Automatically Creating a Lexicon of Verbal Polarity Shifters: Mono- and Cross-lingual Methods for German

Marc Schulder, Michael Wiegand
Spoken Language Systems
Saarland University
Germany
marc.schulder@lsv.uni-saarland.de
michael.wiegand@lsv.uni-saarland.de

Josef Ruppenhofer
Institute for German Language
Mannheim
Germany
ruppenhofer@ids-mannheim.de

Abstract

In this paper we use methods for creating a large lexicon of verbal polarity shifters and apply them to German. Polarity shifters are content words that can move the polarity of a phrase towards its opposite, such as the verb *"abandon"* in *"abandon all hope"*. This is similar to how negation words like *"not"* can influence polarity. Both shifters and negation are required for high precision sentiment analysis. Lists of negation words are available for many languages, but the only language for which a sizable lexicon of verbal polarity shifters exists is English. This lexicon was created by bootstrapping a sample of annotated verbs with a supervised classifier that uses a set of data- and resource-driven features. We reproduce and adapt this approach to create a German lexicon of verbal polarity shifters. Thereby, we confirm that the approach works for multiple languages. We further improve classification by leveraging cross-lingual information from the English shifter lexicon. Using this improved approach, we bootstrap a large number of German verbal polarity shifters, reducing the annotation effort drastically. The resulting German lexicon of verbal polarity shifters is made publicly available.

Title and Abstract in German

Die automatische Erstellung eines Lexikons polaritätsverschiebender Verben:
Einsprachige und sprachübergreifende Methoden für das Deutsche

In dieser Arbeit untersuchen wir Methoden zur Erstellung eines deutschsprachigen Lexikons polaritätsverschiebender Verben. Diese Verben, die vielfach auch Polaritätsshifter genannt werden, sind Inhaltswörter, die die Polarität einer Phrase zu ihrem entgegengesetzten Wert verschieben, wie z.B. das Verb *„aufgeben"* in der Verbalphrase *„alle Hoffnung aufgeben"*. Das Verhalten von Polaritätsshiftern ähnelt somit dem von Negationswörtern wie *„nicht"*. Für robuste Sentimentanalyse werden sowohl Negationswörter als auch Polaritätsshifter benötigt. Während Listen von Negationswörtern in vielen Sprachen verfügbar sind, existiert jedoch ein Polaritätsshifter-Lexikon hinreichender Größe nur für das Englische. Jene Ressource wurde mittels Bootstrapping erzeugt, indem ein überwachter Klassifikator auf einer kleinen Stichprobe von Verben trainiert wurde. Dieser Klassifikator nutzt Daten- und Ressourcen-getriebene Merkmale. Wir reproduzieren diesen Ansatz und passen ihn soweit notwendig für das Deutsche an. Somit weisen wir die Übertragbarkeit dieses Ansatzes auf andere Sprachen nach. Wir verbessern die Qualität der Klassifikation zudem weiterhin, indem wir Informationen aus dem existierenden englischen Polaritätsshifter-Lexikon nutzten. Mittels dieses verbesserten Ansatzes finden wir per Bootstrapping eine große Anzahl deutscher Polaritätsshifter und verringern somit deutlich den manuellen Annotationsaufwand. Das resultierende deutsche Lexikon polaritätsverschiebender Verben ist frei verfügbar.

Proceedings of the 27th International Conference on Computational Linguistics, pages 2516–2528
Santa Fe, New Mexico, USA, August 20-26, 2018.

1 Introduction

Polarity shifters are content words such as verbs, nouns or adjectives that influence the sentiment polarity of an expression in ways similar to negation words. For example, the negated statement in (1) that uses the negation word *nicht* in German and *not* in English can also be expressed using the verbal shifter *unterlassen* in German and *fail* in English, as seen in (2).

(1) Peter hat ihnen **nicht** geholfen.
Peter did **not** help them.
(2) Peter hat es **unterlassen**$_{\textit{shifter}}$ ihnen zu helfen.
Peter **failed**$_{\textit{shifter}}$ to help them.

Polarity shifters can affect both positive and negative expressions, moving their polarity towards the opposite polarity. In (3) the shifter *verweigern/deny* affects the positive polar expression *Stipendium/scholarship*, resulting in a negative polarity for the sentence. On the other hand, the shifter *lindern/alleviate* in (4) creates a positive sentence despite the negative polar expression *Schmerz/pain*.

(3) Ihr wurde das [[Stipendium]$^+$ **verweigert**$_{\textit{shifter}}$]$^-$.
She was [**denied**$_{\textit{shifter}}$ the [scholarship]$^+$]$^-$.
(4) Die neue Behandlung hat ihre [[Schmerzen]$^-$ **gelindert**$_{\textit{shifter}}$]$^+$.
The new treatment has [**alleviated**$_{\textit{shifter}}$ her [pain]$^-$]$^+$.

As can be seen for *verhindern/prevent* in (5) and (6), the same shifter can even affect both positive and negative expressions.

(5) Seine Prinzipien [**verhinderten**$_{\textit{shifter}}$ eine [Einigung]$^+$]$^-$.
His principles [**prevented**$_{\textit{shifter}}$ an [agreement]$^+$]$^-$.
(6) Ihre Maßnahmen [**verhinderten**$_{\textit{shifter}}$ ein [Gemetzel]$^-$]$^+$.
Their measures [**prevented**$_{\textit{shifter}}$ a [slaughter]$^-$]$^+$.

We present a **reproduction** and **extension** to the work of Schulder et al. (2017), which introduced a **lexicon of verbal polarity shifters**, as well as methods to increase the size of this lexicon through bootstrapping. The lexicon lists verb lemmas and assigns a binary label (*shifter* or *no shifter*) to each. The original approach was developed on English. We apply it to **German**, validating the generality of the approach and creating a new resource, a German lexicon of 677 verbal polarity shifters. We also improve the bootstrapping process by adding features that leverage polarity shifter resources across languages.

