

The Independent Encoding of Attribution Relations

Silvia Pareti

School of Informatics / The University of Edinburgh, UK

S.Pareti@sms.ed.ac.uk

Abstract

Attribution relations have been annotated as discourse relations, attributes of discourse relations, structures carrying factuality, frames for the expression of subjective language, quote–speaker relations and classes of temporal references. While this proves their relevance for different domains, whether as disruptive elements to rule out or essential carriers to treasure, it provides only a limited and marginal picture of this relation. In this paper I will overview its interconnection with other domains, in particular its strong connection with discourse relations, and motivate the need for an independent encoding. I will also highlight what the elements that constitute an attribution relation or contribute to its interpretation are and introduce the attribution corpus developed starting from the annotation in the PDTB.

1 Introduction

The annotation of attribution has been addressed by studies from different domains having however a different annotation focus, e.g. discourse relations (Prasad et al., 2006), sentiments (Wiebe, 2002), factuality (Saurí and Pustejovsky, 2009). Attribution relations (ARs) are relevant for other domains as they can be carriers or constitute themselves informative clues for other phenomena. However, the annotation of attribution has been so far tailored to suit the needs of the ‘hosting domain’, thus including only aspects and structures relevant to the annotation purpose. For example, the MPQA Opinion Corpus (Wiebe, 2002) has annotated the attribution

of speech events, however only when these were a vehicle for private states and only intra–sententially.

All these approaches fail to fully encode attribution and are therefore not suitable to provide a solid basis for attribution studies and to train attribution extraction systems. It is therefore beneficial to separate attribution from other annotation domains and build a resource that can encompass a wider range of attribution structures and reach a more structured and deeper representation of this relation.

In this paper I will explore the interconnections between attribution and discourse and investigate what are the essential traits of an attribution and how it should be encoded. I will start with a brief presentation of the range of domains where attribution is relevant and how they have encoded this relation (Sec. 2). That will provide the framework for taking a closer look at the inclusion of attribution in the PDTB (Prasad et al., 2006) and the overlap and mutual effects of attribution and discourse (Sec. 3.1). However close these two domains are, I will show that there is no exact correspondence between attribution and discourse and that attribution should be annotated as its own relation – an “attribution relation” or AR (Sec. 3.2).

Section 4 will then present the constitutive elements of ARs (Sec. 4.1) and other elements linked to an AR that can contribute to its interpretation (Sec. 4.2). Sec. 4.3 will overview attributes of attribution, the ones that have been included in the PDTB and additional relevant ones that have been included or considered for inclusion in the annotation schema developed for attribution. This has been used to build an attribution corpus starting from the

annotation of ARs in the PDTB.

2 Background

ARs have previously been partially annotated in the context of annotating other phenomena of interest to language processing. This work has only marked the portion of attribution of interest for the main task at focus (e.g. the annotation of discourse relations or event factuality). In this section I will survey some of the most prominent annotation efforts that have included attribution and highlight how their approach has encoded attribution and the perspective they have taken at it.

A portion of ARs has been addressed and annotated by studies dealing with ‘subjectivity analysis’. A subset of ARs, namely opinions and beliefs, are part of the ‘private states’ at focus in the MPQA Opinion Corpus (Wiebe, 2002). Despite a strong overlap in scope, the approach is considerably different. While a private state is defined as “an experiencer holding an attitude, optionally toward an object” (Wiebe, 2002, p.4), attribution goes in the opposite direction. The object is not optional, but a fundamental element of the AR, intended as “a relation of ‘ownership’ between abstract objects and individual or agents” (Prasad et al., 2008, p.40).

Discourse studies encode ARs annotating two elements: the *attributed span* and the **attribution span**, as in Ex.(1)¹. When attribution itself is considered as a discourse relation (Carlson and Marcu, 2001; Wolf and Gibson, 2005), these two annotated elements correspond to discourse units. Attribution holds from the attributed span, *nucleus*, towards the attribution span, *satellite*.

- (1) **Mr. Englund added** *that next month’s data isn’t likely to be much better, because it will be distorted by San Francisco’s earthquake.* (wsj_0627)

Studies concerned with the attribution of direct quotes, e.g. the Sydney Morning Herald Corpus (O’Keefe et al., 2012), also annotate attribution as composed by two elements, i.e. *quote–speaker* pairs (Ex.(2)). The element connecting speaker and quote

¹The attribution span is highlighted in bold in the examples, while the attributed span is in italics. Examples taken from the WSJ (WSJ article reference in brackets)

and expressing the type of AR (e.g. assertion or belief) is not annotated. However, the attribution of quotes implies that what is attributed is an assertion.

