Chapter 2

Conveying Attitude with Reported Speech

Sabine Bergler
Concordia University
Department of Computer Science and Software Engineering
1455 de Maisonneuve Blvd. West
Montreal, Quebec, H3G 1M8, Canada
Email: bergler@encs.concordia.ca

Abstract

Attribution is a phenomenon of great interest and a principled treatment is important beyond the realm of newspaper articles. The way natural language has evolved to reflect our understanding of attribution in the form of reported speech can guide investigations into principled representations forming the basis for shallow text mining as well as belief revision or maintenance.

Keywords: attribution, reported speech, reliability of information, argumentative structure, profile structure, potential belief space.

1. Introduction

Society has developed a multitude of mechanisms that serve to authenticate items, and in particular information. Signatures authenticate letters, paintings, and seal contracts. Imprints on money, seals, and forms make them official. Insignia establish membership in certain groups, as do uniforms and religious symbols. Information, likewise, has well established mechanisms of authentication, which vary slightly from society to society. The Native American language Pawnee has four different prefixes that obligatorily have to mark statements for their reliability (hearsay, reasonably certain but not witnessed directly, leaving room for doubt, or mere inference) (Mithun, 1999). And while the number and type of such evidential markers differ in different languages, hearsay is maybe the most widespread one.

English and most European languages do not have a system of evidential morphology, but mark hearsay and other evidentiality at the syntactic level. Reported speech, both in form of direct quotation (*... and then she said "I have to go."*) or indirect paraphrases (*... and then she said that she had to go.*), is the most formalized register. Reported speech is most prominent in newspaper articles, where it can occur in up to 90% of the sentences of an article. Computational linguistic treatments of newspaper articles usually ignore reported speech, either by omitting the material entirely, or by ignoring its evidential status. This paper argues that reported speech segments information into coherent subunits, called *profiles* after (Bergler, 1992). Different profiles can

imply different credibility of the source of the information, different roles of the source in an argumentative structure, or a different context (temporal or other). An extended example illustrates profiling on a product review article. This paper concludes that the mechanism of profiling (and its proper analysis) should be extended beyond reported speech to all explicit attributions, such as newsgroup messages, etc.

2. Evidential Analysis of Reported Speech

Reported speech is characterized by its syntactic structure: a matrix clause, containing at least the source as subject NP and a reporting verb, embeds the information conveyed in a complement clause. The complement is optionally introduced by "that" for indirect reported speech, and it is surrounded by quotation marks for direct reported speech. As argued in (Bergler, 1992), the complement usually conveys the *primary information* in newspaper articles and most other genres. In fact, the case where the matrix clause bears the major information, namely that something had been uttered by somebody under certain circumstances without the utterance itself being of importance, is rare (but see (Clark and Gerrig, 1990) for examples). The syntactic dominance of the matrix clause shows the semantic importance of the contained *circumstantial information* (Bergler, 1992), the who, when, where, and how. But the natural propositional encoding of the complement clause as embedded in the matrix clause is not suitable. Rather, the information of the matrix clause should be seen as a meta-annotation for interpreting the primary information differently in different contexts and for different purposes. Thus the a priori trust a Republican reader has in utterances by Cheney is different from a Democrat's. And a text will be interpreted differently at the time of the events unfolding and after additional information is known. This variability of the pragmatic force of the matrix clause also suggests that it cannot be "resolved" at the time of first text analysis, but has to remain attached in a form close to the original for further analysis. (Bergler, 1992) gives a general linguistic treatment of reported speech. This paper presents, in contrast, one particular implementation of the general representation for further automatic analysis. The underlying assumption is that the further processing will be by an information extraction or mining system that works with shallow, possibly statistical techniques. But the representation does not preclude the deeper linguistic analysis outlined in (Bergler, 1992).

Politics & Policy: Democrats Plan Tactic to Block Tax-Cut Vote
Threat of Senate Filibuster Could Indefinitely Delay Capital-Gains Package

(S_1) Democratic leaders have bottled up President Bush's capital-gains tax cut in the Senate and may be able to prevent a vote on the issue indefinitely.

(S_2) Yesterday, Sen. Packwood acknowledged, "We don't have the votes for cloture today."