As is the case with negation, modeling polarity shifting is important for various tasks in NLP, such as relation extraction (Sanchez-Graillet and Poesio, 2007), recognition of textual entailment (Harabagiu et al., 2006) and especially sentiment analysis (Wiegand et al., 2010). However, while there has been significant research on negation in sentiment analysis (Wiegand et al., 2010), current classifiers fail to handle polarity shifters adequately (Schulder et al., 2017). This is in part due to the lack of lexical resources for polarity shifters. Unlike negation words (*no*, *not*, *never*, etc.), of which there are only a few dozen in a language, polarity shifters are far more numerous. Among verbs alone there are many hundreds (Schulder et al., 2017). Comprehensive shifter lexicons are, therefore, considerably more expensive to create. Once available, they can be used to improve the aforementioned tasks, as has already been shown for the case of English polarity classification (Schulder et al., 2017).

To reduce the cost of creating such polarity shifter lexicons, Schulder et al. (2017) introduced methods to automatically generate a labeled list of words using either a limited amount of labeled training data or no labeled data at all. Their approach includes both features that rely on semantic resources and data-driven ones. They limited their work to English verbs, but expressed the expectation that their methods should also work for other languages. To verify that expectation, we apply their approach to German, for which all resources required to reproduce their experiments are available. Keeping in mind that this is not the case for many other languages, we focus our evaluation on differentiating between features that rely on unstructured data and those requiring rare semantic resources.

While polarity shifters are not restricted to a particular part of speech – shifter nouns (e.g. *downfall*), adjectives (*devoid*) and adverbs (*barely*) also exist – we limit ourselves to verbs. Verbs and nouns are the most important minimal semantic units (Schneider et al., 2016) and verbs are usually the main syntactic predicates of clauses, projecting far-reaching scopes. Focusing on verbs also allows us a closer comparison with Schulder et al. (2017) and to investigate cross-lingual similarities between verbal shifters.

The **contributions** of this paper are:

(i) we introduce a German lexicon of verbal polarity shifters;
(ii) we reproduce and adapt the approach of Schulder et al. (2017) to German to extend our lexicon;
(iii) we introduce additional methods that take advantage of the existence of the English verbal polarity shifter lexicon and improve upon the current state of the art.

The focus of our work is the binary classification of verbal polarity shifters in German. The resulting German lexicon of 677 verbal polarity shifters is made **publicly available**.[1]

2 Related Work

Existing work on negation modeling focuses almost exclusively on negation words (see the survey of Wiegand et al. (2010)). One reason for this is the lack of lexicons and corpora that cover other forms of polarity shifters. Even the most complex negation lexicon for English sentiment analysis (Wilson et al., 2005) includes a mere 12 verbal shifters. So far the only larger resources for polarity shifters are the English-language verbal shifter lexicons recently introduced by Schulder et al. (2017) and Schulder et al. (2018). Schulder et al. (2017) automatically bootstrap a lexicon which covers 980 verbal shifters at the lemma level, while Schulder et al. (2018) manually annotate word senses of verbs, creating a lexicon of 2131 shifter senses across 1220 verbs. As we reproduce and extend the work of Schulder et al. (2017), all further use of and comparison to an English shifter lexicon refers to their bootstrapped lexicon as well.

To create shifter lexicons at a large scale, automation and bootstrapping techniques are required. Danescu-Niculescu-Mizil et al. (2009) propose using *negative polarity items (NPIs)* to extract downward-entailing operators, which are closely related to polarity shifters. Schulder et al. (2017) also make use of NPIs in addition to a number of other features.

Rather than using lexicons, another approach would be to learn polarity shifters from labelled corpora. In the case of negation, this has already been examined for the biomedical domain (Huang and Lowe, 2007; Morante and Daelemans, 2009; Zou et al., 2013), the review domain (Ikeda et al., 2008; Kessler and Schütze, 2012; Socher et al., 2013; Yu et al., 2016) and across domains (Fancellu et al., 2016). Unfortunately, due to the considerably higher lexical diversity of polarity shifters, far larger corpora would be required for learning shifter than for learning negation.

Available corpora that are suitable for negation learning, such as the Sentiment Treebank (Socher et al., 2013) or the BioScope corpus (Szarvas et al., 2008), are fairly small in size. Most verbs occur in them in very few instances or not at all. In the BioScope corpus, for example, there are only 6 verbal shifters (Morante, 2010). Polarity classifiers trained on such corpora, such as the state-of-the-art Recursive Neural Tensor Network tagger (Socher et al., 2013), fail to detect many instances of polarity shifting. Schulder et al. (2017) show that the explicit knowledge provided by a shifter lexicon can improve polarity classification in such cases.

3 Data

We create a gold standard for German verbal shifters, following the approach Schulder et al. (2017) used for their English gold standard. An expert annotator, who is a native speaker of German, labeled 2000 verbs, randomly sampled from GermaNet (Hamp and Feldweg, 1997), a German wordnet resource. The remaining 7262 GermaNet verbs are used to bootstrap a larger lexicon in §5.3.

Each verb is assigned a binary label of being a shifter or not. To qualify as a shifter, a verb must permit polar expressions as its dependents and cause the polarity of the expression that embeds both verb and polar expression to move towards the opposite of the polar expression. For example, in (6) *verhindern* shifts the negative polarity of its dependent *ein Gemetzel*, resulting in a positive expression. Annotation is performed at the lemma level, as word-sense disambiguation tends to be insufficiently robust.

[1] `https://github.com/uds-lsv/coling2018`

Resource Type	German Resource	English Resource
Wordnet	GermaNet (Hamp and Feldweg, 1997)	WordNet (Miller et al., 1990)
Text Corpus	DeWaC Web Corpus (Baroni et al., 2009)	Amazon Product Reviews (Jindal and Liu, 2008)
Polarity Lexicon	PolArt Sentiment Lexicon (Klenner et al., 2009)	Subjectivity Lexicon (Wilson et al., 2005)
Framenet	Salsa (Burchardt et al., 2006)	FrameNet (Baker et al., 1998)
Effects	EffektGermaNet (Ruppenhofer and Brandes, 2015)	EffectWordNet (Choi et al., 2014)

Table 1: Required German resources, compared with English resources used by Schulder et al. (2017).

	Frequency	Percentage
shifter	224	11.2
no shifter	1776	88.8

Table 2: Distribution of verbal shifters in annotated sample of 2000 verbs taken from GermaNet.

