International Standard

ISO 24617-12

First edition
2025-01

Language resource management — Semantic annotation framework (SemAF) —

Part 12: **Quantification**

Gestion des ressources linguistiques — Cadre d'annotation sémantique (SemAF) —

Partie 12: Quantification

Reference number
ISO 24617-12:2025(en)

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 37, *Language and terminology*, Subcommittee SC 4, *Language resource management*.

A list of all parts in the ISO 24617 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

This document is an addition to the ISO 24617 series of standards for annotating various types of semantic phenomena in natural language. Quantification phenomena are particularly important since quantifications occur in every sentence in every language, except in trivial sentences such as “It is raining” in English, “det regner” in Danish or “Llueve” in Spanish. Quantification phenomena are an essential component for the understanding of spoken and textual language and multimodal messages. Annotating such phenomena in an interoperable way improves the re-usability of language resources as a basis for understanding-based applications of language technology, such as factually and contextually reliable information extraction and question answering in human-computer dialogue.

The content of this document builds on earlier studies of aspects and annotation of quantification phenomena, in particular References [3] and [5]. Based on these and other previous studies, this document specifies an annotation scheme with a markup language, called QuantML, which allows a synthesized way of treating a range of quantification phenomena.

This document provides support for the annotation of quantification phenomena in accordance with the principles of semantic annotation laid down in ISO 24617-6, and in a way that is consistent with existing and developing standards for the annotation of semantic information within the ISO semantic annotation framework (SemAF, the ISO 24617 series).

NOTE The explanatory repository of annotated quantification phenomena in the Quantification Bank (see Reference [37]), maintained at Tilburg University, provides background information about the basic concepts in quantification annotation, plus a collection of annotated examples.

Language resource management — Semantic annotation framework (SemAF) —

Part 12:
Quantification

1 Scope

This document specifies a markup language called QuantML for annotating and representing semantic phenomena relating to quantification in natural language. QuantML comprises an extensible markup language (XML)-based representation format, an abstract syntax and a semantics.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
definiteness
language-dependent morphosyntactic feature of a *noun phrase (NP)* (3.12), marked in English and other European languages by a definite or indefinite article or a nominal suffix, by a demonstrative, or by a possessive expression

Note 1 to entry: The definiteness feature has two possible values: "definite" and "indefinite". Being definite is often regarded as an indication of determinacy, indefinite as an indication of indeterminacy.

Note 2 to entry: In some languages it is only possible to express that a NP is definite (NPs are by default indefinite) or to express that an NP is indefinite (NPs are by default definite).

EXAMPLE al (definite article in Arabic languages), -e (suffix as definite article in Farsi), el/la (definite article in Spanish), a/az (definite article in Hungarian, there is no indefinite article), yī (occasionally indefinite article in Chinese; there is no definite article and the definiteness is definite unless an indefinite article or the context indicates otherwise).

Note 3 to entry: For overviews of definite expressions, see References [1] and [44].

3.2
definite description
singular noun phrase with *definiteness* (3.1) 'definite', interpreted as referring to a (contextually) uniquely determined entity

EXAMPLE Jimmy, the chairperson, my house, this idea.

3.3
determinacy
semantic property of referring to some particular and determinate entity or collection of entities

Note 1 to entry: Determinacy can be interpreted as specifying the relation between the *reference domain* (3.16) and the *source domain* (3.18) of a quantification. The reference domain of a determinate quantification is a proper subset of the source domain; for an indeterminate quantification the reference domain coincides with the source domain.

Note 2 to entry: Determinacy and *definiteness* (3.1) are not always clearly distinguished in the linguistic literature. For a discussion of this issue, see Reference [9].

3.4
distributivity
distribution
specification of whether the entities of the *reference domain* (3.16) of a *quantification* (3.15) are individually involved, or as a group (collectively), or as a mixture of the two

Note 1 to entry: Distributivity can be expressed by adverbs, such as *"together", "ensemble"* (French) and "samen" (Dutch), or by certain determiners, such as *"each"* in English, *"chaque"* in French and *"jeder"* in German. Some determiners, such as the English *"each", "all"* and *"both"* can also be used as adverbs.

3.5
event
eventuality
something that can be said to obtain or hold true, to happen or occur

[SOURCE: ISO 24617-1:2012, 3.5, modified — Note 1 to entry deleted.]

3.6
event set
aspect of a *quantification* (3.15), specifying a set of *events* (3.5) in which the members of a certain *participant set* (3.14) are involved

3.7
exhaustivity
semantic property of a *quantification* (3.15), indicating that no other individuals than the elements of the *participant set* (3.14) are involved in elements of the *event set* (3.6)

3.8
genericity
specification of whether the sentence in which a *quantification* (3.15) occurs refers to a certain specific *event set* (3.6) and *participant set* (3.14) or expresses a general statement or question

3.9
individuation
semantic property of the way a nominal expression is used to refer to its denotation as a collection of individual entities, as parts of a homogenous mass, or as a collection of individual entities and their parts

Note 1 to entry: The distinction between referring to a collection of entities and referring to a part-whole structured domain is expressed in many languages by the distinction between count terms and *mass terms* (3.11).

3.10
inverse linking
modification of a *noun phrase head* (3.13) that contains a quantifier with wider scope than the *quantification* (3.15) of the noun phrase head

EXAMPLE Two students from every university participated in the meeting.

3.11
mass term
noun or nominal compound used in such a way that it does not individuate its reference

Note 1 to entry: Typical examples in English are *"footwear"*, *"water"*, *"cattle"*, *"music"*, *"luggage"* and *"furniture"*. By contrast, expressions such as *"shoe"*, *"drop of water"*, *"cow"*, *"sonata"*, *"suitcase"* and *"chair"* are typically used as count terms, i.e. in such a way that it is understood what counts as (for example) one shoe, as two shoes, etc. Some words are commonly used either way, such as "*rope*" and "*stone*". The two possible uses of nouns are also illustrated by: *"There's no chicken in the pen"/"There's no chicken in the stew."* See also Reference [6].

3.12
noun phrase
NP
group of words that function together syntactically as a noun

Note 1 to entry: An NP typically consist of a noun, one or more determiners, and head modifiers. Other cases include NPs consisting of a personal pronoun, a proper name or a conjunction of nouns instead of a single noun.

3.13
noun phrase head
head
noun or a conjunction of nouns that forms the central element of a *noun phrase* (3.12)

3.14
participant set
set of entities involved in the *event set* (3.6) of a *quantification* (3.15)

EXAMPLE The parents gave all the teachers a present.

3.15
quantification
application of a predicate to a set of entities

Note 1 to entry: A particularly important type of predicate in the context of this document is *involved in certain events in a certain semantic role*.

3.16
reference domain
contextually determined set of entities that a quantifying predicate is applied to

3.17
restrictor
part of a *noun phrase* (3.12) consisting of the *head* (3.13) and modifiers (if present)

3.18
source domain
explicitly mentioned maximal set of entities that a quantifying predicate is applicable to

Note 1 to entry: For a quantifier expressed by a noun phrase, the source domain is the extension of the *restrictor* (3.17). Adverbial temporal and spatial quantifiers have their source domains (temporal and spatial entities), specified as part of their lexical semantics.

4 Background

Quantification is linguistically, logically, and computationally highly complex, and has been studied for centuries by logicians, linguists, formal semanticists and computational linguists, from Aristotle to present-day scholars (see, for example, References [3], [4], [10], [11], [15], [17], [24], [25], [26], [30], [32], [34], [35], [42] and [43]).

Partly inspired by studies of quantification in logic, analyses of the way quantifiers are expressed in natural language has led to generalized quantifier theory (GQT) (see References [4], [5], and [26]). GQT interprets quantifiers as properties of a set of entities. Quantifying expressions in natural language are 'restricted' in the sense of containing an indication of the entities to which the quantification is meant to apply. Natural language quantifiers are thus not determiners such as *"all"* and *"some"*, but rather noun phrases (NPs) such as *"all students"*, *"some sonatas"*, *"quelques gens"* and *"mais que cinco melodias"*.

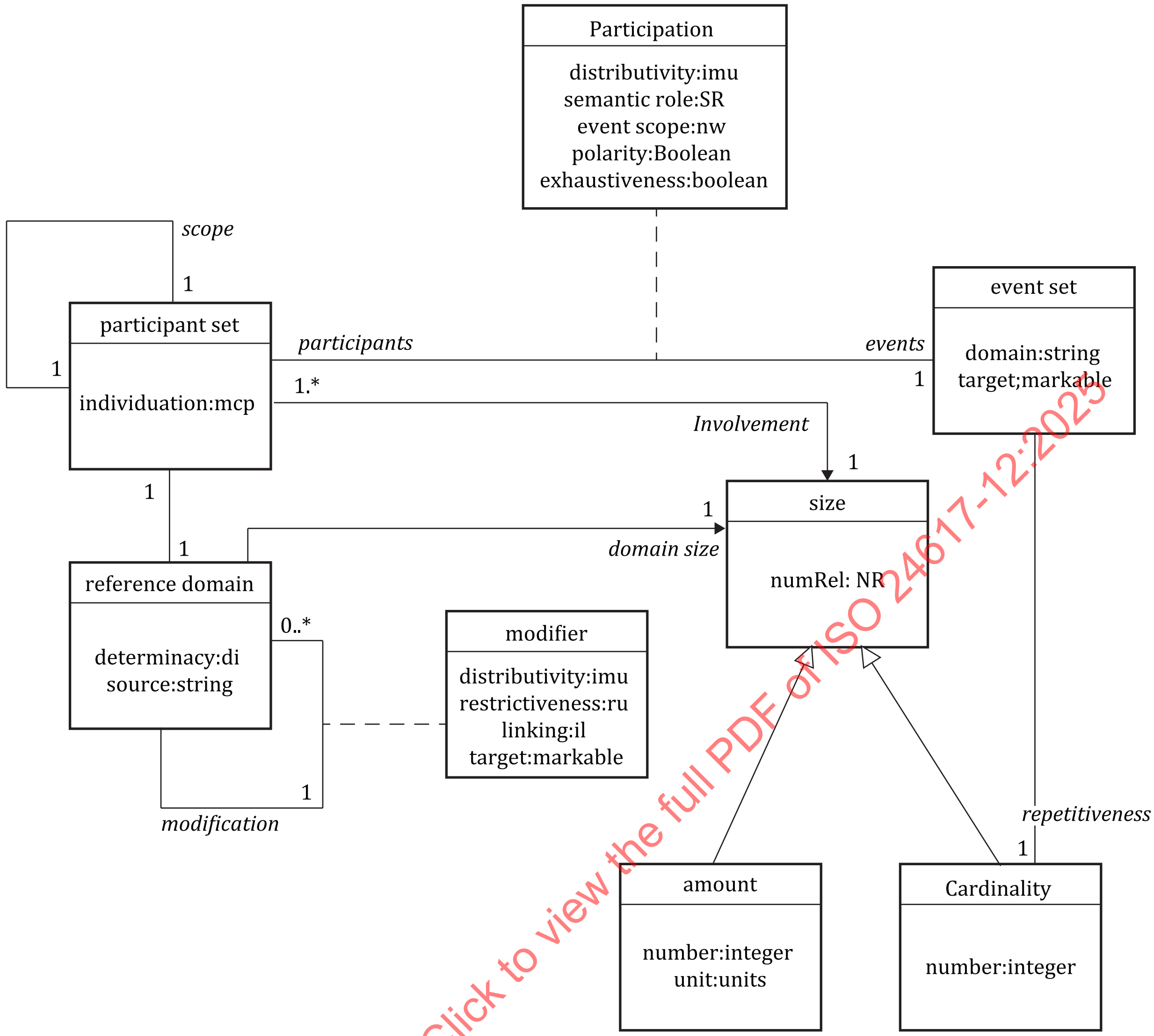
The annotation scheme defined in this document combines GQT with neo-Davidsonian event semantics,[13] [33] which views the combination of a verb and its arguments as the participation in a certain semantic role of the entities denoted by the argument in the events denoted by the verb. This approach is also used in other parts of the SemAF.

The scheme is designed according to the ISO principles of semantic annotation (see ISO 24617-6 and also References [7] and [39]). The QuantML markup language therefore has a triple-layered definition consisting of the following:

a) 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 Figure 1).

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

c) 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), which have a well-established model-theoretic semantics[24] and which are also used in other parts of the SemAF.



Key

imu	{individual, collective, unspecific}	di	{determinate, indeterminate}
SR	semantic role set	ru	{restrictive, unrestrictive}
nw	{narrow, wide}	NR	{greater, equal, less-or-equal}
mcp	{mass, count, count+parts}		

Figure 1 — QuantML metamodel for the annotation of quantification

5 Basic concepts

5.1 Aspects of quantification in natural language and their annotation

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

a) domain;

b) determinacy;

c) distributivity;

d) involvement;

e) individuation;

f) argument role;

g) exhaustivity;

h) polarity;

i) participant scope;

j) event scope;

k) repetitiveness;

l) domain size;

m) restrictiveness of modifiers;

n) linking of modifiers;

o) modality;

p) genericity.

