ISO 24617-12: A New Standard for Semantic Annotation

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

Abstract

This paper presents ISO 24617-12, an annotation schema for quantification phenomena in natural language, as part of the ISO Semantic Annotation Framework (ISO 24617). This schema combines ideas from the theory of generalised quantifiers, from neo-Davidsonian event semantics, and from Discourse Representation Theory. The schema consists of (1) an abstract syntax which defines 'annotation structures' as triples and other set-theoretic constructs made up of quantification-related concepts; (2) a reference representation of annotation structures ('concrete syntax'); and (3) a compositional semantics of annotation structures. Together, these components define the interpreted markup language QuantML. This paper focuses on the identification and structuring of the semantic information useful for the characterisation of quantification in natural language and the interoperable representation of these information structures in QuantML.

Keywords: semantic annotation, quantification, ISO standards, QuantML

1. Introduction

ISO 24617-12 is a proposed new standard[1] for the annotation of quantification phenomena. Such an annotation schema is of interest for at least two reasons. First, quantification occurs in every language in every sentence, except in trivially simple sentences like *"It is raining"* and *"John loves Mary"*. This makes its annotation of obvious interest for corpus-based studies. The ISO Semantic Annotation Framework SemAF (ISO 24617), intending to support the creation and maintenance of corpora with interoperable semantic annotations, has parts for annotating temporal and spatial information, events, semantic roles, discourse relations, dialogue acts, and coreference relations. Adding quantification to this portfolio would greatly enhance the coverage of SemAF. Second, quantification is arguably the most important source of non-lexical ambiguity, and its analytic description is therefore quintessential for applications of natural language processing which rely on deep language understanding and accurate information extraction.

From a semantic point of view, quantification occurs when a predicate is applied to a collection of entities, rather than a single entity. This raises questions like: is the predicate intended to be applied to the individual members of the collection, or to the collection as a whole ('distributivity'); how much or how many entities is the predicate said to be true of ('involvement'); is the predicate said to be true of no other entities than those in the referenced set ('exhaustiveness'); in the case of a predicate that has multiple arguments, which one has widest scope and what is the role of each argument (such as Argument1, Argument 2, Argument 3, or Agent, Theme, Instrument,...). The main challenge in developing an annotation schema for quantification is to identify a set of categories of semantic information that is sufficient for characterising the aspects of quantification that are expressed in natural language, and to define effective building blocks for annotation structures based on these categories.

This paper is organised as follows. Section 2 summarizes the theoretical background of the annotation schema. Section 3 discusses some related and preliminary work. Section 4 describes the categories of semantic information identified in the ISO 24617-12 annotation schema for characterising aspects of quantification. Traditional categories are considered such as scope, distributivity, determinacy, and polarity, as well as some less well-established categories such as event scope and modifier linking Examples are provided of the use of these categories in QuantML. Section 5 indicates a few specific provisions available in QuantML for annotating quantifications that involve parts of individual objects, quantifications with wide event scope, and quantification in copular constructions. Section 6 discusses various forms of optionality of QuantML attributes, convenient for making annotation tasks easier, and the support that annotators can find in the explanatory repository of annotation examples called the 'QuantificationBank'. Section 7 indicates some limitations of the present annotation schema and concludes with an outlook for further work.

[1] At the time of writing, the process of reviewing this proposal and deciding on its acceptance as an ISO standard was still ongoing. The outcome is expected in early 2024.

LREC-COLING 2024, pages 9361–9371

2. Background

2.1. Generalized Quantifier Theory

Quantification is linguistically, logically, and computationally extremely complex, and has been studied for centuries by logicians, linguists, formal semanticists, and computational linguists, from Aristotle (± 350 B.C.), Frege (1879) and Montague, (1974) via Barwise & Cooper (1981) and Hobbs & Shieber (1987) to Szabolcsi (2010), Peters & Westerståhl (2013), Keenan & Paperno (2012, 2017), Szymanik (2016) and Champollion (2015; 2019).
The study of how quantifiers are expressed in natural language has led to *generalised quantifier theory* (GQT, Barwise and Cooper, 1981). GQT interprets quantifiers as properties of a set of entities. Quantifying expressions in natural language are 'restricted', containing an indication of the entities to which the quantification is meant to apply. Natural language quantifiers are thus not determiners like *"all"* and *"some"*, but rather noun phrases like *"all students"*, *"some sonatas"*. Adverbial temporal and spatial quantifiers like *"often", "rwice", "nowhere"* have a restriction to temporal and spatial entities built into their lexical semantics.

