SENTIMENT POLARITY SHIFTERS

CREATING LEXICAL RESOURCES THROUGH MANUAL ANNOTATION AND BOOTSTRAPPED MACHINE LEARNING

Dissertation
zur Erlangung des akademischen Grades eines
Doktors der Philosophie
der Philosophischen Fakultät
der Universität des Saarlandes

vorgelegt von

Marc Schulder

aus Herdecke

Saarbrücken, 2019

DEKAN DER PHILOSOPHISCHEN FAKULTÄT:
Prof. Dr. Heinrich Schlange-Schöningen

ERSTBERICHTERSTATTER:
Prof. Dr. Dietrich Klakow

ZWEITBERICHTERSTATTER:
Prof. Dr. Ingo Reich

TAG DER LETZTEN PRÜFUNGSLEISTUNG:
14. August 2019

That was how it worked. No magic at all.
But that time it had been magic.
And it didn't *stop* being magic
just because you found out how it was done...

— Terry Pratchett, *The Wee Free Men*

That young girl is one of the *least*
benightedly unintelligent organic life forms
it has been my profound *lack* of pleasure
not to be able to *avoid* meeting.

— Marvin, in Douglas Adams, *Life, the Universe and Everything*

ABSTRACT

Alleviating pain is good and abandoning hope is bad. We instinctively understand how words like *alleviate* and *abandon* affect the polarity of a phrase, inverting or weakening it. When these words are content words, such as verbs, nouns and adjectives, we refer to them as *polarity shifters*. Shifters are a frequent occurrence in human language and an important part of successfully modeling negation in sentiment analysis; yet research on negation modeling has focussed almost exclusively on a small handful of closed class negation words, such as *not*, *no* and *without*. A major reason for this is that shifters are far more lexically diverse than negation words, but no resources exist to help identify them.

We seek to remedy this lack of shifter resources. Our most central step towards this is the creation of a large lexicon of polarity shifters that covers verbs, nouns and adjectives. To reduce the prohibitive cost of such a large annotation task, we develop a bootstrapping approach that combines automatic classification with human verification. This ensures the high quality of our lexicon while reducing annotation cost by over 70 percent.

In designing the bootstrap classifier we develop a variety of features which use both existing semantic resources and linguistically informed text patterns. In addition we investigate how knowledge about polarity shifters might be shared across different parts of speech, highlighting both the potential and limitations of such an approach.

The applicability of our bootstrapping approach extends beyond the creation of a single resource. We show how it can further be used to introduce polarity shifter resources for other languages. Through the example case of German we show that all our features are transferable to other languages. Keeping in mind the requirements of under-resourced languages, we also explore how well a classifier would do when relying only on data- but not resource-driven features. We also introduce ways to use cross-lingual information, leveraging the shifter resources we previously created for other languages.

Apart from the general question of which words can be polarity shifters, we also explore a number of other factors. One of these is the matter of shifting direction, which indicates whether a shifter affects positive polarities, negative polarities or whether it can shift in either direction. Using a supervised classifier we add shifting direction information to our bootstrapped lexicon.

For other aspects of polarity shifting, manual annotation is preferable to automatic classification. Not every word that can cause polarity shifting does so for every of its word senses. As word sense disam-

biguation technology is not robust enough to allow the automatic handling of such nuances, we manually create a complete sense-level annotation of verbal polarity shifters.

To verify the usefulness of the lexica which we create, we provide an extrinsic evaluation in which we apply them to a sentiment analysis task. In this task the different lexica are not only compared amongst each other, but also against a state-of-the-art compositional polarity neural network classifier that has been shown to be able to implicitly learn the negating effect of negation words from a training corpus. However, we find that the same is not true for the far more lexically diverse polarity shifters. Instead, the use of the explicit knowledge provided by our shifter lexica brings clear gains in performance.

In der Computerlinguistik befasst sich der Bereich der Sentimentanalyse mit der Erkennung und Verarbeitung von Meinungen und qualitativen Aussagen in menschlicher Sprache. Dies umfasst verschiedene Aufgaben, wie die Bestimmung der Person, die eine Meinung vertritt (*opinion holder detection*), worauf sich die Meinung bezieht (*opinion target detection*) und wie stark eine Meinung ausgeprägt ist (*intensity classification*). Die überwiegende Mehrheit der Forschung konzentriert sich jedoch darauf, die **Polarität** (auch als **Valenz** bezeichnet) eines Textes festzustellen, also, ob dieser eine positive, negative oder neutrale Meinung widerspiegelt.

Die Voraussetzung für die Bestimmung der Polarität eines Textes ist, dass die Polaritäten der individuellen Begriffe, aus denen der Text besteht, bekannt sind. In dem Satz in Beispiel (1) erlaubt uns das Wissen, dass *helfen* ein positiver Begriff ist, die Schlussfolgerung, dass *"bei dem Umzug helfen"* eine positive Phrase ist.

Die Polarität eines Ausdrucks kann auch von einer Reihe von unterschiedlichen Phänomenen beeinflusst werden. Das bestbekannte solche Phänomen ist **Negation**, vor allem in der Form von **Negationswörtern**, wie zum Beispiel *nicht*, *kein*, *weder* und *ohne*. In Beispiel (2) beeinflusst das Negationswort *nicht* die positive Polarität der Phrase *"bei dem Umzug geholfen"*, was in einer negativen Polarität des Satzes resultiert.

(1) Peter hat ihnen [bei dem Umzug [geholfen]$^{+}$]$^{+}$.

(2) Peter hat ihnen [**nicht**$_{\textit{Negation}}$ [bei dem Umzug geholfen]$^{+}$]$^{-}$.

Negationswörter sind jedoch nicht die einzigen Wörter, die die Polarität einer Phrase beeinflussen können. Viele Inhaltswörter, sogenannte **Polaritätsshifter**, können einen ähnlichen Effekt haben. Die negierte Aussage aus (2) kann zum Beispiel auch wie in (3) durch das Verb *unterlassen* ausgedrückt werden. Polaritätsshifter sind zudem nicht nur auf Verben beschränkt. Die nominalen (4) und adjektivischen Formen (5) des Wortes *unterlassen* drücken dieselbe Polaritätsverschiebung aus.

(3) Peter hat es [**unterlassen**$_{\textit{Shifter}}$, ihnen [bei dem Umzug zu helfen]$^{+}$]$^{-}$.

(4) Peters [**Unterlassung**$_{\textit{Shifter}}$ jeglicher [Hilfe bei dem Umzug]$^{+}$]$^{-}$...

(5) Peters [**unterlassene**$_{\textit{Shifter}}$ [Hilfe bei dem Umzug]$^{+}$]$^{-}$...

Genau wie Negationswörter können auch Polaritätsshifter sowohl positive Ausdrücke zu negativen phrasalen Polaritäten verschieben, wie in (6), als auch negative Ausdrücke zu positiven phrasalen Polaritäten, wie in (7).

(6) Die Ihr eintretet, [**lasset** alle [Hoffnung]$^+$ **fahren**$_{Shifter}$]$^-$.

(7) [Die [Wunde]$^-$ **pfleg**$_{Shifter}$ ich dir]$^+$ mit heilgem Kuss.

Bisland wurden Polaritätsshifter jedoch in der Computerlinguistik-Forschung größtenteils übersehen oder ignoriert.

ZIELSETZUNG

Das Ziel dieser Dissertation ist die Erstellung von Klassifikationsmethodiken und Ressourcen für die Arbeit mit Polaritätsshiftern. Der zentrale Kern dieses Unterfangens liegt in der Erstellung eines **allgemeinen Lexikons von englischen Polaritätsshiftern**.

Eine Herausforderung bei dieser Aufgabe liegt darin, den erforderlichen Annotationsaufwand in einem angemessenen Rahmen zu halten: Wie kann Annotation automatisiert werden ohne größere Qualitätsverluste? Haben verschiedene Wortarten unterschiedliche linguistische Anforderungen? Funktionieren dieselben Methoden für jede Wortart? Wie steht es mit anderen Sprachen? Kann Wissen über Shifter zwischen unterschiedlichen Wortarten und zwischen verschiedenen Sprache übertragen werden? Welche Arten von Informationen kann unser Lexikon zur Verfügung stellen? Reicht eine Annotation pro Wort aus oder sollte zwischen verschiedenen Wortbedeutungen unterschieden werden? Können nach Wortbedeutung unterschiedene Daten von computerlinguistischen Anwendungen erfolgreich genutzt werden? Welche Teile eines Satzes werden von einem bestimmten Shifter beeinflusst? Beeinflussen alle Shifter jede Polarität gleich?

Dies sind die Fragen, die im Verlaufe dieser Dissertation beantworten werden.

POLARITÄTSSHIFTER IN DER COMPUTERLINGUISTIK

Die korrekte Verarbeitung von Negation, Polaritätsshifter eingeschlossen, ist von großer Bedeutung für zahlreiche Aufgaben in der Computerlinguistik, zum Beispiel in der Relationsextraktion (Sanchez-Graillet und Poesio, 2007), in Textual Entailment Recognition (zu deutsch *Erkennung textueller Schlussfolgerungen*) (Harabagiu u. a., 2006) und insbesondere in der Sentimentanalyse (Wiegand u. a., 2010).

Die Forschung im Bereich der kompositionalen Polaritätserkennung hat sich bisher zu großen Teilen auf Negationswörter konzentriert (Wiegand u. a., 2010; Schouten und Frasincar, 2016). Ein Grund hierfür ist die Verfügbarkeit lexikalischer Ressourcen für Negationswörter und der Mangel vergleichbarer Ressourcen für Polaritätsshifter. Negationswörter sind üblicherweise Funktionswörter, von denen es nur wenige gibt. Polaritätsshifter hingegen sind Inhaltswörter (d. h. Verben, Adjektive und Substantive), welche deutlich zahlreicher sind. Die lexikalische Datenbank *WordNet* (Miller u. a., 1990) beinhaltet zum Beispiel über 10.000 Verben, 20.000 Adjektive und 110.000 Substantive für die englische Sprache. Zugleich kommen individuelle Inhaltswörter jedoch deutlich seltener vor als individuelle Funktionswörter.

Dies bedeutet, dass die Erstellung eines umfassenden Lexikons von Negationswörtern deutlich einfacher ist als die Erstellung eines Lexikons von Polaritätsshiftern, während zugleich jedes einzelne Negationswort häufiger Verwendung findet, als es ein einzelnes Shifterwort tut. Dies bedeutet jedoch nicht, dass Polaritätsshifter weniger wichtig für erfolgreiche Polaritätsklassifikation seien. Im Verlauf dieser Arbeit wird sogar gezeigt werden, dass Polaritätsshifter vermutlich häufiger Verwendung in englischer Schriftsprache finden als Negationswörter.

INHALT

Diese Dissertation ist in Englisch verfasst und unterteilt sich in drei Teile, die jeweils mehrere Kapitel enthalten. Die inhaltliche Struktur der Dissertation ist wie folgt:

Teil I: einleitung Das generelle Konzept der Polaritätsshifter wird eingeführt, sowie Hintergrundwissen, welches für die Lektüre dieser Arbeit vonnöten ist.

- **Kapitel 1: einführung** Dieses Kapitel führt das Thema Polaritätsshifter allgemein ein, motiviert seine Relevanz für die Sentimentanalyse und erläutert, wieso es bisher weitgehend vernachlässigt wurde.
- **Kapitel 2: hintergrund** Wir präsentieren eine formale Definition für Polaritätsshifter und wie sie in früherer Forschung gehandhabt wurden. Zudem befassen wir uns mit Themen, die dem der Shifter verwandt sind.

Teil II: verbale shifter In diesem Teil werden unsere Beiträge zur Erstellung von Lexika für verbale Shifter, also Verben, die Polaritätsshifter sind, besprochen.

- **Kapitel 3: erstellung eines lexikons von shiftern mithilfe von bootstrapping** Wir erstellen ein Lexikon englischer verbaler Shifter mithilfe eines *Bootstrapping*-Verfahrens. Dieses nutzt eine Kombination aus automatischer Klassifikation und menschlicher Gegenprüfung um sicherzustellen, dass das resultierende Lexikon sowohl umfangreich als auch von hoher Qualität ist. Als Teil dieses Prozesses entwickeln wir auf linguistischen Erkenntnissen basierende Klassifikationsmerkmale. Während einige dieser Merkmale existierende semantische Ressourcen nutzen, extrahieren andere Merkmale Informationen aus einem unannotierten Textkorpus.

Kapitel 4: annotation eines auf wortbedeutungen basierenden shifterlexikons Als eine Alternative zu unserem Bootstrapping-Verfahren erstellen wir ein weiteres Lexikon englischer verbaler Shifter, diesmal rein durch manuelle Annotation. Obgleich dies deutlich mehr Annotationsaufwand bedeutet, erlaubt es uns jedoch, Aspekte des Shiftens in Betracht zu ziehen, die beim Bootstrapping nicht hätten zuverlässig berücksichtigt werden können. Statt für jedes Wort eine einzelne Klassenzuordnung zu verwenden, die angibt, ob das Wort ein Shifter ist, wählen wir nun einen detaillierteren Ansatz, in dem jeder unterschiedlichen Wortbedeutung eine eigene Klassenzuordnung gegeben wird. Zudem annotieren wir auch den textuellen Skopus, der den syntaktischen Rahmen bestimmt, in welchem der Shifter Polaritäten verschieben kann.

Kapitel 5: sprachübergreifendes bootstrapping Nachdem gezeigt wurde, dass Bootstrapping für verbale Shifter im Englischen möglich ist, untersuchen wir nun, ob dies auch für die deutsche Sprache zutrifft. Um dies zu erreichen, erstellen wir deutsche Versionen unserer Klassifikationsmerkmale, wobei die linguistischen Muster angepasst und unser englisches Korpus und andere sprach-spezifische Ressourcen durch deutsche Entsprechungen ersetzen werden. Wir nutzen zudem den Umstand, dass wir bereits ein Lexikon verbaler Shifter in einer anderen Sprache erstellt haben. Mithilfe zweisprachiger Wörterbücher und sprachübergreifender Word Embeddings (zu deutsch *Worteinbettungen*) werden Informationen aus dem englischen Shifterlexikon ins Deutsche übertragen und in unseren Bootstrapping-Klassifikator integriert.

Teil III: erweiterung und anwendung Dieser Teil befasst sich damit, unser englisches Bootstrap-Lexikon zu erweitern. Die Einschränkung auf eine einzelne Wortart und binäre *Shifter/-Nichtshifter* Klassenzuordnungen wird aufgehoben. Zudem wird konkret gezeigt, dass Kenntnisse über Shifter helfen können, Polaritätsklassifikationen zu verbessern.

Kapitel 6: erweiterung des lexikons um nominale und adjektivische shifter In diesem Kapitel wird unser Bootstrapping-Verfahren auf englische Substantive und Adjektive angewendet. Darauf basiert lassen sich Ähnlichkeiten und Unterschiede zwischen den verschiedenen Wortarten beobachten. Welche Merkmale können für Substantive und Adjektive angepasst werden? Erweisen sie sich als genau so wirksam wie für Verben? Wie kann das zuvor erstellte Lexikon verbaler Shifter genutzt werden?

Das Studiums dieser Fragen resultiert in der Erstellung eines großen Bootstrap-Lexikons von englischen Shiftern, welches alle drei Wortarten abdeckt.

Kapitel 7: Erweiterung des Lexikons um Richtungspräferenzen von Shiftern Bisher wurde von uns die Frage, ob ein Wort (oder eine Wortbedeutung) ein Shifter ist, immer wie eine binäre Frage behandelt. Ist es ein Shifter, beeinflusst es alle Arten von polaren Ausdrücken in seinem Skopus. Ist es ein Nichtshifter, beeinflusst es die Polarität nicht. Jedoch beeinflussen nicht alle Shifter sowohl positive als auch negative Polaritäten. Einige Shifter verändern nur eine bestimmte Polarität (positiv *oder* negativ) und lassen die jeweils andere unbeeinflusst. Wir entwickeln einen automatischen Klassifikator, der unser Shifterlexikon um Informationen zu der Richtungspräferenz individueller Shifter erweitert.

Kapitel 8: Anwendung des Lexikons auf Sentimentanalyse Unser Shifterlexikon wurde von uns damit gerechtfertigt, dass es Anwendungen in der Sentimentanalyse verbessern kann. In diesem Kapitel wird diese Behauptung untersucht. Unser Bootstrap-Lexikon wird mit einem kompositionalen Klassifikator verglichen, der Negationsverhalten implizit aus Trainingsdaten erlernt, ohne explizite lexikalische Ressourcen zu benötigen. Zudem werten wir aus, welcher Nutzen daraus gewonnen werden kann, wenn die Kenntniss von Shiftern um die Unterscheidung von Wortbedeutungen oder um Richtungspräferenzen erweitert wird.

Kapitel 9: Fazit In unserem letzten Kapitel werden die Inhalte dieser Dissertation zusammengefasst. Dies beinhaltet einen Überblick über alle Ressourcen, die im Laufe dieser Arbeit erstellt wurden, sowie einen Blick auf mögliche zukünftige Forschungsaufgaben im Bereich der Polaritätsshifter.

Wissenschaftlicher Beitrag

Diese Dissertation leistet die folgenden wissenschaftlichen Beiträge:

Erkenntnisse

(i) Wir stellen ein neuartiges Bootstrapping-Verfahren vor, mit dessen Hilfe der Annotationsaufwand für die Erstellung eines Lexikons von Polaritätsshiftern signifikant reduziert werden kann.

(ii) Wir führen eine Reihe von Klassifikationsmerkmalen ein, die sowohl verfügbare semantische Ressourcen also auch linguistische Erkenntnisse nutzen, um die automatische Klassifikation von Polaritätsshiftern zu unterstützen.

(iii) Wir stellen Ähnlichkeiten und Unterschiede zwischen verschiedenen Sprachen und verschiedenen Wortarten fest und bieten Lösungsansätze für zahlreiche damit verbundene Herausforderungen. Dies beinhaltet die Einführung sprachübergreifender und wortartenübergreifender Methoden.

(iv) Wir untersuchen das Verhalten von Polaritätsshiftern, bestimmen den Skopus ihres Shift-Einflusses, sowie ihre Richtungspräferenz in Bezug auf die Beeinflussung positiver und negativer Polaritäten.

(v) Wir zeigen, dass explizites Wissen über Polaritätsshifter die Ergebnisse einer automatischen Polaritätsklassifikation im Vergleich zu Klassifikatoren ohne dieses Wissen signifikant verbessern kann.

(vi) Wir untersuchen die Nutzung detaillierter Informationen zu Shiftern wie zum Beispiel Richtungspräferenzen oder Klassenzuordnungen für einzelne Wortbedeutungen und analysieren sowohl ihr Potenzial als auch die Herausforderungen, die mit ihnen einhergehen.

Ressourcen

(i) Ein allgemeines Lexikon englischer Polaritätsshifter, welches Verben, Substantive und Adjektive abdeckt. Dieses Lexikon wurde erstellt mithilfe des von uns entwickelten Bootstrapping-Verfahrens.

(ii) Eine Erweiterung des allgemeinen Lexikons, um festzustellen, ob individuelle Shifter Polaritäten in eine oder beide Richtungen verschieben können.

(iii) Ein manuell erstelltes Lexikon englischer verbaler Shifter, welches Shifter-Klassenzuordnungen für jede Wortbedeutung jedes enthaltenen Verbs beinhaltet, mitsamt Angaben zum Shifter-Skopus jeder Wortbedeutung.

(iv) Ein Lexikon deutscher verbaler Shifter, dass mithilfe unseres Bootstrapping-Verfahrens erstellt wurde.

(v) Ein Goldstandard von Verbphrasen in Produktrezensionen zur Evaluation von Polaritätsklassifikatoren.

Alle Ressourcen sind **öffentlich zugänglich**. Angaben zu ihrer ursprünglichen Veröffentlichung finden sich in den entsprechenden Kapiteln. Für den Leser stellen wir zudem eine Sammlung aller Ressourcen in einem einzelnen Datensatz zur Verfügung.[1]

1 https://doi.org/10.5281/zenodo.3365605

PUBLICATIONS

This dissertation is a summarization of my work on sentiment polarity shifters. Individual parts of it have also been published in a number of other venues. Where there is overlap between a chapter of the dissertation and other publications, it is identified at the beginning of the respective chapter.

Overall, the dissertation reproduces findings from the following publications:

Schulder, Marc, 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)*. Taipei, Taiwan: Asian Federation of Natural Language Processing, pp. 624–633. ACL ANTHOLOGY: I17-1063.

Schulder, Marc, Michael Wiegand, and Josef Ruppenhofer (2018a). "Automatically Creating a Lexicon of Verbal Polarity Shifters: Mono- and Cross-lingual Methods for German." In: *Proceedings of the International Conference on Computational Linguistics (COLING)*. Santa Fe, New Mexico, USA: International Committee on Computational Linguistics, pp. 2516–2528. ACL ANTHOLOGY: C18-1213.

Schulder, Marc, Michael Wiegand, Josef Ruppenhofer, and Stephanie Köser (2018b). "Introducing a Lexicon of Verbal Polarity Shifters for English." In: *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*. Miyazaki, Japan: European Language Resources Association, pp. 1393–1397. ISBN: 979-10-95546-00-9. ACL ANTHOLOGY: L18-1222.

Schulder, Marc, Michael Wiegand, and Josef Ruppenhofer (under review). "Bootstrapped Creation of a Lexicon of Sentiment Polarity Shifters." In: *Journal of Natural Language Engineering* Special Issue on Processing Negation.

ACKNOWLEDGMENTS

First of all I would like to thank Dietrich Klakow for being my teacher, supervisor, Doktorvater, guide and support throughout my entire academic career, from my early days as a Bachelor student until now. Thank you for your encouragement, your flexibility and for saying *"You might also be interested in..."* time and time again.

My gratitude also goes to thank Michael Wiegand for his work as project lead and for guiding me in my work on polarity shifters that culminated in this dissertation. One could not ask for a more involved and dedicated supervisor.

I would like to thank my colleagues in the *Robust Fine-grained Sentiment Analysis* project, especially Josef Ruppenhofer for lending me his linguistic expertise, on which many of my classifier features were built. Stephanie Köser for her extensive annotation work (and her precise notes on it), without which my research would not have been possible. I also would like to thank the German Research Foundation (DFG) which funded the work that enabled my research under grants RU 1873/2-1 and WI 4204/2-1.

Thank you to André Miede and Ivo Pletikosić for their typesetting package `classicthesis`, without which this document wouldn't be nearly as pretty.

I am grateful to the battalion of proofreaders that double checked my work these last years. Special thanks to Michael Wiegand and Amy Isard for proofreading the entire dissertation and pointing out my inflationary use of *"use"*. Thanks to Michael Hedderich and Thomas Kleinbauer for checking that I had not forgotten the proper use of my mother tongue when writing the German summary. And many thanks to all the people who corrected the publications on which the dissertation builds: Dave Howcroft, Meaghan Fowlie, Anna Schmidt, Annemarie Friedrich, Andrea Fischer, Clayton Greenberg, Ines Rehbein and Katja Markert.

My gratitude also goes to my sisters and brothers in arms, the many grad students who studied alongside me. I am especially indebted to Dave Howcroft, Anna Schmidt, Mittul Singh, Meaghan Fowley and Amy Isard. Thank you so much for keeping me sane and fed.

Thanks also to my feline work/life balance coaches Danny and Snuggles, who ensured that I would take regular breaks from thesis writing by sitting on my keyboard.

Most importantly of all, I am forever grateful to my parents and brothers, whose unending support is the foundation on which I am built. Thank you for always being there for me.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

TERMINOLOGY

IAA Inter-Anntator Agreement

BOW Bag of Words

LP Label Propagation

NLP Natural Language Processing

NPI Negative Polarity Item

KNN k Nearest Neighbors

POS Part of Speech

VP Verb Phrase

WSD Word Sense Disambiguation

TOOLS & RESOURCES

NLTK Natural Language Toolkit

NOMLEX Nominalization Lexicon

RNTN Recursive Neural Tensor Network

SST Stanford Sentiment Treebank

SVM Support Vector Machine

Part I

PREFACE

1 INTRODUCTION

In natural language processing (NLP), the discipline of sentiment analysis is concerned with the detection and analysis of opinions and qualitative statements in language. While this involves several tasks, such as determining who has the opinion (*opinion holder detection*), what the opinion is about (*opinion target detection*) or how strong the opinion is (*intensity classification*), the vast majority of research focusses on determining the **polarity** (also referred to as **valence**) of a text, i. e. whether it is positive, negative or neutral.

The basis for determining the polarity of a text is knowing the polarity of individual terms within the text. For example, in (1.1) knowing that *to pass* is a positive term allows us to infer that *"pass the exam"* is a positive phrase and that the entire sentence is positive.

The polarity of an expression can also be influenced by a number of different phenomena, the most well known of which is **negation**. The best established cause of negation is **negation words**, such as *no*, *not*, *neither* or *without*. In (1.2) the negation word *not* affects the positive polarity of *"pass the exam"*, resulting in a negative polarity for the sentence.

(1.1) Peter [passed^{+} the exam]$^{+}$.
(1.2) Peter [did **not**$_{negation}$ [pass the exam]$^{+}$]$^{-}$.

Negation words are not, however, the only kind of words that can affect the polarity of a phrase. Many content words, so-called **polarity shifters**, can have a very similar effect. The negated statement in (1.2), for example, can also be expressed using the verb *to fail*, as seen in (1.3). Polarity shifters are also not limited to verbs. The nominal (1.4) and adjectival forms (1.5) of *to fail* exhibit the same kind of polarity shifting.

(1.3) Peter [**failed**$_{shifter}$ to [pass the exam]$^{+}$]$^{-}$.
(1.4) Peter's [**failure**$_{shifter}$ to [pass the exam]$^{+}$]$^{-}$.
(1.5) Peter's [**failed**$_{shifter}$ [attempt to pass the exam]$^{+}$]$^{-}$.

Like negation words, polarity shifters can both shift positive expressions to a negative phrasal polarity, as in (1.6), but also negative expressions to a positive phrasal polarity, such as in (1.7).

(1.6) [**Abandon**$_{shifter}$ all [hope]$^{+}$]$^{-}$, ye who enter here.
(1.7) [To **smooth**$_{shifter}$ that [rough touch]$^{-}$]$^{+}$ with a tender kiss.

However to date, polarity shifters have been widely overlooked or ignored in the NLP community.

1.1 AIMS

Our goal in this dissertation is to create resources and classification techniques for the handling of polarity shifters. The most central resource that we aim to create is a **general lexicon of English polarity shifters**.

The challenge of our task is to create such a large lexicon while keeping the required annotation effort manageable. How can annotation be automated without sacrifices of quality? Are there different requirements for different parts of speech? Do the same methods work for any part of speech? What about different languages? Can shifter information be transferred from one part of speech or language to another? What kinds of information should our lexicon provide? Is a single label per word enough or should we differentiate by word sense? Can word sense disambiguated content be used by NLP applications? Which parts of a sentence are affected by a particular shifter? Do all shifters affect all polarities equally?

These are the questions that we will pose and attempt to answer over the course of this dissertation.

1.2 POLARITY SHIFTERS IN NATURAL LANGUAGE PROCESSING

The correct handling of negation, including polarity shifters, has been shown to be important for various tasks in NLP, such as relation extraction (Sanchez-Graillet and Poesio, 2007), recognition of textual entailment (Harabagiu et al., 2006) and particularly sentiment analysis (Wiegand et al., 2010).

Unfortunately, while significant research has been performed on the topic of compositional polarity, it has widely been focussed on negation words (Wiegand et al., 2010; Schouten and Frasincar, 2016). One reason for this is availability of lexical resources for negation words and lack thereof for polarity shifters. Negation words are usually function words, of which there are few. Polarity shifters, on the other hand, are content words (e. g. verbs, adjectives and nouns), which are far more numerous. *WordNet* (Miller et al., 1990), for example, contains over 10,000 verbs, 20,000 adjectives and 110,000 nouns. At the same time, most individual content words occur far less frequently than individual function words.

This means that creating a comprehensive lexicon of negation words is far cheaper than creating one for polarity shifters, while at the same time each individual negation word helps to detect more polarity changes than a single polarity shifter would. It does not, however, mean that polarity shifters are less essential for polarity classification. In fact, we will show that polarity shifters can be expected to occur more frequently than negation words.

1.3 OVERVIEW

The dissertation is divided into three parts, each of which contains several chapters. The content structure of the dissertation is as follows:

PART I: PREFACE We introduce the general concept of polarity shifters and any background knowledge required for this work.

- CHAPTER 1: INTRODUCTION The chapter you just read. It gives a general introduction to the topic of polarity shifters, why they matter for sentiment analysis and why they have largely been neglected so far.
- CHAPTER 2: BACKGROUND We provide a formal definition of polarity shifters, as well as a look at how they have been handled in previous research. It also touches upon topics that are closely related to shifters, relations that will become relevant during the course of our work.

PART II: VERBAL SHIFTERS In this part we present our efforts for creating lexica of verbal shifters, i. e. polarity shifters that are verbs.

- CHAPTER 3: BOOTSTRAPPING A SHIFTER LEXICON We create a lexicon of verbal shifters using a bootstrapping approach that uses a combination of automatic classification and human verification to ensure a lexicon that is both large and of high quality. As part of this we develop a variety of features, based on linguistic insights. While some of these features make use of existing semantic resources, others extract information from an unannotated text corpus.
- CHAPTER 4: ANNOTATING A SENSE-LEVEL SHIFTER LEXICON As an alternative to bootstrapping, we create a lexicon of verbal shifters entirely through manual annotation. While this requires considerably more annotation work, it also allows us to include aspects of shifting that we could not have handled reliably during bootstrapping. Instead of assigning a single shifter label per word, we now choose a more fine-grained approach, labeling each sense of a word individually. In addition we also annotate the scope within which a shifter can shift polarities.
- CHAPTER 5: CROSS-LINGUAL BOOTSTRAPPING Having shown that bootstrapping verbal shifters is possible in English, we now investigate whether the same is true for German. To achieve this, we establish German versions of our features, adjusting linguistic text patterns and replacing our English text corpus and semantic resources with equivalent German ones. We also make use of the fact that we have already

created a verbal shifter lexicon in another language. Using bilingual dictionaries and cross-lingual word embeddings we transfer the information from our English shifter lexicon to German and integrate it into our bootstrap classifier.

