every assembly-event where one of the machines under considerations is the agent, a collection of more than 50 parts is involved in the theme role. The annotation of this sentence is shown in figure 1.

3.5. Individuation

Studies of quantification in natural language have often been restricted to cases where the NP head is a count noun. Quantification by means of a mass NP is in many respects similar, but there are some interesting differences. Compare the two sentences in (14):

(14) a. The boys polished all the knives in the drawer.

b. The boys drank all the milk in the fridge.

In (14a) a predicate is applied to a set of knives, and likewise in (14b) a predicate is applied to a set of quantities of milk. A difference is that (14a) can be analysed as: *Every knife in the drawer was the theme in a polish-event with one of the boys as the agent*, but it is not clear that the analogous analysis *Every quantity of milk in the fridge was the theme in a drink-event with one of the boys as the agent* would make sense, since the set of quantities of milk in the fridge may include bottles of milk, pints of milk and, other quantities that were not as such the object of a drink-event.

A universal mass noun quantification of the form *"all the M"* does not necessarily refer to *all* the quantities of M. A detailed analysis of mass noun quantification can be found in Bunt (1985), where elements from lattice theory and set theory are formally integrated. Quantities are analysed as having a part-whole structure (just like sets), defining a sum operation Σ such that the sum of two quantities of M forms another quantity of M. One interpretation of expressions of the form *"all the M"* is as referring to a set X of quantities of M that together make up the reference domain M_0, in the sense that their sum equals the sum of all quantities in the reference domain: $\Sigma(X) = \Sigma(M_0)$.

Markables: m1 = “Each machine”, m2 = “machine”, m3 = “assembles”, m4 = “more than fifty parts”, m5 = “parts”

QuantML annotation representation:

```
<entity xml:id=“x1” target=“#m1” domain=“#x2” involvement=“all” definiteness=“det”/>
<sourceDomain xml:id=“x2” target=“#m2” pred=“machine” indiv=“count”/>
<event xml:id=“e1” target=“#m3” pred=“assemble”/>
<entity xml:id=“x3” target=“#m4” domain=“#x2” involvement=“>50” definiteness=“indet’/>
<sourceDomain xml:id=“x4” target=“#m5” pred=“part” indiv=“count”/>
<participation event=“#e1” participant=“#x1” semRole=“agent” distr=“individual” eventScope=“narrow”/>
<participation event=“#e1” participant=“#x3” semRole=“agent” distr=“collective” eventScope=“wide”/>
<scoping arg1=“#x1” arg2=“#x3” scopeRel=“wider”/>
```

Figure 1: Annotation of group quantification

Since mass nouns do not individuate their reference (Quine, 1960), quantification by mass NPs would seem not to allow individual distribution. Yet there is a distinction similar to the individual/collective distinction of count NP quantifiers, as (15) illustrates.

(15) a. All the water in these lakes is polluted.
b. The sand in the truck weighs twelve tons.
c. The boys carried all the sand to the back yard.

In (15a) the predicate of being polluted applies to any sample of *“the water in the lake”*; this distribution is called *homogeneous*. In (15b) the predicate of weighing 12 tons applies to the quantities of sand taken together, so this is a form of collective quantification. In (15c) the boys did not carry every quantity of sand, but certain quantities that together make up “all the sand”; in this case the distribution is called *unspecific*.
These examples illustrate three different ways in which the quantification domain of a mass NP can be completely involved in a predication, corresponding to three different senses of expressions of the form *“all M”* (or *“all the M”*) in English, and similarly in other languages. Complete involvement with homogeneous distribution, as in (15a), where *“all the water”* quantifies over the set of all contextually distinguished quantities of water, is annotated with the @involvement attribute having the value “all”. In cases like (15c), where *“all the sand”* refers to a subset of quantities of sand that together make up all the (contextually distinguished) sand - the @involvement attribute has the value “total”. Finally, on the collective reading of (15b), where *“(all) the sand”* refers to the quantity of sand formed by all contextually relevant quantities of sand together, the involvement will be annotated as “whole”. This is summarized in Table 1.