	Polar Verbs		Positive V.		Negative V.	
	Freq	**%**	**Freq**	**%**	**Freq**	**%**
shifter	81	23.1	12	11.7	69	27.9
no shifter	269	76.9	91	88.3	178	72.1

Table 3: Distribution of verbal shifters in the *PolArt Sentiment Lexicon* (Klenner et al., 2009).

Table 1 provides an overview of the German resources we use in our reproduction, compared to the resources used for the English shifter lexicon. More detailed descriptions of the resources are provided in sections discussing feature design (§4) and experiments (§5).

Table 2 shows that in our gold data 11.2% of verbs are shifters, which is a bit less than the 15.2% of the English gold standard. Table 3 shows the shifter distribution among verbs with sentiment polarity (determined using the *PolArt Sentiment Lexicon* (Klenner et al., 2009)). As was the case for the English gold data, it shows a tendency for shifter verbs to be negative rather than positive terms.

4 Feature Design

In this section we introduce the features that we will use to bootstrap our German verbal shifter lexicon in §5.3. We start by outlining the features proposed by Schulder et al. (2017) and how we adapt them for use with German (§4.1). We further separate them into data-driven features (§4.1.1) and resource-driven features (§4.1.2) to highlight their requirements when applied to a new language.

In §4.2 we introduce new methods that can either be used as stand-alone classifiers or as features for an SVM classifier. Both methods take advantage of existing knowledge about English verbal shifters. One method uses a bilingual dictionary (§4.2.1) and the other cross-lingual word embeddings (§4.2.2).

4.1 Feature Reproduction

In this section we briefly describe how we adapt the features of Schulder et al. (2017) to German language data. We distinguish between features that mainly rely on text data from a corpus (§4.1.1) and those that require complex semantic resources (§4.1.2). When working with languages with scarcer resources, it can be expected that the former will be more readily available than the latter.

4.1.1 Data-driven Features

The main requirement of the following features is a reasonably sized text corpus to detect syntactic patterns and word frequencies. The text corpus was lemmatized using the TreeTagger (Schmid, 1994) and parsed for syntactic dependency structures with ParZu (Sennrich et al., 2009).[2] For features requiring knowledge of polarities we use the *PolArt Sentiment Lexicon* (Klenner et al., 2009).[3]

[2]Lacking an appropriate parser, a part-of-speech tagger may approximate required syntactic structures (Riloff et al., 2013).

[3] We chose to consider features that use a polarity lexicon to still be data-driven features as there exist robust methods to generate them automatically from unlabeled corpora (Turney, 2002; Velikovich et al., 2010; Hamilton et al., 2016). The lexicon we use was created using bootstrapping (Clematide and Klenner, 2010).

Distributional Similarity (SIM): The distributional similarity feature assumes that words that are semantically similar to negation words are also likely to be polarity shifters. Semantic similarity is modeled as cosine similarity in a word embedding space. The word embeddings are created using Word2Vec (Mikolov et al., 2013) on the German web corpus DeWaC (Baroni et al., 2009), using the same hyperparameters as Schulder et al. (2017) and German translations of their negation seeds.

Polarity Clash (CLASH): The polarity clash feature assesses that shifting will often occur when a polar verb modifies an expression of the opposite polarity, such as in (7). The feature is further narrowed down to negative verbs that modify positive nouns, as polar verbal shifters are predominantly of negative polarity (Table 3).

(7) Er hat die [[Hoffnung]$^+$ [**verloren**]$^-$]$^-$.
He [[**lost**]$^-$ [hope]$^+$]$^-$.

Particle Verbs (PRT): Certain verb particles indicate a complete transition to an end state (Brinton, 1985). Schulder et al. (2017) hypothesize that this phenomenon correlates with shifting, which can be seen as producing a new (negative) end state. Therefore, they collect particle verbs containing relevant English particles, such as *away*, *down* and *out*. For our German data we chose the following particles associated with negative end states: *ab*, *aus*, *entgegen*, *fort*, *herunter*, *hinunter*, *weg* and *wider*.

Heuristic using 'jeglich' (ANY): *Negative polarity items (NPIs)* are known to occur in the context of negation (Giannakidou, 2008). Schulder et al. (2017) showed that the English NPI *any* co-occurs with shifters, so its presence in a verb phrase can indicate the presence of a verbal shifter. We expect the same for the German NPI *jeglich*, as seen in (8). We collect all verbs with a polar direct object that is modified by the lemma *jeglich*. The resulting pattern matches are sorted by their frequency, normalized over their respective verb frequency and then reranked using *Personalised PageRank* (Agirre and Soroa, 2009).

(8) Sie [**verwehrten**$_{shifter}$ uns jegliche [Hilfe$_{dobj}$]$^+$]$^-$.
They [**denied**$_{shifter}$ us any [help$_{dobj}$]$^+$]$^-$.

Anti-Shifter Feature (ANTI): This feature specifically targets anti-shifters, verbs that exhibit polar stability instead of causing polar shifting. These are commonly verbs indicating creation or continued existence, such as *live*, *introduce*, *construct* or *prepare*. Such verbs often co-occur with the adverbs *ausschließlich*, *zuerst*, *neu* and *extra*, as seen in (9)–(12). Accordingly, we can create a list of anti-shifters by selecting the verbs that most often co-occur with these adverbs.

(9) Im Winter **leben**$_{antiShifter}$ Schwarzbären ausschließlich von Fisch.
In winter, black bears exclusively **live**$_{antiShifter}$ on fish.

(10) Komplette Tastaturen auf Handys wurden zuerst in 1997 **eingeführt**$_{antiShifter}$.
Full keyboards on cellphones were first **introduced**$_{antiShifter}$ in 1997.

(11) Diese Gebäude wurden neu **gebaut**$_{antiShifter}$.
These buildings have been newly **constructed**$_{antiShifter}$.

(12) Sie haben extra für mich veganes Essen **zubereitet**$_{antiShifter}$.
They specially **prepared**$_{antiShifter}$ vegan dishes for me.

4.1.2 Resource-driven Features

The following features rely on advanced semantic resources which are available in only a few languages.