PART III: EXTENSION AND APPLICATION This part is concerned with extending the lexicon further, leaving behind limitations to a single part of speech and binary shifter labels, as well as providing concrete evidence for our assumption that shifters will help us improve polarity classification.

CHAPTER 6: EXTENDING THE LEXICON BY INTRODUCING NOMINAL AND ADJECTIVAL SHIFTERS After addressing verbal shifters extensively in the previous part, it is time to move on to other parts of speech. In this chapter we apply our bootstrapping workflow to nouns and adjectives. In the course of this, we identify similarities and differences between the various parts of speech. Which features can be adapted to nouns and adjectives? Will they prove to be as efficient as they were for verbs? How can we make use of our previously created lexicon of verbal shifters? As we answer these questions, we create a large bootstrapped shifter lexicon that covers all three parts of speech.

CHAPTER 7: EXTENDING THE LEXICON BY INTRODUCING SHIFTING DIRECTIONS So far we have treated being a shifter as a binary decision. If a word (or word sense) is a shifter, it shifts polar expressions and if it is a non-shifter, it does not. However, not all shifters affect both positive and negative polarities. Some shift only positive or only negative expressions and leave the other polarity unaffected. We develop a classifier that extends our shifter lexicon by labeling each shifter with its possible shifting directions.

CHAPTER 8: APPLYING THE LEXICON TO SENTIMENT ANALYSIS We justified the creation of a shifter lexicon with the improvements it can offer to applications such as sentiment analysis. In this chapter we evaluate this claim. We compare our bootstrapped lexicon to a compositional classifier that has been shown to learn negation patterns from data without the need for lexical resources. Furthermore we evaluate what can be gained by extending shifter knowledge to individual word senses or by including information about shifting directions.

CHAPTER 9: CONCLUSION Our final chapter provides a summarization of the work presented in this dissertation, including an overview of all resources that were created in the course of our work and an outlook to future directions in the topic of polarity shifter research.

1.4 CONTRIBUTIONS

INSIGHTS

(i) We establish a bootstrapping workflow for the creation of polarity shifter lexica that significantly reduces the annotation effort.

(ii) We introduce a variety of features that make use of available semantic resources as well as leveraging linguistic insights to aid the automatic classification of polarity shifters.

(iii) We identify how shifters differ among languages and among various parts of speech, providing solutions to the variety of challenges they pose, including the introduction of cross-lingual and cross-POS features.

(iv) We investigate the behavior of polarity shifters, identifying the scope of their shifting as well as the direction in which they can shift polarities.

(v) We show that explicit knowledge of polarity shifters can provide significant performance gains for polarity classification compared to classifiers without this knowledge.

(vi) We evaluate what gains and challenges are involved when using more fine-grained shifter information, such as labels at the word-sense level or shifting direction labels, for polarity classification.

RESOURCES

(i) A general lexicon of English polarity shifters, created through bootstrapping, covering verbs, nouns and adjectives.

(ii) An extension to the general lexicon that provides shifting direction information for each shifter.

(iii) A hand-crafted lexicon of English verbal shifters that provides shifter labels for each word sense of each verb, including information on the shifting scope of each sense.

(iv) A lexicon of German verbal shifters, created through bootstrapping.

(v) A gold standard of English verb phrases in product reviews for evaluating shifter handling in polarity classifiers.

All resource are **publicly available**. The locations of their original publications are provided in their respective chapters. For the benefit of the reader, a repository compiling all resources is also made available.[1]

1 https://doi.org/10.5281/zenodo.3365605

2 BACKGROUND

In this chapter we establish the background knowledge required for our work on polarity shifters. In section 2.1 we provide a definition of polarity shifters, describe what should or should not be considered a shifter and explain how shifters interact with other phenomena that affect phrasal polarity. Next we discuss resource creation in section 2.2, outlining the vocabularies which our lexica will cover and introducing the human annotators involved in our work. Section 2.3 introduces relevant topics of research that are related to polarity shifters and in section 2.4 we present other research that concerns itself with polarity shifting.

Publication History

The explanations in this chapter are largely based on those in Schulder et al. (2018b) and Schulder et al. (under review).

2.1 POLARITY SHIFTERS

This section introduces the concept of polarity shifters. We begin in section 2.1.1 by defining what shifters are, how they differ from negation words and how they relate to polar expressions. In section 2.1.2 we discuss issues of interpreting the shifting of scalar concepts, such as qualitative statements like *good*, and how they may sometimes result in a shift to a neutral polarity. Another issue, which we address in section 2.1.3, is the question of whether downtoning and intensification should also be considered polarity shifting. The last part of our definition of shifters, section 2.1.4, concerns itself with how exactly polarity shifters and other phenomena affect phrasal polarity and how they influence one another.

2.1.1 *Definition of Polarity Shifters*

Polarity shifting occurs when the sentiment polarity (or valence) of a word or phrase is moved towards the opposite of its previous polarity (i. e. from positive towards negative or vice versa). The notion of polarity shifting (also referred to as *valence shifting*) was brought to broad awareness in the research community by the work of Polanyi and Zaenen (2006), who observed that the prior polarity of individual lexical items could be shifted by a) specific lexical items, b) the discourse structure and genre type of a text and c) socio-cultural factors.

In subsequent research, the meaning of the term *shifter* was narrowed to refer to lexical items that influence polarity.

For the purposes of this work, we differentiate between *polarity shifters* and *negation words*. Both are forms of **negation**, but negation words are *closed class* words, while polarity shifters are *open class* words, such as verbs, nouns or adjectives.

Polarity shifters are defined by their ability to negate or diminish facts or events that were either previously true or presupposed to occur. In (2.1) the speaker presupposes that their daughter would receive a scholarship, as she applied for one. This did not happen, as being denied a scholarship implies not receiving it. In (2.2), the speaker presupposes that their amount of pain would continue at the same level, but due to the medication the amount of pain is reduced. These examples also show that shifting can occur in either direction, as in (2.1) a positive polarity is shifted to negative and in (2.2) a negative polarity is shifted to positive.

(2.1) My daughter was [**denied**$_{shifter}$ the [scholarship]$^{+}$]$^{-}$.

(2.2) The new treatment has [**alleviated**$_{shifter}$ my [pain]$^{-}$]$^{+}$.

While some polarity shifters are also polar expressions themselves, their polarity does not necessarily dictate the direction of its shifting. For example, the shifter *"destroy"* is itself of negative polarity, but can shift both positive and negative words, as seen in (2.3) and (2.4).

(2.3) Smoking [[**destroys**]$^{-}_{shifter}$ your [health]$^{+}$]$^{-}$.

(2.4) The medication [[**destroys**]$^{-}_{shifter}$ [cancer cells]$^{-}$]$^{+}$.

2.1.2 *Shifting of Scalar Concepts*

Most commonly, the terms **negation** and **shifting** are used to refer to a change between discrete polarity classes, e. g. changing from positive to negative or from negative to positive. In cases involving the negation of expressions which are part of a binary opposition (*dead* – *alive*), one can firmly conclude that the complementary state of affairs holds. Any parrot that is dead has ceased to be alive.

The case is less clear when we consider negation affecting scalar notions, as is common in evaluative contexts. Here the understanding that arises depends on which kinds of scalar inferences and default assumptions are made in the context (Paradis and Willners, 2006). Consider for example the polarity of *"wasn't excellent"* in (2.5).[1] There is no consensus on whether the negated expression should be considered negative or neutral.[2]

(2.5) Let's say the movie [**wasn't** [excellent]$^{+}$]$^{-/\sim}$.

1 Following Taboada et al. (2011) we use the negation word *not* in (2.5). A similar sentiment may be expressed via a shifter, e. g. by saying *"the movie failed to be excellent."*

2 In example sentences, we use ~ to denote neutral polarity.

Choi and Cardie (2008) argue that the positive polarity of *excellent* is simply flipped to negative. On the other hand, Taboada et al. (2011) and Kiritchenko and Mohammad (2016) point out that the negation of *excellent* is not synonymous with its antonym *atrocious* and should be considered to be of neutral polarity, due to its weak intensity.

Something being *not good* denies the applicability of an evaluation in the region of good or better, but leaves open just how far in the direction of badness the actual interpretation lies. *"It wasn't good"* may be continued with *"but it was ok"* to yield a neutral or mildly positive evaluation or with *"in fact, it was terrible"* to yield a strongly negative one.[3]

2.1.3 *Downtoning and Intensification*

So far we have talked about polarity shifting as causing changes between discrete states: *Positive*, *negative* and possibly *neutral* polarity. There can also be changes in the intensity of a polar expression without the polarity changing to a different discrete state. When this change reduces the intensity of a polar expression, this is referred to as downtoning. When it increases the intensity, it is called intensification. For example, in (2.6) the downtoner *slightly* reduces the intensity[4] of the adjective *painful*, but the polarity of the expression remains negative. In (2.7) the negative polarity adjective *dangerous* is made even more negative by the intensifier *extremely*.

(2.6) The injection was [slightly [painful]$^{--}$]$^{-}$.
(2.7) Riding a bear is [extremely [dangerous]$^{-}$]$^{--}$.

Some authors, including Polanyi and Zaenen (2006), consider that intensification and downtoning also fall within the scope of shifting, as both affect the polar intensity of a phrase.

We partially agree with this view, in that we also consider downtoning to be shifting, as it moves the polarity of a word or phrase towards its opposite direction. A positive expression, for example, is made less positive (e. g. *"somewhat interesting"*) and a negative one is made less negative (e. g. *"slightly problematic"*). Therefore, while downtoners (e. g. *somewhat*) applied to scalar predicates such as *interesting* do not directly express contradiction, they do give rise to negative entailments and inferences. Moreover, the structure of scales intrinsically provides shifting. Thus, while something being *interesting* allows it to be even more positive in (2.8), something being *somewhat interesting* bounds its positiveness and opens up more negative meanings, as in (2.9).

3 Note that the example also illustrates that distinguishing between items that switch discrete polarities and items that only affect intensity is an idealization. Simple syntactic negation of a polar adjective may influence intensity as well as polarity (e. g. *not terrible* $\neq$ *excellent*) (Kiritchenko and Mohammad, 2016).

4 In examples illustrating changes in intensity we indicate expressions of strong positive intensity with $++$ and expressions of strong negative intensity with $--$.

Considering these properties of scales, one can see shifting (in the sense of negation) at work even in the case of downtoning.

(2.8) [[The movie was interesting.]$^{+}$ In fact, it was fascinating.]$^{++}$
(2.9) [[The performance was somewhat interesting,]$^{+}$ but overall rather dull.]$^{-}$

Intensifiers, on the other hand, strengthen a given polarity and prevent it from being replaced with a different polarity. In (2.8) the *interesting movie* cannot be bad at the same time, but can be even more positive than *interesting* already implies; in this case it is found to be *fascinating*. For our purposes of handling negation, we therefore do not consider intensification to be polarity shifting.

2.1.4 *Compositionality of Phrasal Polarity*

To determine the polarity of a phrase, we observe a) the polarity of its lexical items and b) how their polarity is influenced by contextual elements, including other polarities (Polanyi and Zaenen, 2006; Moilanen and Pulman, 2007). Following the principles of semantic compositionality, the scope of most contextual elements is limited to specific syntactic constituents (Moilanen and Pulman, 2007; Choi and Cardie, 2008). In (2.10) the shifter *defeat* affects the polarity of its direct object while in (2.11) the word *falter* shifts the polarity of its subject.

(2.10) [The hero]$^{+}_{subj}$ [**defeated**$_{shifter}$ [the villain]$^{-}_{dobj}$]$^{+}$.
(2.11) [[[My enthusiasm]$^{+}_{subj}$ **faltered**$_{shifter}$]$^{-}$ **despite** their [encouragement]$^{+}$]$^{-}$.
(2.12) [[The battle was gruesome]$^{-}$, **but** [we prevailed]$^{+}$]$^{+}$.

However, polarity shifters and negation words are not the only factors that can influence phrasal polarity. For example, connectives, such as *although*, *however* or *but* can influence which parts of a sentence can impact the overall polarity of the phrase (Polanyi and Zaenen, 2006). In (2.11) the positive polarity of *encouragement* is counteracted by the connective *despite* and in (2.12) the connective *but* indicates that the positive polarity of the second half of the sentence takes precedence over the negative polarity of the first half.

Matters are complicated even further when one considers modal operators like *if* and *would*, which introduce hypotheticals that do not directly impact the polarity of real events. For example, (2.13) does not convey a negative opinion about Mary (Polanyi and Zaenen, 2006). Modal operators may even shift polarities, e. g. in (2.14) it is used to imply that the cell phone is *not* perfect (Liu et al., 2014).

(2.13) [**If** Mary were [a terrible person]$^{-}$, she **would** [be mean to her dogs]$^{-}$]$^{\sim}$.
(2.14) [This cellphone **would** be [perfect]$^{++}$ **if** it had a bigger screen]$^{+}$.

Determining the polarity of a phrase is therefore not trivially a matter of enumerating all polarities, negations and shifters found in it. Nevertheless, negation and shifters represent an essential factor for determining phrasal polarity.

2.2 RESOURCE CREATION

In the course of the dissertation, a number of lexica and other resources are created, both manually and automatically. In section 2.2.1 we define the English and German vocabularies on which the shifter lexica will be based. In section 2.2.2 we introduce the human annotators that participated in creating our resources and identify their individual contributions.

2.2.1 *Vocabulary*

To create a lexicon of polarity shifters, we first need to define our underlying vocabulary. To this end, we extract all verbs, nouns and adjectives from *WordNet 3.1* (Miller et al., 1990). Furthermore, we remove the following kinds of words:

(i) Words containing digits, such as *"1750s"* and *".38 calibre"*.

(ii) Acronyms and abbreviations, such as *"C.P.U."* and *"Jr."*.

(iii) Proper nouns, such as names.

(iv) Multi-word expressions, with the exception of particle verbs.

The nature of (i), (ii) and (iii) precludes them from being polarity shifters. Multi-word expressions can theoretically be polarity shifters, but most multi-word expressions in *WordNet* are either compositional phrases like *"abandoned infant"* or proverbial expressions like *"bright as a new penny"*. Neither of these can be shifters, as they already contain both the potential shifter (e. g. *"abandoned"*) as well as the shifted expression (e. g. *"infant"*). The polarity of such multi-word expressions should instead be determined using compositional semantic processing and explicit polarity lexicon entries for proverbial expressions.

Particle verbs, on the other hand, are verbs like *"give back"* whose semantic meaning is distinct from their particle-free form (e. g. *"give"*). Particle verbs can potentially be shifters, just like any other verb. Therefore, we include them in our vocabulary.

The resulting English vocabulary consists of 84,174 words: 10,581 verbs, 55,311 nouns and 18,282 adjectives. In addition, we create a German vocabulary of verbs in chapter 5. Following the same approach as for the English vocabulary, we extract 9,262 verbs from *GermaNet* (Hamp and Feldweg, 1997).

From this point onwards, the phrase *"all words"* should be understood to refer to the words belonging to our filtered vocabulary of the language in question.

2.2.2 *Annotators*

The majority of the annotation work presented in this dissertation was performed by Stephanie Köser, who is an expert annotator with experience in both linguistics and annotation work. Other parts where annotated by Dr. Michael Wiegand or the author.

GOLD STANDARDS: The gold standards for the shifter lexica in chapters 3, 5 and 6 and the annotation of the sense-level verbal shifter lexicon in chapter 4 were annotated in their entirety by Stephanie Köser. She also provided the verbal shifter annotation for the shifting direction gold standard in chapter 7, while the author annotated the shifting directions for the nominal and adjectival shifters. The gold standard for the sentiment analysis task in chapter 8 was annotated by the author.

BOOTSTRAP VERIFICATION: The bootstrap verification for the English and German verbal shifters in chapters 3 and 5 was done by Stephanie Köser. The bootstrap verification of the nominal and adjectival shifters in chapter 6 was performed in part by Stephanie Köser and in part by the author.

INTER-ANNOTATOR AGREEMENT: To compute the inter-annotator agreement, several gold standards were partially re-annotated. The re-annotation of all shifter lexica (chapters 3–6) was performed by the author. The re-annotation of the shifting direction gold standard in chapter 7 was done by Dr. Michael Wiegand.

2.3 RELATED TOPICS OF RESEARCH

This section discusses topics of research which, while not identical to polarity shifting, are closely related to it. Section 2.3.1 describes downward-entailing operators and their close relation with negative polarity items, while section 2.3.2 discusses +/−effect theory.

2.3.1 *Downward-Entailing Operators and Negative Polarity Items*

Upward- and downward-entailment are linguistic concepts that describe whether a statement entails either its more relaxed or more restricted forms (Ladusaw, 1980). In (2.15) the upward-entailing operator *know* allows us to infer the relaxed statement *"We know the epidemic spread"*, but not the restricted statement *"We know the epidemic spread quickly via fleas"*. This represents the standard inference assumption that general information can be inferred from more specific information. In (2.16), on the other hand, the inference assumption is inverted by the downward-entailing operator *doubt*, as it allows us to infer the restricted statement *"We doubt the epidemic spread quickly via fleas"*, but not the relaxed one *"We doubt the epidemic spread"*.

(2.15) We **know** the epidemic spread quickly.
(2.16) We **doubt** the epidemic spread quickly.

This inversion of inference assumptions, as is modeled by downward-entailment, is closely related to polarity shifting, as they both relate to the non-existence or limitation of entities (van der Wouden, 1997) (compare to our definition of shifters in section 2.1.1). This overlap in definitions means that downward-entailing operators (like *doubt* in (2.16)) often also qualify as polarity shifters or negation words.

A phenomenon strongly associated with both downward-entailing operators (Ladusaw, 1980) and negation (Baker, 1970; Linebarger, 1980) is negative polarity items (NPIs). NPIs are words that are excluded from being used with positive assertions. For example, the NPI *any* may be used in negated contexts, such as in (2.17), but not in positive assertions like (2.18).

(2.17) They did **not** find any evidence.
(2.18) *They found any evidence.

NPIs have been shown to be strongly connected to downward-entailing operators, usually occurring in their scope (Ladusaw, 1980), although their exact nature is still being controversially discussed (Giannakidou, 2011). Danescu-Niculescu-Mizil et al. (2009) take advantage of the connection when creating a list of downward-entailing operators. Using unsupervised machine learning, they show that the co-occurence of words with NPIs can be leveraged to identify downward-entailing operators. The concept of NPIs is not specific to English and can be found in many other languages (Krifka, 1991). Their connection with downward-entailing operators equally exists and can be leveraged in similar ways as for English (Danescu-Niculescu-Mizil and Lee, 2010).

2.3.2 +/−*Effect Theory*

A semantic phenomenon that is quite similar to polarity shifting is that of +/−effect, which posits that events may have beneficial or harmful *effects* on the objects they influence (Deng et al., 2013; Choi et al., 2014; Choi and Wiebe, 2014). The concept was originally introduced in the context of annotation and lexical acquisition work for opinion inference.

With the release of *EffectWordNet* (Choi and Wiebe, 2014) the terms +*effect* for beneficial effects and −*effect* for harmful effects were introduced.[5] *EffectWordNet* is a lexical resource for verbs that assigns

5 In earlier works, the terms *good-for* and *bad-for* were used.

+/−effect labels to *WordNet* synsets (i. e. sets of words that are synonymous in a specific word sense).

In their original use for opinion inference, the point of +/−effects was that through the interplay of opinion words and effects, information on opinions could be inferred. For example, in (2.19) the event that people are happy about is *"Chavez has fallen"* and *fall* has a harmful −effect on *Chavez*. From this it can be inferred that people have a negative opinion of Chavez, as they show a positive reaction about a harmful event happening to him.

(2.19) I think people are happy because [[Chavez]$^-$ has **fallen**$_{-effect}$]$^+$.

As a semantic concept, −effects bear some similarity to polarity shifters. Many times, the harmful effect that −effect describes is one of removal or weakening, i. e. of shifting. This can be observed in (2.19). However, despite their similarity the two phenomena are not identical. In both (2.20) and (2.21) we observe a −effect, but only *betray* in (2.20) functions as a polarity shifter, affecting the polarity of the verb phrase. On the other hand, *abuse* in (2.21) does not shift the polarity, despite indicating a harmful −effect for the prisoners.

(2.20) He [**betrayed**$_{-effect}$ his [friends]$^+$]$^-$ for money.

(2.21) We don't want the public getting the idea that we [**abuse**$_{-effect}$ our [prisoners]$^-$]$^-$.

2.4 RELATED WORK

Approaches to learning negation from labeled corpora have been examined in a number of domains. Research regarding the automatic handling of negation scopes has focussed a lot on the medical domain (Huang and Lowe, 2007; Morante and Daelemans, 2009; Zou et al., 2013), although more recently this has also been extended to cross-domain approaches (Fancellu et al., 2016). In the context of sentiment analysis, a lot of research is performed in the review domain (Ikeda et al., 2008; Kessler and Schütze, 2012; Socher et al., 2013; Yu et al., 2016). Further information on negation modeling in sentiment analysis can be found in the survey by Wiegand et al. (2010).

The majority of the aforementioned works concern themselves chiefly with the handling of negation words and with determining their scope. In this thesis we shall instead focus our discussion on works that include polarity shifters in their handling of negation.

There are very few resources that provide information about polarity shifters, and even fewer that offer any serious coverage. The most complex general negation lexicon was published by Wilson et al. (2005b) as part of *Opinion Finder* (Wilson et al., 2005a). It contains 30 polarity shifters. The *BioScope* corpus (Szarvas et al., 2008), a collection of texts from the medical domain, has been annotated explicitly for negation cues. Among these negation cues, Morante (2010) identifies 15 polarity shifters.

An alternative way of handling negation words and shifters is to learn about them implicitly from corpora. The *Stanford Sentiment Treebank* (Socher et al., 2013) provides compositional polarity information for 11,855 sentences. Each sentence is provided with a syntactic parse tree, of which each node is annotated with a polarity. Shifting events can be inferred by observing differences between the polarity of a node and the polarities of its children. Socher et al. (2013) show that a neural network polarity classifier trained on the *Stanford Sentiment Treebank* can successfully identify negation words. However, as there is considerably more lexical variety among polarity shifters than there is among negation words, we do not expect the size of the treebank to be sufficient for consistently handling shifters. We evaluate this assumption in chapter 8.

The work that is most closely related to our own effort of bootstrapping lexicon creation is that of Danescu-Niculescu-Mizil et al. (2009) who create a lexicon of downward-entailing operators, which are closely related to polarity shifters (see section 2.3.1). Leveraging the co-occurrence of downward-entailing operators with negative polarity items, they use unsupervised machine learning to generate a ranked list of downward-entailing operators. The 150 highest ranked items are then verified by a human annotator, which yields 90 downward-entailing operators. They include a precision analysis, but lack any information on recall or evaluation of the impact of their created resource on computational applications.

German polarity shifters are also used in Wiegand et al. (2018b), which concerns itself with modeling the scope of negation words and shifters in German sentences. It shows that considering specific syntactic scopes for different kinds of shifters can improve negation modeling. We address the matter of shifting scopes in chapter 4.

Wiegand et al. (2018a) address the fact that shifter annotations for individual word senses, as we provide in chapter 4, are only of use to computational approaches if word sense disambiguation tools are able to distinguish the shifter and non-shifter senses of a word in a given text. While they find that this is generally possible, they conclude that large amounts of labeled training would be required to provide adequate performance.

Part II

VERBAL SHIFTERS

3 BOOTSTRAPPING A SHIFTER LEXICON

Polarity shifters are important for a variety of tasks in natural language processing, however, up until now, their use has rarely gone beyond hand-crafted example cases, as there have been no sufficient resources or heuristics for identifying which words qualify as shifters. Existing resources like the negation lexicon by Wilson et al. (2005b) or the *BioScope* corpus by Szarvas et al. (2008) cover a handful of shifters each. This only scratches the surface, as the number of polarity shifters is more likely to be in the hundreds or thousands.

Attempting to create a comprehensive shifter lexicon entirely through manual annotation would be prohibitively expensive, as it would require the annotation of many tens of thousands of words. Instead we introduce a bootstrapping approach that will allow us to filter out the majority of words that do not cause polarity shifting (non-shifters), which significantly reduces the manual annotation effort.

In this chapter we focus our bootstrapping efforts exclusively on verbs. As the main predicates of phrases they tend to have larger scopes than nouns and adjectives, increasing the impact of their polarity shifting. Focussing on a single part of speech also allows us keep a stronger focus in the design of our classifier, as we do not have to worry about how textual and syntactic patterns differ between parts of speech.

Bootstrapping verbal shifters is of course only a first step in our efforts for the creation of a general polarity shifter lexicon. The bootstrap approach and lexicon will be further extended in chapter 6 to also address polarity shifters among nouns and adjectives.

The structure of our bootstrapping approach is detailed in Figure 3.1. We begin by having a human annotator label a small set of randomly

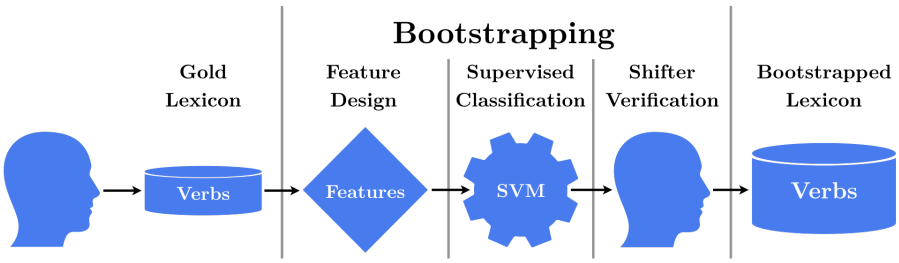

Figure 3.1: Workflow for bootstrapping the verbal shifter lexicon. The manually annotated gold lexicon is used to design the features and train the classifier, which then classifies all unlabeled verbs. Verbs classified as shifters are verified by a human annotator before being included in the final lexicon.

sampled verbs, which are used to evaluate a variety of linguistic features and to train a supervised classifier. This classifier is then used to classify the remaining unlabeled verbs of our vocabulary. Those verbs that the classifier considers shifters are then manually verified by our annotator, while those classified as non-shifters are discarded. As the majority of verbs are non-shifters, this approach significantly reducing the annotation load while still ensuring the high quality of our lexicon.

Contents

We begin the chapter by motivating our choice to first focus entirely on verbal shifters (section 3.1). Following this, we describe the gold standard we create for our task in section 3.2 and the resources required in section 3.3. In section 3.4 we introduce the features which we use for the bootstrap classification. We evaluate the features and the classifiers that use them in section 3.5. Section 3.6 then presents the result of the actual bootstrapping process. The findings of this chapter are then summarized in section 3.7.

Contributions

(i) We present the first high-coverage lexicon of verbal polarity shifters, going substantially beyond what can be extracted from existing phrase-level corpus annotations.

(ii) We develop methods that allow for the high-precision recognition of polarity shifters.

(iii) In addition to using resource-based generic features, we show that we can boost performance with novel task-specific features, many of which are derived from corpora.

(iv) The successful application of linguistically motivated features furthers our understanding of polarity shifters.

Publication History

The work presented in this chapter can also be found in Schulder et al. (2017) and Schulder et al. (under review). The bootstrapped lexicon of verbal shifters has been released publicly.[1]

3.1 VERBAL SHIFTERS

The phenomenon of polarity shifting can be observed among content words in general and is not limited to words of a particular part of

1 https://doi.org/10.5281/zenodo.3364811

LABEL	FREQUENCY	PERCENTAGE
Shifter	304	15.2%
Non-shifter	1,696	84.8%
Total	2,000	

Table 3.1: Distribution of verbal shifters in the gold standard, which consists of a random sample 2,000 words taken from *WordNet*.

speech. Nevertheless we choose to focus our initial work on verbal shifters before addressing nominal and adjectival shifters in chapter 6.

Verbs, together with nouns, are the most important minimal semantic units in text, as shown by the work of Schneider et al. (2016). Verbs are usually the main syntactic predicates of clauses and sentences and thus verbal shifters can be expected to project far-reaching scopes. Most nominal shifters (e. g. *failure*, *loss*), on the other hand, have morphologically related verbs (e. g. *fail*, *lose*). In chapter 6 we will show that this connection can be exploited to spread shifter classification from verbs to nouns.

3.2 GOLD STANDARD

To train and test our classifiers, we create a gold standard of verbal polarity shifters. We extract a random sample of 2,000 verbs from the WordNet vocabulary that we established in section 2.2.1 These words are then labeled by an expert annotator who has experience in linguistics and annotation work. To ensure that all possible senses of a word are considered, they refer to word definitions in a number of different dictionaries. Annotation of the 2,000 verbs takes around 57 hours.

Annotation is handled as a **binary classification** task. Each word is either a *'shifter'* or a *'non-shifter'*. Following our definition of polarity shifters from section 2.1, to qualify as a shifter, a word must allow polar expressions as its dependent and the polarity of the shifter phrase (i. e. the proposition that embeds both the shifter and the polar expression) must move towards a polarity that is the opposite of that of the polar expression.

To measure inter-annotator agreement for the annotation task, 10 percent of the gold standard words are annotated a second time by a second expert annotator. The resulting Cohen's kappa (Cohen, 1960) of $\kappa = 0.66$ indicates substantial agreement (Landis and Koch, 1977).

Table 3.1 shows the distribution of shifters among the set of verbs. At approximately 15 percent, we expect that shifters represent a large enough proportion of verbs to be considered for automatic extraction.

