Veredas: Frame Semantics and Its Technological Applications
Vol. 17, nº 1, 2013

Extending FrameNet for Sentiment Analysis

Josef Ruppenhofer (Hildesheim University)

ABSTRACT: We provide a contribution that points up the potential use of frame semantics and FrameNet in particular for sentiment analysis. We address several key problems in current sentiment analysis, which is characterized by shallow approaches, pragmatic focus, and ad-hoc creation of data sets and methods. We argue that progress towards deep analysis depends on a) enriching shallow representations with linguistically motivated, rich information, and b) focusing different branches of research and combining resources and work forces to join hands with related work in NLP. We propose SentiFrameNet, an extension to FrameNet, as a novel representation for sentiment analysis that is tailored to these aims.

KEYWORDS: sentiment analysis; FrameNet representation; scale structures

Introduction

Automatic sentiment analysis is a burgeoning field of computational linguistics and natural language processing, with interesting challenges and many areas of application. A lot of effort has been put into building up resources such as sentiment/polarity lexicons and evaluation data sets, developing and honing algorithms, and addressing ever new types of texts. The research carried out so far has made a lot of progress on more coarse-grained analysis levels using shallow techniques. However, recent years have seen a trend towards more fine-grained and ambitious analyses requiring more linguistic knowledge and more complex statistical models. While, for instance, one of the most often cited "early" works focused on the binary classification of movie reviews (PANG ET AL., 2002), recent work has tried to produce relatively detailed summaries of opinions expressed in news texts (STOYANOV & CARDIE, 2011); to assess the impact of quotations from business leaders on stock prices (DRURY ET AL., 2011); to detect implicit sentiment (BALAHUR ET AL., 2011); etc. Accordingly, we can expect that greater demands will be made on the amount of linguistic knowledge, its representation, and the evaluation of systems.

Current approaches to sentiment analysis mostly use shallow representations. Efforts to use more complex knowledge often require building it up by heuristically and imperfectly aligning multiple resources. Many researchers use task-specific, non-public and/or non-reusable resources. In terms of evaluation, special data sets are often created for the task at hand, which means that it is not clear how to interpret the results and how to compare them with related work. More general evaluation data sets may not be available or deeply enough analyzed. Finally, emphasis on pragmatic analysis and text-level relevance in annotation and system-building obscures the differences between cases in which sentiment can be readily extracted by lexical and syntactic means and cases for which inferential knowledge is needed.

Against this background, we revisit the question how the problem of sentiment analysis should be posed in the first place, and what linguistic knowledge is needed and how it can be represented. In particular, we contrast what we may call shallow and pragmatic approaches with a deep, lexical-semantics based one, arguing that the latter perspective is a very useful complement to the former in enabling deeper analysis.

Our main contribution is the proposal of SentiFrameNet, an extension of FrameNet (BAKER ET AL., 1998) offering a novel representation for sentiment analysis based on frame semantics. Our representation is linguistically sound, adequately encodes aspects of meaning needed for a deep analysis of sentiment, and makes use of existing resources. We spell out the new information we make available, show how our approach can mitigate sparse data problems, and how it avoids shortcomings of existing data sets used in sentiment evaluation. In section 1, we discuss what we mean by shallow and pragmatic approaches to sentiment analysis. In section 2, we propose a view on the task that is driven by lexicon and syntax, discussing in particular how our representational desiderata for deep analysis are met in the framework of frame semantics. We illustrate our approach to the automatic acquisition of some of the new information in section 3. In section 4, we discuss the impact of our approach before concluding.

Before we continue, we would like to briefly address the somewhat confusing terminology in the field of sentiment analysis as it relates to our purposes. We choose to use the term sentiment analysis here as a cover term for all kinds of what Wiebe usually calls subjective language, including expressions of emotion, epistemic stance, evaluation and so forth. We do not want to imply that differentiation within a frame semantic analysis is not desirable, just that it is not within the scope of the present work to pursue. In fact, we expect modifications to our work to be necessary as the analysis of e.g. emotion and evaluative language is refined.

1. Shallow and pragmatic approaches

Annotation schemes and analysis systems for expression-level sentiment analysis commonly assume that the following components of opinions need to be extracted and interrelated:

i. an expression of opinion
ii. its polarity in context
iii. its strength in context
iv. its (possibly nested) source or holder
v. its target

Although the task and the units of analysis seem intuitive enough, actual approaches to the task differ considerably. The most basic question, of course, is what is to count as a sentiment-bearing expression. One consideration here is simply formal, namely how one determines the boundaries of sentiment expressions. For instance, the annotations in the MPQA corpus (WIEBE ET AL., 2005), which has frequently been used as an evaluation benchmark, were created without giving annotators any lexical or syntactic constraints on what to annotate. While this serves the spirit of discovering the variety of opinion expressions, it makes it difficult to match opinion expressions when using the corpus as an evaluation dataset as the same or similar structures may be treated differently. Another challenge lies in distinguishing so-called polar facts from genuinely sentiment-bearing expressions. For example, out of context, one would not associate any of the words in ((1)) with a particular evaluative meaning. In specific contexts, however, we may understand example ((1)) as reason to either think positively or negatively of Switzerland: employees receiving wages may be drawn to Switzerland based on ((1)), while employers paying wages may view this state of affairs negatively.

(1) Wages are **high** in Switzerland.

As shown by the inter-annotator agreement results reported by (TOPRAK ET AL., 2010), agreement on distinguishing polar facts from inherently evaluative language is low.

