

# Towards an ISO Standard for the Annotation of Quantification

Harry Bunt<sup>1</sup>, Kiyong Lee<sup>3</sup>, and James Pustejovsky<sup>2</sup>

<sup>1</sup>Tilburg Center for Communication and Cognition, Tilburg University, The Netherlands

<sup>2</sup>Brandeis University, Waltham, MA, USA

<sup>3</sup>Korea University, Seoul, Korea

bunt@uvt.nl, pustejovsky@cs.brandeis.edu, ikiyong@gmail.com

## Abstract

This paper presents an approach to the annotation of quantification that is being developed in preparation of the specification of a quantification annotation scheme, as part of an effort of the International Organisation for Standardisation ISO to define interoperable semantic annotation schemes. The paper focuses on the theoretical basis for an ISO standard annotation scheme in this area. It is argued that the combination of Generalized Quantifier Theory, neo-Davidsonian event-based semantics, Discourse Representation Theory, and the ISO Principles of semantic annotation forms a powerful and solid foundation for defining annotations of quantification phenomena with an abstract and a concrete syntax and a compositional semantics. The coverage of the proposed annotation scheme includes both count and mass NP quantifiers, as well as NPs with syntactically and semantically complex heads with internal quantification and scoping structures, such as inverse linking by prepositional phrases and relative clauses.

**Keywords:** semantic annotation, annotation standard, quantification

## 1. Introduction

Quantification occurs in every sentence of written text or spoken discourse. This is because the application of a predicate to one or more sets of objects gives rise to questions of relative scope, of cardinality, and of the distribution (or ‘distributivity’) of the predicate over the sets of arguments. Dealing with these questions is of crucial importance for understanding, translating, or responding to sentences, and for correct information extraction and question answering from natural language text. Example sentence (1a) illustrates the influence of the relative scopes of quantifications on the information that an utterance contains (how many papers were read altogether?); (1b) illustrates the importance of the collective or individual distribution of a predicate over a set of arguments (assuming that a certain piano was carried collectively by certain men, who individually had a beer); and (1c) shows that a quantified relation may involve parts of individuals, such as pizza halves.

- (1) a. All the students read two papers on quantification.
- b. The men had a beer before carrying the piano upstairs.
- c. The three boys ate four and a half pizzas.

The International Organisation for Standardisation ISO is pursuing the establishment of standards for linguistic annotation in general, and semantic annotation in particular, in view of the importance of annotated language resources, both for empirically-based linguistic research and for developing NLP applications. In order to be widely applicable across theories and platforms, annotation standards should on the one hand be theory-neutral, but they should on the other hand also take established theoretical insights into account. This paper outlines an ISO standard annotation scheme under development for quantification that builds on logical and linguistic theories, notably on the theory of generalized quantifiers, on neo-Davidsonian event-based semantics, and on Discourse Representation Theory.

in combination with established principles for semantic annotation.

## 2. Theoretical background

### 2.1. Generalized Quantifier Theory

Philosophers, linguists, and AI researchers have extensively studied quantification phenomena. Logicians from Aristotle to van Benthem have studied quantification for its role in reasoning and thinking. It was noted relatively recently that quantifiers can be viewed as expressing a property of the involvement in a predication of sets of individual objects:  $\forall x$  expresses that all the individual objects in the universe of discourse are involved;  $\exists x$  that at least one such object is involved (Mostowski, 1957; Lindström, 1966). This notion of a quantifier has been generalized to properties such as those expressed in English by “two”, “most”, “less than half of”, “more than 2000”. The concepts in this broader class of quantifiers are called *generalized quantifiers*.

Quantifications in formal logic are statements about all individual objects in a given universe of discourse; natural languages, by contrast, have quantifying expressions like “all the students”, “a book”, “some wine”, and “more than five sonatas”, which indicate a specific domain that the quantification is restricted to. Generalized quantifier theory (GQT) therefore views noun phrases, rather than determiners, as the quantifiers of natural language (Barwise and Cooper, 1981). According to GQT, words like *all* and *some* in English do not form the counterparts of the universal and existential quantifiers of formal logic, and neither do words like *three*, and *most*, which have been called ‘cardinal quantifiers’ and ‘proportional quantifiers’ (Partee, 1988).

### 2.2. Event-based semantics

Some aspects of sentence meaning can be accounted for only if verbs are viewed as introducing events - in a broad sense of ‘event’ that includes states, facts, processes, and

their negations. Examples of such aspects are adverbials, but also quantifying expressions such as “always”, “twice”, “never” and “more than three times”, which refer to sets of events. These expressions say something about the absolute or relative cardinality of a set of events of a certain type. Similarly for expressions of frequency, as in “I will call you twice every day”. Adverbs, such as adverbs of manner, often express a property of events. Observations like these led Davidson (1967) to propose to treat events as individual objects.

