

Quantification annotation in ISO 24617-12

Harry Bunt¹, Maxime Amblard², Johan Bos³, Karen Fort², Philippe de Groote², Bruno Guillaume², Chuyuan Li², Pierre Ludmann², Michel Musiol², Siyana Pavlova², Guy Perrier², Sylvain Pogodalla²

¹Tilburg University ²LORIA, Nancy ³University of Groningen

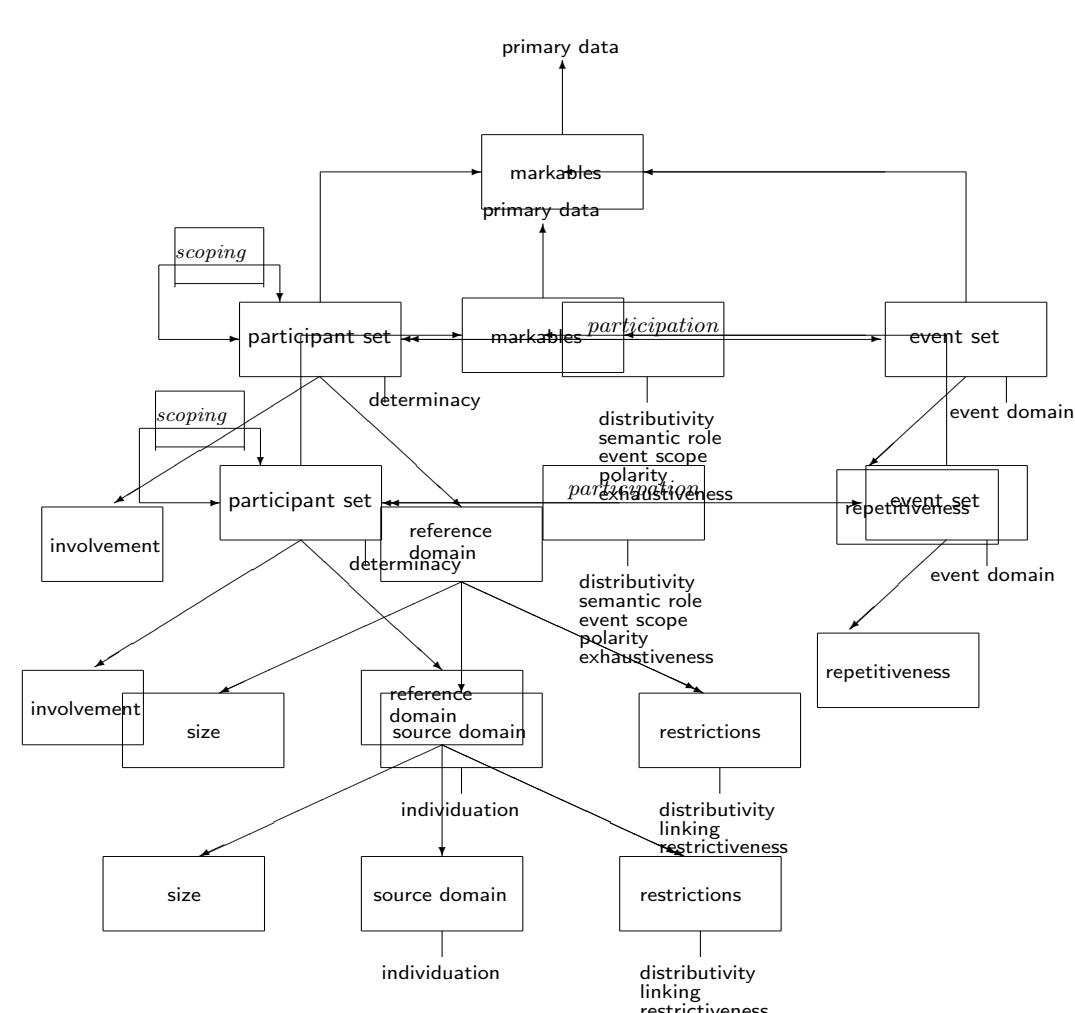
QuantML: Part of ISO Semantic Annotation Framework (ISO 24617). Project start 2019; suspended 2021 (Covid); restart 2022.

Aim: Define an annotation scheme for quantification phenomena, compatible with existing ISO standards for semantic annotation, having an abstract and a concrete syntax with a compositional semantics.