Unsurprisingly, many efforts at automatically building up sentiment lexicons simply harvest expressions that frequently occur as part of polar facts without resolving whether the subjectivity clues extracted are inherently evaluative or merely associated with statements of polar fact. There has, for instance, been some research within NLP on degree adjectives such as *high* and *low* (occurring in phrases like *high/low cost, high/low income*), which treated these adjectives as merely a special subtype of sentiment clue (WU & WEN, 2010).

Some work also excludes certain expressions of sentiment or opinion from analysis. For (WIEBE ET AL., 2005) appraisals by experts are objective, unlike ones by laymen. Seki (2007) annotated sentences as "not opinionated" if they contain indirect hearsay evidence or opinions held by the general public. Most lexical resources do not adequately represent cases where multiple opinions are tied to one expression and where presuppositions and temporal structure come into play. An example is the verb *despoil*: there is a positive opinion by the reporter about the despoiled entity in its former state, a negative opinion about its present state, and (inferrable) negative sentiment towards the despoiler. In most resources, the positive opinion will not be represented.

In terms of how the task is approached, most efforts use an information extraction-like pipeline approach. Expressions of opinion, Sources and Targets are often dealt with separately. Some work such as Kim and Hovy (2006) has explored the connection to role labeling. One reason not to pursue this, articulated by Wiegand (2010, p.121) in work on opinion holder extraction, is that "in many practical situations, the annotation beyond opinion holder labeling is too expensive". Another consideration may be that many predicates – or expressive subjective elements, according to Wiebe et al. (2005) – cannot express their Source as a syntactic dependent. For these, Sources need to be found outside the predicate's syntactic scope. In the case of Targets, the work by Stoyanov and Cardie (2008) shows what we may call a pragmatic focus. These authors suggest a definition of opinion topic – their term for the Source's object of opinion – and present an algorithm for opinion topic identification that casts the task as a problem in topic-coreference resolution. They distinguish between (a) the *topic* of a fine-grained opinion, defined as the real-world object, event or abstract entity that is the subject of the opinion as intended by the opinion holder; (b) the *topic span* associated with an opinion expression, defined as the closest, minimal span of text that mentions the topic; and (c) the *target span*, defined as the span of text that covers the syntactic surface form comprising the contents of the opinion. As the definitions show, Stoyanov and Cardie (2008) focus on text-level, pragmatic relevance by paying attention to what the author intends, rather than concentrating on the explicit syntactic dependent (their target span) as the topic. This pragmatic focus is also in evidence in Wilson's (2008) work on contextual polarity classification, which uses features in the classification that are syntactically independent of the opinion expression such as the number of subjectivity clues in adjoining sentences.

Altogether, current sentiment analysis typically uses shallow representations and focuses on the pragmatic understanding in context, without, however, circumscribing or tracing in detail how the contextualized pragmatic understanding builds up and where it differs from the lexical semantics of the items involved. A disadvantage of this is the lack of a basis for adequately modeling different types of sentiment. As shown above, current approaches are not suitable for encoding multiple opinions, presuppositions and temporal structure, all of which is needed for a detailed, deep analysis of sentiment in text. As a second major drawback we see the impact on evaluation, an issue we will come back to in section 3.

2. The extended frame-semantic approach

In what follows we offer a view of sentiment analysis that performs sentiment analysis on the basis of a frame semantic analysis, using an appropriately extended model of frame semantic representation. The extended frame semantic representation is at the heart of the SentiFrameNet extension to FrameNet that we are working towards.[1]

2.1. Link to semantic frames and roles

Since the possible Sources and Targets of opinion are usually identical to a predicate's semantic roles, as argued by Ruppenhofer et al. (2008), we connect opinion roles such as Source and Target to semantic roles, thereby enabling the use of semantic role labeling systems in the identification of opinion roles. From our point of view, a connection to FrameNet is of particular interest as the FrameNet has been designed to handle semantic frames and their frame elements (FEs).

FrameNet provides for a binary annotation of lexical units (LUs) with a negative or positive semantic type. In SentiFrameNet, we spell out which semantic role is the Source of opinion (if any) and which is the Target. This involves some major changes to FrameNet. First, since sentiment information pertains to specific LUs rather than to whole frames, either the frames have to be split up more finely until all LUs are sentiment-consistent. Alternatively, and this is the path we take here, each lexical unit needs be put into a newly created minimal frame which inherits the frame that the LU currently belongs to. For instance, the verb *denigrate* in the *Judgment communication* frame carries a negative (external) judgment of the Speaker for the unfairness of their criticism. The same external judgment is lacking with the verb *criticize*. The second step is to associate with the newly introduced specialized frames a set of opinion frames with Source and Target roles that can be mapped to the semantic roles by the same mechanisms used to relate roles of different "content" frames.