These categories correspond to elements of annotations. The categories 1 to 11 correspond to 'core attributes', which require a value whenever a quantification is annotated. Some of these attributes are optional and have a default value. Additionally, QuantML has a number of attributes that are relevant only for certain forms of quantification. The attributes 12 to 14 exemplify this: they apply only in case a quantifying expression contains a specification of domain size or a modifier that can restrict the reference domain. The items 15 to 16 are exceptional in that their semantic interpretation is undefined; they have been included solely to allow corpus searches of instances of generic or modal quantification.

The QuantML metamodel, visualized in Figure 1, shows the roles of the categories 1 to 13 and the corresponding attributes in annotations. The metamodel clearly brings out that three components play centre stage in a QuantML annotation: events, participants and the participation relation that links them, each with a number of features corresponding to the information categories 1 to 13. This is illustrated by the annotation fragment in Example 2 in 5.2.

5.2 Quantification domains

NPs, expressing a generalized quantifier, typically consist of three parts:

a) a noun (the 'head');

b) one or more determiners such as *"a", "the", "all", "some"* and *"many";*

c) one or more adjectives, prepositional phrases, possessive phrases or other modifiers.

The head noun with its modifiers, the 'restrictor' of the quantifier, indicates a certain domain that the quantification ranges over. The term *source domain* is used to refer to the set of entities indicated by the restrictor. The domain that a quantification is intended to range over is often not the entire source domain, but a certain part of it, determined by the context. For instance, the sentence in Example 1 is not meant to put an obligation on every person, but only on the students in a certain class.

Example 1 Everybody must hand in his or her essay before Thursday next week.

This more limited domain is called the *reference domain* or 'context set'[16][43]. It is determined by the familiarity, salience, recent mention, physical presence, and other contextual considerations that make certain elements of the source domain stand out as the intended referents. The annotation fragment in Example 2 shows how this is annotated in QuantML.

Example 2 All the students protested.

Markables: m1 = "All the students", m2 = " the students", m3 = "students", m4 = "protested"

```
<entity xml:id="x1" target="#m1" refDomain=#x2" individuation="count"
        involvement="all"/>
<refDomain xml:id="x2" target="#m2" source="#x3" determinacy="det"/>
<sourceDomain xml:id="x3" target="#m3" pred="student"/>
<event xml:id="e1" target="#m4" pred="protest"/>
<participation event="#e1" participant="#x1" semRole="agent" distr="individual"/>
```

5.3 Determinacy

The determinacy of a quantification expresses whether the reference domain is a proper subset of the source domain or coincides with it. Determinacy is sometimes indicated by the morphosyntactic feature of definiteness, which in Germanic and Romance languages is marked by the use of a definite article or a nominal suffix, such as *"the book"* in English, and *"bogen"* in Danish.

NOTE See, for example, Reference [25] on the expression of definiteness in a large number of languages, and References [1] and [40] for overviews of definite expressions in English.

Definite plural NPs are most often determinate and indefinite plural NPs indeterminate, but there is no straightforward relation between definiteness and determinacy.[12] To mark up determinacy in QuantML, the attribute @determinacy in <entity> elements should be used and given either the value "det" or the value "indet".

5.4 Distributivity

The distributivity of a quantification expresses whether a predicate applies to a set of entities as a whole, or to its individual members, or to certain of its subsets. The collective/individual (or 'distributive') distinction is illustrated in Example 3.

Example 3 a) Two men carried a piano upstairs.
b) Two men carried some chairs upstairs.

Besides distributive and collective, QuantML also supports the annotation of distributivity as 'unspecific', meaning that individuals as well as sets of individuals can be involved. The sentence in Example 4, for instance, possibly describes a situation where the boys involved did not necessarily do all the carrying either collectively or individually, but where they carried some boxes collectively and some individually.

Example 4 The boys carried all the boxes upstairs.

Distributivity is a property of the way entities participate in events, and is annotated using the @distr attribute in <participation> elements. This is illustrated in Example 5 (slightly simplified), assuming that each of the men individually had a beer and collectively carried the piano upstairs.

Example 5 The men had a beer before carrying the piano upstairs.

Markables: m1 = "The men", m2 = "men", m3 = "had a beer", m4 = "carrying upstairs", m5 = "the piano", m6 = "piano"

```
<entity xml:id="x1" target="#m1" refDomain="#x2" individuation="count"
        involvement="all"/>
<refDomain xml:id="x2" target="#m1" source="#x3" determinacy="det"/>
<sourceDomain xml:id="x3" target="#m3" pred="man"/>
<event xml:id="e1" target="#m3" pred="drink_beer"/>
<participation event="#e1" participant="#x1" semRole="agent" distr="individual"/>
<event xml:id="e2" target="#m4" pred="carry_upstairs"/>
<entity xml:id="x4" target="#m5" involvement="all" individuation="count" size="1"/>
<refDomain xml:id="x5" target="#m5" source="#x6" determinacy="det"/>
<sourceDomain xml:id="x6" target="#m6" pred="piano"/>
<participation event="#e2" participant="#x1" semRole="agent" distr="collective"/>
<participation event="#e2" participant="#x4" semRole="theme" distr="individual"/>
```

5.5 Involvement, size and exhaustiveness

The members of the reference domain of a quantification that are actually involved in the events of the event set form the *participant set*. Proportional determiners, such as *"many"* and *"most"* and numerical determiners such as *"three"* and *"more than five"*, indicate how many/much of the reference domain constitutes the participant set. Proportional specifications of participant size should be indicated using <relativeSize> elements, numerical specifications using <cardinality> elements in the values of the @involvement attribute of <participation> structures. Both are illustrated in Example (C1) in Annex C.

The use of a numerical determiner in focus, indicated by prosody in spoken language or by typography in written text, gives rise to a partitive determinate interpretation, such as in Example 6 a), where "two salesmen" means *"two of the salesmen"*, different from Example 6 b), where the stress is on *"salesmen"*.

Example 6 a) TWO salesmen came in. (The three others remained outside.)
b) Two SALESmen came in. (Two policemen as well.)

Numerical determiners may also indicate the cardinality of groups of elements from the reference domain that collectively participate in certain events. This is annotated (slightly simplified) as in Example 7.

Example 7 This assembly machine combines twelve parts.
Markables: m1 = "This assembly machine", m2 = "assembly machine", m3 = "combines", m4 = "twelve parts", m5 = "parts"

```
<entity xml:id="x1" target="#m1" refDomain="#x2" individuation="count"
        involvement="all" size="1"/>
<refDomain xml:id="x2" target="#m2" source="#x3" determinacy="det"/>
<sourceDomain xml:id="x3" target="#m2" pred="assembly-machine"/>
<event xml:id="e1" target="#m3" pred="combine"/>
<entity xml:id="x4" target="#m4" involvement="12" refDomain="#x5"
        individuation="count"/>
<refDomain xml:id="x5" target="#m4" source="#x3" determinacy="det"/>
<sourceDomain xml:id="x6" target="#m5" pred="part"/>
<participation event="#e1" participant="#x1" semRole="agent" distr="individual"/>
<participation event="#e1"participant="#x2" semRole="theme" distr="collective"
         evScope="wide"/>
```

This annotation can be read as: 'For this machine it is the case that there is a set of combine events in all of which a collection of twelve parts is assembled'. See also Example (C3) in Annex C.

5.6 Individuation

The expression as well as the interpretation of the distributivity, involvement and domain size of a quantification is different for mass NPs than for count NPs, hence this is a basic aspect of quantification. In QuantML, the attribute @individuation in <refDomain> elements should be used for marking up this aspect, with values 'count' and 'mass'.

Besides these values, a third possibility is 'cParts', which should be used if the reference domain consists of individual objects but parts of individual objects are also considered as potential participants. This possibility is needed for cases such as Example 8 a), but it is also available in the case of Example 8 b), which possibly describes a series of events where Louis had a pizza last Monday, one and a half pizzas last Tuesday, etc., with a total of eight pizzas.

Example 8 a) Louis and Mary had two and a half pizzas.
b) Louis had eight pizzas last week.

Whether a quantification takes parts of individuals into account is a context-dependent matter, and therefore a property of the participant set, represented by means of an attribute of <entity> structures.

For NPs with a mass head noun, the involvement specification requires the use of <measure> elements, which have a @dimension (e.g. 'volume', 'weight'), a @number and a @numRel and attribute, with values such as 'equal' and 'greater_than'.

5.7 Argument roles

The adoption of the neo-Davidsonian view on events and participants means that a certain set of argument roles must be chosen for differentiating between the different arguments of a verb. The specific 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.

5.8 Polarity and modality

The annotation scheme defined in this document specifies a way of marking up the relative scopes of quantifications and negations. Example 9 shows the use of negation with wide scope (case b)) and narrow scope (case c)), respectively, in two readings of the sentence in a).

Example 9 a) The unions do not accept the proposal.
b) It is not the case that the unions all accept the proposal.
c) Each of the unions does not accept the proposal.

Readings with wide and narrow negation scope should be distinguished in annotations by the @polarity attribute in <participation> elements, using the values "neg-wide" and "neg-narrow", respectively.

Modality is defined in ISO 24617-1 as *expressing 'different degrees of epistemic modality, deontic modality, etc.'* (see ISO 24617-1:2012, Table 1). It can be expressed prosodically or lexically by adverbs, such as *"perhaps"* and *"possibly"* in English, or by modal verbs (*"could", "may", "must"*). Since no full semantic treatment of a wide range of modalities is available, their interpretation is regarded as being outside the scope of the annotation scheme defined in this document. QuantML does allow modal quantifications to be marked up as such, using the @modality attribute in <participation> elements, which can be useful for corpus studies, but does not offer a semantic interpretation in such cases.

5.9 Participant scope

The relative scoping of quantifications over sets of participants is a major source of ambiguity.[40][41]. A sentence with *N* noun phrases may have *N!* possible interpretations due to alternative scopings alone, although syntactic constraints usually reduce this number.[18] The relative scope of participants should be represented in QuantML by means of the <scoping> element, with attributes @arg1, @arg2 and @scopeRel.

There are cases where none of the quantifications has wider scope than the other, as in the ‘cumulative’ quantification[39] in Example 10 on the reading where there is a set A of three breweries and a set B of fifteen inns, such that the members of A supplied members of B, and the members of B were supplied by members of A. In this case, the two quantifications can be said to mutually outscope each other. This should be represented by giving the @scopeRel attribute the value ‘dual’.

Example 10 Three breweries supplied fifteen inns.

Scope under-specification is possible in QuantML by omitting one or more <scoping> elements, resulting in an annotation structure interpreted as an underspecified DRS (UDRS).[38]

5.10 Event scope

Issues of scope in quantification also arise between sets of participants and events. The sentence in Example 11 can be read to mean that everyone is mortal, but also read as a prediction of an apocalyptic future event in which everyone will die.

NOTE The latter interpretation requires the consideration of events in which multiple participants occupy the same role. The ISO approach to semantic role annotation (see ISO 24617-4) does allow this.

The latter reading should be annotated as shown in Example 11, with the @evScope value ‘wide’.

Example 11 Everyone will die.

```
<entity xml:id="x1" target="#m1" involvement="all" pred="person"/>
<event xml:id="e1" target="#m2" pred="die"/>
<participation event="#e1" participant="#x1" semRole="theme" distr="individual"
        evScope="wide"/>
```

5.11 Repetitiveness

The events set of a quantification may consist of repetitions of the same event, occurring more than once. Some languages have lexical items for expressing this, such as “twice” and “thrice” in English, “tvisvar” in Icelandic and “dreimal” in German. Other languages express this by a cardinal number and a noun denoting times or turns, such as “deux fois” in French, “vier keer” in Dutch, and “três vezes” in Portuguese. The annotation of quantifications with an indication of a repeating event should specify the number of repetitions as a value of the attribute @repetitiveness in an <event> element.

5.12 Modifiers — Restrictiveness and linking

Quantification in natural language has been studied mostly in relation to the semantics of NPs as arguments of a verb, but quantification issues also take other forms, as in Example 12 a) and Example 13 a), where an adjective is applied to a set of arguments. In both cases, the expression can be interpreted as saying that *“these books”* as a whole are heavy (collective reading) or that each of *“these books”* individually is heavy (distributive reading). To mark up this distinction, the @distr attribute in <adjMod> elements should be used, as shown in Example 12 c) for the collective reading of the sentence in 12 a).