2.2. Event semantics

The QuantML schema combines GQT with neo-Davidsonian event semantics (Davidson, 1989; Parsons, 1990), which views the combination of a verb and its arguments as the involvement in a certain semantic role of the entities denoted by the argument in events denoted by the verb. This allows a semantic interpretation of adverbial modifiers and quantifiers, and it is also used in other parts of SemAF.

Using a neo-Davidsonian approach implies the use of an inventory of semantic roles. For intra-SemAF compatibility, QuantML uses the set of roles defined in ISO 24617-4:2014, which is based on the LRICIS and VerbNet inventories (see Bonial et al. 2011, Bunt & Palmer, 2013).

2.3. Semantic Annotation Principles

QuantML 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 triple-layered definition consisting of:

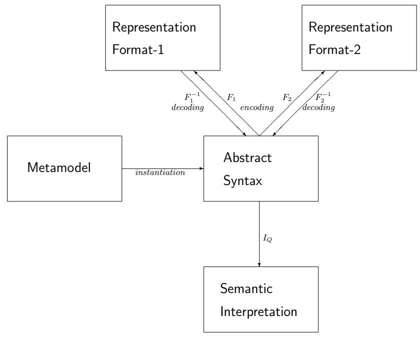

Figure 1: ISO 24617-12 Architecture.

1. An abstract syntax, which specifies the class of well-defined *annotation structures* as pairs, triples, and other set-theoretical constructs containing quantification-related concepts. Annotation structures consist of two kinds of substructures: *entity structures*, which contain information about a stretch of primary data, and *link structures*, which contain information relating two (or more) entity structures. The abstract syntax is visualized in a metamodel (see Fig. 2).

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

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

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

3. Related Work and Preliminary Studies

The annotation schemas of ISO Time-ML (ISO 24617-1:2012) and ISO SpatialML (ISO 24617-7:2020) have some provisions for annotating quantifications over time and events, and for spatial quantification, respectively. These provisions are very limited, and semantically not quite satisfactory (see Bunt & Pustejovsky, 2016) .

The development of a schema for quantification annotation has been on the ISO agenda for some time, suffering delay due to the Covid-19 pandemic. A first, preliminary version of this annotation schema was presented in Bunt et al. (2018) and Bunt (2019a); a second, revised version in Bunt et al. (2022).

Szymanik & Kieras (2022) developed an annotated corpus of Polish quantifier expressions, as a sub-corpus of the National Corpus of Polish. Their annotations contain three categories of information: (1) whether the quantifier is expressed adverbially or by an NP; (2) whether they are universal, existential, proportional, or numerical; (3) whether they are left- or right-monotonous increasing, decreasing, or non-monotone.

Higgins & Sadock (2003) annotated a collection of 893 English sentences that contain two quantifications, taken from the Penn Treebank, and annotated these with scope preferences.

4. Aspects of Quantification in Natural Language

4.1. Overview

For describing properties of quantification in natural language, QuantML takes the following categories of semantic information into account:

(1)
1. domain and determinacy
2. distributivity
3. individuation (count/mass)
4. involvement (absolute and proportional)
5. argument role
6. exhaustivity
7. polarity
8. participant scope
9. event scope
10. repetitiveness
11. cardinality or size of reference domain
12. restrictiveness of modifiers
13. linking of modifiers (inverse or linear)
14. modality (e.g. epistemic)
15. genericity (generic or specific).

