The JDPA Sentiment Corpus for the Automotive Domain

Jason S. Kessler and Nicolas Nicolov

1 Introduction

The expression of sentiment is a complex phenomenon which is intertwined into the semantic structure of text (Polanyi and Zaenen 2006). A document-level label, such as positive or negative, does not present a full representation of all sentiment present in a document. Sentiment, which we define as evaluation, is expressed toward discourse entities by means of individual expressions of sentiment targeted at mentions of those entities. These expressions of sentiment are often rooted in single or multi-word units, whose positive or negativeness may be impacted by the context. Elements in the context that can alter the polarity include negations and terms which can alter the truth-value of an expression of sentiment, as well as less understood phenomena such as sarcasm and tone. While sentiment toward individual mentions of an entity contribute to its overall sentiment, sentiment toward another, related entity such as a part or a feature may also contribute. Sentiment directed toward individual entities can also effect other entities when comparisons among entities are made. An additional dimension of the phenomena is that certain expressions of sentiment may be attributed to discourse participants other than the speaker.

Our goal is to annotate structures pertinent to sentiment that can be combined to formally explain the sentiment that occurs in a document.

The J.D. Power and Associates (JDPA) Sentiment Corpus consists of user-generated content (blog posts) containing opinions about automobiles. Specifically,

Jason S. Kessler
CDK Global, 605 Fifth Ave S, Ste 800, Seattle, WA 98104
e-mail: jason.kessler@gmail.com

Nicolas Nicolov
J.D. Power and Associates / McGraw-Hill, 4888 Pearl East Circle, Boulder, CO 80301
e-mail: nicolas_nicolov@yahoo.com

Work was conducted while both authors were at the Web Intelligence Division, J.D. Power and Associates / McGraw Hill.

we aim to document, in a fine-grained and compositional way, evaluations of automotive related entities. We define entities as discourse representations of concrete objects (e.g., car, door) and non-concrete objects (e.g., handling, power). Our annotation scheme is rooted in manually annotated mentions at the named entity, common NP, and pronoun level. While only a single mention of an entity is typically evaluated at a time, entities that are prominent topics in the discourse and are of domain importance are marked as having an entity-level sentiment. Entity-level sentiment is the author's overall evaluation of the entity, given the entire discourse context.

The examples we give, unless otherwise specified, are taken directly from the corpus and have not been edited.

Mentions referring to the same entity are marked as co-referential. Mentions are assigned semantic types consisting of the Automatic Content Extraction (ACE) (NIST Speech Group 2006) and other mention types and additional domain-specific types:

Type	# Mentions	# Named	# Nominal	# Pronominal	# Coreference groups
CarPart	14128	1704	11791	633	11705
Vehicles.Cars	8729	4259	2723	1747	3618
Person	7407	764	1487	5156	2593
CarFeature	6263	264	5930	69	5804
Organization	4910	4092	346	472	2164
Vehicles.SUVs	2052	1115	567	370	837
Time.Year	1208	928	258	22	1136
Units.Money	813	177	628	8	616
Units	796	246	536	14	763
Vehicles	770	243	431	96	432
Units.Rate	741	298	436	7	720
Facility	649	147	464	38	512
Time	568	347	211	10	549
Vehicles.Trucks	466	228	172	66	205
Time.Duration	315	78	236	1	303
GeoPolitical.City	251	191	56	4	206
GeoPolitical.Countries	184	156	18	10	130
Location	157	22	133	2	148
GeoPolitical.Nationalities	131	127	4	0	115
GeoPolitical.USStates	98	89	7	2	81
Time.Month	87	74	13	0	84
GeoPolitical	82	51	29	2	70
Time.Date	56	44	11	1	55
Units.Age	41	10	26	5	38
Time.DaysOfTheWeek	36	36	0	0	36
Time.OClock	13	10	3	0	13

Table 1 Distribution of mention annotations.

Meronymy (part-of and feature-of) and instance relations are also annotated. Expressions that convey sentiment toward an entity are annotated with the polarity of their prior and contextual sentiment and are linked to the mentions they target. The following modifiers are annotated. These may target other modifiers or sentiment expressions.

- negators (expressions that invert the polarity of a sentiment expression or modifier);
- neutralizers (expressions that do not commit the speaker to the truth of the target sentiment expression or modifier);

- committers (expressions that shift speaker's certainty toward a sentiment expression or modifier);
- intensifiers (expressions that shift the intensity of a sentiment expression or modifier).

Additionally, we have annotated when the opinion holder of a sentiment expression is someone other than the author by linking the expression to the holder. We also annotate when two entities are compared on a particular dimension.

In this overview of the corpus, we aim to not only present the nature of the annotations we have added, their examples, numbers, and inter-annotator agreement, but also to highlight problems/tasks in sentiment analysis and natural language processing that can be addressed using this corpus.

The data was gathered manually by annotators by conducting web searches using a variety of car-related search terms and restricting the retrieved results to certain blog-host sites. The personal blog posts in particular are different in style and sentence structure from professionally edited news texts, with a higher frequency of emotional and colloquial expressions. However, unlike data from Twitter or other microblogging sites, we found the data to adhere for the most part to standard grammatical rules, and disfluencies or incomplete sentences are rare.

We have annotated 335 blog-posts, covering 13,126 sentences and 223,001 tokens.

In this chapter, we cover the annotation of mentions of entities and their semantic relations, the annotation of sentiment expressions and their modifiers, the annotation process and format, how we judged inter-annotator agreement, and discuss some existing usages of the corpus. Descriptions of annotation types are coupled with statistics about their appearance in the corpus, related work, and potential uses.

2 Obtaining the JDPA Sentiment Corpus

Please visit http://verbs.colorado.edu/jdpacorpus/. The corpus is currently licensed for non-commercial use, and hosted at the University of Colorado, Boulder.

3 Annotation types

Evaluative discourse has two, sometimes overlapping components: references to the entities that are being evaluated and terms that are used to express evaluation, or modify its intensity or polarity. We annotate entities that occur in each document, regardless of whether they have any sentiment associated with them. Each entity is represented by coreferring mention span annotations. Furthermore, entities can have relations between each other.

We first discuss our annotation of mentions and the entities they refer to, as well as semantic relations between entities: part-of, feature-of, instance-of, and member-of. Next, we discuss sentiment expression annotations and their modifiers: negators, neutralizers, committers, and intensifiers.

3.1 Entities and their relations

Entities are defined as discourse representations of concrete objects (e.g., car, door) and non-concrete objects (e.g., handling, power).

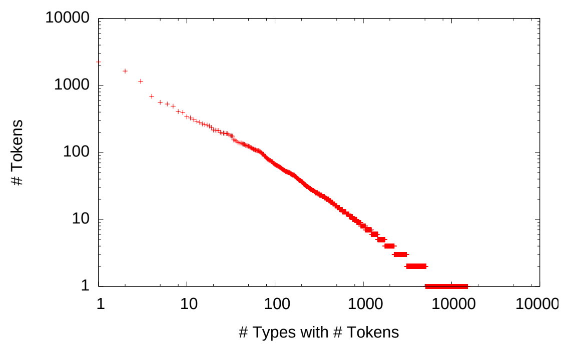

Fig. 1 Types vs. tokens of mentions. The power law exponent is -0.84, with $R^2 = 0.93$.

The most basic relation is **refers-to**. It links together two mentions that are coreferring. The set of coreferent mentions naturally all refer to the same entity.

Winston, Chaffin, and Herrmann (1987) presents six relationships between entities that encompass what humans would consider to be a "part-of" relationship. They annotated for three of these that were found applicable to the automotive domain.