- (2) “*The employment report is going to be difficult to interpret,*” said Michael Englund, economist with MMS International, a unit of McGraw-Hill Inc., New York. (wsj_0627)

The textual anchor establishing the relation is annotated by some studies (Glass and Bangay, 2007; Pouliquen et al., 2007), however as a device helping the identification and therefore extraction of an AR and not as integral part of the relation itself. In particular, speech verbs (e.g. say, report) are identified as their grammatical subject often expresses the source entity of the AR and their object the attributed element.

ARs also affects temporal references, and ‘reporting’ has been included as an event class in TimeML (Pustejovsky et al., 2003) and reporting events have been annotated in TimeBank (Pustejovsky et al., 2006). Accounting for the relation between the time the document was produced and that of the reporting event remained an issue. ARs insert an additional point in time, i.e. that of the enunciation, in case of an assertion or the temporal point where a belief or fact was factual. For example, ‘John thought it was a good idea’ reflects John’s belief at a past point in time. This belief might have changed at the point the article was written or the present time.

Attribution has also strong implications for the factuality of the events expressed in the attributed span. This motivates its partial inclusion in FactBank (Saurí and Pustejovsky, 2009) where the attributed span itself is not marked, but events contained in it (e.g. ‘left’ in Ex.(3)) are linked to their source by source–introducing predicates (SIPs) in order to derive their factuality. The SIP in Ex.(3) implies that the event underlined in the example is considered by the source as just a possibility.

- (3) Berven **suspects** that Freidin left the country in June. (Saurí and Pustejovsky, 2009, p.236)

3 Attribution and Discourse

Attribution is intertwined with other annotation domains. In particular, it overlaps and has implications

relevant to discourse relations, factuality and subjectivity analysis, as briefly introduced in Sec. 2.

The PDTB is the biggest existing resource annotating ARs. However, what makes it a suitable starting point to study attribution is it has not first defined a strict set of rules that attribution should obey to be considered in the scope of the project, thereby restricting attribution to its ‘pretty’ and more standard structures. This, combined with the size of the corpus, means that a wide range of attribution structures can be observed. For example, attributions to unnamed or implicit entities or having no reporting verb. However, I will argue that attribution should be treated and annotated independently and motivate the effort to disjoint it from discourse annotation.

3.1 Intertwined

Attribution relations are closely tied to discourse relations, and have variously been included as a discourse relation itself (Wolf and Gibson, 2005; Carlson and Marcu, 2001) or as an attribute of discourse relations (Prasad et al., 2006). They were included in the PDTB since it was recognised that “a major source of the mismatches between syntax and discourse is the effect of attribution” (Dinesh et al., 2005, p.36).

If the arguments of a discourse connective are taken to be its syntactic arguments, attribution could lead to incorrect semantic interpretation as in Ex.(4) below (Prasad et al., 2008, p.2966). It is therefore important to recognise and exclude attribution in such cases.

- (4) a. Factory orders and construction outlays were largely flat in December [Arg1.]
b. while **purchasing agents said** [Conn.]
c. *manufacturing shrank further in October* [Arg2.]. (wsj_0178)

While attribution is disruptive for discourse relations, these could be of great advantage to the identification of the *content*, i.e. the attributed span when the AR is indirect, i.e. the attributed span, is not surrounded by quote markers. While some studies (Skadhauge and Hardt, 2005; de La Clergerie et al., 2009) have taken an intra-sentential look at attribution and considered as the content of an AR the

grammatical object of a reporting verb, this is not a viable solution when dealing with a wider range of ARs. Here discourse structure may play a role above the level of single sentences.

The ARs collected from the PDTB show that around 17% of ARs extend over more than one sentence (e.g. three sentences in Ex.(5)). Moreover, only half of these are attributions of direct quotes. English does not mark indirect reported speech grammatically, unlike for example German (Ruppenhofer et al., 2010), where this is associated with subjunctive mood. The issue is how to determine the content span boundaries of indirect ARs when the syntactic structure would be of no help. While not always unambiguous also for human readers, recognising a content extending over more sentences could be partly achieved with the help of discourse relations.