On the event-based view underlying QuantML, the backbone of a semantic characterisation of a quantification is formed by (1) a set of events, (2) a set of participants, and (3) the relation between them. These categories of information correspond to attributes in QuantML annotations (notably when representations in the pivotal XML-based concrete syntax are used). The first 10 categories correspond to *'core attributes'*, i.e. attributes that require a value whenever annotating a quantification. Some of these attributes are optional in the sense of having a default value, which is assumed if no value is supplied. (See further Section 6, where different types of optionality are discussed.) In addition to the core attributes, QuantML makes use of *'conditional attributes'*, which are relevant only for certain forms of quantification. The attributes 12 and 13 exemplify this: they apply only in case a quantifying expression contains a modifier that may restrict the reference domain. The items 14 and 15 are exceptional in that they exist only in the concrete syntax; they do not correspond to anything in the abstract syntax or the semantics, and have been added purely to allow corpus searches of generic or modal quantification.

The role of the information categories 1-13 and the corresponding attributes in annotations is visualised in Figure 2, which displays the QuantML data model. The specification of the abstract syntax and its semantics, which are not considered in this paper, provide a formal specification of this data model, a.k.a. 'metamodel'.

The metamodel brings out that three components play center stage in a QuantML annotation: specifications of events, participants, and participation relations, each with a number of features corresponding to the categories 1 - 13.

4.2. Domain and Determinacy

A quantifier expressed by an NP includes an explicit indication of the entities that are considered in the quantification. We call this set of entities, denoted by the NP head, the *source domain*. Quantifications are very often restricted to some particular, contextually determined entity or collection of entities that forms part of the source domain called the *'reference domain'*, a.k.a. 'context set' (Westerståhl, 1985; Partee et al., 1990). For example, the quantifier *"all the students"* typically does not intend to be applied to every person who is a student, but only to the members of a particular group of students, like a class or a team.

Determinacy is the semantic property indicating whether the reference domain is a proper subset of the source domain or coincides with it. Determinacy may be indicated by the morphosyntactic feature of definiteness (Coppock & Beaver, 2015; Peters & Westerståhl, 2013), especially in plural NPs. Indefinite plural NPs are often indeterminate or generic (as in *"Tigers don't eat tomatoes"*); definite plural NPs (like *"the students"*) are typically determinate. Definite singular expressions may well be semantically indeterminate, as in *"Gently, the beggar lay a hand on the queen's knee"*.

In Germanic and Romance languages definiteness can be expressed by a definite or indefinite article. In languages that do not have articles, such as Hungarian and Korean, determinacy can be expressed by different means, such as demonstratives (Chen, 2004; Lee, 1989; Jenks, 2004), but even more than in languages that do have articles, determinacy is determined by the context.

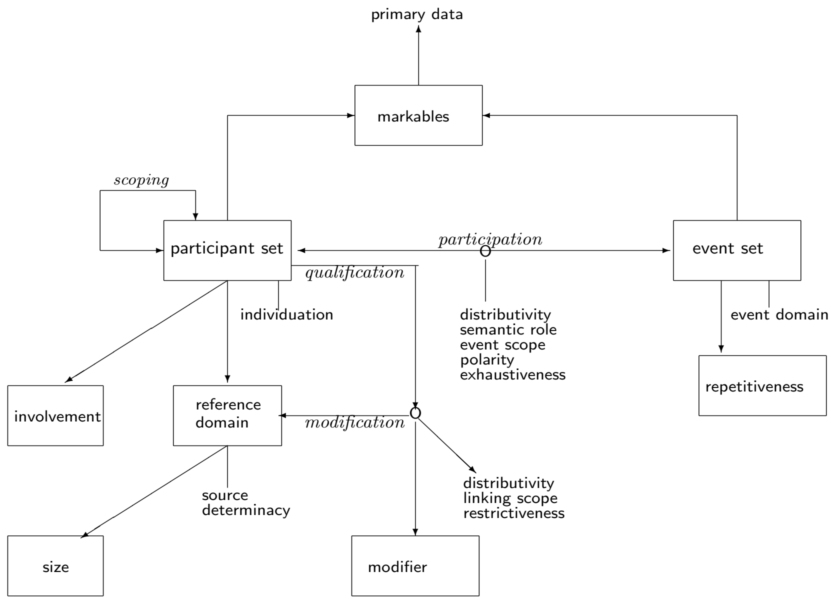

Figure 2: QuantML Metamodel.

4.3. Distributivity

Besides distributive and collective quantification, QuantML also supports the annotation of distributivity as 'unspecific', meaning that individual as well as sets of entities may be involved. The sentence in (2), for example, 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 (lighter) boxes individually.