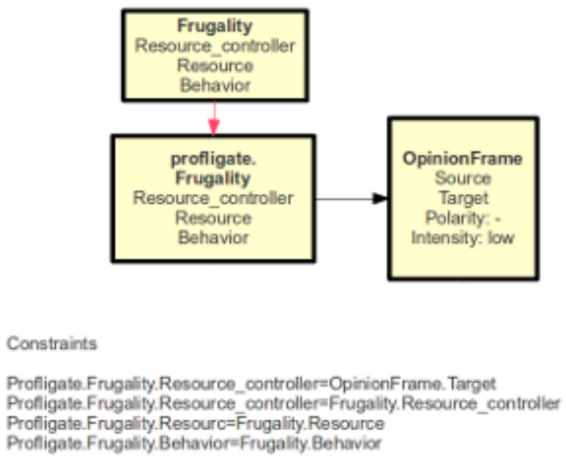

Figure 1: Analysis of the adjective *profligate*

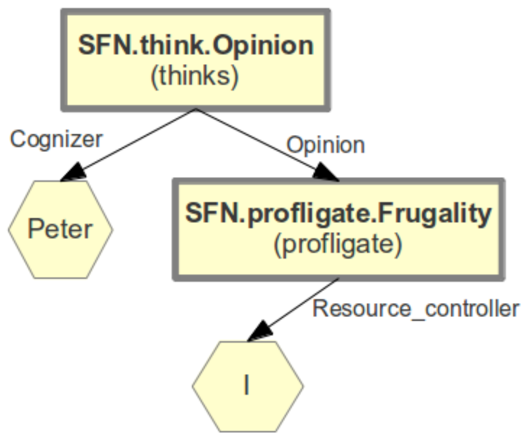

Figure 2: Frame embeddings for *Peter thinks I am profligate*

[1] We would like note that we understand our project as one of explicating lexical knowledge that has always been relevant to the theory of frame semantics, even if not explicitly represented within FrameNet.

In Figure 1, an analysis using an LU-specific frame is given for the adjective *profligate*. A special LU-specific frame is introduced and related by an inheritance relation (indicated by the red arrow) to the more general *Frugality* frame. Note among the constraints, which capture frame-frame and frame element-frame element relations, that the opinion Source is not bound to any of the semantic roles of the frame, indicating that it is left to be bound to a Source of another frame 'higher up' that embeds the *Frugality* frame. (The positively connoted adjective frugal would be treated very similarly, except that the polarity of its associated opinion frame would be positive.)

With *profligate*, to find the Source of opinion towards the Target, we have to look if the frame of the opinion expression is embedded under a frame with a Source. E.g. in *Peter thinks I am profligate*, we guess that Peter is the Source of the opinion expressed by *profligate*, given that the *Frugality* frame of *profligate* is embedded in the OPINION frame element of the *Opinion* frame evoked by *think*, as shown in Figure 2.

In SentiFrameNet all inherently evaluative LUs are associated with opinion frames. When instances of such lexical units are annotated in a text, the associated opinion frames are projected onto the text, too. Determining whether, as in Wiebe et al.'s (2005) work, expert appraisals should be treated differently from lay-people's appraisals is left to later analysis steps. The language of polar facts is not associated with opinion frames. However, we show in section 2.6 how we can support certain types of inferred sentiment. With regard to Targets, our representation selects as Targets of opinion the target spans of Stoyanov and Cardie (2008) rather than their opinion topics. For us, opinion topics that do not coincide with target spans are inferential opinion Targets.

Note that while the idea of having lexical unit-specific frames is somewhat unusual relative to FrameNet's current practice, it should not strike one as completely surprising. First, one might see it as just a further practical step in the direction of greater specificity given that FrameNet has made the same kind of move when it changed from frames that encompassed whole semantic domains / lexical fields to smaller frames that are more consistent in terms of FE expression, paraphrasability, etc. Second, there is, as pointed out e.g. by Busse (2012), nothing in frame semantic theory that would require larger groupings. And finally, as long as the network of frame relations is extended in parallel to the creation of specialized frames, larger groupings with the kind of granularity found in FrameNet at present can always be virtually re-constituted if desired.

2.2. Formal diversity of opinion expressions

For fine-grained sentiment-analysis, handling the full variety of opinion expressions is indispensable. While adjectives in particular have often been found to be very useful cues for automatic sentiment analysis (WIEBE, 2000; BENAMARA ET AL., 2007), evaluative meaning pervades all major lexical classes, as shown by examples such as *admire.v*, *moron.n*, and *regrettably.adv*. There are also many subjective multi-words. The example *give away the store* in ((2)) illustrates nicely that the subjective status and polarity of a multi-word are not necessarily derivable from its constituent parts.

(2) Elected officials **give away the store** to corporations in the form of tax subsidies.

Evaluative meaning also attaches to some grammatical constructions, even ones without obligatory lexical material. An example is the construction exemplified by *Him be a doctor*? The so-called *What, me worry?*-construction (FILLMORE, 1989) involves no particular lexical material, consisting only of a topic NP in objective case and an infinitive phrase that predicates of the topic NP. The rhetorical effect of the construction is to express the speaker's surprise or incredulity about the proposition under consideration. The FrameNet database schema accommodates not only single and multi-words but also handles data for a constructicon (FILLMORE ET AL., 2012) that pairs grammatical constructions with meanings. In SentiFrameNet, constructions are also connected to opinion frames as needed.

2.3. Multiple opinions

We need to accommodate multiple opinions relating to the same predicate as in the case of *despoil* mentioned above. Predicates of this kind are not all that uncommon: in a 100-item random sample taken from the Pittsburgh subjectivity clues, 17 involved multiple opinions. In representing multiple opinions, we draw inspiration from Maks and Vossen (2011). In their treatment of sentiment associated with verbs in a Dutch lexical resource, these authors simply register for all ordered pairs of (suitable) semantic roles if, and what kind of, sentiment exists between the roles in the pair. Additionally, they also introduce pairs for the external reporter of the event and all other semantic roles. Assuming the use of opinion frames as suggested in section 3.1, multiple opinions can be readily represented. Consider the verb *brag* in example (3).

(3) Peter keeps **bragging** about his car.

The verb *brag* in the modified *Bragging* frame has two opinion frames. As shown in *Figure* , the first one has positive polarity and represents the frame-internal point of view. The SPEAKER (Peter) is the Source relative to the TOPIC (car) as the Target. The second opinion frame has negative polarity, representing the reporter's point of view. The SPEAKER (Peter) is the Target but the Source is unspecified, indicating that it needs to be resolved to an embedded Source, namely the person whose report sentence (3) is.