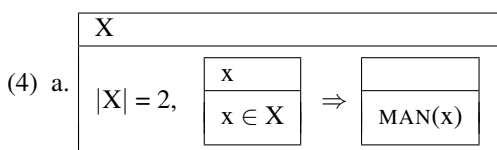
Following Parsons (1990) this view can be expressed in semantic representations by means of one-place predicates applied to existentially quantified event variables, using semantic role relations to indicate the roles of the participants in an event. This approach has been widely accepted in modern semantics, and has been adopted in the ISO annotation standards 24617-1 (Time and events), 24617-4 (Semantic roles), and 24617-7 (Spatial information).

For the semantic annotation of quantification, we propose an approach that combines GQT with the neo-Davidsonian treatment of predicate-argument relations, including the use of semantic roles as defined in ISO 24617-4. This approach allows the expression in annotations of quantification aspects such as the collective/individual distinction as a property of the way in which a set of individuals participates in a set of events. For example, the ISO 24617-4 annotation<sup>1</sup> of example sentence (2) would look as in (3).

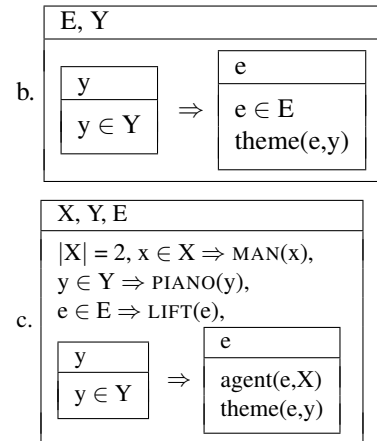
(2) Two men lifted the piano.

(3) (Markables: m1=“Two men”, m2=“lifted, m3=“the piano”.)  
 <entity xml:id=“x1” target=“#m1” pred=“man” involvement=“2” definiteness=“indef”/>>  
 <event xml:id=“e1” target=“#m2” pred=“lift”/>>  
 <srLink event=“#e1” participant=“#x1” semRole=“agent”>  
 <entity xml:id=“x2” target=“#m3” type=“piano” involvement=“1” definiteness=“def”/>>  
 <srLink event=“#e1” participant=“#x2” semRole=“theme”/>>

This theoretical basis is also brought out in the semantics of the annotations, which makes use of Discourse Representation Structures (DRSs) that involve sets of events with sets of participants. For example, the annotation of the NP “Two men” is interpreted as the DRS in (4a), which can be read as follows: *There is a set X of cardinality 2 that consists of men.* A semantic role link, like the one for the theme role with individual distributivity, is interpreted as the DRS in (4b), and the sentence “Two men lifted a piano” is interpreted as the DRS (4c), obtained by combining the DRSs for the NPs, the verb, and the semantic role relations.



<sup>1</sup>The <entity> element of ISO 24617-4 has been extended here with the attributes @involvement and @definiteness, and the attribute @entityType has been renamed @pred.



### 2.3. Principles of Semantic Annotation

A third pillar of the approach to quantification annotation proposed in this paper is formed by the ISO principles of semantic annotation (ISO standard 24617-6; see also (Bunt, 2015) and (Pustejovsky et al., 2011), which require an annotation scheme to have a three-part definition consisting of (1) an abstract syntax that specifies the possible annotation structures in set-theoretical terms, such as pairs and triples of concepts; (2) a concrete syntax, that specifies a representation format of annotation structures as XML expressions; (3) a semantics that specifies the meaning of annotation structures. This formal definition is supported by a metamodel that captures the fundamental concepts used in annotations and the way they are related. This organization ensures that semantic annotations have a semantics, as required by ISO 24617-6 (the principle of ‘semantic adequacy’), and by defining the semantics at the level of the abstract syntax it puts the focus of an annotation standard at the conceptual level, rather than at the level of concrete representations, as required by the ISO Linguistic Annotation Framework (ISO 24612, see also (Ide and Romary, 2004). Annotators have to deal with the concrete representations only, but they can rely on the existence of an underlying abstract syntax and semantics, which can be generated automatically from the XML representations.

A systematic specification of the semantics of annotation structures is hardly possible if it cannot be done in a *compositional* way, i.e., the semantic interpretation of an annotation structure is obtained by combining the interpretations of its components. Two kinds of components are distinguished: ‘entity structures’ and ‘link structures’. An entity structure specifies certain semantic information about a markable, a link structure specifies certain semantic information about the way two or more entity structures are semantically related. For example, the annotation of an NP corresponds to an entity structure, and scope relations correspond to link structures (see below, Section 5).

## 3. Related Work

ISO-TimeML (ISO 24617-1) has certain limited provisions for dealing with time-related quantification. For example, a temporal quantifier like “weekly” is represented as follows, where “P7D” stands for “period of seven days”:

(5) <TIMEX3 xml:id="t5" target="#token0" type="SET" value="P7D" quant="EVERY"/>

Here, @quant is one of the attributes of temporal entities, used to indicate that the entity is involved in a quantification. The limitations of this approach for annotating temporal quantification have been discussed by Bunt and Pustejovsky (2010), and improvements have been suggested by Lee and Bunt (2012).

