

Data Science and Human Data: Annotation

Bonnie Webber

School of Informatics
University of Edinburgh
bonnie@inf.ed.ac.uk

November 24, 2014

Human annotation is valuable

Annotation involves marking up data:

- Identifying regions of interest in images or segments with particular properties in text;
- labelling those regions or segments.

In order to understand

- what makes those regions interesting or gives those segments their properties,
- what the labels follow from or correlate with,

human annotation is better than human intuition in helping to formulate a theoretically sound explanation or simply a reasonably accurate empirical model.

Human annotation is problematic

- People are often **inconsistent**: They say one thing at time t_1 and another at time t_2 [Klebanov & Beigman, 2009, 2010].
- People are often **biased**: They have preferences in how they answer questions and/or annotate data [Passonneau & Carpenter, 2014].
- Either can be the source of *inter-annotator disagreement*, but inter-annotator agreement (IAA) can mask both *inconsistency* and *bias*.
- It's neither efficient or sufficient to simply gather a huge amount of annotation of the same data: Crowd-sourcing alone is not the answer to either inconsistency or bias [Carpenter & Passonneau, 2014; Klebanov & Beigman, 2009, 2010].

- During annotation, one can try to show annotators their earlier annotation of similar tokens.
- After annotation is complete, one can try to assess whether similar tokens have similar annotation.

But both require the ability to identify *similar tokens*: For **discourse annotation**, this can be an interesting problem in its own right.

MSc project 1

Consistency Project – Annotation Tool

The screenshot displays the Annotator tool interface. The main window shows a document titled "wsj_0584" with the following text:

.START

The bolstered cellular agreement between BellSouth Corp. and LIN Broadcasting Corp. carries heightened risks and could fail to fend off McCaw Cellular Communications Inc., the rival suitor for LIN.

Moreover, the amended pact shows how McCaw's persistence has pushed LIN and BellSouth into a corner, forcing huge debt on the proposed new company.

The debt, estimated at \$4.7 billion, could mortgage the cellular company's future earning power in order to placate some LIN holders in the short term.

The plan still calls for LIN to combine its cellular telephone properties with BellSouth's and to spin off its broadcasting operations.

But under new terms of the agreement, announced Friday, LIN holders would receive a special cash dividend of \$42 a share, representing a payout of about \$2.23 billion, shortly before the proposed merger.

LIN said it expects to borrow the money to pay the dividend, but commitments from banks still haven't been obtained.

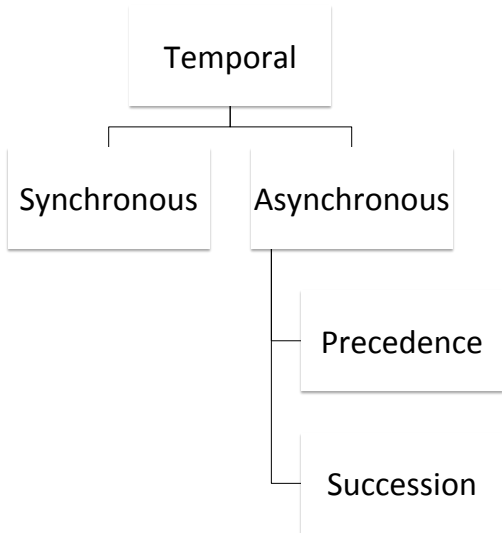
Under previous terms, holders would have received a dividend of only \$20 a share.

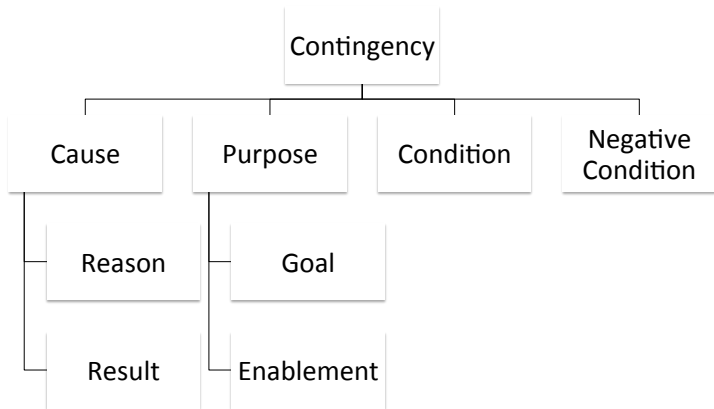
In addition, New York-based LIN would exercise its right to buy out for \$1.9 billion the 55% equity