The remaining relations discussed in this section are annotated, on the surface, as relations between mentions. However, these relations are interpreted as connecting the entities referenced by the mentions. The annotators were free to select any mention to represent the entity in the relation.

What we call the **part-of** relation encompasses the relationship of one entity being a concrete part of another. This is Winston et al.'s "component/integral object" relation. They give the examples of "handle-cup; and "punch line-joke". Some of the part-of relationships that we found in the corpus are:

(1) a. Center console$_1$ Kleenex holder$_{\text{PART-OF-1}}$; I cannot find a tissue box that size to fit in it.

b. The 2009 Mercedes-Benz S600$_2$ is equipped with a twin-turbocharged 5.5 - liter V-12 engine$_{\text{PART-OF-2}}$...

The **feature-of** relation also connects entities, but deals with more abstract entities, where one entity is a property of another. This corresponds to Winston et al.'s "feature-activity" relation. Their examples are "paying-shopping" and "dating-adolescence". In our corpus:

(2) a. I love the comfort$_{\text{FEATURE-OF-1}}$ of interior seating$_1$
b. The speed and fuel gauges$_2$ are very hard to see$_{\text{FEATURE-OF-2}}$

Sometimes entities are defined as being a type of or equivalent to another entity. These definitional and hypernymic relations that we call **instance-of** relations do not appear in Winston, Chaffin, and Herrmann (1987). Some examples are:

(3) a. Hyundai's futuristic proposal$_{\text{INSTANCE-OF-1}}$ for a small three-door crossover$_1$...
b. Cadillac has launched the 2009 Escalade Platinum Hybrid$_{\text{INSTANCE-OF-2}}$, the most technically advanced large luxury SUV$_2$ yet.

Member-of relations exist between an entity that is part of a group represented by another entity. For example, the student-class relationship, or the relationship between the Toyota Corolla and Toyota's line of compact sedans. These correspond to Winston et al's "member/collection" relations; their examples are "tree-forest" and "card-deck". An example is:

(4) a. The peeled back headlamps$_{\text{MEMBER-OF-1}}$, tight front grille$_{\text{MEMBER-OF-1}}$, and stylized tail lamps$_{\text{MEMBER-OF-1}}$ are some of its attractive features$_1$.

The corpus has 61,284 mentions which comprise 42,763 co-reference groups (or entities), averaging 1.43 mentions per group. See Table 3 for inter-annotator agreement among mentions and their relations.

3.2 *Sentiment*

3.2.1 Sentiment expressions

Sentiment expressions are single or multi-word phrases that evaluate an entity. They are linked to the mention they modify through the "target" relation. Our corpus contains 10,425 sentiment expressions, covering 3,545 unique types. 49% of sentiment expressions are headed by adjectives, 22% by nouns, 20% by verbs, and 5% by adverbs. This leads to a diversity of syntactic configurations where sentiment ex-

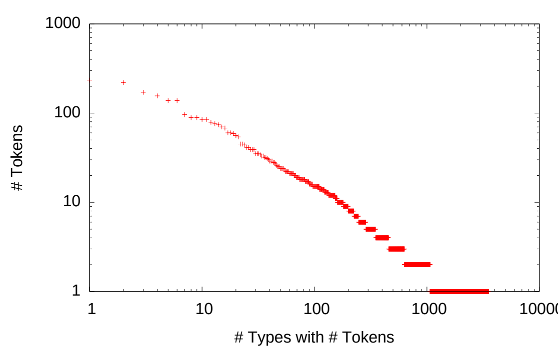

Fig. 2 Types vs. tokens of sentiment expressions. The power law exponent is -0.77, with $R^2 = 0.91$.

pressions are linked to their target mentions (Kessler and Nicolov 2009). 13% of sentiment expressions are more than one word long.

In general, sentiment expressions convey positive or negative evaluations. We use the term **prior polarity** to refer to whether a sentiment expression is positive or negative. The prior polarity is inferred from the meaning of the sentiment expression, given its target, as opposed to its entire context. "Prior polarity" is from Wilson, Wiebe, and Hoffmann (); we allow it to depend on a sentiment expression's sense, figurativeness, and target. Prior polarity contrasts with **contextual polarity** (another term from Wilson, Wiebe, and Hoffmann ()) in that contextual polarity is the polarity of the sentiment expression given any modifiers or contextual information that don't change its inherent meaning or sense. For example, the prior polarity of "good" in Example (5) (invented) is always positive, while its contextual polarity in (5) is respectively positive, negative, and positive. See Table 3 for inter-annotator agreement. We do not annotate contextual polarity directly. Our goal is to make it inferable from modifiers that have been annotated such as negators and others that we discuss below.

(5) a. The car is *good*.
    b. The car is not *good*.
    c. Only an idiot would think the car is not *good*.

The distribution of prior polarities is skewed toward positive, with 74% positive, 24% negative, 1% neutral and well less than 1% of mixed prior polarity. Sentiment expressions having "mixed" prior polarity simultaneously express a positive and negative evaluation. These include "pimped-out", "gangsta", "usable", "subtle," and "curious". Neutral sentiment expressions are evaluations that are not clearly positive or negative, such as "as expected", "average", "conventional", "so-so", and "different". A next step in expanding this corpus is correcting for the skew in positive and negative sentiment expressions.

Table 2(b) shows the 20 most frequently annotated sentiment expressions in the corpus.

Some sentiment expression types have been marked with different prior polarities when they occur in different contexts. For example, the term "increasing" is marked positive in Example (6-a,b) but negative in Example (6-c).

(6) a. ... an electric motor that reduces the load on the engine, *increasing* efficiency.
    b. ... *increasing* combustion efficiency and the torque...
    c. ... *increasing* gas prices and stricter federal emissions regulations...

While prior polarity of "interesting" depends on its topic, other sentiment expressions like "excellent" have a constant prior polarity. Although only 6% of sentiment expression types have tokens with conflicting prior polarities, these account for 25% of sentiment expression tokens in the corpus, making polarity-based disambiguation an important task. Reasons for conflicting prior polarities other than annotator error were the sense of the sentiment expression. For instance, "safe" in Example (7-a) is positive, referring to a vehicle's protectiveness, while "safe" in Example (7-b) is negative, inferring its targets' design is traditional.

(7) a. My family and friends feel extremely *safe* in our Hummer.
    b. I saw two VW Eos last week...... and both looked good, albeit in a *safe*, conservative Solara-sort-of-ways.

# Tokens	Type
2238	i
1639	it
1153	car
687	my
559	engine
527	you
490	its
407	power
395	we
341	vehicle
327	cars
307	one
291	2009
282	interior
266	me
261	they
256	2008
248	ford
235	toyota
216	honda

(a) Mentions

# Tokens	Type
234	good
220	new
171	great
156	like
138	comfortable
138	better
96	love
89	problems
89	fun
85	well
85	unique
79	nice
76	best
74	excellent
70	difficult
68	smooth
60	powerful
60	expensive
59	easy
56	poor

(b) Sentiment expressions

# Tokens	Type
18	seems
18	felt
16	still
16	think
14	seemed
14	feel
14	definitely
13	looks
13	feels
12	certainly
12	may
11	actually
11	might
10	really
10	probably
9	sure
8	seem
8	overall
8	looks like
7	always

(c) Committers

# Tokens	Type
63	if
34	would
27	should
18	could
14	want
13	when
11	optional
10	can
9	needs
8	how
8	may
7	?
6	might
6	or
6	expected
6	need
5	wanted
4	feels
4	expect
4	supposed

(d) Neutralizers

# Tokens	Type
325	very
227	more
122	most
111	much
84	really
77	so
76	top
64	too
58	pretty
39	extremely
38	quite
36	enough
35	even
32	!
32	just
28	less
28	bit
28	a bit
27	completely
26	a little

(e) Intensifiers

# Tokens	Type
299	not
122	no
45	doesn't
44	without
36	don't
28	never
27	isn't
26	didn't
20	don t
20	wasn't
19	doesn t
19	can't
14	aren't
13	won't
13	didn
13	wouldn't
13	nothing
8	lack
8	wasn
8	isn t

(f) Negators

# Tokens	Type
71	like
69	says
66	told
52	owner-reported
30	according
26	ranked
23	said
21	top-ranked
20	ranks
19	according to
12	from
9	reported
9	say
9	calls
8	think
8	rated
7	love
6	rating
6	likes
5	ranking

(g) OPOs

Table 2 Top 20 annotated items in different categories.