(S_3) The Republicans contend that they can garner a majority in the 100-member Senate for a capital-gains tax cut.

(S_4) They accuse the Democrats of unfairly using Senate rules to erect a 60-vote hurdle.

(S_5) Democrats asserted that the proposal, which also would create a new type of individual retirement account, was fraught with budget gimmickry that would lose billions of dollars in the long run.

Figure 1. Adapted from Jeffrey H. Birnbaum, The Wall Street Journal, 10/27/89.

As Figure 1 demonstrates, the role of reported speech is attribution: the statement does not assert as `true' what amounts to the information content of the sentence, but a situation in which this content was proffered by some source. This device can be used both to bolster a claim made in the text already, and to distance the author from the attributed material, implicitly lowering its credibility (Anick and Bergler, 1992). Thus the credibility or reliability of the attributed information is always in question for reported speech and other attributions. If the attribution is used to bolster a claim already made by citing a particularly strong source for endorsement, ignoring the fact that an explicit attribution was made will do no great harm. This is in fact a frequent case in the type of newspaper articles typically used for large-scale system development and testing (as in MUC, TREC, DUC, etc.) and this is why ignoring attribution has been tolerable. But when a text is argumentative (opposing two or more points of view on a topic), speculative (when the final outcome of an event is not yet decided and the text uses different sources as predictors), or presents a personal opinion or experience, text meaning depends on proper attribution recognition (Bergler, 1995a). Argumentative or speculative text structure is not limited to newspaper articles. Scientific articles, too, use reported speech for this purpose, but in a different rhetorical style. And multi-participant political analysis segments on newscasts form the same phenomenon: different opinions are identified with different individuals and contrasted, even though we might term it broadcast speech, rather than reported speech. Interestingly, broadcast speech retains the required elements of reported speech, in that it is always anchored by the identity of the source and the circumstances of the utterance (date, occasion, place, etc.) as they are relevant to its analysis. The reported material is always literal and quoted, of course, but has still undergone an editing process, extracting the broadcast speech from a larger interview and potentially juxtaposing material that the source did not intend to. Thus the simple fact that no paraphrasing is involved does not make broadcast speech necessarily truer to the original than reported speech.

Reported speech in newspaper articles can be detected and analyzed without a complete syntactic analysis, using shallow means and standard tools. In a feasibility study, Doandes (2003) presents a knowledge-poor system that identifies sentences that contain reported speech in Wall Street Journal texts and analyzes them into structures inspired by (Bergler, 1992) and illustrated in Figure 2.

The system works in a shallow environment: Built on top of ERS (Witte and Bergler, 2003), it has access to slightly modified versions of the Annie tools distributed with GATE (Cunningham, 2002) and an in-house NP chunker and coreference resolution module. The NP chunker relies on the Hepple tagger (Hepple, 2000) and Annie Gazetteer, the coreference module has access to WordNet (Fellbaum, 1998).

BP	basic profile
OTHERCIRC	circumstantial information other than source and reporting verb
PARAPHRASE	paraphrased material, usually complement clause in case of indirect reported speech
REPSOURCE	source, in active voice the matrix clause subject
REPORTEDSPEECH	complement clause
REPVERB	reporting verb, main verb in matrix clause
QUOTEDSPEECH	material in quotation marks

Figure 2. Template for representing reported speech sentences in (Doandes, 2003).

Doandes uses part-of-speech tags to identify main verb candidates. In a detailed analysis of verb clusters, she determines main verbs and compares them against a list of likely reported speech

verbs. In case a reported speech verb is found, the sentence pattern (with complete part-of-speech annotations, annotations for NPs, and annotations for verb groups) is compared to the possible patterns for reported speech constructions as described in (Quirk et al., 1985). Figure 3 gives the resulting *basic profile* for the sentence: *Yesterday, Sen. Packwood acknowledged, "We don't have the votes for cloture today."*

```
BP
OTHERCIRC Yesterday,
PARAPHRASE
REPSOURCE Sen. Packwood
REPORTEDSPEECH, "We don't have the votes for cloture today."
REPVERB acknowledged
QUOTEDSPEECH "We don't have the votes for cloture today."
```

Figure 3. Example representation in (Doandes, 2003).

