

QuantML: A new semantic annotation standard

Harry Bunt, Tilburg University

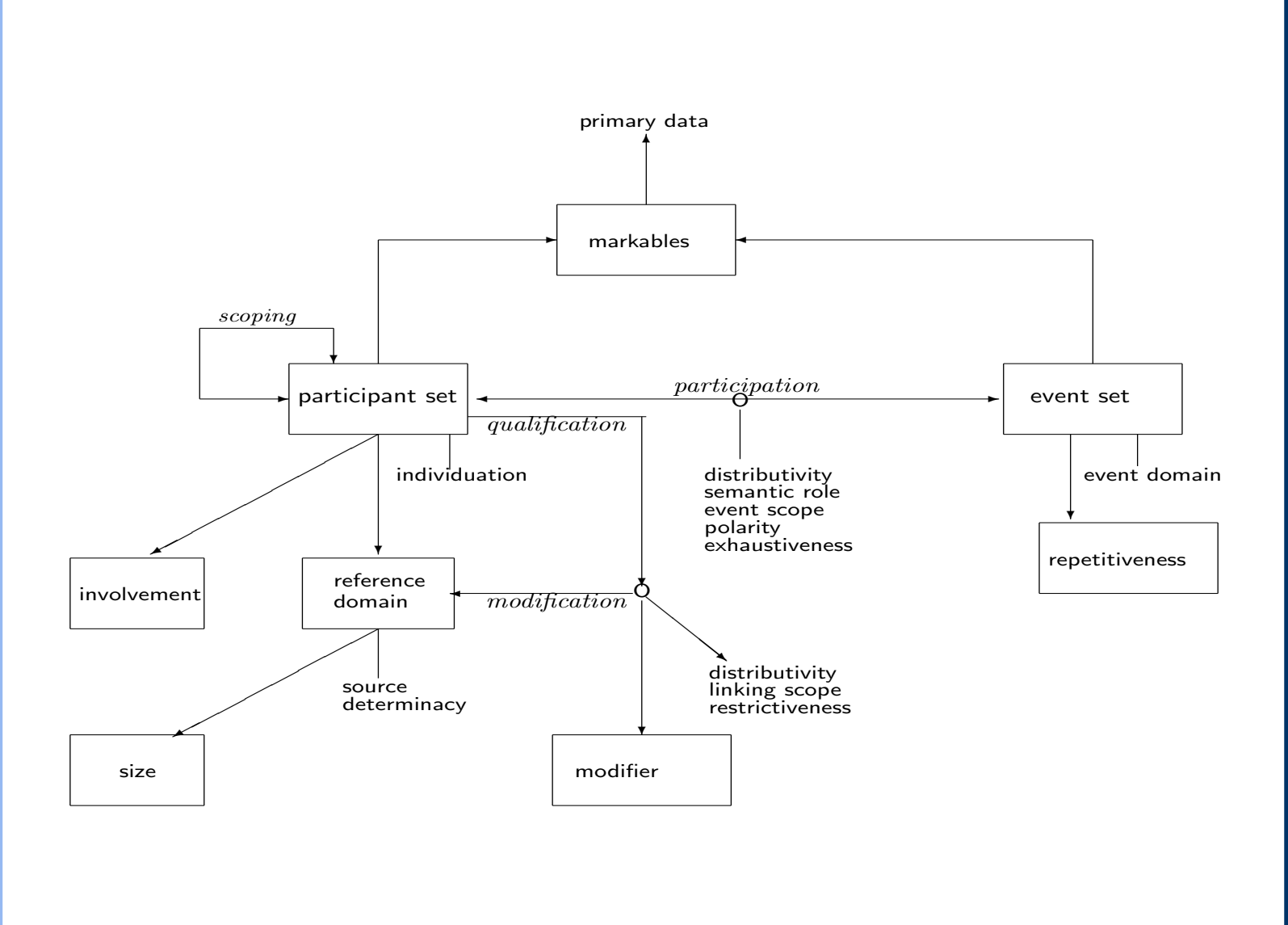
QuantML: Accepted as ISO Draft International Standard, May 8, 2024

Aim: Interoperable annotation scheme for quantification phenomena, compatible with existing ISO standards for semantic annotation, having besides a reference XXML-based representation format also an abstract syntax with a compositional semantics.