Although count nouns do individuate their reference in terms of individuals, there is a form of quantification with count NPs that resembles the ‘total, unspecific’ quantification with mass NPs (Bunt, 2017). Consider the example *“Mario ate three pizzas for dinner”*. The standard interpretation would go something like this: There is a set of three pizzas that were the object in an eat-event at dinner time with Mario as the agent. But now consider: *“Mario ate five pizzas last week”*. A plausible interpretation could now be: Last week Mario ate in total 5 pizzas in some eat-events (for example, 2.75 pizza in one event and 2.25 pizza in another). This interpretation requires the consideration of pizza parts as the participants in eat-events, and a notion of summation of parts (in this example adding up to 5 pizzas). Quantifications of this kind are annotated in QuantML by the @individuation attribute in <sourceDomain> elements having the value “count/parts”.

3.6. Modification scope

Relative scope is an issue not only between two participant quantifications, or between a participant quantification and an event quantification, but also when the head noun of a quantifying NP is modified by a relative clause, a prepositional phrase, or a possessive phrase that contains quantifiers. In that case a quantifier in the modifier may outscope the quantification over the head noun. The following examples illustrate this phenomenon, which is known in the linguistic literature as ‘inverse linking’ (May, 1977; May and Bale, 2007; Szabolcsi, 2010; Ruys and Winter, 2011; Barker, 2014).

(16) a. Two students from every college participated.
b. The children’s toys were stolen.

The relative scoping of the two quantifiers is in these case annotated as a property of the modifying relation, expressed by the value “inverse” of the attribute @linking in a <ppMod> or a <possRestr> element, as shown in Fig. 2.
Possessive expressions introduce a relation that is not made explicit, or that is expressed using a rather vague preposition like *“of”* in English and *“de”* in Romance languages. Typical examples are shown in (17). All these (and other) forms have in common that they express some sort of possession relation between a (set of) possessor(s) and a set of possessions. Possessive expressions involve quantification over possessions (and possibly also over possessors). A case like (17a1) can be analysed schematically as in (17b), introducing a generic ‘Poss’ relation as proposed by Peters and Westerståhl (2013).

distribution	involvement	interpretation	example
homogeneous	all	For all quantities of M	(15a)
unspecific	total	For the elements in a set of quantities of M that together make up the whole of M	(15c)
collective	whole	For M as a whole	(15b)

Table 1: Involvement and distributivity in mass NP quantification.

Markables in sentence (16a):
m1=“Two students from every university”, m2=“students”, m3=“students from every university”, m4=“from every university”, m5=“every university”, m6=“university”, m7=“participated”

QuantML annotation:

```
<entity xml:id=“x1” target=“#m1” domain=”#x2” involvement=”2” definiteness=”indef”/>
<refDomain xml:id=“x2” target=“#m3” source=“#x3” restrs=”#r1”/>
<sourceDomain xml:id=“x3” target=“#m2” individuation=”count” pred=“student”/>
<ppMod xml:id=“r1” target=“#m4” pRel=“from” pEntity=“#x4” distr=”individual” linking=”inverse”/>
<entity xml:id=“x4” target=“#m5” domain=”#x5” involvement=”all” definiteness=”det”/>
<sourceDomain xml:id=“x5” target=“#m6” individuation=”count” pred=“university”/>
<event xml:id=“e1” target=“#m7” pred=“participate”/>
<participation event=“#e1” participant=“#x1”semRole=“agent” distr=”unspecific” eventScope=”narrow”/>
```

Figure 2: QuantML annotation of modification scope

(17) a. 1. Tom’s house
2. John and Mary’s two children
3. two of my books
4. the headmaster’s children’s toys
5. the children of the headmaster
6. every student’s library card

b. house(x) ∧ tom(y) ∧ Poss(x,y)