The development corpus consisted of 65,739 sentences from the Wall Street Journal, the test corpus of 2,404 sentences taken mainly from the Wall Street Journal, with a few articles from the DUC 2003 corpus of newspaper articles (DUC, 2003). 513 occurrences of reported speech were found and precision is 98.65%, recall is 63%. The analysis into basic profiles incurred some mistakes (such as retaining only part of the subject NP in the SOURCE slot). Using a strict notion of correctness for the entire basic profile, the performance drops to 87% precision and 55% recall.

Many recall problems are linked to limitations of the particular implementation, such as tagging errors, the NP chunking process (the NP chunker splits heavy NPs into several smaller chunks, thus occasionally obfuscating the reported speech pattern), and an incomplete list of reported speech verbs. (Doandes, 2003) works from a simple list of candidate reported speech verbs with no attempt at word sense disambiguation. The results seem satisfactory for the evaluation corpus, but will not necessarily hold outside the newspaper genre. (Wiebe et al., 1997) report on the difficulty of distinguishing *private state, direct speech, mixed direct and indirect speech, other speech event, other state or event*. Most of these categories describe attributions and thus do not need to be distinguished for profile structure at the level described here, even though their distinction would refine the use of the profile for subsequent processing.

3. Profile Structure

Figure 1 is typical for newspaper articles: information from two different points of view, here Democrats and Republicans, is interleaved. Ideally, an automatic system would group S_1 and S_5 into one profile, and S_2, S_3, and S_4 into another, effectively grouping Democrats versus Republicans. This is, however, not possible with shallow techniques. S_1 is not a reported speech sentence and thus does not generate a profile. World knowledge is required to infer that *Sen. Packwood* speaks for the *Republicans* in this article, but pronoun resolution techniques allow *they* to be resolved to *Republicans*, creating a merged profile from S_3 and S_4, enabling interpretation of *a 60-vote hurdle* in the context of S_3.

Profile structure is complementary to both rhetorical structure (cf. Marcu, 1997) and text structure (cf. Polanyi et al., 2004). It creates another type of context, which is coherent with respect to underlying processing assumptions, such as reliability of the source, or, as seen above, inferential assumptions (*60-vote hurdle*). For a more detailed discussion, see (Bergler, 1995a). The profile structure for the text in Figure 1 is given in Figure 4.

The use of profiles is simple: profiles provide a partition of the text according to the source of the information transmitted. This local context can be used for different reasoning. As seen above, a follow-up statement (*60-vote hurdle*) may make (more) sense when interpreted in the context of the previous utterance of the same source (*100-member senate*), even if other text had interfered.

BP
OTHERCIRC Yesterday,
PARAPHRASE
REPSOURCE Sen. Packwood
REPORTEDSPEECH, "We don't have the votes for cloture today."
REPVERB acknowledged
QUOTEDSPEECH "We don't have the votes for cloture today."

MERGED-PROFILE:

BP
OTHERCIRC
PARAPHRASE they can garner a majority in the 100-member Senate for a capital - gains tax cut.
REPSOURCE The Republicans
REPORTEDSPEECH they can garner a majority in the 100-member Senate for a capital - gains tax cut.
REPVERB contend
QUOTEDSPEECH

BP
OTHERCIRC
PARAPHRASE the Democrats of unfairly using Senate
rules to erect a 60-vote hurdle.
REPSOURCE They
REPORTEDSPEECH the Democrats of unfairly using Senate rules to erect a 60-vote hurdle.
REPVERB accuse
QUOTEDSPEECH

BP
OTHERCIRC
PARAPHRASE the proposal, which also would create a new type of individual retirement account, was fraught with budget gimmickry that would lose billions of dollars in the long run.
REPSOURCE Democrats
REPORTEDSPEECH the proposal, which also would create a new type of individual retirement account, was fraught with budget gimmickry that would lose billions of dollars in the long run.
REPVERB asserted
QUOTEDSPEECH

