The Annotators Did Not Agree on Some of the Guidelines Examples

Maxime Amblard, Karën Fort, Bruno Guillaume, Philippe de Groote, Chuyuan Li, Pierre Ludmann, Michel Musiol, Siyana Pavlova, Guy Perrier, Sylvain Pogodalla

The Annotators Did Not Agree on Some of the Guidelines Examples

Maxime Amblard[⋆], Karën Fort[†,⋆], Bruno Guillaume[⋆], Philippe de Groote[⋆], Chuyuan Li[⋆], Pierre Ludmann[⋆], Michel Musiol[‡,⋆], Siyana Pavlova[⋆], Guy Perrier[⋆], Sylvain Pogodalla[⋆]
[⋆] Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France
[†] Sorbonne Université / STIH - EA 4509, France
[‡] ATILF, UMR 7118, Université de Lorraine, CNRS, 54000 Nancy, France
`<first name>.<surname>@loria.fr`

Abstract

We propose the annotation of 7 sentences out of the 31 provided in the ISA-17 shared task, according to our understanding of the guidelines. We include here several remarks to improve the annotation and provide some tools to make the task easier.

1 Introduction

We annotated a fifth of the proposed sentences in the ISA-17 shared task. We tried and covered the interactions, within the QuantML annotation scheme, between quantification and various phenomena: negation, relative clauses, cardinal determiners, etc. Our annotations of the following sentences are provided as supplementary material:

(1) *Some of the students failed the exam.*
(5) *The editors didn't see a misprint.*
(7) *A man who walks in the park whistles.*
(8) *Mary visits a museum every day.*
(11) *Not all the students passed the exam.*
(12) *Most of the students passed the exam.*
(31) *More than four hundred ships are waiting to pass through the Suez Canal.*

In Section 2, we describe the methodology of our participation to the ISA-17 shared task; in Section 3, we comment on the task per se. After these practicalities, Sections 4 and 5 provide feedback on the QuantML annotation scheme and its guidelines. Section 6 draws a parallel between QuantML and the Parallel Meaning Bank (PMB) scheme. In the final Section, we detail theoretical suggestions as well as the tools we deployed for the shared task.

2 Methodology

A Collaborative Effort We organized eight different collective meetings, of approximately two hours each, in addition to individual work. The results we provide were therefore obtained through collaborative discussions. It has to be noted that, although most of us have a very good command of the language, there is no native speaker of English in the team.

We based our annotations on the documentation resources directly mentioned in the Call for Papers of the ISA-17 Shared Task: (1) the provided working document for the ISO at hand (ISO WD 24617-12:2019(E)); (2) Semantic Annotation of Quantification in Natural Language (Bunt, 2020); (3) A Semantic Annotation Scheme for Quantification (Bunt, 2019); and, (4) Towards an ISO Standard for the Annotation of Quantification (Bunt et al., 2018). Whenever discrepancies showed up in the documentation, we tried and used the resources in that preference order.

We noted an important learning curve effect, as we spent quite a long time on our second annotation (Sentence 5), as compared to the following ones.

Tooling Up In order to be as consistent as possible in our annotation and easily share them in the group, we used several tools (see Section 7): (1) XML processing tools (well-formedness and validation); (2) visualization tools to provide a graphical structure of the annotated elements; (3) a framework dedicated to automatic building of semantic representations. Applying this set of tools to a subset of 10 annotation examples provided in the guidelines also helped identify inconsistencies.

3 Remarks on the Shared Task

Some issues in the shared task organization hindered our participation, without fully concerning the annotation scheme nor its guidelines.

The first issue is the lack of a unique and explicit reference document (guidelines), as is the case for other resources such as the Penn Treebank (Santorini, 1990). The call for papers pointed to some relevant ones, but their selection and ranking is left to the participants, even though the standard is not published yet and there are discrepancies in the pointed documents. Section 5 develops this issue further.

Markables The proposed markables biased our annotation. For instance, in Sentence 7, we needed to mark *in the park*, but we were reluctant to use a markable that is not provided. Yet, Example (66) in the ISO draft features such a markable.