3.7. Polarity and scope of negation

The QuantML scheme does not offer a general treatment of the annotation of polarity and modality, but it provides devices for dealing with the relative scopes of quantifications and negations. The example sentence in (18) illustrates the possible scopes of a negation at clause level, the negation scoping either over the entire clause, over the clause minus *“the unions”*, or just over the determiner in *“the unions”*. The first two readings can be distinguished in annotations by means of a @polarity attribute in <participation> elements with the value “neg-wide” for wide-scope negation and “neg-narrow” for the second reading, while the third reading is distinguished by the value of the @involvement attribute in the corresponding <entity> element indicating that less than all of the individuals in the reference domain are involved.

(18) a. The unions do not accept the proposal.

b. It is not the case that all the unions accept the proposal [some of them don’t]
<participant event=“#e1” participant=“#x1” semRole=“agent” distr=“individual” polarity=“neg-wide”/>

c. All the unions do not accept the proposal [none of them does]
<participant event=“#e1” participant=“#x1” semRole=“agent” distr=“individual” polarity=“neg-narrow”/>

d. Not all the unions accept the proposal [though most of them do]
<entity xml:id=“x1” target=“#m1” pred=“union” involvement=“not-all”/>

Note that this way of annotating negation scope makes it possible to handle cases of double or triple negation, as in *“Not all the unions do not accept the proposal”*.

3.8. Repetitiveness

The annotation of repeated participation in recurring events has been treated in ISO 24617-1 as a quantification over temporal objects, but in spite of the suggestion that comes from the word “times” in the English language, expressions like *“once”, “twice”* and *“three times”* do not really quantify over time, but rather over sets of eventualities (Lewis, 1975). The QuantML scheme does not provide a complete proposal for dealing with adverbial temporal or spatial quantification, but repetitiveness can be covered in a natural way by using the concepts available in QuantML. Participation in a k-times repetitive event is annotated by means of a <participation> element with @repetitiveness = “k”, the semantics of which is given by (19) for individual, non-exhaustive participation in the Agent role with narrow event scope and positive polarity.[3] (Note that k can

---

[3] Alternatively, repetitiveness could be annotated in <event> elements, but that would make the formulation of the semantics of annotation structures slightly more complex.

be any numerical predicate that identifies a range of natural numbers, such as 'only once', 'more than three times' or 'two or three times'.)

(19) I_Q(Agent, individual, narrow, non-exhaustive, k, positive) =
$[\, X \mid x \in X \rightarrow [\, E \mid k(E), e \in E \rightarrow agent(e,x) \,]]$

This leads for example to the semantic interpretation (20b) for the sentence (20a), where 'child$_0$" designates the predicate 'child' restricted to the reference domain formed by the contextually distinguished children:

(20) a. Two of the children called twice.

b. $[\, X \mid |X|=2, x \in X \rightarrow [\, E \mid |E|=2, child_0(x), e \in E \rightarrow [\, call(e), agent(e,x) \,]]$

4. Conclusions

Although the development of QuantML as an ISO standard is still in a preliminary stage, the scheme as developed so far supports the annotation of quite a variety of forms and aspects of quantification in a way that is interoperable (a) in the sense that its XML-based representation format is just one possible encoding of the underlying abstract annotation structures with their formal semantics, and (b) in the sense of sharing a view on sentence meaning rooted in Neo-Davidsonian event semantics, and DRT with other parts of ISO SemAF.

Current limitations of QuantML have to do with the limitations of the events-and-participants view and with lack of agreement on the analysis of certain forms of quantification. The events-and participants approach seems to be stretched to its limits for verbs that take abstract concepts like thoughts, beliefs, desires, etc. as their arguments, as in *"Bob wants to catch a fish"*.

