Observations on the use of QuantML

Harry Bunt
Department of Cognitive Science and Artificial Intelligence
School of Humanities and Digital Sciences
Tilburg University, Tilburg, Netherlands
bunt@ilburguniversity.edu

Abstract

This note contains some observations of the use of QuantML in annotating and interpreting some of the sentences in the Quantification Challenge test suite, used in the ISA-17 shared task on quantification annotation.[1]

1 Introduction

1.1 QuantML annotation

Inspired by the theory of generalized quantifiers (GQT), QuantML (IS0 WD 24617-12, 2020) focuses on quantifiers as expressed by noun phrases. This raises the question how lexical and morphosyntactic properties of noun phrases relate to aspects of quantification.

Section 2 discusses (1) the @definiteness attribute and the issue of determinacy vs. definiteness of an NP; (2) the annotation and interpretation of anaphoric possessive nominals, like 'their apartment'; (3) the annotation of deictic personal pronouns (I, you) and locative deictic expressions ('here', 'everywhere'); (4) the annotation and interpretation of the scope of negations relative to quantifier scopes.

2 Issues emerging from the Quantification Challenge

2.1 Definiteness and Determinacy

Definiteness is a morphological category with a language-dependent marking. Besides NPs with a definite determiner or suffix, other expressions that are also considered to be definite include NPs with a demonstrative pronoun (*"those shoes"*) or a 'universal determiner'(*"every man"*), and singular NPs with a possessive pronoun (*"my house"*) or a genitive (*"Mary's dog"*). Proper names and personal pronouns such as *"she"* and *"you"* are also usually counted as definite (for an overview see Zwarts, 1994). . NPs with a possessive expression in central determiner position and without a pre-determiner are definite when the possessive expression is a pronoun or a proper name, as in *"my house"* or *"Toms two children"*, but in general NPs with a possessive determiner expression are indefinite (Peters and Westerståhl, 2013), contrary to widespread belief (see e.g. Abbott, 2004).

Determinacy, on the other hand, is the semantic property of referring to some particular and determinate entity or collection of entities (Coppock and Beaver, 2015; Peters and Westerståhl, 2013; Westerståhl, 1985). Definiteness and determinacy are related since definite expressions are ordinarily used in a determinate sense, but the relation between definiteness and determinacy is not straightforward. The semantic difference between definite and indefinite expressions has been discussed in terms of familiarity and novelty (e.g. Heim, 1982), salience (Lewis, 1979), uniqueness, and existence presuppositions (see e.g. Coppock and Beaver, 2015). The familiarity/salience intuition about definite NPs can be accommodated in a GQT framework by assuming the reference domain of a quantification to contain familiar or particularly salient entities.

Indefinite NPs, definite plural NPs, and mass NPs differ from definite singular NPs in not carrying a uniqueness presupposition, but when used to quantify over event participants they all carry an existence presupposition. This is reflected in the semantics of QuantML annotations by the use of discourse referents that designate non-empty sets.

Conceptually, a quantifier denoted by an NP is characterised in QuantML by four components: (1) a quantification domain, (2) a distributivity, (3) a (proportional or absolute) involvement, and optionally (4) a domain size. For the quantification domain, a distinction is made between (a) the do-

[1] https://sigsem.uvt.nl/isa17/QuantificationChallenge.docx

main denoted by the nominal head (the 'source domain'), (b) the contextually determined smaller domain that the quantification is meant to apply to (the 'reference domain' or 'context set'), and (c) the set of individuals or quantities in the reference domain that are actually involved as participants in a set of events, called the 'participant set'.

For a definite plural NP, interpreted as a determinate quantifier with individual distributivity (and without domain size specification), this leads to a DRS of the form (2), where P_0 designates the contextual restriction of the source domain predicate P that restricts the quantification domain to the contextually determined reference domain.

(1) $[X|x \in X \leftrightarrow P_0(x)]$

A definite plural NP that includes a reference domain size specification N, as in *"the five judges"* (see sentence 12 of the Quantification Challenge test suite), where $N = \lambda z.|z| = 3$, corresponds to a DRS like:

(2) $[X|x \in X \leftrightarrow [P_0(x), |P_0| = 3]]$

The additional condition $|P_0| = 3$ expresses a presupposition about the size of the reference domain.

A definite singular NP can be treated in the same way (Scha, 1881), with $|P_0| = 1$, as quantifying over a reference domain that is a singleton, or alternatively as directly introducing a single individual, as is done in QuantML and leads to a DRS of the simpler form (3).

