

The semantic annotation of quantification

Harry Bunt
TiCC, Tilburg Center for Cognition and Communication
Tilburg University, The Netherlands
harry.bunt@uvt.nl

Abstract

This paper presents an approach to the annotation of quantification, developed in the context of an ISO project aiming to define standards for semantic annotation. The approach is illustrated for a range of quantification phenomena, including cumulative, collective, and group quantification.

1 Introduction

In 2012, two ISO standards for semantic annotation were established, one for time and events (ISO 24617-1), and one for dialogue acts (ISO 24617-2); others are under development for semantic roles, spatial information, and discourse relations. Quantification turns up as a problem in nearly all of these efforts. ISO 24617-1 has some provisions for dealing with quantification (see Pustejovsky et al., 2010), but these are too limited and do not always give correct results (Bunt and Pustejovsky, 2010).

The annotation of quantification faces three main issues:

1. Which set of semantic features, expressed most likely as XML attributes and values, adequately characterize a wide range of forms of quantification;
2. Is it possible to define a formal semantics of these expressions;
3. How can the relevant features (attributes and values) be defined as components of the semantic annotations in a way that respects compositional semantics?

Bunt (2005) proposed a way to representing quantifications in terms of feature structures, however, in this proposal the properties of a quantification are all expressed as properties of an event, which is inconvenient for annotators and is hard to combine with a compositional NP semantics. This paper presents an approach where quantification features are distributed over annotation structure components in a way that corresponds to their linguistic expression in syntactic structures, and shows the semantic adequacy of the proposal by a compositional translation into discourse representation structures.

2 Aspects of quantification

Quantification occurs when a predicate is applied to sets of arguments. Questions then arise concerning the precise way that the predicate is applied to members of these sets. As an example, consider the sentence (1a); some of the questions that may be asked (and answered) are (1b) - (1f):

- (1) a. *Although a threat had been made before, three men rather unexpectedly moved both pianos.*
 - b. How many men were involved? (Answer: *Three.*)
 - c. How many pianos were involved? (Answer: *Two.*)
 - d. Did the same men move both pianos? (Answer: *Yes.*)
 - e. Did the men act collectively or individually? (Answer: *Collectively, probably.*)
 - f. Were the pianos acted upon collectively or individually? (Answer: *Individually, probably.*)

Given the *restriction* part of a quantification, that specifies a domain from which elements can be taken that participate in certain events individually, in groups, or collectively, the *distribution* is a function that computes the set of those entities that participate in the events as agents, as themes, as instruments, etc. We call such a set of participants a *predication domain*, and the domain, defined by the restriction, the *reference domain*. In the case of individual distribution, the two are the same.

A common function of proportional quantifier words like *all*, *some*, and *most*, and of absolute quantifiers like *three*, *more than five*, *2 litres of* is to specify the fraction of the reference domain that is involved in the events under consideration. Numerical and amount quantifiers may also be used to indicate the size of a reference domain, like *twelve* in *The twelve students in this room all speak two languages*.

Proportional and absolute quantifiers can also be used to indicate the number/amount of a predication domain per element of another predication domain, like *five* in *Each of the dogs ate five sausages*.

Some of the most important aspects of quantification, distinguished in (Bunt, 1985) are:

- (2) 1. the quantifier's restriction, describing the reference domain of the quantification;
2. the distribution, defining the predication domain;
3. size of the reference domain;
4. involvement of the reference domain (in absolute or relative terms);
5. relative scoping of the quantifications associated with argument NPs;
6. scoping of NP-quantifications relative to quantified events;
7. size of groups of elements from a reference domain;
8. number of elements of a reference domain involved per element of a predication domain.

3 Events, participants, and quantification annotation

In a davidsonian approach to meaning, we may view the combination of a verb with its arguments as introducing a set of events ('eventualities', more generally) of the type indicated by the verb, and with a number of properties concerning the way in which the participants of these events are involved. Applying this view to sentence (1a), we obtain a descriptions in terms of a set of move-events, a set of men, participating in these events as agents, and a set of pianos that participate as themes. This is exactly what is expressed in an annotation of semantic roles according to the ISO standard 24617-4 under development (see Bunt & Palmer, 2013):

- ```
(3) [v1:] <event xml:id="e1" target="#m2" eventType="move"/>
 [p1:] <participant xml:id="x1" target="#m1" entityType="man"/>
 [p2:] <participant xml:id="x2" target="#m3" entityType="piano"/>
 [L1:] <srLink event="#e1" participant="#x1" semRole="agent"/>
 [L2:] <srLink event="#e1" participant="#x2" semRole="theme"/>
```

