Supplementary Material

Yi Luan

1 Annotation Guideline

1.1 Entity Category

- **Task**: Applications, problems to solve, systems to construct.

  E.g. information extraction, machine reading system, image segmentation, etc.

- **Method**: Methods , models, systems to use, or tools, components of a system, frameworks.

  E.g. language model, CORENLP, POS parser, kernel method, etc.

- **Evaluation Metric**: Metrics, measures, or entities that can express quality of a system/method.

  E.g. F1, BLEU, Precision, Recall, ROC curve, mean reciprocal rank, mean-squared error, robustness, time complexity, etc.

- **Material**: Data, datasets, resources, Corpus, Knowledge base.

  E.g. image data, speech data, stereo images, bilingual dictionary, paraphrased questions, CoNLL, Panntreebank, WordNet, Wikipedia, etc.

- **Other Scientific Terms**: Phrases that are a scientific terms but do not fall into any of the above classes

  E.g. physical or geometric constraints, qualitative prior knowledge, discourse structure, syntactic rule, discourse structure, tree, node, tree kernel, features, noise, criteria

- **Generic**: General terms or pronouns that may refer to a entity but are not themselves informative, often used as connection words.

  E.g model, approach, prior knowledge, them, it...

1.2 Relation Category

Relation link can not go beyond sentence boundary. We define 4 asymmetric relation types (*Used-for*, *Feature-of*, *Hyponym-of*, *Part-of*), together with 2 symmetric relation types (*Compare*, *Conjunction*). B always points to A for asymmetric relations

- **Used-for**: B is used for A, B models A, A is trained on B, B exploits A, A is based on B. E.g.

  The TISPER system has been designed to enable many text applications.

  Our method models user proficiency.

  Our algorithms exploits local soothness.

- **Feature-of**: B belongs to A, B is a feature of A, B is under A domain. E.g.

  prior knowledge of the model

  genre-specific regularities of discourse structure

English text in science domain

- **Hyponym-of**: B is a hyponym of A, B is a type of A. E.g.

  TUIT is a software library

  NLP applications such as machine translation and language generation

- **Part-of**: B is a part of A... E.g.

  The system includes two models: speech recognition and natural language understanding

  We incorporate NLU module to the system.

- **Compare**: Symmetric relation (use blue to denote entity). Opposite of conjunction, compare two models/methods, or listing two opposing entities. E.g.

  Unlike the quantitative prior, the qualitative prior is often ignored...

  We compare our system with previous sequential tagging systems...

- **Conjunction**: Symmetric relation (use blue to denote entity). Function as similar role or use/incorporate with. E.g.

  obtained from human expert or knowledge base

  NLP applications such as machine translation and language generation

1.3 Coreference

Two Entities that points to the same concept.

- **Anaphora and Cataphora**:

  We introduce a machine reading system... The system...

  The prior knowledge include...Such knowledge can be applied to...

- **Coreferring noun phrase**:

  We develop a part-of-speech tagging system...The POS tagger...

1.4 Notes

1. Entity boundary annotation follows the ACL RD-TEC Annotation Guideline [**?**], with the extention that spans can be embedded in longer spans, only if the shorter span is involved in a relation.
2. Do not include determinators (such as the, a), or adjective pronouns (such as this,its, these, such) to the span. If generic phrases are not involved in a relation, do not tag them.
3. Do not tag relation if one entity is:
   - Variable bound:
     We introduce a neural based approach.. *Its* benefit is...
   - The word *which*:
     We introduce a neural based approach, *which* is a...
4. Do not tag coreference if the entity is