Since deciding on the markables is the very first annotation step, it determines the finesse of the final result: a more careful examination and explanation is necessary for the scheme and its guidelines. One suggestion is to use markable variables including offsets, as `m<n>-<p>` for a markable spanning from the $(n+1)$-th token to the p-th, so that annotators are not biased by the pre-identified markables.

Unavailability of a Resource We had a doubt when annotating the semantic roles (namely, a `@semRole` in Sentence 31), as the proposed semantic roles are defined in ISO 24617-4, a document which is not freely available. We thus referred to VerbNet for a suitable role. But we are not sure if it is presented in the ISO list.

4 Adequacy of the Scheme

The QuantML annotation scheme tackles important and intricate phenomena about quantification. It relies on a thorough and helpful introduction and state of the art, allowing non-specialists of the domain to better understand the logic underlying the annotation (Fort, 2016) and therefore to evolve from a teacher-like to a peer relation (Akrich and Boullier, 1991). Some phenomena we encountered during the annotation phase could, however, be more developed. For instance, more space could be devoted to the interaction between quantification and negation, in particular when the polarity induced by a negation on an event differs between participants in that event.

Tricky Events We also encountered difficulties with the `@clause` attribute possible values of the **`relClause`** element (see for instance Sentence 7). In particular, if **`event`** elements are the expected concrete values, the guidelines should be explicit about it. Besides, we wondered whether a **`participation`** element could be sufficient, as both elements share most attributes.

Having events as attribute values also occur for instance in Sentence 31, with the attitude verb *to wait*, where the *theme* or *goal* participant should be annotated with the clause or the event the *agent* is waiting for. It is not clear whether attributing an event is possible, as events and participants seem clearly disjoint in the QuantML metamodel.

Genericity Finally, QuantML declares genericity as a loose end, which is unfortunate since they induce major variations in quantification. The scheme could also account for generics (with dedicated involvement values, like `"generic a"`, or an attribute `@genericity`), even though it could not interpret them to their full extent. Doing so would boost the understanding of generics and their semantics. Otherwise, users will have to retrace all their annotations when if the scheme introduces genericity at a later stage. Worse, users may tinker and tamper with their annotations, e.g., with `@involvement`, but in non-standard ways then.

5 About the Annotation Guidelines

We will use the term *guidelines* here to refer to the four documents listed in Section 2. As a matter of fact, there is an appendix holding that title of guidelines in both the technical report and the ISO working document, but it was insufficient to annotate the provided sentences.

The biggest issue with this, apart from the difficulty of searching into several documents, is the inconsistency between and among them.

Scattered Examples In terms of the assignments of different attributes, explanations and examples are so scattered that we spent a lot of time comparing the sentences and the semantic differences. A comprehensive list of all the possible values for each attribute would help to ease the work. A precise definition of the different attributes should also be provided. Some of them are only glossed (e.g., the `@rep` attribute for events, while the `@repetition` attribute for participation links are described in different places), some of them are only defined in the QuantML semantics (e.g., the `@semRole` attribute of the **`relClause`** element), some of them are never mentioned but in the list of attributes of a given element (e.g., the `@clause` attribute of the **`relClause`** element).

Terminology Issues The terminology used in the annotation scheme did not appear very intuitive to us. For example, the value **`unspecific`** for the `@distr` attribute does not tell us much about what it is supposed to mean. Our understanding of the guidelines leads us to believe that this should be interpreted as what we consider as a *cumulative* distribution and propose that the attribute value be changed accordingly. Also, the guidelines raise a distinction between *definiteness* and *determinacy*, but mix them up in both the abstract and the concrete syntax. Since the only values are **det** and **indet**, a boolean attribute `@determinacy` seems a better choice of terminology.

6 A Parallel with a Meaning Bank

With the interpretation to DRT, the QuantML annotation scheme allows developing resources coupling raw text and corresponding DRS. The Parallel Meaning Bank (PMB, Abzianidze et al., 2017) is such a resource, yet built on an automated annotation pipeline. In a nutshell, the pipeline builds syntactic analysis in a Combinatory Categorial Grammar and associates λ-DRS to tokens (Bos, 2009). This yields DRS from a compositional syntax-semantics interface.