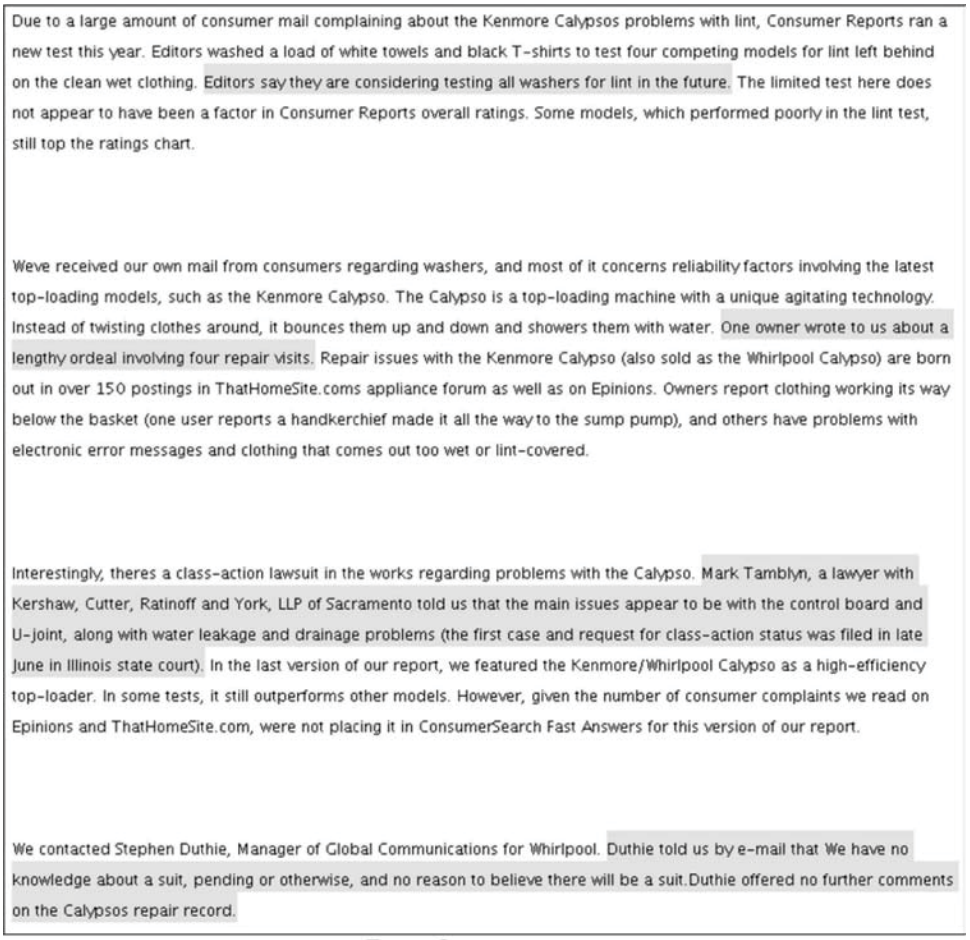
Figure 4. Profile structure for the text in Figure 3.

Statements from different sources cannot necessarily be assumed to be coherent in the same way, since beliefs and assumptions may differ. Secondly, profile structure facilitates evaluation of the reliability or credibility of several attributions together and in context. Moreover, different sources in a text can influence the evaluation of their respective credibilities: compare the status of *Police Officer XYZ* first to *the thief* (status: high reliability) and then to *disciplinary commission* (status:

neutral). Thus the basic frames representing a single utterance should be grouped by coreference resolution on the source into merged profiles. Different sources may be aligned for the sake of the argument in one article, as are *Sen. Packwood* and *Republicans* in Figure 1. This is called *supporting group structure* in (Bergler, 1992). The respective lexicalizations of the sources indicate in part the reliability or credibility of the primary information (from the point of view of the reporter).

4. Extended Example

As mentioned in the introduction, reported speech is not limited to newspaper articles and occurs also in other genres. This section demonstrates the usefulness of the proposed representations and processing strategies on an extended example from an online product review found on ConsumerSearch.



Due to a large amount of consumer mail complaining about the Kenmore Calypsos problems with lint, Consumer Reports ran a new test this year. Editors washed a load of white towels and black T-shirts to test four competing models for lint left behind on the clean wet clothing. Editors say they are considering testing all washers for lint in the future. The limited test here does not appear to have been a factor in Consumer Reports overall ratings. Some models, which performed poorly in the lint test, still top the ratings chart.