Forms of quantification that have so far escaped a generally agreed analysis include generics and habituals, whose theoretical status has not been fully resolved; see e.g. Kamp and Reyle (1993), Section 3.7.4. Krifka et al. (1995) analyse generics in terms of a special default quantifier; others introduce a notion of 'normal' or 'prototypical' into the interpretation framework (cf. Eckhardt, 2000; van Rooij and Schulz, 2020).

Another issue for further work concerns the overlaps between QuantML and schemes for annotating other phenomena, such as events and coreference. The recently introduced notion of an annotation scheme plug-in with its interface (Bunt, 2019b) may provide a mechanism for dealing with such overlaps.

Most importantly, the QuantML annotation scheme needs to be validated in manual and automatic annotation. For manual annotation, the scheme reflects the fact that quantification in natural language is an extremely complex matter. To do justice to this complexity, the annotation scheme is inevitably quite complex itself, and impossible for use by untrained annotators, except perhaps if annotators are supported by an interactive annotation tool that for example asks questions like "Did the men act together or each one by himself?", to distinguish between collective and distributive readings, and suggests appropriate default values of certain attributes. An extensive user manual and a repository of annotated examples would also seem to be indispensable for training annotators, and such material could be useful as well as training material for automatic annotation.

References

Abzianidze, L. and J. Bos (2019). Thirty musts for meaning banking. In *Proceedings of ACL 2019 First International Workshop on Designing Meaning Representations*, pp. 15–27.

Barker, C. (2014). Scope. In S. Lappin and C. Fox (Eds.), *The Handbook of Contemporary Semantic Theory*, pp. 40–76. New York: John Wiley.

Barwise, J. (1978). On Branching Quantifiers in English. *Journal of Philosophical Logic 8*, 47–80.

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

Bonial, C., 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.

Bos, J. (1995). Predicate logic unplugged. In *Proceedings 10th Amsterdam Colloquium*, Amsterdam, pp. 133–142. ILLC.

Bunt, H. (1985). *Mass Terms and Model-Theoretic Semantics*. Cambridge University Press.

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. (2017). Towards interoperable annotation of quantification. In *Proceedings ISA-13, Thirteenth International Workshop on Interoperable Semantic Annotation*, Montpellier, pp. 84–94.

Bunt, H. (2018). *Semantic Annotation of Quantification in Natural Language. TiCC Technical Report* 2018-15. Tilburg Center for Cognition and Communication, Tilburg University.

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

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

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

Bunt, H. and M. Palmer (2013). Conceptual and representational choices in defining an ISO standard for semantic role annotation. In *Proceedings of the 9th Joint ISO - ACL Workshop on Interoperable Semantic Annottiion (ISA-9)*, Potsdam, pp. 41–50.

Bunt, H. and J. Pustejovsky (2010). Annotating temporal and event quantification. In *Proceedings ISA-5, Fifth International Workshop on Interoperable Semantic Annotation*, pp. 15–22. City University of Hong Kong.

Bunt, H., J. Pustejovsky, and K. Lee (2018). Towards an ISO Standard for the Annotation of Quantification. In *Proceedings 11th International Conference on Language Resources and Evaluation (LREC 2018)*, Myazaki, Japan. ELRA.

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

Champollion, L. (2019). Distributivity in Formal Semantics. *Annual Review of Linguistics 5*, 289–308.

Cooper, R. (1983). *Quantification and syntactic theory*. Dordrecht: Reidel.

Copestake, A., D. Flickinger, C. Pollard, and I. Sag (2005). Minimal Recursion Semantics: An Introduction. *Research on Language and Computation 3*, 281–332.

Coppock, E. and D. Beaver (2015). Definiteness and determinacy. *Linguistics and Philosophy 38*, 377–435.

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.

Dowty, D. (1989). On the Semantic Content of the Notion of 'Thematic Role'. In G. Chierchia, B. Partee, and R. Turner (Eds.), *Properties, Types and Meaning*, Volume 2. Dordrecht/Boston/London: Kluwer Academic Publishers.