Since quantification occurs when two or more sets of arguments are related by a predicate, we can view the quantifications in (1a) as due to the Agent and Theme predicate relating sets of events and participants. This information can thus be represented as properties of the semantic role links (<srLink>), adding features to these links as follows, where the feature @signature="set" expresses that the sentence is about sets of events and participants :

- ```
(4) [v1:] <event xml:id="e1" target="#m2" eventType="move" signature="set"/>
      [p1:] <participant xml:id="x1" target="#m1" entityType="man"
           signature="set" involvement="3"/>
      [p2:] <participant xml:id="x2" target="#m3" entityType="piano"
           signature="set" definiteness="def" involvement="2"/>
      [L1:] <srLink event="#e1" participant="#x1" semRole="agent"
           distribution="collective"/>
      [L2:] <srLink event="#e1" participant="#x2" semRole="theme"
           distribution="individual"/>
```

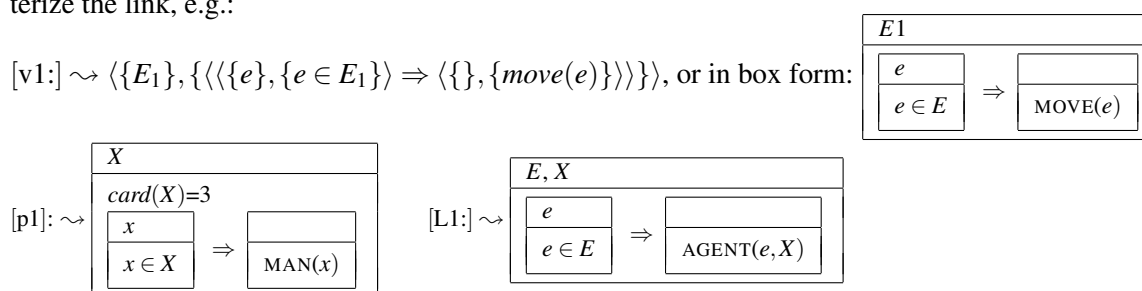
The correctness and usefulness of annotating quantification this way depends on how well it deals with the three issues mentioned in Section 1: expressive adequacy of attributes and values; their semantic adequacy; and compatibility with compositional phrase semantics. These issues are addressed next.

4 Representational and semantic adequacy

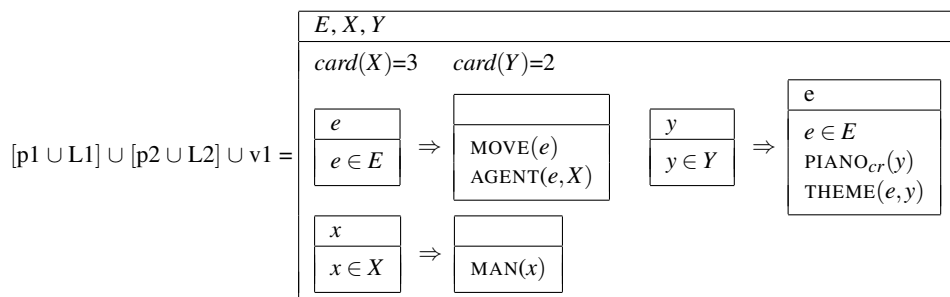
Of the information types listed in (2), those numbered 1, 2, 4, and 8 can be represented by the attributes and values shown in (4). For types 3, 5, 6 and 7, the attribute @cardinality is defined for <event> (for “say twice” etc.) and <participant> elements; the attribute @outScoping allows the expression of relative scope restrictions; and the values of the @groupCard attribute can be used to indicate group sizes in group quantifications. This provides the expressive power to represent a wide range of quantification phenomena.

The issue of compatibility with compositional NP semantics arises because we propose to represent some of the properties of a quantification as parts of semantic role links, where traditionally the semantic representation of quantification is considered to be part of NP semantics. This is in particular the case with distribution, as in (4). Having @distribution as an attribute of <participant> elements would run into problems for a sentence such as *The men had a beer before they moved the piano*, where the subject NP should be interpreted individually for the drinking, but collectively for the lifting.