(3) $[x|P_0(x), |P_0| = 1]]$

The upshot of this is that the annotation of an NP in QuantML as an entity structure which, in the concrete syntax, has the definiteness value "det", is interpreted as having a reference domain which consists of one or more 'contextually distinguished' elements of the source domain. If the distributivity is "single", then the presupposition is that there exists exactly one entity that the quantification refers to; if the distributivity has another value, then there is a tacit presupposition that the reference domain is not empty, and if the NP includes a size specification, then that is interpreted as a specific numerical presupposition. Since, as noted, the relation between definiteness and determinacy is not straightforward, annotators should be aware of this interpretation of the definiteness and distributivity attributes, and this should guide their choices.

2.2 Nonstandard cases of distribution

QuantML uses besides the commonly used distributions "collective" and "individual" (a.k.a. "distributive"), the nonstandard cases "single" for (singular) proper names, (singular) definite descriptions, and singular deictic pronouns (*"I"*, French *"tu"*, etc.); the value "parts" for non-collective mass NPs as quantifiers; and "unspecific" for count NP quantifiers whose reference domain includes both individual entities and collections of entities.

Plural proper names, like *"the Marx Brothers"*, are in some contexts intended to quantify individually or collectively, as in sentence 3 of the Quantification Challenge test suite: *"Are the Marx Brothers famous?"* The Marx Brothers are clearly famous as a collectivity, and Chico, Groucho and Harpo are individually famous, but Gummo and Zeppo less so. The distribution "single" is appropriate on the collective reading (but "collective" too). Now consider the following example:

(4) The Marx Brothers performed in several movies.

This sentence is true both on the collective and on the individual reading. And the phrase *"the Marx Brothers"* can also be used not as a proper name for the group of five brothers but in order the describe the domain formed by the sons of their mother Minnie Marx, which included a sixth brother who died in infancy. In the latter case, the distribution "single" would obviously be inappropriate.

2.3 Special cases of event scope

The scope of a quantification over events nearly always has narrow scope relative to their quantified participants. (Singular) Proper names, (singular) definite descriptions and singular deictic pronouns do not quantify over a domain with multiple individuals, but refer to a single individual and hence do not really have a scope relative to the events in which the single individual participates. This is indicated in QuantML by using the event scope: "free" (and the inter-NP scope relation "unscoped"). In the QuantML semantics this has the effect that, when compositionally interpreting an annotation structure, the quasi-quantifications over single individuals can freely move around in constructing a nested DRS.

2.4 Deictic NPs

QuantML supports a simple treatment of NPs in the form of deictic personal pronouns, like *"I"* and *"you"* by assuming conceptual predicates $speaker_0$ and $addressee_0$ to be defined in the semantics, identifying a contextually distinguished speaker and one or more contextually distinguished addressees. Additional predicates of this kind could be introduced if a wider coverage of deictic expressions would be desirable (cf. Clarke, Montague).

2.5 Polarity

In the annotation of a sentence containing a negation, all the participation link structures must have their polarity specified as being either wide-negative or narrow-negative. Specifying the polarity of just one participation link (typically the Agent or Theme link) is not articulate enough in the case of more than one quantified participant set. Together with the relative scoping of participation links, this gives the necessary articulation.

Consider the following example:

(5) The editors didnt see a misprint.

This sentence allows three readings that differ only in scope of the negation; another plausible reading is the one where NP2 scopes over NP1:

a. every one of the editors didnt see a misprint, but not all the editors missed the same misprint (in other words: for each editor there was some misprint that this editor dit not see)
NP1 - NP2- NEG

b. each of the editors didn't see any misprint (in other words: for each editor it is not the case that (s)he saw a misprint
NP1 - NEG - NP2

c. it is not the case that each of the editors saw a misprint (in other words, not all the editors saw a misprint)
NEG - NP1 - NP2

d. a misprint was not seen by (any of) the editors.
NP2 - NP1 - NEG (or NP2 - NEG - NP1)

In terms of participation link polarity and quantifier scoping, these interpretations can be represented schematically, as follows:

a. NP1 $>$ NP2; NP1: neg-narrow, NP2: neg-narrow

b. NP1 $>$ NP2; NP1: neg-narrow, NP2: neg-wide

c. NP1 $>$ NP2; NP1: neg-wide, NP2: neg-wide

d. NP2 $>$ NP1; NP1: neg-narrow, NP2: neg-narrow (or NP1: neg-wide)

Annotation: see Appendix. Markables: m1=The editors, m2=editors, m3=see, m4=a misprint, m5=misprint.