Example 12

a) (I’m carrying) these heavy books (to the library).

b) Markables: m1 = these heavy books, m2 = heavy, m3 = heavy books, m4 = books

c)

```
<entity xml:id="x1" target="#m1" refDomain="#x2" involvement="all"/>
<refDomain xml:id="x2" target="#m3" source="#x3" restrictions="#r1"/>
<sourceDomain xml:id="x3" target="#m3" pred="book"/>
<adjMod xml:id="r1" target="#m2" distr="collective" pred="heavy"/>
```

When an adjective is used predicatively in a quantifying copular construction, as in Example 13 a), an event-based semantic analysis can be obtained by positing a *'be'* state with the predicate and its argument as participants, leading to an annotation as given in Example 13 b) for the distributive reading of the sentence in Example 13 a). This approach has the advantage of generalizing to any copular verb (such as *"appear", "seem", "look"*) and of going along seamlessly with other verbs in the semantics of annotation structures.

Example 13 a) These books are heavy.

b)

```
<entity xml:id="x1" target="#m1" refDomain="#x2" involvement="all"/>
<refDomain xml:id="x2" target="#m2"source="#x3"/>
<sourceDomain xml:id="x3" target="#m2" pred="book"/>
<event xml:id="e1" target="#m3" pred="be"/>
<participation event="#e1" participant=#x1" distr="individual" semRole="theme"/>
<predication participant="#x1" event="#e1" predicate="heavy" distr="individual"/>
```

When a quantifier's reference domain is restricted by an adjective, a noun, a prepositional phrase, a possessive phrase or a relative clause (see Example 14), this is annotated by using the attribute @restrs in <refDomain> structures. The possible values of this attribute are the modifier structures defined in QuantML: <adjMod>, <nnMod>, <ppMod>, <possMod> and <relClause>.

Example 14 a) Alice showed me her archaeology books/ Timmy's books.

b) Alice showed me two rare books from Chengdu/ that she'd bought in Chengdu.

The quantifier expressed by an NP in a prepositional phrase (PP) can have wider scope than a quantifier in the main clause, as illustrated in Example 15. On the most plausible reading of this sentence, the quantifier *"every university in the country"* takes scope over the quantifier *"a student"*. This phenomenon is known as "inverse linking" (see, for example, References [2], [30], [31] and [37]).

Example 15 President Kay met with a student from every university in the country.

Modifiers can also be used in a non-restrictive way, which in English is sometimes indicated using commas, as in *"The children, who were having a jolly good time at the birthday party, didn't notice the approaching thunderstorm"*. In such a case, the modifier, called a 'qualifier' in this document, does not restrict the reference domain but provides additional information about the participant set. Occurrences of non-restrictive modifiers should be annotated in QuantML as values of the attribute @qualifiers in <entity> structures; see Example (C7) in Annex C.

5.13 Genericity

Generic quantification occurs in sentences that make general statements without referring to any specific events at a particular time and place, as in Example 16.

Example 16 a) Tigers don't eat tomatoes.

b) A self-respecting German businessman drives a Mercedes.

A fundamental question is whether such sentences do express quantifications. One view is that 'generic' NPs do not quantify but refer to a single 'prototypical' individual (see, for example, Reference [14]). Alternatively, generics have been analysed in terms of a special quantifier (see Reference [25]). Within the framework of Discourse Representation Theory (DRT), the use of a special implication has been proposed that allows exceptions.[24]

Since there is no well-established, generally accepted semantic treatment of genericity, this is treated in QuantML in a similar way as modality (see 5.8): it can be marked up, using the @genericity attribute, for which no semantics is defined. This can be useful for corpus studies.

6 QuantML specification

6.1 Abstract syntax

6.1.1 General

An abstract syntax is a formalization in set-theoretical terms of a metamodel. It provides a theoretical basis for specifying various alternative representation format (by means of a concrete syntax), and for providing a semantic interpretation of annotation structures in any representation format supported by the same abstract syntax. Annotation structures consist of two types of substructure: *entity structures* and *link structures.* An entity structure contains semantic information about a segment of primary data and is formally a pair ⟨*m, s*⟩ consisting of a markable and certain semantic information. A link structure contains information about the semantic relation between two or more segments of primary data.

6.1.2 Conceptual inventory

The basic components of annotation structures are instances of the concepts that make up the metamodel and form a store of concepts called the 'conceptual inventory'. The conceptual inventory of QuantML contains the following concepts:

a) Predicates, including the following:

   1) Predicates that correspond to (senses of) lexical items of the language of the primary data, notably nouns, verbs, adjectives and prepositions. Such predicates are designated by canonical forms of lexical items, such as verb stems. These predicates form an open class, the content of which depends on the language of the primary data, the subject matter of the annotated material, and the use of lexical resources such as WordNet and VerbNet.

   2) The numerical relations 'equal', 'les-than' and 'greater-than-or-equal'.

   3) Non-numerical quantitative predicates corresponding to determiners such as *"a few", "a little", "several"* and *"many"* in English, or *"beaucoup"* and *"plusieurs"* in French. This is a language-dependent open class which always contains the language-independent predicates 'all', 'no' and 'some'.

   4) Predicates corresponding to proportional determiners such as 'most', 'all', 'half' and 'two-thirds'.

   5) Semantic role predicates as defined in ISO 24617-4, such as 'Agent', 'Theme' and 'Pivot', for indicating argument roles.

   NOTE 1 QuantML does not prescribe the use of any particular set of semantic roles. Role sets other than ISO 24617-4 (see Reference [6]) can be used as plug-ins (see Reference [8]).

   6) The possessive relation 'Poss'.

b) The non-negative real numbers.

c) Dimensions, such as 'weight', 'volume' and 'length', and units of measurement, which may be *basic* (e.g. 'kilo', 'litre', 'mile') and *derived* (e.g. 'square meter', 'meter per second').

   NOTE 2 See ISO 80000-1:2022, 3.10 and 3.11, and ISO 24617-11:2021.

d) Concepts that correspond the specification of the various aspects of quantification, including the following:

   1) 'determinate' and 'indeterminate' for specifying a determinacy;

   2) 'collective', 'individual', 'sampled' and 'unspecific' for specifying distributivity;

   3) 'count', 'mass' and 'count+parts' for specifying individuation;

   4) 'exhaustive' and 'non-exhaustive' for specifying exhaustivity;

5) ‘wide’ and ‘narrow’ for specifying event scope;

6) ‘wider’, ‘dual’ and ‘equal’ for specifying relative scopes of participants;

7) ‘positive’, ‘wide negative’ and ‘narrow negative’ for specifying the scope of a negation;

8) ‘inverse’ and ‘linear’ for specifying whether a modifier applies with scope inversion.

6.1.3 Entity structures

The following types of entity structure ⟨m, s⟩ are defined:

a) Participant structures: s = ⟨ε_D, v, q, [N], [w]⟩, specifying a reference domain (‘ε_D‘), its individuation (‘v’), an involvement (‘q’), and optionally a domain size (‘N’) and non-restrictive modifiers (‘w’).

b) Domain specification structures:

1) Reference domain specification: s = ⟨D, [r], dt⟩, specifying a list of component domains (‘D’), a list of restrictions ‘r’ (possibly empty) and a determinacy (‘dt’).

2) Subdomain specification: s = <D, [r]>, specifying a list of component domains and a list of restrictions; used for annotating conjunctive NPs.

3) Source domain specification: s = <P>, specifying the characteristic predicate of a certain domain.

c) Involvement specifications: either numerical, where s is a pair ⟨numerical relation, real number⟩ or a measure structure (see g)), or relative, i.e. s is a proportional or an approximative quantitative predicate.

d) Event structures: s = ⟨event domain, repetitiveness⟩, where the repetitiveness is an optional element (by default ‘greater than 0’).

e) Clause structures: s = ⟨event structure, list of entity structures, list of participation or predication links, list of scope relations⟩.

f) Modifier structures, including the following:

1) Adjectival structure: s = ⟨property, distributivity⟩.

NOTE 1 Only intersective modifiers are considered in this document, ruling out adjectives such as “fake” and “former”.

2) NN structure: s = ⟨property⟩.

NOTE 2 Noun-noun modification is always distributive.

3) PP structure: s = ⟨relation, participant set, distributivity, linking⟩ (following Reference [35]); the semantic relation is expressed by the PP’s preposition; the linking is ‘inverse’ or ‘linear’.

4) RC structure: s = ⟨ semantic role, clause structure, distributivity, linking⟩, where ‘semantic role’ is the role that the participants, indicated by the head, play in the events in the relative clause.

5) Possessive structure: s = ⟨Poss, participant set, distributivity, linking⟩.

g) Measure structures: s is a pair consisting of a real number and a (basic or derived) unit.

6.1.4 Link structures

The abstract syntax defines link structures for:

— participation in an event:

— participation in an adjectival predication;

— the relative scoping of participants.

Participation structures connect participants to events; predication link structures do the same for copular constructions. Scope link structures indicate a scope relation between two participant entity structures.

a) Participation links: A septet ⟨event structure, participant structure, semantic role, distributivity, exhaustiveness, event scope, polarity⟩.

b) Predication links: A sextet ⟨participant structure, event structure, adjective, distributivity, exhaustiveness, polarity⟩.

c) Scope links: A triple ⟨participant structure, participant structure, scope relation⟩.

6.2 Concrete syntax — A reference representation format

6.2.1 Representation formats

A concrete syntax is specified in the form of an XML representation of annotation structures. These structures are built up from atomic attribute values, which are XML constants that name elements of the conceptual inventory of the abstract syntax, such as 'det' and 'indet'. For each type of entity structure of the abstract syntax, an XML element is defined which has an attribute @xml:id, whose value is a unique name for the information in the element, and an attribute @target, whose value anchors the annotation in the source data through markables. For convenience, in this document the same predicate names are used as in the abstract syntax. NP heads may be complex due to the combination of modifiers and conjunctions, as *in "precious(ancient(Chinese figurines and drawings) and Thai sculptures)"*; the element <complexDomain> is used for representing such structures (see also A.3.3.2).

6.2.2 Entity structure representations

XML elements for representing entity structures are as follows:

a) <entity>, for representing participant structures:
@refDomain – value type IDREF, @individuation – (count | mass | count+parts),
@involvement – IDREF, optionally: @size – IDREF and @qualifiers - IDREF;

b) <event>: @pred – CDATA, optionally @rep – IDREFS and @genericity – (generic | specific);

c) <refDomain>: @components - IDREFS, @restrictions - IDREFS, @determinacy – (det | indet);

d) <sourceDomain>: @pred - CDATA;

e) <subDomain>: @components - IDREFS, @restrictions - IDREFS;

f) <adjMod>: @pred – CDATA, @distr – (individual | collective | sampled | unspecific);

g) <nnMod>: @pred – CDATA, optionally @restrictions; - IDREFS;

h) <ppMod>: @pRel - CDATA, @pEntity - IDREF, @distr – (individual | collective | sampled | unspecific), @ linking – (linear | inverted);

i) <possMod>: @possessor - IDREF, @distr – (individual | collective | sampled | unspecific), @linking – (linear | inverted);

j) <relClause>: @semRole - CDATA, @distr – (individual | collective | sampled | unspecific), @linking – (linear | inverted);

k) <cardinality>: @numRel - (greater_than-or_equal | equal | less-than), @number – CDATA;

l) <relativeSize>: @pred – CDATA;

m) <measure>: @numRel - (greater_than-or_equal | equal | less-than), @number - CDATA, @unit - CDATA.

6.2.3 Link structure representations

XML elements for representing link structures are as follows:

a) <participation>: @event - IDREF, @participant - IDREF, @semRole - CDATA, @distr – (individual | collective | sampled | unspecific), @evScope: (wide | narrow), and optionally @exhaustiveness – (exhaustive | non-exhaustive), @polarity – (positive | wide-negative | narrow-negative), and @modality - CDATA. The following values are default and may be suppressed: exhaustiveness = “non-exhaustive”, abbreviated “nex”, event scope = “narrow”, polarity = “positive”.

b) <predication>: @participant - IDREF, @event – IDREF, @predicate – CDATA, @distr – (individual | collective | sampled), and optionally @exhaustiveness – (exhaustive | non-exhaustive), @polarity – (positive | wide-negative | narrow-negative), and @modality - CDATA. Default values, which may be suppressed: exhaustiveness = “non-exhaustive”, polarity = “positive”.

c) <scoping>: @arg1, @arg2, both with values of type IDREF, @scopeRel – (wider | dual | equal).

Some attributes, such as @pred and @pRel, have values derived from nouns, verbs, adjectives or prepositions in the annotated data. These values are represented by canonical forms of the lexical items of the language of the data such as verb stems and singular forms of nouns. (For convenience the same forms are used in the abstract and in the concrete syntax.) Since every use is associated with a markable, it is in principle possible to associate different word senses with different occurrences.

Example 17 illustrates the use of the various annotation components. Default values of attributes have been suppressed.