- (5) **According to Audit Bureau of Circulations**, *Time*, the largest newsweekly, had average circulation of 4,393,237, a decrease of 7.3%. *Newsweek’s circulation for the first six months of 1989 was 3,288,453, flat from the same period last year. U.S. News’ circulation in the same time was 2,303,328, down 2.6%.* (wsj_0012)

In Ex.(5), the last two sentences are a continuation of the content but they bear no syntactic relation with the first sentence. Instead, they are two discourse relations (both entailing an implicit connective *and*, of type Comparison:Contrast:Juxtaposition) binding the first part of the content span with the second and the third sentence. Discourse alone might not provide sufficient evidence to determine the content extension. Nonetheless, in combination with other triggers, e.g. verb tense and mood, this could allow the correct identification of inter-sentential indirect ARs.

3.2 Distinct

The PDTB is rich in attribution annotation and represents a great starting point for the collection of a large resource for the study of attribution. However, what is annotated is not attribution itself but the attribution of discourse connectives and their arguments. Attribution is therefore subordinate to discourse and reconstructing a full AR can be rather complex.

The content of an AR might not be fully corresponding to a discourse relation or one of its arguments, but be composed of several discourse connectives and their arguments. We can consider, for example, the marked AR that corresponds to the second paragraph of the excerpt below (wsj_0437):

The reports, attributed to the Colombian minister of economic development, said Brazil would give up 500,000 bags of its quota and Colombia 200,000 bags, the analyst said.

(HOWEVER) *These reports were later denied by a high Brazilian official, who said Brazil wasn't involved in any coffee discussions on quotas, the analyst said.* (wsj_0437_12²)

(BUT) The Colombian minister was said to have referred to a letter that he said President Bush sent to Colombian President Virgilio Barco, and in which President Bush said it was possible to overcome obstacles to a new agreement.

The content span of this AR, the text in italics, is partially included in all three discourse relations below: the two implicit ones, having *however* and *but* as connectives, and the one with discourse connective *later*. In order to reconstruct the full AR from the annotation, it was necessary to take all three discourse relations into account and merge together the text spans they were attributing to 'the analyst said'.

1. The reports said Brazil would give up 500,000 bags of its quota and Colombia 200,000 bags (Arg1)

HOWEVER (Implicit connective)

These reports were later denied by a high Brazilian official (Arg2)

2. The reports said Brazil would give up 500,000 bags of its quota and Colombia 200,000 bags (Arg1)

LATER (Connective)

These reports were denied by a high Brazilian official (Arg2)

3. *who said Brazil wasn't involved in any coffee discussions on quotas* (Arg1)

BUT (Implicit connective)

²Examples from the attribution corpus report the AR unique ID.

The Colombian minister was said to have referred to a letter that he said President Bush sent to Colombian President Virgilio Barco, and in which President Bush said it was possible to overcome obstacles to a new agreement (Arg2)

This shows that there is no exact correspondence between ARs and discourse arguments and therefore some ARs are partially or not annotated. This happens if part of their content is not corresponding to a discourse argument or when the whole AR is included in a discourse argument as in Arg1 of *But* (relation 3 above). The nested AR (i.e. '**who said Brazil wasn't involved in any coffee discussions on quotas**') in this attribution argument is just not annotated.

While the PDTB is a great resource for attribution, attribution cannot be handled as a mere attribute of discourse connectives and their arguments as there is no exact correspondence between ARs and discourse relations. I have therefore disjoint the annotation of discourse and attribution by collecting the ARs in the PDTB and reconstructing incomplete ARs, thus creating a separate level of annotation.

4 The Independent Encoding of Attribution

ARs are encoded in the PDTB as formed by two elements, the attributed material, i.e. abstract object or discourse units, and the attribution span. I will argue that this encoding of ARs is not sufficient and cannot suit the variety of purposes attribution could serve. It does not allow, for example, to easily identify attributions to a specific source. In the next section I will present which are the core elements of this relation, which are additional and the attributes that we can associate with ARs.