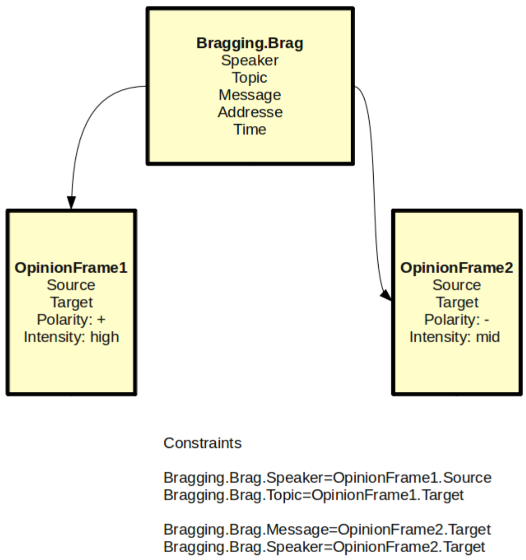

Figure 3: Multiple opinion frames associated with *brag.v*

2.4. Event structure and presuppositions

A complete representation of subjectivity needs to include event and presuppositional structure. This is necessary, for instance, for predicates like *come around (on)* in ((4)), which involve changes of opinion relative to the same Target by the same Source. Without the possibility of distinguishing between attitudes held at different times, the sentiment associated with these predicates cannot be modeled adequately – the proposal by Maks and Vossen (2011) does not take such cases into account. Along similar lines, modeling event structure is also necessary for cases where different Targets of opinion are valued differently and at different times. An example is the verb *mourn*, as shown in ((5)). It refers to an established ongoing positive feeling towards an entity and a negative feeling about having lost access to it.

(4) Newsom is still against extending weekday metering to evenings, but has **come around** on Sunday enforcement.

(5) Family and friends are **mourning** the death of a popular County Councillor.

For *come around (on)*, we want to to distinguish its semantics from that of predicates such as *ambivalent* and *conflicted*, where a COGNIZER simultaneously holds opposing valuations of (aspects of) a Target. Following FrameNet's practice, we model presupposed knowledge explicitly in SentiFrameNet by using additional frames and frame relations. A partial analysis of *come around* is sketched in *Figure* .

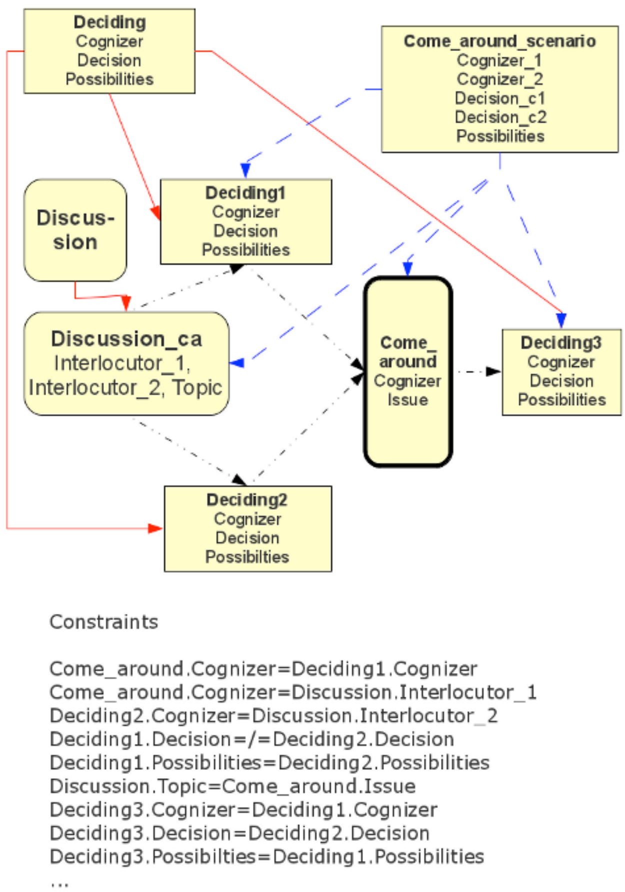

Figure 4: *Come around scenario*

We use the newly added *Come around scenario* frame as a background frame that ties together all the information we have about instances of coming around. Indicated by the blue dashed lines are the sub-frames of the scenario. Among them are three instances of the *Deciding* frame (tied by solid red lines to the general *Deciding* frame), all related temporally (black dashed-dotted lines) and in terms of content to an ongoing *Discussion*. The initial difference of opinion is encoded by the fact that Deciding1 and Deciding2 share the same POSSIBILITIES but differ in the filler of the DECISION FE. The occurrence of *Come around* leads to Deciding3, which has the same COGNIZER as Deciding1 but its DECISION is now identical to that in Deciding2, which has been unchanged. The sentiment information we need is encoded by simply stating that there is a sentiment of positive polarity of the COGNIZER (as Source) towards the DECISION (as Target) in the general *Deciding* frame – this opinion frame is not displayed in the graphic. The *Come around* frame itself is not associated with sentiment information, which seems right given that it does not include a DECISION as a frame element but only includes the ISSUE. Factuality presuppositions are relevant to sentiment analysis, too. For instance, in ((6)), the verb *know* presupposes that the speaker reporting on John' state of awareness also shares the belief (and the evaluation) that the move was a mistake. This is lost on a system that is not aware of presuppositions.

(6) John **knows** that the move was a mistake.