Theoretical background:

- Generalised Quantifier Theory:
 - quantification in NL is *restricted to a certain domain*, often expressed by NP head
 - quantifiers express *properties of sets of individuals*
- (Neo-)Davidsonian event semantics: *evnts viewed as individuals*
- Discourse Representation Theory

QuantML metamodel:



information to be contained 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 a predicate-argument relation (collective, individual. Unspecific);
7. relative scopes
8. modifier linking (inverse, linear)
9. polarity, exhaustiveness, repetitiveness

Main issues in 2022 Working Draft:

- ◆ determinacy, also in article-less languages. Determinate quantifiers: reference domain is a proper part of the source domain. Indeterminate quantifiers: reference domain = source domain.
- ◆ restrictive and non-restrictive modification of quantification domain by relative clauses, PPs and possessives.
- ◆ optimising transparency and readability of annotation representations

Example: *One of the 5 men who came in whistled.*

Markables: m1=one of the 5 men who came in, m2=the 5 men who came in, m3=men, m4=who came in, m5=came in, m6=whistled

```

Annotation: <entity xml:id="x1" target="#m1" domain="#x2" involvement="one" definiteness="indet"/>
<refDomain xml:id="x2" target="#m2" source="#x3" restrictions="#r1"/>
<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="#m3" pred="whistle"/>
<event xml:id="e2" target="#m5" pred="come-in"/>
<entity xml:id="x3" target="#m4" domain="#x4" involvement="6.5" definiteness="indef"/>
<sourceDomain xml:id="x4" target="#m5" individuation="count/parts" pred="pizza"/>
<participantLink event="#e1" participant="#x1" semRole="agent" distr="individual" eventScope="narrow"/>
    
```

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

Annotation representations: [1] QuantML/XML, [2] NSG graphs (Nancy Semantics Group, Amblard et al. 2021)), [3] DRS/SBN (Simplified Box Notation, Bos 2021), [4] DRS Graphs (Abzianidze et al. 2020), [5] Model-instance graphs (Bunt 2022)

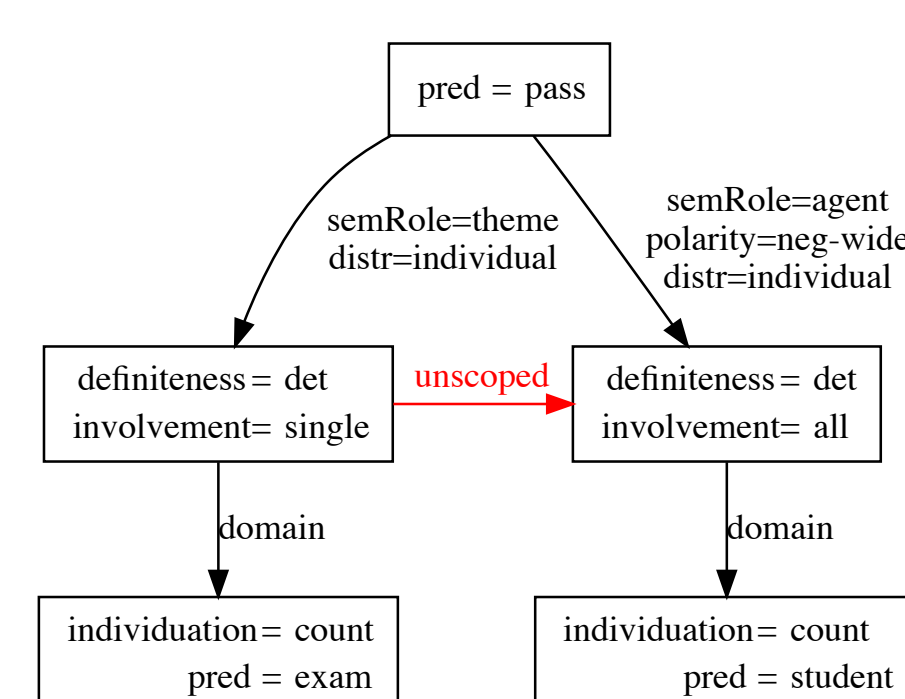
Alternative representations for "Not all the students passed the exam":