Example 17 The three men moved both pianos

Markables: m1 = The three men, m2 = The men, m3 = three, m4 = men, m5 = moved, m6 = both, m7 = both pianos, m8 = pianos

QuantML/XML:

```
<entity xml:id="x1" target="#m1" refDomain="#x2" individuation="count"
        involvement="#q1" size="#c1"/>
<refDomain xml:id="x2" target="#m2" components="#x3" determinacy="det"/>
<sourceDomain xml:id="x3" target="#m4" pred="man"/>
<relativeSize xml:id="q1" target="" pred="all"/>
<cardinality xml:id="c1" target="#m3" numRel="equal" number="3"/>
<event xml:id="e1" target="#m5" pred="move"/>
<entity xml:id="x4" target="#m7" refDomain="#x5" individuation="count"
        involvement="#q2" size="#c2"/>
<relativeSize xml:id="q2" target="#m6" pred="all"/>
<cardinality xml:id="c2" target="#m6" numRel="equal" number="2"/>
<refDomain xml:id="x5" target="#m7" components="#x6" determinacy="det"/>
<sourceDomain xml:id="x6" target="#m8" pred="piano"/>
<participation event="#e1" participant="#x1" semRole="agent" distr="collective"/>
<participation event="#e1" participant="#x4" semRole="theme" distr="individual"/>
<scoping arg1="#x1" arg2="#x4" scopeRel="wider"/>
```

6.3 Semantics

QuantML annotations have a compositional semantics, in the sense that the interpretation of an annotation structure is obtained by combining the interpretations of its entity structures and participation link structures, in a manner determined by its scope link structures. The specification of the semantics in this document has the form of translating annotation structures to DRSs, as defined in DRT. This form of semantics is convenient for combining annotations of quantification with other types of semantic information, using the

SemAF (see the ISO 24617 series), which also uses DRSs in some of its parts; otherwise, second-order logic would be equally suitable. This subclause gives a brief outline of the semantics; a systematic specification is provided in Annex B.

According to the metamodel of Figure 1, the main components of an annotation structure are the structures that describe participant sets, event sets and the participation relations between them. Example 18 shows the DRS representing the quantifier expressed by the NP *'Thirty-two Chinese students"* (in b)) and the event set expressed by the verb *"enrolled"* (in c)). The DRS for the NP introduces a discourse referent (X) for the participant set, and includes conditions expressing that the source domain of the participants is Chinese students, and that there are 32 of them in the participant set; the DRS for the verb introduces a discourse referent (E) that refers to a set of 'enroll' events.

NOTE All discourse referents for sets of participants or events are required to be non-empty. Discourse referents for events have default repetitiveness of 1. The conditions |X| ≥ 1 and |E| = 1 are therefore suppressed in all DRSs.

Example 18

a) Thirty-two Chinese students enrolled.

b) [X | x ∈ X → [student(x), Chinese(x)], |X| = 32]

c) [E | e ∈ E → enroll(e)]

d) [X | x ∈ X → [E | e ∈ E → agent(e, x)]]

The DRS for the participation link (see d)) introduces two discourse referents, one for a set of events (E) and one for a set of participants (X), and relates these sets through the semantic role Agent, applied to individual members of X. Since the event scope is narrow, the event set referent is within the scope of the quantification over the participant set.

The DRSs of Example 18 b), c) and d) are combined using the 'glue merge' operation, defined in Annex B, with the DRS shown in Example 19 as result.

Example 19 [X | |X| = 32, x ∈ X → [student(x), Chinese(x),
[E | e ∈ E → [enroll(e), agent(e, x)]]]]

For a verb with multiple arguments, the interpretations of the link structures are combined in a way that reflects the relative scoping of the arguments, similar to the way event scope is reflected in Example 20. Example 20 illustrates this for the wide-scope reading of *"Some students".* The DRS for this NP is quantified over a reference domain (Y) that is a subset of the source domain 'student', saying that for each member of this reference domain there is a set of three papers, for each of which there is a 'read' event with the student as Agent and the paper as Theme.

Example 20

a) Some students read three papers.

b) [Y | Y ⊆ student, y ∈ Y → [Z | Z ⊆ paper, |Z| = 3, z ∈ Z → [E |E ⊆ read, e ∈ E →
[| agent(e, x), theme(e,y)]]]]]

Annex A
(informative)

Annotation guidelines

A.1 Overview

This annex contains guidelines for using the XML-based referential representation format of QuantML, also referred to as 'QuantML/XML', with a focus on language-independent issues, while illustrative examples are taken mostly from the way quantification is expressed in English.

The natural unit of text (or speech) to annotate in QuantML is what in linguistics is called a '(simple) clause', i.e. a structure consisting of a verbal predicate and its arguments, since this is where quantifications primarily arise. Clause A.2 explains, step by step, the creation of the QuantML annotation for a simple clause in English that contains several quantifications. Clause A.3 provides a systematic discussion of the use of the elements and attributes of QuantML/XML.

A.2 From source text to QuantML — Worked example

QuantML annotations contain two kinds of structures, corresponding to the 'entity structures' and 'link structures' of the abstract syntax (see 6.1.1). Entity structures contain information about a stretch of source text, and their XML representations therefore contain an attribute whose values identify such a stretch, called "markables". The **@**target attribute is used for this purpose. The first step in the annotation process is the identification of a predicate and its arguments. This is illustrated for the sentence given in Example (A1a).

(A1a) Santa gave the children a present.

The predicate *"gave"* has the following arguments:

— Santa;

— a certain set of children;

— an indeterminate set of presents.

Markables are introduced for the corresponding NPs and the verb, as well as for the NP heads, as shown in (A1b), since these correspond to certain lexical semantic content. For the verbal predicate, an <event> element is created, by convention using the infinitival verb form to denote the lexical semantic content, as in (A1c). Tense and aspect are not considered in QuantML.

(A1b) m1 = "Santa", m2 = "gave", m3 = "the children", m4 = "children", m5 = "a present",
m6 = "present".

(A1c) <event xml:id="e1" target="#m2" pred="give"/>

Predicate arguments, i.e. participant sets, are described by <entity> elements such as (A1d) for the quantifier *"the children"*. Proper names are treated as quantifying over a singleton set that consists of a contextually determinate individual, as illustrated in (A1e) for the proper name "*Santa*":

NOTE For the use of the @individuation attribute, see A.3.3.3.

(A1d)
```
<entity xml:id="x4" target="#m3" refDomain="#x5" individuation="count"
        involvement="#q2"/>
<refDomain xml:id="x5" target="#m3" components="#x6" determinacy="det"/>
<sourceDomain xml:id="x6" target="#m4" pred="child"/>
<relativeSize xml:id="q2" target="" pred="all"/>
```

(A1e)
```
<entity xml:id="x1" target="#m1" refDomain="#x2" individuation="count"
        involvement="#q1" size="#n1"/>
<refDomain xml:id="x2" target="#m1" components="x3" determinacy="det"/>
<sourceDomain xml:id="x3" target="#m1" pred="santa"/>
<relativeSize xml:id="q1" target="" pred="all"/>
<cardinality xml:id="n1" target="" numRel="equal" number="1"/>
```

In the absence of information about the context of use, the sentence is ambiguous, in view of the possible answers to the following questions:

a) Did Santa give each of the children a different present?

b) Was there a particular present that was given to each of the children?

c) Was a present given to the children as a group (such as a new set of toys for the schoolyard)?

In other words:

— What is the relative scope of the quantifications over children and presents?

— Were the children involved individually or collectively?

Semantic annotations often aim to disambiguate or to express constraints on interpretation. In the reading of Example (A1a) in which Santa gave each child a different present, all the individuals in the reference domain of the quantifier *"the children"* are involved. They participate individually in the role of beneficiary of a *"give"* event. This should be annotated as follows:

(A1f)
```
<participation event="#e1" participant="#x4" semRole="beneficiary" distr="individual"/>
<participation event="#e1" participant="#x1" semRole="agent" distr="individual"/>
```

The quantifier expressed by *"a present"* refers on this reading to an indeterminate set of presents, of which at least one (represented by the size indicator "some") is given to one of the children. Together with the assumption that each child received a different present, i.e. that the quantification over children outscopes the one over presents, this is expressed by the following QuantML fragment:

(A1g)
```
<entity xml:id="x7" target="#m5" refDomain="#x8" individuation="count" involvement="#q3"/>
<refDomain xml:id="x8" target="#m5" components="#x9" determinacy="indet"/>
<sourceDomain xml:id="x9" target="#m4" pred="present"/>
<relativeSize xml:id="q3" target="" pred="some"/>
<participation event="#e1" participant="#x7" semRole="theme" distr="individual"/>
<scoping arg1="#x4" arg2="#x7" scopeRel="wider"/>
```

For a quantification with a reference domain that contains only one element, as in the case of *"Santa"*, the scope relative to other quantifications does not matter. In order not to make the QuantML semantics unnecessarily complex, it is convenient to assign proper names maximal scope. Since the quantifier of *"the children"* already outscopes the one of *"a present"*, this can be expressed by stipulating that the quantifier of *"Santa"* outscopes the one of *"the children"*:

(A1h)
```
<scoping arg1="#x1" arg2="#x4" scopeRel="wider"/>
```

Together, (A1b) to (A1h) represent the QuantML annotation for (A1a) upon the chosen reading.

A.3 QuantML/XML elements, attributes and values

A.3.1 Markables

The use of markables is characteristic for standoff annotation, and is of semantic importance for content words, the meaning of which can vary with context and which can be expressed in the value of the @pred attribute. The specification of a markable allows different occurrences of a content word to be identified, and thus to distinguish different, context-specific interpretations.

The segments for which markables are specified may overlap and may be discontinuous. Overlapping occurs due to phrasal markables containing content words. Discontinuity is illustrated by the Dutch sentence in (A2), containing the discontinuous verbal cluster *"heb (...) gekocht"* (have bought) and the discontinuous NP *"een electrische auto met vijf deuren"* ('an electric car with five doors').

(A2) Ik heb *een electrische auto* gekocht *met vijf deuren*.
(Word-for-word translation: "I have an electric car bought with five doors.")

Markables: m1 = Ik, m2 = heb gekocht, m3 = een electrische auto met vijf deuren, m4 = electrische, m5 = electrische auto met vijf deuren, m6 = auto, m7 = auto met vijf deuren, m8 = met vijf deuren, m9 = vijf deuren, m10 = deuren

The numbering of markables is obtained by going from left to right through the text and applying the left-right order relation of overlapping and discontinuous segments as defined in discontinuous phrase structure grammar,[7] according to which segment S1 precedes S2 if the leftmost constituent word of S1 precedes that of S2, and if these words are the same, then if the rightmost word of S1 precedes that of S2.

A.3.2 The <event> element

The term 'event' is used in QuantML in a similar way as in ISO 24617-1 (ISO-TimeML), i.e. as a cover term for expressions that denote either:

a) situations that can be said to happen or occur, which can be punctual or last for some time; or

b) states or circumstances in which something obtains or holds true, possibly for an indefinite time.

Given a clause in the source text to be annotated, the QuantML/XML element <event> is used to mark up the (main) verb with values for the attributes @xml:id, @target, @pred, and optionally @rep and @genericity. These values specify, respectively, a unique identifier, a markable that identifies the (occurrence of the) verb, and an event type of which the verb occurrence denotes a token (such as *"sleep"*, *"appear"* or *"give"*) and optionally restrictions on the number of repetitions of the event (interpreting expressions such as *"twice"*, *"more than six times"* and *"many times"*). If no @rep value is specified, its assumed default value is ≥ 1. Moreover, the optional attribute @genericity can be used to indicate that a sentence should be understood as generic, rather than about specific events.

The element <event> can be viewed as a simplified version of the <event> element of ISO-TimeML, which has additional attributes including tense, aspect, polarity and modality. The @pred attribute is defined in ISO-TimeML as denoting 'the content related to the event through the indication of a lexical predicate', using infinitive verb forms for this purpose.

A.3.3 The <entity> element

A.3.3.1 General

The **<entity>** element is used to describe sets of participants in events. As the metamodel in Figure 1 clearly shows, the annotation of aspects of quantification is concentrated in participant sets and participation relations. This information is found mostly in NPs and adverbs (such as *"together"*, *"individually"*, *"never"*, *"each"* and *"only"*).

Besides the @xml:id and @target attributes, an <entity> element has three obligatory attributes: @refDomain, @individuation and @involvement, and two optional ones: @size and @qualifiers. Roughly speaking, the @refDomain value specifies a kind participant; the @individuation value makes the count/mass distinction; and the @involvement value specifies how much/many of the reference domain belongs to the participant set. Of the optional attributes, the @size value provides information about the comprehensiveness of the reference domain, and @qualifiers values refer to non-restrictive modifiers. The use of the attributes is discussed in A.3.3.2 to A.3.3.6.