Much work (Ding, Liu, and Yu ; Fahrni and Klenner 2008; Choi, Kim, and Myaeng 2009) has focused on identifying the target-dependent polarity of sentiment expressions,[1] while Wiebe and Mihalcea (2006) and Su and Markert (2008) have looked at the problem of polysemy from the perspective of disambiguating subjective and objective senses. The annotations available in the JDPA corpus lend themselves to the task of contextually determining the polarity of sentiment expressions.

Similar annotations exist in the MPQA corpus (Wiebe, Wilson, and Cardie 2005), however; such annotations tend to include modifiers that, in the JDPA corpus, would be annotated separately from the sentiment expression.

For example, in (8) "not happy" is marked as a single subjective expression with a negative attitude type, while in the our annotation scheme "happy" would be marked as a sentiment expression with positive prior polarity, and "not" would be marked as a negator which targets it.

(8) If we're *not happy*$_{\text{ATTITUDE-TYPE: SENTIMENT-NEG}}$, that goes double for our public affairs babysitters. (MPQA corpus, non_fbis/08.46.28-13637)

Wilson, Wiebe, and Hoffmann (2005) present a system to determine the contextual polarity of subjective expressions in the MPQA corpus.

Some expressions are only sentiment-bearing when in the right context. For example the term "usable" occurs nine times in the corpus, four of which are annotated as sentiment expressions. Example (9-a) illustrates an example of "usable" being a sentiment expression, and Example (9-b) illustrates a case where it is not.

[1] We draw the distinction between the immediate target of a sentiment expression and a document-level topic. Other work, such as Nowson (2009), has addressed the problem of developing topic-dependent feature-sets for supervised classification of document-level polarity.

(9) a. . . . a comfortable and *usable* interior. . .
b. . . . 5,800 pounds (2,631 kg) of *usable* towing capacity. . . .

In fact, 44% of sentiment expression types occurring in the corpus also match non-sentiment bearing sequences of words. These account for 74% of all sentiment expression tokens, motivating the need for sentiment expression detection which can disambiguate candidates based on their context. However, 10% of sentiment multi-word units types have a non-sentiment bearing occurrence but are observed to be sentiment-bearing more than half the time. These account for a substantial 40% of all sentiment expression tokens. 34% of sentiment expression types are not sentiment-bearing in more than half their occurrences. These account for 34% of all sentiment-expression tokens.

Breck, Choi, and Cardie (2007) applied sequence labeling techniques to the similar task of identifying subjective expressions, a problem which involves the contextual disambiguation of sentiment bearing and non-sentiment bearing phrases.

Given the 10,000+ sentiment expressions annotated, the corpus is a powerful resource for building and evaluating tools to detect whether a given phrase or sequence of words carries sentiment in context.

Sentiment expressions are linked to the mention they describe through the **target** relation. This forms an important connection between sentiment expressed in a document and the entities discussed. For inter-annotator agreement, we treat the **target** relation as sentiment expression span to entity link, although annotators are instructed to link to the mention that is directly targeted.

Figures 1 and 2 show the comparative types vs. tokens distributions of mentions and sentiment expressions. Both are nearly similar but sentiment expressions, having a larger exponent, have a fatter tail and thus might be more difficult to automatically recognize.

3.3 Contextual polarity and modifiers

There has been considerable work on identifying the contextual polarity of sentiment expressions (Kim and Hovy 2004; Choi and Cardie 2008; Wilson, Wiebe, and Hoffmann ; Wiegand and Klakow 2009; Moilanen and Pulman 2009).

A sentiment expression's context can change or modify its polarity, as illustrated by Example (5). We annotate several types of modifiers, which act to change the polarities of sentiment expressions and change the properties of other modifiers. Similar sets of modifiers have been discussed in the literature, but ours is the first attempt at manually annotating occurrences of these modifiers (Polanyi and Zaenen 2006; Shaikh, Prendinger, and Ishizuka 2008; Choi and Cardie 2008; Moilanen and Pulman 2009).

Negators invert the polarity of the sentiment expression they target.[2] While "not" is the most well known negator, many other expressions act the same way toward

[2] Called "negatives" in Polanyi and Zaenen (2006).

sentiment expressions and other modifiers. For example, in Example (10) “avoids” acts to invert the polarity of the sentiment expression “reduction”. Other counter-factives, like “pretend”, would also be marked as negators.[3]

(10) This layout *avoids* any *reduction* in the interior space…

In addition to targeting sentiment expressions, negators can also target other modifiers (11-a) and even mentions as indicating the absence of an entity. For instance, in Example (11-b) “suppressed” indicates the absence of the entity invoked by the mention “noise”.

(11) a. …*not*$_{\text{TARGET-VERY}}$ a *very quick* car.
b. Road and engine <u>noise</u> have been *suppressed*…

Example (11-a) can be read two ways, as indicated by (12).

(12) a. The car is fast but not very fast.
b. The car is not fast.

(12-a) is the literal reading, while (12-b) expresses what is perhaps the illocutionary force of (11-a). Regardless of the reading, the negator *not* would be annotated as targeting *very*. The ContextualSentiment (see Section 3.4) property of *car* would be annotated as positive in the case of (12-a) an neutral in the case of (12-b).

While negations can introduce scope-related ambiguity, our annotation framework is generally able to be scope-neutral w.r.t. to the polarity of sentiment expressions. For instance, (13-a) (invented) has narrow and wide scope readings, illustrated respectively in (13-a,i) (if something is a part of the car then it is bad) and (13-a,ii) (it is not the case that every part of the car is bad; some may be good).

(13) a. *Not*$_{\text{NEGATOR, TARGET-BAD}}$ every *part* of the car is *bad*$_{\text{SE, TARGET-PART}}$.
(i) $\forall p.\text{part-of-car}(p) \rightarrow \neg\text{bad}(p)$
(ii) $\neg\forall p.\ \text{part-of-car}(p) \rightarrow \text{bad}(p) \equiv \exists p.\ \text{part-of-car}(p) \wedge \neg\text{bad}(p)$

Our annotation scheme does not identify a reading.

1014 negator annotations appear in the corpus, tokens of 160 unique types.

Intensifiers act to amplify or dampen the intensity of the sentiment expressed by a sentiment expression or the force of another modifier. Unlike other annotation schemes (Wiebe, Wilson, and Cardie 2005; Hu and Liu 2004) which record the intensity of sentiment, we do not record the final intensity of sentiment toward an entity, only the polarity. However, recording intensifiers is important, because their interaction with other modifiers has the potential to change the polarity of sentiment, as shown in (13-a).