	POSITIVE VERBS		NEGATIVE VERBS	
LABEL	FREQUENCY	PERC.	FREQUENCY	PERC.
Shifter	4	5.5%	49	25.9%
Non-shifter	69	94.5%	140	74.1%
Total	73		189	

Table 3.2: The distribution of sentiment polarities among verbal shifters from the gold standard. Polarities are automatically determined using the *Subjectivity Lexicon* (Wilson et al., 2005b).

Our gold standard is annotated at the lemma level. In the case of words with multiple word senses, we consider a word a shifter when at least one of its senses qualifies as shifter. Both regarding our bootstrapping efforts and regarding the later computational use of the resulting shifter lexicon, shifter labels for individual word senses would only be of use if the texts they were applied to were also word sense disambiguated. We do not believe that general word sense disambiguation applications are sufficiently robust to make such an approach worthwhile at this point. Initial investigations into designing specialized word sense disambiguation for polarity shifters has found that such systems would require significant (and so far unavailable) amounts of training data (Wiegand et al., 2018a).

We revisit the topic of sense-level shifter annotation in chapter 4 and the challenges of word sense disambiguation in chapter 8.

3.3 RESOURCES

We rely on a number of additional resources to create the classifier features that we will introduce in section 3.4. Resources used by more than one feature are described here. Others will be introduced together with the feature that uses them.

3.3.1 *Sentiment Polarity*

As polarity shifters affect the polarity of words, it does not come as a surprise that some of our features require knowledge about **sentiment polarity**. To determine the polarity of individual words, we use the *Subjectivity Lexicon* (Wilson et al., 2005b). Table 3.2 shows the distribution of shifters among polar verbs. While a large number of shifters are themselves negative polar expressions, not all are.

3.3.2 *Text Corpus*

Many of our features require a **text corpus**, for example to determine word frequencies or for pattern matching. We use the *Amazon Product Review Data* (Jindal and Liu, 2008), a corpus of 5.8 million product reviews. The corpus was chosen both for its large size and the thematic domain of its texts. Product reviews are a typical domain for sentiment analysis applications, as they are rich in opinions and polar statements and very focussed on communicating the opinion of the author (Liu, 2012, p. 16). We expect that using such a sentiment-rich text corpus will help avoid issues of sparsity that might arise in other corpora that consist more of neutral factual statements that cannot be affected by polarity shifters.

3.3.3 *Word Embeddings*

Some of our features make use of the distributional hypothesis that *"a word is characterized by the company it keeps"* (Firth, 1957). The assumption of distributional semantics is that words that occur in similar contexts will have similar meanings. To determine this distributional similarity, we compute a word embedding vector space from our text corpus using *Word2Vec* (Mikolov et al., 2013). Following the work of Wiegand and Ruppenhofer (2015), who used word embeddings in the related task of verb category induction for sentiment roles, we set *Word2Vec* to use the *continuous bag of words* algorithm and to generate a vector space with 500 dimensions. All other settings are kept at their default. To determine the distributional similarity between two specific words A and B in the embedding space, we compute the cosine similarity of their vector representations $\vec{A}$ and $\vec{B}$:

$$\cos(\vec{A}, \vec{B}) = \frac{\vec{A} \cdot \vec{B}}{\|\vec{A}\|\|\vec{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \tag{3.1}$$

3.3.4 *Syntactic Structure*

Syntactic dependency relations are often used to extract information from corpora through the application of text patterns (Jiang and Riloff, 2018) and to collect complex distributional information (Shwartz et al., 2016). We determine dependency relations and other syntactic information using the *Stanford Parser* (Chen and Manning, 2014).

3.4 FEATURE DESIGN

To bootstrap our lexicon in section 3.6 our classifiers require knowledge in the form of classifier features. In section 3.4.1 we introduce features that we specifically designed for the task of determining verbal shifters. Section 3.4.2 presents more generic features that have already been established as useful for a number of sentiment analysis tasks.

3.4.1 *Task-Specific Features*

In this section we present task-specific features for both verbal shifters and their counterpart, verbal non-shifters. Each feature creates a verb ranking, indicating how likely each verb is to be considered a verbal (non-)shifter.

EFFECTWORDNET (EFFECT): For this feature we leverage the relatedness of polarity shifters to the concept of +/−effects that we discussed in section 2.3.2. +/−effect theory uses the idea that events may have beneficial or harmful *effects* on their objects. Harmful events are said to have a −effect. Often, this harmfulness is caused by the destruction, removal or diminishment that also causes polarity shifting. While −effects and polarity shifting are not equivalent, their strong relatedness may still make knowledge about −effects a useful source of information. We use *EffectWordNet* (Choi and Wiebe, 2014), which provides effect labels for *WordNet* verb synsets. We generalize the synsets to lemmas by marking a lemma as −effect if at least one word sense of the lemma is marked as −effect and none of its word senses are marked as +effect. We rank the resulting list of −effect verbs by their word frequency (−EFFECT).

DISTRIBUTIONAL SIMILARITY (SIM): As we observed when we defined polarity shifters in section 2.1, polarity shifters and negation words are closely related. Often, the same sentence can be expressed using either a negation word or a polarity shifter, as we can see in (3.1) and (3.2).

(3.1) Peter [did **not**$_{negation}$ [pass the exam]$^{+}$]$^{-}$.
(3.2) Peter [**failed**$_{shifter}$ to [pass the exam]$^{+}$]$^{-}$.

Verbs that occur in contexts that are distributionally similar to negation words might therefore be more likely to be shifters. Using our word embedding we rank all verbs by their cosine similarity to a given negation word. The most highly ranked verbs are considered shifters.

To create these rankings, we also need to define which negation words we want to consider. For this we take the intersection of negation words found in the *negation lexicon* by Wilson et al. (2005b) and in the *negation signals* from Morante and Daelemans (2009). The resulting set of negation words comprises *neither*, *never*, *no*, *none*, *nor*, *not* and *without*.

In addition to calculating the similarity to a specific negation word, we also want to create a ranking that represents the general concept of negation, rather than a specific negation word. To achieve this, we compute the **centroid** of all individual negation words. Given n negation words, each of which is represented as a vector $\vec{v}$, we define the centroid C as:

$$C = \frac{1}{n} \sum_{i=0}^{n} \vec{v}_i \tag{3.2}$$

POLARITY CLASH (CLASH): As we observed in section 3.3.1, many shifters have a polarity of their own. These often shift expressions that are of the opposite polarity. For example, in our text corpus the negative polarity verb *ruin* is a shifter that frequently has expressions of positive polarity like *career*, *reputation*, or *enjoyment* in its shifting scope:

(3.3) It [**ruined**$^{-}_{shifter}$ her [career]$^{+}$]$^{-}$.

(3.4) This may [**ruin**$^{-}_{shifter}$ her [reputation]$^{+}$]$^{-}$.

(3.5) The constant coughing [**ruined**$^{-}_{shifter}$ my [enjoyment]$^{+}$ of the play]$^{-}$.

We expect that the more often a polar verb occurs with direct objects (which is the most common scope of verbal shifters) of the opposite polarity, the more likely it is that this verb is a polarity shifter. Due to the rarity of verbal shifters with positive polarity in the *Subjectivity Lexicon* (see Table 3.2), we look exclusively for negative verbs with positive nouns as direct objects.[2]

We rank these polar verbs by the frequency of occurring with positive nouns (CLASH), normalized by the overall frequency of the verb ($\text{CLASH}_{\text{NORM}}$).

VERB PARTICLES (PRT): Particle verbs are phrasal constructs that combine a verb and an adverbial particle, such as *"tear down"* or *"lay aside"* in (3.6) and (3.7). In many cases, the particle indicates a particular aspectual property, such as the complete transition to an end state (Brinton, 1985). In *"dry (something) out"*, for example, the particle *out* indicates that we *"dry (something) completely"*.

Polarity shifting often involves the creation of a new (negative) end state of some entity, for example through its destruction, removal or diminishment (see section 2.1). Therefore we expect a significant

2 For the feature design in this chapter, we simplify the question of determining the scope of a verbal shifter by assuming it is always the direct object. For a more thorough exploration of verbal shifting scopes, see chapter 4

number of particle verbs to be shifters. Examples for this can be seen in (3.6) and (3.7).

(3.6) This [**<u>tore down</u>**$_{shifter}$ our great [dream]$^{+}$]$^{-}$.

(3.7) Please [**lay <u>aside</u>**$_{shifter}$ all your [worries]$^{-}$]$^{+}$.

We only consider particles which typically indicate a complete transition to a negative end state: *aside*, *away*, *back*, *down*, *off* and *out*. The list of these particle verbs is ranked via the frequency of the particle verb in our text corpus, normalized over the frequency of its particle.

HEURISTIC USING *"any"* (ANY): In section 2.3.1 we discussed how **negative polarity items (NPIs)**, such as *any*, have been shown to typically appear in the context of negation (Ladusaw, 1980; Giannakidou, 2011). We hypothesize that the connection between NPIs and negation words, as in (3.8), can similarly occur between NPIs and polarity shifters, as in (3.9). This hypothesis is closely related to the work of Danescu-Niculescu-Mizil et al. (2009), who use NPI co-occurence patterns to detect downward-entailing operators, which are closely related to polarity shifters.

(3.8) They did [**not** give us <u>any</u> help$^{+}_{dobj}$]$^{-}$.

(3.9) They [**denied**$_{shifter}$ us <u>any</u> help$^{+}_{dobj}$]$^{-}$.

The feature we design collects all verbs that take a direct object that has the NPI *any* as its determiner.[3] An example of a matching occurrence can be seen in (3.10). The scope of the verbal shifter *deny* is its direct object, the noun phrase *"any help"*. The noun phrase is headed by the noun *help*, which in turn has the determiner *any*.

dobj

det

(3.10) They [**denied**$_{shifter}$ us <u>any</u> help^{+}]$^{-}$.

We sort the verbs by their frequency of co-occurrence with this particular textual pattern (ANY). We normalize the pattern frequency by the general frequency of the respective verb (ANY$_{\text{NORM}}$). As a further constraint we require that the direct object is a polar expression (ANY$_{\text{NORM+POLAR}}$). This constraint is fulfilled in (3.10) since the noun *help* is a positive polar expression.

ANTI-SHIFTERS (ANTI): Up until now all of our features have been aimed at identifying polarity shifters. However, classifiers can also profit from negative examples. In our case, these would be examples

3 Preliminary investigations of sentences containing different NPIs suggested that the NPI *any* was most promising for identifying verbal shifters.

of non-shifters. To provide strong negative examples we look for **anti-shifters**, words that are not just non-shifters, but represent the extreme opposite of shifters by creating a strong polar stability that prevents shifting. These can be used as a negative feature that indicates the absence of shifting.

To determine anti-shifters, we look for words co-occurring with specific adverbials that show attraction to verbs of creation while at the same time being repelled by verbs of destruction. Verbs of creation usually entail a positive end state (i. e. something being created), which is expected to be antithetical to shifting and indicative of anti-shifters. Verbs of destruction, on the other hand, tend to entail negative end states and polarity shifting, a phenomenon that we already made use of in the design of our particle verb feature. Using the log-likelihood collocation measure of *SketchEngine*[4] we identify the adverbials *exclusively*, *first*, *newly* and *specially*. Typical examples of co-occurrence with anti-shifters are shown in (3.11)–(3.14).

(3.11) In winter, black bears exclusively **live**$_{antiShifter}$ on fish.
(3.12) Full keyboards on cellphones were first **introduced**$_{antiShifter}$ in 1997.
(3.13) These buildings have been newly **constructed**$_{antiShifter}$.
(3.14) They specially **prepared**$_{antiShifter}$ vegan dishes for me.

To create a ranked list of anti-shifters we sort all words by their frequency of co-occurring with these adverbs, normalized by their general word frequency.

3.4.2 *Generic Features*

In addition to the task-specific features presented in section 3.4.1 we examine generic features derived from common lexical resources. Unlike the features in section 3.4.1, the generic features do not produce a ranking. Therefore, we will only be able to evaluate them in the context of a supervised classifier.

WORDNET (WN): *WordNet* (Miller et al., 1990) is the largest available ontology for the English language. It organizes words by their word senses, collecting words that share a meaning into so called **synsets** (i. e. synonym sets). An example of a synset entry can be seen in Table 3.3.

Each synset is given a **gloss**, which is a brief description of the word sense that it represents. For example, the third word sense of the verb *eliminate*, as shown in Table 3.3, has the gloss *"kill in large numbers"*. Each synset is also assigned a **supersense** (also known as **lexicographer sense**), which is a coarse semantic category, such as *change*, *body* or *cognition*. Synsets are connected to each other by a number of semantic relations. The most prominent relation is that

4 http://www.sketchengine.co.uk/

FIELD	VALUE
Synset	`eliminate.v.03`
Lemmas	eliminate, annihilate, extinguish, eradicate, wipe out, decimate, carry off
Gloss	*"kill in large numbers"*
Supersense	Change
Hypernym	`kill.v.01`

Table 3.3: Excerpt from the *WordNet* synset entry for the third word sense of the verb *"eliminate"*.

of **hypernymy**, which connects word senses with other senses that represent a more general form of the same concept. For example, the hypernym of the third word sense of the verb *eliminate* is the first word sense of the verb *kill*.

We want to leverage the information provided by *WordNet* to find semantic patterns that differentiate shifters from non-shifters. As we assign categories to words rather than senses (and due to the lack of robust word sense disambiguation) we represent a word as the union of the synsets that contain it.

For each verb we provide features based on its glosses, its supersenses and its hypernyms. Glosses are a common way to leverage *WordNet* for lexicon induction tasks in sentiment analysis (Esuli and Sebastiani, 2005; Gyamfi et al., 2009; Choi and Wiebe, 2014; Kang et al., 2014). Our expectation is that the descriptive texts of shifters will share similar word choices. To model this, we represent glosses as bags of words. Both supersenses and hypernyms have also been found to be effective features for sentiment analysis tasks, as shown by Flekova and Gurevych (2016) and Breck et al. (2007) respectively.

FRAMENET (FN): *FrameNet* (Baker et al., 1998) is a semantic resource that is based on the theory of frame semantics (Fillmore, 1967). It has successfully been used for a number of sentiment-related tasks, such as opinion spam analysis (Kim et al., 2015), opinion holder and target extraction (Kim and Hovy, 2006) and stance classification (Hasan and Ng, 2013).

FrameNet provides over 1,200 semantic frames that collect words with similar semantic behavior. We assume that polarity shifters will cluster together in specific frames. For example, the frame AVOIDING consists exclusively of verbal shifters, such as *desist*, *dodge*, *evade*, *shun*, *shirk*, etc. In our classification, we use the **frame memberships** of a verb as its features.

We used *FrameNet* version v1.6. It covers only 31.4 percent of verbs from our gold standard. To **extend coverage**, we use the semantic-parser *SemaFor* (Das et al., 2010), which can infer frames for verbs missing from FrameNet (Das and Smith, 2011). For each missing verb, we let *SemaFor* label 100 sentences from our corpus and use the frame that is assigned most often. In our exploratory experiments with supervised classification, this expansion caused a significant increase of 6 percent in F-score (paired t-test, $p < 0.05$).

3.5 EXPERIMENTS

We will now experimentally evaluate the features introduced in section 3.4. In section 3.5.1 we analyze the high-precision potential of individual task-specific features which we introduced in section 3.4.1. Section 3.5.2 introduces the classifiers that we consider for our bootstrapping task. These classifiers use both task-specific and generic features. The performance of the classifiers is evaluated in section 3.5.3 through a recall-oriented evaluation of our entire gold standard. The best classifier from this evaluation will later be used in section 3.6 to bootstrap the remaining unlabeled verbs to create a large lexicon of verbal shifters.

3.5.1 *Analysis of Task-Specific Features*

We designed the task-specific features introduced in section 3.4.1 specifically for use in polarity shifter classification. To determine the quality of these features, we perform a precision-based evaluation. Given the 2,000 verbs in our gold standard (section 3.2), each feature must generate a ranked list of potential shifters from them. Features are then evaluated on the precision of high-ranking elements from their list. Generic features (section 3.4.2) will not be evaluated in this phase as they do not generate ranked lists.

Due to the nature of the precision metric and the uneven distribution between shifters and non-shifters, we are not able to directly compare features that rank the former with those that rank the latter. Instead, we first analyze features that rank shifters in section 3.5.1.1 and then features that rank non-shifters in section 3.5.1.2.

3.5.1.1 *Ranked Lists of Shifters*

Table 3.4 shows the number of verbs retrieved by each feature, as well as the precision of the 20, 50, 100 and 250 highest ranked verbs. We compare our features against two baseline features. The first baseline is the list of all gold standard verbs ranked by their frequency in our text corpus (FREQ). The second baseline restricts that frequency-ranked list to negative polar expressions (NEGATIVE), as the ratio of

FEATURE	RETRIEVED	PRECISION@n			
		20	50	100	250
FREQ	2,000	10.0	18.0	22.0	22.0
NEGATIVE	189	30.0	30.0	29.0	*n/a*
SIM_{NOR}	1,901	15.0	24.0	16.0	18.4
$\text{SIM}_{\text{NEITHER}}$	1,901	20.0	18.0	18.0	21.6
SIM_{NONE}	1,901	25.0	24.0	22.0	21.6
SIM_{NOT}	1,901	25.0	24.0	23.0	23.2
$\text{SIM}_{\text{NEVER}}$	1,901	20.0	30.0	30.0	32.8
SIM_{NO}	1,901	35.0	28.0	36.0	28.8
$\text{SIM}_{\text{WITHOUT}}$	1,901	40.0	36.0	34.0	27.6
$\text{SIM}_{centroid}$	1,901	45.0	30.0	29.0	27.6
CLASH	107	40.0	52.0	39.0	*n/a*
$\text{CLASH}_{\text{NORM}}$	107	45.0	46.0	37.0	*n/a*
−EFFECT	175	45.0	44.0	46.0	*n/a*
PRT	165	60.0	64.0	58.0	*n/a*
ANY	539	30.0	28.0	29.0	34.0
ANY_{NORM}	539	65.0	60.0	53.0	38.8
$\text{ANY}_{\text{NORM+POLAR}}$	272	75.0	66.0	62.0	41.2
$\text{ANY}_{\text{NORM+POLAR+PAGER}}$	1,901	**80.0**	**70.0**	**63.0**	**45.2**

Table 3.4: Analysis of task-specific features (section 3.4.1) for the shifter classification of verbs. Features generate a ranked list of potential shifters and are evaluated on the precision of the list for the 20, 50, 100 and 250 highest ranked verbs. Best results are depicted in bold.

FEATURE	RETRIEVED	PRECISION@n			
		20	50	100	250
FREQ	2,000	90.0	82.0	78.0	78.0
POSITIVE	73	90.0	94.0	*n/a*	*n/a*
+EFFECT	95	90.0	92.0	*n/a*	*n/a*
ANTI	725	**95.0**	**96.0**	**93.0**	**87.4**

Table 3.5: Analysis of features for the classification of verbal non-shifters. Complement to the shifter analysis in Table 3.4. Best results are depicted in bold.

shifters to non-shifters was the greatest among these expressions (see Table 3.2).

Our distributional similarity feature (SIM) provides mixed results. The quality of the ranking strongly depends on the choice of negation word and how many of the highest ranked verbs are considered. The large number of retrieved verbs for SIM is due to the fact that embedding-based methods like SIM can rank all verbs found in the word embedding. The only gold standard verbs missing from the embedding are those that were discarded by *Word2Vec* for occurring less than five times in our text corpus. The list generated from a centroid of all considered negation words ($\text{SIM}_{centroid}$) is one of the best-performing versions of the SIM feature. As it also does not require manually selecting specific negation words, we consider it the most promising version of the SIM feature.

Our other features show better results. The polarity clash (CLASH), *EffectWordNet* (−EFFECT) and verb particle (PRT) features all clearly outperform the baselines, regardless of the cut-off value of their ranking. The number of words that each individual feature retrieves is fairly small. This shows that in order to create a shifter lexicon which is not only of high precision, but also of sufficient recall, we will have to rely on more than a single feature.

The heuristic using the negative polarity item *any* (ANY) is our strongest feature. Its performance is greatly increased by using normalized frequencies to sort its output (ANY_{NORM}) and by adding polarity restrictions ($\text{ANY}_{\text{NORM+POLAR}}$). To increase the quality of the feature even further, we apply *personalized PageRank* (Haveliwala, 2002; Agirre and Soroa, 2009) This allows us to filter out false positives and also rank verbs that were not retrieved by $\text{ANY}_{\text{NORM+POLAR}}$ by clustering together verbs with strong similarity to the majority of retrieved verbs. In its original form, *PageRank* ranks all nodes in a graph by how highly connected they are. *Personalized PageRank* extends this idea by allowing prior information to be added to the process. This information is used to assign specific re-entrance weights to nodes in the graph. The resulting non-uniform distribution causes nodes with stronger weights to be visited more often during randomized graph traversal. This introduces a ranking bias towards regions of the graph (i. e. specific sets of nodes) that are of greater relevance according to the prior information.

To apply *personalized PageRank* to ANY, we create a word-similarity graph in which all our gold standard verbs are nodes, while edges represent the distributional similarity between them. Just as for similarities to negation words (SIM, see section 3.4.1) we compute cosine similarities based on our word embedding. As prior information we give the nodes that represent words retrieved by $\text{ANY}_{\text{NORM+POLAR}}$ a uniform re-entrance weight probability, while all other nodes receive a weight of zero. Following Manning et al. (2008, ch. 21.2), we set the restart probability to $\alpha = 0.1$. The reranked list ($\text{ANY}_{\text{NORM+POLAR+PAGER}}$), which includes all gold standard verbs found in the word embedding, does indeed improve performance.

3.5.1.2 *Ranked Lists of Non-shifters*

Now that we have determined the performance of our shifter rankings, we take a look at the rankings for non-shifters. We again use a frequency-ranked list of all verbs (FREQ) as a baseline, as well as a version that is limited to verbs with positive polarity (POSITIVE), as these are rarely shifters (see Table 3.2). Analogous to −EFFECT for shifters, we consider words with +effect as possible non-shifters (+EFFECT), as the beneficial effect they have is expected to intensify or stabilize existing polarities, which is rules them out as shifters. These features are compared against our anti-shifter feature (ANTI).

In Table 3.5 we see that due to the strong bias towards non-shifters in the distribution of labels, even the baselines achieve high precision. POSITIVE and +EFFECT perform equally well, but both suffer from a small number of retrieved verbs. ANTI clearly outperforms all other features and offers a fairly large number of retrieved verbs.

3.5.2 *Classifiers*

In preparation for our bootstrapping task, we perform a recall-oriented evaluation to consider the classification of **all** verbs from our gold standard as opposed to the n-best rankings used in section 3.5.1. We consider a simple majority-class baseline ($\text{BASELINE}_{\text{MAJORITY}}$) and two types of classifiers: **graph-based classifiers** and **supervised classifiers**.

GRAPH-BASED CLASSIFIERS (LP & KNN): As graph-based classifiers, we use one based on **label propagation (LP)** (see example in Figure 3.2) as well as a **k-nearest neighbor (kNN)** classifier. Given a number of seed words and a word-similarity graph, the classifiers propagate the labels of the seeds across the graph, labeling the remaining words in the process. This means LP and kNN both require no labeled training data if the seeds are automatically determined by heuristics.

As word-similarity graph we use a graph based on cosine similarities in our word embedding. This is the same graph as we already used for the *PageRank* computation in section 3.5.1.

We provide 250 shifter words and 500 non-shifters as seeds. We use twice as many non-shifters as we use shifters to account for the higher frequency of non-shifters while avoiding overfitting to the statistics from Table 3.1.

As shifter seeds we use the highest-ranked words from our best task-specific shifter feature, $\text{ANY}_{\text{NORM+POLAR+PAGER}}$ (see section 3.5.1.1).

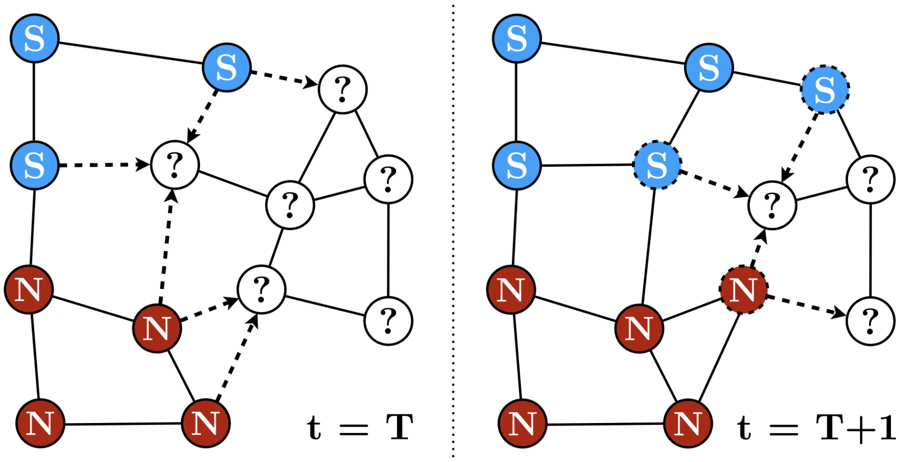

Figure 3.2: Simplified example of the label propagation algorithm (LP). Seeds for shifters (S) and non-shifters (N) are propagated to unlabeled nodes (?), based on their word similarity. This process is reiterated until all nodes are labeled.

For non-shifter seeds we use the highest-ranked words from the anti-shifter feature ANTI.

In order to examine whether anti-shifters are actually necessary to get negative seeds of sufficient quality, we also run an alternative setting ($\text{LP}_{\text{No Anti-shifter}}$ and $\text{kNN}_{\text{No Anti-shifter}}$) in which the same number of negative seeds is simply extracted from the ranking of frequent verbs. The reasoning behind this is that the proportion of frequent verbs not being shifters is already fairly high, as shown by FREQ in Table 3.5.

For the label propagation classifiers we use the **Adsorption** label propagation algorithm (Baluja et al., 2008) as it was implemented in *Junto* (Talukdar et al., 2008). For the k-nearest neighbor classifiers, we set $k = 10$.

SUPERVISED CLASSIFIERS (SVM): Apart from the graph-based classifiers, we also consider a **supervised classifier**, namely **Support Vector Machines (SVMs)** as implemented in *SVM*$^{\text{light}}$ (Joachims, 1999). This classifier requires manually labeled training data, but, unlike LP and kNN, we may combine arbitrary feature sets.

We train a classifier using the task-specific features which we defined in section 3.4.1 and another classifier which uses the generic features described in section 3.4.2. For the task-specific features we use their most complex configurations from Table 3.4 (e. g. $\text{SIM}_{\textit{centroid}}$ rather than SIM_{NOR} or $\text{SIM}_{\text{WITHOUT}}$). For a third classifier, we combine the two feature sets, creating a classifier that makes use of the entire available range of features.

CLASSIFIER	PREC	REC	F1
Baseline$_{\text{majority}}$	42.4	50.0	45.9
κNN$_{\text{No Anti-shifter}}$	54.9	56.4	55.6*
κNN	58.3	59.6	58.9*
LP$_{\text{No Anti-shifter}}$	63.0	56.6	59.6*
LP	68.6	56.7	62.0*
SVM$_{\text{task-spec. features (section 3.4.1)}}$	65.5	69.7	67.5*
SVM$_{\text{generic features (section 3.4.2)}}$	79.6	74.4	76.9*
SVM$_{\text{all features}}$	**80.7**	**77.6**	**79.1***

*: F1 is better than previous classifier (paired t-test with $p < 0.05$).

Table 3.6: Evaluation of classifiers (section 3.5.3) on the 2,000 verbs from the gold standard (Table 3.1). The evaluation is run as a 10-fold cross validation and all reported metrics are macro-averages. Best results are depicted in bold.

3.5.3 *Classifier Evaluation*

To evaluate the classifiers, we perform 10-fold cross validation. This means we perform ten iterations of training on 1,800 words and evaluating on the remaining 200, replacing the set of evaluated words each time until all words have been evaluated. We compute the averaged performance across the 10 iterations, reporting macro-average precision, recall and F-score.

Table 3.6 shows that among the graph-based classifiers, LP is notably better than kNN. Both classifiers benefit from anti-shifter seeds. Supervised classification outperforms graph-based classification, so using labeled training data is beneficial. It also means that the *full* set of task-specific shifter features (section 3.4.1) is more effective than just the strongest feature, ANY$_{\text{norm+polar+pageR}}$, which is used to determine shifter seeds for the graph-based classification). While the generic features outperform the task-specific features (in supervised classification), combining them results in another significant improvement, demonstrating the importance of the task-specific features.

