ISO 24617-12:2025(en)

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

Contents

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*, Working Group 2, Semantic annotation. Project leader was Harry Bunt, other contributors include James Pustejovsky, Kiyong Lee, Maxim Amblard, Johan Bos, Philppe de Groote, Bruno Guillaume, Chuyuan Li, Pierre Ludmann, Michel Musiol, Guy Perrier, Siyana Pavlova, and Sylvain Pogodalla.

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”*.

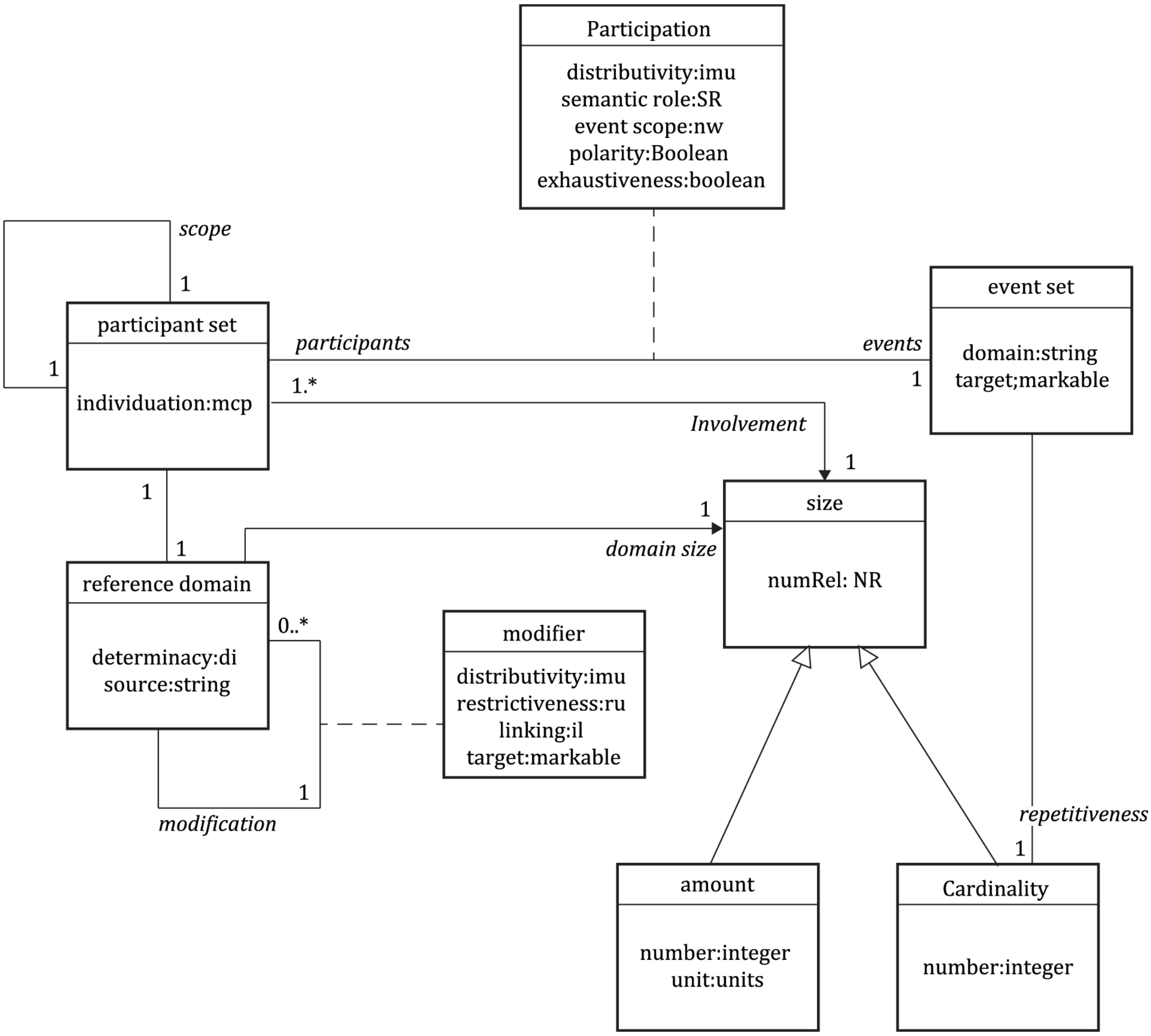
The annotation scheme defined in this document combines GQT with neo-Davidsonian event semantics,[13][33] which views the combination of a verb and its arguments as the participation in a certain semantic role of the entities denoted by the argument in the events denoted by the verb. This approach is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