ISO-Space (ISO 24617-7) uses the @quant attribute as well, applying it to spatial entities, and in addition uses the attribute @scopes to specify a scoping relation. The following example, taken from ISO 24617-7:2014, illustrates this:

(6) a. There's a computer<sub>se1</sub> on<sub>ss1</sub> every desk<sub>se2</sub>.  
 b. <spatialEntity id="se1" target=""#token2" form="nom" countable="true" quant="1" scopes=""#""/>  
 <spatialEntity id="se2" target="#token5" form="nom" pred="desk" countable="true" quant="every" scopes="#se1"/>  
 <spatialSignal id="ss1" target="#token3" type="dirTop" />  
 <qsLink id="qs1" relType="EC" figure="#se1" ground="#se2" trigger="#ss1"/>  
 <oLink id="o1" relType="above" figure="#se1" ground="#se2" trigger="#ss1" frameType="intrinsic" referencePt="#se2" projective="false" />

From a semantic point of view, this use of the @scopes attribute is not very satisfactory since the relative scoping of quantifications over different sets of entities is not a local property of one of these quantifications; therefore an annotation such as (6) does not have a compositional semantics. Instead, we propose to use a link structure to represent scope relations among quantifying NPS, which would come down to adding an element like the following:

(7) <scopeLink arg1="#se2" arg2="#se1" relType="wider"/>

Indirectly related to the definition of an annotation scheme for quantification is the Groningen Meaning Bank project (Bos et al., 2017), which is developing a resource consisting of sentences paired with DRSs that represent their meanings. This work cannot be compared directly with the usual kind of annotation work, which associates pieces of semantic information with markables like individual words and small stretches of text, whereas in the Groningen Meaning Bank DRSs are associated with full sentences. It may however be interesting to compare these DRSs with those that come out of the compositional interpretation of annotations as proposed here.

## 4. Aspects of Quantification

### 4.1. Restrictors

A natural language quantifier, a noun phrase (NP), consists in its full form of three parts: (1) the head noun; (2) a sequence of determiners; and (3) pre-nominal or post-nominal adjectives, prepositional phrases (PPs), relative

clauses and other modifiers of the head noun. The head noun together with its modifiers is called the *restrictor* of the quantifier, and indicates a domain that is considered in the quantification, the *'source domain'*. Quantification in natural language is nearly always restricted to a contextually determined part of the source domain, called the *'reference domain'* or *'context set'* (Westerståhl, 1985). For example, the quantifier *"all the students"* in (8) does not apply to every student, but only to those present at some meeting or performance.

(8) All the students applauded.

The definiteness of an NP is an indication that the domain of a quantification is restricted to a certain reference domain, rather than to its source domain; definiteness is therefore an item of quantification information that should be annotated, as illustrated above in (3).

### 4.2. Scope

Relative scope is one of the most studied aspects of quantification in natural language – see e.g. (Cooper, 1983); (Kamp and Reyle, 1993; Szabolcsi, 2010); (Ruys and Winter, 2011). Studies of scope have focused almost exclusively on the relative scopes of sets of participants, as in the classical example *"Everybody in this room speaks two languages"*. Not only the relative scoping of sets of quantified *participants* is a semantically important issue, but also the relative scoping of participants and *events*. This is illustrated by the two possible readings of the sentence in (9):

(9) All the students protested

On one reading, each of the students protested, for instance by sending a letter of protest; on another, all students participated in a single protest event, such as a demonstration. (Note that the latter interpretation involves the consideration of events in which multiple participants occupy the same role. Several approaches, such as those of the VerbNet and PropBank frameworks allow only a single occupant for each semantic role; the ISO approach to semantic role annotation (ISO 24617-4), does allow multiple participants with the same semantic role.)

The relative scopes of events and participants can be marked up on the link structure that expresses the participation in a set of events, as shown in (10) with the attribute 'eventScope' added to the <srLink> element.

(10) (Markables: m1="All the students", m2="protested")  
 <entity xml:id="x1" target="#m1" pred="student" involvement="all" definiteness="def"/>  
 <event xml:id="e1" target="#m2" pred="protest"/>  
 <srLink event="#e1" participant="#x1" semRole="agent" eventScope="wide"/>

In logic it is customary to assume that the relative scopes in a sequence of quantifiers are linearly ordered (but see e.g. Hintikka, 1973 on 'branching quantifiers'); in natural language sentences it may happen that two or more quantifications have equal scope, see e.g. (11):

(11) Three breweries supplied five inns

The intended reading here is that in total three breweries supplied in total five inns. In this total-total, or *cumulative* reading (Scha, 1981) the two quantifications have equal scope; the two cardinal determiners both indicate the amount of involvement of the respective reference domains in the predication. This can be represented in annotations as follows:

(12) `<scopeLink arg1="#se2 arg2="#se1 relType="equal"/>`

The sentence in (13a) has the same syntactic form as the one in (11), but here the intended reading is not cumulative; it is from a report about a tournament of (European) football where teams of boys and teams of girls participated, and whenever a team of boys played against a team of girls, its size would be reduced from 11 to 7. So the two cardinal determiners are indicators not of reference domain involvement or of scoped involvement of subsets of the reference domain, but rather of group size associated with the participation of groups of boys and girls. This can be accounted for in annotations by introducing a “group” value for the @distr attribute of a participation link, as shown in (13b). Semantically, an annotation with such a distributivity will be interpreted as describing a quantification over groups of seven boys and eleven girls.