Theoretical background:

- Generalised Quantifier Theory:
 - quantification in natural language is *restricted to a certain domain*
 - quantifiers denote *properties of sets of individuals*
 - *noun phrases* are natural language quantifiers
- (Davidsonian) Event Semantics:
 - events are viewed as individuals
 - verb – NP combinations are viewed as *participation relations between sets of events and sets of (other) individuals*

QuantML metamodel:



Aspects of quantification to be expressed in annotations;

1. domain of a quantification
 - ❖ source domain (typically expressed by NP head)
 - ❖ reference domain (contextually determined part of source domain)
2. individuation of the source domain (count, mass, count-parts)
3. size of the reference domain
4. involvement of the reference domain (absolute or proportional)
5. size of groups of elements from a reference domain
6. distribution of an event participation relation (collective, individual, unspecific);
7. relative scopes of quantifications over participant sets
8. relative scope of quantifications over events and participants
9. modifier linking (inverse, linear)
10. polarity, exhaustiveness, repetitiveness

Main Requirements:

- ◆ Coverage: sufficiently expressive to represent a wide range of quantification phenomena.
- ◆ Flexibility: logically precise where appropriate as well as supporting the annotation of semantically vague or unspecific quantifications.
- ◆ Interoperability: supporting alternative representation formats
- ◆ Compatibility: as part of the ISO Semantic Annotation Framework, QuantML should have a triple-layered architecture consisting of (1) an abstract syntax, (2) a representation format, and (3) a semantics. (All annotation should have a well-defined semantics.)

Example: *One of the five men who came in whistled.*

Markables: m1 = one of the five men who came in, m2 = five men who came, m3 = men, m4 = men who came in, m5 = came in, m6 = whistled

```

Annotation: <entity xml:id="x1" target="#m1" domain="#x2" involvement="one"/>
<refDomain xml:id="x2" target="#m2" source="#x3" restrictions="#r1" determinacy="indet"/> />
<sourceDomain xml:id="x3" target="#m3" individuation="count" pred="man"/>
<relClause xml:id="r1" target="#m4" event="#e2" semRole="agent" restrictiveness="restrictive"/>
<event xml:id="e1" target="#m5" pred="come-in"/>
<participantLink event="#e1" participant="#x1" semRole="agent" distr="individual" eventScope="narrow"/>
<event xml:id="e2" target="#m6" pred="whistle"/>
<participantLink event="#e2" participant="#x1" semRole="agent" distr="individual" eventScope="narrow"/>
    
```

Semantic interpretation (simplified): $[\{X, X'\} \mid X \subseteq X', |X|=1, |X'|=5, x \in X' \leftrightarrow [\text{man}(x), \text{come-in}(x)], x \in X \rightarrow \text{whistle}(x)]$

More examples: see the QuantificationBank (<https://sigsem.uvt.nl/QuantificationBank/>)

Special features of QuantML: coverage includes:

- ◆ mass NP quantification and quantification over parts of individuals - “*The girls ate most of the chocolate*”; “*Tony ate two and a half pizzas.*”
- ◆ quantification with wide event scope and quantification in copular constructions - “*Some student demonstrated twice*”; “*These books look heavy*”
- ◆ restrictive and non-restrictive modification - “*The men who came in whistled*” / “*The men, who came in, whistled*”

Future work:

- ◆ Further development of the QuantificationBank.
- ◆ Explorations of the possibilities to combine (the use of) QuantML and other ISO annotation schemes (ISO-TimeML, SpaceML, reference annotation,...)

References:

- Amblard, M., K. Fort, B. Guillaume, Ph. De Groote, C. Li, P. Ludmann, M. Musiol, S. Pavlova, G. Perrier & S. Pogodalla: ‘The annotators did not agree on some of the guidelines examples’. In *Proceedings ISA-17@IWCS 2021, Nancy*.
- Bunt, H.: ‘Combining annotation schemes through interlinking’. In *Proceedings ISA-20,@LREC-COLING 2024, Turin*.
- Bunt, H. ‘The compositional semantics of QuantML annotations’. In *Proceedings ISA-19@IWCS 2023, Nancy*.
- Bunt, H., J. Pustejovsky & K. Lee: ‘Towards an ISO Standard for the Annotation of Quantification’, In *Proceedings LREC 2018, Miyazaki*.
- Malchanau, A., V. Petukhova & H. Bunt: ‘Fusing ISO 24617-2, Dialogue Acts and Application-Specific Semantic Content Annotations’ In *Proceedings ISA-20@LREC-COLING 2024, Turin*.