

Checking annotation data in the FrameNet database

Russell Lee-Goldman

June 8, 2006

The FrameNet Annotation Database contains all of the sentences collected and annotated for target lexical units (LUs). Sentences marked with annotation (annotation sets) contain labels for a target LU, the FEs it occurs with (including null-instantiated FEs), and additional labels such as Cop and Supp. FrameNet has developed several policies for determining the accuracy and style of sentence annotation, and these policies have been implemented in the form of data-checking algorithms.

Due to the size of the annotation data, the most efficient method we have found for checking and managing it has been through programs, largely written in Perl or Java, which algorithmically determine the quality of each annotation set. Because the action that should be taken to repair annotation is usually not clear in advance, these programs simply output their results in a human-readable format so that manual correction can be applied. However, in some rare cases an automatic fix (in the form of native database commands) has been possible.

The annotation data is checked, broadly, in three areas: data integrity, data completeness/consistency, and style. In the remainder of this document, for each of these areas, we give a brief explanation, list the checking programs, and show a summary table of the outputs of recent runs. In many cases, the output list some cases in which the annotation is “acceptable”, that is, it is correct according to the annotation principles, but the rules implemented by the checking program are not able to recognize this, and implementing rules to handle all the cases correctly would be impractical. Therefore, the summary table contains estimates of the number of cases that are marked as questionable by the checker, but are in fact “acceptable”. For a more detailed discussion of how this sort of data checking fits in with the larger picture of quality assurance in FrameNet, see (Scheffczyk & Ellsworth 2006).

1 Integrity

Due mostly to technical errors involved in changes in data infrastructure, parts of the annotation data can lose integrity. For instance POS tags may be lost, or Phrase Type tags may be deleted, leaving FE and GF labels behind. In the past such errors were not uncommon, but current practice and in-house software prevents their occurrence. Currently the following integrity checks are done on the annotation data:

- FE/GF/PT triples. With the exception of labels on target LUs or on nonprimary FE layers, all instances of FE, GF, or PT labels should coincide and be coterminous with labels of the other two types. Potential exceptions are FE labels on target words and FE labels on non-primary layers of annotation.

- Missing/Zero-length labels. FE, GF, or PT labels with a length of 0 or 1 (i.e., that span 0 or 1 character only) are reported. Exceptions include legitimate one-letter proper names, or *a* as a numeral as in *a cup of flour*.

The following table lists these checkers and their most recent results.

name	most recent results
FE/GF/PT triples	Unknown but small number of errors, fixable.
Missing/Zero-length labels	06/01/13: 31 annotation sets, all acceptable

Several data-checking programs that automatically fix data integrity errors have been implemented. These correct errors such as anomalous duplications of labels (e.g., two identical FE labels that covered the exact same span of text) or lexemes (e.g., two identical lemmas with the exact same lemmas and POS associated with them). Such errors should no longer occur.

2 Consistency/Completeness

The second sort of error that is checked for relates to data consistency and completeness. This is the largest class, including errors that clash with theoretical or conceptual ideas behind FrameNet and Frame Semantics. Most of the checkers (with the exception of the Core FE checker) are not completely rigid: the constraints they encode are meant to find potential errors, but some systematic exceptions may exist.