A.3.3.2 The @refDomain attribute and the <refDomain>, <subDomain> and <sourceDomain> elements

The value of the @refDomain attribute specifies the reference domain of a quantification. Its value is a <refDomain> element, which in turn refers to a <sourceDomain> element, and for complex NP heads with conjunctions and modifiers, to a general <subDomain> element. Example (A3) illustrates their use.

(A3) ancient [[Chinese charms] and [terracotta figurines]]

Markables: m1 = ancient, m2 = ancient Chinese charms and terracotta figurines,
m3 = Chinese, m4 = Chinese charms, m5 = Chinese charms and terracotta figurines,
m6 = charms, m7 = terracotta,m8 = terracotta figurines, m9 = figurines

```
<entity xml:id="x1" target="#m2" refDomain="#x2" individuation="count"
        involvement="some"/>
<refDomain xml:id="x2" target="#m5" components="#x3 #x5" restrs="#r1"
        determinacy="indet"/>
<subDomain xml:id="x3" target="#m5" components="#x4" restrs="#r2"/>
<sourceDomain xml:id="x4" target="#m6" pred="charm"/>
<adjMod xml:id="r2" target="#m3" pred="chinese/>
<subDomain xml:id="x5 target="#m8" components="#x6" restrs= "#r3"/>
<sourceDomain xml:id="x6" target="#m9" pred="figurine"/>
<nnMod xml:id="r3" target="#m7" pred="terracotta"/>
<adjMod xml:id="r1" target="#m1" pred="ancient"/>
```

@determinacy: This attribute of <refDomain> elements should be assigned the value "det" if the NP is interpreted as quantifying over a contextually determined reference domain that is a proper part of the source domain (defined by the NP head). The occurrence of an indefinite article is often a cue for the quantification being indeterminate (in which case the reference domain coincides with the source domain). Definite articles, possessive phrases and demonstratives can be a cue for being determinate, but whether this is the case depends on the context. Proper names are determinate: the use of a name such as "John" carries the assumption that there is one contextually distinguished person named "John".

@pred: The value of this <sourceDomain> attribute is the characteristic function of the source domain, for which a language-dependent canonical form of the head noun is used. See also 6.2.3.

@restrs: An optional attribute in <subDomain> and <refDomain> elements, whose values refer to restrictive modifiers of an NP head noun or component of a conjunctive NP head. Absence of a value indicates that no such modifiers are present. Five types of modifier are distinguished in QuantML: adjectives, nouns, preposition phrases, possessive structures and relative clauses, used as follows:

a) Restrictively used adjectives are annotated with <adjMod> elements, in which the @distr attribute should be assigned the value "individual" if the adjective applies to the individual members of the reference domain, and "collective" if it applies to these members together; the @pred attribute gets is value in the same way as NP head nouns.

b) Nouns as modifiers are annotated with <nnMod> elements, in which the @pred attribute gets is value in the same way as NP head nouns.

c) Modification by a prepositional phrases is annotated with a <ppMod> element, in which the @distr attribute should be assigned the value "individual" if the PP applies to individual members of the reference domain, and "collective" if it applies to these members together; the @linking attribute should be assigned the value "inverse" if the quantification of the NP in the PP outscopes the quantification of the NP head, otherwise it has the default value "linear".

d) Possessive modifications are annotated with <possMod> elements, in which the @distr attribute is assigned the value "individual" if the possessive restriction applies to the individual members of the reference domain, and "collective" if it applies to these members together; the @linking attribute should have the value "inverse" if the quantification of the NP that refers to the possessor outscopes the one of the NP head, otherwise it should have the value "linear".

e) Relative clauses are annotated with <relClause> elements, where the @distr attribute should be assigned the value "individual" if the clause applies to individual members of the reference domain, and "collective" if it applies to these members together; @linking should have the value "inverse" if the relative clause contains a quantification that outscopes the quantification of the NP head, else it should have the value "linear".

A.3.3.3 The @individuation attribute

The value of the @individuation attribute specifies whether the NP head noun is used as a mass noun or as a count noun, and in the latter case whether parts of the individuals in this domain are also considered. In the latter case, @individuation should get the value "cParts". It is recommended to use this value only if there is evidence, either in the text to be annotated or in the context, that parts of individuals are relevant to consider as participants, as in the text *"We ate three and a half pizzas"*.

A.3.3.4 The @involvement attribute

The value of the @involvement attribute indicates how many/much or which fraction of the reference domain is contained in the participant set. Such an indication may be expressed by a quantitative predeterminer or by a proportional predeterminer, such as *"all', "each", "some", "several", "many", "much", "half"* and *"most"* in English. A quantitative specification is annotated by the @involvement value referring to a <cardinality> element or a <measure element>. Proportional specification of involvement is annotated by means of <relativeSize> elements. This element has a @pred attribute whose values are the lexical items corresponding to determiners (i.e. values such as *"many", "plusieurs"* and *"mucho"*), with the following exceptions:

a) The string value "all" is used to indicate that all the elements of a countable reference domain are members of the participant set. This includes the use of predeterminers in English such as *"all", "each", "every"* and some uses of *"any"*.

b) The string "some" is used to indicate that at least one element of a countable domain is involved.

c) The strings "total" and "some-m" are used to indicate complete and non-zero involvement of a mass reference domain, respectively.

These string values are also used for marking complete or non-zero involvement in cases where the involvement is not explicitly indicated but understood, as *"the children?"* in (A1c) and in determinerless NPs, e.g. *"Do you have <u>fresh pasta</u>?" ≈ "Do you have <u>some fresh pasta</u>?"*. A predeterminer indicating zero involvement should be annotated as the negation of non-zero involvement, using the @polarity attribute of a <participation> link element (see A.3.4).

A.3.3.5 The @size attribute

Where the @involvement attribute contains information about the size of a *participant set*, the optional @ size attribute does the same for a *reference domain*. In English, this is expressed by a numerical expression in post-determiner position (see 5.2), as in *"Two of the <u>five</u> students"*.

Singular proper names and definite description (e.g. *"Joe", "the president"*) are treated as quantifying over a singleton set, which means that the values shown in (A4) should be used for the @involvement and @size attributes, and for the @determinacy of the <refDomain> element. This use of attribute values allows the annotations of proper names and definite descriptions to be interpreted semantically in the same way as other NPs.

(A4)
```
<entity xml:id="x1" target="#m1" refDomain="#x2" individuation="count" involvement="all"
        size="n1"/>
<refDomain xml:id="x2" target="#m1" components="#x3" determinacy="det"/>
<sourceDomain xml:id="x3" target="#m2" pred="president"/>
<cardinality xml:id="n1" target="" numRel="equal" number="1"/>
```

A.3.3.6 The @qualifiers attribute

This optional attribute should be assigned a value only for NPs of which the head includes a non-restrictive modifier, as in *"My parents, who lived in Arizona, went out for lunch every Sunday"*, and in a plausible reading of *"The curious student asked many questions".*

A.3.4 The <participation> link

The relations between events and participant sets are not expressed by words, hence a <participation> element does not contain a markable, and since these relations do not occur as values of any attributes, they do not need an @xml:id with an identifier as value. It does have obligatory attributes for indicating its two arguments, indicated by the value of the semantic role of the participants, and by four properties of the relation: distributivity, polarity, exhaustiveness and event scope.

@semRole: Assign as value one of the semantic roles defined in ISO 24617-4.

@distr: If each member of the participant set participates in the events individually, then assign the value "individual". If they act together or are acted upon together, then assign the value "collective". If some of them act (or are acted upon) together, and others by themselves, assign the value "unspecific". For a mass NP, use the value "sampled" if parts of a mass noun denotation participate, and "collective" if the parts act (or are acted upon) together.

@polarity: Optional attribute, default value "positive". Only use the values "neg-wide" and "neg-narrow" if there is evidence that the annotated material makes an overall negative claim, as in *"Nobody is here"*; or that a negative element occurs within the scope of a quantifier (see 5.8). Note that all participation links to the same event set must have the same value.

@modality: Optional attribute used to annotate adverbs such as *"perhaps", "possibly", "probably"* and *"necessarily",* and modal verbs such as *"must", "can", "could", "shall"* and *"may".* QuantML has three possible values for the scope of a modality specification, as explained in 5.8. If no value of this attribute is specified, then it is assumed not to apply.

@exhaustiveness: Optional attribute, default value "non-exhaustive". Use the value "exhaustive" only if there is linguistic or contextual evidence indicating that no other participants are involved with the semantic role under consideration than those in the participant set.

@evScope: Optional attribute; the quantification over participants nearly always outscopes the one over events, therefore "narrow" is the default event scope. Use the value "wide" only when there are multiple participants with the same semantic role (see also 5.10).

A.3.5 The <predication> link

Predication links are used to annotate the relation between individuals and properties as described by copular verbs, and have the same attributes as participation links, except they do not have an @evScope attribute, and they do have a **@pred** attribute with adjectives as values. More complex predicates can be taken into consideration, but do not seem of great interest for the annotation of quantification.

A.3.6 Scope links: <scoping> elements

Being representations of link structures, just like <participation> elements, <scoping> elements are not associated with a markable and do not require an identifier, but just have attributes for referring to two quantifications and the scope relation between them.

@arg1 and **arg2**: Values refer to the quantifications whose relative scoping is indicated.

@scopeRel: Use “dual” in the case of cumulative quantification (mutual outscoping), and “equal” in the case of group quantification; see Example (C3) and Example (C4) in Annex C.

Annex B
(informative)

QuantML semantics

B.1 Overview

Conforming to the ISO principles of semantic annotation (see ISO 24617-6:2016), QuantML has a triple-layered annotation scheme, with a concrete syntax, an abstract syntax and a semantics. These layers are connected by three functions:

a) an encoding function F_{AC} which assigns to every well-formed structure of the abstract syntax a representation using the concrete syntax;

b) a decoding function $F_{AC}{}^{-1}$, which assigns to every structure of the concrete syntax a structure of the abstract syntax;

c) an interpretation function I_Q that assigns a semantic interpretation to the structures of the abstract syntax.

This architecture, visualized in Figure B.1, supports the interoperability of annotations, as it allows semantically equivalent alternative representation formats, indicated by 'Representation Format 2' in Figure B.1.

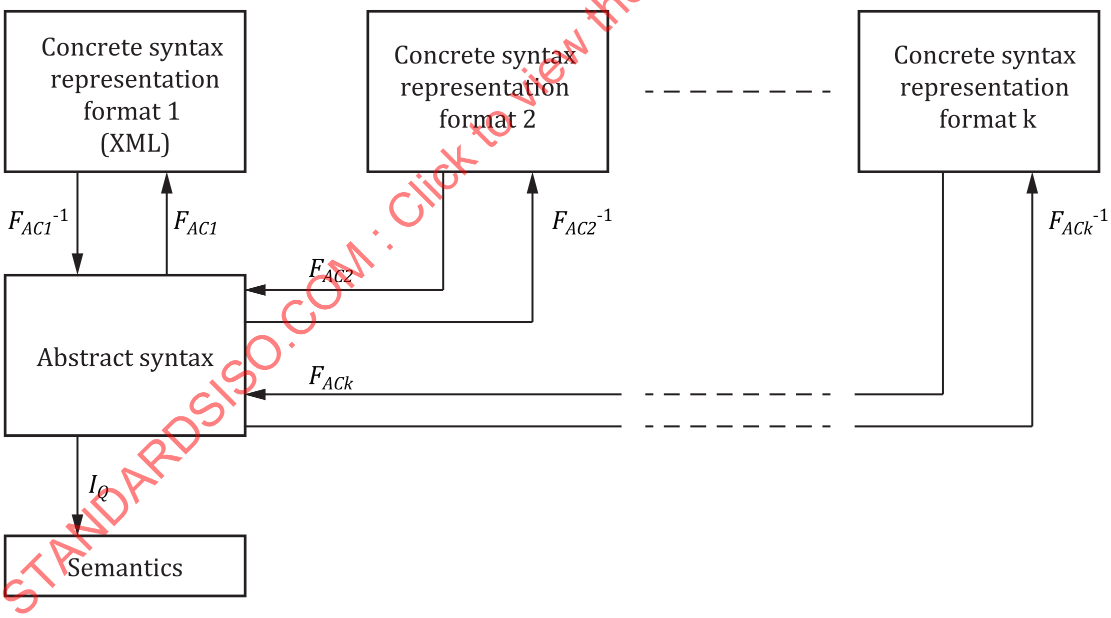


Figure B.1 — Three-layer architecture of QuantML

Annotators deal only with the concrete syntax, as they make annotations in the defined representation format by the concrete syntax (or some other, equivalent format). They can rely on the existence of the mappings of these representations to the underlying abstract structures and their semantics.