(2) The boys carried all the boxes upstairs

Distributivity is a relational property, rather than a property of a participant set. This is illustrated by example (3), assuming that *"the men"* individually had a beer, and collectively carried the piano upstairs.

(3) The men had a beer after carrying the piano upstairs.

Distributivity is thus marked up by an attribute of the <participation> relational element in the drinking and carrying events, as in the annotation fragment shown in Fig. 3 for the sentence (2).

Group quantification is treated as a case of wide event scope in combination with collective distributivity; see Section 4.7. Other issues of scope concern the interaction between quantifiers and modifiers, and between quantifiers and negations; these are discussed below in the sections 4.9 and 4.10.

4.4. Involvement and exhaustivity

Not all the members of the reference domain of a quantification have to participate in the events under cconsideration. Those members which are actually involved form the *participant set*. Numerical and proportional determiners such as *many, more than five, most, a little, all, no* indicate how many/much of the reference domain constitutes the participant set. The meaning of a numerical determiner depends on the speaker's intention, as may be expressed by the stress pattern of a spoken utterance. Used with focal stress, *"two"* may give rise to a partitive interpretation; for example, in (4a) *"two salesmen"* means 'two of the (contextually distinguished) salesmen', different from (4b) where the stress is on *"salesmen"*.

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

The meaning of a numerical determiner relates to the much debated issue whether a numeral like *"two"* should be interpreted as 'exactly two' or as 'two or more'. Consider the following examples:

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

It is widely assumed (e.g. Partee, 1988; Kamp and Reyle, 1993; Krifka, 1999) that a numeral like *"two"* indicates that the cardinality of a set 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 pragmatic inferences (Kadmon, 2001). Quantifier readings of the type 'exactly two N' are called 'exhaustive', and can be

```
<entity xml:id="x1" target="#m1" domain="#x2" involvement="all"/>
<refDomain xml:id="x2" target="#m1" source="#x3" determinacy="det"/>
<sourceDomain xml:id="x3" target="#m2" individuation="count" pred="boy"/>
<event xml:id="e1" target="#m3" pred="carry-up"/>
<participation event="#e1" participant="#x1" semRole="agent" distr="unspecific"
      eventScope="narrow"/>
```

Figure 3: Annotation (fragment) of a quantification with 'unspecific' distributivity

thought of as generated by a covert operator, an 'exhaustivizer', that could be lexicalized as *"only"* (Szabolcsi, 2010). Exhaustivity is marked up in QuantML by means of an (optional) attribute of a <participation> element.

4.5. 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 *"all the students speak two languages"*. Relative scope is a relational property of two quantifications, and is expressed in QuantML with a <scoping> element with the attributes @arg1, @arg2, and @scopeRel.

Cumulative quantification, a case of branching quantification (Barwise, 1978, Hintikka, 1973; Scha, 1981), illustrated in (6)[2], is treated in QuantML as mutual outscoping of the quantifiers. That is, the reading of (6) in which 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 in QuantML by the @scopeRel attribute in a <scoping> element that has the value "dual".

(6) Three breweries supplied fifteen inns.

Scope underspecification is possible in QuantML by simply omitting one or more <scoping> elements. The semantics of a QuantML structure with incomplete scoping is an underspecified DRS (UDRS, Reyle 1993).

4.6. Argument Roles

Using neo-Davidsonian semantics means that a certain set of argument roles must be chosen. The particular choice of roles is as such not an issue for the annotation of quantification. For convenience and intra-SemAF consistency, QuantML uses the role set defined in ISO 24617-4:2014, Semantic roles (see also Bonial et al., 2011).

4.7. Size and cardinality

Numerical determiners indicate the size of a set, notably of a participant set, of a reference domain, or of a group of individuals. In (7), *"twenty-seven"* indicates the cardinality of the reference domain, *"twenty-five"* that of the participant set.

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

Group size is an aspect of *'group quantification'*, illustrated in (8) interpreted as saying that in assemble-events with one of these machines as the agent, a set of twenty parts is collectively participating as the theme.

(8) These machines assemble twenty parts each.

Upon the 'group' reading, in every assembly-event with one of theese machines as the agent, a collection of 20 parts is involved as the theme. The annotation of this example is shown in Figure 4.