Weve received our own mail from consumers regarding washers, and most of it concerns reliability factors involving the latest top-loading models, such as the Kenmore Calypso. The Calypso is a top-loading machine with a unique agitating technology. Instead of twisting clothes around, it bounces them up and down and showers them with water. One owner wrote to us about a lengthy ordeal involving four repair visits. Repair issues with the Kenmore Calypso (also sold as the Whirlpool Calypso) are born out in over 150 postings in ThatHomeSite.coms appliance forum as well as on Epinions. Owners report clothing working its way below the basket (one user reports a handkerchief made it all the way to the sump pump), and others have problems with electronic error messages and clothing that comes out too wet or lint-covered.

Interestingly, theres a class-action lawsuit in the works regarding problems with the Calypso. Mark Tamblyn, a lawyer with Kershaw, Cutter, Ratinoff and York, LLP of Sacramento told us that the main issues appear to be with the control board and U-joint, along with water leakage and drainage problems (the first case and request for class-action status was filed in late June in Illinois state court). In the last version of our report, we featured the Kenmore/Whirlpool Calypso as a high-efficiency top-loader. In some tests, it still outperforms other models. However, given the number of consumer complaints we read on Epinions and ThatHomeSite.com, were not placing it in ConsumerSearch Fast Answers for this version of our report.

We contacted Stephen Duthie, Manager of Global Communications for Whirlpool. Duthie told us by e-mail that We have no knowledge about a suit, pending or otherwise, and no reason to believe there will be a suit.Duthie offered no further comments on the Calypsos repair record.

Figure 5. ConsumerSearch
http://www.consumersearch.com/www/house_and_home/washing_machines/ fullstory.html.

The abbreviated text in Figure 5 is a screenshot of Doandes' system, highlighting the reported speech occurrences it found. The reported speech clearly segments the text into topical areas. Here, the profile structure contains a much smaller proportion of the text than it did for the text in Figure 1. The automatic system missed two instances of reported speech in the last sentence of the second paragraph (introduced by the reporting verb *report*, which was not in the list of reporting verbs). More interestingly, the text contains an instance of explicitly reported material that does not follow the reported speech pattern in *In the last version of our report, we featured the Kenmore/Whirlpool Calypso as a high-efficiency top-loader.* These functionally equivalent constructions have to be represented and analyzed in the same manner and the prototype has extended the treatment already to *according to.*

We know from the genre of text that in fact it contains only secondary information, taken from other sources. The explicit reported speech sentences, however, appear to ground an entire paragraph. We see the reported speech here playing two distinct roles: in the first paragraph, rather than openly criticizing that the tests for lint had not been done for all washers, anonymous editors are reported as stating that *they are considering testing all washers for lint in the future.*

In the second paragraph, the reported speech (both the detected and the undetected instances) serves to ground a general point in a particular experience (*four repair visits*). In the third paragraph, detailed statements about problematic parts are attributed to a lawyer involved in preparing a class-action lawsuit. Both explicit attributions shift the (legal) responsibility from the editors to the cited sources, but at the same time serve to increase their credibility.

The basic profiles for this text are straightforward, but they do not fall into an interesting profile structure (each reported speech occurrence stands alone, not connected to the others in profile structure). The basic profiles are more interesting here as goalposts in the rhetorical and text structure: each reported speech occurrence has a privileged relationship to the rest of the information within the same paragraph. It is beyond the scope of this paper to show how to integrate text structure and profile structure. The role the profile structure should play in follow-up analysis of the text, however, can be sketched in isolation in the next section.

5. Source List Annotation

In a nutshell, (Bergler, 1992, 1995) advocates using the primary information for further processing (such as information extraction or summarization), but to keep an annotation attached that reports the evidential chain, or in this case, the list of sources and the other circumstantial information encoded in the matrix clause. This information can then be interpreted as needed in the context of the ultimate use of the information. This is a most important point: reporters use reported speech not because they couldn't make the necessary inferences themselves and write down the interpretation of what was said, but because this interpretation depends on the context and the 'user'. The text in Figure 5 indicates this very phenomenon: the ranking of a certain washing machine was lowered from the year before, because of information gained about repair issues and consumer complaints, not because of different test results.