The direction property can be set to strengthen or weaken. “Considerable” in Example (14-a) has a direction strengthen, and Example (14-b)’s direction is weaken.

(14) a. … it also adds *considerable* benefits …

[3] The TimeML corpus (Pustejovsky et al. 2003) has explicit annotations for counter-factive events and treats negation as a property of an event. We believe that both act the same way w.r.t. contextual polarity. The TimeML corpus is presented elsewhere in this volume.

b. It is *kind of* fun to drive.

The direction strengthen is far more common than weaken, with 2,159 occurrences (84%) of strengthening intensifiers (covering 396 types) and 422 occurrences (16%) of weakening intensifiers, accounting for 155 types.

Committers are used to express the author's certainty toward a modifier or sentiment expression.[4] They often express epistemic modality (as in the case of Examples (15-a,b,d)) or hedges (15-c). Committers have a property, **direction**, upward or downward, indicating whether the commitment is being strengthened or weakened. Examples (15-a,b) are all labeled as upward committers, while (15-d) is downward:

(15) a. It was discovered that the switch itself was *DEFINITELY* cracked ...
b. I'm *sure* this will drive well ...
c. A good looking car *in itself* ...
d. The interior *looks* to be in nice condition ...

The distribution of direction is relatively even with 417 upward committers (covering 202 types) and 379 downward committers (covering 235 types). The high types-to-tokens ratio and sparsity of the annotation type indicates that this type may be difficult to recognize automatically.

Agreement for committer spans is weak—31%. Some committers have been marked as neutralizers or intensifiers and vice versa. In fact, "may" occurs in the top 20 neutralizers and committers (Table 2).

Neutralizers are used to place sentiment expressions or other modifiers into a context where their truth-value is unknown, as occurs in hypothetical or conditional sentences.[5] For the purposes of simplification, in our annotation scheme, neutralizers only target sentiment expressions and not states or events. The targets of the neutralizers in Examples (16-a,b,c) have been shown for clarity. Example (16-a) shows a hypothetical neutralizer, "if" targeting the sentiment expression "poor". That sentiment expression now has a neutral contextual polarity. The neutralizer "tried" in (16-b) is a verb that neutralizes the veridicity of its complement clause, headed by the sentiment expression "like". Example (16-c) is similar, except the neutralized argument is the direct object.

(16) a. ... $\textit{if}_{\text{TARGET-1}}$ ... the interior is $\textit{poor}_1$ ...
b. I $\textit{tried}_{\text{TARGET-2}}$ to get used to it and $\textit{like}_2$ it ...
c. Aimed at young couples and families who $\textit{look for}_{\text{INTENSIFIER; TARGET-2}}$ a $\textit{higher level}_{2,\text{SENT. EXP.; TARGET-3}}$ of $\textit{performance}_3$ ...

437 neutralizers (covering 150 types) are annotated in the corpus.

---

[4] Rubin (2007) presents a corpus containing "certainty markers", or expressions indicating commitment to a sentence or a clause and its level of certainty, on a scale from uncertain through absolute certainty. Our committers are judged on a binary scale: do they raise or lower the author's commitment to a sentiment expression or modification.

[5] The problem of determining when an event is asserted as true, false or unknown truth-value is called veridicity (Karttunen and Zaenen 2005). Kessler (2008) developed a rule-based system for recognizing the veridicity of some clauses which is tailored to the blogosphere and released a lexicon which includes "neutral veridicality elements" which neutralize their argument clauses.

Due to the scarcity and difficulties in annotating, we feel that committers and neutralizers should be treated with caution when used as training and evaluation examples.

3.4 Entity and mention-level sentiment

Sentiment is marked for certain mentions. Most sentiment is inferable from the structure of sentiment expressions and their modifiers, as all sentiment expressions target mentions. However, in the case where sentiment expressions of conflicting contextual polarities target a mention or in similarly ambiguous cases, annotators mark the **ContextualSentiment** property of mentions. Other mentions carry some inherent sentiment, which we refer to as **MentionPriorPolarity**. For example, referring to a car as a “lemon” would convey a negative mention prior polarity.

Entities that are judged to be prominent are assigned an **EntityLevelSentiment**, which summarizes the author’s sentiment toward that entity and its meronyms. A mention of a prominent entity is annotated with entity-level sentiment. 873 entities are assigned entity-level sentiment. These entities have an average of eight either direct or indirect meronyms (e.g., the seats in a car’s interior.) Many singletons and entities which are not invoked by many mentions exist in the corpus. Thus, the average prominent entity has 13 mentions referring to it, or referring to one of its direct or indirect meronyms. An average of four sentiment expressions targeted any of these mentions.

3.5 Other person’s opinions

Reported speech has been a prominent topic in subjectivity and sentiment analysis (Breck and Cardie 2004; Kim and Hovy 2006; Ruppenhofer, Somasundaran, and Wiebe 2008; Krestel, Witte, and Bergler 2008).

We chose to annotate the source of reported speech when a direct or indirect quotation contains a sentiment expression, and the source of the reported speech is not the author. In this case, the source of the reported speech can also be called the opinion holder.

We annotate a word or an expression indicating reported speech as an OPO, an abbreviation for “Other Person’s Opinion”. An OPO annotation takes two slots, one being the source of the reported speech, and the other called the target. The target is how we represent the sentiment-relevant quotation. It consists of a list of all sentiment expressions, modifiers, and other OPOs within the quotation.

(17) consists of some invented examples illustrating how OPOs are annotated. (17-a,b) illustrate that direct and indirect quotation are handled identically. (17-c) illustrates how OPOs can target other OPOs in the case of nested sentiment-bearing quotations.

(17) a. Bill *said*SOURCE-BILL, TARGET-GOOD "the car is *good*".
b. Bill *thinks*SOURCE-BILL, TARGET-GOOD the car is *good*.
c. Bill *said*SOURCE-BILL, TARGET-THINKS Mary *thinks*SOURCE-MARY, TARGET-GOOD the car is *good*.

These constraints (the quotation contains sentiment and the source of the quotation is not the author) were added in order to make the best use of our finite annotation resources.

Our handling of other person's opinions contrasts with the MPQA annotation scheme (Wiebe, Wilson, and Cardie 2005), where all reported speech and subjectivity was attributed a source.

We annotate speech events or sentiment expressions that select for a source (i.e., Wiebe, Wilson, and Cardie (2005)'s direct subjective expressions) with the OPO or other person's opinion annotation. Example (18-a) illustrates an objective speech event sourcing a sentiment expression to someone other than the author, while (18-b) shows an example of a speech event that is also a sentiment expression. In (18-b), *love* is annotated both as an OPO and as a sentiment expression. The sentiment expression annotation of *love* targets *cars*, while the OPO annotation of *love* is sourced by *kids* and targets the sentiment expression annotation of *love*.

(18) a. The guards at Indian Point *told*TARGET-NICE, SOURCE-TOLD me [that I have a] *nice* car ...
b. My kids *love*SEE BELOW cars ...

Love annotations in (18-b).
OPO annotation: SOURCE-KIDS, TARGET-LOVE (sentiment expression annotation)
Sentiment expression annotation: TARGET-kids

792 OPOs are annotated in the corpus, covering 250 unique types. Agreement for OPO spans is 53%, for OPO targets is 67%, and for OPO sources 85%.

4 Annotation process

Annotators were trained by reviewing detailed written annotation guidelines and being trained on and having annotated a pilot project, and having their annotations be reviewed by a manager or experienced annotator. Annotators were instructed to mark up text that appeared to fit the criteria for a particular annotation type regardless of its syntactic properties. The annotation scheme was developed by collectively annotating several documents, and reviewing them in meetings. Seven annotators contributed to the corpus.