6.2 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).

6.3 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 $|X| \geq 1$ and $|E| = 1$ are therefore suppressed in all DRSs.

Example 18

a) Thirty-two Chinese students enrolled.

b) $[\ X \mid x \in X \rightarrow [\ student(x), Chinese(x)\ ], |X| = 32\ ]$

c) $[\ E \mid e \in E \rightarrow enroll(e)\ ]$

d) $[\ X \mid x \in X \rightarrow [\ E \mid e \in E \rightarrow agent(e, x)\ ]\ ]$

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

$$[\ X \mid |X| = 32, x \in X \rightarrow [\ student(x), Chinese(x), [\ E \mid e \in E \rightarrow [\ enroll(e), agent(e, x)\ ]]]]$$

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) $[\ Y \mid Y \subseteq student, y \in Y \rightarrow [\ Z \mid Z \subseteq paper, |Z| = 3, z \in Z \rightarrow [\ E \mid E \subseteq read, e \in E \rightarrow [\ \mid agent(e, x), theme(e,y)\ ]\ ]\ ]\ ]\ ]$

on; see Example (C3) and Example (C4) in Annex C.

Annex A
(informative)

QuantML semantics

A.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.

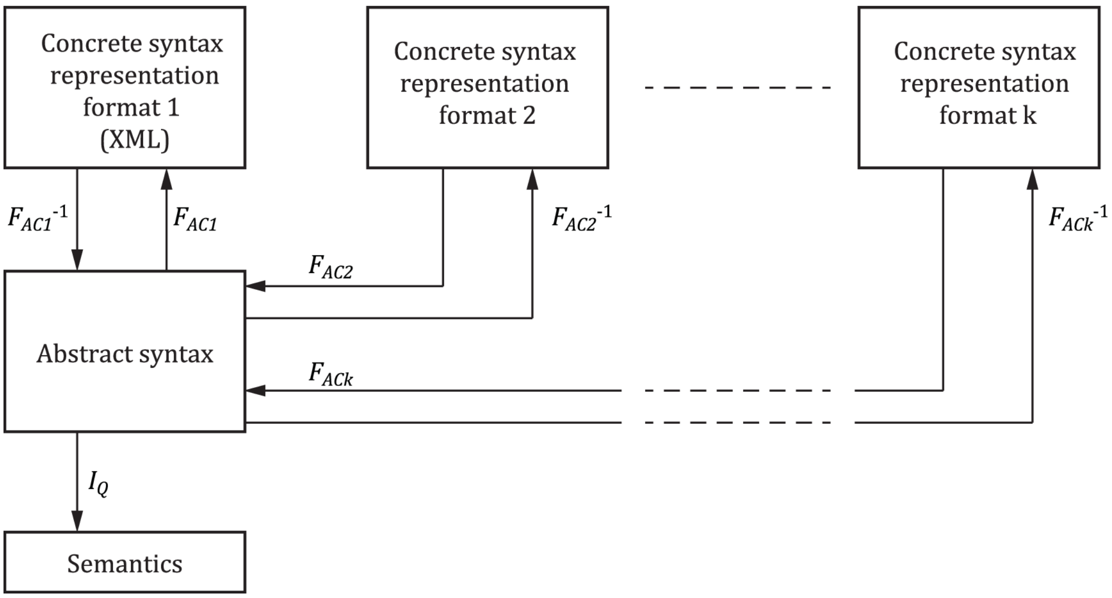


Figure B.1 — Three-layer architecture of QuantML

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.