The interpretation function, I_Q, defined in this annex translates annotation structures into DRSs. This semantics is compositional in the sense that the translation of a clause annotation structure is obtained by combining the interpretations of its component entity structures and link structures. The operations for combining interpretation structures are defined in Clause B.3. In Clause B.2, some naming and notation conventions are introduced. Clauses B.4 to B.7 outline the interpretation of entity structures, link structures and clause annotation structures. More details about the semantic layer of QuantML can be found online in the Quantification Bank[37] and in Reference [9].

B.2 Naming and notation conventions

The notation X* is used (adopting the notations introduced in References [24] and [26]) to designate the set consisting of the members of X and the subsets of X, and if P is the characteristic function of X, then P* designates the characteristic function of X*, and ‘P^’ designates the property of being a part of an individual that has the characteristic property P.

If P is the characteristic function of the source domain D of a quantification, then P_0 designates the characteristic function of the reference domain of that quantification.

Short notations are used to represent frequently occurring semantic structures: [X ⊆ P | C] for [X | C, x ∈ X → P(x)], [X = P | C] for [X | C, x ∈ X ↔ P(x)], and [x ∈ P | C] for [x | C, P(x)].

B.3 Constructing interpretations compositionally

The interpretation function, I_Q, defined in this annex translates annotation structures recursively into DRSs. This semantics is compositional in the sense that the translation of a clause annotation structure is obtained by combining the interpretations of its component entity structures and link structures. The recursion in I_Q comes to an end at the minimal building blocks of annotation structures, which are elements of the QuantML conceptual inventory. For these building blocks, an assignment function F_Q provides semantic values that can be used in DRS conditions: I_Q(a) = F_Q(a).

An important subclass of the conceptual inventory is formed by ‘lexical’ predicates which correspond to natural language content words (nouns, adjectives, verbs, prepositions, etc.). These predicates occur in annotation structures as elements of a pair ⟨*m, P*⟩, where *m* is a markable that identifies an occurrence of the corresponding content word. Different occurrences that correspond to different word senses can thus be distinguished. For convenience, in this document, the same strings that are used in annotation structures to name ‘lexical’ conceptual predicates are also used to name predicates in DRS conditions. The same convention is used for approximative quantitative determiners such as in English *“many”*, *“some”* and *“several”.*

Combining the information in two or more DRSs into a single DRS occurs at three places in the semantics:

a) In the interpretation of participant structures. Head-modifier constructions are interpreted by combining the DRSs of the head and the modifier. This is achieved by the classical DRS-merge operation familiar from DRT, symbolized ‘∪’.

b) In the interpretation of participation links. The information in three DRSs is combined:

   1) one for a participant set;

   2) one for an event set;

   3) a ‘glue DRS’ that specifies their thematic relation, the distributivity of the relation, and other properties.

   This combination is performed by an operation called *glue merge*, symbolized ‘∪⁺‘.

c) At clause level, the DRSs for the individual participation links are combined into a DRS that takes the scope relations into account. This involves the use of two operators, called *‘scoped merge’* and *‘dual-scoped merge’,* respectively, and symbolized ’∪*’and ‘$\cup^{\blacklozenge}$’.

The *glue merge* operation ∪⁺ is defined as follows:

— Of its three arguments, the first is a DRS that forms the interpretation of a participant structure, which has the general form [X, X’ | X ⊆ X’, C(X, X’), N(X’), x’∈ X’ ↔ K_1] or the simpler form [X | C(X), x ∈ X → K_1] (see

(B6)), with X referring to the participant set and X' to the reference domain, C(X, X') to a (possibly empty) set of conditions on X and X', N(X') to a quantitative condition on the size of X', and K_1 to any sub-DRS.

— The second argument is a DRS interpreting an event structure, which has the same form as a simple participant structure interpretation, namely: $[E \mid C(E), e \in E \rightarrow K_2]$.

— The third argument, the glue DRS, interprets the distributivity, event scope, exhaustivity and polarity of a participation link. The most important forms that this third argument can take, namely for non-exhaustive individual or collective participation with narrow event scope and positive polarity, are those in (B1) (where R is a semantic role).

(B1) a. (individual distributivity) $[X \mid x \in X \rightarrow [E \mid e \in E \rightarrow R(e,x)]]$
b. (collective distributivity) $[X, E \mid e \in E \rightarrow R(e,X)]$

The glue DRS functions as a mould into which the information in the first two arguments is cast by the glue merge operation, exploiting the fact that the discourse referents in the glue DRS refer to the participant set and event set of the first and second arguments to unify the corresponding discourse referents. For example, for the sentence *"More than two thousand students protested twice"*, where the 'glue DRS' has the form (B1a), the glue merge has the following result:

$$\cup^{+}([S \mid |S| > 2000, s{\in}S \rightarrow student(s)], s{\in}S \rightarrow [P \mid |P| = 2, p{\in}P \rightarrow protest(p)],$$

$$[X \mid x{\in}X \rightarrow [E \mid e{\in}E \rightarrow agent(e,x)]]) =$$

$$= [X \mid |X| > 2000, x{\in}X \rightarrow [student(x), [x{\in}X \rightarrow [E \mid |E| = 2, e{\in}E \rightarrow [protest(e), agent(e,x)]]$$

A slightly different operation, called *scoped merge* ('∪*'), combines the information in two DRSs that each represent the semantics of a participation link, forming a single DRS which reflects the relative scope of the arguments. This operation exploits the fact that both DRSs refer to the same event set, therefore the discourse referents for events can be unified. This is illustrated in (B2). More details can be found in B.5.3 and in Reference [9].

(B2) Some students read three papers twice.

$$[X \mid x{\in}X \rightarrow student(x), x{\in}X \rightarrow [E \mid |E| = 2, e{\in}E \rightarrow [read(e), agent(e,x)]]] \cup^{*}$$

$$[Y \mid |Y| = 3, y{\in}Y \rightarrow paper(y), [y{\in}Y \rightarrow E \mid |E| = 2, e{\in}E \rightarrow [read(e), theme(e,y)]]] =$$

$$= [X \mid x{\in}X \rightarrow student(x), x{\in}X \rightarrow [Y \mid |Y| = 3, y{\in}Y \rightarrow paper(y), y{\in}Y \rightarrow$$

$$[E \mid e{\in}E \rightarrow [read(e), agent(e,x), theme(e,y)]]]]$$

The *dual scoped merge* operation ∪* combines two participant structure interpretations in the case of cumulative quantification, i.e. mutual outscoping. It essentially applies the scoped merged operation twice, for the two scoping relations. More details can be found in B.5.3 and in Reference [9].

B.4 Conceptual inventory items

The proportional specification of a quantification's domain involvement, e.g. by *"most"*, depends on the reference domain, e.g. *"most (of the) books are new"* means that more than half of the books in the reference domain are new. This can be captured by interpreting *"most"* as the two-place predicate λY. λX. X > (Y/2), which can be applied to an argument such as '$|book_0|$' to produce the predicate $\lambda X.X > (|book_0|/2)$, i.e. the numerical property of being greater than half the number of books in the reference domain. Proportional involvement specifications such as *"all"* can be treated similarly: $F_Q(all) = \lambda Z. \lambda X. Z \subseteq X$.

Specification of involvement in a mass NP quantification requires a way of measuring quantities of a mass domain. Quantities of milk are, for example, typically measured in terms of volume, quantities of sugar in weight and quantities of rope in length. A dimension that is ordinarily used to measure the size of a quantity of M, specified in the QuantML semantics, is designated by Dim(M). The use of dimensions is implemented in the QuantML semantics by defining the unit in an amount expression ⟨N, u⟩ as including a dimension in its semantics.

NOTE See Reference [6] for a definition of dimensions within the framework of model-theoretic semantics.

For example, a kilogram is a unit in the weight dimension, hence F_Q(kilogram) = ⟨Weight, kg⟩. This is exploited in the semantics of an involvement specification according to (B3), where the subscripts 1 and 2 designate the first and second element of a pair.

(B3) $I_Q(N, u) = \lambda P. \lambda x. P((F_Q(u))_1(x), (F_Q(u))_2)\ (I_Q(N))$

For example, *"more than five kilograms"* (where $I_Q(N) = \lambda z.z > 5$) has the semantics λx.(Weight(x),kg) > 5.

The size of a participant set that includes parts of individuals can be measured by adding up the sizes of the participating parts, expressed as fractions of a (prototypical) individual. For a set X consisting of individuals as well as parts of individuals of type D (i.e. from the source domain D), the sum of these fractions is designated by Count(X, D). For more details, see the semantics provided in the Quantification Bank[37].

B.5 Entity structures

B.5.1 General

As noted in 6.2.1, an entity structure is a pair ⟨*m, s*⟩ consisting of a markable that identifies a segment of primary data and semantic information about that segment. As markables do not form part of the semantic information, for any type of entity structure $I_Q(\langle m, s\rangle) = I_Q(s)$.

B.5.2 Participant structures

The semantic information in a participant structure is expressed in a tuple $\langle \varepsilon_{RD}$, v, q, [N], [w]⟩, consisting of a reference domain, an individuation, an involvement, a domain size (optional) and non-restrictive modifiers (optional). The specification of a reference domain is a triple $\varepsilon_{RD} = \langle D, [\mu], dt\rangle$, consisting of a domain description, a (possibly empty) set of restrictions and a determinacy. The semantic interpretation of participant structures makes use of two auxiliary functions:

a) σ, which determines the source domain of a complex domain specification;

b) π, which adds parts of individuals to the reference domain in the case of a count+parts individuation (π).

These functions are defined in (B4).

(B4) a) (i) For any domain specification, $\sigma(D, \mu, dt) = \sigma(D, \langle\rangle, dt)$
(ii) if D is a single, unstructured predicate, then $\sigma(D, \langle\rangle, dt) = F_Q(D)$
(iii) for a conjunctive specification: $\sigma(\langle D_1, .., D_k\rangle, \langle\rangle, dt) = \{\sigma(D_1, \langle\rangle, dt), .., \sigma(D_k), \langle\rangle, dt)\}$

b) (i) For any domain specification, $\pi(D, \mu, dt) = \pi(D, \langle\rangle, dt)$
(ii) if D is a single, unstructured predicate, then $\pi(D, \langle\rangle, dt) = (F_Q(D))\hat{}$
(iii) for a conjunctive specification: $\pi(\langle D_1, .., D_k\rangle, \langle\rangle, dt) = \langle \pi(D_1, \langle\rangle, dt), .., \pi(D_k), \langle\rangle, dt)\ \rangle, \langle\rangle, dt\rangle$.

The semantic interpretation of a participant structure without optional components is defined by (B5), where $q' =_D I_Q(q)$.

(B5) a) $I_Q(\varepsilon_{RD}, \text{count}, q) = [\ X \mid q'(X), x{\in}X \rightarrow I_Q(\varepsilon_{RD})(x)\ ]$
b) $I_Q(\varepsilon_{RD}, \text{mass}, q) = [\ X \mid q'(\text{Dim}(\sigma(\varepsilon_{RD})(\Sigma X))), x \in X \rightarrow I_Q(\varepsilon_{RD})(x)\ ]$
c. $I_Q(\varepsilon_{RD}, \text{count+parts}, q) = [\ X \mid q'(\text{Count}(X, \sigma(\varepsilon_{RD}))), x \in X \rightarrow I_Q(\pi(\varepsilon_{RD}))(x)\ ]$.

The DRS in the right-hand side of (B5) introduces a set-type discourse referent (X) which is used to refer to the participant set of a quantification. Participant structures with a specification of domain size (N) require the introduction of *two* set-type discourse referents, one for the participant set (X), and one for the reference domain (X'), as shown in (B6), where N' designates $I_Q(N)$.

(B6) a) For a quantitative involvement specification q (e.g. 'six' or 'many'):
$I_Q(\varepsilon_{RD}$, count, q, N) = [X, X' | X ⊆ X', N'(X'), q'(|X|), x'∈ X' ↔ $I_Q(\varepsilon_{RD})$(X')(x')]
$I_Q(\varepsilon_{RD}$, mass, q, N) = [X, X' | X ⊆ X', N'(X'), q'(ΣX), x'∈ X' ↔ $I_Q(\varepsilon_{RD})$(X')(x')]
$I_Q(\varepsilon_{RD}$, count+parts, q, N) = [X, X' | X ⊆ X', q'(Count(X,σ(ε_{RD}))), x'∈ X' ↔ $I_Q(\pi(\varepsilon_{RD}))$(X')(x')]

b) For a proportional involvement specification q, e.g. 'half' (in which case ε_{RD} is determinate):
$I_Q(\varepsilon_{RDRD}$ count, q, N) = [X, X' | X ⊆ X', N'(X'), q'(|X'|)(|X|), x'∈ X' ↔ $I_Q(\varepsilon_{RD})$(X')(x')]
$I_Q(\varepsilon_{RD}$, mass, q, N) = [X, X'| X ⊆ X', N'(X'), q'(Dim(σ(ε_{DS})))(ΣX')(Dim(σ(ε_{RD})(ΣX)),

x'∈X'↔ $I_Q(\varepsilon_{RD})$(X')(x')]

$I_Q(\varepsilon_{RD}$, count+parts, q, N) = [X, X' | X ⊆ X', q'(Count(X',σ(ε_{RD})))(C ount(X,σ(ε_{RD}))),

x'∈ X' ↔ $I_Q(\pi(\varepsilon_{RD}))$(X')(x')]

B.5.3 Domain specifications

For domain specifications consisting of a single 'bare' predicate P, with F_Q(D) = P, $I_Q(\varepsilon_{RD})$ is either P or P_0 (P_0 designating as before the characteristic predicate of a reference domain with source domain P), depending on whether the determinacy is indeterminate or determinate. This is expressed by (B7).