- (13) a. Seven boys played against eleven girls.
- b. (Markables: m1=“Seven boys”, m2=“played against”, m3=“eleven girls”)
- ```

<entity xml:id="x1" target="#m1" pred="boy"
  involvement="7" definiteness="indef"/>
<event xml:id="e1" target="#m2" pred="play"/>
<srLink event="#e1" participant="#x1"
  semRole="agent" distr="group"/>
<entity xml:id="x1" target="#m3" pred="girl"
  involvement="11" definiteness="indef"/>
<srLink event="#e1" participant="#x2"
  semRole="agent" distr="group"/>

```

### 4.3. Distributivity

Distributivity comes in an obvious form in the distinction between individual (or ‘distributive’) and collective participation; the group distribution illustrated by (13) forms another case. Yet another form of distributivity occurs in (14), where the three boys involved did not necessarily do all the carrying either collectively or individually, but where they may have carried some heavy boxes collectively and some lighter boxes individually:

(14) The boys carried all the boxes upstairs.

The quantifications in this sentence over sets of boys and sets of boxes have ‘unspecific’ distributivity (Bunt, 1985); the sentence just says that all the boxes were *somehow* carried upstairs by the boys. Following (Kamp and Reyle, 1993), we use the notation  $X^*$  to designate the set consisting of the members of  $X$  and the subsets of  $X$ , and  $P^*$  to designate the characteristic function of the set  $X^*$ , where  $P$  is the characteristic function of  $X$ . Using moreover the notation  $R_C$  to indicate the characteristic function of a reference domain that forms a subset of a source domain with

characteristic function  $R$ , the interpretation of (14) can be represented in second-order predicate logic as follows:

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

Note that 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 was already illustrated above by example (1c), assuming that “*the men*” individually had a beer, and collectively carried the piano upstairs. Distributivity should thus be marked up on the participation relation in the drinking and carrying events.

- (16) (Markables: m1=“The men”, m2=“had”, m3=“a beer”, m4=“carrying” upstairs, m5=“the piano”)
- ```

<entity xml:id="x1" target="#m1" pred="man"
  involvement=">1" definiteness="def"/>
<event xml:id="e1" target="#m2" pred="drink"/>
<entity xml:id="x2" target="#m3" pred="beer"
  involvement="1" definiteness="indef"/>
<event xml:id="e2" target="#m4" pred="carry"/>
<entity xml:id="x3" target="#m5" pred="piano"
  involvement="1" definiteness="def"/>
<srLink event="#e1" participant="#x1" semRole="agent"
  distr="individual"/>
<srLink event="#e2" participant="#x1" semRole="agent"
  distr="collective"/>

```

### 4.4. Involvement and size/cardinality

The prenominal part of an NP may contain a sequence of determiners of different type. Grammars commonly distinguish different classes of determiners, with different possible sequencing and co-occurrence restrictions. For example, in English grammar it is customary to distinguish between predeterminers, central determiners, and postdeterminers (e.g. Quirk et al., 1972), each of which has a different function:

- predeterminers express the (absolute or proportional) quantitative involvement of the reference domain, and may, additionally, provide information about the distribution of a quantification over the reference domain;
- 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 (for count NPs) or its size (for mass NPs).

This is illustrated by the NP “*All my nine grandchildren*”, where “*all*” is a predeterminer, “*my*” a central determiner, and “*nine*” a postdeterminer. Note also that the value of an @involvement attribute in an entity structure that has widest scope represents the total involvement of the reference domain, but in an entity structure within the scope of another, it represents a ‘scoped’ involvement, like the determiner “*two*” in “*Everybody in this room speaks two languages*”.

<i>involv.</i>	<i>distributivity</i>	<i>interpretation</i>	<i>example</i>
all	homogeneous	For all quantities of $M$	(21a)
total	unspecific	For the elements in a set of quantities of $M$ that together make up the whole of $M$	(21c)
all	collective	For $M$ as a whole	(21b)

Table 1: Involvement and distributivity in mass NP quantification.

#### 4.5. Quantification in modifications

The restrictor part in a full-fledged NP may be simply a noun, but in general may contain adjectives and other expressions that modify the noun, such as other nouns, as in “*waste dump*”, prepositional phrases, or relative clauses. Moreover, conjunctions of nouns (possibly with modifications) may add further complexity to restrictors.