The source list annotation is the product of a percolation process over the embedded structures formed by the successive attribution of the material. For instance, the basic profile of Figure 3 shows one level of embeddedness: if we call the primary information Ø = "*We don't have the votes for cloture today.*", then the basic profile encodes one level of embedding, C(Ø), where C stands for the circumstantial information provided regarding Ø. But the description of the

circumstantial information and the selection of the material has as its source the reporter R, in this case *Jeffrey H. Birnbaum,* leading to another level of embedding R(C(Ø)). And the newspaper where the reporter published the story has its own influence on style and credibility, leading to a third level of embedding, P(R(C(Ø))). And for an automated reasoner or agent, this would be embedded in the context of the agent's beliefs, A(P(R(C(Ø)))). This now properly represents that the agent A read in the paper P an article written by reporter R indicating that under the circumstances C (which include the source S) Ø was asserted. These levels of embedding have to be chosen for each agent, since some agent may not consider different newspapers differently and can thus drop one level of embedding. For agent A, Ø has been asserted under (P',R',C'), where P' is agent A's assessment of paper P's credibility, R' is the agent's assessment of the reporter's credibility, and C' is the agent's assessment of the source S's credibility combined with the pragmatic force of the reported speech verb and other circumstantial information. (P',R',C') is called the *source list* of Ø.

(Gerard, 2000) implemented a proof of concept system called *Percolator*, based on the process of percolation introduced by (Wilks and Ballim, 1991). Percolator assumes a particular agent's interpretation and constructs the representation of Figure 6 for the text of Figure 1.

System believes
  Reader believes
    Reporter Jeffrey H. Birnbaum believes
      Sen. Packwood said
        [``We don't have the votes for cloture today.''
          Source-list(Sen. Packwood, h, n, n, n)]
      Republicans said
        [Republicans can garner a majority in the 100-member Senate for a
          capital-gains tax cut.
          Source-list(Republicans, n, n, n, n)]
        [the Democrats are unfairly using Senate rules to erect a 60-vote hurdle.
          Source-list(Republicans, n, h, n, n)]
      Democrats said
        [the proposal, which also would create a new type of individual retirement
          account, was fraught with budget gimmickry that would lose billions of
          dollars in the long run.
          Source-list(Democrats, n, n, n, n)]

Figure 6. Percolator's representation of profile structure for Figure 1 with source lists.

In this representation, each reported speech complement is indexed by its *Source-list*. The Source-list (which is a particular implementation of the more general concept of a source list elaborated above) holds the description of the source (as given in the subject noun phrase of the matrix clause) and four evaluation features, which represent in turn: the reporter's confidence in the source(s) as expressed by the lexicalization of the matrix clause's subject NP, the reporters confidence in the reported information as expressed by the lexicalization of the reporting verb, the reader's confidence in the source (potentially from prior knowledge or beliefs), and the reader's confidence in the reported information (again, potentially from previous knowledge or beliefs). Important here is the translation of a strength feature encoded in the lexical entries of reporting verbs (Bergler, 1995b) into a credibility rating of the reported information: *acknowledge* as a reporting verb here carries an implication that the information of the complement clause is negative for the subject. If a source is reported to *acknowledge* information this has to be rated as information of high reliability (possible values for the prototype were **h**igh, **n**eutral, and **l**ow), because sources will not falsely make detrimental statements.

The Source-list pairs the reporter's apparent evaluation of the source and the reported information from lexical semantics with the reader's evaluation of the source and the reported information. This reflects a reader's ability to immediately discount the reporter's apparent evaluation based on previous knowledge or on previous beliefs (about the reporter, the source, or the information). But a reader with no relevant previous beliefs has to rely solely on the intrinsic evaluation of the reporter.

For any specific agent, previous beliefs about the reporter, the source, or the topic are already encoded (or absent, as in Figure 7). Gerard assumes one such agent's strategy to further process the Source-list annotated representation in Figure 6 into the belief-space annotation of Figure 7, loosely modelled on the nested belief environments in (Wilks and Ballim, 1991). Figure 7 illustrates an evolved source list in a flattened-out embedding structure. By *percolating* out of the embedding for the source and the reporter (i.e. deleting it) the lost information is transformed in additional information in the evolving source list, which has now an additional source, namely the reporter. Gerard showed on several examples that an extension of Wilks' and Ballim's percolation mechanism allows to properly combine subjective evaluations of each level of attribution.