3.5.4 *Training Size Requirements*

One of our motivations for investigating means of automatically classifying polarity shifters is to reduce the amount of human annotation that is required to create a lexicon of adequate size. At the same time, our SVM classifiers require human-annotated words to train their models. The quality of these models will then determine both how

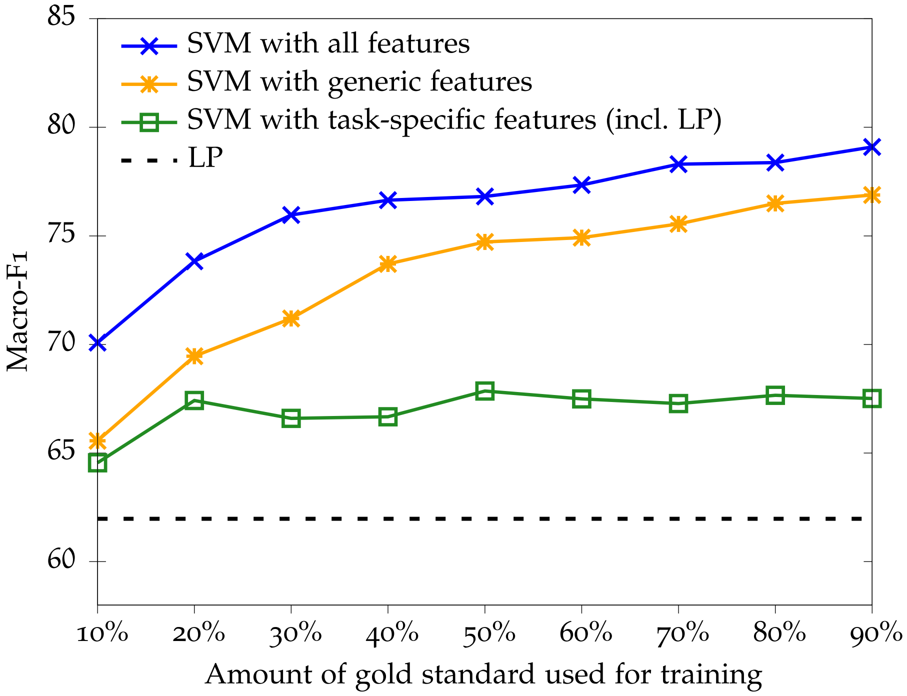

Figure 3.3: Learning curve for supervised training. This repeats the evaluation of section 3.5.3, but reduces the amount of training data. At 90 percent training data this task is identical to the one reported in Table 3.6.

many shifters are detected and how many classification errors need to be removed by a human annotator in our bootstrap verification step in section 3.6 (see also Figure 3.1). We are therefore faced with a trade-off between pre- and post-processing annotation efforts. Pre-annotating too many words weakens the advantage in work reduction that automatic classification offers, but pre-annotating too few words runs the risk of missing too many shifters or causing large amounts of post-annotation.

As we use 10-fold cross validation in our experiments, each supervised classifier is usually trained on 90 percent of our gold standard, i. e. on 1,800 verbs. To get a sense of how much pre-annotation is indeed required to achieve acceptable classifier performance, we now evaluate our classifiers on reduced amounts of training data. Otherwise, all experimental parameters are identical to those used in section 3.5.2.

Figure 3.3 displays the learning curve of the major feature set configurations of the SVM classifier. While the task-specific features on their own are always worse than the generic features, a classifier combining those feature groups always outperforms the classifier solely trained on the generic features. This improvement is particularly large when little labeled training data is available, which is a typical scenario for lexical bootstrapping tasks. Figure 3.3 also shows that the SVM classifier has reached roughly the point of saturation when using all features and the maximal amount of labeled training data. This amount should be sufficient for bootstrapping our gold standard lexicon on further unlabeled verbs, as will be shown in section 3.6.

3.6 BOOTSTRAPPING THE LEXICON

We now bootstrap a larger list of shifters from the remaining unlabeled 8,581 *WordNet* verbs not included in our gold standard (section 3.2). On this verb set we run an SVM trained on the gold standard (2,000 verbs) with the best performing feature set (Table 3.6). The classifier predicts 1,043 verbs as shifters. The remaining 7,538 instances predicted as non-shifters will not be considered further.

As our classifier reached a precision of 93.1 percent on non-shifters on our gold standard, we are confident that the predicted non-shifters include few actual shifters. As our precision for shifters is lower, i. e. 68.3 percent, we manually check the predicted shifter instances. Using our classifier to pre-filter the data (Choi and Wiebe, 2014) reduced the number of verbs that still had to be annotated manually by 87.8 percent from 8,581 to just 1,043 instances. Across the entire vocabulary of verbs, counting both gold standard and bootstrapped verbs, we manually annotated 3,043 verbs. This is an enormous reduction of over 70 percent in annotation effort.

Figure 3.4 shows the precision on different intervals ranked by confidence score of the SVM on the predicted 1,043 shifters. Since the top quarter words reach a very high precision, with hindsight, a manual annotation of at least these instances would not even have been necessary.

Among the 1,043 predicted shifters, manual annotation confirmed 676 actual shifters. In total we produced a novel list of 980 verbal shifters (304 gold standard + 676 bootstrapping) in this chapter.

3.7 CONCLUSION

We took a first step toward producing a comprehensive lexicon of polarity shifters by bootstrapping a large list of verbal polarity shifters. Using a random sample of 2,000 manually annotated verbs, we built a supervised classifier to pre-filter the remaining verbs. Verbs that were predicted to be shifters were then verified by a human annotator to ensure that the lexicon is of high precision. This reduced the number of verbs that needed to be annotated manually by over 70 percent

Our bootstrapping approach makes use of a variety of linguistic phenomena and resources. These include distributional similarity to negation words, the co-occurrence of opposite polarities and the semantic properties of specific verb particles. General semantic information is gathered from *WordNet* glosses, hypernyms and supersenses

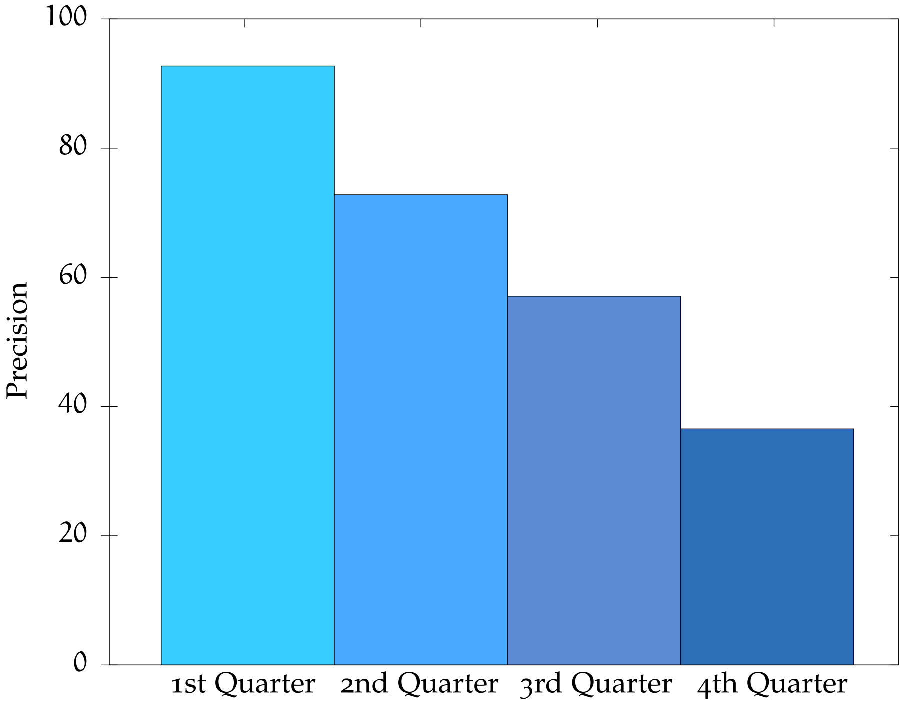

Figure 3.4: Evaluation of the bootstrapping of English verbal shifters that were not part of the gold standard (compare Table 3.1). SVM classifiers provide a confidence value for each label they assign. We rank the 1,043 potential shifters from highest to lowest confidence and group them, so that the first quarter contains the highest and the fourth quarter the lowest confidence candidates.

as well as from *FrameNet* frame relations. We also successfully build on insights from related fields of research, observing the close relatedness of polarity shifters with both −effects and downward-entailing operators. In the case of −effects we make use of an existing lexicon for +/−effect. From downward-entailing operators we adapt the insight that they often co-occur with negative polarity items and can therefore be learned in an unsupervised fashion by observing which words occur most frequently with NPIs. The resulting heuristic of detecting shifters through their co-occurrence with the NPI *any* proves to be our strongest unsupervised feature.

Our task-specific features were shown to already work well with small amounts of training data. Generic features, on the other hand, reach their true potential when provided with larger amounts of training data. Best results are always achieved when employing all presented features.

In future chapters, we will investigate a number of ways in which our lexicon can be built upon and extended. In chapter 4, we repeat our annotation of verbal shifters, this time with a focus on distinguishing different word senses and identifying the scope of the shifting effect. In chapter 5 we transfer our lexicon and the bootstrapping process to another language. For chapters 6 and 7 we return to the lexicon we began building in this chapter. Chapter 6 adds nominal and adjectival shifters, creating a general lexicon of polarity shifters, and chapter 7 adds information on whether individual shifters can affect both positive and negative polarities or just one or the other. Finally, in chapter 8 we put the lexicon to the test and evaluate it in a sentiment analysis task.

4 ANNOTATING A SENSE-LEVEL SHIFTER LEXICON

In chapter 3 we created a lexicon of verbal polarity shifters using a combination of manual annotation and automatic classification. The classification was used to filter the vocabulary of verbs, removing words that were likely to be non-shifters. This filtered out 87.8 percent of the verbs under consideration. The remaining verbs, which were deemed by the classifier to be shifters, were then verified by a human annotator.

While the bootstrap approach greatly reduced the annotation effort involved, it also resulted in a number of trade-offs. Firstly, its reliance on automatic filtering means that some shifters are likely to have been removed due to false negative classifications. This would have resulted in an incomplete lexicon that is missing some verbal shifters. Secondly, to allow for data-driven features without the reliance on word sense disambiguation, the annotation was restricted to labels at the lemma-level, ignoring the fact that some words only function as polarity shifters in some of their word senses. Lastly, we only annotated whether a word is a shifter, but no further information, such as the scope of its shifting effect, was provided. Wiegand et al. (2018b) have shown that considering specific syntactic scopes for polarity shifters can improve negation modeling for polarity classification.

In this chapter we present an alternative lexicon of verbal polarity shifters. This lexicon was annotated entirely by hand. For each word, every word sense is annotated separately for whether it causes polarity shifting. Senses labeled as shifters are furthermore annotated for the scope of their shifting effect.

Contents

The lexicon in this chapter differs significantly from the one we presented in chapter 3. One of the most significant differences, which we motivate in section 4.1, is that we choose to annotate individual word senses, rather than lemmas. In addition we explicitly specify the syntactic scope of the shifting effect, as explained in section 4.2. The annotation process itself, described in section 4.3, is performed entirely by hand. In section 4.4 we describe the data format of the resulting lexicon and in section 4.5 we provide a statistical analysis of its contents. Section 4.6 concludes the chapter.

Contributions

(i) We create a complete lexicon of verbal polarity shifters, providing explicit annotations for each verb found in *WordNet 3.1*.

(ii) Our lexicon provides a fine grained annotation, labelling every word sense of a verb separately.

(iii) Each word sense is also annotated for its shifting scope, indicating which parts of a sentence are affected by the shifter.

Publication History

The work presented in this chapter has previously been published in Schulder et al. (2018b). The resulting sense-level lexicon of verbal shifters has been released publicly.[1]

4.1 WORD SENSES

Many words that shift polarities only do so for some of their word senses. For example, *mark down* acts as a shifter in (4.1), where it has the sense of *"reducing the value of something"*, but the sense of *"writing something down to have a record of it"* in (4.2) causes no shifting. In our work we found that among shifter lemmas with multiple word senses, only 23 percent caused shifting in each of their senses. An annotation on the basis of individual word senses is therefore required.

(4.1) The agency [**marked down**$_{shifter}$ [their assets]$^{+}$]$^{-}$.

(4.2) She [**marked down**$_{non\text{-}shifter}$ [his confession of guilt]$^{-}$]$^{-}$.

To differentiate the senses of a verb, we use its synset affiliations found in *WordNet*. Synsets are a collection of lemmas that share the same word sense. Each synset therefore represents a word sense and lists lemmas that can be used to express that word sense.

Words within the same synset share a shifter label. On the other hand, the scope of a shifter, i. e. the part of an expression that can have its polarity changed by the shifter, can differ among words of the same synset (see section 4.2). This is because the scope of a shifter depends on the syntactic properties of the lemma, but synsets can contain lemmas with different syntactic behaviors.

The annotation which we introduce in section 4.3 is therefore applied to individual lemma-synset pairs, i. e. separately for each lemma in a specific synset. This allows us to provide both shifter labels and shifting scope labels in the same annotation.

1 https://doi.org/10.5281/zenodo.3365287

4.2 SHIFTING SCOPE

A verbal shifter usually only affects the parts of a sentence that are syntactically governed by the verb through its valency. However, not every argument of a verbal shifter is subject to polarity shifting. Which argument is affected by polarity shifting depends on the verb in question. In (4.3), *surrender* shifts only the polarity of its subject, but does not affect the object. Conversely, *defeat* shifts only its object in (4.4). The polarity of the subject of *defeat* does not play a role, as can be seen in (4.5).

subj

(4.3) [[The villain]$^-$ **surrendered**$_{shifter}$]$^+$ to [the hero]$^+$.

dobj

(4.4) [The villain]$^-$ [**defeated**$_{shifter}$ [the hero]$^+$]$^-$.

dobj

(4.5) Chance [**defeated**$_{shifter}$ [the hero]$^+$]$^-$.

In the following, we present the shifting scopes we observed, the abbreviations we use for them in the annotation and examples for each scope.

SUBJECT (SUBJ): The verbal shifter affects its subject.

subj

Example: [[His confidence]$^+$ **decreased**$_{shifter}$]$^-$.

DIRECT OBJECT (DOBJ): The verbal shifter affects its direct object.

dobj

Example: The storm [**ruined**$_{shifter}$ [their party]$^+$]$^-$.

PREPOSITIONAL OBJECT (POBJ_*): The verbal shifter affects the object within a prepositional phrase. The preposition in question is included in the annotation.

Example for `pobj_from`:

pobj prep

The wall [**shielded**$_{shifter}$ them [from the explosion]$^-$]$^+$.

Example for `pobj_for`:

pobj prep

The company [**reimbursed**$_{shifter}$ him [for his expenses]$^-$]$^+$.

CLAUSAL COMPLEMENT (COMP): The verbal shifter affects a clausal complement, such as an infinitive clauses or gerund.
Example for infinitive clause complement:

comp

He [**failed**$_{shifter}$ [to pass the exam]$^{+}$]$^{-}$.

Example for gerund complement:

comp

She [**stopped**$_{shifter}$ [using drugs]$^{-}$]$^{+}$.

The given scopes assume that verb phrases are in their active form. In passive phrases, subject and direct object roles are inverted. To avoid this issue, sentence structure normalization should be performed before computing the shifting scope.

Synsets in *WordNet* only capture the semantic similarity of words, but almost no syntactic properties (Ruppenhofer and Brandes, 2015). The shifting scope of a verb depends on its syntactic arguments, which can differ between verbs of the same synset. For example, *discard* and *dispose* share the sense *"throw or cast away"*, but while *discard* shifts its direct object (4.6), *dispose* requires a prepositional object (4.7). For this reason we annotate lemma-synset pairs individually, instead of assigning scope labels to an entire synset.

(4.6) He [**discarded** [the evidence]$^{+}_{dobj}$]$^{-}$.

(4.7) He [**disposed** [of the evidence]$^{+}_{pobj}$]$^{-}$.

We also consider cases where a verbal shifter has more than one potential scope for the same lemma-synset pair. For example, *infringe* can shift its direct object or various prepositional objects, as seen in (4.8)–(4.10). Therefore, *infringe* receives the scope labels `dobj`, `pobj_on` and `pobj_upon`.

(4.8) The inquiry [**infringes** [people's privacy]$^{+}_{dobj}$]$^{-}$.

(4.9) The inquiry [**infringes** [on people's privacy]$^{+}_{pobj}$]$^{-}$.

(4.10) The inquiry [**infringes** [upon people's privacy]$^{+}_{pobj}$]$^{-}$.

A particular verbal shifter instance will only ever shift the polarity of one of its scopes. The scope affected by the shifting depends on the sentence context.

4.3 ANNOTATION

The entire dataset was labelled by an expert annotator with experience in linguistics and annotation work. To measure inter-annotator agreement, a second annotator re-annotated 400 word senses for their shifter label. They achieved a Cohen's kappa agreement of $\kappa = 0.73$, indicating substantial agreement (Landis and Koch, 1977).

LABEL	LEMMAS FREQ.	LEMMAS PERC.	SYNSETS FREQ.	SYNSETS PERC.	LS PAIRS FREQ.	LS PAIRS PERC.
Shifter	1,220	11.53%	924	6.88%	2,131	8.88%
Non-shifter	9,357	88.47%	12,502	93.12%	21,855	91.12%
Total	10,577		13,426		23,986	

Table 4.1: The percentage of verbal shifters in *WordNet*. Lemmas are counted as shifters when at least one sense is a shifter. *"LS Pairs"* represents lemma-synset pairs.

The annotation progressed as follows: Given a complete list of *WordNet* verb lemmas, the annotator would inspect one lemma at a time. For this lemma, all applicable word senses were retrieved from the set of synsets. For each such pair of a lemma and a synsets, the annotator decided whether it was a shifter or not. Decisions were based on the sense definition of the synset and whether sentences using this sense of the lemma cause shifting. If a word sense was labelled as a shifter, it was subsequently also annotated for its potential shifting scopes.

In cases where label conflicts between different lemma-synset pairs from the same synset were encountered, these labels were reconsidered by the annotator. This introduced an additional robustness to the annotation as it let the annotator revisit challenging cases from a new perspective.

The resulting list of lemma-synset pairs provides more fine-grained information than what could be achieved by either an annotation that only covered word lemmas or one that only covered synsets (see sections 4.1 and 4.2).

4.4 DATA

In this section we describe the data formats in which we provide the lexicon. Section 4.4.1 describes the format of our main file, which contains the entire lexicon in its most detailed form. Based on this main lexicon we also derive two auxiliary lexica in section 4.4.2. These provide complete labelled lists of all verb lemmas and all verb synsets respectively.

4.4.1 *Main Lexicon File Format*

We provide our main lexicon as a comma-separated value (csv) file in which each line represents a specific lemma-synset-scope triple of a verbal shifter. Each line follows the format "`LEMMA,SYNSET,SCOPE`".

The fields are defined as follows:

`lemma`: The lemma form of the verb.

`synset`: The numeric identifier of the synset, commonly referred to as *offset* or *database location*. It consists of 8 digits, including leading zeroes (e. g. 00334568).

`scope`: The scope of the shifting. Given as `subj` for subject position, `dobj` for direct object position and `comp` for clausal complements. Prepositional object positions are given as `pobj_*`, where `*` is replaced by the preposition in question, e. g. `pobj_from` for objects with the preposition *'from'* or `pobj_of` for the preposition *'of'*.

When a lemma has multiple word senses, a separate entry is provided for each lemma-synset pair. When a lemma-synset pair has multiple potential shifting scopes, a separate entry is provided for each scope. Any combinations not provided are considered not to exhibit shifting. Take, for example, the set of entries for *"blow out"*, which occurs in the synsets 00436247, 02767855 and 02766970:

(4.11)
```
blow out,00436247,subj
blow out,02767855,dobj
```

It tells us that *blow out* in the sense 00436247 (*"melt, break, or become otherwise unusable"*) is a shifter that affects its subject. The sense 02767855 (*"put out, as of fires, flames, or lights"*) also exhibits shifting, but this time affects the direct object. It is, however, not a shifter for sense 02766970 (*"erupt in an uncontrolled manner"*).

For an example of multiple scopes for the same word sense, consider *cramp*:

(4.12)
```
cramp,00237139,dobj
cramp,00237139,pobj_in
```

Its sense 00237139 (*"prevent the progress or free movement of"*) can shift the polarity of either its direct object (e. g. *"it cramped his progress"*) or that of a prepositional object with the preposition *'in'* (e. g. *"it cramped him in his progress"*). The three other senses of *cramp* given by *WordNet* are not considered shifters.

4.4.2 *Auxiliary Lexica*

Our main lexicon is labelled at the lemma-synset pair level to provide the most fine-grained level of information possible. It can, however, easily be used in more coarse-grained applications. As a convenience measure for the reader, we provide lemma- and synset-level auxiliary lexica that list all *WordNet* lemmas and all *WordNet* synsets respectively, accompanied by their shifter label. A lemma is labelled as a shifter if at least one of its senses is considered a shifter in our main lexicon.

scope	frequency	percentage
subj	402	18.11%
dobj	1,574	70.90%
pobj_*	212	9.55%
comp	32	1.44%
Total	2,220	

Table 4.2: Distribution of shifting scopes for individual word senses. Total is higher than number of lemma-synset pairs (Table 4.1) as 4 percent of shifters have multiple potential scopes.

Similarly, a synset is labelled as a shifter if at least one of its lemma-realizations is a shifter. Shifter scopes are only provided in the main lexicon.

4.5 statistics

In Table 4.1 we present the percentage of shifters among the verbs contained in *WordNet*. While only about 10 percent of verbs are shifters, this still results in 924 synsets and 1,220 lemmas, which is 240 lemmas more than the bootstrapped verbal shifter lexicon from chapter 3 contains.

49 percent of verbs in *WordNet* are polysemous, i. e. they have multiple meanings. Among verbal shifters, this ratio is considerably higher, reaching 73 percent. Of these, only 23 percent are shifters in all of their word senses.

To get an idea of how common verbal shifters are in actual use, we computed lemma frequencies over the *Amazon Product Review Data* corpus (Jindal and Liu, 2008), which comprises over 5.8 million reviews. We found this corpus suitable due to its size, sentiment-related content and our successful previous use of it in chapter 3.

We observe 1,163 different verbal shifter lemmas with an overall total of 34 million occurrences. Correcting for non-shifter senses of shifter lemmas[2], we still estimate 13 million occurrences, accounting for 5 percent of all verb occurrences in the corpus. For comparison, the 15 negation words included in the negation lexicon by Wilson et al. (2005b) occur 13 million times as well. While the frequency of individual negation (function) words is unsurprisingly higher, the total number of verbal shifter occurrences highlights that verbal shifters are just as frequent and should not be ignored.

2 Due to the lack of robust word-sense disambiguation tools, we estimate the likelihood that a lemma instance functions as a shifter by counting the numer of shifter word senses of the lemma. A lemma with 3 shifter senses and 1 non-shifter sense would, therefore, be given a likelihood of 0.75 of being a shifter.

Statistics on the distribution of shifting scopes can be found in Table 4.2. 70.90 percent of verbal shifters have a direct object scope and 9.55 percent a prepositional object scope. Among these, *'from'* is the most common preposition at 51 percent, followed by *'of'* with 22 percent. 18.11 percent shift the polarity of their subject and only 1.44 percent shift that of a clausal complement. This distribution shows that shifting cannot be trivially assumed to always affect the direct object and thus explicit knowledge of shifting scopes could be useful for judging the polarity of a phrase.

4.6 CONCLUSION

In this chapter, we introduced a second lexicon of verbal polarity shifters. Unlike the lexicon from chapter 3, this lexicon is annotated entirely by hand, instead of relying on a supervised classifier to pre-filter the list of words to be inspected by a human annotator. This means that it provides explicit labels for the entire verb vocabulary of *WordNet*.

While the lexicon from chapter 3 contains a single label for each word lemma, the lexicon presented in this chapter takes a more fine-grained approach, providing labels for each individual word sense. In addition, it also explicitly identifies the potential syntactic scopes of each shifter.

In chapter 8, we will investigate whether the fine-grained sense-level annotation of this shifter lexicon provides an advantage over the bootstrapped lemma-level annotation. Section 8.4 contains a direct comparison of the two lexica in the context of a polarity classification task.

An issue that remains is the sheer number of working hours such annotation tasks cost. The annotation of 10,000 verbs in this chapter alone took over 370 hours. While it is possible to fund such work for one language, creating resources for more languages without leveraging the knowledge gained in the first would be excessive. In the upcoming chapter 5 we will address this concern during the creation of a lexicon of verbal shifters for the German language.

5 CROSS-LINGUAL BOOTSTRAPPING

So far, our work has focussed on polarity shifters in the English language. However, shifters also occur in other languages. For example, the negated statement in (5.1) 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 (5.2).

(5.1) Peter hat ihnen [**nicht**$_{negation}$ [geholfen]$^{+}$]$^{-}$.
Peter [did **not**$_{negation}$ [help them]$^{+}$]$^{-}$.

(5.2) Peter hat es [**unterlassen**$_{shifter}$, ihnen [zu helfen]$^{+}$]$^{-}$.
Peter [**failed**$_{shifter}$ to [help them]$^{+}$]$^{-}$.

As in English, German shifters can affect both positive and negative expressions. In (5.3) 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 (5.4) creates a positive sentence despite the negative polar expression *Schmerz*/*pain*.

(5.3) Ihr wurde das [[Stipendium]$^{+}$ **verweigert**$_{shifter}$]$^{-}$.
She was [**denied**$_{shifter}$ the [scholarship]$^{+}$]$^{-}$.

(5.4) Die neue Behandlung hat ihre [[Schmerzen]$^{-}$ **gelindert**$_{shifter}$]$^{+}$.
The new treatment has [**alleviated**$_{shifter}$ her [pain]$^{-}$]$^{+}$.

As can be seen for *verhindern*/*prevent* in (5.5) and (5.6), the same shifter can affect both positive and negative expressions.

(5.5) Seine Prinzipien [**verhinderten**$_{shifter}$ eine [Einigung]$^{+}$]$^{-}$.
His principles [**prevented**$_{shifter}$ an [agreement]$^{+}$]$^{-}$.

(5.6) Ihre Maßnahmen [**verhinderten**$_{shifter}$ ein [Gemetzel]$^{-}$]$^{+}$.
Their measures [**prevented**$_{shifter}$ a [slaughter]$^{-}$]$^{+}$.

As is the case with English, there is a lack of resources that identify polarity shifters in German. To improve this situation, we adapt and extend the shifter lexicon bootstrapping approach that we introduced in chapter 3 and use it to create a lexicon of German verbal polarity shifters.

The adapted approach is shown in Figure 5.1. As before, we had a human annotator label a random sample of verbs, which is later used as a gold standard. Unlike in our previous bootstrapping, we now also have a pre-existing lexicon of verbal shifters available, albeit not in our target language. We leverage this advantage by introducing the English lexicon of verbal shifters from chapter 3 as a resource and create features that make use of the lexicon in addition to German

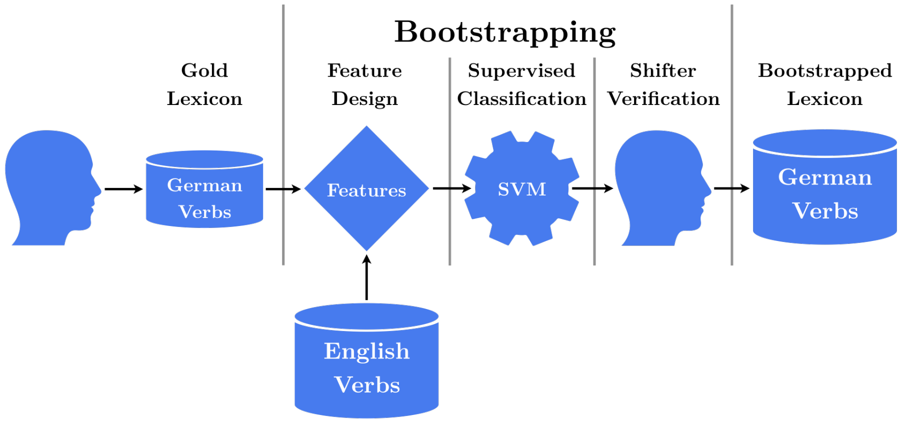

Figure 5.1: Workflow for bootstrapping the German lexicon of verbal polarity shifters. Like in the English workflow (Figure 3.1), a human annotator creates a gold standard for training a supervised classifier and later verifies verbs that were classified as shifters. As a new addition to the process, the English shifter lexicon from chapter 3 is also used as an input resource for classifier features.

versions of the previously established features. Using our German gold standard and features, we train the supervised classifier and use it to label the remaining German verbs. Finally, verbs labeled as shifters are confirmed by the human annotator before being added to our lexicon of German verbal shifters.

The bootstrapping approach from chapter 3 includes features that rely on semantic resources as well as ones that are data-driven. One reason we chose German as the second language to which to apply the approach is that the resources required to reproduce all features are available for German. 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 semantic resources which may not be available for many under-resourced languages.

Applying our approach to a second language also serves to validate whether the features we use in our bootstrap classifier are language-independent and to show what changes are necessary to adapt them to a new language. In addition we improve the bootstrapping process by introducing features that leverage available knowledge about shifters across languages.

Contents

We briefly introduce our gold standard for this new bootstrapping task in section 5.1, followed by a description in section 5.2 of the German-language resources that we use. In section 5.3 we explain how we adapt the features from chapter 3 to German and introduce entirely new features as well. The features are then used to train classifiers,

LABEL	FREQUENCY	PERCENTAGE
Shifter	224	11.2%
Non-shifter	1,776	88.8%
Total	2,000	

Table 5.1: Distribution of verbal shifters in annotated sample of 2,000 verbs taken from *GermaNet*.

which are evaluated in section 5.4. Using the best classifier from that evaluation, we bootstrap our lexicon of German verbal shifters in section 5.5. The findings of this chapter are then summarized in section 5.6.

Contributions

(i) We introduce a German lexicon of verbal polarity shifters.

(ii) We adapt the bootstrapping approach from chapter 3 to German.

(iii) We introduce cross-lingual methods that take advantage of the existence of the English verbal shifter lexicon and improve upon the current state-of-the-art.

Publication History

Work presented in this chapter has previously been published in Schulder et al. (2018a). The bootstrapped lexicon of German verbal shifters has been released publicly.[1]

5.1 GOLD STANDARD

We created a gold standard for German verbal shifters, following the approach we used in section 3.2 for our English gold standard. An expert annotator labeled 2,000 verbs, randomly sampled from *GermaNet* (Hamp and Feldweg, 1997), a German wordnet resource. The remaining 7,262 *GermaNet* verbs in our vocabulary (see section 2.2.1) are used to bootstrap a larger lexicon in section 5.5.