Head noun modifications bring certain issues of quantification, such as scope and distributivity, e.g. the restrictor “*heavy books*” in the sentence “*Peter carried some heavy books*” may be interpreted as referring to certain books that are heavy each (individual reading) or to a heavy pile of books (collective reading). To express this in annotations, an <adNLink> structure is introduced with the attribute @distr:

```
(17) (Markables: m1="heavy", "books")
    <entity id="x1" target="#m2" pred="book"/>
    <entity id="x2" target="#m1" pred="heavy"/>
    <adNLink head="#x1" mod="#x2" distr="collective"/>
```

An adjective can be used not only for modifying an NP head noun, but also for modifying a noun that modifies another noun, as in “(*toxic waste*) *dump*” or “(*natural language*) *processing*”. The QuantML representation (18) shows how the adjective scope in “(*toxic waste*) *dump*” can be indicated (see also example (22)).

```
(18) (Markables: m1="toxic", m2="toxic waste",
    m3="toxic waste dump", m4="waste", m5="dump")
    <qDomain xml:id="x1" target="#m3"
    source="#x2" restrictions="#r1"/>
    <sourceDomain xml:id="x2" target="#m5"
    pred="dump" />
    <nnLink xml:id="r1" target="#m2" pred="waste"
    restr="#r2" />
    <adNLink xml:id="r2" target="#m1" pred="toxic"
    distr="individual"/>
```

Quantifier scope issues arise when a head noun is modified by a PP or a restrictive clause, as in (19), where a quantification inside the PP takes scope over the one of the head noun.

(19) The committee spoke with two students from every university.

This phenomenon is known as ‘inverse linking’ (May, 1977; May and Bale, 2005; Szabolcsi, 2010; Ruys and Winter, 2011; Barker, 2014). Inverse linking in PP modification is widespread; especially the case of an existentially quantified main NP and universally quantified PP, as in (19), is quite common.

To capture the relevant information related to quantification within a complex restrictor, the annotation of complex

restrictors needs to be articulated in marking up the head noun that is central to the restrictor, and the various possible modifiers with indications of their distributivity and possible scope inversion.

#### 4.6. Mass Quantifiers

Most studies of quantification in natural language have been restricted to cases where the NP head is a count noun. Quantification by NPs where the head is a mass noun are in many ways similar; compare, for example for example, (20a) and (20b) :

- (20) a. The salesmen sold all the cars.  
b. The vendors sold all the ice cream.

In (20a) a predicate is applied to sets of salesmen and cars, and in (20b) to sets of vendors and quantities of ice cream. A difference is that (20a) can be analysed as: “Every one of the cars was the object in a sell-event with one of the salesmen as the agent”, but it is not clear that the analogous analysis where every quantity of ice cream was the object in a sell-event would make sense. A universal quantification like “*all the ice cream*” does not necessarily refer to *all* the quantities of ice cream that the vendors possessed, but rather to a certain subset of quantities that has the property of together making up the whole of the vendors’ ice cream. Such a situation commonly arises for mass NP quantifiers, and may arise also for count NPs in case the individuals in the quantification domain have an internal part-whole structure, as in “*The boys ate all the pizzas*”.

A detailed analysis of quantification in relation to mass terms can be found in (Bunt, 1985), which analyses the notion ‘quantity of’ as a part-whole relation, and defines a ‘merge’ operation  $\cup$  on quantities such that the merge of two or more quantities of  $M$  is again a quantity of  $M$ . An expression of the form ‘all the  $M$ ’ with a mass noun  $M$  can be interpreted as referring to a set of quantities of  $M$  whose merge forms the whole of all  $M$ .

Properties of mass quantification, like those of count NP quantification, are distributivity, scope, definiteness, domain involvement, and size of the reference domain (or of parts of it), but there are some notable differences in distributivity and in the expression of involvement and size.

Since mass nouns do not individuate their reference, quantification by mass NPs would seem not to allow individual distribution. Yet there is a distinction somewhat similar to the individual/collective distinction of count NP quantifiers, as illustrated by (21).

- (21) a. The water in these lakes is polluted.  
b. The sand in the truck weighs 20 tons.  
c. The boys carried all the sand to the back yard.