(B7) $I_Q(\varepsilon_{RD}) = I_Q(D, \langle\rangle, dt) = I_Q(d)(I_Q(D)) = F_Q(dt)(F_Q(D))$;
F_Q(indeterminate) = λZ.Z; F_Q(determinate) = λZ.Z_0

A conjunctive NP introduces a sequence of domain specifications in the annotation structure, and a disjunctive condition in the semantics.

NOTE 1 In DRT, it is common practice not to use disjunctions but negated conjunctions. For better readability, in this document, disjunctions are used, e.g. in the right-hand side of (B7).

(B8) $I_Q(\langle D_1,...D_k\rangle, \langle\rangle, dt) = \lambda z.\ I_Q(\langle D_1\rangle, \langle\rangle, dt)(z) \vee, ..., \vee I_Q(\langle D_k\rangle, \langle\rangle, dt)(z)$

A restrictive modifier in a domain specification restricts the domain to those elements that have both the characteristic properties of the source domain and those of the modifier. The semantics of a modification merges the information in the DRSs interpreting the head and the modifier into a single DRS as shown in (B9). For any determinacy 'dt':

(B9) $I_Q(D, \langle\mu_1, ..., \mu_k\rangle, dt) = \lambda z.([\ |\ I_Q(D)(z)] \cup [\ |\ I_Q(r_1)(z)] \cup ... \cup [\ |\ I_Q(r_k)(z)])$.

The use of (B7) and (B9) is illustrated by the interpretation in (B11) and (B12) of the sentence in (B10).

(B10) *Two of the five men who entered the bar whistled.*

The entity structure for "*Two of the five men who entered the bar*" has the reference domain "*the men who entered the bar*", the individuation 'count', the involvement '2' and the domain size specification '5'.

(B11) I_Q(⟨man, I_Q(*who entered the bar*)⟩, count, 2, 5) = [X, X' | X ⊆ X', |X| = 2, |X'| = 5,

x'∈ X' ↔ [man_0'(x'), (*who entered the bar*)'(x')]

where I_Q(*who entered the bar*) is (see B.5.4.4):

(B12) λz.[Y | y ∈ Y ↔ bar_0(y), |Y| = 1, y∈ Y → [E ⊆ enter | e∈ E → [agent(e,z), theme(e,y)]]]

Used non-restrictively (as a 'qualifier'), a modifier does not restrict the reference domain but provides additional information about it.

NOTE 2 Non-intersective adjectives, such as *fake*, *former* and *alleged*, present a notorious problem for compositional semantics, and are not considered in QuantML.

This is expressed in (B13) for modification with linear linking; inverse linking is considered in B.5.4.3.

(B13) a. For w = ⟨r, individual, linear⟩ or w = ⟨r, sampled, linear⟩, and involvement specification q:

$$I_Q(\varepsilon_D, \text{count}, q, N, w) = [\, X, X' \mid X \subseteq X', q'(|X|), N'(X'), x' \in X' \leftrightarrow I_Q(\varepsilon_D)(x'), x \in X \rightarrow I_Q(r)(x) \,]$$
$$I_Q(\varepsilon_D, \text{mass}, q, N, w) = [\, X, X' \mid X \subseteq X', q'(\Sigma X), N'(X'), x' \in X' \leftrightarrow I_Q(\varepsilon_D)(x'), x \in X \rightarrow I_Q(r)(x) \,]$$
$$I_Q(\varepsilon_D, \text{count+parts}, q, N, w) = [\, X, X' \mid X \subseteq X', q'(\text{Count}(X,\sigma(\varepsilon_D))), N'(\text{Count}(X,\sigma(\varepsilon_D))), x' \in X' \leftrightarrow I_Q(\pi(\varepsilon_D))(x'), x \in X \rightarrow I_Q(r)(x) \,]$$

b. For w = ⟨r, collective, linear⟩, the semantics is the same except that the right-hand side contains a condition of the form $I_Q(r)(X)$ instead of $x \in X \rightarrow I_Q(r)(x)$.

c. For w = ⟨r, unspecific, linear⟩, the semantics is the same except that the right-hand side contains a condition of the form $x \in X \rightarrow [\, x' \in X^* | \, x \in x' \vee x = x', I_Q(r)(x') \,]$ instead of $x \in X \rightarrow I_Q(r)(x)$.

B.5.4 Modification structures

B.5.4.1 General

The semantic interpretation of a restrictive modification with linear linking is shown in (B14), where r is any type of modifier; inverse linking is considered in B.5.4.3.

(B14)
$$I_Q(r, \text{individual}, \text{linear}) = I_Q(r, \text{sampled}, \text{linear}) = I_Q(r)$$
$$I_Q(r, \text{collective}, \text{linear}) = \lambda z. ([\, Y, E \mid |Y| > 1, z \in Y, e \in E \rightarrow I_Q(r)(Y) \,])$$
$$I_Q(r, \text{unspecific}, \text{linear}) = \lambda z. ([\, Y, E \mid |Y| \geq 1, (I_Q(r)(z) \vee (z \in Y, e \in E \rightarrow I_Q(r)(Y))) \,])$$

B.5.4.2 Relative clauses

A relative clause (RC) combines a set of events with sets of participants, just like a main clause, but with one of the arguments missing, whose role is played by the modified NP head. The interpretation of a clausal annotation structure, if it is fully scoped, has a most deeply nested sub-DRS embedded within the scope of all the quantifiers in the clause, in which the participants are linked to events in their respective semantic roles. This is the so-called 'nucleus' of the DRS. To construct the interpretation of the RC as a one-place predicate, the condition $R_a(e,z)$ that links a 'missing' participant (z) to an event (e) is inserted in the nucleus, unifying the event variables, and this participant variable is abstracted over. This is expressed schematically in (B15), where 'd' is any distributivity, 'IN(K,C,z)' is the operation of inserting in the nucleus of K the condition C with abstraction variable z, and '$v_n(K)$' is the nuclear variable of K.

NOTE A relative clause that contains quantifiers with equal scope has more than one nucleus, see Example (B18). These nuclei all have the same variable, so the value of 'v_n' is still uniquely defined; the insertion operation 'IN' is in that case repeated for each nucleus.

(B15) $I_Q(R_a, a_{rc}, d, \text{linear}) = \lambda z.\text{IN}(I_Q(a_{rc}), R_a(v_n(I_Q(a_{rc})), z))$

Predicates that are formed in this way can be used as $I_Q(r_i)$ terms in (B9) for interpreting a domain specification with a linearly linked RC-restriction. Inverse linking is considered in B.5.4.3.

B.5.4.3 Prepositional phrases

The semantic content of a PP structure includes a relation R_P, expressed by the preposition, and a participant entity structure corresponding to the NP, with the usual form $[\, X, X' \mid X \subseteq X', C_1, x' \in X' \leftrightarrow K_1]$.

The interpretation of the PP-structure is obtained by applying the insertion operation introduced in (B15) for K_1, using the relation R_P, with the result:

(B17) $I_Q(\langle m, \langle R_P, \varepsilon_P, d, linear\rangle\rangle) = \lambda z.\ [\ X, X' \mid X \subseteq X', C_1, x \in X \rightarrow (K_1 \cup [\ \mid R_P(x,z)\ ])\ ]$

If a PP modifier is inversely linked to an NP head, as in (B19), then the quantifier in the PP outscopes the one of the head. In such a case, the PP components R_P and ε_P are used in the construction of the modified head interpretation as shown in (B18), where 'd' is any distributivity.

(B18) a. $I_Q(\langle D, \langle R_P, \varepsilon_P, d, inverse\rangle\rangle) = \cup^{+}(I_Q(\varepsilon_P), I_Q(\varepsilon_D), I_Q\ (R_P, \varepsilon_P, d, inverse))$

b. $I_Q(R_P, \varepsilon_P, d, inverse) = [\ X \mid x \in X \rightarrow [\ Y \mid y \in Y \rightarrow R_P(x,y)\ ]]$

This is illustrated in (B19) for the NP *"Three students from all eight universities"*.

(B19) Three students from all eight universities [participated in the talks.]

$I_Q(\langle\langle student, \langle\langle from, \langle university, count, all, det, 8\rangle\rangle, individual, inverse\rangle\rangle, 3, indet\rangle)$

$= [\ U \mid U = university_0 \mid |U| = 8, u \in U \rightarrow [\ X \subseteq student \mid |X| = 3, x \in X \rightarrow from(x,u)\ ]]$

B.5.4.4 Possessives

Possessive structures can be formed with possessive pronouns and with genitives, which may include quantifiers, as in Example (B20). This gives rise to issues of distributivity and linking. Semantically, possessive structures can be analysed in terms of a possessor, a possessee and a binary relation 'Poss' (Reference [35]). In QuantML they are treated like PPs, where the relation between the discourse referents of the modified NP and the possessor is the Poss relation. Examples of the annotation of possessives and their semantics can be found in the Quantification Bank[37]. Inverse linking of possessive modifiers can be handled in the same way as in PP-modification, with 'Poss' instead of the relation denoted by the PP's preposition.

(B20) a. Two of every student's essays were lost.

b. The headmaster's children's toys all disappeared.

B.5.5 Proper names and definite descriptions

Proper names and definite singular NPs are annotated in QuantML using entity structures which are interpreted as introducing a singleton set as the reference domain. For example, the NP *"John Smith"* is annotated as in (B21a), which encodes the entity structure ε_P in (B21b), interpreted as in (B21c).

(B21) a.

```
<entity xml:id="x1" target="#m1" refDomain="#x2" individuation="count"
        involvement="all" size = "#n1"/>
<refDomain xml:id = "x2" target="#m1" components="#x3" determinacy="det"/>
<sourceDomain xml:id ="x3" target="#m1" pred="johnsmith"/>
<cardinality xml:id="n1" target="" numRel="equal" number="1"/>
```

b. $\varepsilon_P = \langle m1, \langle johnsmith, determinate\rangle, count, all, 1\rangle$

c. $I_Q(\varepsilon_P) = [\ X, X' \mid X \subseteq X', X' \subseteq X, |X'| = 1, x \in X' \leftrightarrow johnsmith_0(x)\ ]$

$= [\ X \mid x \in X \leftrightarrow johnsmith_0(x), |X| = 1\ ]$

Singular definite expressions such as *"the president"* are interpreted in the same way.

B.5.6 Event structures

If P_E is the characteristic predicate of a certain event domain, and k is the repetitiveness of a set of events, then the semantics of an event structure with that event domain and repetitiveness is specified by (B22):

(B22) $I_Q(P_E, k) = [\ E \mid k(|E|),\ e \in E \rightarrow P_E(e)\ ]$

An adjective or PP used predicatively, as in *“These boxes look heavy”*, is annotated using a <predication> element, and interpreted as describing a state, denoted by a copular verb. The semantics of a predication structure is thus similar to that of an event structure. For examples, see the Quantification Bank[37].

B.6 Link structures

B.6.1 Participation and predication links

B.6.1.1 Links with positive polarity

The semantic interpretation of a participation link structure is formed by the glue merge of the interpretations of the event structure, the participant structure and a glue DRS that contains the linking information. This is expressed in (B23), where ‘R’ is a semantic role, ‘d’ a distributivity, ‘s’ an event scope and ‘ξ’ an exhaustiveness, and the polarity is positive.

(B23) $I_Q(\langle \varepsilon_E, \varepsilon_P, R, d, s, \xi, \text{positive}\rangle) = \cup^{+}(I_Q(\varepsilon_P), I_Q(\varepsilon_E), I_Q(R, d, s, \xi, \text{positive}))$

The most common form of participation is non-exhaustive and has positive polarity and narrow event scope. For this form, the glue DRS depends only on the distributivity, as specified in (B24).