Eckhardt, R. (2000). Normal objects, normal worlds, and the meaning of generic sentences. *Journal of Semantics 16*, 237–278.

Frege, G. (1897). *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle: Nebert.

Hintikka, J. (1973). Quantifiers vs. quantification theory. *Dialectica 27*, 329–358.

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

Ide, N. and H. Bunt (2010). Anatomy of annotation schemes: Mapping to GrAF. In *Proceedings 4th Linguistic Annotation Workshop (LAW IV)*, Uppsala.

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

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

ISO (2014b). *ISO 24617-7: 2014, Language Resource Management - Semantic Annotation Framework (SemAF) - Part 7: Spatial information*. 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*. Geneva: International Organisation for Standardisation ISO.

ISO (2016). *ISO 24617-8:2016, Language Resource Management - Semantic Annotation Framework (SemAF) - Part 8: Semantic relations in discourse, Core annotation scheme (DR-Core)*. Geneva: International Organisation for Standardisation ISO.

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

Kadmon, N. (2001). *Formal Pragmatics*. Oxford: Blackwell.

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

Keenan, E. and J. Stavi (1986). A semantic characterization of natural language determiners. *Linguistics and Philosophy 9*, 253–326.

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.

Krifka, M., F. Pelletier, G. Carlson, A. ter Meulen, G. Chierchia, and G. Link (1995). Genericity: An introduction. In G. Carlson and F. Pelletier (Eds.), *The generic book*, pp. 1–124. Chicago: University of Chicago Press.

Leech, G. and J. Svartvik (1975). *A communicative grammar of English*. London: Longman.

Lewis, D. (1975). Adverbs of quantification. In E. Keenan (Ed.), *Formal Semantics of Natural Language*, pp. 3–15. Cambridge: Cambridge University Press.

Lindström, P. (1966). First order predicate logic with generalized quantifiers. *Theoria 32*, 186–195.

May, R. (1977). *The Grammar of Quantification*. Ph.D. Dissertation, MIT.

May, R. and A. Bale (2005). Inverse linking. In M. Everaert and H. van Riemsdijk (Eds.), *The Blackwell Companion to syntax, Vol. 2*, pp. 639–667. Oxford: Blackwell.

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

Mostowski, A. (1957). On a generalization of quantifiers. *Fundamentae Mathematicae 44*, 12–36.

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

Partee, B. (1988). Many quantifiers. In *ESCOL 89: Proceedings of the Eastern States Conference on Linguistics, Reprinted in Compositionality in Formal Semantics: Selected Papers by Barbara Partee*, Oxford, pp. 241–158. Blackwell 2004.

Partee, B., A. ter Meulen, and R. Wall (1990). *Mathematical Models in Linguistics*. Berlin: Springer.

Peters, S. and D. Westerståhl (2006). *Quantifiers in Language and Logic*. New York: Oxford University Press.

Peters, S. and D. Westerståhl (2013). The semantics of possessives. *Language 89(4)*, 713–759.

Petukhova, V. 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.

Pustejovsky, J., 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.

Pustejovsky, J., H. Bunt, and A. Zaenen (2017). Designing annotation schemes: From theory to model. In N. Ide and J. Pustejovsky (Eds.), *Handbook of Linguistic Annotation*, pp. 21–72. Berlin: Springer.

Quine, W. (1960). *Word and Object*. Cambridge, Mass.: MIT Press.

Quirk, R., A. Greenbaum, G. Leech, and J. Svartvik (1972). *A Grammar of Contemporary English*. London: Longman.

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

Robaldo, L. (2011). Distributivity, collectivity, and cumulativity in terms of (in)dependence and maximality. *Journal of Logic, Language and Information 23*(1), 233–271.

Rooij, R. v. and K. Schulz (2020). Generics and typicality: a bounded rationality approach. *Linguistics and Philosophy 43(1)*, 83–117.