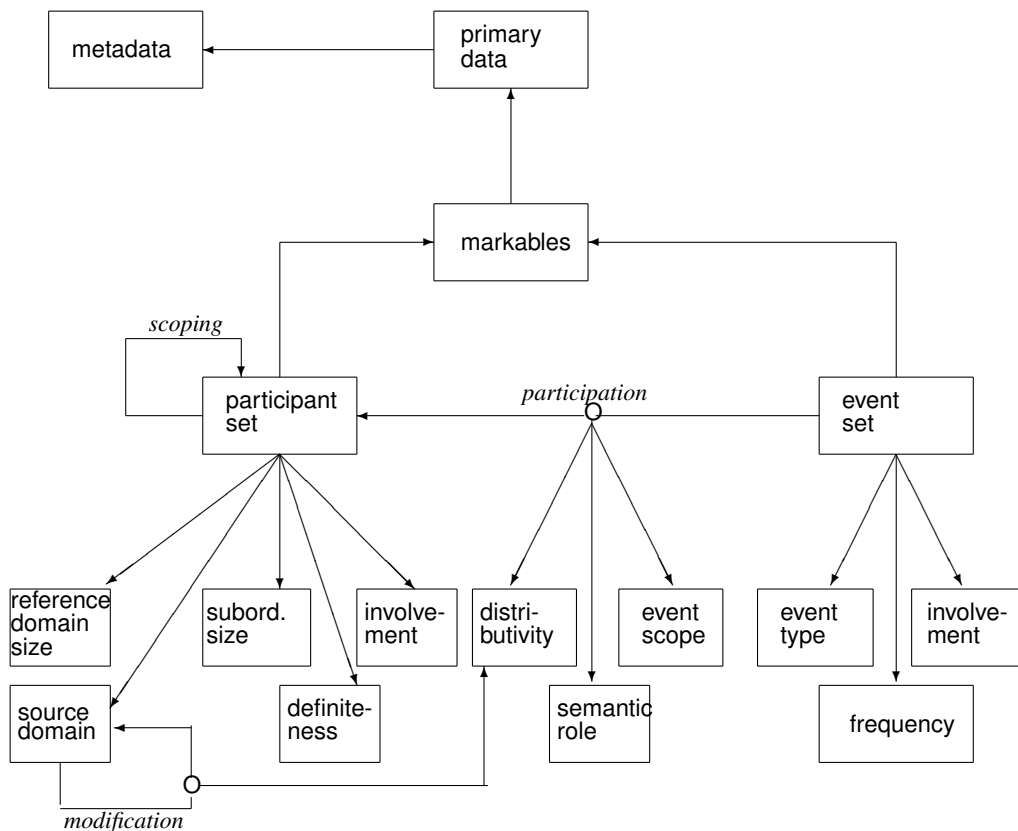


Figure 1: Metamodel for the annotation of quantification

The quantification in (21b) is collective, as the weight applies to the whole formed by all the quantities of sand. In (21a) the predicate of being polluted applies to any sample of “the water in the lake”; this distribution is called *homogeneous*. In (21c) the boys carried certain quantities of sand that together make up “all the sand”; in this case the distribution is called *unspecific*, as this case is rather similar to the count NP case where both individuals, sets of individuals, and parts of individuals may be involved.

Expressions of proportional involvement, like “some pasta”, “most of the pasta”, “all the pasta” cannot be interpreted in terms of number of quantities. As the examples in (21) illustrate, complete involvement of a mass NP reference domain means that the merge of the quantities involved forms the entire domain: if  $Q_M$  is the set of all quantities of  $M$ , and  $X_M$  is the set of quantities of  $M$  involved in the quantifying predication, then  $\cup X_M = \cup Q_M$ . Non-zero involvement means that the merge of the quantities in  $M$  has non-zero size, and a “most  $M$ ” quantification over reference domain  $M$  means that, for any unit of measurement  $u$ ,  $|\cup X|_u > |M|_u/2$ , where  $|\cdot|_u$  indicates size measured in units  $u$ . (See (Bunt, 1985) for a calculus of size measurement.)

The examples in (21) 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*) the  $M$  in English, and similarly in other languages. In (21a), “The water” refers to the set of all (contextually relevant) quantities of water; this involvement will be indicated in annotations as “all”. In (21c), “all the sand” refers to a subset of quantities of sand that together make up all the (contextu-

ally relevant) sand; this involvement is indicated as “total” in annotations of reference domain involvement. Finally, in (21b), “The sand” refers to the single quantity of sand formed by all contextually relevant quantities of sand. This involvement will be annotated as “all”, just as in the case of collective count NP quantification. Table 1 summarizes these possible annotation choices.

## 5. QuantML Annotation Scheme

### 5.1. ISO scheme organization

An ISO standard annotation scheme for quantification should fit within the series of semantic annotation standards known collectively as the Semantic Annotation Framework (SemAF), ISO 24617. It should as such be compatible with the existing parts of SemAF: Part 1, Time and events; Part 2, Dialogue acts; Part 4: Semantic roles; Part 7: Spatial information, and Part 8: Discourse relations. Moreover, it should be defined according to the ISO Principles for semantic annotation, as mentioned in Section 2.3. This means in particular that the definition of the annotation scheme includes an abstract syntax, a concrete syntax, and a semantics, supported by a conceptual metamodel. Figure 1 shows the proposed metamodel, reflecting the conceptual analysis of quantification in the previous sections.