4.1 Constitutive Elements of ARs

There are three elements necessary to define the relation of attribution based on textual evidence. These elements are the two that are related, i.e. the attributed material or *content* and the entity this is attributed to, the source, which may or may not correspond to the author of the article, but also the link connecting them, i.e. the **cue**. Annotating the cue is fundamental as this represents the key to the correct identification and interpretation of the relation

it establishes. Is the AR in Ex.(6a)³ a statement or an opinion? Is it factual or just a speculation? Does the AR in Ex.(6b) entail that the source or the author believe in the truth of the proposition in the content?

- (6) a. Network officials involved in the studio talks **may hope** *the foreign influx builds more support in Washington*, but that seems unlikely. (wsj_2451.pdtb_09)
- b. *“He taught me how to play like a gypsy,”* **jokes** the musician. *“I didn’t learn to count until I got to Juilliard.”* (wsj_1388.pdtb_02)

Although source, cue and content are constitutive elements of ARs, they can possibly be only implicitly or anaphorically expressed as in Ex.(7), where the source is implicit and the content anaphorically recalled by a pronoun.

- (7) [...] *profound change toward free-market economics, especially in the statist countries.* **Having said** *that*, we must caution against an apparent tendency to overstate the case. (wsj_1529)

In order to encode all the constitutive elements of an AR independently, I had to further annotate the attribution corpus collected from the PDTB. The text labelled as attribution span was therefore further annotated with the source and cue elements of the AR. However, these were not the only elements constituting the attribution span.

4.2 Other Relevant Components of ARs

Beside the constitutive elements of ARs, the surrounding context can carry further information relevant to the AR, although optional. When the attribution span contains relevant elements that are neither part of the source nor of the cue, these should be marked as SUPPLEMENTAL. In particular, supplemental elements are those providing a context for interpreting an AR, including its:

- setting (time, place, audience) (Ex.(8)⁴);
- topic (Ex.(9));

³From now on, examples will mark the cue of an AR in bold, the source underlined and the content in italics.

⁴Supplements are represented in the examples in small capitals.

- communication medium (Ex.(10));
- relevance to the author’s argument (Ex.(11));
- manner (Ex.(12)).

- (8) *“Ideas are going over borders, and there’s no SDI ideological weapon that can shoot them down,”* **he told** [A GROUP OF AMERICANS] [AT THE U.S. EMBASSY] [ON WEDNESDAY]. (wsj_0093_07)
- (9) **OF SONY, Mr. Kaye says:** *“They know there’s no way for them to lose. They just keep digging me in deeper until I reach the point where I give up and go away.”* (wsj_2418_15)
- (10) **Trade and Supply Minister Gerhard Briksa said** IN A LETTER PUBLISHED IN THE YOUTH DAILY JUNGE WELT *that the rise in alcohol consumption in East Germany had been halted;* (wsj_1467_05)
- (11) **AS AN INDICATOR OF THE TIGHT GRAIN SUPPLY SITUATION IN THE U.S., market analysts said** *that late Tuesday the Chinese government, which often buys U.S. grains in quantity, turned instead to Britain to buy 500,000 metric tons of wheat.* (wsj_0155_16)
- (12) *“A very striking illusion,”* **Mr. Hyman says** NOW, HIS VOICE DRIPPING WITH SKEPTICISM, *“but an illusion nevertheless.”*(wsj_0413_14)

If part of the attribution span, these elements have been included in the annotation of the attribution corpus, with the label ‘supplement’. The information contained in the supplement might still not be sufficient to fully evaluate and fully understand an AR. In Ex.(12) we don’t know what the source considers an ‘illusion’, i.e. the topic this assertion is about. Nonetheless, the supplement usually provides enough elements for the interpretation of the AR. This without having to process the whole article or resorting to external knowledge.

4.3 Features of Attribution Relations

There are several features relevant for encoding ARs. Features that can capture if an AR is factual or contribute to determine whether the attributed

proposition is truthful, differentiate sources and attributions. These features can enable applications of attribution beyond the retrieval of ARs having a specific source or cue. The PDTB annotates four such features. One is the *type of attribution*, i.e. belief, assertion, fact or eventuality. This affects the factuality of the content since in an AR of type ‘fact’ this is higher, and it usually implies that the source and author believe it is truthful, while in an attributed belief the level of factuality is much lower as in Ex.(13). The source is not sure about the proposition expressed in the content being really true.