GermaNet: Wordnets are large lexical ontologies providing various kinds of semantic information and relations. Schulder et al. (2017) used glosses, hypernyms and supersenses taken from the English WordNet (Miller et al., 1990) as features in their work. We use GermaNet (Hamp and Feldweg, 1997), a German wordnet resource that provides all these features. In the case of glosses, called paraphrases in GermaNet, GermaNet offers two variations: the paraphrases originally written for GermaNet, and a more extensive set of paraphrases harvested from Wiktionary (Henrich et al., 2014). To improve coverage we use this paraphrase extension in our experiments.

Salsa FrameNet: Framenets provide semantic frames that group words with similar semantic behavior. Schulder et al. (2017) use the frame memberships of verbs as a feature, hypothesizing that verbal shifters will be found in the same frames. We reproduce this feature using frames from the German FrameNet project *Salsa* (Burchardt et al., 2006).

EffektGermaNet: Wiebe and colleagues (Deng et al., 2013; Choi et al., 2014) introduced the idea that events can have harmful or beneficial *effects* on their objects. These *effects* are related but not identical to polarity shifting. Choi et al. (2014) provide lexical information on *effects* in their English resource EffectWordNet. We use its German counterpart, EffektGermaNet (Ruppenhofer and Brandes, 2015), to model the *effect* feature in our data.

4.2 New Features

In §4.1 we described how we reproduce features already used for English shifter classification. Next we introduce new features that have not yet been used for the creation of a verbal shifter lexicon.

4.2.1 Bilingual Dictionary

The motivation behind the work of Schulder et al. (2017) was to introduce a large lexicon of verbal polarity shifters. Now that such a lexicon exists for English, it is an obvious resource to use when creating verbal shifter lexicons for other languages. We hypothesize that a verb with the same meaning as an English verbal shifter will also function as a shifter in its own language. All that is required is a mapping from English verbs to, in our case, German verbs. We choose to use the bootstrapped lexicon of Schulder et al. (2017), rather than the manually created one of Schulder et al. (2018), to show that bootstrapping is sufficient for all stages of the learning process.

One potential source for such a mapping is a bilingual dictionary. We use the English-German dataset by DictCC[4], as it is large (over one million translation pairs) and publicly available. It covers 76% of German verbs found in GermaNet and 77% of English verbs found in WordNet.

Mapping the shifter labels of the English verbs to German verbs is performed as follows: For each German verb, all possible English translations are looked up. Using the English verbal shifter lexicon, we confirm whether the English translations are shifters. If the majority of translations are shifters, the German word is also labeled as a shifter, otherwise as not a shifter. This approach provides explicit labels for 1368 of our 2000 gold standard verbs (68%). Less than 6% of these are tied between *shifter* and *no shifter* translations. Ties are resolved in favor of the *shifter* label. The remaining verbs are labeled with the majority label *no shifter*.

While this bilingual dictionary mapping approach makes for a promising feature, we refrain from considering it for generating a gold standard. Using a dictionary instead of annotating a random sample would introduce biases existing in the dictionary, e.g. more translation pairs being available for frequent words, which can in turn favor features that work better for frequent words. Schulder et al. (2017) also observe in their error analysis that some verbs act as shifters in only some of their word senses. As different word senses often do not translate into the same foreign word, indiscriminate translation may introduce non-shifting senses of English shifter words as false positives. Evaluating the dictionary mapping as a feature will allow us to judge its usefulness for high-precision lexicon induction in future works.

4.2.2 Cross-lingual Word Embeddings

As an alternative to using bilingual dictionaries we investigate transferring English shifter labels to German using cross-lingual word embeddings. These are word embeddings which provide a shared vector space for words from multiple languages. Similar to how the SIM feature (see §4.1.1) compares negation words to verbs in a mono-lingual word embedding, a cross-lingual word embedding allows us to compare English verbs to verbs of another language based on their distributional similarity without having labeled data for the other language. These comparisons can then be used to apply the labels of the English lexicon of verbal shifters to the other language.

Mapping shifter labels cross-lingually with a bilingual dictionary, as described in §4.2.1, requires a dictionary with good coverage for both languages. For many languages, publicly available dictionaries of adequate size are hard to come by. For instance, the second largest English dictionary on DictCC is only 40% the size of the English-German dataset and only a few others have more than 2% its size.

[4]https://www.dict.cc

In §5.2 we explore the effect of dictionary size on mapping performance and how cross-lingual word embeddings fare in comparison.

Methods for creating cross-lingual word embeddings can be grouped into *cross-lingual training* and *monolingual mappings*. *Cross-lingual training* learns joint embeddings from parallel corpora. However, such corpora are far smaller and rarer than monolingual corpora and, therefore, not ideal for us.[5]

Monolingual mappings take preexisting monolingual word embeddings and learn linear transformations to map both embeddings onto the same vector space. Commonly, these approaches use bilingual dictionaries to initialize this mapping, which would rather defeat our goal of using embeddings as a data-driven alternative to dictionaries. The *VecMap* framework (Artetxe et al., 2017) provides an initialization method that relies on numerals instead of a dictionary. The idea behind this is that Arabic numerals are used in most languages, even across different writing systems (e.g. Cyrillic, Chinese, etc.), and, therefore, can function as a dictionary without requiring actual bilingual knowledge.

For our experiments, we train Word2Vec word embeddings for English and German, using the Amazon Product Review (Jindal and Liu, 2008) and DeWaC (Baroni et al., 2009) corpora, respectively. Ideally, product review corpora would be used for both languages, but available German review corpora are considerably smaller than their English counterparts. For example, the German corpus Webis-CLS (Prettenhofer and Stein, 2010) contains only 33 million words, while the English-language Amazon Product Review Corpus consists of 1.2 billion words. When generating word embeddings, the size of the corpus is very important for the quality of the resulting embedding, so we choose instead to use DeWaC, a web corpus of 1.7 billion words.

Training is performed using the same hyperparameters as used by Artetxe et al. (2017).[6] We use VecMap to create a cross-lingual word embedding using the default configuration for numeral-based mappings. The resulting cross-lingual embedding covers 79% of German GermaNet verbs as well as 79% of English WordNet verbs. It covers 1598 of our 2000 gold data verbs (80%).