A model for capturing such presuppositions exists in Saurí's (2008) work on event factuality, which concerns the factual status of eventualities mentioned in text, that is, whether states and events are presented as facts, possibilities, or situations that do not hold at all. Saurí's (2008) representation of factuality keeps track of two types of sources, her *anchor* corresponds to what we have called reporter, and her *cognizer* to a frame-internal role such as COGNIZER or EXPERIENCER in frames that concern the relation of such a role to some mental content, proposition, or modalized event. In addition, the representation registers the level of commitment of the two sources towards the event. *Figure* shows how factuality is recorded for the predicate know in the *Awareness* frame, using SentiFrameNet's new construct of a factuality frame.

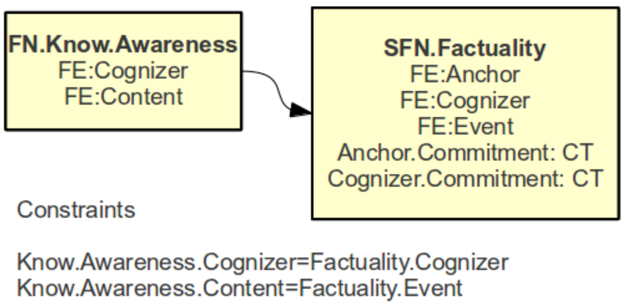

Figure 5: Factuality analysis for *know.v*

2.5. Modulation, coercion and composition

A key aspect in analyzing subjectivity is that speakers can modulate simple expressions: they can shift the valence or polarity of sentiment-bearing expressions through some kind of negation operator, or intensify or attenuate the impact of an expression ((7)).

(7) I [doubt$_{\text{Negation}}$] they're **great**, but I bet they're [quite$_{\text{Intensification}}$] **enjoyable**.

Despite this, it is desirable to have at least a partial ordering among predicates related to the same semantic scale; we want to be able to find out from our resource that *good* is less positive than *excellent*, while there may be no ordering between *terrific* and *excellent*. In SentiFrameNet, an ordering between the polarity strength values of different lexical units is added on the level of frames. This ordering specifies the particular content/opinion frame mappings involved, since multi-level predicates may co-occur in the same frame with predicates with only one level of subjectivity. We provide some discussion on methods to derive scalar orderings from corpus data in section 3.

The frame semantic approach also offers new perspectives on sentiment composition. One benefit is that we can recognize cases of presupposed sentiment, as in the case of the noun revenge, which are not amenable to shifting by negation: *She did not take revenge* does not imply that there is no negative evaluation of some injury inflicted by an offender. By always tracking Sources and Targets we can also readily distinguish cases of mixed sentiment by one Source (*You're a good liar*) from cases of differing opinions of two Sources, as in *Sue loves that idiot*, where *idiot* reflects the view of the reporter, not *Sue*. Further, many cases of what has been called valence shifting for us are cases where the evaluation is wholly contained in a predicate.

(8) Just barely **avoided** an accident today.
(9) I had served the bank for 22 years and had **avoided** a promotion since I feared that I would be transferred out of Chennai city.

If we took the view that *avoid* is a polarity shifter and further treated nouns like promotion and accident as sentiment-bearing (rather than treating them as denoting events that affect somebody positively or negatively) we should expect to find that while ((8)) has positive sentiment, ((9)) has negative sentiment. But the latter is not true: accomplished intentional avoiding is always positive for the avoider. Also, the reversal analysis for *avoid* does not know what to do with complements that have no inherent polarity. It readily follows from the coercion analysis that *I avoid running into her* is negative but that cannot be derived in Moilanen and Pulman's (2007) compositional model which takes into account inherent lexical polarity, which *run (into)* lacks. The fact that *avoid* imposes a negative evaluation by its subject on its object can easily be modeled using the idea of opinion frames presented in section 2.1 above. Another interesting case is the degree modifier *too*. On Moilanen and Pulman's (2007) analysis, *too* is said to be able to reverse only positive heads. The treatment as a shifter is problematic, however. Basically, there are cases where *too* combines with negative predicates but the overall interpretation is positive ((10)) and cases where *too* does not seem to reverse positive predicates ((11)).

(10) A: Will they be able to steal our piano while we're gone?
B: No, it's **too** heavy.
(11) A: Will John fail?
B: No, he is **too** smart.

We, therefore, do not treat *too* as a shifter and consider its interaction with sentiment-bearing predicates a matter of complex pragmatic inference.

2.6. Polar facts and implicit sentiment

A key challenge in analyzing subjectivity consists in differentiating between the lexical denotation of predicates and speakers' world knowledge about the referents and situations involved. Recall example ((1)) above. In the case of wages, there is a large difference between the number of readers having one association and those having the opposite: most people are employees rather than self-employed or employers. This kind of asymmetry seems to make polar facts difficult to distinguish from evaluation proper, and to make them of interest in certain application settings. In product review mining, for instance, where reviewers and their readers share the same values and interests as actual and potential buyers, one can reasonably assume that polar facts such as *The laptop case is made of cheap plastic* are meant to express and to invite a negative appraisal. A conscious step in the direction of performing sentiment analysis that aims at detecting implicit sentiment is taken by Balahur et al. (2011). These authors propose an approach towards automatically detecting emotions as underlying components of sentiment from contexts in which no clues of sentiment appear, based on commonsense knowledge. The system is not only meant to recognize polar facts – statements that express appraisal of an entity indirectly by asserting some property – but to recognize the emotional state of event participants. The system should, for instance, know that going to a party is something that produces joy. The resource that the authors are assembling – EmotiNet – is intended to be a knowledge base of concepts with their associated affective value. The goal is to model situations as chains of actions and their corresponding emotional effect using an ontological representation. We adopt a more limited approach to implied sentiment that makes use of lexically specified positive or negative impacts on some of the participants, that is, it uses information about affectedness and causality. Consider the following example discussed in Ruppenhofer et al. (2008).