As in chapter 3, annotation is performed at the lemma level. While we showed in chapter 4 that some words are only shifters in some of their word senses, we would require automatic word sense disambiguation (WSD) to take this into account for the automatic classification in our bootstrapping approach. We are not aware of WSD tools

1 https://doi.org/10.5281/zenodo.3365369

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 Review Data (Jindal and Liu, 2008)
Polarity Lex.	PolArt Sentiment Lexicon (Klenner et al., 2009)	Subjectivity Lexicon (Wilson et al., 2005a)
Framenet	Salsa (Burchardt et al., 2006)	FrameNet (Baker et al., 1998)
Effects	EffektGermaNet (Ruppenhofer and Brandes, 2015)	EffectWordNet (Choi et al., 2014)

Table 5.2: German resources used, compared with English resources used in chapter 3.

for German that would be sufficiently robust for this task. The matter of using WSD for shifter identification will be revisited in section 8.4.

Table 5.1 shows that in our gold data 11.2 percent of verbs are shifters. This is a bit less than the 15.2 percent of the English bootstrapping gold standard from chapter 3, but is very close to the shifter frequency of 11.5 percent found in the manually annotated English shifter lexicon from chapter 4.

5.2 RESOURCES

To perform bootstrapping for German verbal shifters, we naturally need to replace the English-language resources we used in chapter 3 with German-language ones. An overview of the resources we use in this chapter and their English-language equivalents can be found in Table 5.2.

Note that to differentiate between the general concept of wordnets and the eponymous *Princeton WordNet*, which is a wordnet for the English language, we spell the general concept as a regular word (*"wordnet"*) and the specific resource for English with camelcase (*"WordNet"*). The same goes for the concept of framenets and the specific resource for English, *FrameNet*.

For *WordNet*, *FrameNet* and *EffectWordNet* there exist equivalent German resources: *GermaNet* (Hamp and Feldweg, 1997), *Salsa* (Burchardt et al., 2006) and *EffektGermaNet* (Ruppenhofer and Brandes, 2015), respectively. As a polarity lexicon, we use the *PolArt Sentiment Lexicon* by (Klenner et al., 2009). The distribution of positive and negative verbs in the *PolArt Sentiment Lexicon* with respect to polarity shifters can be seen in Table 5.3 As was the case for the English gold standard,

	POSITIVE VERBS		NEGATIVE VERBS	
LABEL	FREQUENCY	PERC.	FREQUENCY	PERC.
Shifter	12	11.7%	69	27.9%
Non-shifter	91	88.3%	178	72.1%
Total	103		247	

Table 5.3: The distribution of sentiment polarities among verbal shifters from the German gold standard. Polarities are automatically determined using the *PolArt Sentiment Lexicon* (Klenner et al., 2009).

we find a tendency for shifter verbs to be negative rather than positive expressions.

As a text corpus, we use the *DeWaC Web Corpus* (Baroni et al., 2009), a web corpus of 1.7 billion words. The corpus was lemmatized using *TreeTagger* (Schmid, 1994) and parsed for syntactic dependency structures with *ParZu* (Sennrich et al., 2009).

Ideally, we would have used a product review corpus as our text corpus, as we did for English, especially in view of our efforts to create cross-lingual features, one of which employs both a German and an English text corpus (see section 5.3.2.2). Unfortunately, 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 Data* consists of 1.2 billion words. Several of our features rely on word embeddings. When generating such embeddings, the size of the corpus is very important for the quality of the resulting embedding. Therefore we believe it to be of greater importance to choose a sufficiently large corpus, like *DeWaC*, rather than sticking to a specific domain.

Other than the choice of corpus, we use the same approach for generating our German word embeddings as we did for the English embedding. We use the *Word2Vec* tool by Mikolov et al. (2013) and use the same hyperparameters as in chapter 3. Using the *continuous bag of words* algorithm we generate a vector space with 500 dimensions. All other settings are kept at their default.

5.3 FEATURE DESIGN

In this section we introduce the features that we use to bootstrap our German verbal shifter lexicon in section 5.5. We start by outlining how we adapt the features that were introduced in chapter 3 for use with German (section 5.3.1). We further separate them into data-driven features (section 5.3.1.1) and resource-driven features (section 5.3.1.2) to highlight their requirements when applied to a new language.

In section 5.3.2 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 (section 5.3.2.1) and the other cross-lingual word embeddings (section 5.3.2.2).

5.3.1 *Adapted Features*

In this section we briefly describe how we adapt the English verbal shifter features from chapter 3 to German language data. We distinguish between features that mainly rely on text data from a corpus and those that require complex semantic resources. When working with languages with scarcer resources, it can be expected that the former will be more readily available than the latter.

5.3.1.1 *Data-driven Features*

The main requirement of the following features is a reasonably sized text corpus to detect syntactic patterns and word frequencies. The corpus must to be lemmatized and parsed for syntactic dependency structures.

For languages in which an appropriate dependency parser is not available, a part of speech tagger may be used to approximate the required syntactic structures (Riloff et al., 2013).

Some features also require knowledge about word polarities. 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, the *PolArt Sentiment Lexicon*, 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 the vector space of our word embedding (see section 5.2). As negation seeds we use German translations of the English negation seeds used in section 3.4.1.

POLARITY CLASH (CLASH): The polarity clash feature is based on the expectation that shifting will often occur when a polar verb modifies an expression of the opposite polarity, such as in (5.7). The feature is further narrowed down to negative verbs that modify positive nouns, as polar verbal shifters are predominantly of negative polarity (Table 5.3).

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

PARTICLE VERBS (PRT): Certain verb particles indicate a complete transition to an end state (Brinton, 1985). We previously hypothesized in chapter 3 that this phenomenon correlates with shifting, which can be seen as producing a new (negative) end state. In chapter 3 we collected 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): As we discussed in section 2.3.1, **negative polarity items (NPIs)** are known to occur in the context of negation (Giannakidou, 2011). In chapter 3 we 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 expected the same for the German NPI *jeglich*, as seen in (5.8). We collected 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 *personalized PageRank* (Agirre and Soroa, 2009).

(5.8) Sie [**verwehrten**$_{shifter}$ uns <u>jegliche</u> [Hilfe$_{dobj}$]$^{+}$]$^{-}$.
They [**denied**$_{shifter}$ us <u>any</u> [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 *leben/live*, *einführen/introduce*, *bauen/construct* or *zubereiten/prepare*. Such verbs often co-occur with the adverbs *ausschließlich/exclusively*, *zuerst/first*, *neu/newly* and *extra/specially*, as seen in (5.9)–(5.12). Accordingly, we can create a list of anti-shifters by selecting the verbs that most often co-occur with these adverbs.

(5.9) Im Winter **leben**$_{anti\text{-}shifter}$ Schwarzbären <u>ausschließlich</u> von Fisch.
In winter, black bears <u>exclusively</u> **live**$_{anti\text{-}shifter}$ on fish.

(5.10) Komplette Tastaturen auf Handys wurden <u>zuerst</u> in 1997 **eingeführt**$_{anti\text{-}shifter}$.
Full keyboards on cellphones were <u>first</u> **introduced**$_{anti\text{-}shifter}$ in 1997.

(5.11) Diese Gebäude wurden <u>neu</u> **gebaut**$_{anti\text{-}shifter}$.
These buildings have been <u>newly</u> **constructed**$_{anti\text{-}shifter}$.

(5.12) Sie haben <u>extra</u> für mich veganes Essen **zubereitet**$_{anti\text{-}shifter}$.
They <u>specially</u> **prepared**$_{anti\text{-}shifter}$ vegan dishes for me.

5.3.1.2 *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. In chapter 3 we 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. In chapter 3 we used the frame memberships of verbs as a feature, hypothesizing that verbal shifters will be found in the same frames. We now adapt this feature to German, using frames from the German FrameNet project *Salsa* (Burchardt et al., 2006).

EFFEKTGERMANET: +/−Effect theory (Deng et al., 2013; Choi et al., 2014) represents the idea that events can have harmful or beneficial *effects* on their objects (see section 2.3.2). 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.

5.3.2 *New Features*

In section 5.3.1 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.

5.3.2.1 *Bilingual Dictionary*

The motivation behind our work in chapter 3 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 lexica 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 from chapter 3, rather than the manually created one from chapter 4, 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*[2], as it is large (over one million translation pairs) and publicly available. It covers 76 percent

2 https://www.dict.cc

of German verbs found in *GermaNet* and 77 percent of English verbs found in *WordNet*.

The mapping of 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 1,368 of our 2,000 gold standard verbs (68 percent). Less than 6 percent of these are tied between shifter and non-shifter translations. Ties are resolved in favor of the *'shifter'* label. The remaining verbs are labeled with the majority label *'non-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. In section 4.1 we showed 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 allows us to judge its usefulness for high-precision lexicon induction in future work.

5.3.2.2 *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. Similarly to the way in which the SIM feature (see section 5.3.1.1) 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 section 5.3.2.1, 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 percent of the size of the English-German dataset and only a few others have more than 2 percent of its size. In section 5.4.4 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**. In **cross-lingual training** joint embeddings are learned from parallel corpora. However, such corpora are far smaller and rarer than monolingual corpora and, therefore, not ideal for us.[3]

Monolingual mappings take two 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. Training is performed using the same hyperparameters as used by Artetxe et al. (2017). *Word2Vec* uses the *continuous bag of words* algorithm to generate a vector space of 300 dimensions with a context window of 5 words, sub-sampling at $1e-05$ and negative samples at 10. The vocabulary of the embedding is restricted to the 200,000 most frequent words. We also experimented with using the full vocabulary, but this resulted in lower quality embeddings.

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 percent of German *GermaNet* verbs as well as 79 percent of English *WordNet* verbs. It covers 1,598 of our 2,000 gold data verbs (80 percent).

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.

3 *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 Upadhyay et al. (2016) and Artetxe et al. (2017).

CLASSIFIER	FEATURES	SHIFTER LEXICON	TEXT CORPUS	TRAINING DATA
SIM	Data-driven	*n/a*	German	*n/a*
$\text{LP}_{\text{ANY+ANTI}}$	Data-driven	*n/a*	German	*n/a*
CROSS-LING. EMBEDDING	*n/a*	English	German, English	*n/a*
DICTIONARY	Bilingual Dictionary	English	*n/a*	*n/a*
$\text{SVM}_{\text{DATA+RESOURCE}}$	Resource- & data-driven	*n/a*	German	German

Table 5.4: Classifiers used in Table 5.6 and their resource requirements.

5.4 EXPERIMENTS

We evaluate our German bootstrap classifiers in the same manner as we evaluated the English classifiers in section 3.5.3. The task is the classification of all verbs from the given gold standard in a 10-fold cross validation. We report macro-average precision, recall and F-score.

The classifiers are defined in section 5.4.1 and their general performance evaluated in section 5.4.2. In section 5.4.3 we evaluate how the classifiers perform with more limited amounts of training data. In addition to this, we also investigate in section 5.4.4 how the size of the bilingual dictionary affects performance.

5.4.1 *Classifiers*

Analogously to section 3.5.3 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 (section 5.3.2.2) and our dictionary classifier (section 5.3.2.1), 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 5.4.

For the LP classifier we use the list of verbs returned by the ANY feature as seeds for the positive label (*'shifter'*) and the output of the ANTI feature as seeds for the negative label (*'non-shifter'*). For SVM we group features into data-driven and resource-driven feature sets as outlined in sections 5.3.1.1 and 5.3.1.2, as well as introducing the outputs of the cross-lingual word embedding and dictionary classifiers as additional separate features. An overview of the resulting groups is given in Table 5.5.

GROUP	FEATURES
DATA	$LP_{ANY+ANTI}$, SIM, CLASH, PRT
RESOURCE	GERMANET, SALSA, EFFEKTGERMANET
EMBED	CROSS-LINGUAL EMBEDDING
DICT	DICTIONARY

Table 5.5: Lists of the features included in the SVM feature groups in Table 5.6. All features in DATA and RESOURCE are German equivalents of the English language features that were used in chapter 3.

CLASSIFIER	PREC	REC	F1
$\text{BASELINE}_{\text{MAJ}}$	44.4	50.0	47.0
SIM	58.0	67.6	62.4
$LP_{ANY+ANTI}$	67.2	65.0	66.1
CROSS-LINGUAL EMBEDDING	67.6	74.6	$70.9^{*\dagger}$
DICTIONARY	69.2	77.3	$73.0^{*\dagger}$
SVM_{DATA}	60.8	72.6	66.2
$SVM_{RESOURCE}$	79.4	73.9	$76.4^{*\dagger}$
$SVM_{DATA+RESOURCE}$	79.0	76.7	$77.7^{*\circ\dagger}$
$SVM_{DATA+RESOURCE+EMBED}$	79.6	78.9	$79.2^{*\circ\dagger}$
$SVM_{DATA+RESOURCE+DICT}$	78.0	80.9	$79.4^{*\circ\dagger}$
$SVM_{DATA+RESOURCE+DICT+EMBED}$	**80.3**	**82.0**	$\mathbf{81.0}^{*\circ\dagger\ddagger}$

*: F1 is better than that of $LP_{ANY+ANTI}$ (paired t-test with $p < 0.05$).
$^\circ$: F1 is better than that of DICTIONARY (paired t-test with $p < 0.05$).
†: F1 is better than that of SVM_{DATA} (paired t-test with $p < 0.05$).
‡: F1 is better than that of $SVM_{DATA+RESOURCE}$ (paired t-test with $p < 0.05$).

Table 5.6: Evaluation of classifiers (section 5.4.1) on the 2,000 verbs from the gold standard (Table 5.1). The evaluation is run as a 10-fold cross validation and all reported metrics are macro-averages. Best results are depicted in bold.

5.4.2 *Classifier Evaluation*

Table 5.6 shows the performance of our various classifiers. All classifiers clearly outperform the baseline and resource-based features outperform data-driven ones. This is similar to the performance observed for English (see chapter 3). Cross-lingual embeddings and dictionaries as stand-alone classifiers both outperform the label propagation approach due to their better recall coverage of shifters. Interestingly,

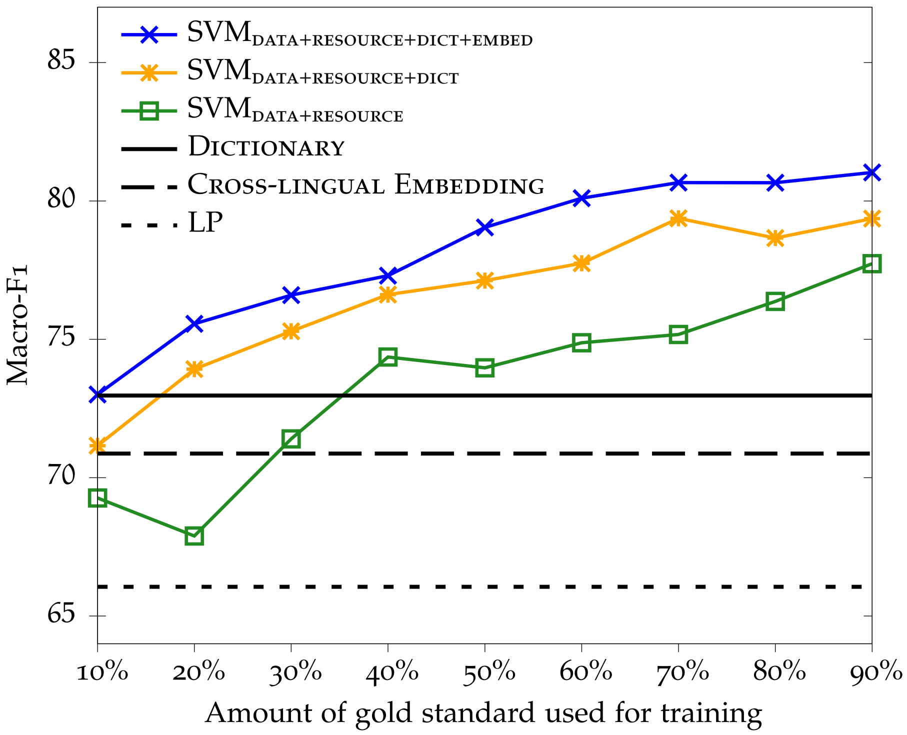

Figure 5.2: Learning curve for supervised training. This repeats the evaluation of section 5.4.2, but reduces the amount of training data. $\text{SVM}_{\text{DATA+RESOURCE}}$ represents the configuration of the best classifier for English verbal shifters from section 3.5.3. At 90 percent training data this task is identical to the one reported in Table 5.6.

the cross-lingual embedding classifier performs far better than SIM, despite both relying on distributional similarity. Comparing similarity among verbs, even cross-lingually, works better than across parts of speech, which is 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, $\text{SVM}_{\text{DATA+RESOURCE+DICT+EMBED}}$, provides a significant improvement over $\text{SVM}_{\text{DATA+RESOURCE}}$, the best classifier of chapter 3. 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.

5.4.3 *Training Size Requirements*

One of our aims in this chapter is to establish whether our bootstrapping approach is still viable for under-resourced languages. In such languages, acquiring funding for the creation of larger amounts of labeled training data will often be challenging. Accordingly we investigate the performance of our classifiers with smaller amounts.

Figure 5.2 shows the learning curve of select SVM configurations, compared to the classifiers that work without labeled German data, i. e. LP, Cross-lingual Embedding and Dictionary. Cross-lingual Embedding and Dictionary classifiers provide a stronger baseline than LP, outperforming $\text{SVM}_{\text{DATA+RESOURCE}}$ 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 percent. Combining all available sources of information as SVM features is therefore the preferred approach if any amount of training data is available.

5.4.4 *Dictionary Size Requirements*

The dictionary mapping approach (section 5.3.2.1) has been shown to be a strong stand-alone classifier and SVM feature (Table 5.6), slightly outperforming the cross-lingual word embedding approach. 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 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 8,000 of the English verbs found in *WordNet*. Of the 2,000 German verbs in our gold standard, *DictCC* covers 1,368. 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 4,000 most frequent verbs from *WordNet* ($\text{Dictionary}_{\text{VOC\_SIZE=4K}}$). We also create even smaller versions with only the 1,000 ($\text{Dictionary}_{\text{VOC\_SIZE=1K}}$) and 500 most frequent English verbs ($\text{Dictionary}_{\text{VOC\_SIZE=0.5K}}$).

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 1,168 of German words in our gold data, i. e. 85 percent of the full dictionary of 1,368 words. Reducing it further to 1,000 English verbs drops the size of the German vocabulary to 52 percent. Using only the 500 most frequent English words leaves a German coverage of 33 percent.

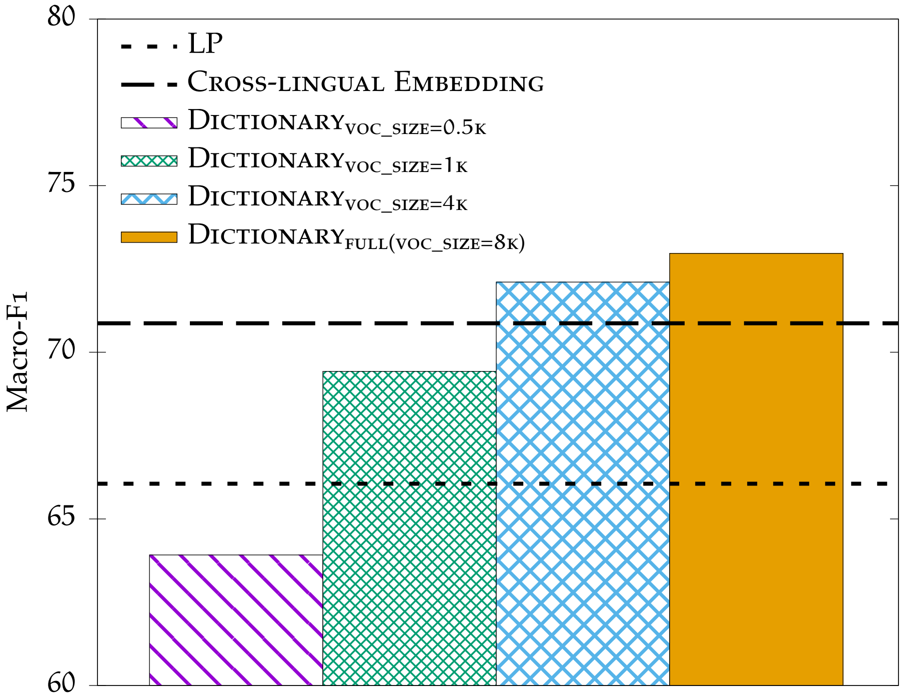

Figure 5.3: Comparison of dictionary-based classifiers using dictionaries with different vocabulary sizes. Classifiers use *no labeled training data*. Dictionary$_{\text{full(voc_size=8k)}}$ is equivalent to the dictionary shown in Table 5.6.

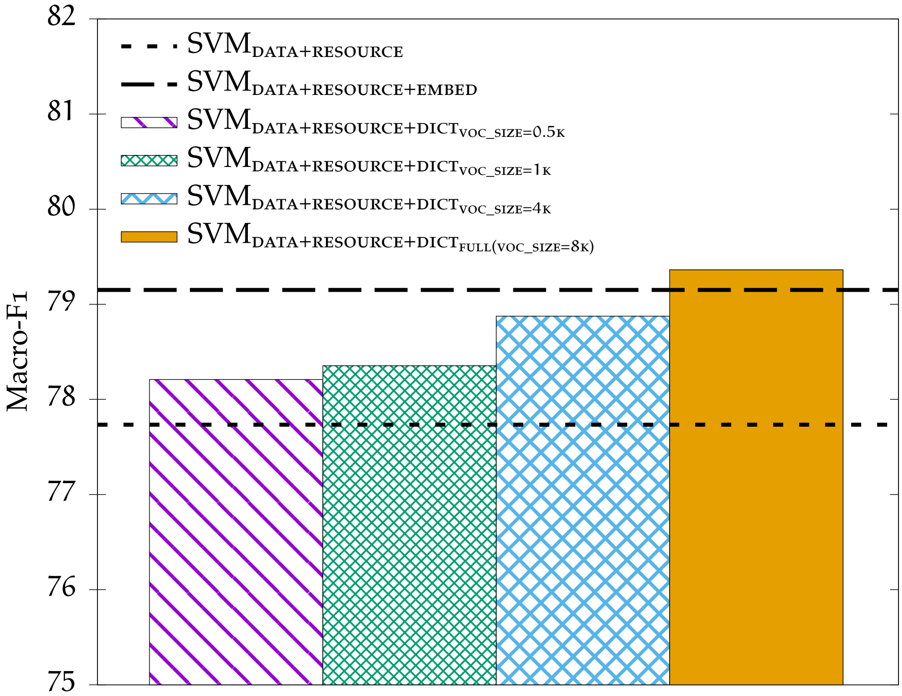

Figure 5.4: Comparison of SVM classifiers using dictionaries with different vocabulary sizes. SVM$_{\text{data+resource+dict}_{\text{full(voc_size=8k)}}}$ is equivalent to the classifier SVM$_{\text{data+resource+dict}}$ in Table 5.6.

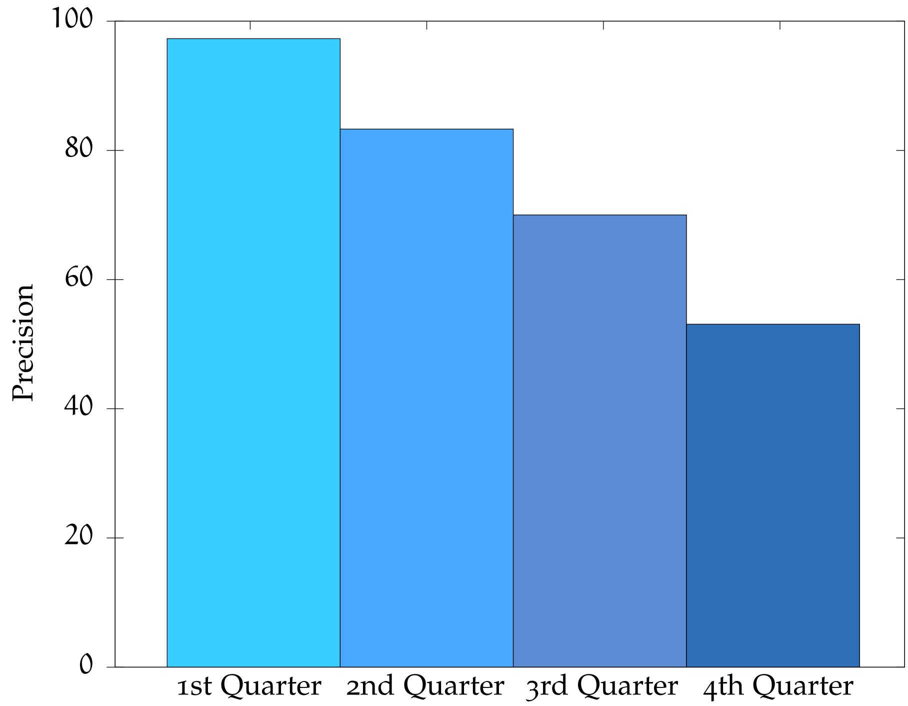

Figure 5.5: Evaluation of the bootstrapping of German verbal shifters that were not part of the gold standard (compare Table 5.1). SVM classifiers provide a confidence value for each label they assign. We rank the 595 potential shifters from highest to lowest confidence and group them, so that the first quarter contains the highest and the fourth quarter the lowest confidence candidates.

Figure 5.3 shows the performance of the differently sized dictionaries as stand-alone classifiers, while Figure 5.4 shows how much they can improve the best classifier of chapter 3, 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.

5.5 BOOTSTRAPPING THE LEXICON

To increase the size of our lexicon, we bootstrap additional shifters. For this we use the same approach as we used in section 3.6. We train our best classifier (Table 5.6) on the 2,000 verbs from our gold standard (section 5.1) and then use it to classify the remaining 7,262 *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 percent.

Figure 5.5 shows the classifier precision at different confidence intervals. Just as in section 3.6, we see very high performance for the first quarter, matching our observation in chapter 3 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)*.

5.6 conclusion

We confirmed that the bootstrapping process that we initially introduced for English verbal shifters in chapter 3 can successfully be applied to German as well. 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 2,000 verbs taken from *GermaNet*, we trained a supervised classifier with various data- and resource-driven features.

The performance of the bootstrapping process was further improved by leveraging information from the existing English lexicon of verbal shifters from chapter 3. To transfer knowledge from the English lexicon we used bilingual dictionaries and cross-lingual word embeddings. The resulting improved classifier allowed us to triple the number of confirmed German shifters in our lexicon, compared to their number in the gold standard.

To take into account the fact that different languages have different kinds and amounts of resources available, we differentiated features by whether they require only unlabeled data and basic linguistic tools or whether they depend on semantic resources that may not be available for many languages. In addition, we introduced 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 provided the best results.

Our recommendation for creating shifter lexica in new languages is to start out with cross-lingual label transfer, but also to 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 adapting the bootstrap approach from chapter 3, we limited ourselves to verbs. In part III we will leave this limitation behind and also address nominal and adjectival shifters. We take the first step towards this in chapter 6, by extending the bootstrap approach to be able to handle these new parts of speech.

Part III

EXTENSION AND APPLICATION

6 EXTENDING THE LEXICON BY INTRODUCING NOMINAL AND ADJECTIVAL SHIFTERS

In part II we covered the creation of several lexica of verbal polarity shifters. This was motivated by the prominent role of verbs as the main syntactic predicate of most clauses and sentences (see section 3.1).

Now that verbal shifters have been covered sufficiently, it is time to move on to other parts of speech. Nouns and adjectives[1] can function as polarity shifters just as well as verbs can, as we see in (6.1)–(6.3)

(6.1) Peter [**failed**$_{shifter}$ to [pass the exam]$^{+}$]$^{-}$.

(6.2) Peter's [**failure**$_{shifter}$ to [pass the exam]$^{+}$]$^{-}$...

(6.3) Peter's [**failed**$_{shifter}$ [attempt to pass the exam]$^{+}$]$^{-}$...

When creating lexica for nominal shifters and adjectival shifters the challenge of covering a large vocabulary efficiently becomes even more pressing. While the verbal vocabulary of *WordNet* covers *just* 10,000 words, it contains more than 110,000 nouns and 20,000 adjectives. Manually annotating the entirety of these would be prohibitively expensive. Instead we extend the bootstrapping approach that we introduced in chapter 3 for use with nouns and adjectives.

Apart from adapting existing features for the new parts of speech, we also introduce features that map information from the bootstrapped verbal shifter lexicon to nouns and adjectives, similar to our approach from chapter 5 that mapped English shifter labels to German verbs.

By combining the bootstrapped lexicon of verbal shifters with those for nominal and adjectival shifters, we are able to create a **general lexicon of polarity shifters**. The combined bootstrapping workflow for all parts of speech can be seen in Figure 6.1.

To allow for an easy comparison of our methodology for verbs with that for nouns and adjectives, this chapter provides information on all three parts of speech.

Contents

Section 6.1 introduces the gold standards for nouns and adjectives and section 6.2 presents the other resources that we require for the creation of our features. Section 6.3 describes how we define (for the computational purposes of our features) the scope that a polarity shifter can affect. Section 6.4 then describes the features that we use

1 We omit adverbial shifters at this point, as there is a strong overlap with the typical negation words (such as *not*) and downtoners (such as *barely*) that are already covered in other resources.