We use this new word embedding to apply English shifter labels to German. To achieve this, we go through our list of German verbs, look up the most similar English verb for each and apply its label. We also investigated majority voting using k-nearest neighbors, but this did not improve performance.

5 Experiments

5.1 Classifier Evaluation

We start our evaluation by reproducing the classifier evaluation of Schulder et al. (2017). The task is the classification of all verbs from the given gold standard in a 10-fold cross validation.

Analogous to Schulder et al. (2017) we evaluate a supervised SVM classifier as well as a graph-based label propagation (LP) classifier that requires no labeled training data. In addition, we evaluate our cross-lingual word embedding classifier (§4.2.2) and our dictionary classifier (§4.2.1), which both make use of the pre-existing English lexicon, but require no additional labeled German data. For an overview of the classifiers and their data requirements, see Table 4.

For the LP classifier we use the ANY features as seeds for the positive label (*shifter)* and the ANTI feature as negative label (*no shifter)* seeds. For SVM we group features into data-driven and resource-driven feature sets (see Table 6) as outlined in §4.1.1 and §4.1.2, as well as introducing the outputs of the cross-lingual word embedding and dictionary classifiers as additional separate features.

Table 5 shows the performance of our various classifiers. All classifiers clearly outperform the baseline[7] and resource-based features outperform data-based ones. This is similar to performance observed

[5]BilBOWA (Gouws et al., 2015) seeks to improve the coverage problem of parallel corpora by incorporating additional monolingual corpora into the training process. However, our experiments with it did not provide satisfactory results. This is in line with reports by Artetxe et al. (2017) and Upadhyay et al. (2016).

[6]Word2Vec configuration: CBOW, 300 dimensions, context window of 5 words, sub-sampling at $1e-05$, negative samples at 10 and vocabulary restricted to the 200,000 most frequent words. We also experimented with using the full vocabulary, but this resulted in lower quality embeddings.

[7]As in Schulder et al. (2017), accuracy proves to be a problematic measure, as it has a strong majority label bias. The *no shifter* label makes up 88.8% of our gold annotation (Table 2), which explains the strong performance of the majority baseline on this metric.

Classifier	Features	Shifter Lex	Text Corpus	Training Data
SIM	Data-driven	—	German	—
$LP_{ANY+ANTI}$	Data-driven	—	German	—
Cross-ling. Embedding	—	English	German, English	—
Dictionary	Bilingual Dictionary	English	—	—
$SVM_{data+resource}$	Data-driven, Resource-driven	—	German	German

Table 4: Classifiers used in Table 5 and their resource requirements.

Classifier	Acc	Prec	Rec	F1
$Baseline_{majority}$	88.1	44.4	50.0	47.0
SIM	70.7	58.0	67.6	62.4
$LP_{ANY+ANTI}$	87.1	67.2	65.0	66.1
Cross-lingual Embedding	85.1	67.6	74.6	$70.9^{*\dagger}$
Dictionary	86.5	69.2	77.3	$73.0^{*\dagger}$
SVM_{data}	74.6	60.8	72.6	66.2
$SVM_{resource}$	91.3	79.4	73.9	$76.4^{*\dagger}$
$SVM_{data+resource}$	91.4	79.0	76.7	$77.7^{*\circ\dagger}$
$SVM_{data+resource+embed}$	91.6	79.6	78.9	$79.2^{*\circ\dagger}$
$SVM_{data+resource+dict}$	91.3	78.0	80.9	$79.4^{*\circ\dagger}$
$SVM_{data+resource+dict+embed}$	**92.1**	**80.3**	**82.0**	$\mathbf{81.0}^{*\circ\dagger\ddagger}$

statistical significance (paired t-test with $p < 0.05$):
* better than LP; $^\circ$ better than Dictionary; † better than SVM_{data}; ‡ better than $SVM_{data+resource}$

Table 5: Evaluation of classification (§5.1) on the 2000 verb gold standard (Table 2). Precision, recall and f-score are macro-averages.

Group	Features
data	$LP_{ANY+ANTI}$, SIM, CLASH, PRT
resource	GermaNet, Salsa, EffektGermaNet
embed	Cross-lingual Embedding
dict	Dictionary

Table 6: Features included in SVM feature groups in Table 5. All features in *data* and *resource* were also used in Schulder et al. (2017).

for English (Schulder et al., 2017). Cross-lingual embeddings and dictionaries as stand-alone classifiers both outperform the label propagation approach due to their better recall coverage of shifters.

Interestingly, the cross-lingual embedding classifier performs far better than SIM, despite both relying on word embeddings to judge distributional similarity. Comparing similarity among verbs, even cross-lingually, works better than across parts-of-speech, as required for negation-shifter comparisons.

Adding both cross-lingual features to the SVM classifier improves performance further. This shows that they are not only complementary to the existing features, but also to each other, as using only one cross-lingual feature does not improve performance as much. The most feature-rich SVM configuration, $SVM_{data+resource+dict+embed}$, provides a significant improvement over $SVM_{data+resource}$, the best classifier of Schulder et al. (2017). We conclude that cross-lingual shifter information is useful even when the same bootstrapping process and feature set is used in both the source and target language.

Figure 1 shows the learning curve of select SVM configurations, compared to the classifiers that work without labeled German data, i.e. LP, Embedding and Dictionary. Cross-lingual embedding and dictionary classifiers provide a stronger baseline than LP, outperforming $SVM_{data+resource}$ when training data is sparse. However, adding them as features to the SVM results in a classifier that consistently improves upon all other systems, even at small training sizes of only 20%. Combining all available sources of information as SVM features is therefore the preferred approach if any amount of training data is available.