4.8. Individuation

Studies of quantification have often focused on cases where the NP head is a count noun. Quantification with a mass NP is in many respects similar, but there are some interesting differences. Compare the two sentences in (9):

(9) a. The boys polished all the forks in the drawer.

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

In (9a) a predicate is applied to a set of forks, and likewise in (9b) a predicate is applied to a set of quantities of milk. A difference is that (9a) can be analysed as: *Every fork 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, drops 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 *each* of the quantities of M. A detailed analysis of mass noun quantification can be found in Bunt (1985), where elements from mereology and set theory are formally integrated. Quantities are analysed as having a part-whole structure, with a sum operation such that the sum of two quantities of M forms another quantity of M. Expressions of the form *"all*

[2]Example due to Reyle 1983.

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 4: Annotation of group quantification

the M" may be used to refer to every element in a set of quantities of M (the participant set) in the reference domain, but also to refer to the sum of these quantities, or to a set of quantities whose sum forms the reference domain. The examples in (10) illustrate this.

(10) a. All the water in these lakes is polluted.

b. The sand in the truck weighs twelve tons.

c. All the sand was carriedto the back yard.

4.9. Modifier Linking

When the head noun of an NP is modified by a relative clause, a prepositional phrase, or a possessive phrase that contains quantifiers, it may happen that a quantifier in the modifier takes scope over the quantification over the head noun. The following example illustrates this phenomenon, known in the linguistic literature as 'inverse linking' (May, 1977; May and Bale, 2007; Szabolcsi, 2010; Ruys and Winter, 2011; Barker, 2014).

(11) Two students from every college attended.

The relative scoping of the two quantifiers is in such a case annotated as a property of the modifying relation using the attribute @linking.

Possessive expressions indicate 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 (12). They can be analysed semantically with a generic 'Poss' relation, proposed by Peters and Westerståhl (2013). This approach is followed in QuantML, where they are treated like preposition phrases with this 'Poss' relation. Like PPs, possessive phrases may contain quantifiers and be inverse linked, as illustrated in (12).

(12) a. Lisa's eyes
b. John and Mary's children
c. the headmaster's children's toys
d. every student's library card

4.10. Polarity

The QuantML schema does not offer a general treatment of polarity, but it provides devices for dealing with the relative scopes of quantifications and negations. The example sentence in (13) illustrates some possible scopes of a negation at clause level, the negation scoping either over the entire clause or over the clause minus *"the unions"*. The 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.

(13) The unions do not accept the proposal.
a. at least one of them doesn't
`<participation polarity="neg-wide"/>`
b. none of them does
`<participation polarity="neg-narrow"/>`

4.11. Repetitiveness

Participation in recurring events is treated in ISO 24617-1 (Time and events) as a quantification over temporal objects, but in spite of the suggestion coming from the word *"times"* in English, expressions like *"once"*, *"twice"* and *"three times"* do not really quantify over time, but rather over sets of eventualities (Lewis, 1975). In QuantML a k-times recurring event is annotated by the @repetitiveness attribute in an <event> structure having the value"k".

5. Special Features of QuantML

This section describes a few specific provisions available in QuantML for annotating mass NP quantifications, quantifications that involve parts of individual objects, quantifications with wide event scope, and quantification in copular constructions.

- **Individuation with parts of individuals.** Although count nouns do individuate their reference in terms of individuals, there is a form

Markables in sentence (11a): m1="Two students from every university", m2="students", m3="students from every university", m4="from every university", m5="every university", m6="university", m7="attended"

Annotation:
```
<entity xml:id="x1" target="#m1" domain="#x2" involvement="2"/>
<refDomain xml:id="x2" target="#m3" source="#x3" determinacy="indet" individuation="count"
      restrs="#r1"/>
<sourceDomain xml:id="x3" target="#m2" pred="student"/>
<ppMod xml:id="r1" target="#m4" pRel="from" pEntity="#x4" distr="individual"
      inking="inverse"/>
<entity xml:id="x4" target="#m5" domain="#x5" involvement="all"/>
<refDomain xml:id="x5" target="#m5" source="#x6" individuation="count" determinacy="det"/>
<sourceDomain xml:id="x6" target="#m6" pred="university"/>
<event xml:id="e1" target="#m7" pred="attend"/>
<participation event="#e1" participant="#x1" semRole="agent" distr="unspecific"
      eventScope="narrow"/>
```