(13) Meanwhile, some U.S. officials **fear** *PLO chief Arafat is getting cold feet and may back off from his recent moderation and renunciation of terrorism.*(wsj_1682_00)

A second feature of ARs in the PDTB is the *type of source*, i.e. writer, other or arbitrary. This aspect allows to distinguish between real and ‘pseudo-attributions’. In the latter the attribution is not to a third party but to the writer or author of the article, who is the default source of the whole article, and thus redundant.

There are other two attributes, *determinacy* and *scopal polarity*, accounting for the factuality of the AR (Ex.14a) and the polarity of its content respectively (Ex.14b). While in the first example the AR is just an hypothesis, therefore not factual, in the second one the AR itself is factual, the content being in the scope of the negation instead.

(14) a. [...] BY NEXT WEEK the network **may announce** *“Teddy Z” is moving to 8:30 p.m. from its 9:30 time slot*[...] (wsj_1150_00)

b. DEPOSITS **aren’t expected** *to exceed withdrawals in the foreseeable future*, as the industry continues to shrink. (wsj_1293_03)

Beside the features already included in the PDTB, ARs carry other relevant ones worth annotating. As noted by (Karttunen and Zaenen, 2005), the attribution cue can indicate the *authorial stance*, i.e. the position the author takes towards the truth of the proposition expressed in the content. By choosing to use a factive (e.g. admit, regret, realise)

or counter-factive cue (e.g. lie, joke (Ex.6b)), the author implies a certain degree of commitment or non-commitment towards the truth of the attributed statement. Using a non-factive cue (e.g. say, claim, suggest), the author remains instead more neutral. The authorial stance is a relevant feature of ARs as the commitment the author expresses towards the statement can be employed to uncover ideological biases or, if we assume the author to be trustworthy, to determine if the statement is truthful.

Attribution cues can also express the *source attitude*, i.e. the sentiment the source itself expresses towards the proposition, e.g. ‘negative’ in Ex.(13) and positive in Ex.(15). While the most frequent reporting verbs (e.g. say) tend to be neutral, other verbs normally not associated with a reporting meaning, and in particular manner verbs (e.g. smile, quip, purr), can express this feature.

(15) *“We’ve had the Russians and Chinese, and people from India visiting us,”* Mr. Iverson beams. *“Everyone in the world is watching us very closely.”* (wsj_2153_01)

These features have not yet been included in the annotation as a preliminary inter-annotator agreement study showed that their definition needs further investigation. In this study, two expert annotators applied the annotation schema (Pareti, 2011) to 14 articles from the WSJ corpus (380 jointly identified ARs) and assigned them values for the four features annotated in the PDTB and the two additional ones I have proposed. Cohen’s Kappa values for the correct selection of the value for authorial stance (i.e. committed, not.committed, neutral) and source attitude (i.e. positive, negative, tentative, neutral, other) were .48 and .20 respectively.

Other features do not require manual annotation as they can be derived from lexical and syntactic clues of the AR elements — for example, whether a source is a group or an individual, named or unnamed. Another automatically derivable feature is whether the attribution content is completely (direct AR), partly (mixed AR) or not at all (indirect AR) surrounded by quotation markers. This feature was called “quote status” (Pareti, 2012) and included in the attribution corpus developed. It is relevant not only because direct quotes are generally used to reflect the exact words uttered by the source, and are

thus more faithful to the original statement, but also because they tend to occur with different syntactic structures and lexical choices. For example, in the attribution corpus collected, the verb cue ‘suggest’ never occurs in the context of a direct attribution, while ‘joke’ always associates with a direct quote.

5 Conclusion and Future Work

This paper overviews the importance of ARs in different domains. ARs can carry temporal events and subjective expressions, affect the factuality of events and cause a mismatch between syntactic and discourse arguments of discourse connectives. However, annotating ARs ‘ad hoc’, as part of other annotation projects, is rather detrimental as it prevents attribution from being encoded in an independent and more complete way.

While the PDTB represents a fundamental source of attribution annotation, I have shown the limitations of such annotation and proved the need for an independent encoding of attribution. For this reason, I have created an independent corpus of ARs starting from the annotation in the PDTB. This was done by separating the annotation of ARs from that of discourse relations and further annotating each AR according to a previously developed annotation schema. This resource could enable reaching a deeper understanding of ARs and allow the development of AR extraction systems that can be reliably employed (e.g. for information extraction or multi-perspective QA). The independent encoding would also allow projects from other domains to rely on the annotation for the portion relevant to the phenomenon at study.