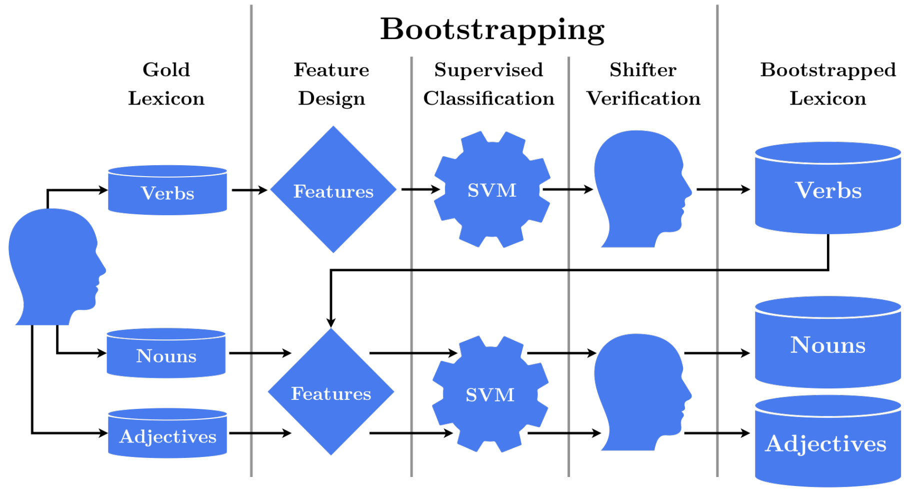

Figure 6.1: Workflow for creating a general lexicon of polarity shifters. In a first phase the lexicon of verbs is created (chapter 3), while in the second phase it is used to help in the creation of the noun and adjective lexica (this chapter). Human annotation is used to create the initial gold lexica and to verify the bootstrapped shifters.

in our bootstrap classifier. These features are used in classifiers that are presented and evaluated in section 6.5. The best of these classifiers are used to bootstrap shifter lexica for their respective parts of speech in section 6.6. In section 6.7 these POS-specific lexica are merged to create a single general lexicon of polarity shifters. Our findings are summarized in section 6.8

Contributions

(i) We create a general lexicon of polarity shifters that covers verbs, nouns and adjectives by extending the previously bootstrapped lexicon of verbal shifters.

(ii) We extend the bootstrapping approach from chapter 3 to also work for nouns and adjectives.

(iii) We introduce features that leverage similarities between shifters across different parts of speech.

Publication History

Work presented in this chapter is also contained in Schulder et al. (under review). Some text passages were also adapted from Schulder et al. (2017). The general lexicon of English polarity shifters is publicly available.[2]

2 https://doi.org/10.5281/zenodo.3365601

LABEL	VERBS FREQ.	VERBS PERC.	NOUNS FREQ.	NOUNS PERC.	ADJECTIVES FREQ.	ADJECTIVES PERC.
Shifter	304	15.20%	107	5.35%	129	6.45%
Non-shifter	1,696	84.80%	1,893	94.65%	1,871	93.55%
Total	2,000		2,000		2,000	

Table 6.1: Distribution of polarity shifters in the gold standard. For each part of speech, a random sample of 2,000 words was taken from *WordNet*.

6.1 GOLD STANDARD

We extend the polarity shifter gold standard that we introduced in chapter 3 to also contain nouns and adjectives, following the annotation procedure established in section 3.2. A random sample of 2,000 nouns and 2,000 adjectives was taken from the vocabulary (see section 2.2.1) and annotated by the same expert annotator. Each word lemma is labeled as either *'shifter'* or *'non-shifter'*. This gives us a general gold standard of 6,000 words (2,000 per part of speech) for bootstrapping polarity shifters. Overall the gold standard annotation cost about 170 hours of annotation work.

To measure inter-annotator agreement for the annotation task, 10 percent of the gold standard words for each part of speech were annotated a second time by another expert annotator. The Cohen's kappa (Cohen, 1960) agreement was $\kappa = 0.66$ for verbs, $\kappa = 0.77$ for nouns and $\kappa = 0.71$ for adjectives. All of these scores indicate substantial agreement (Landis and Koch, 1977). The higher agreement for nouns and adjectives compared to verbs is due to the fact that their annotation was performed after the one for verbs had been completed, by which time both annotators had gained more experience in the annotation of polarity shifters.

Table 6.1 shows the distribution of shifters in our gold standard. We can see that shifters are far more frequent among verbs than among nouns (almost three times as many) and adjectives (more than twice as many). Unsurprisingly, the majority of words are non-shifters. However, extrapolating from the shifter frequencies in Table 6.1, we can still expect to find several thousand shifters in our vocabulary.

6.2 RESOURCES

In this section we briefly describe the resources required by the features presented in section 6.4. The resources are the same as were previously used for bootstrapping verbal shifters in chapter 3. For more detailed descriptions of the resources, see section 3.3. This overview only

	VERBS		NOUNS		ADJECTIVES	
LABEL	FREQ.	PERC.	FREQ.	PERC.	FREQ.	PERC.
Positive Words						
Shifter	4	5.5%	1	1.9%	5	2.3%
Non-shifter	69	94.5%	52	98.1%	216	97.7%
Negative Words						
Shifter	49	25.9%	30	24.4%	68	20.4%
Non-shifter	140	74.1%	93	75.6%	266	79.6%

Table 6.2: Distribution of sentiment polarities among polarity shifters from the gold standard according to the *Subjectivity Lexicon* (Wilson et al., 2005b).

describes resources used for multiple features. Resources required by only a single feature are introduced in definition of that feature.

As text corpus for determining word and pattern frequencies, we choose the *Amazon Product Review Data* (Jindal and Liu, 2008), which consists of 5.8 million product reviews. We re-use the word embedding that was originally used for bootstrapping verbal shifters. It was created using *Word2Vec* (Mikolov et al., 2013). Syntactic structures, such as dependency relations, are determined with the help of the *Stanford Parser* (Chen and Manning, 2014).

To determine the polarity of a word, we use the *Subjectivity Lexicon* (Wilson et al., 2005b). Table 6.2 shows the distribution of polarities among words in our gold standard, separated by part of speech. We see that a fair number of polarity shifters are themselves polar words. Most of these are of negative polarity. Among words of positive polarity there are very few shifters.

6.3 SHIFTING SCOPE

Before we describe the features that will be used in our classification task, we must define how we handle shifting scope, i. e. which part of a sentence is affected by the polarity shifter.

Usually, a shifter can be expected to only affect expressions that are syntactically governed by the valency of the shifter word. However, the shifting scope does not necessarily include every syntactic argument of a shifter word. For example, (6.4) shows how the verbal shifter *to defeat* affects its direct object, while the polarity of its subject remains unaffected. The polarity of expressions outside the scope of the shifter do not affect the shifting process, as we can see when considering (6.4) and (6.5).

(6.4) [The villain]$^{-}_{subj}$ [**defeated**$_{shifter}$ [the hero]$^{+}_{dobj}$]$^{-}$.

(6.5) Chance$_{subj}$ [**defeated**$_{shifter}$ [the hero]$^{+}_{dobj}$]$^{-}$.

To determine the scope of a shifter, we rely on its dependency relations, such as the direct object relation between *defeat* and *the hero*. The part of speech of a shifter determines which dependency relation indicates its scope:

1. Shifter is a verb:
   a) `dobj`: If the shifter is a verb, then its direct object is the scope.

      dobj

      Example: The storm [**ruined**$_{shifter}$ [their party]$^{+}$]$^{-}$.
2. Shifter is a noun:
   a) `nn`: If the shifter is the head of a noun compound, then the compound modifier is the scope.

      nn

      Example: It is a [[cancer]$^{-}$ **cure**$_{shifter}$]$^{+}$.

   b) `prep_of`: If the shifter is a noun that is the head of the preposition **of**, then the object of that preposition is the scope.

      prep pobj

      Example: The [**destruction**$_{shifter}$ of [my dreams]$^{+}$]$^{-}$.
3. Shifter is an adjective:
   a) `amod`: If the shifter is an attributive adjective, then the modified noun is the scope.

      amod

      Example: The [**exonerated**$_{shifter}$ [convict]$^{-}$]$^{+}$ walked free .

   b) `nsubj`: If the shifter is a predicative adjective, then its subject is the scope.

      nsubj

      Example: The [[hero]$^{+}$ is **dead**$_{shifter}$]$^{-}$.

Note that this definition of shifting scopes is a simplified representation designed to fit the needs of our data-driven features. For more detailed discussions that also address less frequent kinds of scopes we refer the reader to our discussion of verbal shifting scopes in section 4.2 and to Wiegand et al. (2018a).

6.4 FEATURE DESIGN

This section introduces the features which we will use in our classifier evaluation in section 6.5 and when bootstrapping our lexicon in section 6.6.

We introduce features that were specifically designed for the classification of polarity shifters in section 6.4.1. Generic features for supervised classification in sentiment analysis tasks are presented in section 6.4.2. In section 6.4.3 we discuss means of applying shifter information across different parts of speech. Finally, section 6.4.4 briefly lists previously used features that are exclusive to verbs and cannot be applied to nouns or adjectives.

6.4.1 *Task-Specific Features*

We begin our feature discussion with features that are specifically designed to identify polarity shifters. Each of these features creates a ranked list, indicating how likely each word is to be a shifter. For each part of speech separate ranked lists are created.

DISTRIBUTIONAL SIMILARITY (SIM): We leverage the close relation between polarity shifters and negation words (see section 2.1) by extracting words that are distributionally similar to negation words. For this we use the word embedding representation of our text corpus that we computed with the help of *Word2Vec* (Mikolov et al., 2013) (see section 3.3.3). All words are ranked by their cosine similarity to a given negation word. The highest ranking words are considered more likely to be polarity shifters.

As negation words we consider the intersection of two negation word lists: the *negation* category in the negation lexicon by Wilson et al. (2005b) and the negation signals from Morante and Daelemans (2009). The negation words are *neither*, *never*, *no*, *none*, *nor*, *not* and *without*. As we found in section 3.5.1 that relying on the similarity to a specific negation word resulted in unpredictable performance, we compute the centroid of the aforementioned negation words.

POLARITY CLASH (CLASH): We found that when polarity shifters are polar words themselves, they are often used to shift words of the opposing polarity. Examples for shifters of different parts of speech can be seen in (6.6)–(6.8).

(6.6) That could [**harm**$^{-}_{shifter}$ your [defense]$^{+}$]$^{-}$.

(6.7) This is [**obstruction**$^{-}_{shifter}$ of [justice]$^{+}$]$^{-}$.

(6.8) I'm sorry about the [**ruined**$^{-}_{shifter}$ [celebration]$^{+}$]$^{-}$.

The more often a polar word co-occurs with expressions of the opposite polarity in a shifter scope dependency relation (see section 6.3), the more likely is it that the word is a shifter.

As we saw in Table 6.2, the *Subjectivity Lexicon* contains very few polarity shifters of positive polarity, so we limit our search to negative shifter candidates that occur with positive polar expressions.

HEURISTIC USING 'ANY' (ANY): Our final shifter feature relies on the co-occurence of negation words with negative polarity items (NPIs), such as the word *any*. The theoretical basis for this was discussed in section 2.3.1 and it was successfully implemented for verbal shifters in chapter 3.

Our feature collects all occurrences of potential shifters in which the NPI *any* is a determiner within the scope of the shifter. Shifting scopes are determined as defined in section 6.3. An example can be seen in (6.9). The scope of the nominal shifter *lack* is its prepositional object *"support"* and the determiner of *"support"* is *any*.

prep pobj det

(6.9) There was a clear [**lack**$_{shifter}$ of any support^{+}]$^{-}$.

As an additional constraint we require that the head word of the scope must be a polar expression. In (6.9) this requirement is met, as the noun *support* is of positive polarity.

6.4.2 *Generic Features*

The task-specific features introduced in section 6.4.1 were designed to produce high-precision lists of shifters. Another source of features are general purpose semantic resources.

WORDNET (WN): *WordNet* is the largest available English ontology. It provides various kinds of semantic information for individual word senses and their relation to one another. As in chapter 3 we make use of three kinds of information provided by *WordNet*:

- **Glosses** are texts that clarify the meaning of a word sense.
- **Supersenses** are coarse semantic categories, such as *'Act'* or *'Emotion'*.
- **Hypernyms** connect a word sense with a more general form of the same concept (e. g. *vehicle* is the hypernym of *car*).

A more detailed description of these information sources and motivations for their use can be found in section 3.4.2.

FRAMENET (FN): *FrameNet* (Baker et al., 1998) is a semantic resource used for various sentiment related tasks, such as opinion holder and target extraction (Kim and Hovy, 2006), stance classification (Hasan and Ng, 2013) and opinion spam analysis (Kim et al., 2015). It provides over 1,200 semantic frames that comprise words with similar semantic behavior. We use the **frame memberships** of a word as its features, expecting that polarity shifters are grouped in the same frames.

6.4.3 *Cross-POS Features*

Following the workflow outlined in Figure 6.1, the verb component of our shifter lexicon is created first, followed by the other parts of speech. This means when we bootstrap nouns and adjectives, the verb lexicon is available to us as a resource.

Our hypothesis is that nominal and adjectival forms of a verbal shifter will equally be shifters, as can be seen in (6.10)–(6.12).

(6.10) Smoking [$\textbf{damages}^{V}_{\textit{shifter}}$ his [health]$^{+}$]$^{-}$.

(6.11) Beware the [[health]$^{+}$ $\textbf{damage}^{N}_{\textit{shifter}}$]$^{-}$ caused by smoking.

(6.12) Constant chain smoking is the reason for his [$\textbf{damaged}^{A}_{\textit{shifter}}$ [health]$^{+}$]$^{-}$.

Using our bootstrapped lexicon of verbal shifters we can assign shifter labels to related nouns and adjectives. To determine which words should be considered related, we apply the following two approaches:

DERIVATIONAL RELATEDNESS ($\text{VERBLEX}_{\text{DERIV}}$): The *derivational relatedness* relation by *WordNet* provides connections to both nouns and adjectives. For nouns we also use the *Nominalization Lexicon* (NOMLEX) (Macleod et al., 1998) in cases where *WordNet* does not provide derivational relatedness information.

STEM RELATEDNESS ($\text{VERBLEX}_{\text{STEM}}$): As a lightweight alternative to $\text{VerbLex}_{\text{DERIV}}$ without dependence on any lexicon resource we use word stems to approximate relatedness. In word stemming, the inflectional suffix of a word is removed, leaving only the stem, which is not specific to a particular part of speech. Taking, for example, the shifters in (6.10)–(6.12), the stem of the verb *damages*, the noun *damage* and the adjective *damaged* is always *damag*.

We classify each noun and adjective that shares a stem with a known verbal shifter as a shifter. To determine word stems we use the *Porter Stemmer* (Porter, 1980) as implemented in the *Natural Language Toolkit* (NLTK) (Bird et al., 2009).

6.4.4 *Features limited to verbs*

A few features are only available for use with verbs, either due to their nature or due to the availability of resources. They are excluded from the classifiers for nominal and adjectives shifters.

Detailed descriptions of these features can be found in section 3.4.1. As a brief reminder of the affected features and to explain why they cannot be used with other parts of speech, we list them here:

EFFECTWORDNET (EFFECT): *EffectWordNet* covers only verbs, so our feature that relies on +/−effect is limited in the same way.

VERB PARTICLES (PRT): As the name suggests, verb particles only occur with verbs.

ANTI-SHIFTERS (ANTI): The patterns used for determining verbal anti-shifters are only suitable for capturing verbs. We were unable to determine anti-shifter patterns for nouns and adjectives that performed sufficiently well.

6.5 EXPERIMENTS

In this section we will evaluate the performance of the features introduced in section 6.4. In section 6.5.1 we introduce a number of classifiers, which are then evaluated in section 6.5.2 As some of the noun and adjective classifiers rely directly on the output of the verb classifier (see Figure 6.1), we evaluate the three parts of speech separately. Following this we investigate in section 6.5.3 how much training data is actually required to generate high quality polarity shifter classifications. All these evaluations are in preparation for bootstrapping a complete polarity shifter lexicon in section 6.6.

6.5.1 *Classifiers*

We consider a number of classifiers that work *with* and *without labeled training data*. All classifiers are compared against a majority-class baseline that labels all words as non-shifters.

LABEL PROPAGATION (LP): The label propagation classifier was introduced in chapter 3 to provide a classifier that uses no training data. It used the output of the *"any"* heuristic and the anti-shifter feature as seeds with which to propagate shifter and non-shifter labels across a word-similarity graph based on our word embedding.

We choose to not use the label propagation classifier for nouns or adjectives, as the seeds for are considerably weaker for these parts of speech. The *"any"* heuristic provides fewer than 200 nouns and adjectives, as opposed to over 500 for verbs. The anti-shifter feature is not available for nouns and adjectives.

MAPPING FROM VERB LEXICON (VERBLEX): To make up for the lack of a label propagation classifier for nouns and adjectives we introduce the cross-POS feature as a stand-alone classifier. While it of course relies heavily on information about verbal shifters, it does not require any additional labeled training data for nouns or adjectives.

We provide two versions of this classifier. The first relies on linguistic resources and uses the derivational-relatedness relation of *WordNet* as its main source of information and in case of no relation being provided falls back on NOMLEX nominalizations ($\text{VerbLex}_{\text{DERIV}}$). The second version uses similarity of word stems to create its mapping ($\text{VerbLex}_{\text{STEM}}$). This version requires no resources other than the verbal shifter lexicon.

SUPPORT VECTOR MACHINES (SVM): As a supervised classifier we choose **support vector machine** as implemented in *SVM*$^{\text{light}}$ (Joachims, 1999). Supervised classification relies on the availability of manually labeled training data. On the other hand, using support vector machines allows us to combine arbitrary features. This is in contrast to label propagation, for which all information had to be encoded in the choice of seeds and graph weights.

We evaluate a number of different feature combinations. We train one classifier using the task-specific features introduced in section 6.4.1 (SVM_T). In the case of the verbal shifter classifier, SVM_T also contains the verb-exclusive task-specific features from section 6.4.4. A second classifier uses the generic features from section 6.4.2 (SVM_G). A third classifier combines both sets of features ($\text{SVM}_{\text{T+G}}$). For nouns and adjectives we also use a fourth classifier that adds the output of the best-performing cross-POS feature from section 6.4.3 to the combined feature set ($\text{SVM}_{\text{T+G+V}}$).

6.5.2 *Classifier Evaluation*

Table 6.3 shows the performance of the classifiers on our polarity shifter gold standard (section 6.1). Training and testing is performed using 10-fold cross validation on the 2,000 gold standard words of the respective part of speech. Results are presented as macro-averaged precision, recall and F-score.

CLASSIFICATION OF VERBS: Both label propagation (LP) and support vector machines (SVM) clearly outperform our baseline. Furthermore, all feature combinations of SVM outperform label propagation. This indicates that using labeled training data is beneficial and that using a combination of task-specific features (section 6.4.1) is better

CLASSIFIER	VERBS			NOUNS			ADJECTIVES		
	PREC	REC	F1	PREC	REC	F1	PREC	REC	F1
$\text{BASELINE}_{\text{MAJ}}$	42.4	50.0	45.9	47.3	50.0	48.6	46.8	50.0	48.3
LP	68.6	56.7	62.0*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*	*n/a*
$\text{VERBLEX}_{\text{STEM}}$	*n/a*	*n/a*	*n/a*	76.9	65.3	70.5*	66.0	56.9	61.0*
$\text{VERBLEX}_{\text{DERIV}}$	*n/a*	*n/a*	*n/a*	82.6	**74.1**	78.1*	46.8	49.9	48.3
SVM_{T}	65.5	69.7	67.5*	62.1	61.9	61.9	59.4	**66.9**	62.9*
SVM_{G}	79.6	74.4	76.9*	70.1	56.6	62.4	**74.4**	60.5	66.3†
$\text{SVM}_{\text{T+G}}$	**80.7**	**77.6**	**79.1***	70.4	57.6	63.1	72.8	62.1	**66.6†**
$\text{SVM}_{\text{T+G+V}}$	*n/a*	*n/a*	*n/a*	**84.6**	73.8	**78.7***	70.0	62.9	66.1†

*: F1 is better than previous classifier (paired t-test with $p < 0.05$).
†: F1 is better than best VerbLex classifier of same part of speech (paired t-test with $p < 0.05$).

Table 6.3: Classification of polarity shifters for individual parts of speech. SVM features are grouped as task-specific (T), generic (G) and VerbLex (V) (see sections 6.4.1, 6.4.2 and 6.4.3, respectively). The evaluation is run as a 10-fold cross validation and all reported metrics are macro-averages. Best results are depicted in bold.

than using only the strongest feature (which was used to determine the shifter seeds for label propagation).

While the generic features (SVM_G) outperform the task-specific features (SVM_T), combining both feature sets provides another significant performance boost (SVM_{T+G}). This shows that the feature sets are of a complementary nature.

CLASSIFICATION OF NOUNS: Comparing the performance of the SVM classifiers between verbs and nouns in Table 6.3, we see that the noun classifier is falling behind a little. This is mainly due to the considerably lower performance of the generic features (SVM_G), which is most likely caused by the lower frequency of shifters among nouns. While 15 percent of verbs in our gold standard are shifters, only 5 percent of nouns are. As the generic features (see section 6.4.2) used in SVM_G heavily rely on sufficient amounts of training data for each class label, available instances of nominal shifters might simply not be enough.

Looking at the VerbLex classifier, transferring the labels of verbal shifters to nouns proves to be a strong feature. In fact, $\text{VerbLex}_{\text{deriv}}$ provides performance similar to that of the best verb classifier. This proves that in most cases, the nominal forms of verbal shifters are also shifters.

Unfortunately combining $\text{VerbLex}_{\text{deriv}}$ with SVM_{T+G} results in no significant performance gains (SVM_{T+G+V}). We will present a possible reason for this in section 6.5.3.

CLASSIFICATION OF ADJECTIVES: For adjectives the performance of task-specific and generic features is similar to that for nouns. Unlike for nouns, the VerbLex classifier does not perform particularly well. $\text{VerbLex}_{\text{deriv}}$ fails to produce results above those of the majority baseline due to massive sparsity issues, as *WordNet* and NOMLEX together provide mappings to only five of the considered adjectives. $\text{VerbLex}_{\text{stem}}$ shows that the idea of relatedness of polarity shifters across part of speech is still generally valid. However, it does not perform as well as SVM_{T+G}. Instead, the SVM classifier provides the best performance, as was the case for verbs.

6.5.3 *Training Size Requirements*

To get a sense of how much training data is required to achieve acceptable classifier performance for individual parts of speech, we extend the learning curve experiment from section 3.5.4, applying it now to nouns and adjectives as well as verbs.

Figure 6.2 shows a learning curve of how the classifiers from Table 6.3 perform with different amounts of training data. The curve ranges from training on 200 to 1,800 words, i. e. 10 to 90 percent of our

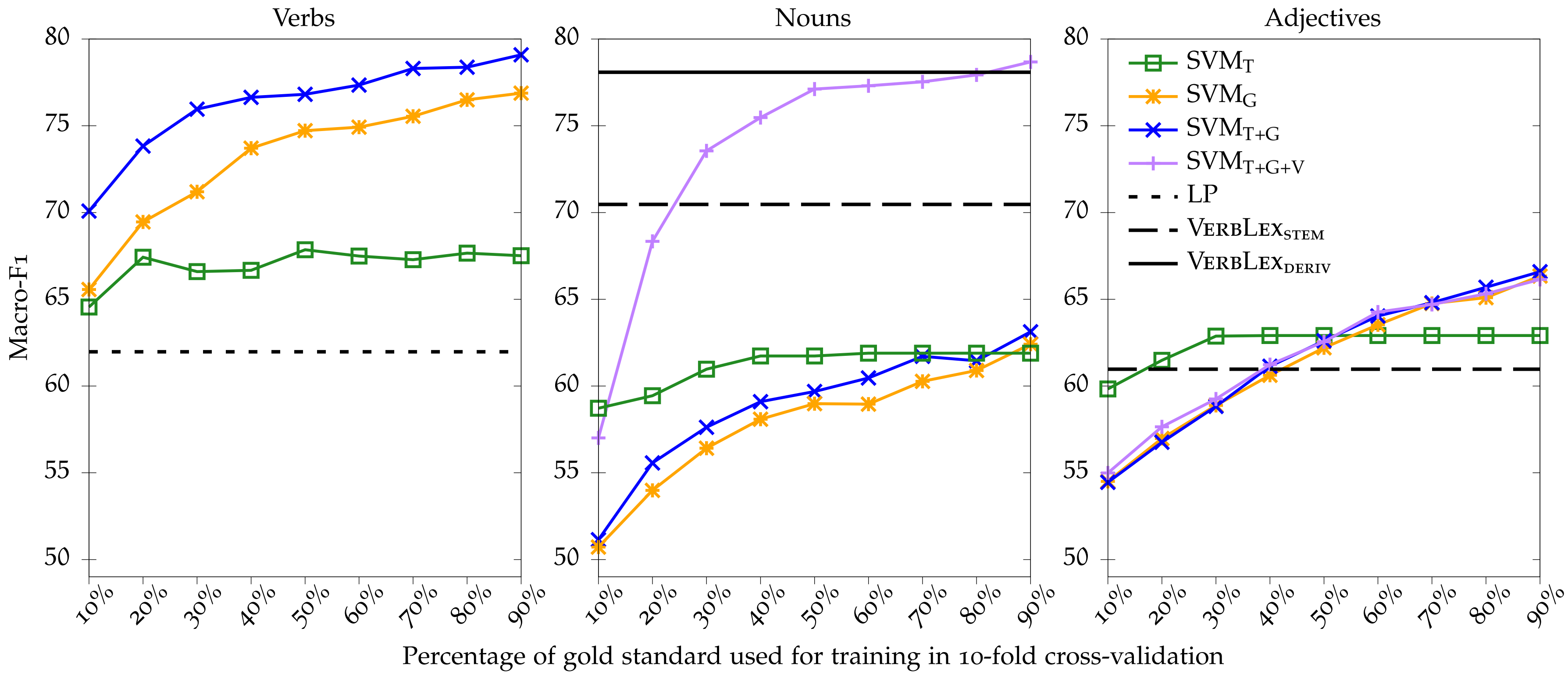

Figure 6.2: Learning curves for supervised training. This repeats the evaluation of section 6.5.2, but reduces the amount of training data. At 90 percent training data this task is identical to the one reported in Table 6.3.

gold standard for a specific part of speech. Each score is the average of a 10-fold cross validation. Accordingly, the results on 90 percent training data are identical to those reported in section 6.5.2. Label propagation (LP) and mapping from the verb lexicon (VerbLex) require no labelled training data. Accordingly, their performance is the same regardless of training size.

We can see that the task-oriented features (SVM_T) reach their maximum performance early on, meaning that they can provide good performance with little training data, but also that adding more data will not improve performance. Generic features (SVM_G) behave in the opposite way, clearly profiting from larger amounts of training data. Especially for nouns and adjectives the learning curves of SVM_G suggest that the classifier would profit from even larger amounts of training data than we have available at this point. This matches the conclusion we drew during the discussion of noun classification in section 6.5.2 that the usefulness of generic features is hampered by the low frequency of nominal and adjectival shifters in the training data.

Comparing the different parts of speech and their performance relative to SVM_T, we observe clear differences. For verbs, SVM_G clearly outperforms the task-specific features even with small amounts of training data, while for nouns it requires the full training set to perform even equally well. For adjectives, SVM_G surpasses SVM_T at training amounts above 50 percent.

Combining generic and task-specific features ($\text{SVM}_\text{T+G}$) provides clear improvements for verbs at every stage of the learning curve. In the case of nouns and adjectives, $\text{SVM}_\text{T+G}$ appears to mostly ignore the task-specific features, performing similar to SVM_T. For nouns we see that the supposed lack of difference between $\text{VerbLex}_\text{deriv}$ and $\text{SVM}_\text{T+G+V}$ that we observed in Table 6.3 might be due to the chosen amount of training data. Judging by the learning curve, we expect that adding further training data would show $\text{SVM}_\text{T+G+V}$ soon outperforming $\text{VerbLex}_\text{deriv}$. For adjectives, on the other hand, the inclusion of $\text{VerbLex}_\text{stem}$ in $\text{SVM}_\text{T+G+V}$ brings no advantage. Nevertheless, $\text{VerbLex}_\text{stem}$ still presents a viable option for cases where supervised classification is not an option.

6.6 BOOTSTRAPPING THE LEXICON

In section 6.5 we trained classifiers on our verb, noun and adjective gold standards of 2,000 word each. In this section we use those classifiers to bootstrap a lexicon from the remaining unannotated 8,581 verbs, 53,311 nouns and 16,282 adjectives. All words that the classifiers *predict* to be shifters are verified by a human annotator to filter out false positive classifications. All words predicted to be non-shifters will not be considered further. This classifier-based pre-filtering approach

(Choi and Wiebe, 2014) allows us to ensure the high quality of the lexicon while keeping the annotation workload manageable.

In section 6.6.1 we motivate which classifiers for each part of speech we will include in our bootstrapping evaluations (sections 6.6.2 and 6.6.3). Section 6.6.2 presents a quantitative evaluation in which we judge the classifiers by how many shifter candidates they find and by how few erroneous classifications this introduces. Section 6.6.3 is a qualitative evaluation in which we inspect whether high classifier confidence equals high quality of classification.

The verified bootstrapped shifters of the various parts of speech will then be combined in section 6.7 with the shifters from the gold standard to create a single large polarity shifter lexicon.

6.6.1 *Choosing Classifiers for Bootstrapping*

For bootstrapping verbs, we use $\text{SVM}_{\text{T+G}}$, as it is clearly the best available classifier (see Table 6.3). For nouns and adjectives, the discussion in section 6.5.2 showed that the case was not clear-cut. The VerbLex heuristic, which can be used either as an unsuper stand-alone classifier or as a feature, introduces a very strong new resource: our own verbal shifter lexicon.

For nouns, the stand-alone $\text{VerbLex}_{\text{deriv}}$ classifier outperformed our supervised classifier $\text{SVM}_{\text{T+G}}$. Adding it as a feature to the SVM classifier ($\text{SVM}_{\text{T+G+V}}$) did not provide significant improvements either, suggesting training data for nominal shifters may be unnecessary. For adjectives, on the other hand, the question is whether $\text{VerbLex}_{\text{stem}}$ can contribute to the supervised classifier at all, as $\text{SVM}_{\text{T+G}}$ outperformed $\text{SVM}_{\text{T+G+V}}$. However, before reaching a final verdict, let us investigate both these questions further. To this end, we run three separate bootstrapping classifiers for nouns and adjectives: $\text{SVM}_{\text{T+G}}$, $\text{SVM}_{\text{T+G+V}}$ and the best-performing VerbLex feature of the respective part of speech. For nouns this is $\text{VerbLex}_{\text{deriv}}$ and for adjectives it is $\text{VerbLex}_{\text{stem}}$.

