Appendix C

Annotation Guidelines for Kovatchev et al. [2020]

C.1 Presentation

This document sets out the guidelines for the annotation of atomic types using the Extended Typology for Relations. The task consists of annotating pairs of text that hold a textual semantic relation (paraphrasing, entailment, contradiction, similarity) with a textual label, and the atomic phenomena they contain. These guidelines have been used to annotate the ETRC corpus. For the purpose of the annotation, the WARP-Text annotation tool has been used.

N.B.: The task definition, tagset definition and annotation of linguistic phenomena in these Guidelines overlap with those for the ETPC corpus. The reader is encouraged to consult the ETPC guidelines presented in Appendix A or the full SHARel guidelines available online. Here I only provide the guidelines for the reason-based types.

C.2 Annotating reason-based Phenomena

Reason-based phenomena account for relations that cannot be expressed and processed using only linguistic knowledge. Like the linguistic phenomena, the reason-based phenomena can be sense-preserving or non-sense preserving. Our goals with the annotation of reason-based phenomena are twofold:

1) we want to make a precise and explicit annotation of the units involved in the inference

2) we want to determine the kind of reason-based and background knowledge required.

1a and 1b show an example of an "existential" reason-based – "speaking X" entails "X exists". 2a and 2b show an example of "causal" reason-based – "X is written in Y (language)" entails "reading X requires Y (language)" .

1a Speaking more than one language is imperative today.

1b There is more than one language.

2a Reading the Bible requires studying Latin.

2b The Bible is written in Latin.

When annotating reason-based phenomena, there are several important things:

- Annotate all possible phenomena separately. The aim is to annotate every token that is not already annotated as linguistic or addition-deletion.
- The scope of some phenomena can overlap. That means some tokens may be part of multiple scopes.
- When choosing the scope, we choose the smallest scope possible. Unlike the sense-preserving, in this part of the annotation, the goal is to choose the most specific scope possible. For example, in 1a and 1b we could annotate "Speaking more than one language" and "There is more than one language", but in order to be as specific as possible, we choose to only annotate "Speaking" and "There is".
- When choosing the scope, if possible, try to annotate whole linguistic units without breaking them. For example in 2a and 2b, we could only annotate "Reading requires" and "is written in"
- Like in the linguistic phenomena – the sense preserving reason-based phenomena need not relate units that have similar syntactic or semantic role; however, the non-sense preserving reason-based phenomena must relate units that have similar syntactic or semantic role.

C.3 List of reason-based phenomena

1. Cause and Effect: T causes H to be true [neg. sense preserving: T causes H to be FALSE]

a. "Once a person is welcomed into an organization, they belong to that organization"

2. Conditions and Properties: A very general type where H containing facts (and properties) implied by T [neg. sense preserving: H contains facts and properties that contradict the implied from T (ex.: "There is only one language")]

a. Existential – T entails H exists (pre-requirement)

b. "To become a naturalized citizen, one must not have been born there" (pre-requirement)

c. "The type of thing that adopts children is person" (argument type)

d. "When a person is an employee, that organization pays his salary" (simultaneous conditions)

3. Functionality: Relationships which are functional [neg. sense preserving: mutual exclusivity – types of things that do not participate in the same relationship]

a. "A person can only have one father (or two arms)" (+)

b. "Government and media sectors usually do not employ the same person" (-)

4. Transitivity: If R is transitive and R(a,b) and R(b,c) are true, then R(a,c)

a. "The "support" is transitive. If Putin supports United Russia party, and United Russia party supports Medvedev, then Putin supports Medvedev"

5. Numerical Reasoning

6. Named Entity Reasoning: reasoning that goes beyond substitution; relations between multiple entities (i.e. not just Trump – president, but rather Trump – Clinton)

7. Temporal and Spatial Reasoning

8. Other (World Knowledge)