(12) I am not a Colts fan I am a Bears fan but I am **glad** [the Colts beat the Patriots].

In example ((12)), the stable basic sentiment conveyed by the sentence is that the speaker is glad about the reported event. The understanding that the speaker dislikes the Patriots is an inference from our point of view. Recognizing such inferences requires knowing for as many predicates as possible which of their arguments are negatively or positively affected by an event and which participants are causally responsible. Of course, we also need to assess what contextual support there is for the possible inferences. In ((12)), for instance, the speaker blocks the possible inference of positive sentiment towards the Colts by pointing out *I am not a Colts fan*. Information about affectedness and causal responsibility can be added semi-automatically to frames using heuristics. Roles mapping to subject are good candidates for being causally responsible participants, assuming that the frame as a whole involves causation. Frame names starting with "Cause to" and their descendants by Inheritance tend to be causative, and further frames containing FEs named THEME or PATIENT etc are also good candidates for being causative and involving affected participants, respectively. The frame and frame element relations within the database can be used to propagate classifications from seed FEs in certain frames to other FEs in the same or other frames. We can also apply Greene and Resnik's (2009) strategy of extracting dependency relations from the annotated data as proxy information about volition, causality, and affectedness and use it to supplement or correct the conclusions we come to on the basis of the frame hierarchy. We also have to allow for multiple frame elements in a frame to be marked positively as affected or causally responsible. For instance, in the *Killing* frame, both the KILLER and the CAUSE frame elements can be marked as causally responsible.

3. Automatically acquiring scalar information

In order to speed up the effort of adding information about affectedness; scale structures; FE to opinion role-linkings etc to FrameNet, we would like to develop methods that can initialize the relevant features in an extended FrameNet version based on the information already in the FrameNet database and based on corpus evidence. The first steps in this direction that we have been undertaking concern the addition of scalar information to FrameNet. Specifically, we have been experimenting with ordering adjectives in the *Mental property* and *Dimension* frames. For instance, how can one find out from corpus data that the lack of intelligence and the corresponding negative judgment conveyed by *moronic* is greater than that conveyed by *stupid*?

Various researchers have worked out ways to automatically assign valence scores to evaluative predicates. One prominent approach uses the relations of a predicate within a lexical resource as evidence for its valence (e.g. BACCIANELLA ET AL., 2010). Another considers the distribution of gradable predicates in product or service reviews that come with associated star ratings (e.g. RILL, 2012). Both methods tend to produce distinct scores, and thereby absolute orderings, for all predicates, even though it is far from obvious that speakers would agree on complete orders between predicates such as e.g. *dumb*, *foolish*, *stupid*.

Our corpus-based work builds on a range of linguistic work from Bolinger (1972) to Kennedy and McNally (2005), using distributional facts to assign scale structures and establish partial orders for gradable predicates. We do not currently aim at establishing absolute orderings between predicates but only relative ones. Our first approach uses simple co-occurrence frequencies of the relevant adjectives with end-of-scale and non-end-of-scale modifiers. The former group includes adverbs such as *utterly, completely, partially* while the latter includes adverbs such as *quite, rather, very*. According to Kennedy and McNally's theory of scale structure, the end-of scale modifiers should co-occur only with adjectives that have a closed scale or a semi-closed scale with a lower or upper bound, while the other adverbial modifiers should co-occur with adjectives that have an open scale. Being able to partition our sets of adjectives based on their co-occurrence with the various modifiers would be a first step in ordering them. While Kennedy and McNally intended to make a categorical distinction, the evidence from the UKWAC and BNC corpora is less clear cut: one does find unexpected combinations. Our decision criterion for the raw frequencies was a simple one: the adjective would be classified depending on the kind of modifier it occurred with more often.

Our second approach to classifying adjectives as having an open or closed scale uses a statistical test: for all our adjectives in a given frame, we measure their association with normal and end-of-scale modification, performing a multiple distinctive collexeme analysis as suggested by Gries and Stefanowitsch (2004). The idea is basically that rather than doing chi-square tests for lexical material, we perform Fisher's exact tests on the association between lexical material – the collexemes – and constructions. The constructions in our case are normal and end-of-scale degree modification and the collexemes are the adjectives that we want to order. Having performed the analysis, we end up with orderings that indicate which adjectives prefer one or the other construction and how strongly; adjectives could also be found to have no preference.

Comparison of the two methods shows that the collexeme analysis accords better with our intuition on the scales associated with the adjectives than does mere counting. For instance, *moronic* would be considered an open-scale adjective based on raw frequencies: there are 10 instances with an end-of-scale modifier in the UKWAC, and 17 instances with normal modifiers. In the BNC, *moronic* occurs just once, with a normal modifier. The collexeme analysis shows, however, that the adjective has a stronger association with the end-of-scale construction than the normal modification construction and should thus be considered an end-of-scale adjective. In ongoing work, we are collecting data from human raters to evaluate the correlation between corpus-based classifications following the results of collexeme analysis and human judgments.