Ruys, E. and Y. Winter (2011). Quantifier Scope in Formal Linguistics. In D. Gabbay and F. Guenthner (Eds.), *Handbook of Philosophical Logic*, pp. 195–225. Dordrecht: Foris.

Scha, R. (1981). Collective, distributive and cumulative quantification. In J. Groenendijk and M. Stokhof (Eds.), *Formal Methods in the Study of Language*, pp. 483–512. Amsterdam: Mathematisch Centrum.

Schwarzschild, R. (1996). *Pluralities*. Dordrecht: Kluwer Academic Publishers.

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

Westerståhl, D. (1985). Determiners and context sets. In J. van Benthem and A. ter Meulen (Eds.), *Generalized Quantifiers in Natural Language*, pp. 45–71. Dordrecht: Foris.

Appendix:
Outline of QuantML specification

Metamodel

The metamodel underlying QuantML annotations shows the concepts that make up annotation structures corresponding to the information categories discussed in Section 3, with their grouping into entity structures and link structures – see Figure 3.

Abstract syntax

The structures defined by the abstract syntax are n-tuples of elements that are either basic concepts, taken from a store called the 'conceptual inventory', or, recursively, of such n-tuples. Two types of structure are distinguished: entity structures and link structures.

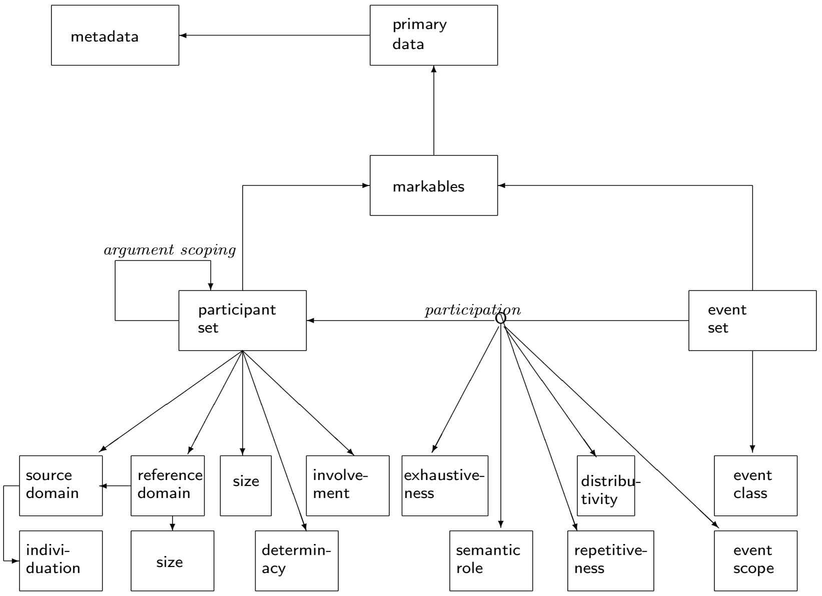

Figure 3: QuantML Metamodel.

An entity structure contains semantic information about a segment of primary data and is formally a pair ⟨m, s⟩ consisting of a markable, which refers to a segment of primary data, and certain semantic information. A link structure contains information about the way two or more segments of primary data are semantically related.

QuantML conceptual inventory:

- unary predicates that characterize source domains (such as 'book', 'student', and 'water) or event domains (such as 'lift', 'carry', 'drink'), or that correspond to adjectives or to prepositions;
- binary predicates that correspond to semantic roles, notably the roles as defined in ISO 24617-4 (Semantic roles);
- numerical predicates for specifying reference domain involvement, reference domain size, size of certain parts of a reference domain, or number of repetitions or frequency of recurrence in event structures;
- predicates for specifying proportional reference domain involvement, such as 'all', 'some', 'most';
- parameters for specifying definiteness: 'determinate' and 'indeterminate'; domain individuation: 'count', 'mass', 'count/parts'; and distributivity: 'collective', 'individual', 'homogeneous', 'single' (used for singular proper names and definite descriptions), 'unspecific';
- basic units of measurement, such as 'meter', 'kilogram', 'litre', and the operators 'division' and 'multiplication' for forming complex units;
- the polarity values 'positive' and 'negative';
- the values 'exhaustive' and 'non-exhaustive';
- parameters for specifying event scope: 'wide' and 'narrow' (default value), and for specifying whether scope inversion occurs: 'inverse' or 'linear' (default value).
- ordering relations for specifying the relative scopes of quantifiers over sets of participants: 'wider', 'dual', and 'unscoped'.