The attribution corpus in its first version is in a flat CoNLL style — i.e. each line corresponds to one AR and each column to one element, feature or pointer of the AR. I am currently developing an XML format for AR annotation, which allows for the representation of nested ARs.

Acknowledgements

The author is supported by a Scottish Informatics & Computer Science Alliance (SICSA) studentship.

References

- Lynn Carlson and Daniel Marcu. 2001. Discourse tagging reference manual. Technical report ISITR- 545. Technical report, ISI, University of Southern California, September.
- Eric de La Clergerie, Benot Sagot, Rosa Stern, Pascal Denis, Gaelle Recource, and Victor Mignot. 2009. Extracting and visualizing quotations from news wires. In *Proceedings of L&TC 2009, Poznan, Poland*.
- Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Rashmi Prasad, Aravind Joshi, and Bonnie Webber. 2005. Attribution and the (non-)alignment of syntactic and discourse arguments of connectives. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky, CorpusAnno ’05*, pages 29–36, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Kevin Glass and Shaun Bangay. 2007. A naive, saliencebased method for speaker identification in fiction books. In *In Proceedings of the 18th Annual Symposium of the Pattern Recognition Association of South Africa (PRASA 07)*, pages 1–6, November.
- Lauri Karttunen and Annie Zaenen. 2005. Veridicity. In Graham Katz, James Pustejovsky, and Frank Schilder, editors, *Annotating, Extracting and Reasoning about Time and Events*, number 05151 in Dagstuhl Seminar Proceedings, Schloss Dagstuhl, Germany. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI).
- Tim O’Keefe, Silvia Pareti, James Curran, Irena Koprinska, and Matthew Honnibal. 2012. A sequence labelling approach to quote attribution. Manuscript submitted for publication.
- Silvia Pareti. 2011. Annotating attribution relations and their features. In Kamps J. Karlgren J. Alonso, O., editor, *ESAIR’11: Proceedings of the CIKM’11 Workshop on Exploiting Semantic Annotations in Information Retrieval*. ACM Press, October.
- Silvia Pareti. 2012. A database of attribution relations. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Mehmet Ug(ur Dog(an, Bente Maegaard, Joseph Mariani, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC’12)*, Istanbul, Turkey, may. European Language Resources Association (ELRA).
- Bruno Pouliquen, Ralf Steinberger, and Clive Best. 2007. Automatic detection of quotations in multilingual news. In *Proceedings of the International Conference Recent Advances In Natural Language Processing (RANLP 2007)*, pages 487–492.
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Aravind Joshi, and Bonnie Webber. 2006. Annotating attribution in

- the Penn Discourse TreeBank. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, SST '06, pages 31–38.
- Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn Discourse Treebank 2.0. In *Proceedings of the 6th International Conference on Language Resources and Evaluation LREC08*.
- James Pustejovsky, Jos Castao, Robert Ingria, Roser Saur, Robert Gaizauskas, Andrea Setzer, and Graham Katz. 2003. Timeml: Robust specification of event and temporal expressions in text. In *Proceedings of IWCS-5, Fifth International Workshop on Computational Semantics*.
- James Pustejovsky, Jessica Littman, Roser Saur, and Marc Verhagen. 2006. Timebank 1.2 documentation. Technical report.
- Josef Ruppenhofer, Caroline Sporleder, and Fabian Strötker. 2010. Speaker attribution in cabinet protocols. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, *Proceedings of the Seventh conference on International Language Resources and Evaluation LREC10*. European Language Resources Association (ELRA).
- Roser Saurí and James Pustejovsky. 2009. Factbank: A corpus annotated with event factuality. In *Language Resources and Evaluation*, (43):227–268.
- Peter R. Skadhauge and Daniel Hardt. 2005. Syntactic identification of attribution in the RST treebank. In *Proceedings of the Sixth International Workshop on Linguistically Interpreted Corpora*.
- Janyce Wiebe. 2002. Instructions for annotating opinions in newspaper articles. Technical report, University of Pittsburgh.
- Florian Wolf and Edward Gibson. 2005. Representing discourse coherence: A corpus-based study. *Comput. Linguist.*, 31:249–288, June.