While we consider the corpus-based approach promising, challenges remain. One is, unsurprisingly, polysemy. Since we do not have large random samples of frame-annotated data for all senses of our lemmas available to us, we must simply ignore this issue in any automatic approach. The only problematic constellation for us is polysemy such that while one or more word senses have an open scale, one or more others have a closed scale. Cases where some word senses are simply not gradable such as *dumb* in the sense of 'unable to speak' versus *dumb* 'stupid' are not a problem since the non-gradable sense would not impact the corpus statistics for the gradable sense(s) in the modification constructions. A second concern is, necessarily, access to large enough corpora. As the discussion of *moronic* above suggests, some lexical units will be rare in some smaller corpora (possibly due to associations with particular varieties of English) and no reliable decisions can be made then. Large corpora are also necessary if, in addition to partitioning adjectives into closed and open-scale sets, we want to determine relative orderings. Prime evidence for such ordering relations could come from constructions such as *X let alone Y*, or *X or even Y*, where X and Y are adjectives relating to the same scale. Our exploration of instances of these constructions have shown, however, that typically scale-switches are involved or non-gradable adjectives are ordered pragmatically as in ((13)).

(13) That has never been fully tested by any degree of analysis or qualitative, let alone quantitative, research …

On the other hand, it is not clear how consistently speakers impose orderings on open-scale adjectives. For instance, do speakers of English agree on the relative degrees of (lack of) intelligence associated with *foolish*, *stupid*, and *dumb*? If speakers cannot agree, then we also do not have to keep looking for a method that could induce an ordering. The elicitation of human ratings for sets of adjectives should help answer this question.

4. Impact

Deep analysis One benefit of connecting sentiment analysis with frame semantics is immediate access to a deeper lexical semantics. This is an improvement over resources such as the Pittsburgh subjectivity clues, which just list word or lemma forms and do not allow one to distinguish, for instance, moral evaluations from aesthetic ones. Given particular application-interests, for instance, identifying state-ments of uncertainty, frames and lexical units relevant to the task can be pulled out easily from the general resource, while ignoring expressions related to other types of subjectivity. A frame-based treatment also improves over resources such as SentiWordNet (BACCIANELLA ET AL., 2008), which, while representing word meanings, lacks any representation of semantic roles. However, pursuing sentiment analysis in a frame-semantic context makes sense only in the context of deep analysis. In our own work, we therefore initially focus on a domain-specific context where reasonable coverage can be achieved.

Theoretical insights In terms of research, new questions await to be addressed. For instance, whether predicates with multiple opinions can be distinguished automatically from ones with only one. A related question is whether predicates carrying factivity or other sentiment-related presuppositions can be discovered automatically. Even if they could not be analyzed correctly by a system, it would still be useful to identify them and bring them to the attention of human analysts. Also, our approach allows us to ask how contextual sentiment is, and how much of the analysis of pragmatic annotations can be derived from lexical and syntactic knowledge. It is particularly suited to exploring sentiment that involves inferred Targets that are not syntactic dependents of the opinion expression (cf. the discussion of 11).

Evaluation Importantly, a frame-based representation impacts the possibilities of system evaluation. First, the units of annotation are pre-defined by a general frame semantic inventory. They simply are the recognized frame-evoking lexical units and systems can readily know what kind of units to target as potential opinion-bearing expressions. This is not necessarily the case in other corpora that are frequently used for annotation purposes. For instance, in the MPQA corpus sentence ((14)) contains an opinion-bearing span *may annoy* that is annotated both as a direct subjective element (DSE), that is an expression that can take the Source as a grammatical dependent, and an expressive subjective element (ESE), which cannot do so, even though one could reasonably have annotated the two pieces separately, *may* as an ESE and *annoy* as a DSE. Systems that do not match the combined span will be penalized either for reduced overlap or for not matching, if identical spans are required. By contrast, on a SentiFrameNet-style analysis of the same sentence, there is no alternative to the separate annotation with two different frames. Further, the correctness of inferred Targets and the polarity towards them can be weighted differently in evaluation from that of getting the immediate Targets and the polarity towards them right.

(14) The US naming of North Korea [**may annoy** ESE/DSE] China … (source document:20.21.49-25548)

Sparse data Anchoring sentiment analysis in frame semantics naturally provides more data for training as frame semantic annotations can be recast as sentiment annotation. In fact, FrameNet's full-text data could be enriched with pragmatic annotations.

Synergy The connection between frame semantics and sentiment analysis will create cooperation between NLP efforts that are currently laboring independently, and allow re-use of existing resources. In our own work, we use the Semafor SRL system (DAS ET AL., 2010) to pre-label text with semantic frames, which we then correct manually. Sentiment information that is lexically inherent need not be annotated explicitly at all: it can be added automatically once the semantic frame's roles are annotated. The lexical sentiment information can then be compared to a gold standard that is pragmatic, or we can enrich it with annotations of polar facts or other kinds of inferred sentiment. By expanding the FrameNet inventory and creating annotations, we improve a lexical resource and create role-semantic annotations as well as doing sentiment analysis.

Conclusion

In this contribution, we have argued that current approaches to sentiment analysis suffer from three main shortcomings: a) ad-hoc development of resources and methods which do not result in long-term advances in the field, b) data sparseness and c) the lack of a proper theoretical understanding which is necessary for proper modeling sentiment. We proposed SentiFrameNet as a linguistically sound, deep representation for sentiment analysis, extending an existing resource. Our approach allows us to join forces with related work in NLP (e.g. role labeling, event factuality) and it should also enable new insights into the theoretical foundations of sentiment analysis. With respect to FrameNet and frame semantics, the current proposal advances the idea of breaking frames down even more finely than has been done so far in FrameNet. On a practical level, in implementing this proposal, we have begun exploring ways to (semi-)automatically acquire some of the sentiment-relevant information that we want to incorporate into an extended FrameNet.

References

BACCIANELLA, S.; ESULI, A.; SEBASTIANI, F. SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In: *Proceedings of the Seventh conference on International Language Resources and Evaluation*. La Valletta, Malta, 2008. p. 2200-2204.