(B24) $I_Q(R, \text{individual, narrow, non-exhaustive, positive}) =$

$$I_Q(R, \text{sampled, narrow, non-exhaustive, positive}) = [\ X \mid x{\in}X \rightarrow [\ E \mid e{\in}E \rightarrow R(e,x)\ ]]$$

$$I_Q(R, \text{collective, narrow, non-exhaustive, positive}) = [\ E, X \mid e{\in}E \rightarrow R(e,X)\ ]$$

$$I_Q(R, \text{unspecific, narrow, non-exhaustive, positive}) = [\ X \mid x{\in}X \rightarrow [\ E \mid e{\in}E \rightarrow [\ x'{\in}X^{*} \mid x = x' \vee x{\in}x',\ R(e,x')\ ]]]$$

The structure of a participation link interpretation, also called a *‘plint structure’*, is determined by the information in a glue DRS. The glue merge has the effect that the discourse referents referring to a participant set (X) and an event set (E) are incorporated into a glue DRS, unifying the information in the DRSs representing the event set and the participants set.

For example, the annotation of the reading with individual distributivity of the sentence “*More than five hundred students protested twice”* includes a participation link which gives rise to the plint structure in (B25).

(B25) $[\ X \subseteq \text{student} \mid |X| > 500,\ x{\in}X \rightarrow [\ E \subseteq \text{protest} \mid |E| = 2,\ e{\in}E \rightarrow \text{agent}(e,x)\ ]]$

Predication links, used to annotate the semantic relation between a participant and a predicate expressed by a copular verb, have in common with participation links the specification of a distributivity, an exhaustiveness and a polarity. Their semantic interpretation is defined by (B26), identical to (B23) for participation links, but with a glue DRS containing a predicate that corresponds to the copular verb. (B27) specifies this glue DRS. Since these glue DRSs have the same structure as those for participation links, the semantic representations computed by (B26) are also structurally identical. The term ‘plint structure’ is therefore also used to refer to predication link interpretations.

(B26) $I_Q(\langle \varepsilon_P, \varepsilon_E, P, d, \xi, \text{positive}\rangle) = \cup^{+}(I_Q(\varepsilon_P), I_Q(\varepsilon_E), I_Q(P, d, \xi, \text{positive}))$

(B27) $I_Q(P, \text{individual/sampled}, \xi, \text{positive}\rangle) = [X \mid x\in X \rightarrow [E \mid |E| = 1, e\in E\rightarrow [\text{theme}(e,x), \text{attribute}(e,I_Q(P))\ ]]]$

(B28) $I_Q(P, \text{collective}, \xi, \text{positive}\rangle) = [\ X, E \mid |E| = 1, e\in E \rightarrow [\ \text{theme}(e,X), \text{attribute}(e, I_Q(P))\ ]]$

B.6.1.2 Links with negative polarity

Participation or predication links with negative polarity can have wide-scope or narrow-scope negation, as in the readings (B29b) and (B29c) of the sentence (B29a).

(B29) a. The unions do not accept the proposal.
b. It is not the case that (all) the unions accept the proposal.
c. (All) the unions do not accept the proposal (i.e. none of them accepts it).

The meaning of a link with wide-scope negation is the negation of the same positive link. This is expressed in (B30) for participation links, using the top-level DRS-negation introduced in Reference [27], symbolized as '~'. Similarly, for predication links.

(B30) $I_Q(\langle \varepsilon_E, \varepsilon_P, R, d, s, \xi, \text{neg-wide}\rangle) = \sim I_Q(\langle \varepsilon_E, \varepsilon_P, R, d, s, \xi, \text{positive}\rangle)$

Narrow-scope negation is interpreted as negation of the quantification over events and differs from wide-scope negation only for narrow event scope and non-collective distributivity. This is expressed in (B31), using a combinator $\cup^n$ that brings the event information of a participation link within the scope of a negation.

(B31) $I_Q\langle \varepsilon_E, \varepsilon_P, R, d, s, \xi, \text{neg-narrow}\rangle) = I_Q(\varepsilon_P) \cup (I_Q(R, d, s, \xi, \text{neg-narrow}) \cup^n I_Q(\varepsilon_E))$, with $\cup^n$ defined as:
$[\ X\ (Y, Z,...) \mid C_1, x \in X \rightarrow \neg K_1] \cup^n K_2 =_D [\ X\ (Y, Z,...) \mid C_1, x \in X \rightarrow \neg(K_1 \cup K_2)\ ]$

The glue DRS $I_Q(R, d, s, \xi, \text{neg-narrow})$ for narrow-scope negation is specified in (B32) for participation links, and follow by analogy for predication links.

(B32) a. For d = individual or sampled: $I_Q(R, d, \text{narrow}, \xi,\text{neg-narrow}) = [X \mid x\in X \rightarrow \neg\ [E \mid e\in E \rightarrow [R(e,x)]]$
b. $I_Q(R, \text{unspecific}, \text{narrow}, \xi, \text{neg-narrow}) = [X \mid x\in X \rightarrow \neg\ [E \mid e\in E \rightarrow [\ x'\in X^* \mid x = x' \vee x\in x', R(e,x')]]]$

B.6.1.3 Exhaustive participation

Exhaustive linking expresses that a predicate holds of no other individuals than those in the participant set. This is expressed by adding a condition of the form $x \in X \leftarrow P(x)$ to the set of conditions on the reference domain in case of non-collective distribution. (The notion of exhaustiveness does not apply in the case of collective distributivity.)

The most important cases, for positive and negative linking with individual distributivity, are shown in (B33). For unspecific distributivity, the interpretation follows by analogy with (B24).

(B33) If d = individual or d = sampled then

a. $I_Q(R, d, \text{narrow}, \text{exhaustive}, \text{pos}) = [\ X \mid x \in X \leftrightarrow [\ E \mid e \in E \rightarrow R(e,x)\ ]]$
b. $I_Q(R, d, \text{narrow}, \text{exhaustive}, \text{neg-narrow}) = [\ X \mid x \in X \leftrightarrow \neg[\ E \mid e \in E \rightarrow R(e,x)\ ]]$
c. $I_Q(R, d, \text{narrow}, \text{exhaustive}, \text{neg-wide}) = \sim[\ X \mid x \in X \leftrightarrow [\ E \mid e \in E \rightarrow R(e,x)\ ]]$

B.6.2 Scope links

Scope links determine how the interpretations of participation structures combine to form the semantic annotation of the clause. This is expressed in (B34), where σ is a scope relation, $\sigma' = I_Q(\sigma)$, and $L_i' = I_Q(L_i)$.

(B34) $I_Q(L_1, L_2, \sigma) = \sigma'(L_1', L_2')$

Three scope relations are distinguished:

a) wider: the first argument outscopes the second;

b) dual: the two arguments mutually outscope each other;

c) equal: the two arguments have equal scope.

The semantics of the scope links makes use of the three forms of DRS merging introduced in B.3; besides the standard DRS merge (∪) also the scoped merge (∪*) and the *dual-scoped merge* (∪♦), both defined below. Using these operations, (B35) describes the semantics of the scope relations.

(B35) a. $I_Q(\text{wider}) = \lambda x.\ \lambda y.\ x \cup^* y$
b. $I_Q(\text{equal}) = \lambda x.\ \lambda y.\ x \cup y$
c. $I_Q(\text{dual}) = \lambda x.\ \lambda y.\ x \cup^{\blacklozenge} y$

The scoped merge of two plint structures L_1' and L_2', of which the first represents a quantification that outscopes the second, combines the content of the respective DRSs into a single DRS, bringing the quantification over the participant set in the second argument within that of the first argument and unifies the two nuclei. The formal specification of the scoped merge is formulated in terms of pattern-matching based operations, exploiting the schematic structure of plint structures according to (B26); see Reference [9] for details.

As discussed in B.5.2 and B.5.3, quantifications with a specification of reference domain size or a restrictive modification require the introduction of a discourse referent to refer to the reference domain. This second discourse referent is just dragged along when plint structures are merged, hence their possible presence is not considered explicitly. Since plint structures for exhaustive quantification have the same schematic form as those for the non-exhaustive cases, exhaustiveness has no effect on the scoped merge operation.

A dual-scope relation between participation structures is interpreted as mutual outscoping. The dual-scoped merge operation, involved in the semantics of this relation, is a two-way application of the scoped merge operation, as shown in (B36). From the DRSs of its arguments a single DRS is constructed which branches out into two sub-DRSs, corresponding to the two instances of mutual outscoping; both sub-DRSs have as their nucleus the merge of the nuclei of their arguments (see Reference [9] for details).

(B36) $L'_1 \cup^{\blacklozenge} = [\, X_1, X_2 \,|\, C_1, C_2, x_1 \in X_1 \rightarrow [\, U \subseteq X_2 \,|\, u \in U \rightarrow (K_i \cup K_2) \,],$

$x_2 \in X_2 \rightarrow [\, V \subseteq X_1 \,|\, v \in V \rightarrow (K_i \cup K_2) \,]\,]$

B.7 Clause-level annotation structures

The semantics of the annotation of a clause with two scoped quantifications is defined by (B24) and (B26) to (B28) plus the definitions of the merge operations. For clauses with more than two scoped quantifications, the definitions of the scoped merge and the dual-scoped merge can be generalized to allow not only two plint structures as arguments, but also one plint structure and one structure which is the result of merging two plint structures (and this recursively). The latter approach is adopted here, thus keeping all scope relations and merge operations binary.

A fully connected clause annotation which *n* participation links includes *n*–1 scope relations, which together define a chain $[L_1, L_2,..\ L_n]$. Using the generalized scoped merge and dual-scoped merge, such a chain can be interpreted semantically by generalizing (B34) to (B36) for a chain of length $n > 2$, with scope relations σ_{12}, σ_{23}, ... $\sigma_{n\text{-}1,n}$. For details, see Reference [9].

(B37) $I_Q([L_1', L_2',.., L_n'] = \sigma_{12}(L_1', \sigma_{23}(L_2', ...\ \sigma_{n\text{-}1,n}(L_{n\text{-}1}', L_n')...)$

Annex C
(informative)

Example annotations with semantic interpretations

C.1 Overview and notation

This annex provides examples of the use of QuantML, including the semantics as described in Annex B. Default values of attributes are mostly not specified. For better readability of the structures defined by the abstract syntax, sharp brackets enclosing single items will often be suppressed, leading to, for example, ⟨m3,move⟩ rather than ⟨m3,⟨move⟩⟩. To simplify representations of involvement, size and repetitiveness, simple numerical values such as “3” and relative values such as “all” will sometimes be used, leading to, for example, @involvement="3" rather than @involvement="#n1", <cardinality xml:id="n1" target="#m2" numRel="equal" num="3"/>. More examples can be found in the Quantification Bank[37].

C.2 Participant scoping

(C1) All the students read at least three papers twice.

Reading a: Each of the students read three (possibly) different papers

Markables: m1 = All the students, m2 = the students, m3 = students, m4 = read, m5 = at least three, m6 = at least three papers, m7 = papers, m8 = twice

QuantML/XML annotation:

```
<entity xml:id="x1" target="#m1" refDomain="#x2" individuation="count" involvement="#n1"/>
<refDomain xml:id="x2" target="#m2" components="#x3" determinacy="det"/>
<sourceDomain xml:id="x3" target="#m3" pred="student"/>
<relativeSize xml:id="n1" pred="all"/>
<event xml:id="e1" target="#m4" pred="read" rep="#n2"/>
<cardinality xml:id="n2" target="#m8" numRel="equal" num="2"/>
<participation event="#e1" participant="#x1" semRole="agent" distr="individual"
        evScope="narrow"/>
<entity xml:id="x4" target="#m6" refDomain="#x5" individuation="count" involvement="#n3" />
<cardinality xml:id="n3" target="#m5" numRel="greater_or_equal" num="3"/>
<refDomain xml:id="x5" target="#m7" components= "#x6`’ determinacy="indet"/>
<sourceDomain xml:id="x6" target="#m7" pred="paper"/>
<participation event="#e1" participant="#x4" semRole="theme" distr="individual"
        evScope="narrow"/>
<scoping arg1="#x1" arg2="x4" scopeRel="wider"/>
```

Abstract syntax:

$A_{C1a} = \langle \varepsilon_E, \langle \varepsilon_{P1}, \varepsilon_{P2} \rangle, \langle L_{P1}, L_{P2} \rangle, sc_{12} \rangle$

$\varepsilon_E = \langle m3, read, 2 \rangle$, $\varepsilon_{P1} = \langle m1, \langle\langle m2, \langle student, determinate \rangle\rangle, count, all \rangle\rangle$,

$\varepsilon_{P2} = \langle m5, \langle\langle m6, \langle paper, indeterminate \rangle\rangle, count, \langle m4, \langle greater\_or\_equal, 3 \rangle\rangle\rangle\rangle$

$L_{P1} = \langle \varepsilon_E, \varepsilon_{P1}, Agent, individual, narrow \rangle$, $L_{P2} = \langle \varepsilon_E, \varepsilon_{P2}, Theme, individual, narrow \rangle$

$sc_{12} = \langle L_{P1}, L_{P2}, wider \rangle$