5.2 Evaluation of Dictionary Size

The dictionary mapping approach (§4.2.1) has been shown to be a strong stand-alone classifier and SVM feature (Table 5), slightly outperforming the cross-lingual word embedding approach (§4.2.2). However, the underlying English-German dictionary by DictCC is of considerable size, consisting of over 1.1 million translation pairs. Even then, almost a quarter of WordNet and GermaNet verbs are not

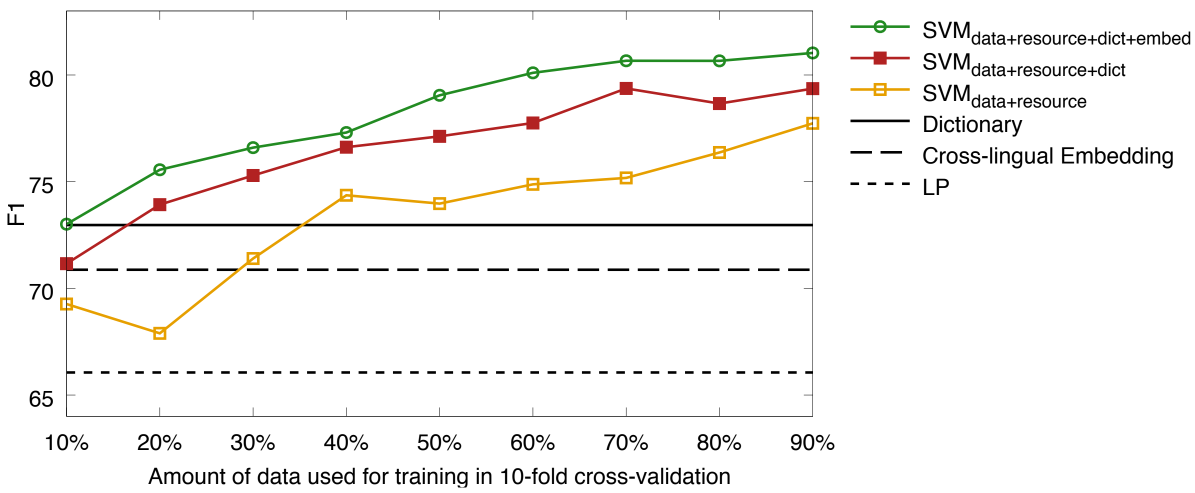

Figure 1: Learning curve on gold standard. $\text{SVM}_{\text{data+resource}}$ represents the previously best classifier (Schulder et al., 2017).

covered. For many other languages, finding a publicly available dictionary of comparable size may pose a challenge. Therefore, we investigate how smaller dictionaries may perform in our classifiers.

The English-German DictCC dictionary covers slightly over 8000 of the English verbs found in WordNet. Of the 2000 German verbs in our gold standard, DictCC covers 1368. To simulate bilingual dictionaries of smaller size, we create a version of the DictCC dictionary with half the English vocabulary by limiting it to the 4000 most frequent verbs from WordNet ($\text{Dict}_{\text{voc\_size=4k}}$). We also create even smaller versions with only the 1000 ($\text{Dict}_{\text{voc\_size=1k}}$) and 500 most frequent English verbs ($\text{Dict}_{\text{voc\_size=0.5k}}$).

As bilingual dictionaries provide a many-to-many mapping, having half the English vocabulary does not necessarily mean that we receive only half the German translations. Many German words receive multiple translations, all of which we then use to determine their shifter label via majority vote. Reducing the English vocabulary, therefore, first reduces the number of label votes for each German word, until, eventually, German words are removed as there are no more votes for them. Having fewer votes per German output label can, however, still affect the robustness of the labeling process. In our case, reducing the English vocabulary by half still provides translations for 1168 of German words in our gold data, i.e. 85% of the full dictionary. Reducing it further to 1000 English verbs drops the size of the German vocabulary to 52%. Using only the 500 most frequent English words leaves a German coverage of 33%.

Figure 2 shows the performance of the differently sized dictionaries as stand-alone classifiers, while Figure 3 shows how much they can improve the best classifier of Schulder et al. (2017), i.e. $\text{SVM}_{\text{data+resource}}$. In both cases we see that while even smaller dictionaries can still provide acceptable performance, using cross-lingual embeddings is preferable to using a dictionary of insufficient size.

5.3 Bootstrapping

In their intrinsic evaluation Schulder et al. (2017) showed that explicit knowledge of a large number of polarity shifters can improve sentiment analysis. To increase the size of our lexicon, we bootstrap additional shifters following their approach. We train our best classifier (Table 5) on the 2000 verbs from our gold standard (§3) and then use it to classify the remaining 7262 GermaNet verbs that had not been labeled so far. Of these, the classifier labels 595 verbs as shifters. A German native speaker manually checks these predicted shifters and confirms 453 to be true verbal shifters. Limiting our annotation effort to predicted shifters and discarding all others reduces the cost of annotation by 92%.

Table 7 shows the classifier precision at different confidence intervals. Like Schulder et al. (2017), we see very high performance for the first quartile, matching their observation that manual confirmation is not strictly necessary for high confidence labels. Combining the 453 bootstrapped shifters with the 224 shifters from the gold standard we produce a **novel list of 677 German verbal shifters** *(see footnote 1)*.

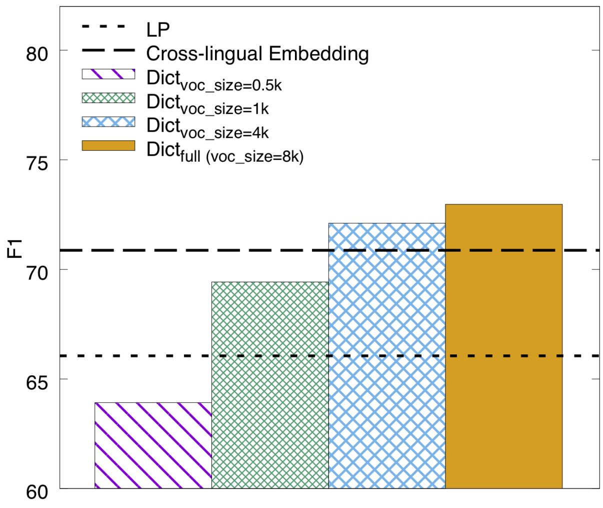

Figure 2: Comparison of dictionaries with different vocabulary sizes. Classifiers use *no labeled training data*. $\text{Dict}_{\text{full}}$ is equivalent to the dictionary shown in Table 5.