Bibliography

[1] ABBOTT, B. (2004) Definiteness and indefiniteness. In: HORN, L., WARD, G. (eds.) *Handbook of Pragmatics*. Oxford: Blackwell, pp. 122–149

[2] BARKER, C. (2014). Scope. In: S. Lappin and C. Fox (eds.), *The Handbook of Contemporary Semantic Theory*. John Wiley, Chapter 2, pp. 40–76

[3] BARWISE, J. (1979). On branching quantifiers in English. *Journal of Philosophical Logic*. 8, pp. 47–80

[4] BARWISE, J., COOPER, R. (1981). Generalized Quantifiers and Natural Language. *Linguistics and Philosophy*. 4, pp. 159–219

[5] VAN BENTHEM, J. (1984). Questions about quantifiers. *Journal of Symbolic Logic*. 49 pp. 443–466

[6] BUNT, H. (1985). *Mass terms and model-theoretic semantics*. Cambridge University Press

[7] BUNT, H. (1996). Formal tools for describing and processing discontinuous constituency structure. In: BUNT, H., VAN HORCK, A. (eds.) *Discontinuous Constituency*. Berlin: Mouton de Gruyter, pp. 63–83

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

[9] BUNT H. (2023). The compositional semantics of QuantML annotations. In: *Proceedings 19th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-19)*, Nancy, France, pp. 3–13

[10] CHAMPOLLION, L. (2015) The interaction of compositional semantics and event semantics. *Linguistics and Philosophy*. 2015, 38, pp. 31–66. DOI: 10.1007/s10988-014-9162-8

[11] COOPER, R. (1983). *Quantification and Syntactic Theory*. Dordrecht: Reidel

[12] COPPOCK, E, BEAVER, D. (2015) Definiteness and determinacy. *Linguistics and Philosophy*. 38, pp. 377–435

[13] DAVIDSON, D. (1967). The logical form of action sentences. In: N. Rescher (ed.) *The Logic of Decision and Action*. Chapter 3. University of Pittsburgh Press

[14] ECKHARDT, R. (2000). Normal objects, normal worlds, and the meaning of generic sentences. *Journal of Semantics.* 16, pp. 237–278

[15] FREGE, G. (1879). *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens.* Halle: L. Nebert Verlag

[16] GROENENDIJK J., STOKHOF M., VELTMAN F. (1991). Contextually Restricted Quantification. *5th Conference on Semantics and Linguistic Theory,* pp. 112–129. DOI: 10.3765/salt/v0i0.2691

[17] HINTIKKA, J. (1973). Quantifiers vs. quantification theory. *Dialectica*. 27, pp. 329–358

[18] HOBBS, J. SHIEBER, S. (1987). An Algorithm for generating quantifier scopings. *Computational Linguistics*, 13(1–2), pp. 47–63

[19] ISO 24617-1:2012, *Language resource management — Semantic annotation framework (SemAF) — Part 1: Time and events (SemAF-Time, ISO-TimeML)*

[20] ISO 24617-4:2014, *Language resource management — Semantic annotation framework (SemAF) — Part 4: Semantic roles (SemAF-SR)*

[21] ISO 24617-6:2016, *Language resource management — Semantic annotation framework — Part 6: Principles of semantic annotation (SemAF Principles)*

[22] ISO 80000-1:2022, *Quantities and units — Part 1: General*

[23] ISO 24617-11:2021, *Language resource management — Semantic annotation framework (SemAF) — Part 11: Measurable quantitative information (MQI)*