Figure 5: QuantML annotation of modification scope

of quantification with count NPs that resembles the 'total, unspecific' quantification with mass NPs. 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 ten pizzas last week"*. A plausible interpretation could now be that Mario ate in total 10 pizzas in some eat-events that took place last week, for example, 2.75 pizzas on Monday in one event, 3.25 pizzas on Wednesday, and so on. This interpretation requires the consideration of pizza parts as the participants in eat-events (as would be needed for it "Mario had half a pizza for lunch", 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".

- **Use of event scope.** Quantifications over participant sets tend to outscope those over events, but contra Champollion (2015), events may take wider scope. This is illustrated by the two possible readings of the sentence *"Everybody will die"*[3]. 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 the relative scope of events and participants is marked up by means of the attribute @eventScope in <participation> elements.

- **Quantification in copular constructions.** Quantified predicate-argument relations arise not only when a verb is combined with its arguments, but also when an adjective is used, as in (14), where both the collective and distributive interpretations are possible.

(14) These books are heavy.

Proposals for an event-based semantics for copular expressions like (14) have been put forward in two directions: (a) by positing a *be* state with the predicate (like *heavy*) and its argument (like *these books*) as participants (Kamp & Reyle, 1993), and (2) by positing a state corresponding to the adjectival predicate (like a *be heavy* state) with the argument as participant. While one may have misgivings about the ontological and linguistic status of *be* events (Maienborn, 2003), from a semantic point of view the first approach has the advantage of generalising to other copular verbs (such as *"appear", "seem", "look"*) and of going along smoothly with other verbs. Following this approach, QuantML uses a <predication> element as illustrated in Figure 5 for the annotation of the collective reading of (14).

6. Annotation Support

A semantic annotation schema for quantification cannot be simple, given the many aspects of quantification discussed in Section 4. QuantML is indeed not very simple. However, the use of QuantML is simplified greatly by the definition of many attributes as *optional*, and is further supported by the availability of the Quantification-Bank , an explanatory repository of annotated examples[4] The present section describes these forms of support for annotators.

[3]Example due to Bart Geurts (p.c.)

[4]https://sigsem.uvt.nl/QuantificationBank/

```
<entity xml:id="x1" target="#m1" domain="#x2" involvement="all"/>
<refDomain xml:id="x2" target="#m1" source="#x3" determinacy="det"/>
<sourceDomain xml:id="x3" target="#m2" individuation="count" pred="book"/>
<event> xml:id="e1" target="#m3" pred="be">
<participation event="e1" participant="#x1" distr="collective" semRole="theme">
<predtication event="#e1" participant="#x1" predicate="heavy" distr="collective" />
```

Figure 6: QuantML annotation quantifying adjective in copular construction

6.1. Optional attributes

Optionality with defaults: The specification of default values is especially useful for those attributes for which one of its possible values occurs much more often than the others, which often corresponds to the linguistic phenomenon of *markedness*. Polarity is an example: in natural language positive polarity is unmarked; negative needs to be marked explicitly,. The unmarked value is used in QuantML as a default value, implemented in the 3-level architecture as the value that the encoding function assigns in the abstract syntax of the annotations, and this value is also used in the semantic interpretation (see Fig. 1). The following QuantML attributes are optional in this sense:

- polarity (default: positive)
- event scope (default: narrow)
- exhaustiveness (default: non-exhaustive)
- repetitiveness (default: at least 1)
- restrictiveness (default: restrictive)
- linking (default: linear)

Optionality without defaults: A different form of optionality is when the values of an attribute allow a more specific semantic interpretation than when the attribute does not have a value. This is the case for the attribute @size, used to specify the size of a reference domain, and for the attribute @restrs in <refDomain> elements. These attributes have a value only when a quantification specifies a reference domain size or an NP head noun modifier is present, respectively.

Uninterpreted optionality: A third type of optionality is formed by attributes that only have values at the level of annotation representation, not in the abstract syntax and thus not interpreted semantically.