Most documents were annotated independently, and were not peer-reviewed. Some documents were annotated by multiple people in order to compute inter-annotator agreement metrics. The annotations we chose to release were those of the most experienced annotator.

During the process of corpus creation, some annotation concepts became more concise, some proved to be not clearly enough defined to be accurately annotated, and others required the addition or deletion of slots. A new batch was started when a change to the annotation schema became necessary, or if an existing batch became too large. The following is a description of the individual batches.

- Batch 001: First batch. Size: 78,604 tokens.
- Batch 004: Addition of Mention.CarFeature to distinguish concrete, removable or purchasable CarParts from more abstract CarFeatures such as *power, acceleration* and *drive*. Size: 7,643 tokens.
- Batch 005: Batch consists of J.D. Power and Associates car review files. These were selected because they were felt to have a higher density of auto-related sentiment than the blogs that were examined in prior batches. Size: 42,019 tokens.
- Batch 006: Addition of Mention.Descriptor[6] for adjectives preceding mention nouns, such as ***heated***, ***power*** *seats*; MemberOf slot added to link individual mentions to a plural mention. Size: 95,864 tokens.
- Batch 007: Removal of Mention.Descriptor and addition of Descriptor class to reflect the fact that descriptors do not refer to discourse entities. Size: 11,221 tokens.
- Batch 008: Same format as Batch 007. Size: 30,612 tokens.

5 Release Format

The annotations are stored as XML-encoded, stand-off mark-ups produced by the Protégé plug-in Knowtator (Ogren 2006), the tool which as used to annotate documents.

We provide stand-off annotation files in XML format outputted by Knowtator. These XML files are in
`car/batch<batch number>/annotation/<file identifier>.xml`

The corresponding text files, copied from their original sources are in
`car/batch<batch number>/txt/<file identifier>.txt`

Some files have accompanying metadata, which includes the URL of the file's text. These are in
`car/batch<batch number>/meta/<file identifier>-meta.xml`

In Knowtator's XML format annotations span two or more tags, within a document's `<annotations>` tag.

The first tag is `<annotation>`, containing the `<mention>` subtag, specifying the id of the annotation. Next is the `<annotator>` subtag, giving an anonymized annotator's id and pseudonym. `<span>` specifies the start and end byte-offsets of the annotation and the text it spans while `<spannedText>` contains the text covered by the annotation. `<spannedText>` is optional and may omit some leading/trailing whitespace (or multiple whitespaces). See the `<annotation>` tag below for an example.

The second tag is `<classMention>`, linked to the annotation tag's id by the 'id' attribute. The only required subtag is `<mentionClass>`, whose content and

[6] For details on mention descriptors see the sentiment annotation guidelines (Eckert et al. 2010).

'id' attribute are the semantic type of the annotation. A `<classMention>` tag may have zero or more `<hasSlotMention>` subtags. Each of these corresponds to a property of the annotation, detailed in either a `<stringSlotMention>` tag or a `<complexSlotMention>` tag. The *SlotMention tags are linked via the 'id' attribute in `<hasSlotMention>`.

`<stringSlotMention>` is used for slots that have properties which are nominal, numeric or textual. The slot's name is in the 'id' attribute of the subtag `<mentionSlot>` while the value of the slot is in the 'value' attribute of the `<stringSlotMentionValue>` subtag.

Some slots are used to refer to other annotations. These "complex" slots are specified through the `<complexSlotMention>` tag. Like `<stringSlotMention>` this tag requires the `<mentionSlot>` subtag, whose 'id' attribute specifies the name of the slot. However, its value is specified through the 'value' attribute of `<complexSlotMentionValue>` subtag. The value is always the id of the annotation that the slot refers to. Some `<complexSlotMention>` tags have multiple `<clomplexSlotMentionValue>` subtags, each containing an annotation id.

The following example shows how these tags fit together to form a single annotation:

```
<annotations textSource="car-001-xxx.txt">
...
<annotation>
  <mention id="car-001--xxx-20755" />
  <annotator id="A3">Annotator 3</annotator>
  <span start="0" end="6" />
  <spannedText>Nissan</spannedText>
</annotation>

<classMention id="car-001--xxx-20755">
  <mentionClass id="Mention.Organization">Mention.Organization</mentionClass>
  <hasSlotMention id="car-001-20759" />
  <hasSlotMention id="car-001-21156" />
</classMention>

<stringSlotMention id="car-001--xxx-20759">
  <mentionSlot id="EMLevel" />
  <stringSlotMentionValue value="Named" />
</stringSlotMention>

<complexSlotMention id="car-001--xxx-21156">
  <mentionSlot id="RefersTo" />
  <complexSlotMentionValue value="car-001--xxx-21145" />
</complexSlotMention>
...
</annotations>
```

This generic annotation format allows for the representation of the many different annotation types and their various parameters.

5.1 Inter-annotator agreement

Assessing inter-annotator agreement on the corpus involves analyzing several types of annotations: *spans*, *properties*, *span-span links*, *span-entity links*, and *entity-entity links*.

Spans are markings of consecutive sequences of tokens. Annotators assign these spans one of the annotation types, we define in the Section 3, Annotation types (also see Table 1). We consider two spans to match if they have at least one overlapping token and are of the same annotation type. Text-spans might be annotated with properties. Two spans can still match even if they have conflicting property annotations. We explain how we assess inter-annotator agreement on properties shortly. For instance, the span annotations, denoted by underlining in Examples (19) and (20) match, while those in Example (21) do not:

(19) a. My Honda Civic coupe . . .
b. My Honda Civic coupe . . .

(20) a. My Honda Civic coupe . . .
b. My Honda Civic coupe . . .

(21) a. My Honda Civic coupe . . .
b. My Honda Civic coupe . . .

To assess agreement on spans, we employ the *agr* metric, introduced by Wilson and Wiebe (2003), as a means of determining agreement of their subjective expression span annotations. $agr(A||B)$, where A and B are sets of spans marked by different annotators, gives the precision of A's annotations against B's. Formally, $arg(A||B) = \frac{|A \text{ matches } B|}{|A|}$.

Agreement on span properties is only measured on matching spans. Although Cohen's κ (Cohen 1960) has been used to measure inter-annotator agreement on nominal coding tasks such as this, our situation is complicated by heavily skewed distributions and the fact that multiple annotators have marked distinct sets of documents. Therefore, we only report observed agreement, or given annotators A and B, $obs(A,B) = \frac{|\text{Spans where A and B overlap and share the same property}|}{|\text{Spans where A and B overlap}|}$. The final agreement score is micro-average of all *obs* over all pairs of annotators, weighted by the number of properties annotated.

Span-span links are directed relations between spans (e.g., a negator [source span] and a sentiment expression [destination span]). We only measure inter-annotator agreement on links where both annotators have marked overlapping source spans and have marked the spans as having the same link relation (e.g., both source spans are negators and have target relations). Two span-span links match when the destination spans overlap, and mismatch when the destination spans do not.

We calculate agreement between two annotators, A and B in the following way. Let $aligned(A,B,t,r)$ be the number of spans of type t (e.g., negator) that A and B annotated that overlapped and were annotated with the link relation r (e.g., target). Let $correct(A,B,t,r)$ be the subset of $aligned(A,B,t,r)$ where the r relations' destination spans overlapped. We define the pair-wise annotator agreement metric, *agr* as:

$$agr(A,B,t,r) = \frac{correct(A,B,t,r)}{aligned(A,B,t,r)}$$

To compute global agreement, we take a weighted average of each annotator's *agr* scores in the following way: let S be the set of annotators, t be the source span annotation type, and r be the relation type:

$$agreement(t,r) = \frac{1}{\sum_{A,B \in S: A \neq B} aligned(A,B,t,r)} \sum_{A,B \in S: A \neq B} correct(A,B,t,r)$$

The *agreement* weights the output of annotators roughly by the number of valid annotations they contributed.