6.6.2 *Quantitative Evaluation*

Figure 6.3 illustrates the number of *predicted* shifters returned by the bootstrap classifiers and how many of them are correct. For each classifier, the overall size of its bar indicates the number of words that it predicted to be shifters. Furthermore, the bar is divided into true positives (actual shifters, as confirmed by a human annotator) and false positives (non-shifters that were mislabelled by the classifier).

In Figure 6.3 we see that the $\text{SVM}_{\text{T+G}}$ classifier is considerably more conservative for nouns and adjectives than it is for verbs, labelling only 408 nouns and 360 adjectives as shifters, compared to the 1,043 verbs. This is likely due to the lower frequency of shifters among those parts of speech in our training data (see Table 6.1). Adding VerbLex

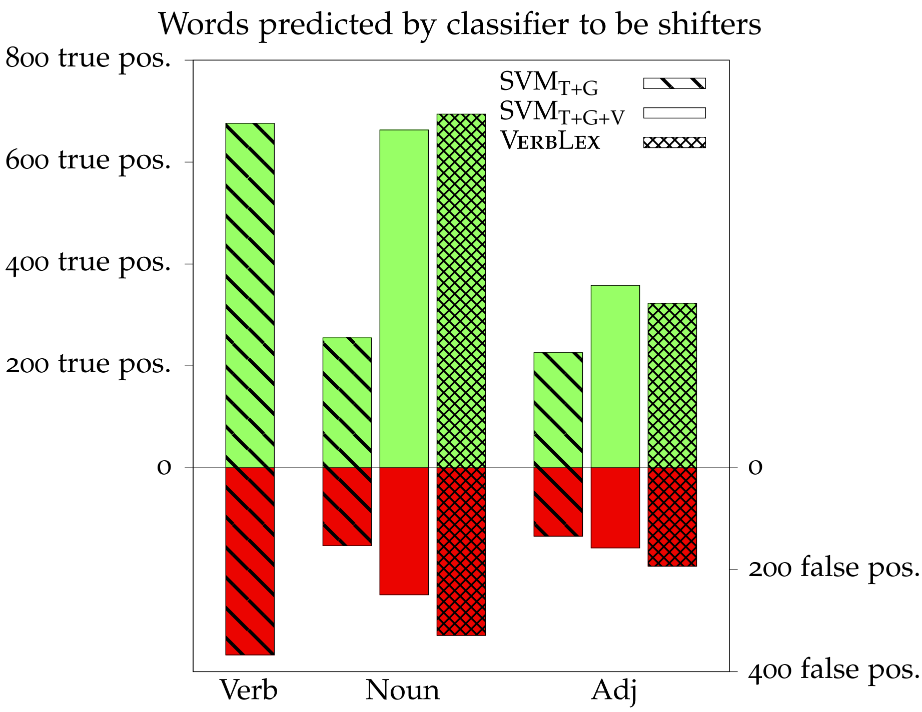

Figure 6.3: Quantitative evaluation of bootstrapped shifters. Each bar represents the number of words that a classifier predicted to be shifters, split by how many are actually shifters (true positives) and how many are misclassified non-shifters (false positives).

information to the classifier helps with this issue, as the output of SVM$_{T+G+V}$ shows. Output for nouns is more than doubled and that for adjectives is increased by over 40 percent.

Comparing SVM$_{T+G+V}$ to VerbLex, we see that, for nouns, the SVM classifier succeeds in filtering out false positives without removing many true positives. Therefore, its output has a higher precision, reducing the verification effort without incurring large losses in recall. For adjectives this is even more obviously true, as SVM$_{T+G+V}$ has the same number of predicted shifters as VerbLex but contains 35 more true positives.

In conclusion we can say that the question of which classifier can be considered the most useful depends on whether a gold standard for all parts of speech is available or not. If it is, combining all available resources in a supervised classifier is a good way to balance precision and recall. In cases where a gold standard would have to be created from scratch, it is advisable to only create one for verbs and move the remaining annotation effort to the verification phase.

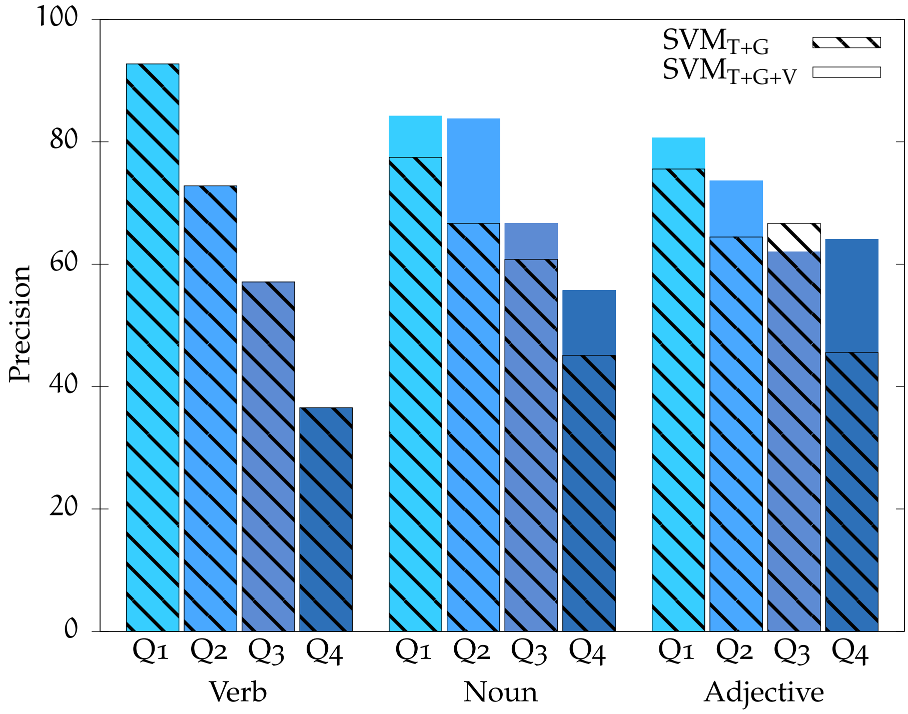

Figure 6.4: Qualitative evaluation of bootstrapped shifters. The list of words predicted as shifters is sorted by the classification confidence and split into four groups. The first quarter (Q1) contains the highest and the fourth quarter (Q4) the lowest confidence candidates.

6.6.3 *Qualitative Evaluation*

Our SVM classifiers provide a confidence value for each label they assign, indicating how certain they are that their choice is correct. In Figure 6.4 we inspect whether higher confidence also translates into higher precision. For this we rank the bootstrapped shifters of each classifier by their confidence value and then split them into four groups, from highest to lowest confidence. As the VerbLex heuristics do not provide confidence values, we limit this evaluation to the two best SVM classifiers.

For verbs, we can see a clear trend that high confidence also means high precision. Precision is at a strong 92.7 percent for the highest confidence quarter, but drops to 36.5 percent for the lowest quarter. This shows that verification for low confidence items is certainly recommended, but also suggests that manual verification may be skipped for high confidence items. In cases where employing a human annotator for a manual verification step is not feasible, a high precision lexicon can be ensured by limiting it to only items of high confidence.

The SVM$_{T+G}$ classifier, which was used for verb classification, also shows similar behavior when classifying nouns and adjectives, although the difference between high and low confidence items are less distinct.

	VERBS		NOUNS		ADJECTIVES	
	FREQ.	PERC.	FREQ.	PERC.	FREQ.	PERC.
Gold						
Annotated	2,000		2,000		2,000	
Shifter	304	15.20%	107	5.35%	129	6.45%
Non-shifter	1,696	84.80%	1,893	94.65%	1,871	93.55%
Bootstrap						
Annotated	1,043		1,270		832	
Shifter	676	64.81%	793	62.44%	512	61.54%
Non-shifter	367	35.19%	477	37.56%	320	38.46%
Complete						
Annotated	3,043		3,270		2,832	
Shifter	980	32.21%	900	27.52%	641	22.63%
Non-shifter	2,063	67.79%	2,370	72.48%	2,191	77.37%

Table 6.4: Result of the lexicon generation workflow outlined in Figure 6.1. Gold and bootstrap lexicon are merged to create a single large shifter lexicon. For each version we provide information on how many words were annotated by a human annotator and their shifter distribution.

$\text{SVM}_{\text{T+G+V}}$ profits from the addition of VerbLex information, improving precision over $\text{SVM}_{\text{T+G}}$ in all but one quarter (while significantly increasing recall, as we saw in Figure 6.3).

6.7 CREATING THE COMPLETE LEXICON

With the bootstrapping process complete, all that is left to do is to consolidate all our data into a single polarity shifter lexicon. The complete lexicon will contain all words verified by a human, i. e. all words from the gold standard (section 6.1) and all bootstrapped words (section 6.6). In the case of nouns and adjectives, for which the bootstrap output of three different classifiers was evaluated, we combine the verified output of all three.

Table 6.4 shows the annotation effort for each dataset and its balance of shifters versus non-shifters. The benefit of the bootstrapping process is clearly visible. The percentage of shifters among bootstrap data is far higher than that among the randomly sampled gold standard. While the amount of bootstrap data that had to be annotated was roughly half of what was annotated for the gold standard, it contains more than double as many verbal shifters, over seven times as many nominal shifters and four times as many adjectival shifters.

Overall, the bootstrap process produced 1,981 shifters among 3,145 words. Based on the gold standard shifter frequencies, we must assume that to find as many shifters by blindly annotating random parts of the vocabulary, we would have had to annotate 24,000 additional words. Taking into account the 6,000 words annotated for the gold standard, our approach reduces the annotation effort by over 72 percent, a significant saving of over 680 work hours. In total our complete polarity shifter lexicon contains 2,521 confirmed shifters and 6,624 confirmed non-shifters.

6.8 conclusion

In expanding our bootstrapping efforts from verbs to nouns and adjectives, we observe both similarities and differences between the parts of speech. There exists a far greater number of nouns and adjectives in English than there are verbs, making it even more important to aid the annotation process through bootstrap classification. At the same time, the number of shifters relative to the size of the vocabulary is considerably lower in these parts of speech, making supervised training of classifiers more challenging.

To overcome this challenge, we introduce new features that use the lexicon of verbal shifters which we had already bootstrapped, transferring its shifter labels from verbs to nouns and adjectives. This proves to be an excellent feature for labeling nouns, in part due to the availability of derivational relatedness mappings between verbs and nouns. Results for adjectives are more mixed, as there are no equivalent mapping resources available, so using a combination of features for supervised classification remains the best approach.

Combining the results of our bootstrapping efforts in this chapter with those from chapter 3, we compile a **general lexicon of polarity shifters**. Now that this is accomplished, it is time to address a detail that we have so far ignored in our work: shifting directions. Many shifters can both shift positive polarities to negative and negative polarities to positive. Some shifters, however, only do one or the other. They only shift in one direction and when they encounter a word of the other polarity, their polarity remains unaffected. We will look into this phenomenon in the upcoming chapter.

7 EXTENDING THE LEXICON BY INTRODUCING SHIFTING DIRECTIONS

The lexica we have created in previous chapters approach polarity shifting as a binary state of affairs. Either a word (or word sense) is a shifter or not. However, some shifters only shift polarities in one direction.

Many polarity shifters can affect both positive and negative expressions. In (7.1), the verbal shifter *destroy* shifts a positive polar expression to negative, while in (7.2) it shifts from negative to positive.

(7.1) It [**destroyed**$_{shifter}$ their [hopes]$^{+}$]$^{-}$.
(7.2) That would [**destroy**$_{shifter}$ the [cancer]$^{-}$]$^{+}$.

Other shifters, however, are unidirectional and only affect expressions of a single polarity (Wilson et al., 2005b). The verbal shifter *to risk*, for example, shifts only positive polar expressions like *good health* in (7.3), while the polarity of negative polar expressions like *war* in (7.4) remains unaffected. Similarly, the adjectival shifter *antiquated* shifts the positive noun *ideal* in (7.5), but not the negative noun *stereotype* in (7.4)

(7.3) You [**risk**$_{shifter}$ your [good health]$^{+}$]$^{-}$.
(7.4) Their actions [**risk**$_{shifter}$ a [war]$^{-}$]$^{-}$.
(7.5) The "American dream" is an [**antiquated**$_{shifter}$ [ideal]$^{+}$]$^{-}$.
(7.6) Women belonging in the kitchen is an [**antiquated**$_{shifter}$ [stereotype]$^{-}$]$^{-}$.

Conversely there are shifters that only affect negative expressions but not positive ones, such as *recoup* in (7.7) and (7.8) and *amend* in (7.9) and (7.10).

(7.7) She must [**recoup**$_{shifter}$ her [losses]$^{-}$]$^{+}$.
(7.8) I could [**recoup**$_{shifter}$ a [fortune]$^{+}$]$^{+}$.
(7.9) Let us [**amend**$_{shifter}$ that [problem]$^{-}$]$^{+}$.
(7.10) We can [**amend**$_{shifter}$ the [solution]$^{+}$]$^{+}$ to improve its clarity.

In its current state, the general shifter lexicon that we created in chapters 3 and 6 does not take the **shifting direction** of a polarity shifter into account. As a result, the polarities of sentences such as (7.4) and (7.8) would erroneously be assumed to have shifted. To prevent such mistakes, this chapter introduces a supervised classifier for extending the lexicon to include information on the potential shifting directions of each polarity shifter.

Contents

To begin this new classification task, we first define a gold standard for it in section 7.1. In section 7.2 the classifiers and the features that they include are described. They are then evaluated in section 7.3 and the best classifier is used in section 7.4 to assign direction labels to the remaining shifters that were not part of the gold standard for the general shifter lexicon from chapter 6. Section 7.5 concludes the chapter.

Contributions

(i) We extend the general lexicon of English polarity shifters (chapters 3 and 6) to include information on the shifting direction of each shifter.

(ii) We introduce a supervised classifier to automatically determine the shifting direction of a shifter, using features established in previous chapters, as well as newly introduced ones.

(iii) We discover significant differences between the shifting direction tendencies of different parts of speech.

Publication History

The work in this chapter is also contained in Schulder et al. (under review). The shifting direction extension of the general shifter lexicon is publicly available.[1]

7.1 GOLD STANDARD

We again use supervised classification to support our annotation efforts. To train and evaluate the classifier, we create a gold standard for shifting directions. It consists of the 304 verbal, 107 nominal and 129 adjectival shifters that are part of the gold standard introduced in section 6.1.

All annotations are performed by an expert annotator[2]. In addition, 200 words were independently labeled by another annotator, resulting in a Cohen's kappa inter-annotator agreement (Cohen, 1960) of $\kappa = 0.65$, indicating substantial agreement.

This gives us a list of 540 polarity shifters, which are now annotated for their shifting direction. Each shifter is given one of three labels:

AFFECTS POSITIVES: The shifter only affects positive polarities

1 https://doi.org/10.5281/zenodo.3365601

2 For organizational reasons, the annotation of verbal shifters and that for nominal and adjectival shifters had to be performed by different annotators

	VERBS		NOUNS		ADJECTIVES	
AFFECTS	FREQ.	PERC.	FREQ.	PERC.	FREQ.	PERC.
Positives	47	15.46%	36	33.64%	85	65.89%
Negatives	86	28.29%	15	14.02%	8	6.20%
Both	171	56.25%	56	52.34%	36	27.91%
Total	304		107		129	

Table 7.1: Distribution of shifting directions among the 540 polarity shifters found in the shifter gold standard (see Table 6.1).

AFFECTS NEGATIVES: The shifter only affects negative polarities

AFFECTS BOTH: The shifter is bidirectional

The resulting label distribution can be seen in Table 7.1. Interestingly, the individual parts of speech show distinctly different distributions. About half the verbs and nouns are bidirectional, but among adjectives only a quarter are bidirectional, while two thirds affect only positive words and almost none affect only negative words. Among verbs, almost 30 percent affect only negatives and 15 percent positives. For nouns, this distribution is reversed. For a classifier to perform well in our task, it will have to take these differences in label distribution into account.

7.2 CLASSIFIERS AND FEATURE DESIGN

As with our polarity shifter classification in section 6.5, we use an SVM classifier to label our remaining data. Unlike then, we must choose among three labels, rather than two, so we replace SVM^{light}, which is a binary classifier, with the multi-class classifier $SVM^{\text{multiclass}}$ (Tsochantaridis et al., 2005). We use most of the features we defined in sections 6.4.1 and 6.4.2 as well as additional ones that we will introduce in this section.

BASELINES: As baselines we define two majority classifiers and a word embedding classifier. $\text{Baseline}_{\text{MAJ}}$ assigns the overall majority label *'affects both'* to all words, based on the label distribution of shifter directions we observe in our gold standard (Table 7.1).

$\text{Baseline}_{\text{POS\_MAJ}}$ assigns to each word the majority label for its respective part of speech, i. e. verbs and nouns are still labeled as *'affects both'*, but adjectives receive the label *'affects positives'*. This is a stronger baseline than Baseline$_{\text{MAJ}}$, as it takes into account the label distributions of individual parts of speech.[3]

For Baseline$_{\text{Word Embedding}}$ we train an SVM classifier using the word embedding we created in section 3.3. We use the 500 dimensions of the embedding vector of a word as its features for the classifier. The result is that, similarly to the label propagation classifiers from chapters 3, 5 and 6 distributionally similar words are assigned the same label.

BASIC SET OF ESTABLISHED FEATURES: Most of the features that we designed for detecting polarity shifters can also be used for classifying shifting directions. True to their name, the generic features from section 6.4.2 can all be used again. This covers the *WordNet* gloss bag of words, hypernym relation and lexicographer senses, as well as the *FrameNet* frame memberships. From the set of task-specific features from section 6.4.1, we use the particle verb feature (PRT) and the $+/-$effect feature (EFFECT).

UNUSED SHIFTER FEATURES: Some of the polarity shifter features introduced in chapters 3 and 6 will not be used for the classification of shifting directions.

Distributional similarity to negation words (NEGATIVE) and the *any* heuristic (ANY) were both designed to indicate words that were used in similar contexts as negation words. This only tells us whether a word is likely to shift polarities, but not which polarities it affects. Similarly, derivational relatedness to verbal shifters (VerbLex) only provides information on shifting, not direction, and requires a full coverage lexicon of verbs from which to map to nouns and adjectives. As all words in this task have already been confirmed to be shifters, these features convey irrelevant information.

The polarity clash feature (CLASH) proved to be too sparse for direction classification, as including the polarity of the shifter removes many shifters that are either neutral or for which the polarity is not known. We replace it with the *scope polarity* feature (described below), which models the polarity of the shifting scope, rather than the interplay of polarities between shifter and scope.

SCOPE POLARITY: Many unidirectional shifters are far more frequently used in contexts that involve the polarity that they affect than those with the unaffected polarity. For example, the verbal shifter *fend off*, which affects only negative expressions, occurs almost five times as often with negative expressions than with positive ones. Similarly, the verb *spoil* occurs almost three times as often with (affected) positive expressions as with (unaffected) negative expressions.

3 We also investigated providing part of speech information as a feature for our SVM classifier. Unfortunately, this did not integrate well enough with the other features.

CLASSIFIER	PRECISION	RECALL	F1
BASELINE$_{\text{MAJ}}$	16.2	33.3	21.8
BASELINE$_{\text{POS_MAJ}}$	40.5	45.7	42.9*
BASELINE$_{\text{Word Embedding}}$	58.1	58.1	58.1*
SVM$_{\text{Basic}}$	67.2	57.1	61.6
SVM$_{\text{Basic + Scope Polarity}}$	67.7	58.5	62.7†
SVM$_{\text{Basic + WordNet Extension}}$	69.7	62.0	65.6†
SVM$_{\text{Basic + Scope Polarity + WordNet Extension}}$	**72.5**	**64.8**	**68.4**$^{\dagger *}$

*: F1 is better than previous classifier (paired t-test with $p < 0.05$).
†: F1 is better than all baselines (paired t-test with $p < 0.05$).

Table 7.2: Results of shifting direction classification. The evaluation is run as a 10-fold cross validation and all reported metrics are macro-averages. Best results are depicted in bold.

We count in our text corpus how often the scope of a shifter has a positive or negative polarity. The scope of the shifter is defined as the dependency relations outlined in section 6.3.

WORDNET EXTENSION: To discern the directions in which a word may shift, we need additional ways to distinguish semantic nuances. For this reason we increase the amount of information we extract from *WordNet*.[4] In addition to the bag of words of glosses, we also provide a bag of words that contains words from both the glosses and the usage examples given for a word. (Using a bag of words composed only of examples resulted in worse performance, due to too little overlap in word-choice between examples for different words.)

Similar to our use of *hyperonymy* relations, which indicate more general meanings of a word, we now add *hyponymy* relations, which indicate terms that are more specific forms of the given word.

To establish further semantic connections between words, we also include relations that define *similarity*, *antonymy*, *entailment causation*, *attributes*, as well as the *WordNet* relations *also see* and *verb group* and the syntactic relation of *derivational relatedness*.

7.3 EXPERIMENTS

Table 7.2 shows the performance of our shifting direction classifiers. Like previously, we evaluate them using 10-fold cross validation. All reported metrics are macro-averages.

4 The additional *WordNet* features described here were also considered for bootstrapping shifter lexica in chapters 3, 5 and 6, but were omitted from those chapters as they did not improve performance.

	VERBS		NOUNS		ADJECTIVES	
AFFECTS	FREQ.	PERC.	FREQ.	PERC.	FREQ.	PERC.
Positives	66	9.76%	158	19.92%	471	91.99%
Negatives	154	22.78%	20	2.52%	5	0.98%
Both	456	67.46%	615	77.55%	36	7.03%
Total	676		793		512	

Table 7.3: Distribution of shifting directions among the 1,981 bootstrapped shifters (see section 6.6). Labels were determined using the best shifting direction classifier from Table 7.2.

As we only have limited amounts of training data, we train all three parts of speech together to avoid sparsity issues. This is in contrast to chapter 6, in which parts of speech were trained individually, due to the availability of larger amounts of training data and the use of cross-POS features.

$\text{Baseline}_{\text{Word Embedding}}$ represents a strong supervised baseline. It clearly outperforms the two majority label baselines and shows that the semantic information provided by a word embedding is a solid feature for classifying shifting directions. Unfortunately, we found that using the word embedding as a feature in our multi-feature SVM classifier did not integrate well with the other features.

Instead, we extend the set of features that we previously introduced in section 6.4 ($\text{SVM}_{\text{Basic}}$) by adding scope polarity and extending our set of *WordNet* features. The classifier combining all those features ($\text{SVM}_{\text{Basic + Scope Polarity + WordNet Extension}}$) significantly outperforms all other classifiers and raises the F-score by 10 points above that of the best baseline.

7.4 CLASSIFICATION OF UNLABELED SHIFTERS

We use our best directions classifier, which contains the Basic, Scope Polarity and WordNet Extension feature groups, to classify all 1,981 polarity shifters that were not part of the directions gold standard. This covers the 676 verbs, 793 nouns and 512 adjectives that were confirmed to be shifters in the verification phase of the shifter bootstrapping process (see section 6.6).

Table 7.3 shows the distribution of shifting direction labels among the automatically classified shifters. We see the same trends as for our gold standard (see Table 7.1), albeit with a stronger bias towards the majority label of each part of speech.

As part of our previous polarity shifter bootstrapping workflow (Figure 6.1), we performed a shifter verification step, in which all words predicted to be shifters were verified by a human annotator, while those predicted to be non-shifters were discarded. Such a verification step is not possible for the current classification task of shifting directions, as we are equally interested in all three labels. Verifying *all* words manually would of course defeat the purpose of performing an automatic classification. Therefore we use the classifier labels without further verification.

Using the labels from our gold standard and the automatic classification, we extend the general shifter lexicon presented in section 6.7 by adding shifting direction labels to all shifters. We also include information for each item regarding whether its label was determined by a human annotator as part of the shifting direction gold standard or automatically assigned by our classifier. They are freely available as part of the bootstrapped shifter lexicon[5].

7.5 CONCLUSION

In this chapter we addressed a specific attribute of polarity shifters: their shifting direction, which determines which polarities they affect.

We developed a new gold standard for this task and found that while many shifters affect both positive and negative polarity expressions, a significant number of them shift only one of the two polarities. We find that the distribution of shifting directions is strongly dependent on the part of speech of the shifters in question. While over half of verbs and nouns are bidirectional, adjectives were found to show a strong tendency to only shift from positive to negative polarity. Verbs and nouns also showed diverging preferences, with more verbs affecting only negative polar expressions and more nouns affecting only positive polarity expressions.

To equip the entirety of our bootstrapped shifter lexicon with shifting direction labels, we trained an SVM classifier using a combination of features from previous chapters as well as newly introduced ones, such as scope polarity and additional *WordNet* attributes and relations.

As a result, our general lexicon of polarity shifters can now provide information on shifting directions for every shifter. Whether this information will help to improve applications, we will investigate in the upcoming chapter 8, in which we finally apply a number of our shifter lexica to a sentiment analysis task, comparing their performance to a neural network classifier that learns negation implicitly.

5 https://doi.org/10.5281/zenodo.3365601

8 APPLYING THE LEXICON TO SENTIMENT ANALYSIS

One of our main motivations for creating a large lexicon of polarity shifters has been to provide a resource that can help improve natural language applications. This of course begs the question whether knowledge of polarity shifters can bring such improvements. We must also ask whether efforts to make the shifter lexicon more fine grained, such as the word sense annotation in chapter 4 or the shifting direction extension in chapter 7, are of use. In this chapter we investigate these questions for the task of **phrase-level polarity classification**.

Classifying polarities at the phrase-level is not only an intermediate step in compositional sentence-level classification, but is also independently used for applications like knowledge base population (Mitchell, 2013), summarization (Stoyanov and Cardie, 2011) and question answering (Dang, 2008). To investigate the importance of explicit knowledge about polarity shifters for such tasks, we create a gold standard of sentences from the product review domain that is annotated for changes of polarity between polar expressions and the phrases that they are contained in.

Contents

In section 8.1 we describe the experimental design of our classification task and introduce our gold standard. Section 8.2 compares the performance of a state-of-the-art compositional polarity classifier (without explicit knowledge of shifters) with an approach that uses the knowledge provided by our bootstrapped shifter lexicon. In section 8.3 we investigate whether using the knowledge about shifting directions that we added to the lexicon in chapter 7 can improve performance further. Finally, in section 8.4 we compare our bootstrapped lexicon from chapter 3, which is lemma-based, to the sense-level lexicon from chapter 4. Section 8.5 concludes the evaluation.

Contributions

(i) We introduce an evaluation specifically designed to judge whether classifiers correctly detect when polarities shift in the context of a verb phrase. To perform this evaluation, we create a gold standard of 2,631 verb phrases, annotated for in-context occurrences of polarity shifting.

(ii) We evaluate how well our bootstrapped lexicon identifies instances of shifting, comparing it to a state-of-the-art compositional polarity classifier that has previously been shown to perform well at handling negation.

(iii) We provide a comparison of our different English shifter resources, evaluating whether shifting directions or sense-level shifter labels can improve performance for in-context detection of shifting.

Publication History

The work presented in this chapter is also contained in Schulder et al. (under review). The experimental setup and comparison to existing methods (sections 8.1 and 8.2) were previously published in Schulder et al. (2017). The phrase-level polarity classification gold standard has been released publicly.[1]

8.1 EXPERIMENTAL SETUP

The question we seek to answer in this experiment is whether a given classifier can correctly decide if the polarity of a word has changed in the context of a phrase. For example, the noun *passion* is of positive polarity, but in (8.1) it is affected by the verbal shifter *lack*, resulting in the negative polarity phrase *"lack her usual passion"*.

(8.1) The book seemed to [$\textbf{lack}_{\mathrm{V}}$ [her usual $\textbf{passion}^{+}_{\mathrm{N}}$]$_{\mathrm{NP}}$]$^{-}_{\mathrm{VP}}$.

We limit this evaluation to cases involving (potential) shifting through verbs. This ensures a fair comparison of the bootstrapped lexicon from chapter 3 with the sense-level lexicon from chapter 4, as the latter only covers verbs.

We determine the scope of the shifter as the direct object of the (potential) shifter, defined in section 6.3. The direct object must be a noun with non-neutral polarity. As in previous chapters, this is determined using the *Subjectivity Lexicon* (Wilson et al., 2005b) (see section 3.3.1). We do not consider sentences that also contain a negation word to avoid the complication of multiple polarity shifts cancelling each other out. This gives us a verb phrase (VP) that contains a verb (the potential shifter) and a polar noun. The question that must be answered in our evaluation is whether the polarity of the polar noun and the VP are the same or different. In other words, whether the polarity has *'shifted'* or *'not shifted'*. We can therefore pose it as a binary classification task

To create a dataset for the evaluation we extract sentences from the *Amazon Product Review Data* text corpus (see section 3.3.2) that contain

1 https://doi.org/10.5281/zenodo.3364811

FIELD	CONTENT	POLARITY
Verb	*soothe*	
Sentence	*Norah Jones' smooth voice and soft jazz piano work could soothe any savage beast.*	
Polar Noun	*beast*	negative
Verb Phrase	*soothe any savage beast*	positive
Shifting Label	Shifted	

Table 8.1: Annotation example of a shifted phrase in the gold standard for the sentiment analysis evaluation. Given the context of the sentence, the annotator must determine the sentiment polarities of the polar noun and the verb phrase. Based on these, the shifting label is set according to the mapping described in Table 8.3.

FIELD	CONTENT	POLARITY
Verb	*mutter*	
Sentence	*The elderly crowd could be heard muttering their shocked disapproval as we left.*	
Polar Noun	*disapproval*	negative
Verb Phrase	*muttering their shocked disapproval*	negative
Shifting Label	Not shifted	

Table 8.2: Annotation example of a phrase whose polarity was not shifted.

a VP headed by a verb that has a polar noun as a dependent. The polarity of the noun is determined using the *Subjectivity Lexicon*.