Note that Figure 7 does not model a traditional belief space, because there is not enough evidence to transform this information into a belief. Rather, it introduces the notion of a *potential* belief, defined as information that might or might not turn into a *held* belief given further evidence. This reflects a reader's ability to accommodate contradictory information: having read an article that presents two contradictory theories, one can argue both ways if one has no own opinion on the matter until there is evidence that "settles" the issue. This representational device permits to delay the decision as to whether information is believed until a certain threshold of comfort is reached. (Gerard, 2000) explores ideas on when to transform potential beliefs into held beliefs.

System believes
  Reader believes
    NIL
  Reader potentially believes
    ["We don't have the votes for cloture today."
     Source-list(Sen. Packwood, Reporter Jeffrey H. Birnbaum, h, n, n, n)]
    [Republicans can garner a majority in the 100-member Senate for a capital-gains tax cut.
     Source-list(Republicans, Reporter Jeffrey H. Birnbaum, n, n, n, n)]
    [the Democrats are unfairly using Senate rules to erect a 60-vote hurdle.
     Source-list(Republicans, Reporter Jeffrey H. Birnbaum, n, h, n, n)]
    [the proposal, which also would create a new type of individual retirement account, was fraught with budget gimmickry that would lose billions of dollars in the long run.
     Source-list(Democrats, Reporter Jeffrey H. Birnbaum, n, n, n, n)]

Figure 7. Potential belief spaces for the text of Figure 1 for an agent with no prior knowledge or beliefs.

Note that this work has important differences with traditional work on *belief* reports. Utility texts such as newspaper articles expressly avoid using belief reports because they represent an evaluation by the reporter which the reader might not share. Instead, newspaper articles offer *evidence* reports which only give rise to beliefs through an additional interpretation step. (Gerard, 2000) presents one possible way to transform these evidence reports into first potential, and eventually held beliefs. The focus of this paper is on the extraction of a proper representation of

the evidence reports that will enable the intricate reasoning about beliefs discussed in the literature (see for instance (Rapaport et al., 1997, Rapaport, 1986)).

6. Extension to Other Attribution

Reported speech is an important issue in its own right, but it can moreover serve as a model for attribution in different registers, as well. Reported speech is a culturally determined vehicle for evidential annotation of hearsay. It requires at least a description of the source and the minimal pragmatic characterization of the speech act (encoded in the reporting verb). This then is a minimal requirement for all attribution in this cultural context and we can see it borne out in email and chat groups, where the sender is always identified (at least by their email aliases) and the "speech act" is implicit in the vehicle (email). The importance of this authentication vehicle for information is also felt when one hits on a Web page with interesting content but no hint or link as to its author and purpose: this is a highly unsettling situation and may well lead to dismissing the page for further use. It incurs the same stigma as anonymous messages by phone or traditional mail.

Like reported speech, information of any kind should have a source list annotation detailing the path it took from the original source to the current user, in addition to other circumstantial information that impacts its interpretation. Moreover, it is convenient, especially when compiling many sources into a comprehensive overview over different opinions (like in product reviews), to compile structures akin to profile structures, that group sources with similar points of view into larger supporting groups, for easier structuring of the information and coherence of interpretation. Because information may be easier to interpret in the pertinent context, profile structures restore this original context. Finally, a reasoner that has to make sense of many differing accounts must have a reasoning zone like potential beliefs, where contradictory information is stored together with its authenticating information (akin to the source list) until enough evidence is accumulated to adopt one side as the predominant one, which can be transformed into actually held beliefs.

Reported speech as the most widely used example of evidential coding can thus serve as the guiding model for the treatment of second hand information in general.

7. Conclusion

Attribution is a phenomenon of great interest and a principled treatment is important beyond the realm of newspaper articles. The way natural language has evolved to reflect our understanding of attribution in the form of reported speech can guide investigations into representations forming the basis for shallow text mining as well as belief revision or maintenance systems.