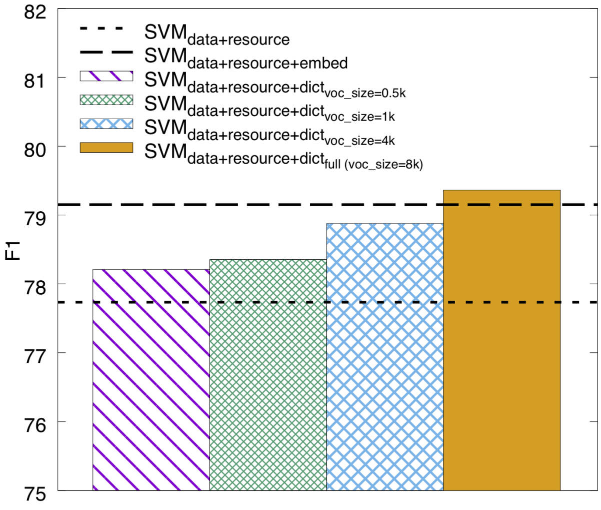

Figure 3: Comparison of SVM classifiers using dictionaries with different vocabulary sizes. $\textit{dict}_{\textit{full}}$ is equivalent to the SVM feature *dict* in Table 5.

Confidence Rank	1–150	151–300	301–450	451–595
Precision	97.3	83.3	70.0	53.1

Table 7: Classification of GermaNet verbs that were ***not*** part of gold standard (§3); verbs are ranked by confidence-score of classifier and evaluated at intervals by precision of *shifter* label.

6 Conclusion

We confirm that the bootstrapping process for creating a large lexicon of verbal polarity shifters can successfully be applied to German. Given appropriate resources, the effort for adjusting to a new language is minimal, mostly requiring translating seed words and adjusting syntactic patterns, while the underlying concepts of the features remain the same. Using a manually annotated sample of 2000 verbs taken from GermaNet, we train a supervised classifier with various data- and resource-driven features. Its performance is further improved by leveraging information from an existing English lexicon of verbal shifters using bilingual dictionaries and cross-lingual word embeddings. The resulting improved classifier allows us to triple the number of confirmed German shifters in our lexicon.

We differentiate features by whether they require only unlabeled data and basic linguistic tools or whether they depend on rare semantic resources that may not be available for many languages. In addition, we introduce the possibility of using cross-lingual resources to reduce the dependence on resources in the target language. This shows promise, improving performance for both unsupervised and supervised classification, especially for scenarios where only small amounts of training data are available. However, supervised learning that combines all features still provides the best results.

Our recommendation for creating shifter lexicons in new languages is to start out with cross-lingual label transfer, but to also invest in annotating a random sample of verbs if possible, especially if advanced semantic resources like a wordnet are available, as they require supervised learning to be leveraged.

In reproducing the work of Schulder et al. (2017), we limited ourselves to verbs. In the future, we would like to investigate methods to extend the shifter lexicon to also cover nouns and adjectives.

While we have shown that the same approach for classifying verbal shifters works for German and English, future work will expand the number of languages, especially to verify that these methods can also be applied to non-Indo-European languages, such as Chinese, Japanese or Arabic. In this context it will also be interesting to see whether using shifter lexicons from several languages can further improve the dictionary and cross-lingual word embedding classifiers.

Acknowledgements

The authors would like to thank Stephanie Köser for annotating the German gold standard lexicon presented in this paper. For proofreading the paper the authors would also like to thank Meaghan Fowlie and David M. Howcroft.

The authors were partially supported by the German Research Foundation (DFG) under grants RU 1873/2-1 and WI 4204/2-1.

References

Eneko Agirre and Aitor Soroa. 2009. Personalizing PageRank for Word Sense Disambiguation. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 33–41, Athens, Greece.

Mikel Artetxe, Gorka Labaka, and Eneko Agirre. 2017. Learning bilingual word embeddings with (almost) no bilingual data. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 451–462, Vancouver, Canada.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, pages 86–90, Vancouver, Canada.

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetti. 2009. The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation*, 43(3):209–226.

Laurel J. Brinton. 1985. Verb Particles in English: Aspect or Aktionsart. *Studia Linguistica*, 39:157–68.

Aljoscha Burchardt, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Padó, and Manfred Pinkal. 2006. The SALSA Corpus: a German corpus resource for lexical semantics. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, pages 969–974, Genoa, Italy.

Yoonjung Choi, Lingjia Deng, and Janyce Wiebe. 2014. Lexical Acquisition for Opinion Inference: A Sense-Level Lexicon of Benefactive and Malefactive Events. In *Proceedings of the Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA)*, pages 107–112, Baltimore, Maryland, USA.

Simon Clematide and Manfred Klenner. 2010. Evaluation and Extension of a Polarity Lexicon for German. In *Proceedings of the Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA)*, pages 7–13, Lisbon, Portugal.

Cristian Danescu-Niculescu-Mizil, Lillian Lee, and Richard Ducott. 2009. Without a 'doubt'? Unsupervised Discovery of Downward-Entailing Operators. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (HLT/NAACL)*, pages 137–145, Boulder, Colorado, USA.

Lingjia Deng, Yoonjung Choi, and Janyce Wiebe. 2013. Benefactive/Malefactive Event and Writer Attitude Annotation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 120–125, Sofia, Bulgaria.

Federico Fancellu, Adam Lopez, and Bonnie Webber. 2016. Neural Networks for Negation Scope Detection. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 495–504, Berlin, Germany.

Anastasia Giannakidou. 2008. Negative and Positive Polarity Items: Licensing, Compositionality and Variation. In Claudia Maienborn, Klaus von Heusinger, and Paul Portner, editors, *Semantics: An International Handbook of Natural Language Meaning*, pages 1660–1712. Mouton de Gruyter.

Stephan Gouws, Yoshua Bengio, and Greg Corrado. 2015. BilBOWA: Fast Bilingual Distributed Representations without Word Alignments. In *International Conference on Machine Learning (ICML)*, pages 748–756, Lille, France.

William L Hamilton, Kevin Clark, Jure Leskovec, and Dan Jurafsky. 2016. Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 595–605, Austin, Texas, USA.

Birgit Hamp and Helmut Feldweg. 1997. GermaNet - a Lexical-Semantic Net for German. In *Proceedings of the ACL Workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, pages 9–15, Madrid, Spain.

Sanda Harabagiu, Andrew Hickl, and Finley Lacatusu. 2006. Negation, Contrast and Contradiction in Text Processing. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 755–762, Boston, Massachusetts, USA.