Lexical attributes: The attribute @pred has values that correspond to lexical items, which therefore do not need to be specified. It is only for convenience that in practice annotators are invited to specify the appropriate lexical items (or synset items, for example).

Altogether, only 5 of the 'core attributes' must always be given a value: (1) distributivity, (2) determinacy, (3) semantic role, (4) individuation, (5) involvement. The 'conditional attributes' are evidently used only in the annotation of certain specific forms of natural language quantifiers.

6.2. The QuantificationBank

For supporting annotators, the QuantificationBank[5] provides a repository of annotated examples, with annotation guidelines and explanations of the concepts.

7. Limitations and Outlook

ISO 25617-12 supports the annotation of a wide variety of forms and aspects of quantification, including mass NP quantification, group quantification, quantifications with unspecific distributivity or with a reference domain that includes parts of individuals, NP quantifiers with quantifying modifiers of the head NP and inverse linking. QuantML does not cover certain forms of quantification that have so far escaped a generally agreed analysis, including generics and habituals. Krifka et al. (1995) analyse generics in terms of a special default quantifier; other proposals introduce a notion of 'normal' or 'prototypical' entity into the interpretation framework (cf. Eckhardt, 2000; van Rooij and Schulz, 2020).

The main limitation of the present annotation schema is that it considers only the quantifications within a single clause (relative clauses included). This limitation seems natural, since the combination of verbal, adjectival, adverbial, or nominal predicates with their arguments occurs within a clause, but it precludes a treatment of quantifiers that contain references to elements outside the clause, as in the case of anaphoric possessive modifiers. It therefore seems interesting to develop an annotation schema that integrates the present one with the ISO 24617-9:2019 schema for reference annotation. The concepts of 'annotation schema plug-in' (Bunt, 2019b) and schema interlinking (Bunt, 2024) may provide a mechanism for achieving such an integration, as well as for dealing with overlaps with standards for the annotation of time and events (ISO 24617-1:2012) and spatial information (ISO 24617-10:2020).

[5]https://sigsem.uvt.nl/QuantificationBank

8. Bibliographical 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.

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. (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. (2024). Combining annotation schemes through interlinking. In *Proceedings ISA-20, Twentieth Joint ACL-ISO Workshop on Interoperable Semantic Annotation*, Turin, Italy.

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.

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

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

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

Bunt, H., J. Pustejovsky, and K. Lee (2018b). 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.

Chen, P. (2004). Identifiability and Definiteness in Chinese. *Linguistics 42 (6)*, 1129–1184.

Cheng, L. and R. Sybesma (1999). Bare and Not-So-Bare Nouns and the Structure of NP. *Linguistic Inquiry 30 (4)*, 509–542.

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.

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.

Higgins, D. and J. Sadock (2003). A machine learning approach algorithm to modeling scope preferences. *Computational Linguistics 29*(1), 73–96.

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.

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.

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.

Jenks, P. (2018). Articulated definiteness without articles. *Linguistic Inquiry 49*, 501–536.

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

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

Keenan, E. and D. Paperno (Eds.) (2012). *Handbook of Quantifiers in Natural Language, Vol. 1.* Cham, Switzerland: Springer.

Keenan, E. and D. Paperno (Eds.) (2017). *Handbook of Quantifiers in Natural Language, Vol. 2.* Cham, Switzerland: Springer.

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. 1–24. Chicago: University of Chicago Press.

Lee, C. (1989). (In)definites, Case Markers, Classifiers and Quantifiers in Korean. In *Harvard Studies in Korean Linguistics, Vol. III.* Dept. of Linguistics, Harvard University.

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

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.

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 (2013). The semantics of possessives. *Language 89(4)*, 713–759.

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.

Reyle, U. (1993). Dealing with ambiguities by underspecification: Construction, representation, and deduction. *Journal of Semantics 10*, 123–179.

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.

Szymanik, J. (2016). *Quantifiers and Cognition: Logical and Computational Perspectives*, Volume 96 of *Studies in Language and Philosophy*. Berlin: Springer.

Szymanik, J. and W. Kieras (2022). The semantically annotated corpus of polish quantificational expressions. *Language Resources and Evaluation 56 56*, 1057–1074.

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

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.