[24] KAMP, H. and U. REYLE (1993). *From Discourse to Logic*. Dordrecht: Kluwer Academic Publishers

[25] KEENAN, E. and D. PAPERNO (eds.) (2017). *Handbook of Quantifiers in Natural Language*. Cham, Switzerland: Springer

[26] KEENAN, E. and D. WESTERSTAHL (1997). Generalized quantifiers in Linguistics and Logic. In: VAN BENTHEM, J., TER MEULEN, A. (eds.) *Generalized Quantifiers in Natural Language*. Foris, Dordrecht, pp. 837–993

[27] KRAHMER, E., MUSKENS, R. (1995) Negation and disjunction in discourse representation theory. *Journal of Semantics*. 12(4), pp. 357–376

[28] KRIFKA, M., PELLETIER, F.J. CARLSON, G. TER MEULEN, A. CHIERCHIA, G., LINK, G. (1995) Genericity: An Introduction. In: CARLSON, G., PELLETIER, F.J. (eds) *The generic book*. University of Chicago Press, pp. 1–124

[29] LEWIS, D. (1979) Scorekeeping in a Language Game. *Journal of Philosophical Logic*. 8 pp. 339–359

[30] MAY R. (1977) *The Grammar of Quantification*. Ph.D. Dissertation, MIT

[31] MAY, R. and A. BALE (2005) Inverse linking. In: M. EVERAERT and H. VAN RIEMSDIJK (eds.) The Blackwell Companion to syntax, Vol. 2. Oxford: Blackwell, Chapter 6, pp. 639–667

[32] MONTAGUE, R. (1974) The proper treatment of quantification in ordinary language. In: HINTIKKA, J., MORAVCSIK, J., SUPPES, P. (eds.) *Approaches to Natural Language*. Dordrecht: Reidel, pp. 221–242

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

[34] PARTEE B. (1988) Many Quantifiers. In: POWERS, J., DE JONG, K. (eds.) *Proceedings of the Eastern States Conference on Linguistics (ESCOL)*, pp. 383–402

[35] PETERS, S. WESTERSTÅHL, D. (2013) The Semantics of Possessives. *Language*. 89(4), pp. 713–759

[36] PUSTEJOVSKY, J., BUNT, H. ZAENEN, A. (2017) Designing Annotation Schemes: From Theory to Model. In: IDE, N., PUSTEJOVSKY, J. (eds) *Handbook of Linguistic Annotation*. Springer, Berlin

[37] Quantification Bank: https://sigsem.uvt.nl/QuantificationBank/

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

[39] SCHA, R. (1981) Collective, distributive and cumulative quantification. In: GROENENDIJK, J, STOKHOF, M. (eds.) *Formal Methods in the Study of Language*. Amsterdam: Mathematisch Centrum, pp. 483–512

[40] SZABOLCSI, A. (1997) *Ways of Scope Taking*. Dordrecht: Kluwer

[41] SZABOLCSI, A. (2008) Scope and binding. In: MAIENBORN, C., VON HEUSINGER, K, PORTNER, P. (eds.) *Semantics: An International Handbook of Natural Language Meaning*. Berlin: Mouton de Gruyter

[42] SZABOLCSI, A. (2010) *Quantification*. Cambridge University Press, Cambridge, UK

[43] WESTERSTÅHL, D. (1985) Determiners and context sets. In: VAN BENTHEM J., TER MEULEN, A. (eds.) *Generalized Quantifiers in Natural Language*. Dordrecht: Foris, pp. 45–71

[44] WINTER, Y. RUYS, E. (2011) Scope ambiguities in formal syntax and semantics. In: GABBAY, D., GUENTHNER, F. (eds.) *Handbook of Philosophical Logic* (2nd edition). Springer

[45] ZWARTS, F. (1994) Definite Expressions. In: ASHER, R.E. (ed.) *The Encyclopedia of Language and Linguistics*. Oxford: Pergamon