Specifying the semantics of annotations at the level of abstract syntax ensures that changes in representation format do not need to require adaptations in the semantics. The ISO principles require representation formats to be ‘ideal’, i.e. (1) complete, in the sense that every annotation structure defined by the abstract syntax has a representation; (2) unambiguous, i.e. every representation is a rendering of only one annotation structure defined by the abstract syntax

- (22) a. Alex owns some (valuable ancient (Chinese books and Japanese paintings)).  
 b. Markables: m1=Alex, m2=owns, m3=some valuable ancient Chinese books and Japanese paintings, m4= valuable, m5=valuable ancient Chinese books and paintings, m6=ancient, m7=Chinese, m8=Chinese books, m9=books, m10=Japanese, m11=Japanese paintings, m12=paintings

c. QuantML Representation:

```
<entity xml:id="x1" target="#m1" domain="#x1" involvement="1" definiteness="def"/>
<sourceDomain xml:id="x1" target="#m1" pred="alex"/>
<event xml:id="e1" target="#m2" pred="own"/>
<entity xml:id="x2" target="#m3" domain="#x3" involvement="some" definiteness="indef"/>
<qDomain xml:id="x3" target="#m5" source="#x4 #x6" restrictions="#r1 #r2"/>
<qDomain xml:id="x4" target="#m8" source="#x5" restrictions="#r3"/>
<sourceDomain xml:id="x5" target="#m9" pred="book"/>
<qDomain xml:id="x6" target="#m11" source="#x7" restrictions="#r4"/>
<sourceDomain xml:id="x7" target="#m12" pred="painting"/>
<adNLink xml:id="r1" target="#m4" pred="valuable" distr="individual"/>
<adNLink xml:id="r2" target="#m6" pred="ancient" distr="individual"/>
<adNLink xml:id="r3" target="#m7" pred="chinese" distr="individual"/>
<adNLink xml:id="r4" target="#m10" pred="japanese" distr="individual"/>
<participationLink event="#e1" participant="#x1" semRole="theme" distr="individual" eventScope="narrow"/>
```

(see Bunt, 2010). This approach supports the design of alternative user-friendly representations, allowing for example to use tabular forms or other formats that are more convenient for human annotators and researchers than XML representations. This has been exploited in the DialogBank (Bunt et al., 2016), a resource of dialogues annotated according to the ISO 24617-2 annotation scheme, with alternative representations and the possibility to convert from one representation to another.

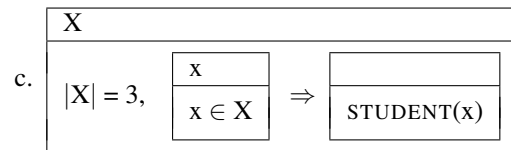
### 5.2. QuantML Abstract Syntax, Concrete Syntax, and Semantics

The annotation structures defined by the QuantML abstract syntax consist of entity structures and link structures. The concrete syntax specifies a pivot XML format for representing these structures. The examples earlier in this paper gave a slightly simplified impression of these representations. The example in Fig. 2 provides more detail in case the quantifier has a structured source domain, caused by the NP head being a modified conjunction of two modified nouns. (The `<srLink>` used in the examples above has been replaced here by the QuantML element `<participationLink>`.) The semantics specifies an interpretation-by-translation by means of a compositional interpretation function that defines DRSs for all well-formed annotation structures; compositional in the sense that the DRS expressing the meaning of an annotation structure is constructed out of the DRSs interpreting the component entity and link structures.

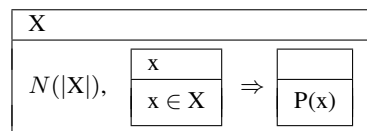
Example (23) shows the abstract entity structure for the quantifier “Three students” with its concrete XML representation and its semantics. The abstract annotation structure pairs the markable for an occurrence of the expression “three students” with a quadruple of concepts as shown in (23a); a concrete representation in XML may look as in (23b); and the semantic interpretation as shown in (23c), obtained by applying the interpretation function as partly defined by (24), where ‘P’ stands for the characteristic predicate of a source domain, and *N* is a predicate that ex-

presses reference domain involvement.

- (23) a.  $\langle \text{man, some, indef, } \lambda z. |z| = 3 \rangle$   
 b. `<entity xml:id="x1" target="#m1" type="student" involvement="3"/>`



- (24)  $I_A(\langle P, 3, \text{indef}, C \rangle) =$



## 6. Conclusions and Further Work

In this paper we have outlined a theoretically well-founded approach to the annotation of quantification in natural language which observes the most important requirements for establishing an ISO annotation standard. These requirements concern:

1. theoretical adequacy, taking well-established results of studies in logic, linguistics, and computation into account (notably those of Generalized Quantifier Theory, of neo-Davidsonian event-based semantics, and the ISO Principles of semantic annotation);
2. empirical adequacy, providing a wide coverage of quantification phenomena;
3. semantic adequacy, in specifying a well-defined semantics of annotations;
4. compliance with the ISO Linguistic Annotation Framework<sup>2</sup> and the ISO Principles of semantic anno-

<sup>2</sup>ISO 24612; cf. Ide and Romary, 2004

tation<sup>3</sup>, resulting in the separation of the abstract and concrete syntax of annotations, with a compositional semantics that provides interpretations in the form of DRSs for the structures of the abstract syntax;

5. compatibility with existing standards for the semantic annotation of time and events, spatial information, and semantic roles.

