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)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

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"*.

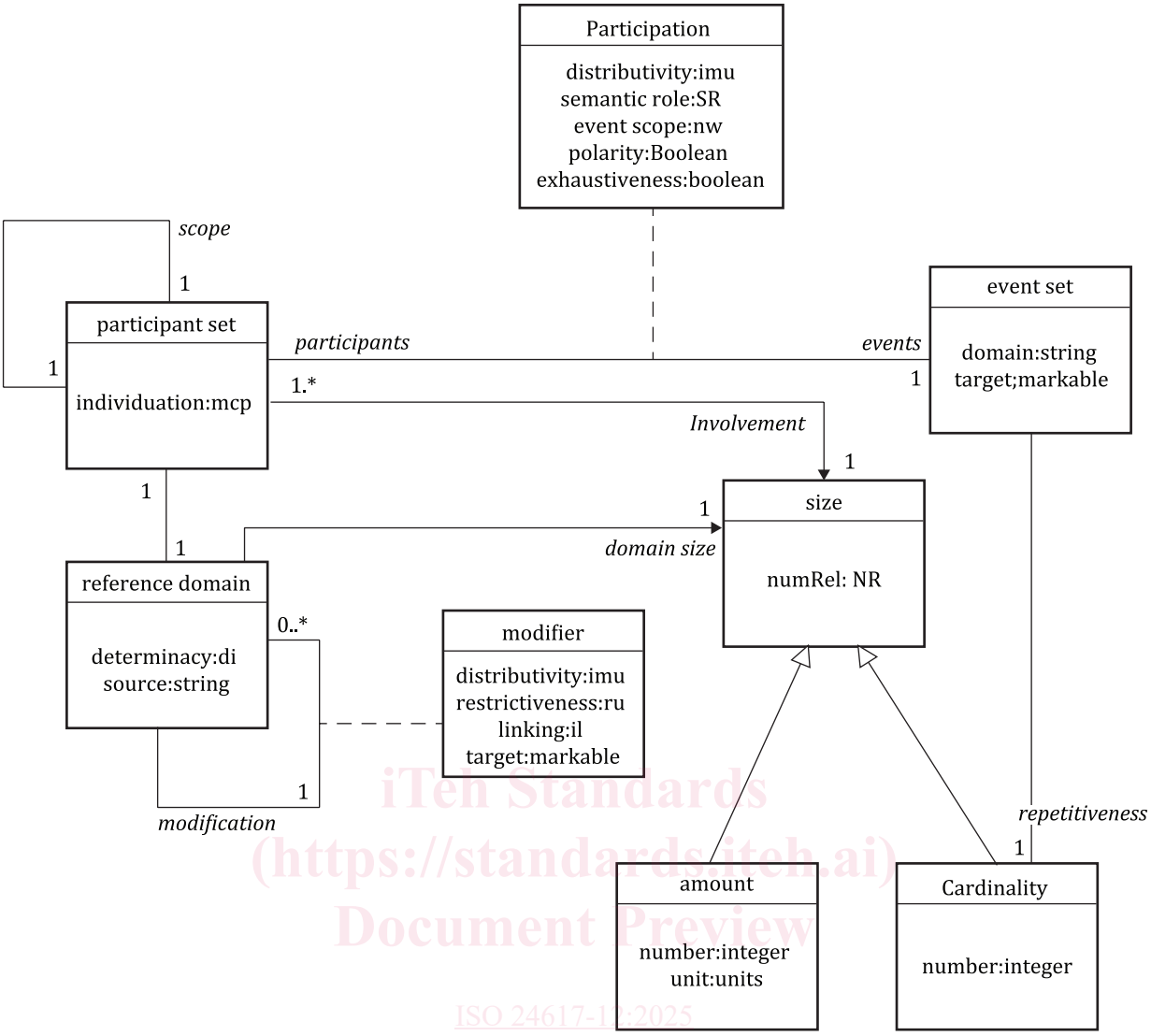
The annotation scheme defined in this document combines GQT with neo-Davidsonian event semantics,[13] [33] which views the combination of a verb and its arguments as the participation in a certain semantic role of the entities denoted by the argument in the events denoted by the verb. This approach is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