Who said that? The information, the source...and the rest: Developing an annotation schema for attribution

Silvia Pareti (silvia_prt@yahoo.it) and Bonnie Webber (bonnie@inf.ed.ac.uk)

Università degli Studi di Pavia

Istituto di Linguistica Computazionale-CNR-Pisa

THE UNIVERSITY of EDINBURGH **informatics**

What is attribution?

Fiona | **says** | "This afternoon it will rain"

Source
Type
- Writer
- Other ✓
- Arbitrary

Content

Cue
Type
- Assertion ✓
- Fact
- Belief
- Eventuality

Scopal change
- None ⊙
- Scopal_change ○

Factuality
- Factual ⊙
- Non-factual ○

Why annotate attribution?

The reliability of information strongly depends on its source. Different sources can differ in bias and reliability.

?

Fiona

BBC NEWS 24 — This afternoon

Related works

[Prasad et al., 2007] – attribution of discourse connective and their arguments

[Skadhauge&Hardt, 2005] – sentence level attribution

[Wiebe, 2002] – word level attribution

[Murphy, 2005] – analysis of English and Italian attribution markers

The annotation schema developed for the annotation of attribution in the PDTB (Penn Discourse TreeBank) has been adopted and adapted to the present study.

Factuality

L'umanità | deve proclamare | uno sciopero a oltranza | fino alla distruzione di
The world | *should proclaim* | *a non-stop strike* | *till the destruction of*

tutti gli armamenti nucleari.
all nuclear armaments

This accounts for the cases where a negation, an imperative, a question, a future tense or another element, such as an adverb or a modal expressing possibility or probability and scoping over the cue, changes the attribution into **non-factual**. This means, the relation between content and source is not presented as a fact.

Hidden sources

E' questa la forchetta del prezzo di vendita | che | è stata resa nota …
It is this the selling price range | *that* | *has been announced*

Si | sa soltanto che | l'offerta italiana sarà compresa tra 1.000 e 1.500 milioni
One knows only that | *the Italian offer will be included between 1.000-1.500 ml*

La gente | dice | : | nessuno è indispensabile.
The people | *say* | : | *no one is essential.*

When the source is hidden by means of a passive or impersonal construction, this is marked as **Arbitrary**.

Coreference resolution

Ad | affermarlo | *LO | sono | i magistrati milanesi L. Orsi e V. Perozziello
Saying | *it* | *are* | *the Milanese judges L. Orsi and V. Perozziello*

Very often the attribution relation recalls sources and contents by means of an anaphoric personal or demonstrative pronoun. Fundamental is a bidirectional pointer in order to make the attribution relation retrievable from both referents. Although ideally this issue should be addressed separately, the lack of a suitable tool at present requires the addition of another layer of annotation in order to allow coreference resolution if part of an attribution relation.

Probabilmente | Vialli | non ha dimenticato | le voci sulla sua presunta vita
Probably | *Vialli* | *has not forgotten* | *the rumors about his presumed*

allegra durante i Mondiali del 1990 | rivelate | su Italia1 | da M. Mosca
'happy life' during the 1990 World Cup | *revealed* | *on Italia1* | *by M. Mosca*

E | Ø | non crede che | la recente alleanza tra Juventus e Milan possa
And | *(he)* | *doesn't believe that* | *the recent alliance Juventus-Milan could*

cambiare molto il comportamento dei commentatori sulle emittenti di B.
change much the commentators' behaviour on Berlusconi's televisions.

Scopal change

Se c'è, cioè, una maggioranza in Parlamento [...] | , | Ø | penso che
If there is a majority at the Parliament [...] | , | *(I)* | *think that*

la legislatura possa utilmente proseguire.
the legislature could usefully continue.

The attribution is factual as negation or other particles scope on the content.

Scope and nested attribution

Senza dimenticare le qualità del Moretti attore, | notevolissimo | secondo
Without forgetting the qualities of Moretti actor, | *remarkable(m.s.) according to*

Ghezzi | : | "Nel Portaborse Nanni ha dimostrato tutto il suo valore …"
Ghezzi | : | *"In Portaborse Nanni has proved all his value …"*

Secondo | quanto | riportato | da un quotidiano di Bergamo
According to | *what* | *was reported* | *by a newspaper from Bergamo*

Ø | avrebbe sibilato | ai suoi: | "…"
he | *hissed(quotative conditional)* | *to his folks:* | *"…"*

Type of cues

Determining the type a cue belong to is not a trivial task. A taxonomy of possible cues can't be employed to surely determine the type as this is highly dependent on the context as many verbs, for example, are **polysemous**. This will surely reduce interannotator agreement. Unfortunately the PDTB can't provide data in this regard as the whole attribution annotation was done by a single person.

assertion		belief		facts		eventualities	
affermare	to assert	credere	to believe	ricordare	to remember	permettere	to allow
sostenere	to claim	pensare	to think	sapere	to know	sostenere	to support
osservare	to observe	dubitare	to doubt	osservare	to observe	desiderare	to wish

---

The ISST (Italian Semantic-Syntactic Treebank) Corpus

305,547 word tokens from Italian newspaper articles (1985-1995)

5 layers of annotation:
- orthographic
- syntactic (constituents)
- syntactic (dependency)
- semantics
- morpho-syntactic

References

Murphy, A. C., Markers of attribution in English and Italian opinion articles: A comparative corpus-based study. *ICAME Journal* vol. 29 pp. 131-150, 2005..

Prasad, R., Dinesh, N., Lee, A., Joshi, A., Webber, B., Attribution and its Annotation in the Penn Discourse TreeBank. In *Traitement Automatique des Langues, Special Issue on Computational Approaches to Document and Discourse*, vol. 47, no. 2:43-64, 2007.

Skadhauge, P. R., Hardt, D., *Syntactic Identification of Attribution in the RST Treebank*. In Proceedings of the 2nd International Joint Conference on Natural Language Processing, Jeju Island, Korea, 11-13 October, 2005.

Wiebe, J., *Instructions for annotating opinions in newspaper articles*. Technical report TR-02-101, Department of Computer Science, University of Pittsburgh, 2002.