In all the readings where the quantification over editors scopes over the one over misprints (i.e., all readings except readings d, where a specific misprint wasn't seen by all or any of the editors), the semantics of the annotation is obtained by the scoped merge of the interpretations of the two participation link structures:

(6) $I_Q(A) = I_Q(L_{P1}) \cup^* I_Q(L_{P2})$

The interpretations of these link structures are as follows:

(7) a. $I_Q(L_{P1}) = [X|x \in X \leftrightarrow \text{editor}_0(x), x \in X \rightarrow \neg[E \subseteq \text{see} \,|e \in E \rightarrow \text{agent}(e,x)]]$
b. $I_Q(L_{P2}) = \sim[Y|y \in Y \rightarrow [E \subseteq \text{see} \,|e \in E \rightarrow \text{theme}(e,y)]]$

In order to be able to represent clause level negation in a DRT-based framework, QuantML makes use of the top-level negation introduced by Krahmer & Muskens (1995), symbolised as $\sim$. This is illustrated in (7b).

This has the consequence that the various operators used in the QuantML semantics for combining DRSs have to be extended to the combination of (top-level) negated DRSs. For the standard DRS merge this can be done as follows.

(8) For any two DRSs K_1 and K_2:
$\sim K_1 \cup \sim K_2 =_D \sim (K_1 \cup K_2)$.

The same extension is needed in the definition of the combinators $\cup'$ ('unscoped merge'), $\cup^D$ ('dual merge') and $\cup^*$ ('scoped merge'). Except the scoped merge, all these combinators are defined only of either both or non of the arguments are top-level negated. The scoped merge combinator can apply also to a first argument that is not negated and the a second argument that is.

The scoped merge $\cup^*$ moves (and merges) a DRS K_2 to a position within the scope of another DRS K_1, as illustrated in (9).

(9) $[X|C_1, x{\in}X \rightarrow K_1] \cup^* [Y|C_2, y{\in}Y \rightarrow K_2] = [X|C_1, x{\in}X \rightarrow [Y|C_2, y{\in}Y \rightarrow (K_1 \cup K_2)]]$

The definition of this combinator has to be extended additionally to the case that the second argument is a top-level negated DRS, in which case two things happen: (1) as the argument moves into the first argument, its negation changes from top-level to DRS-internal, and thus from $\sim$ to $\neg$; (2) the nuclear information in the first argument (i.e. K_1) moves within the scope of the negation of the second argument, which is allowed (only) if it is itself negated. So the additional extension of the definition is as follows.

(10) $[X|C_1,x{\in}X{\rightarrow}\neg K_1] \cup^*$
$\sim[Y|C_2,y{\in}Y{\rightarrow}K_2] =$
$[X|C_1,x{\in}X{\rightarrow}\neg[Y|C_2,y{\in}Y{\rightarrow}(K_1 \cup K_2)]]$

If K_2 is a (top-level) negated DRS, then moving it inside K_1 has the effect that its negation is no longer at top level, hence the $\sim$ negation should be replaced by the ordinary DRS-negation $\neg$. This is defined in (12).

(12) For any two DRSs K_1 and K_2:
$K_1 \cup^* \sim K_2 =_D K_1 \cup^* \neg K_2.$

2.6 Complex numerical specifications

A possible QuantML treatment of complex numerical specifications like *"two or three times"* is shown by the annotation in (11). The @numericalPred element would then be allowed two specify two numericals as arguments of a numerical relation, besides the argument indicated by the element's identifier.

This presupposes the availability of ternary relations, which would also be needed for dealing with between twenty and twenty-five, or 10-12, or 7 a 8. These are also needed for numerical domain involvement specification.

The corresponding structure in the abstract syntax would be a 'numerical size specification $\langle m, \langle r, n_1, n_2\rangle\rangle$, with the semantics $\lambda z.r(n_1, n_2)$.