Span-entity links are directed relations between a span and a co-reference group (i.e., an entity). For example, consider the target relation between a sentiment expression and its target. While the sentiment expression is linked through the target relation to a specific mention, we are interested, for the purpose of detecting entity-level sentiment, in the co-reference group it targets. This means that we would treat the following two invented annotations (22-a,b) as matching, although they would not match if we treated sentiment expression-target relations as span-span links.

(22) a. Bob bought a new $\underline{\text{Malibu}}_1$ and *loves*$_{\text{SENT, TARGET-1}}$ $\underline{\text{it}}_{\text{2,REFERS-TO: 1}}$.
b. Bob bought a new $\underline{\text{Malibu}}_1$ and *loves*$_{\text{SENT, TARGET-2}}$ $\underline{\text{it}}_{\text{2,REFERS-TO: 1}}$.

We use the same formulas as span-span links to compute agreement, but we relax the definition of *correct* to include the case when two destination spans both refer to the same entity. This requires us to align co-reference annotations across documents. Consider Example (23) (invented), the same sentence annotated by two different annotators:

(23) a. I bought a new $\underline{\text{R8}}_1$ and drove $\underline{\text{it}}_2$ home. The $\underline{\text{Audi}}_{\text{3REFERS-TO:2}}$ *rocks*$_{\text{TARGET:3}}$.
b. I bought a new $\underline{\text{R8}}_1$ and drove $\underline{\text{it}}_{\text{2REFERS-TO:1}}$ home. The $\underline{\text{Audi}}_3$ *rocks*$_{\text{TARGET:1}}$.

We consider the target relations of "rocks" to match, because the two targeted co-reference groups both have one matching mention ("it"). In (23-a), "it" and "Audi" are co-referent, while in (23-b), "R8" and "it" are co-referent. We consider these two annotations of co-reference groups aligned, since they both match on the mention "it". Thus, even though "rocks" in (23-a) and (23-b) targets different mentions, the target entities are treated as matching.

Entity-entity links are directed relations between two co-reference groups (e.g., part-of or feature-of). We calculate aggregate agreement for each annotation type using the same processes we use for span-span links and span-entity links.

Annotation	Property	Type	Agreement	# Matched
Mention	–	span	0.83	21,518
Mention	Semantic Type	property	0.83	17,923
Mention	MentionPriorPolarity	property	1.00	7
Mention	ContextualSentiment	property	0.95	13
Mention	EntitySentiment[1]	property	0.85	87
Mention	Inferred Contextual Sentiment[2]	property	0.87	18,706
Mention	Refers-to	span-entity link	0.68	5,684
Mention	Part-of	entity-entity link	0.35	1,178
Mention	Feature-of	entity-entity link	0.23	294
Mention	Member-of	entity-entity link	0.81	34
Mention	Instance-of	entity-entity link	0.73	184
SentimentExpression	–	span	0.75	3,976
SentimentExpression	PriorPolarity	property	0.95	3,712
SentimentExpression	Target	span-entity link	0.66	2,879
Negator	–	span	0.66	384
Negator	NegatorTarget	span-span link	0.85	335
Neutralizer	–	span	0.36	70
Neutralizer	NeutralizerTarget	span-span link	0.78	64
Intensifier	–	span	0.60	729
Intensifier	IntensifierDirection	property	0.96	690
Intensifier	IntensifierTarget	span-span link	0.95	737
Committer	–	span	0.33	93
Committer	CommitterDirection	property	0.91	79
Committer	CommitterTarget	span-span link	0.82	75

[1] Because this is a span property, matches are only counted when both annotators marked EntitySentiment toward matching mentions.
[2] This was automatically determined through a heuristic that accounted for targeting sentiment expressions, modifiers, and annotated prior polarity or contextual sentiment.

Table 3 Inter-annotator agreement on annotation types and their properties.

5.2 Comparison to other resources

We know of two other publicly available corpora that contain opinion-related information in English that include targets of opinions.

The first was presented in Hu and Liu (2004), in which the topic of each sentence is annotated and its contextual sentiment value is given. The sentences are drawn from online reviews of five consumer electronics devices. It contains 113 documents spanning 4,555 sentences and 81,855 tokens. While our corpus is larger and contains much richer annotations, it does not contain annotations for implicit sentiment expressions which are indirectly covered by their approach. Additionally, they annotate sentences containing comparisons

The second is the subset of the MPQA v2.0 corpus containing target annotations (Wilson 2008). The documents are mostly news articles. It contains 461 documents spanning 80,706 sentences and 216,080 tokens. It contains 10,315 subjective expressions (annotated with links) that link to 8,798 targets. These subjective expressions are annotated with “attitude types” indicating what type of subjectivity they invoked. 5,127 of these subjective expressions convey sentiment.

The MPQA corpus has been an important resource in sentiment analysis, and is presented elsewhere in this book. Its annotation scheme captures forms of private states beyond entity-targeted evaluations, such as speculations and beliefs. It is discussed in this volume (Wilson, Wiebe, and Cardie 2015).

There are a number of other co-reference annotation efforts. Two are included in this volume (Poesio, Chamberlain, and Kruschwitz 2015; Iida 2015).

6 Usage

This corpus has been used to develop novel algorithms for finding targets of sentiment expressions (Kessler and Nicolov 2009). This was the initial usage of the corpus, showing how a supervised learning system trained on the corpus was used to link sentiment expressions to their target mentions. This substantially outperformed existing rule and heuristic-based systems. Vaswani (2012) explored a similar approach in a cross-domain setting. Ginsca (2012) looked at the same problem in a supervised setting, and archived state-of-the art results using tree kernels. Linking negators to sentiment expressions was also explored in Ginsca (2012). Jbara (2013) also looks at supervised and unsupervised targeting of sentiment expressions.

Internally, at J.D. Power and Associates we used this corpus to create statistical sentiment expression identification systems. We extended the corpus with other domains and additional relations. The importance of co-reference and entity-level sentiment was explored on a precursor of the corpus and is discussed in Nicolov, Salvetti, and Ivanova (2008). Kessler and Nicolov (2009), mentioned above, made heavy use of dependency parsing. To support this effort, we investigated new dependency parsing techniques (Choi and Nicolov 2009; 2010).

We also investigated unsupervised in sentiment analysis to deal with domains that were not annotated using the scheme presented in this chapter. We analyzed large data sets of user-generated content referring to a company, its products and services, and would extract hierarchical topics and determine sentiment about them. Topics change over time and detecting topic drift becomes crucial (Knights, Mozer, and Nicolov 2009a; 2009b). We also investigated a data-driven way for identifying topics and multi-word expressions associated with them (Lindsey, Headden, and Stipicevic 2012).

Social media data contains considerable amount of spam and we explored fast approaches for spam detection (Nicolov and Salvetti 2007; Salvetti and Nicolov 2006).

Brown (2011) used the corpus to create a supervised system to label part-of, feature-of, instance-of, and member-of relations between mentions. He also labels the ‘produces’ relation, which is not discussed in this chapter.

Yu and Kübler (2011) used the corpus to experiment with supervised, semi-supervised, and cross-domain learning to improve sentence-level opinion identification. The automotive and digital camera review portion (not discussed in this chapter) served as separate domains for the cross-domain learning setting.