This paper has shown the importance of a two step process in interpreting reported speech and other explicit attributions: interpreting first the shallow semantics of reported speech, then interpreting it in the pragmatic and semantic context. The first step can be achieved with acceptable performance using shallow techniques as has been demonstrated in a proof of concept system. The importance of analyzing the profile structure even for shallow further analysis lies in the contexts it constructs: interpretation of material in different profiles may differ (this point has also been raised by Polanyi and Zaenen (this volume) discussing reported speech, among others, as a conceptual valence shifter).

The second step is largely dependent on the particular use for the system and can thus not be described in general. Yet, the general principle of a potential belief space that can accommodate conflicting information without disabling the underlying reasoning system should be part of any particular system that deals with attitudes and affect in text.

8. Acknowledgements

Monia Doandes' help is gratefully acknowledged, as are the valuable comments from Jan Wiebe and anonymous reviewers. This work was funded in part by the National Science and Engineering Research Council of Canada and the Fonds nature et technologies of Quebec.

9. Bibliography

Anick, P. and Bergler, S. (1992) Lexical Structures for Linguistic Inference, In *Pustejovsky, J. and Bergler, S. (Eds.), Lexical Semantics and Knowledge Representation,* Springer, Berlin.

Bergler, S. (1992) *The Evidential Analysis of Reported Speech.* PhD Dissertation, Brandeis University, Massachusetts.

Bergler, S. (1995a) From Lexical Semantics to Text Analysis, in Saint-Dizier, P. and Viegas, E. (Eds.), *Computational Lexical Semantics*, Cambridge University Press, Cambridge.

Bergler, S. (1995b) Generative Lexicon Principles for Machine Translation: A Case for Meta-lexical Structure, *Journal of Machine Translation 9(3)*.

Clark, H. and Gerrig, R. (1990) Quotations as Demonstrations. *Language 66(4),* 764-805.

Cunningham, H. (2002) Gate, a General Architecture for Text Engineering. *Computers and the Humanities, 36*, 223-254.

Doandes, M. (2003) *Profiling for Belief Acquisition from Reported Speech.* Master's thesis, Concordia University. Montreal.

DUC (2003) Document Understanding Conference. NIST. http://www-nlpir.nist.gov/projects/duc/index.html

Fellbaum, C. (Ed.) (1998) *WordNet: An Electronic Lexical Database*. MIT Press.

Gerard, C. (2000) *Modelling Readers of News Articles Using Nested Beliefs*. Master's thesis, Concordia University. Montreal.

Hepple, M. (2000) Independence and Commitment: Assumptions for Rapid Training and Execution of Rule-based Part-of-Speech Taggers. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL-2000)*. 278-285. Hong Kong.

Marcu, D. (1997) The Rhetorical Parsing of Natural Language Texts. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the Association for Computational Linguistics*. 96-103. Madrid.

Mithun, M. (1999) *The Languages of Native North America*. Cambridge University Press, Cambridge.

Polanyi, L., Culy, C., van den Berg, M., Thione, G.L., and Ahn, D. (2004) A Rule Based Approach to Discourse Parsing. In *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*. Cambridge.

Polanyi, L. and Zaenen, A. (2004) Contextual Valence Shifters. In Qu, Y., Shanahan, J.G., Wiebe, J. (Eds.) *Exploring Attitude and Affect in Text: Theories and Applications AAAI-EAAT 2004*, AAAI Press Report SS-04-07.

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985) *A Comprehensive Grammar of the English Language*. Longman, London.

Rapaport, W. J. (1986) Logical Foundations for Belief Representation, *Cognitive Science 10*. 371-422.

Rapaport, W.J., Shapiro, S.C., and Wiebe, J. (1997) Quasi-Indexicals and Knowledge Reports, *Cognitive Science 21(1)*. 63-107.

Wiebe, J., Bruce, R., and Duan, L. (1997) Probabilistic Event Categorization, In *Recent Advances in Natural Language Processing (RANLP-97).* 163-170. Tsigov Chark.

Wilks, Y. and Ballim, A. (1991) *Artificial Believers*. Erlbaum, Norwood.

Witte, R., and Bergler, S. (2003) Fuzzy Coreference Resolution for Summarization. In *Proceedings of the International Symposium on Reference Resolution and Its Applications on Question Answering and Summarization (ARQAS 2003)*. 43-50. Venice.