The semantic adequacy of the proposed annotation format can be shown by defining a systematic translation of annotations into DRSs, following Bunt (2011; 2013). XML elements describing sets of events or participants, like those in (4), are translated to a DRS which introduces a higher-order (i.e. set-valued) discourse marker,¹ and conditions translating the other features. DRSs interpreting linking elements introduce discourse markers for the linked sets of elements, and conditions that further characterize the link, e.g.:



The merge of these DRSs plus those translating [p2] and [L2] yields the satisfactory result:²



Note that in this example both NPs outscope the verb; their relative scoping is irrelevant since the group of three men acted collectively, as a single entity. While verbs very often have narrow scope relative to argument NPs, this is not always the case, as (5) illustrates. Using the attributes and values introduced so far, the wide-scope interpretation of “die” is easily annotated:³

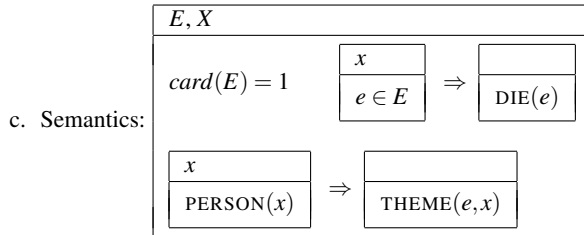
- (5) a. *Everybody will die.*
 b. Annotation for wide-scope ‘die’:

¹The presentation is simplified here; see Bunt (2013) for the use of pairs $\langle m, x \rangle$ of markables and discourse markers, where the markables make sure that only the intended marker variables are unified upon DRS-merging.

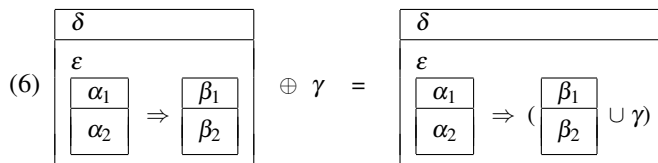
²The subscript ‘cr’ (for ‘contextually relevant’) indicates the interpretation of the definiteness of the NP *both pianos*.

³We assume here an approach to semantic roles which allows an event to have more than one theme, such as the LIRICS annotation scheme (Petukhova & Bunt, 2007).

```
[v1:] <event xml:id="e1" target="#m1" eventType="die" signature="set"
      cardinality="1"/>
[p1:] <participant xml:id="x1" target="#m2" entityType="person"
      signature="set" involvement="all"/>
[L1:] <srLink event="#e1" participant="x1" semRole="theme"
      outScoping="#e1 #x1"/>
```



Cases of quantified NPs with unequal scope are readily annotated in the format outlined here, but in order to support these annotations by a well-defined semantics a new kind of merge operation on DRSs is needed, the *scoped merge*, which is defined as follows:



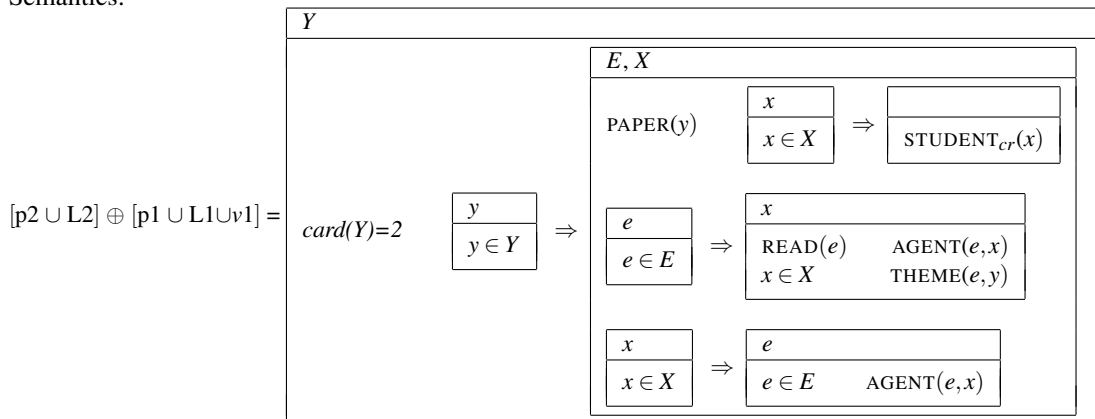
For instance, applied to the classical scoping example (7a), we obtain the semantics (7c) of the annotation (7b):