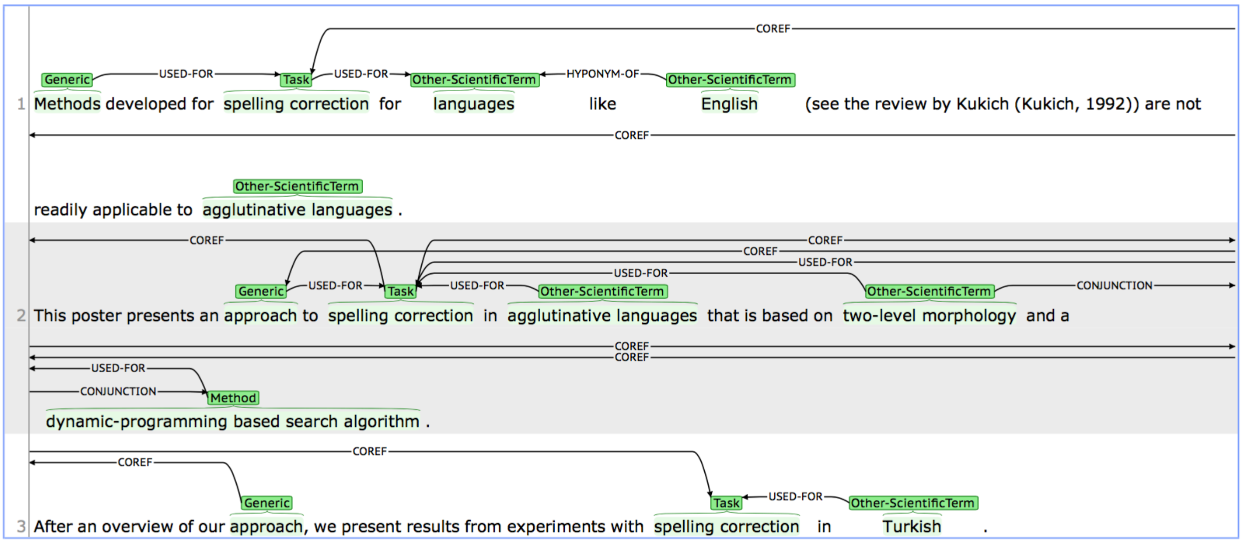

Figure 1: Annotation example 1 from ACL

- Generically-used Other-ScientificTerm:
  ...advantage gained from *local smoothness* which... We present algorithms exploiting *local smoothness* in more aggressive ways...
- Same scientific term but refer to different examples:
  We use a *data structure*, we also use another *data structure*...

5. Do not label negative relations:
   X is not used in Y or X is hard to be applied in Y

2 Annotation and Knowledge Graph Examples

Here we take a screen shot of the BRAT interface for an ACL paper in Figure 1. We also attach the original figure of Figure 3 in Figure 2. More examples can be found in the project website[1].

[1] http://ssli.ee.washington.edu/tial/projects/sciIE/

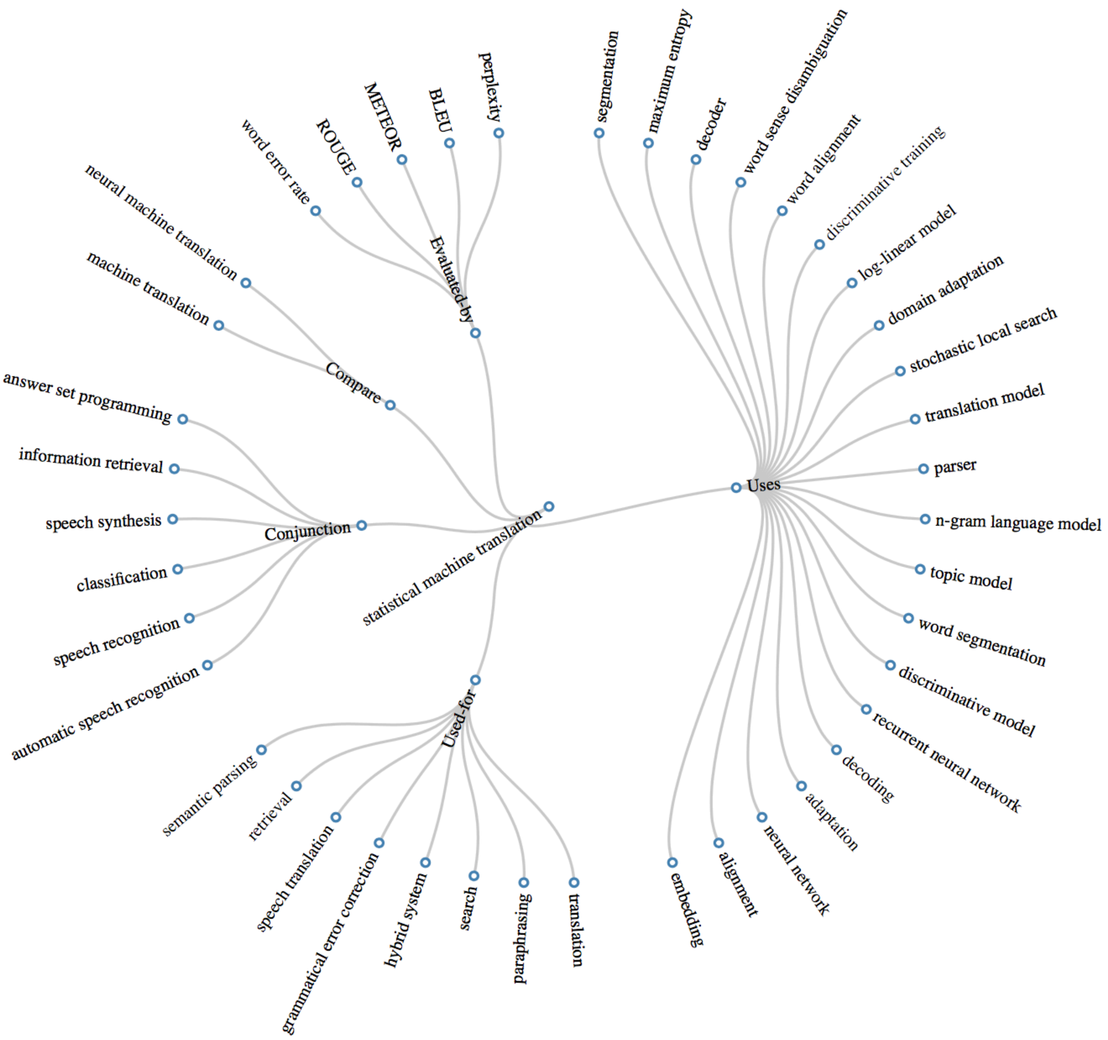

Figure 2: An example of our automatically generated knowledge graph centered on *statistical machine translation*. This is the original figure of Figure **??**.