

# Annotating Attribution Relations

## Towards source sensitivity within text

**Silvia Pareti**

S.Pareti@sms.ed.ac.uk

http://homepages.inf.ed.ac.uk/s1052974/

### Attribution

(1) Mr. Smith **said**: “There is no correlation between smoking and lung cancer.”

<b>SOURCE</b>	e.g. Mr.Bretz, the newspaper, scientists, she, some, NONE
<b>CUE</b>	e.g. say, according to, idea, reportedly, PUNCTUATION
<b>CONTENT</b>	direct/indirect/mixed attributions

Attribution as the “relation ascribing the ownership of an attitude towards some linguistic material, i.e. the text itself, a portion of it or its semantic content, to an entity.” [1]. This comprises ASSERTIONS as well as BELIEFS, FACTS and EVENTUALITIES.

### Why we should care

Sources differ in bias, expertise, reliability and this affects the relevance and truth value of the attributed content.



Mr. Smith 1      Mr. Smith 2      Mr. Smith 3  
 president of Cigarettes&Co    oncologist at NHS Cancer trust    a smoker

Mr. Smith **said**  
**ventured** : “There is no correlation ... ”  
**joked**

Identifying attribution relations would enable:

- Retrieving attributions to a specific source
- Determining relevance/reliability for IR tasks
- Multi-perspective QA and summarisation
- Profiling

### Where we are now

Previous attribution extraction has addressed opinions or direct and indirect reported speech. The scope of reported speech attribution extraction is however limited to:

non-anaphoric NE sources  
 a limited set of verb cues  
 intra-sentential attributions

Considering only a portion of this relation, current attribution extracting systems suffer from:

**VERY LOW RECALL** (e.g. 13% for NewsExplorer [2])

**LOW PRECISION** (e.g. 52% [4])

Thus robust applications cannot be developed.

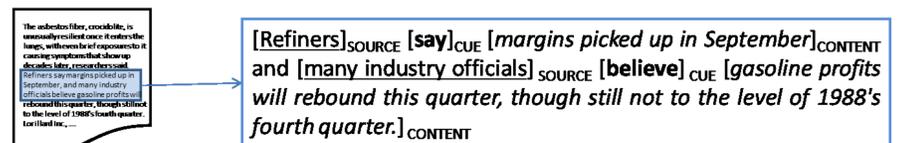
### A database of attribution relations

I have collected and further annotated a large attribution database from the PDTB [3] that supports further investigation of attribution and the training and testing of attribution extraction systems:

- Around 10,000 attributions
- Annotation of source, cue, content
- No pre-assumptions (all type of attributions are annotated)

### Where we want to go

#### Automatic attribution extraction



Wall Street Journal excerpt.

Issues: Attribution cues alone are not sufficient to identify attribution: e.g. **add** has a predictivity of attribution relations of 0.43. (Of 722 instances of the verb *add* in the PDTB, only 312 occur with attribution.)

#### Automatic feature recognition

Attribution relations carry features relevant for the interpretation of the content (i.e. determining its relevance and reliability).

**Authorial stance**, author’s commitment to the truthfulness of the content:

- committed: e.g. *acknowledge, admit*
- not committed: e.g. *lie, joke*
- neutral: e.g. *say, think*

**Source attitude**, e.g.:

- positive: e.g. *beam, smile*
- negative: e.g. *angrily, fume*

**Illocutionary act**: assertive (e.g. *say*), directive (e.g. *request*), commissive (e.g. *promise*), expressive (e.g. *think*), declarative (e.g. *nominate*)

**Sources type**, e.g.:

SOURCE TYPE	EXAMPLE	FEATURES
group	<i>scientists</i>	plural, mass noun
individual	<i>a man</i>	singular
named	<i>Mr. McMillin</i>	NE
unnamed	<i>a spokesman</i>	not NE
collectivity	<i>analysts</i>	plural
aggregation	<i>some entrepreneurs</i>	plural and quantifiers

**Level of nesting** of an attribution into another (the more passages a piece of information goes through, the higher the chances that it is less true to the original, as in the whisper game).

(2) [many industry officials]<sub>SOURCE</sub> [**believe**]<sub>CUE</sub> [*gasoline profits will rebound this quarter, though still not to the level of 1988’s fourth quarter*]<sub>CONTENT</sub>

Source type: group, unnamed, aggregation | Illocutionary act: expressive  
 Authorial stance: neutral | Level of nesting: 2  
 Source attitude: tentative

### References

- [1] S. Pareti and I. Prodanof. Annotating attribution relations: Towards an Italian discourse treebank. In *Proceedings of LREC10*. European Language Resources Association (ELRA), 2010.
- [2] B. Pouliquen, R. Steinberger, and C. Best. Automatic detection of quotations in multilingual news. In *Proceedings of RANLP*, 2007.
- [3] R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber. The Penn Discourse Treebank 2.0. In *Proceedings of LREC08*, 2008.
- [4] N. Schneider, R. Hwa, P. Gianfortoni, D. Das, M. Heilman, A. W. Black, F. L. Crabbe, and N. A. Smith. Visualizing topical quotations over time to understand news discourse. Technical Report T.R. CMU-LTI-10-013, Carnegie Mellon University, Pittsburgh, PA, July 2010.