(7) a. *All the students read two papers*

b. Annotation for wide-scope *two papers*:

```
[v1:] <event xml:id="e1" target="#m1" eventType="read" signature="set"/>
[p1:] <participant xml:id="x1" target="#m2" entityType="student"
      definiteness="def" signature="set" involvement="all"/>
[p2:] <participant xml:id="x2" target="#m3" entityType="paper"
      signature="set" cardinality="2" involvement="all"/>
[L1:] <srLink event="#e1" participant="x1" semRole="agent"
      distribution="individual" outScoping="#x1 #e1"/>
[L2:] <srLink event="#e1" participant="x2" semRole="theme"
      distribution="individual" outScoping="#x2 #x1"/>
```

c. Semantics:



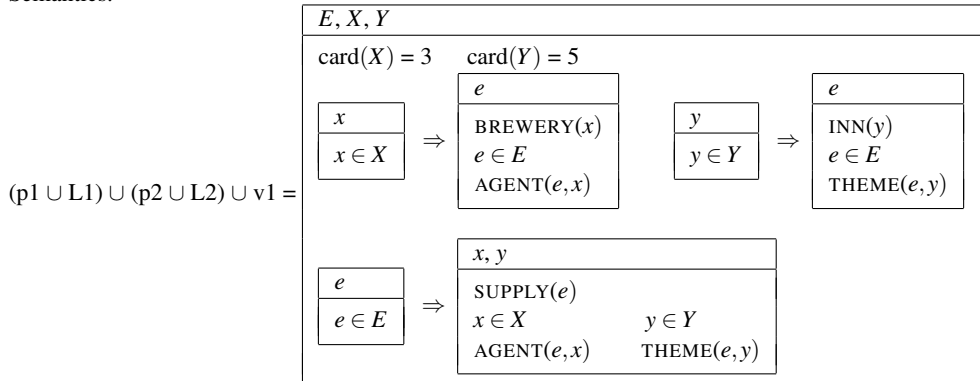
Besides scoped quantifications, also unscoped, partially scoped, and equally scoped quantifications have to be considered. Partially and unscoped cases, where there is no or incomplete information about relative scoping, are easily annotated by not specifying values for @outScoping attributes, and interpreted with underspecified DRSs (see Reyle 1993; 1994). Equally scoped quantifications, as occurring in *cumulative quantification* (Scha, 1981) and in group quantification (Bunt, 1985), can be annotated using the attribute @eqScope, as shown in (8) and (9). The semantics is obtained simply by using the ordinary rather than the scoped merge of the sub-DRSs.

(8) a. *Three breweries supplied fifteen inns.*

b. Annotation of cumulative reading (*In total 3 breweries supplied in total 15 inns*):

```
[v1:] <event xml:id="e1" target="#m2" eventType="supply" signature="set" />
[p1:] <participant xml:id="x1" target="#m1" entityType="brewery"
signature="set" involvement="3" />
[p2:] <participant xml:id="x2" target="#m3" entityType="inn"
signature="set" involvement="15" />
[L1:] <srLink event="#e1" participant="#x1" semRole="agent"
distribution="individual" eqScope="#x1 #x2" />
[L2:] <srLink event="#e1" participant="x2" semRole="theme"
distribution="individual" eqScope="#x2 #x1" />
```

c. Semantics:



Group quantification, as in *Three boys played soccer against five girls* on the reading where teams of 3 boys played against teams of 5 girls, can be annotated by using, besides the @eqScope attribute, the @groupCard attribute. The semantics is obtained by interpreting the participant annotations as DRSs introducing discourse markers for sets of sets of 3 boys and 5 girls.

```
[e1:] <event xml:id="e1" target="#m2" eventType="play" signature="set" />
[p1:] <participant xml:id="x1" target="#m1" entityType="boy"
signature="set" groupCard="3" />
[p2:] <participant xml:id="x2" target="#m3" entityType="girl"
signature="set" distribution="group" groupCard="5" />
(9) [L1:] <srLink event="#e1" participant="x1" semRole="agent"
distribution="group" eqScope="#x1 #x2" />
[L2:] <srLink event="#e1" participant="x2" semRole="co-agent"
distribution="group" eqScope="#x2 #x1" />
```

5 Concluding Remarks

We have described the essentials of a way of annotating quantification phenomena where the features that characterize different forms of quantification and their properties are distributed over components of annotation structures in a way that corresponds to their linguistic expression (e.g., involvement and cardinality are features of <participant> components, corresponding to their expression in NPs).