However, syntactically similar sentences lead to similar relative scopes of quantifiers, which can be erroneous as Examples (71/1608) and (56/3237) from the PMB show:

(71/1608) She has a bottle of milk every morning

(56/3237) I am keeping a diary every day

They both are annotated with the universal quantification (over *morning* or *day*) outscoping the existential one (over *bottle* or *diary*). It should, however, be the opposite in Example (71/1608), i.e., the diary is the same every day.

On the other hand, QuantML is not bound to a specific syntactic framework, and formulates the semantic difference between these two examples with the dedicated structure **scoping** (see Sentence 31).

There also are divergences between QuantML and PMB. For instance, looking at prepositional modifications of events, specifically locations with *in* and *under* for instance, the PMB features spatial operators to discriminate them (00/2412 and 06/1633) whereas QuantML mixes them up.

Apart from the differences mentioned above, QuantML and PMB also differ on other aspects, for instance on dealing with determinacy and predicate distribution over plurals. In the QuantML scheme, determinacy is interpreted as reference domains over source domains while it triggers presuppositions through pointers in the PMB (Venhuizen et al., 2013). Furthermore, tense is annotated in PMB whereas QuantML does not offer this option. Finally, the PMB works at the moment with the simplifying assumption that any predicate is distributive.

7 Going Further

Theoretical Suggestions In the annotations we provide, we propose two different solutions for Sentence 5 where the difference is at the level of the scope of the negation. We could not find in the guidelines how this kind of ambiguity should be handled.

Similarly, we did not find whether specifying forbidden scoping and reading is possible and/or desirable. Finally, the annotation process led us into interesting theoretical questions about the effect of collective/distributive readings under the scope of a negation (see the `@distr` attribute in both annotations we provided for Sentence 5).

Providing Tools In addition to the sentences we annotated, we provide a Document Type Definition (DTD) to check the produced XML, and an implementation of the scheme's interpretation. This helped us identify inconsistencies among the XML attributes, e.g., `@individuation` or `@indiv` (ISO WD 24617-12:2019(E)). The DTD regulates the XML format as it requires a single and coherent way to represent these attributes. Further, it regulates values for closed classes: e.g., it would reject **def** and **indef**, as typos for **det** and **indet** respectively ((C62) and (C21) in ISO WD 24617-12:2019(E)).

We have noted that the same attribute `@involvement` is currently used both with a reference to an XML identifier or with a string value (both are used in Sentence (B59) of the guidelines). This prevents from a precise definition of `@involvement` the DTD.

In order to investigate the semantics, we also used the framework of Abstract Categorial Grammars (ACG) to define the compositional syntax-semantic interface I_Q, and implemented it in ACGtk. We use Continuation-Based Dynamic Logic instead of Discourse Representation Theory, for it features no free variables. At the moment, our implementation ignores individuation, and treats all merges as dynamic conjunctions.

While unrefined, this implementation allows us to investigate the semantic interpretations of annotation choices. For instance, the

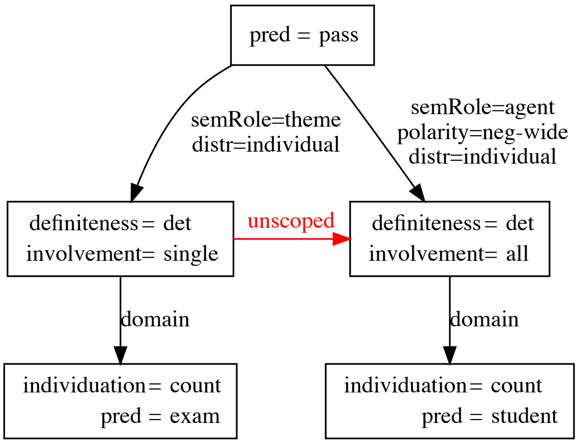

Figure 1: Graph representation of Sentence 11 (*Not all the students passed the exam*).