The paper specifies a metamodel that specifies a number of concepts relating to properties of quantification in natural language (including quantification phenomena within head noun modifications) that have to be taken into account in an adequate annotation scheme, including those that occur in noun modification structures with quantifier restrictors, such as the distributivity of adjectival modification and the relative scoping of quantifiers when inverse linking occurs in modification by preposition phrases.

There is still work to be done: further development of the approach outlined here will try to incorporate solutions for some of the well-known semantically hard puzzles relating to quantification, such as the treatment of reciprocals, reflexives, generics and habituals.

## 7. Bibliographical References

- Barker, C. (2014). Scope. In S. Lappin and C. Fox, editors, *The Handbook of Contemporary Semantic Theory*, pages 40–76. John Wiley.
- Barwise, J. and Cooper, R. (1981). Generalized quantifiers and natural language. *Linguistics and Philosophy*, pages 159–219.
- Bos, J., Basile, V., Evang, K., Venhuizen, N., and Bjerva, J. (2017). The Groningen Meaning Bank. In Nancy Ide and James Pustejovsky, editors, *Handbook of Linguistic Annotation*, pages 463–496. Springer, Berlin.
- Bunt, H. and Pustejovsky, J. (2010). Annotating temporal and event quantification. In *Proceedings ISA-5, Fifth International Workshop on Interoperable Semantic Annotation*, pages 15–22. City University of Hong Kong.
- Bunt, H., Petukhova, V., Malchanau, A., Fang, A., and Wijnhoven, K. (2016). The DialogBank. In *Proceedings 10th International Conference on Language Resources and Evaluation (LREC 2016)*, Portoroz.
- Bunt, H. (1985). *Mass terms and model-theoretic semantics*. Cambridge University Press, Cambridge, UK.
- Bunt, H. (2010). A methodology for designing semantic annotation languages exploiting syntactic-semantic isomorphisms. In Alex Fang, Nancy Ide, and Jonathan Webster, editors, *Proceedings of ICGL 2010, Second International Conference on Global Interoperability for Language Resources*, pages 29–45. City University of Hong Kong.
- Bunt, H. (2015). On the principles of semantic annotation. In *Proceedings 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-11)*, London.
- Cooper, R. (1983). *Quantification and syntactic theory*. Reidel, Dordrecht.
- Davidson, D. (1967). The logical form of action sentences. In N. Rescher, editor, *The Logic of Decision and Action*. University of Pittsburgh Press.
- Hintikka, J. (1973). Quantifiers vs. quantification theory. *Dialectica*, 27:329–358.
- Ide, N. and Romary, L. (2004). International standard for a Linguistic Annotation Framework. *Natural Language Engineering*, 10:221–225.
- ISO. (2010). *ISO 24612: Language resource management: Linguistic Annotation Framework (LAF)*. International Organisation for Standardisation ISO, Geneva.
- Kamp, H. and Reyle, U. (1993). *From Discourse to Logic*. Kluwer Academic Publishers, Dordrecht.
- Lee, K. and Bunt, H. (2012). Counting time and events. In *Proceedings 8th Joint ISO-ACL Workshop on Interoperable Semantic Annotation (ISA-8)*. ILC-CNR, Pisa.
- Lindström, P. (1966). First order predicate logic with generalized quantifiers. *Theoria*, 32:186–195.
- May, R. and Bale, A. (2005). Inverse linking. In Martin Everaert and Henk van Riemsdijk, editors, *The Blackwell Companion to syntax, Vol. 2*, pages 639–667. Blackwell, Oxford.
- May, R. (1977). *The Grammar of Quantification*. Ph.D. Dissertation, MIT.
- 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*. MIT Press, Cambridge, MA.
- 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*, pages 241–158, Oxford. Blackwell 2004.
- Pustejovsky, J. and Bunt, H., and Zaenen, A. (201). Designing annotation schemes: From theory to model. In Nancy Ide and James Pustejovsky, editors, *Handbook of Linguistic Annotation*. Springer, Berlin.
- Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1972). *A grammar of contemporary English*. Longman, London.
- Ruys, E. and Winter, Y. (2011). Scope ambiguities in formal linguistics. In D. Gabbay et al., editors, *Handbook of Philosophical Logic*. Foris, Dordrecht.
- Scha, R. (1981). Collective, distributive and cumulative quantification. In J. Groenendijk et al., editors, *Formal Methods in the Study of Language*, pages 483–512. Mathematisch Centrum, Amsterdam.
- Szabolcsi, A. (2010). *Quantification*. Cambridge University Press, Cambridge, UK.
- Westerståhl, D. (1985). Determiners and context sets. In Johan van Benthem and Alice Ter Meulen, editors, *Generalized Quantifiers in Natural Language*, pages 45–71. Foris, Dordrecht.

<sup>3</sup>ISO 24617-6; cf. Bunt, 2015