Taking a davidsonian approach, we have introduced attributes and values for events, event participants, and the semantic roles of participants in events. However, the view of quantification which underlies this is more general; when predicate and argument structures are used, rather than events with participants and semantic roles, that is easily accommodated, by introducing <predicate>, <argument>, and <argLink> elements, and e.g. attaching distribution and scoping features to the letter, as in:

```
(10) <argLink pred="#P1" arg="#x1" argNum="arg1" distr="collective" />
```

Quantification over events, time and place can be annotated in a similar way. As suggested by Lee & Bunt (2012), Temporal quantification can be annotated by adding features to the <event> and <timeX3> elements defined in ISO 24617-2 and to the <tLink> that represents the semantic relation between a set of events and a set of temporal entities (as in *Most of the professors teach every Monday*).

Clearly, something similar can be done for the annotation of quantifications over events and space, as occurring in *Policemen can be found at every streetcorner*, adding the same features to the elements defined in the ISO-Space language under development (see Pustejovsky et al., 2012).

References

- Bunt, H. (1985). *Mass Terms and Model-Theoretic Semantics*. Cambridge University Press.
- Bunt, H. (2005). Quantification and Modification Represented as Feature Structures. In *Proceedings 6th International Workshop on Computational Semantics (IWCS-6)*, Tilburg, Netherlands, pp. 54–65.
- Bunt, H. (2011). Abstract syntax and semantics in semantic annotation, applied to time and events. In E. Lee and A. Yoon (Eds.), *Recent Trends in Language and Knowledge Processing*. Seoul: Hankuk-munhwasa.
- Bunt, H. (2013). A Methodology for Designing Semantic Annotations. *Language Resources and Evaluation (forthc.)*.
- Bunt, H. and M. Palmer (2013). Conceptual and Representational Choices in Defining an ISO Standard for Semantic Role Annotation. In *Proceedings of the Ninth Joint ACL-ISO Workshop on Interoperable Semantic Annotation ISA-9*, Potsdam.
- Bunt, H. and J. Pustejovsky (2010). Annotation of temporal and event quantification. In *Proceedings of the Fifth Joint ACL-ISO Workshop on Interoperable Semantic Annotation ISA-5*, Hong Kong, pp. 15–22.
- ISO24617-1:2012 (2012). *Language Resource Management - Semantic Annotation Framework, Part 1: Time and Events (SemAF/Time)*. ISO Standard 24617-1, March 2012, Geneva. March 2012.
- ISO24617-2:2012 (2012). *Language Resource Management - Semantic Annotation Framework, Part 2: Dialogue Acts (SemAF/Dialogue acts)*. ISO Standard 24617-2, September 2012, ISO, Geneva.
- Lee, K. and H. Bunt (2012). Counting time and events. In *Proceedings of the Eighth Joint ACL-ISO Workshop on Interoperable Semantic Annotation ISA-8*, Pisa, pp. 34–41.
- Petukhova, V. and H. Bunt (2008, May). LIRICS Semantic Role Annotation: Design and Evaluation of a Set of Data Categories. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC 2008)*, Genova, Italy.
- Pustejovsky, J., K. Lee, H. Bunt, and L. Romary (2010). ISO-TimeML: An International Standard for Semantic Annotation. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, Malta, Paris, pp. 394–397. ELDA.
- Pustejovsky, J., J. Moszkowics, and M. Verhagen (2012). The Current Status of ISO-Space. In *Proceedings of Joint ISA-7, SRSL-3 and I@MRT Workshop at LREC 2012*, Istanbul, Paris. ELDA.
- Reyle, U. (1993). Dealing with Ambiguity by Underspecification. *Journal of Semantics* 10(2), 123–179.
- Reyle, U. (1994). Monotonic Disambiguation and Plural Pronoun Resolution. In H. Kamp (Ed.), *Ellipsis, Tense, and Questions. Esprit project 6852 DYANA-2, Deliverable R2.2.B*. Stuttgart: IMS.
- Scha, R. (1981). Distributive, Collective and Cumulative Quantification. In J. Groenendijk and M. Stokhof (Eds.), *Formal Methods in the Study of Language*. Amsterdam: Mathematical Center.