- Missing External Argument. Annotation sets that contain a verb or preposition target, are checked to ensure they have some linguistic material marked as External Argument (GF: Ext). Several classes of exceptions exist, including sentences with expletive subjects, with a Constructionally Null-instantiated (CNI) FE (usually Agent), or sentences where the verb acts as a prenominal (adjectival) modifier.
- FE Coreness. According to the definition of core FEs, each annotation set must contain at least one instance of each core FE, or at least one FE in each core set. Further, if one FE requires another, both must occur, and if one FE excludes another, then both may not occur together. There are no exceptions to these constraints.
- Duplicate Ext/Obj. Annotation sets with more than one Ext and/or one Obj label are reported. Classes of exceptions include doubly-instantiated FEs (such as non-adjacent modifiers of nouns (as in **a dispute AROSE between the buyer and the seller**) and floated quantifiers.
- Multiple FE instantiations. Annotation sets containing multiple instantiations of a given FE are reported. In general, a given FE should not be overtly instantiated (i.e., associated with linguistic material) on one layer and also null-instantiated on another, or overtly instantiated on multiple layers. Non-null instantiation on multiple layers is possible, and common, in constructions with Agents and Bodypart-possessors in sentences like ***he** broke **his** leg*.
- Improper 2nd layer annotation. Instances of (non-NI) 2nd-layer annotation that spans over material with no primary layer annotation is reported. Also checks for NI annotation of a FE without primary-layer instantiation of the same FE.

- Layer of Extrathematic FEs. Instances of extrathematic FEs occurring on a non-primary layer of annotation are reported. Some extrathematic FEs like Cause, Purpose, Explanation and Reason, may commonly occur on the second layer, with Time, Place, Depictive, et al., on the primary layer.
- Gov/Ctrlr/X marking. Governor and Controller labels should co-occur with the X label, and all X labels should include all Targets, and all FEs should be contained by an X label. Violating labels are reported.
- Ext/Obj and non-core FEs. In most cases the Grammatical Functions Ext and Obj should only occur with core Frame Elements. Instances of Ext or Obj on non-core FEs are reported. Potential exceptions include constructionally-introduced objects, as in the double-object or benefactive constructions.

name	most recent results
Missing External argument	05/12/12: 214 annotation sets, nearly all acceptable.
FE Coreness	06/01/19: under 78 instances of missing or illicit core FEs; will be fixed in future releases.
Duplicate Ext/Obj	06/06/15: 137 annotation sets; nearly all acceptable.
Multiple FE instantiations	06/01/12: about 700 sets of duplicate FEs, nearly all acceptable.
Improper 2nd later annotation	06/06/15: 6 instances, all acceptable.
Layer of Extrathematic FEs	06/06/15: 234 instances, nearly all acceptable.
Gov/Ctrlr/X marking	05/11/28: 59 instances, manual evaluation necessary.
Ext/Obj and noncore FEs	06/06/15: 128 labels, mostly acceptable. Some manual evaluation necessary.

3 Style

The final set of checkers deal mostly with style. We have determined that each core FE in a frame should be (non-null-)instantiated at least once for each LU, and at each peripheral FE in a frame should be instantiated at in at least one annotation set in the frame. Finally, LUs with the Finished_initial, Finished_checked, or FN1_sent status should have at least 5 annotated sentences. There are no “exceptions” to these guidelines, though meeting them is not as stringent as it is for the above-described constraints.

- Expression of Core FEs per LU. Each core FE should be instantiated once per LU.
- Expression of Peripheral FEs per Frame. Each peripheral FE should be instantiated once per Frame.
- Annotation set count. Finished_initial, Finished_checked, and FN1_Sent LUs should have at least 5 annotation sets.

name	most recent results
Expression of Core FEs per LU	05/11/16: 1343 instances of an unexpressed core FE. Many of these will likely be fixed by correcting the coreness status of several LUs, or by adding annotation so that the FEs are instantiated properly.
Expression of Peripheral FEs per Frame	06/01/13: 1925 instances of an unexpressed peripheral FE.
Annotation set count	06/01/31: 66 LUs. All of these are due to movement of annotation to new LUs, and other temporary statuses. Eventually these LUs will be deleted. Other LUs will be moved to appropriate statuses (such as Insufficient attestations).

References

- SCHEFFCZYK, JAN, & MICHAEL ELLSWORTH. 2006. Improving the quality of framenet. In *Proceedings of the Workshop on Quality assurance and quality measurement for language and speech resources*, ed. by Steven Krauwer & Uwe Quasthoff, 8–13, Genoa, Italy. LREC.