QuantML has three kinds of entity structures: (1) for events; (2) for participants; (3) for restrictions on sets of participants. A quantified set of participants is characterized by the following properties:

- the source domain, from which the participants are drawn, and its individuation;
- the reference domain, typically a subset of the source domain;
- the quantitative (absolute or proportional) involvement of the reference domain;
- the size of the reference domain, or of groups, subsets, or parts of the reference domain involved in the quantified predication.

The entity structure $\langle m, s\rangle$ for a set of participants thus contains a triple s = ⟨⟨D, v⟩, q, d⟩ with D = characteristic domain predicate, v = individuation, q = reference domain involvement, and d = determinacy, with possibly an additional size specification. The domain component is more complex when the restrictor of an NP contains head noun modifiers and/or multiple, conjoined heads (see Bunt 2018 for details). Entity structures for sets of events are very simple; they contain just a predicate that characterizes a domain of events.

Modifier structures come in five varieties, depending on whether the head noun of an NP is modified by an adjective, noun, PP, relative clause, or possessive restriction. These are not spelled out in Fig. 3.

Two kinds of link structure are defined: participation structures, which link participants to events, and scope link structures. Participation structures are a 7- or 8-tuple, specifying (1) a set of events; (2) a set of participants; (3) a semantic role; (4) a distributivity; (5) the exhaustiveness of the participation; (6) the relative scope of the event quantification; (7) the polarity, which is "positive" by default; and possibly (8) a repetitiveness. Scope link structures specify the relative scope of two participant entity structures.

Annotation structures for quantification are associated mostly with clauses and their constituent NPs and verbs. The annotation structure for a clause is a quadruple consisting of an event structure, a set of participant structures, a set of participation link structures, and a set of scope link structures. In a complete clause annotation structure all participant structures are linked to the verb's event structure, and the relative scopes of all participant entity structures are specified.

Concrete syntax

A concrete syntax is specified here in the form of an XML representation of annotation structures. For each type of entity structure, defined by the abstract syntax, a corresponding XML element is defined; each of these elements has an attribute @xml:id whose value is a unique identifier (unique within the annotation structure), and an attribute @target, whose value anchors the annotation in the primary data, having a markable as value (or a sequence of markables). In addition, these elements have the following attributes:

<**entity**>: @domain, @involvement, @definiteness and @size (optional);
<**event**>: @pred (event class);
<**refDomain**>: @source (with multiple values in the case of a conjunctive head) and @restrictions;
<**sourceDomain**>: @pred, @individuation;
<**adjMod**>: @pred, @distr, and @restrictions (optional);
<**nnMod**>: @pred, @distr, and @restrictions (optional);
<**ppMod**>: @pRel, @pEntity, @distr, @linking;
<**relClause**>: @semRole, @clause, @distr, @linking;
<**amount**>: @num, @unit;
<**complexUnit**>: @unit1, @operation, @unit2.

For each of the two types of link structure defined by the abstract syntax, an XML element is defined:

<**participation**> with attributes @event, @participant, @semRole, @distr, @eventScope, @exhaustiveness, [optionally: @repetitiveness]; and @polarity;

<**scoping**> with the attributes @arg1, @arg2, @scopeRel.