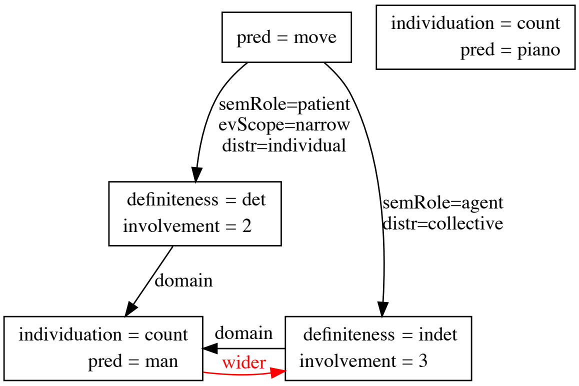

Figure 2: Wrongly annotated sentence A6 (*Three men moved both pianos*) in the TiCC report.

guidelines lead us to annotate *some of the students* in Sentence 1 with <**entity** ... involvement="some" definiteness ="det"/> since it features a definite plural. Yet, it yields an unexpected logical formula stating *all* the students in the reference domain failed their exam. A similar situation arises for Sentence 12.

We also used a graph-based concrete syntax representation annotation with an automatic conversion from the XML format into a graph (see Figure 1). By converting examples[1] from the TiCC report (Bunt, 2020), we noticed some annotation errors (for instance, **participation** links are missing in Examples (A5) and (A7) and some errors in attributes in Example (A6), see Figure 2).

[1]The first 11 examples, provided in the file `annotation.pdf` in the supplementary material

8 Conclusion and Perspectives

Our participation in this shared task allowed us to spend time working on difficult issues in a collaborative way. This was also an opportunity to deploy a range of tools that assisted our understanding of the QuantML and our participation to the ISA-17 shared task.

References

Lasha Abzianidze, Johannes Bjerva, Kilian Evang, Hessel Haagsma, Rik van Noord, Pierre Ludmann, Duc-Duy Nguyen, and Johan Bos. 2017. The Parallel Meaning Bank: Towards a Multilingual Corpus of Translations Annotated with Compositional Meaning Representations. In *15th Conference of the European Chapter of the Association for Computational Linguistics, Volume 2, Short Papers*, pages 242 – 247, Valencia, Spain.

Madeleine Akrich and Dominique Boullier. 1991. Le mode d'emploi, genèse, forme et usage. In Denis Chevallier, editor, *Savoir faire et pouvoir transmettre*, pages 113–131. éd. de la MSH (collection Ethnologie de la France, Cahier 6).

Johan Bos. 2009. Towards a large-scale formal semantic lexicon for text processing. In *From Form to Meaning: Processing Texts Automatically. Proceedings of the Biennal GSCL Conference*, pages 3–14.

Harry Bunt. 2019. A semantic annotation scheme for quantification. In *Proceedings of the 13th International Conference on Computational Semantics - Long Papers*, pages 31–42, Gothenburg, Sweden. Association for Computational Linguistics.

Harry Bunt. 2020. Semantic annotation of quantification in natural language. Technical report, TiCC Technical Report 2020-2. Tilburg Center for Cognition and Communication and Department of Cognitive Science and Artificial Intelligence, Tilburg University.

Harry Bunt, James Pustejovsky, and Kiyong Lee. 2018. Towards an ISO standard for the annotation of quantification. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Karën Fort. 2016. *Collaborative Annotation for Reliable Natural Language Processing*. Focus series. ISTE Wiley.

ISO WD 24617-12:2019(E). 2021. Language management — Semantic annotation framework (SemAF) — Part 12: Quantification. Working document, International Organization for Standardization, Geneva.

Beatrice Santorini. 1990. Part-of-speech tagging guidelines for the Penn Treebank Project. Technical Report MS-CIS-90-47, Department of Computer and Information Science, University of Pennsylvania.

Noortje Venhuizen, Johan Bos, and Harm Brouwer. 2013. Parsimonious semantic representations with projection pointers. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)–Long Papers*, pages 252–263.