Annotation as DRS in Simplified Box Notation

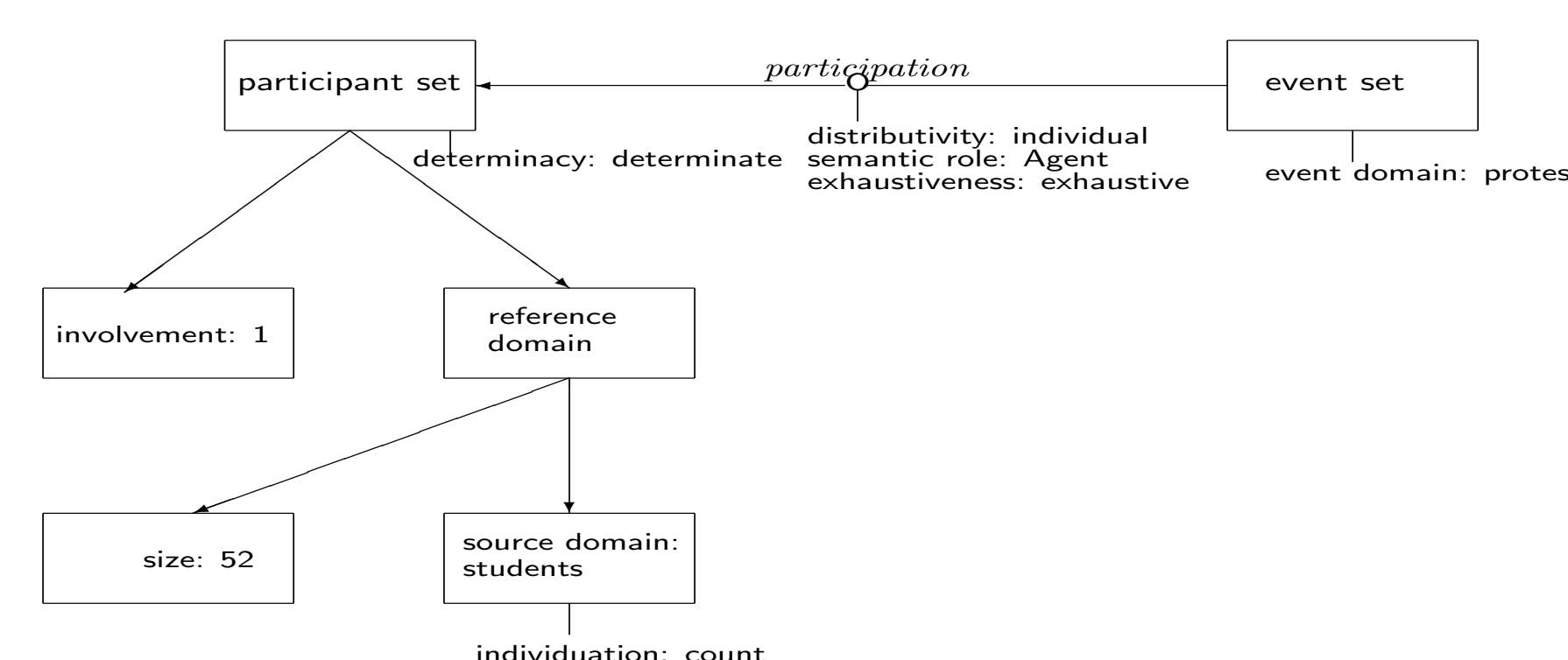
```

student.n.01  NEGATION -1  % Not
              NEGATION -1  % all
              % the
              % students
              NEGATION -1  %
              pass.v.14  Agent -1  % passed
              Theme +1   %
exam.n.01     % the exam.
    
```

Annotation as NSG Graph



Annotation as model-instance graph



Next steps: Implementation of latest insights in abstract syntax, semantics, and representation formats. Streamlining of metamodel and representation formats to make QuantML optimally transparent and intuitive. Development of online repository of annotated examples, the QuantML Bank.

References:

- Abzianidze, L., J. Bos & S. Oepen: 'Dressing up discourse representation structures as graphs.' In *Proc. CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing*.
- 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 *Proc. ISA-17 Shared Task: Quantification Annotation Challenge*, 2021.
- Bos, J. 'Variable-free discourse representation structures'. In *SemArchiv*, 2021.
- Bunt, H., J. Pustejovsky & K. Lee: 'Towards an ISO Standard for the Annotation of Quantification', In *Proceedings LREC 2018*.
- Bunt, H.: 'Intuitive and Formal Transparency in Annotation Schemes'. In *Proceedings ISA-18, 2022*.