BAKER, C. F.; FILLMORE, C. J.; LOWE, J. B. The Berkeley FrameNet project. In: *Proceedings of the 17th International Conference on Computational Linguistics*. Montreal, Canada, 1998. p. 86-90.

BALAHUR, A.; HERMIDA, J.M; MONTOYO, A. Detecting implicit expressions of sentiment in text based on commonsense knowledge. In: *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*. Portland, Oregon, 2011. p. 53-60.

BARONI, M.; BERNARDINI, S.; FERRARESI, A.: ZANCHETTA, E. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, vol. 43, p. 209-226.

BENAMARA, F.; CESARANO, C.; PICARIELLO, A.; REFORGIATO, D; SUBRAHMANIAN, V. Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In: *Proceedings of the International Conference on Weblogs and Social Media*. Boulder, Colorado, 2007. p 1-4.

BOLINGER, D. *Degree words*. The Hague: Mouton, 1972.

BUSSE, D. *Frame-Semantik*. Berlin: De Gruyter, 2012.

DAS, D; SCHNEIDER, N.; CHEN, D.; SMITH, N. A. Probabilistic frame-semantic parsing. In: *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Los Angeles, CA, 2010. p. 948–956.

DRURY, B.; DIAS, G.; TORGO, L. A contextual classification strategy for polarity analysis of direct quotations from financial news. In: *Proceedings of the International Conference Recent Advances in Natural Language Processing 2011*. Hissar, Bulgaria, 2011. p. 434-440.

FILLMORE, C. J.; LEE-GOLDMAN, R.; RHODES, R. The FrameNet Constructicon. In: BOAS H. C.; SAG, I. (Eds.). *Sign based construction grammar*. Stanford, CA: CSLI, 2012.

FILLMORE, C. J. Grammatical construction theory and the familiar dichotomies. In: DIETRICH, R.; GRAUMANN, C. F. (Eds.). *Language processing in social context*. Amsterdam: North-Holland/Elsevier, 1998. p 17–38.

GREENE, S.; RESNIK, P. More than words: syntactic packaging and implicit sentiment. In: *Proceedings of HLT: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*. Boulder, CO, 2009. p. 503-511.

GRIES, S. T.; STEFANOWITSCH, A. Extending collostructional analysis: A corpus-based perspective onalternations. *International journal of corpus linguistics*, vol.9, n. 1, p. 97-129, 2004.

KENNEDY, C.; MCNALLY, L. Scale structure, degree modification, and the semantics of gradable predicates. *Language*, vol. 81, n. 2, p.345-381, 2005.

KIM, S.M; HOVY E. Extracting opinions, opinion holders, and topics expressed in online news media text. In: *Proceedings of the Workshop on Sentiment and Subjectivity in Text*. 2006. Sydney, Australia, 2006. p. 1-8.

MAKS, I.; VOSSEN, P. A verb lexicon model for deep sentiment analysis and opinion mining applications. In: *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis.* Portland, Oregon, 2011. p. 10–18.

MOILANEN, K.; PULMAN, S. Sentiment composition. In: *Proceedings of RANLP 2007*. Borovets, Bulgaria, 2007. p. 258-63.

PANG, B.; LEE, L.; VAITHYANATHAN, S. Thumbs up? sentiment classification using machine learning techniques. In: *Proceedings of EMNLP02, the Conference on Empirical Methods in Natural Language Processing.* Philadelphia, PA, 2002. p. 79-86.

RUPPENHOFER, J.; SOMASUNDARAN, S.; WIEBE, J. Finding the sources and targets of subjective expressions. In: *Proceedings of LREC*, Marrakech, Morocco, 2008. p. 2781-2788.

SAURÍ, R. 2008. *A Factuality Profiler for Eventualities in Text*. Doctoral dissertation. Brandeis University. 2008.

SEKI, Y. Crosslingual opinion extraction from author and authority viewpoints at ntcir-6. In: *Proceedings of NTCIR-6 Workshop Meeting*. Tokyo, Japan, 2007. p. 336-343.

STOYANOV, V.; CARDIE, C. Topic identification for fine-grained opinion analysis. In: *Proceedings of the 22nd International Conference on Computational Linguistics*. Manchester, UK, 2008. p. 817–824.

STOYANOV, V.; CARDIE, C. Automatically creating general-purpose opinion summaries from text. In: *Proceedings of RANLP 2011*. Hissar, Bulgaria, 2011. p.202–209.

TOPRAK, C.; JAKOB, J; GUREVYCH, I. Sentence and expression level annotation of opinions in user-generated discourse. In: *Proceedings of ACL-10, the 48th Annual Meeting of the Association for Computational Linguistics*. Uppsala, Sweden, 2010.p. 575-584.

WIEBE, J.; WILSON, T. A.; CARDIE, C. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, vol. 39, n. 2/3, p.164-210, 2005.

WIEBE, J. Learning subjective adjectives from corpora. In: *Proceedings of the Seventeenth National Conference on Artificial Intelligence*. Austin, Texas, 2000. p. 735-740.

WIEGAND, M. *Hybrid approaches to sentiment analysis*. Doctoral dissertation. Saarland University, Saarbrücken. 2010.

WILSON, T.A. *Fine-grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of Private States*. Doctoral dissertation. University of Pittsburgh, 2008.

WU, Y.; WEN, M. Disambiguating dynamic sentiment ambiguous adjectives. In: *Proceedings of the 23rd International Conference on Computational Linguistics*. Beijing, China, 2010. p. 1191-1199.

Received: 30/01/2013
Accepted: 03/06/2013
Published: 31/10/2013