Verena Henrich, Erhard Hinrichs, and Tatiana Vodolazova. 2014. Aligning Germanet Senses with Wiktionary Sense Definitions. In Zygmunt Vetulani and Joseph Mariani, editors, *Human Language Technology Challenges for Computer Science and Linguistic (LTC)*, pages 329–342. Springer.

Y. Huang and H. J. Lowe. 2007. A Novel Hybrid Approach to Automated Negation Detection in Clinical Radiology Reports. *Journal of the American Medical Infomatics Association*, 14:304–311.

D. Ikeda, H. Takamura, L. Ratinov, and M. Okumura. 2008. Learning to Shift the Polarity of Words for Sentiment Classification. In *Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP)*, pages 296–303, Hyderabad, India.

Nitin Jindal and Bing Liu. 2008. Opinion Spam and Analysis. In *Proceedings of the International Conference on Web Search and Data Mining (WSDM)*, pages 219–230, Palo Alto, California, USA.

W. Kessler and H. Schütze. 2012. Classification of Inconsistent Sentiment Words using Syntactic Constructions. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, pages 569–578, Mumbai, India.

Manfred Klenner, Angela Fahrni, and Stefanos Petrakis. 2009. PolArt: A Robust Tool for Sentiment Analysis. In *Proceedings of the Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 235–238, Odense, Denmark.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. In *Proceedings of Workshop at International Conference on Learning Representations (ICLR)*, Scottsdale, Arizona, USA.

George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. 1990. Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*, 3:235–244.

R. Morante and W. Daelemans. 2009. A Metalearning Approach to Processing the Scope of Negation. In *Proceedings of the Conference on Computational Natural Language Learning (CoNLL)*, pages 21–29, Boulder, CO, USA.

R. Morante. 2010. Descriptive Analysis of Negation Cues in Biomedical Texts. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, pages 1429–1436, Valletta, Malta.

Peter Prettenhofer and Benno Stein. 2010. Cross-Language Text Classification using Structural Correspondence Learning. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1118–1127, Uppsala, Sweden.

Ellen Riloff, Ashequl Qadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert, and Ruihong Huang. 2013. Sarcasm as Contrast between a Positive Sentiment and Negative Situation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 704–714, Seattle, Washington, USA.

Josef Ruppenhofer and Jasper Brandes. 2015. Extending effect annotation with lexical decomposition. In *Proceedings of the Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA@EMNLP)*, pages 67–76, Lisboa, Portugal.

Olivia Sanchez-Graillet and Massimo Poesio. 2007. Negation of protein–protein interactions: analysis and extraction. *Bioinformatics*, 23(13):i424–i432.

Helmut Schmid. 1994. Probabilistic Part-of-Speech Tagging using Decision Trees. In *Proceedings of the International Conference on New Methods in Language Processing (NeMLaP)*, pages 44–49, Manchester, United Kingdom.

Nathan Schneider, Dirk Hovy, Anders Johannsen, and Marine Carpuat. 2016. SemEval-2016 Task 10: Detecting Minimal Semantic Units and their Meanings (DiMSUM). In *Proceedings of the International Workshop on Semantic Evaluation (SemEval@NAACL-HLT)*, pages 546–559, San Diego, California, USA.

Marc Schulder, Michael Wiegand, Josef Ruppenhofer, and Benjamin Roth. 2017. Towards Bootstrapping a Polarity Shifter Lexicon using Linguistic Features. In *Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP)*, pages 624–633, Taipei, Taiwan.

Marc Schulder, Michael Wiegand, Josef Ruppenhofer, and Stephanie Köser. 2018. Introducing a Lexicon of Verbal Polarity Shifters for English. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, pages 1393–1397, Miyazaki, Japan.

Rico Sennrich, Gerold Schneider, Martin Volk, and Martin Warin. 2009. A New Hybrid Dependency Parser for German. In *Proceedings of the German Society for Computational Linguistics and Language Technology (GSCL)*, pages 115–124, Potsdam, Germany.

R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts. 2013. Recursive Deep Models for Semantic Compositionality over a Sentiment Treebank. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1631–1642, Seattle, Washington, USA.

G. Szarvas, V. Vincze, R. Farkas, and J. Csirik. 2008. The BioScope Corpus: Annotation for Negation, Uncertainty and their Scope in Biomedical Texts. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing (BioNLP@ACL-HLT)*, pages 38–45, Columbus, Ohio, USA.

Peter Turney. 2002. Thumbs up or Thumbs down?: Semantic Orientation Applied to Unsupervised Classification of Reviews. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 417–424, Philadelphia, Pennsylvania, USA.

Shyam Upadhyay, Manaal Faruqui, Chris Dyer, and Dan Roth. 2016. Cross-lingual Models of Word Embeddings: An Empirical Comparison. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1661–1670, Berlin, Germany.

Leonid Velikovich, Sasha Blair-Goldensohn, Kerry Hannan, and Ryan McDonald. 2010. The Viability of Web-derived Polarity Lexicons. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (HLT/NAACL)*, pages 777–785, Los Angeles, California, USA.

Michael Wiegand, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, and Andrés Montoyo. 2010. A Survey on the Role of Negation in Sentiment Analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing (NeSp-NLP)*, pages 60–68, Uppsala, Sweden.

Theresa Wilson, Paul Hoffmann, Swapna Somasundaran, Jason Kessler, Janyce Wiebe, Yejin Choi, Claire Cardie, Ellen Riloff, and Siddarth Patwardhan. 2005. OpinionFinder: A System for Subjectivity Analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP) on Interactive Demos*, pages 34–35, Vancouver, Canada.

H. Yu, J. Hsu, M. Castellanos, and J. Han. 2016. Data-driven Contextual Valence Shifter Quantification for Multi-Theme Sentiment Analysis. In *Proceedings of the Conference on Information and Knowledge Management (CIKM)*, pages 939–948, Indianapolis, Indiana, USA.

B. Zou, G. Zhou, and Q. Zhu. 2013. Tree Kernel-based Negation and Speculation Scope Detection with Structured Syntactic Parse Features. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 968–976, Seattle, Washington, USA.