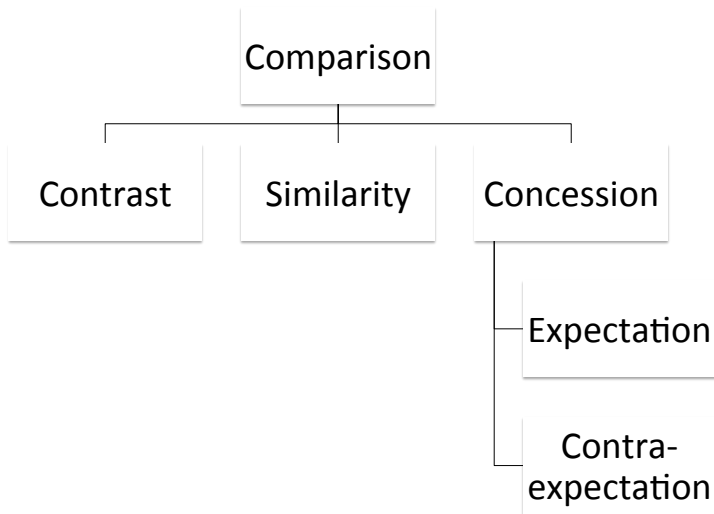
The right-hand panel shows the configuration for a relation:

- Relation Type: Implicit
- Conn1: [Dropdown]
- Conn2: [Dropdown]
- SClass1A: [Dropdown]
- SClass2A: [Dropdown]
- SClass1B: [Dropdown]
- SClass2B: [Dropdown]
- +Conn Span
- +Conn Feat Span
- Conn Src: Wr
- Conn Type: Comm
- Conn Pol: Null
- Conn Det: Null
- +Arg1 Span
- +Arg1 Feat Span
- Arg1 Src: Inh
- Arg1 Type: Null
- Arg1 Pol: Null
- Arg1 Det: Null
- +Arg2 Feat Span
- Arg2 Src: Inh
- Arg2 Type: Comm
- Arg2 Pol: Null
- Arg2 Det: Null
- +Sup1 Span
- +Sup2 Span

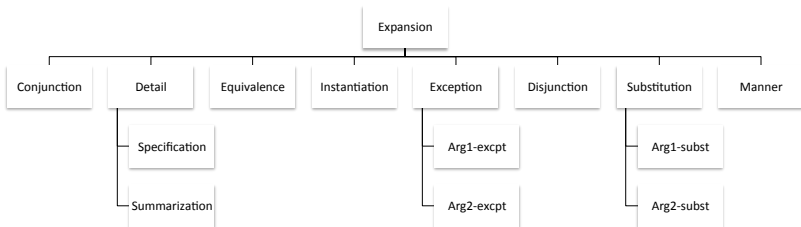
Buttons at the bottom of the relation configuration panel include: New Relation, Save Relation, Cancel Changes, and Delete Relation.







Consistency Project – Sense Annotation



- Standard practice relies on inter-annotator agreement (IAA) to recognize biases that lead to annotators to assign different labels to the same token.
- Standard practice then relies on reconciliation to either reach agreement or make an executive decision.
- Neither ensures a high-quality corpus.
- Probabilistic models of agreement are more promising.

We'd like to experiment with this on **discourse annotation** that we are about to crowdsource.

MSc project 2

Crowdsourcing Experiment

Experiment

University of Edinburgh

Trial

Instructions The grey box below contains a paragraph with an underlined blank space.
This is for you to fill in.
From the list of words below the box, please select the one you think fits best in this space. If you believe that no word should be inserted, choose 'None'.
If you feel that only some other word would be right, select 'Other!.'

Word Selection

"He loves sports and all the guy stuff," Ms. Oaklander **said.** " _____ **on**
the other hand, he loves to cook and he loves design."

* Conjunction:

- None
- Because
- Before
- So
- But
- And
- Or
- Other

- Beata Beigman Klebanov and Eyal Beigman (2010) Some Empirical Evidence for Annotation Noise in a Benchmarked Dataset *Proc. Annual Conference of the North American Chapter of the Association for Computational Linguistics* (NAACL), pp. 438–446.
- Beata Beigman Klebanov and Eyal Beigman (2010) From Annotator Agreement to Noise Models *Computational Linguistics*, 35(4), pp. 495–503.
- Rebecca J. Passonneau and Bob Carpenter (2014). The Benefits of a Model of Annotation *Transactions of the Association for Computational Linguistics* (TACL), 2, pp. 311–326.