Bloom (2011) used the corpus to evaluate an appraisal expression recognizer, where appraisal expressions are semantic structures often corresponding to opinion holder/sentiment expression/target relations. He provides some insights into corpus inconsistencies and annotation issues.

W. Kessler (2013) used the comparison annotation set (not discussed in this chapter) to train and test a system to recognize semantic structures representing comparisons between entities.

7 Discussion

We are interested in annotating domains beyond automotive. So far we have annotated around 100,000 tokens in the consumer electronics domain (digital cameras) which we are also making available.

We have designed this corpus to be used as training and testing data for machine learning experiments. Detecting span annotations may be cast as sequence labeling (e.g., Breck, Choi, and Cardie (2007)) while detecting span properties may be simultaneously cast as an aspect of a sequence labeling problem (e.g., the semantic type of a named entity in named entity recognition) or as a separate task, along the lines of word-sense-disambiguation. Learning the refers-to relation can be cast as a coreference resolution problem (Ng and Cardie 2002). Systems to identify span-span links can be trained through supervised ranking. For example, in our previous work, Kessler and Nicolov (2009) used this technique to identify the targets of sentiment expressions in a previous version of the corpus, considering it a span-span relation. We used the output of a dependency parser in conjunction with other linguistic information as features in our supervised learning system. Entity-entity links such as part-of relations can be identified through methods such as Girju, Badulescu, and Moldovan (2006), and have been explored in Brown (2011).

The corpus was created with the intention of exploring how sentiment toward parts and features of products ultimately registers as sentiment toward the larger, topical product. In other words, we show how sentiment toward the durability of floor mats affects the overall evaluation expressed toward the car. However, this annotation scheme doesn't explain why sentiment toward one part may be more important than sentiment toward another. For example in (24) (invented), the safety record of the manufacturer is shown to be much more important in the writer's sentiment toward the car than its comfort.

(24) While the car's leather seats are luxurious, I can't buy the car because of the manufacturer's pitiful safety record.

Asher, Benamara, and Mathieu (2008) describes how some discourse representations can help elucidate these effects. Adding these to the corpus would be worthy future work.

We currently have no way, besides the contextual sentiment annotation of mentions, to account for issues such as tone and sarcasm. Recent work (Tsur, Davidov, and Rappoport 2010), makes inroads into addressing these difficult aspects of sentiment.

8 Conclusion

In this chapter we have introduced a sentiment corpus with rich annotations, described the various annotation types and relations, presented statistics including inter-annotator agreement, and we have cataloged components of sentiment that occur naturally. We have also assessed their prevalence and have found a very diverse form of linguistic expression that demonstrates many issues in semantics and discourse. We have discussed some uses of the corpus, and potential future work. We hope this corpus will be of interest to researchers building the next-generation of sentiment analysis systems.

Acknowledgements We would like to thank Miriam Eckert and Lyndsie Clark who assisted with an earlier iteration of the corpus description (Kessler et al. 2010).

We would like to thank Martha Palmer, James Martin, Michael Mozer at University of Colorado, and Michael Gasser at Indiana University. We would also like to thank our fellow team members at J.D. Power and Associates: Laura Meredith Green, Hanna Lind, Whitney Zimmer, Claire Bonial, George Figgs, Steliana Ivanova, Ron Woodward, William Headden, Dan Knights, Shumin Wu, Jinho D. Choi, Jingjing Li, and Greg Brown.

References

Asher, N.; Benamara, F.; and Mathieu, Y. Y. 2008. Distilling Opinion in Discourse: A Preliminary Study. In *Proceedngs of the International Conference on Computational Linguistics (COLING-2008): Companion volume: Posters*, 7–10.

Bloom, K. 2011. *Sentiment Analysis Based on Appraisal Theory and Functional Local Grammars*. Ph.D. Dissertation, Illinois Institute of Technology.

Breck, E., and Cardie, C. 2004. Playing the Telephone Game: Determining the Hierarchical Structure of Perspective and Speech Expressions. In *Proceedings of COLING-2004*, 120–126.

Breck, E.; Choi, Y.; and Cardie, C. 2007. Identifying expressions of opinion in context. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI)*, 2683–2688.

Brown, G. I. 2011. An Error Analysis of Relation Extraction in Social Media Documents. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT-2011) Student Session*, 64–68.

Choi, Y., and Cardie, C. 2008. Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2008)*, 793–801.

Choi, J. D., and Nicolov, N. 2009. K-best, Locally Pruned, Transition-based Dependency Parsing Using Robust Risk Minimization. In *Recent Advances in Natural Language Processing V*, Current Issues in Linguistic Theory (CILT 309). Amsterdam & Philadelphia: John Benjamins. 205–216.

Choi, J. D., and Nicolov, N. 2010. K-best, Locally Pruned, Transition-based Dependency Parsing Using Robust Risk Minimization and Automatic Feature Reduction. In Tufiş, D., and Forăscu, C., eds., *Multilinguality and Interoperability in Language Processing*. 288–302.

Choi, Y.; Kim, Y.; and Myaeng, S.-H. 2009. Domain-specific Sentiment Analysis Using Contextual Feature Generation. In *1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion Measurement (TSA)*, 37–44.

Cohen, J. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20(1):37–46.

Ding, X.; Liu, B.; and Yu, P. S. A Holistic Lexicon-based Approach to Opinion Mining. In *Proceedings of the 1st ACM International Conference on Web Search and Web Data (WSDM 2008)*, 231–240.

Eckert, M.; Clark, L.; Lind, H.; Kessler, J.; and Nicolov, N. 2010. *Structural Sentiment and Entity Annotation Guidelines*. J. D. Power and Associates Technical Report.

Fahrni, A., and Klenner, M. 2008. Old Wine or Warm Beer: Target-Specific Sentiment Analysis of Adjectives. In *Symposium on Affective Language in Human and Machine, AISB 2008 Convention*, 60–63.

Ginsca, A.-L. 2012. Fine-Grained Opinion Mining as a Relation Classification Problem. In Jones, A. V., ed., *2012 Imperial College Computing Student Workshop, ICCSW 2012*, volume 28 of *OASICS*, 56–61. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany.

Girju, R.; Badulescu, A.; and Moldovan, D. 2006. Automatic Discovery of Part-Whole Relations. *Computational Linguistics* 32(1):83–135.

Hu, M., and Liu, B. 2004. Mining and Summarizing Customer Reviews. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2004)*, 168–177.

Iida, R. 2015. NAIST Text Corpus. In Ide, N., and Pustejovsky, J., eds., *Handbook of Linguistic Annotation*. Springer.

Jbara, A. A. 2013. *Using Natural Language Processing to Mine Multiple Perspectives from Social Media and Scientific Literature*. Ph.D. Dissertation, The University of Michigan.

Karttunen, L., and Zaenen, A. 2005. Veridicity. In Katz, G.; Pustejovsky, J.; and Schilder, F., eds., *Annotating, Extracting and Reasoning about Time and Events*, number 05151 in Dagstuhl Seminar Proceedings. Dagstuhl, Germany: Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany.

Kessler, W., and Kuhn, J. 2013. Detection of Product Comparisons – How Far Does an Out-of-the-Box Semantic Role Labeling System Take You? In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2013)*, 1892–1897.

Kessler, J. S., and Nicolov, N. 2009. Targeting Sentiment Expressions through Supervised Ranking of Linguistic Configurations. In *Proceedings of the 3rd International Conference on Weblogs and Social Media (ICWSM-2009)*, 90–97. San Jose, CA, USA: Association for the Advancement of Artificial Intelligence (AAAI).