3 Limitations and plug-ins

As it is, QuantML is limited in a number of respects, partly due its intended future role as a part of the ISO SemAF series of annotation standards. In QuantML for example the semantic information related to verb tense is not taken into account; QuantML follows a recommendation from (Bos and Abzianidze, 2019) in this respect, but also because ISO-TimeML (ISO 24617-1, 2012) already takes such information into account in considerable detail. Adverbial temporal quantifiers (*"always", "sometimes", "never"*) are left out of consideration for the same reason, and similarly for adverbial spatial quantifiers. (Temporal quantification by means of an NP, like *"every hour"* is covered by QuantML, though - see sentence 20 of the Quantification Challenge.) Similarly, QuantML does not include a treatment of anaphoric quantifiers, since the ISO Reference Annotation Framework (ISO 24617-9, 20??) is devoted to the annotation of anaphoric relations.

It would of course be interesting to combine the annotation of quantification, time, space, events and anaphora into a single coherent scheme. A possible move in this direction is to employ 'annotation scheme' plug-ins, as proposed by (Bunt, 2019).

4 Annotation guidelines

The precise definition of QuantML, with the specification of the meanings of the elements of its abtract syntax and the attributes and values of its concrete syntax, could in theory be sufficient for an annotator for applying QuantML in the annotation of a given text, but in practice it is more helpful for annotators to have additional guidelines for how to apply the concepts of the annotation scheme. The documents that formally define QuantML currently (ISO WD 24617-12, Bunt 2021) contain highly incomplete guidelines, since a separate document is planned specifically for this purpose.

References

B. Abbott. 2004. Definiteness and indefiniteness. In *Handbook of Pragmatics*, pages 122–149. Blackwell, Oxford.

J. Bos and L. Abzianidze. 2019. Thirty musts for meaning banking. In *Proceedings First International Workshop on Designing Meaning Representations*, pages 15–27, Florence, Italy.

H. Bunt. 2019. Plug-ins for content annotation of dialogue acts. In *Proceedings 15th Joint ISO-ACL Workshop on Interoperable Semantic Annotation (ISA-15)*, pages 34–45, Gothenburg, Sweden.

H. Bunt. 2021. *Semantic Annotation of Quantification in Natural Language, second edition. TiCC Technical Report* 2021-2. Tilburg Center for Cognition and Communication, Tilburg University.

H. Clark. 1996. *Using Language*. Cambridge University Press, Cambridge, UK.

(11) <participation event="#e1" participant="#x1" semRole="agent" distr="individual" eventScope="narrow" rep="#n1"/
<numericalPred xml:id="n1" target="#m9" numRel="or" num1="2" num2="3"/>

Figure 1: Example of a QuantML complex numerical specification

E. Coppock and D. Beaver. 2015. Definiteness and Determinacy. *Linguistics and Philosophy*, 38:377–435.

I. Heim. 1982. *The semantics of definite and indefinite noun phrases*. Ph.D. Thesis, University of Massachussetts at Amherst.

ISO. 2012. *ISO 24617-1: Language resource management – Semantic annotation framework – Part 1: Time and events ('ISO-TimeML')*. International Organisation for Standardisation ISO, Geneva.

ISO. 2016. *ISO WD 24617-2: Language resource management – Semantic annotation framework – Part 12: Quantification ('QuantML', Working Draft)*. International Organisation for Standardisation ISO, Geneva.

ISO. 2019. *ISO 24617-9:2019, Language Resource Management: Semantic Annotation Framework (SemAF) - Part 9: Reference Annotation Framework*. International Standard. International Organisation for Standardisation ISO, Geneva.

E. Krahmer and R. Muskens. 1995. Negation and disjunction in Discourse Representation Theory. *Journal of Semantics*, 12(4):357–376.

D. Lewis. 1975. Adverbs of quantification. In *Formal Semantics of Natural Language*, pages 3–15. Cambridge University Press, Cambridge, UK.

R. Montague. 1970. Pragmatics and intensional logic. *Synthese*, 22:68–94.

S. Peters and D. Westerståhl. 2013. The Semantics of Possessives. *Language*, 89(4):713–759.

R. Scha. 1981. Distributive, Collective and Cumulative Quantification. In J. Groenendijk and M. Stokhof, editors, *Formal Methods in the Study of Language*. Mathematical Center, Amsterdam.

D. Westerståhl. 1985. Determiners and context sets. In Johan van Benthem and Alice ter Meulen, editors, *Generalized Quantifiers in Natural Language*, pages 45–71. Foris, Dordrecht.

F. Zwarts. 1994. Definite expressions. In *The Encyclopedia of Language and Linguistics*. Pergamom, Oxford.