Semantics

The QuantML semantics specifies a recursive interpretation function I_Q that translates annotation structures into DRSs in a compositional way: the interpretation of an annotation structure is obtained by combining the interpretations of its component entity structures and participation link structures, in a way that is determined by scope link structures (if any). For a full specification of the QuantML semantics see Bunt (2019c). Here we outline the overall approach and present some interesting parts of the definition of I_Q.

The QuantML interpretation function translates every participant entity structure, event entity structure, and participation link structure into a DRS and combines these. Consider the example in (22). The entity structures for *"More than two thousand students"*, and *"protested"* are translated into the DRSs shown in (22)b and c. For the participant entity structure this is achieved by applying an instance of clause (21a) in the I_Q definition, which interprets entity structures with source domain D, individuation *v*, involvement *q*, and definiteness *indef*. The interpretation q' of domain involvement specification *q* is defined in (21b-c), and that of the domain specification in (21d-e).

(21) a. $I_Q(\langle \text{m1}, \langle\langle\langle \text{m2}, D\rangle, v\rangle, q, \textit{indef}\,\rangle\rangle)$ = [X | $q'(X)$, [x ∈ X → D'(x)]]
b. $q' = I_Q(q) \circ F_Q(v)$
c. $F_Q(v)$: $F_Q(\textit{count}) = \lambda\text{X}.|\text{X}|$; $F_Q(\textit{mass}) = F_Q(\textit{count/parts}) = \lambda\text{X}.|\Sigma\text{X}|$
d. $D' = I_Q(\langle D, v\rangle)$
e. $I_Q(\langle D,\ v\rangle)$: $I_Q(\langle D,\ \textit{count}\rangle) = I_Q(\langle D,\ \textit{mass}\rangle) = I_Q(D)$; $I_Q(\langle D, \textit{count/parts}\rangle) = I_Q(D)^+$

(22) a. More than two thousand students protested.
b. $I_Q(\langle \text{m1}, \langle\langle\langle \text{m2, student}\rangle, \text{count}\rangle, \lambda z.|z| > 2000$, indet $\rangle\rangle\rangle$) = [X | |X|=2000, [x ∈ X → student(x)]]
c. $I_Q(\langle \text{m3}, \langle\text{protest}\rangle\rangle)$ = [E | [e ∈ E → protest(e)]]

The DRS in (22b) says that there exists a set with the property of containing two thousand students, reflecting the GQT approach to NP interpretation. The DRS in (22c) together with (24) illustrates the adoption of neo-Davidsonian event semantics.

The participation link structure has in this example the form $\langle \epsilon_E,\ \{\epsilon_{P1}\},$ R, d, $\xi,\ \sigma,$ p$\rangle$, where ϵ_E and ϵ_{P1} are the participant and event entity structures that are linked in the Agent role (R = Agent), with d = collective, ξ = non-exhaustive, σ (event scope) = narrow, and p (polarity) = positive. The semantic interpretation of such a structure is defined as follows, where '∪' designates the familiar merge operation for DRSs:

(23) $I_Q\langle \epsilon_E, \{\epsilon_{P1}\}$, R, d, $\sigma\rangle = I_Q\langle \epsilon_{P1}\rangle \cup (I_Q\langle \epsilon_E\rangle \cup I_Q($R, d, ξ, σ, p)

A triple like ⟨ R, d, σ⟩ is interpreted as shown in (24):

(24) a. I_Q(R, individual, narrow) = [X | x ∈ X → [E | e ∈ E → agent(e, x)]
b. I_Q(R, individual, wide) = [E | e ∈ E → [X | x ∈ X → agent(e, x)]
c. I_Q(R, collective, σ) = [X, E | x ∈ X → [e | e ∈ E, R(e, X)]

Applying rule (23) to the right-hand sides of (22) and (24c), with the values for R, d and σ substituted, gives the desired result shown in (25):.

(25) [X | |X| >2000, [x ∈ X → student(x)], [E | e ∈ E → [protest(e), agent(e,X)]]]