Kessler, J. S.; Eckert, M.; Clark, L.; and Nicolov, N. 2010. The 2010 ICWSM JDPA Sentiment Corpus for the Automotive Domain. In *4th International AAAI Conference on Weblogs and Social Media Data Workshop Challenge (ICWSM-DWC 2010)*.

Kessler, J. S. 2008. Polling the Blogosphere: a Rule-Based Approach to Belief Classification. In *2nd International Conference on Weblogs and Social Media (ICWSM 2008)*, 68–75.

Kim, S.-M., and Hovy, E. 2004. Determining the Sentiment of Opinions. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING-2004)*, 1367–1373.

Kim, S.-M., and Hovy, E. 2006. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. In *ACL Workshop on Sentiment and Subjectivity in Text*, 1–8.

Knights, D.; Mozer, M.; and Nicolov, N. 2009a. Detecting Topic Drift. In *Recent Advances in Natural Language Processing V*, Current Issues in Linguistic Theory (CILT 309). Amsterdam & Philadelphia: John Benjamins. 113–130.

Knights, D.; Mozer, M.; and Nicolov, N. 2009b. Detecting Topic Drift with Compound Topic Models. In *Proceedings of the 3rd International Conference on Weblogs and Social Media (ICWSM 2009)*, 242–245.

Krestel, R.; Witte, R.; and Bergler, S. 2008. Minding the Source: Automatic Tagging of Reported Speech in Newspaper Articles. In *Proceedings of the 6th International Language Resources and Evaluation Conference (LREC'08)*, 2823–2828. Marrakech, Morocco: European Language Resources Association (ELRA).

Lindsey, R.; Headden, W. P.; and Stipicevic, M. J. 2012. A Phrase-Discovering Topic Model Using Pitman-Yor Processes. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL'12)*, 214–222.

Moilanen, K., and Pulman, S. 2009. Multi-Entity Sentiment Scoring. In *Proceedings of Recent Advances in Natural Language Processing (RANLP'09)*, 258–263.

Ng, V., and Cardie, C. 2002. Improving Machine Learning Approaches to Coreference Resolution. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)*, 104–111.

Nicolov, N., and Salvetti, F. 2007. Efficient Spam Analysis for Weblogs through URL Segmentation. In *Recent Advances in Natural Language Processing (RANLP)*, Current Issues in Linguistic Theory (CILT 292). Amsterdam & Philadelphia: John Benjamins. 125–136.

Nicolov, N.; Salvetti, F.; and Ivanova, S. 2008. Sentiment Analysis: Does Coreference Matter? In *Affective Language in Human and Machine*, AISB 2008 Convention: Communication, Interaction and Social Intelligence, 37–40.

NIST Speech Group. 2006. The ACE 2006 Evaluation Plan: Evaluation of the Detection and Recognition of ACE Entities, Values, Temporal Expressions, Relations, and Events.

Nowson, S. 2009. Scary Movies Good, Scary Flights Bad: Topic Driven Feature Selection for Classification of Sentiment. In *1st International CIKM Workshop on Topic-Sentiment Analysis for Mass Opinion Measurement (TSA)*, 17–24. Hong Kong, China: ACM.

Ogren, P. V. 2006. Knowtator: A Protégé plug-in for annotated corpus construction. In *Proceedings of the 2006 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume: demonstrations (NAACL-HLT'06)*, 273–275.

Poesio, M.; Chamberlain, J.; and Kruschwitz, U. 2015. Phrase Detectives. In Ide, N., and Pustejovsky, J., eds., *Handbook of Linguistic Annotation*. Springer.

Polanyi, L., and Zaenen, A. 2006. Contextual Valence Shifters. In *Computing Attitude and Affect in Text: Theory and Applications*. 1–10.

Pustejovsky, J.; Hanks, P.; Sauri, R.; See, A.; Gaizauskas, R.; Setzer, A.; Radev, D.; Sundheim, B.; Day, D.; Ferro, L.; and Lazo, M. 2003. The TIMEBANK Corpus. In *Proceedings of Corpus Linguistics*, 647–656.

Rubin, V. L. 2007. Stating with Certainty or Stating with Doubt: Intercoder Reliability Results for Manual Annotation of Epistemically Modalized Statements. In *Proceedings of NAACL-HLT'07*, 141–144.

Ruppenhofer, J.; Somasundaran, S.; and Wiebe, J. 2008. Finding the sources and targets of subjective expressions. In Calzolari, N. et al., ed., *Proceedings of the 6th International Language Resources and Evaluation (LREC'08)*, 2781–2788.

Salvetti, F., and Nicolov, N. 2006. Weblog Classification for Fast Splog Filtering: A URL Language Model Segmentation Approach. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL*. 137–140.

Shaikh, M. A. M.; Prendinger, H.; and Ishizuka, M. 2008. Sentiment Assessment of Text by Analyzing Linguistic Features and Contextual Valence Assignment. *Applied Artificial Intelligence* 22(6):558–601.

Su, F., and Markert, K. 2008. From Words to Senses: A Case Study of Subjectivity Recognition. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, 825–832.

Tsur, O.; Davidov, D.; and Rappoport, A. 2010. ICWSM – A Great Catchy Name: Semi-Supervised Recognition of Sarcastic Sentences in Product Reviews. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM 2010)*, 162–169.

Vaswani, V. 2012. *Predicting sentiment-mention associations in product reviews*. Ph.D. Dissertation, Kansas State University, Department of Computing and Information Sciences.

Wiebe, J., and Mihalcea, R. 2006. Word Sense and Subjectivity. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING-ACL 2006)*, 1065–1072.

Wiebe, J.; Wilson, T.; and Cardie, C. 2005. Annotating Expressions of Opinions and Emotions in Language. *Language Resources and Evaluation* 39(2-3):165–210.

Wiegand, M., and Klakow, D. 2009. Topic-Related Polarity Classification of Blog Sentences. In Lopes, L. S.; Lau, N.; Mariano, P.; and Rocha, L. M., eds., *Proceedings of the 14th Portuguese Conference on Artificial Intelligence (EPIA'09)*, Lecture Notes in Computer Science (LNAI 5816), 658–669. Springer-Verlag.

Wilson, T., and Wiebe, J. 2003. Annotating Opinions in the World Press. In *Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue*.

Wilson, T.; Wiebe, J.; and Cardie, C. 2015. MPQA Opinion Corpus. In Ide, N., and Pustejovsky, J., eds., *Handbook of Linguistic Annotation*. Springer.

Wilson, T.; Wiebe, J.; and Hoffmann, P. Recognizing Contextual Polarity: An Exploration of Features for Phrase-level Sentiment Analysis. *Computational Linguistics* 35(3):399–433.

Wilson, T.; Wiebe, J.; and Hoffmann, P. 2005. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *Joint Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing (HLT-EMNLP 2005)*, 347–354.

Wilson, T. A. 2008. *Fine-grained Subjectivity and Sentiment Analysis: Recognizing the Intensity, Polarity, and Attitudes of Private States*. Ph.D. Dissertation, University of Pittsburgh.

Winston, M. E.; Chaffin, R.; and Herrmann, D. 1987. A Taxonomy of Part-Whole Relations. *Cognitive Science* 11(4).

Yu, N., and Kübler, S. 2011. Filling the Gap: Semi-Supervised Learning for Opinion Detection across Domains. In *Proceedings of the 15th Conference on Computational Natural Language Learning (CoNLL'11)*, 200–209.