We began by annotating 400 sentences in which the verb is a polarity shifter according to our shifter lexicon. Next, we annotated 2,231 sentences where the verb is a non-shifter. This way the ratio of shifters and non-shifters in the sentences matches the ratio of verbal shifters and non-shifters from the gold standard (Section 3.2).[2] To cover a variety of different shifters, rather than only the most frequent ones, each shifter may only occur once in our data set.

Tables 8.1 and 8.2 show examples of annotated sentences, including the fields of information provided to the annotator during creation of the dataset. In Table 8.1 the polarity was found to have shifted, while in Table 8.2 it had not. For each sentence, the annotator must choose the polarities of the given noun and VP, labeling each as either *'positive'*, *'negative'* or *'neutral'*. The full sentence that the VP was taken

2 We rely on the shifter distribution from the bootstrapping gold standard of chapter 3, rather than on that of the manually created lexicon from chapter 4 for the simple reason that the shifting direction gold standard was created before the manually created lexicon was available.

NOUN POLARITY ⇒ VERB PHRASE POLARITY = ?				
		VERB PHRASE POLARITY		
		Positive	Neutral	Negative
NOUN POLARITY	Positive	*Not shifted*	*Shifted*	*Shifted*
	Neutral	*Shifted*	*Not shifted*	*Shifted*
	Negative	*Shifted*	*Shifted*	*Not shifted*

Table 8.3: Mapping for labeling polarity transitions in the gold standard. The shifting label is based on the agreement of polarities between the polar noun and the verb phrase that contains the noun.

from is provided to clarify the context in which the phrase appears. The verb that is the head of the VP (and therefore the potential shifter) is also explicitly identified to avoid confusion in cases where more than one verb occurs in the phrase.

The field *shifting label* shows the label that classifiers will have to determine in our evaluation. It is automatically determined for the annotated dataset, based on the polarities of the polar noun and the VP annotated by the annotator. If both polarities are identical, the label is *'not shifted'*, in all other cases the label is *'shifted'*.

In the example in Table 8.1, the annotator labels the noun *beast* as negative and the VP *soothe any savage beast* as positive. Based on these polarities, the shifting label of the sentence is determined to be *'shifted'*. In Table 8.2, on the other hand, the annotator labels both the noun *disapproval* and the VP *muttering their shocked disapproval* as negative, resulting in a shifting label of *'not shifted'*.

The shifting label is used as the classification label for our evaluation. Classifiers may either provide the shifting label directly or provide noun and VP polarities, from which the shifting label is then inferred.

As we discussed in section 2.1.2, there is no consensus on how to model how far the polarity of a shifted word moves. Expressions like *"it wasn't excellent"* have been argued to convey either positive (Choi and Cardie, 2008) or neutral polarity (Taboada et al., 2011; Kiritchenko and Mohammad, 2016).

Our evaluation is concerned with whether shifting occurs, rather than with the exact polarities or polar intensities involved. To accommodate both legitimate interpretations, we count both behaviors as shifting. As long as the polarity of the polar noun and that of the VP are not identical, we consider it to be shifting. Our own approach gains no advantage from this decision, as its decisions are solely driven by its knowledge of shifters and not by the polarities involved.

8.2 COMPARISON TO EXISTING METHODS

In this section we will answer the question of whether knowledge of polarity shifters can be used to improve polarity classification.

As baselines we use a majority classifier that labels all sentences as *'not shifted'* and the *Recursive Neural Tensor Network* (RNTN) tagger by Socher et al. (2013), which is considered to be the state-of-the-art for handling negation at the phrase-level.

RNTN is a compositional sentence-level polarity classifier that achieves strong performance on polarity classification datasets. Given the constituency parse of a sentence, it determines the polarity of each tree node. This allows us to extract the polarities it assigns to the relevant nouns and VPs in our data.

One of the major strengths of RNTN is that it can learn polarity shifting effects (as caused by negations and shifters) implicitly from labeled training data, rather than requiring explicit knowledge of shifters or shifting rules. It does, however, rely on labeled training data in the form of sentences with a constituency parse tree, each node of which has been labeled with polarity information. Such data is expensive to create. To this date, the only manually annotated dataset that provides such fine-grained polarity information is the *Stanford Sentiment Treebank* (SST) (Socher et al., 2013).

Unfortunately, resources like SST do not contain most shifters with sufficient frequency to either train or test the ability of a classifier to handle polarity shifters. For example, SST only contains 30 percent of the polarity shifters from our lexicon. Over a third of these occur only a single time. RNTN, which was trained on SST, has been shown to successfully model the shifting effect of negation words, as these occur with considerably higher frequency. We do not, however, expect it to be able to handle any but the most frequent polarity shifters.

Our own approach (LEX) first determines the polarity of a given noun using the *Subjectivity Lexicon* and infers the polarity of the VP through our knowledge of polarity shifters. If the head verb of the VP is a shifter according to our lexicon, then the polarity of the VP is set to be the opposite of the polarity of the noun. Is the verb a non-shifter, then the VP receives the same polarity as the noun.

We evaluate our approach with three versions of the verbal shifter lexicon that we bootstrapped in chapter 3. LEX_{SVM} uses the list of shifters that was output by our best SVM classifier, but without the verification step later performed by a human annotator. Similarly, LEX_{LP} uses the output of the best LP classifier without human verification. LEX_{GOLD} uses the final verified version of the shifter lexicon. As LEX_{GOLD} was also used to determine the ratio of shifters and non-shifters for the gold standard, it should not be considered as a competing classifier, but rather as an upper bound to the expected performance of the LEX approach.

	CLASSIFIER	PRECISION	RECALL	F1
Baseline	MAJORITY	39.95	50.00	44.41
	RNTN	50.81	51.16	50.98*
Lemma Lexicon	LEX_{LP}	77.71	67.38	72.18*
	LEX_{SVM}	81.63	80.95	81.29*
	LEX_{GOLD}	**88.85**	**81.18**	**84.84***

*: F1 is better than previous classifier (paired permutation test with $p < 0.05$).

Table 8.4: Classifier performance for sentiment analysis task of determining whether shifting occurs between a polar noun and the VP that contains it (see section 8.1). The head verb of the VP is either a verbal shifter or a non-shifter. All metrics are macro-averages. Best results are depicted in bold.

Results in Table 8.4 show that our approach clearly outperforms both baselines. While RNTN performs slightly better than the majority classifier, it still fails to detect most instances of shifting. Even the classifiers using automatically induced lexica, LEX_{LP} and LEX_{SVM}, provide a significant improvement over RNTN. As could be expected, considering the difference in output quality between the SVM and LP classifiers, LEX_{SVM} outperforms LEX_{LP} significantly. In fact, LEX_{SVM} comes fairly close to the upper bound of LEX_{GOLD}.

Even LEX_{GOLD} still contains a number of misclassifications, however. These can be caused by a number of factors:

1. Unidirectional shifters that do not affect the polarity of the given noun, like the verbal shifters *risk* in *"risk a war"* and *recoup* in *"recoup a fortune"*.

2. Verbs that are shifters in some word senses, but not the one encountered in the given sentence, such as the verbal shifter *bring down* in *"bring down a curse"*.

3. Additional phenomena that affect the polarity of the verb phrase.

In the following sections we will evaluate possible solutions to the first two factors: Section 8.3 will add shifting directions to our lexicon and section 8.4 will explore the issue of word senses. For a discussion of additional phenomena that influence polarity, see section 2.1.4.

8.3 COMPARISON TO LEXICON WITH SHIFTING DIRECTIONS

In chapter 7 we discussed the fact that some shifters are unidirectional, i. e. they affect only either positive or negative expressions but not both. When they occur with expressions of the unaffected polarity, they do not function as shifters. Our LEX classifier, however, is unaware of this

	CLASSIFIER	PREC.	RECALL	F1
Directions Lexicon	LEX+DIR$_{SVM}$	88.92	80.33	84.41
	LEX+DIR$_{GOLD}$	**89.26**	**81.18**	**85.03***
Lemma Lexicon	LEX$_{GOLD}$	88.85	**81.18**	84.84

*: F1 is different from previous classifier (paired permutation test with $p < 0.05$).

Table 8.5: Comparison of shifter lexica with and without shifting direction information on sentiment analysis task (see section 8.1). All metrics are macro-averages. Best results are depicted in bold.

phenomenon and will always shift the polarity. Take, for example, the verb *to free*. While it is a shifter, as can be seen in (8.2), it only affects negative expressions. In our sentiment analysis gold standard *to free* occurs in the context of the sentence seen in (8.3). As the affected expression here is of positive polarity, no shifting occurs. The LEX classifier, however, would erroneously consider it to have shifted to negative.

(8.2) Police [**freed**$_{shifter}$ the [hostages]$^{-}$]$^{+}$.
(8.3) It makes you think of the breakup of TBS and how it [**freed**$_{shifter}$ this [great talent]$^{+}$]$^{+}$ to go on his own.

To remedy this flaw, we extend our gold lexicon classifier (LEX$_{GOLD}$) to take shifting directions into account (LEX+DIR). For this we use the direction labels that we generated in chapter 7 (LEX+DIR$_{SVM}$). In addition we also manually create a **gold** direction lexicon (LEX+DIR$_{GOLD}$). This is to allow us to differentiate between the question of whether our bootstrapped direction lexicon can improve performance and the question of whether knowledge of shifting directions is helpful in general, regardless of errors in the automatic classification.

Table 8.5 shows the performance of the direction-enhanced lexicon classifiers (LEX+DIR) and compares them to the best classifier without shifting direction information (LEX$_{GOLD}$). The bootstrapped directions (LEX+DIR$_{SVM}$) appear to be too noisy, marking several shifters as not affecting the polarity of the given noun when in fact they do. While the gold direction lexicon (LEX+DIR$_{GOLD}$) does remove five false positive classifications, this is not a significant difference (given a two-tailed paired permutation test with $p < 0.05$).

We now must ask ourselves why we encounter so few cases in our dataset in which unidirectional shifters co-occur with scopes of an unaffected polarity. The most apparent contributing factor is the size of the dataset. Of the 2,631 sentences in the gold standard, knowledge of directions can only help with the 400 sentences that LEX$_{GOLD}$ marks as *'shifted'*. Among these, only 173 are unidirectional according to our direction gold annotation, which matches the ratio of direction labels as described in Table 7.1.

The second factor is the balance between sentences in which unidirectional shifters occur with the affected and unaffected scopes. Affected scopes are instances in which the shifting scope has a polarity that the shifter can affect (i. e. shift). Unaffected scopes are instances in which no shifting occurs, e. g. because the shifter is unidirectional and only affects negative polarities, while its scope is of positive polarity. As we described in our motivation of the scope polarity feature in section 7.2, many unidirectional shifters occur far more frequently with affected scopes than with unaffected scopes. When observed across our entire text corpus, we find 142,439 occurrences of the 173 unidirectional shifters in which their scope is polar according to the *Subjectivity Lexicon*. In only 15.63 percent of theses cases does the scope have the unaffected polarity.

Another source of mistakes for our direction-aware shifter lexicon are errors in the assumed polarity of the noun. LEX_{GOLD} uses the *Subjectivity Lexicon* to determine polarities of words. These do not necessarily match the true polarities of a sentence. In (8.4), for example, the word *monster* is of positive polarity, as it is in the context of a video game where monsters fight for the user. The *Subjectivity Lexicon* marks *monster* as a negative term, however. As the shifter *to waste* only affects positive expressions, a direction-aware classifier relies on receiving the correct polarity.

(8.4) [..] half the time you [$\textbf{waste}_{\textit{shifter}}$ a [monster]$^{+}$]$^{-}$ by making a false fusion.

Considering these various factors, we must conclude that a considerably larger gold standard of shifter occurrences is required to draw statistically significant conclusions about the modeling of unidirectional polarity shifters. The question of whether shifting direction can help avoid false positive classifications of polarity shifters therefore remains open.

8.4 COMPARISON TO SENSE-LEVEL LEXICON

Some words only work as shifters in some of their word senses. Take the verb *to bring down*, for example. In its meaning of *"to cause the downfall of rulers"* it is clearly a shifter, as seen in (8.5). On the other hand, the sense of *"to impose something unpleasant"* used in (8.6) causes no shifting.

(8.5) The revolution [[**brought down**]$_{\mathbf{V}}$ the $\textbf{tyrant}^{-}_{\mathbf{N}}$]$^{+}_{\text{VP}}$.
(8.6) She [[**brought down**]$_{\mathbf{V}}$ a $\textbf{curse}^{-}_{\mathbf{N}}$]$^{-}_{\text{VP}}$ on the village.

This issue is different from that of shifting directions (see chapter 7 and section 8.3). For both (8.5) and (8.6) the polarity of their direct object (which would be their shifting scope if they were shifters) is negative, but only (8.5) shifts it.

We addressed the matter of differentiating between shifter and non-shifter word senses of the same lemma when we created a verbal shifter lexicon through manual annotation in chapter 4. In the remaining chapters, we chose not to differentiate between word senses, but to rather assign a single shifter label per lemma (see section 3.2), as we were using automatic classification as part of our bootstrapping effort.

To label individual word senses during bootstrapping, we would have had to rely extensively on the use of word sense disambiguation (WSD). In addition, any application seeking to make use of a sense-level lexicon would have to use WSD, too. We argued at the time that available WSD tools would not be robust enough for our needs. In this section we will put this claim to the test by applying our sense-level shifter lexicon from chapter 4 to our sentiment analysis task and evaluating both its general potential and its performance in conjunction with automatic word sense disambiguation.

For this evaluation we define a new classifier SENSE, which works like LEX except that it uses the sense-level shifter lexicon from chapter 4 and requires a word sense to be chosen for each potential shifter. We use the SENSE classifier with a number of different word sense selection mechanisms to evaluate how much performance can be improved by sense-specific shifter labels and whether word sense disambiguation is sufficiently reliable to make use of this potential.

As the sense-level shifter lexicon is annotated with *WordNet* word senses, we choose a WSD tool that uses the same sense inventory to avoid noisy mappings between different sense inventories.[3] The tool we use is *WordNet::SenseRelate::AllWords* (*SenseRelate*) by Pedersen and Kolhatkar (2009), which implements ten different similarity and relatedness measures on the *WordNet* sense inventory. We evaluate each one and report the results of the best performing one, the **vector** relatedness measure, which defines relatedness as the overlap of *WordNet* glosses using a vector space model.

Apart from the output of *SenseRelate*, we also evaluate a baseline which always chooses the first word sense for each word ($\text{SENSE}_{\text{FIRST}}$). As *WordNet* sense numbers are ordered to list common uses before uncommon ones, this makes for a stronger baseline than randomly choosing a sense. To establish an upper bound for the performance of the sense-level lexicon, we also provide an ORACLE classifier that knows the correct label for each sentence and tries to choose word senses that result in that label ($\text{SENSE}_{\text{ORACLE}}$). Note that this does not guarantee perfect performance, as the oracle can only influence the label if the verb in question has both shifter and non-shifter senses.

3 An alternative is the supervised shifter disambiguation approach of Wiegand et al. (2018a). Training data consists of sentences in which words are labeled as shifters or non-shifters by manually annotating their *WordNet* sense and then selecting their label from a sense-level shifter lexicon like SENSE. However, the gold standard in Wiegand et al. (2018a) only covers 20 shifters and creating a larger one is exactly the kind of large scale annotation effort that we seek to avoid in this dissertation.

	CLASSIFIER	PREC.	RECALL	F1
Sense Lexicon	$\text{SENSE}_{\text{FIRST}}$	85.71	67.10	75.27
	$\text{SENSE}_{\text{SenseRelate}}$	86.30	67.17	75.55
	$\text{SENSE}_{\text{ORACLE}}$	**92.45**	74.34	82.41*
Lemma Lexicon	LEX_{GOLD}	88.85	**81.18**	**84.84***

*: F1 is better than previous classifier (paired permutation test with $p < 0.05$).

Table 8.6: Comparison between lemma- and sense-level shifter lexica on the sentiment analysis task (see section 8.1). $\text{SENSE}_{\text{FIRST}}$ assigns the first *WordNet* sense to each verb, while $\text{SENSE}_{\text{SenseRelate}}$ assigns the sense determined by the *SenseRelate* WSD tool. $\text{SENSE}_{\text{ORACLE}}$ always chooses an appropriate word sense where possible. All metrics are macro-averages. Best results are depicted in bold.

Table 8.6 shows the performance of the SENSE classifiers compared to LEX_{GOLD}. Looking at $\text{SENSE}_{\text{ORACLE}}$ we can see that differentiating by word senses does indeed improve precision by avoiding false positive hits. At the same time, recall suffers (relative to LEX_{GOLD}) from systematic flaws in the coverage of the sense-level lexicon that are due to its reliance on the sense inventory of *WordNet*. In some cases, specific meanings of a word are simply not defined in *WordNet* and can therefore also be missing from the sense-level lexicon. *WordNet*, for example, only lists the literal senses of the verb *derail* that pertain to trains, but not the metaphoric sense found in *"derail your chance of success"*. In other cases, the definition of a sense is so broad or vague that annotating them correctly is challenging, as it is unclear which uses of the lemma should be considered.

Comparing $\text{SENSE}_{\text{ORACLE}}$ with $\text{SENSE}_{\text{FIRST}}$ we see that to reap the potential precision improvements of the sense-level lexicon, we require actual word sense disambiguation. Unfortunately, the sense predictions offered by $\text{SENSE}_{\text{SenseRelate}}$ are not sufficient. While it chooses a different sense than the first in 47 percent of all cases, this introduces almost as many errors as it resolves and provides no significant improvement over $\text{SENSE}_{\text{FIRST}}$. This confirms our expectation that automatic word sense disambiguation is not sufficiently reliable for use in polarity shifter detection.

8.5 CONCLUSION

We evaluated our bootstrapped lexicon and by extension the usefulness of shifter knowledge for sentiment analysis tasks. To this end, we developed a classification task in which the goal is to determine whether the polarity of a noun has shifted in the context of a specific verb phrase.

We compared our bootstrapped lexicon from chapter 3 against a state-of-the-art compositional polarity neural network classifier, showing that polarity shifters, unlike negation words, cannot easily be learned implicitly from available corpora. Even the automatically labeled versions of our shifter lexicon for which the labels had not yet been verified by a human annotator clearly outperform the neural network classifier.

Next, we compared the basic version of our bootstrapped shifter lexicon with a version enhanced with the shifting direction information we introduced in chapter 7. We found here that there was very little potential for improvement, at least on our gold standard. It is not clear whether evaluations on a larger dataset would yield different results.

As a final evaluation, we took a look at the manually annotated sense-level shifter lexicon that we created in chapter 4. This lexicon is annotated for individual word senses, which introduced the new challenge of integrating word sense disambiguation into our classification. Using an oracle classifier, that sense-level shifter knowledge could introduce clear improvements for classifier precision. We also observed losses in recall, which may be due to missing word senses in *WordNet*.

Replacing the oracle decisions with actual automatic word sense disambiguation showed us that currently available WSD tools that are compatible with the *WordNet* sense inventory perform no better than a baseline that always selects the first word sense of a word. This finding supports our decision to forego sense-level annotation in our machine-supported bootstrapping approach, as word sense disambiguation does not provide the kind of robust performance that we would require.

9 CONCLUSION

In the past, polarity shifters have been largely ignored by the natural language processing community. Not because they are less important than other forms of negation, but because their lexical diversity made the creation of resources with sufficient coverage far more challenging than, for example, for negation words. This is turn inhibited their use in NLP applications, such as polarity classification or knowledge extraction. To remedy this issue, we have created several resources revolving around polarity shifters, as well as computational methods to create these resources efficiently.

9.1 BOOTSTRAPPING A LEXICON

The most prominent challenge in creating a comprehensive lexicon of polarity shifters is one of scale. To create such a lexicon by hand, one would have to annotate many tens of thousands of words. We introduced a bootstrapping approach that reduced the annotation effort by over 70 percent without sacrificing the precision quality of the resulting lexicon.

In our bootstrapping approach, a human annotator first creates a small set of training data for a supervised classifier, which is then used to label the remaining words. Words assumed by the classifier to be shifters are then verified by the human annotator to avoid false positive classifications. This allows us to skip manual annotation on a large number of likely non-shifters.

To create a high quality classifier for our bootstrapping approach, we used a variety of different features. Many of these were specifically designed for the task at hand, making use of various linguistic phenomena to extract information from a corpus. For the initial bootstrapping of verbal shifters, the most successful feature identified shifters by their co-occurrence with the word *any*, a known negative polarity item. Correlation with negative polarity items had previously been shown to exist for downward-entailing operators, which, we argued, have a strong overlap with polarity shifters.

Other useful features made use of the semantic properties of verb particles, the similarities of polarity shifters to both negation words and $+/-$effect theory and the clash of opposing polarities in phrases. Furthermore we relied on *WordNet* and *FrameNet*, both of which are known to be useful resources for sentiment analysis. We also designed a feature to identify anti-shifters, words that are antithetical to shifting, through their co-occurrence with specific adverbials.

With an eye to under-resourced languages, we tested a classifier that requires no labeled training data at all, which presents good performance. However, even better results are achieved when making use of both our gold standard and our entire collection of features. Using these, we trained a supervised classifier whose performance is strong enough to almost make human verification of high confidence classifications superfluous.

This was also confirmed in our extrinsic evaluation in which our lexicon was used to identify shifting polarities in product reviews. In this task, the automatically classified lexicon performed nearly as well as the annotator-verified version. Regardless of which version of the lexicon was used, all of them clearly outperformed a state-of-the-art compositional classifier that had previously been shown to successfully learn negation caused by negation words from a polarity-annotated treebank. The lesson learned here is that while negation words are comparatively easy to learn, the large lexical variety of polarity shifters is far more challenging, lending greater importance to the availability of lexical resources on shifters.

9.2 TRANSFERRING KNOWLEDGE

As one of our main aims in this work was to limit annotation efforts, we began our work by focussing only on English verbal shifters. Once that resource was created, we used it to support us in the creation of further resources.

We showed how the verbal shifter lexicon could be used to help in the creation of lexica for other parts of speech as well as other languages. In the course of this, we identified both the strengths and the limitations of such an approach. Transferring shifter labels from verbs to nouns worked extremely well, suggesting a strong semantic connection between these parts of speech in matters of negation. Adjectives, on the other hand, proved to be more of a challenge. This was in part due to a lack of cross-POS mappings, but later investigations into shifting directions also suggested that there are some general differences in how adjectival shifters work compared to verbal and nominal shifters, as their shifting direction distribution is significantly different.

Possibly even more relevant than extending the lexicon across parts of speech is transferring it to other languages. We investigated this for German as an exemplary case. Bearing in mind that many languages have far fewer resources available than exist for English, we not only identified how resource-hungry our existing features are, but also introduced low-resource approaches for mapping polarity shifter labels from English to another language. Apart from using bilingual dictionaries, which may vary greatly in size and coverage depending on the language in question, we also investigated the use of word embeddings. While cross-lingual word embeddings have traditionally relied on parallel corpora, which are rare and expensive to create, developments in recent years have introduced cheaper methods. With the help of *VecMap*, which can align word embeddings from different corpora without even the need for seed mappings, we successfully transferred shifter labels from English to German with almost the same quality as a large dictionary. The implication of this is that given enough raw text (and ideally a lemmatizer) to train a word embedding, a shifter lexicon can be bootstrapped from lexica of other languages. If this can be combined with supervised training and other features, results can be improved even further.

9.3 ADDING DEPTH

Apart from expanding coverage, we also worked on adding more nuanced information about shifters. Using supervised classification, we determined for the entire English bootstrapped lexicon which shifters actually shift polarities in both directions and which only affect either positive or negative polarities. As part of this we found that different parts of speech exhibit decidedly different directional tendencies.

We provided other kinds of information through manual annotation. Labeling individual word senses, we found that shifters are 50 percent more likely than non-shifters to be polysemous (i. e. to have multiple word senses). Among shifter words with more than one word sense, less than a quarter are actually shifters in all their senses. Using this data, we also showed that the number of potential negation events in the review domain is as large for verbal shifters as for negation words. This figure does not even include nominal and adjectival shifters, so it is likely that negation through shifters is more frequent than negation through negation words.

To determine which parts of a phrase are affected by a shifter, we had to determine its shifting scope. We found that scope could be determined using syntactic relations, that the set of possible relations were dependent on the part of speech of the shifter and that specific shifter words could have one or several potential scopes. While we used a limited list of scopes in our bootstrapping efforts, e. g. verbs were assumed to shift the polarity of their direct object, our manual annotation for verbal shifters determined all potential scopes for each word. This showed that a significant number of verbal shifters instead affected their subject, specific prepositional objects or in rare cases clausal complements.

9.4 THE RESOURCES

We created a total of four resources in the course of this work: A general lexicon of English polarity shifters, a lexicon of English verbal shifter senses, a lexicon of German verbal shifters and a gold standard for shifting polarities in product review expressions. All resources are publicly available.[1]

9.4.1 *The General Lexicon of English Polarity Shifters*

The general lexicon of shifters was created using our bootstrap approach, combining supervised classification and human verification. Its individual components are described in chapters 3, 6 and 7.

The lexicon contains 2,521 verified polarity shifters and 6,624 verified non-shifters. This includes 980 verbal shifters, 900 nominal shifters and 641 adjectival shifters.

All shifters are also labeled with their shifting direction. A quarter of the shifting directions were verified by a human annotator, while the rest were automatically classified. Whether a direction label was verified by an annotator is marked in the lexicon.

9.4.2 *The Lexicon of English Verbal Shifter Senses*

The lexicon of verbal shifter senses was created through manual annotation of the *WordNet* vocabulary of verbal word senses. This process was described in chapter 4. It covers 23,986 lemma-synset pairs, i. e. word senses of a specific verb. This involves 10,577 lemmas and 13,426 synsets. The lexicon identifies 924 shifter synsets, resulting in 2,131 shifter lemma-synset pairs. 1,220 verbs are shifters in at least one of their word senses.

The lexicon also determines the scope of each shifter. As shifting scope depends on the syntactic properties of a verb lemma and shifters are determined per synset, the scope value is determined for individual lemma-synset pairs. Some lemma-synset pairs can also have more than one potential scope.

9.4.3 *The Lexicon of German Verbal Shifters*

The lexicon of German verbal shifters was created in chapter 5 using an adapted and extended form of the bootstrap approach that had previously been used for the creation of the general lexicon of English polarity shifters. It contains 2,595 German verbs, of which 677 verbs were verified as polarity shifters and 1,918 verbs were verified as non-shifters.

1 https://doi.org/10.5281/zenodo.3365605

9.4.4 *Shifted Polarities in Product Review Expressions*

The gold standard for shifted polarities consists of 2,231 verb phrases from the *Amazon Product Review Data*. These verb phrases are annotated for the polarity of the verb phrase itself and for the polarity of the direct object of the verb that is heading the verb phrase. From these two polarities, it is determined whether the polarity has shifted between them. Shifting occurs when a polarity moves to its polar opposite or when a shift to or from neutral polarity occurs.

Among the 2,631 annotated verb phrases, shifting occurs in 529 cases. In the remaining 2,102 cases, the polarity between verb phrase and direct object is unchanged.

9.5 FUTURE DIRECTIONS

We successfully created several resources for polarity shifters. However, the value of these resources does not lie in their mere existence, but in how they are used. Our hope is that our shifter lexica will be integrated into classifiers for tasks such as polarity classification, relation extraction and the recognition of textual entailment. It may also have uses in other domains, such as in natural language applications for education, in which the lexica could be used to detect whether language learners are able to handle negation correctly.

Use of shifters in natural language applications will also show which nuances of their behavior will require further research. Determining the scope of negation for negation words has long been a challenging focus of research. It remains to be seen whether polarity shifters pose a similar challenge or will be found to be more regular in their behavior.

Our evaluation of the use of shifting direction in determining whether shifting occurs in a phrase indicated that shifters are rarely used in non-shifting circumstances. As our data was of limited size and of a specific opinion-rich domain, this observation should not be considered conclusive. Hopefully use of our lexicon in other tasks will provide additional insights to this question.

We found that differentiating between shifter and non-shifter senses of a word was not yet feasible in natural language applications due to the insufficient robustness of word sense disambiguation methods. A second issue were systematic flaws related to using *WordNet* word senses as the basis for the sense-level shifter lexicon. This does not, however, invalidate the fact that many words shift in only some of their senses. Development of new word sense disambiguation techniques may make it worthwhile to refine the sense-level shifter lexicon further and to revisit the topic of determining shifter senses automatically.

Of all future directions, however, the one which we are most interested in is the spread of shifter resources to further languages. The need for resources and applications in languages other than English is just as large, but limitations in funding, public attention and scalability pose a great challenge. In creating a German lexicon of verbal shifters, we showed that our bootstrapping approach was adaptable and cross-lingual. We also provided guidance on how to approach lexicon creation depending on the availability of resources and annotation capabilities, making our approach fit for use with languages of all levels of infrastructure. Now all that remains is to make use of these means to spread the knowledge.

BIBLIOGRAPHY

Agirre, Eneko 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)*. Athens, Greece: Association for Computational Linguistics, pp. 33–41. ACL ANTHOLOGY: E09-1005.

Artetxe, Mikel, 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)*. Vancouver, British Columbia, Canada: Association for Computational Linguistics, pp. 451–462. ACL ANTHOLOGY: P17-1042.

Baker, C. Lee (1970). "Double Negatives." In: *Linguistic Inquiry* 1.2, pp. 169–186. JSTOR: 4177551.

Baker, Collin F., Charles J. Fillmore, and John B. Lowe (1998). "The Berkeley FrameNet Project." In: *Proceedings of the International Conference on Computational Linguistics (COLING)*. Vancouver, British Columbia, Canada: Association for Computational Linguistics, pp. 86–90. DOI: 10.3115/980451.980860.

Baluja, Shumeet, Rohan Seth, Dandapani Sivakumar, Yushi Jing, Jay Yagnik, Shankar Kumar, Deepak Ravichandran, and Mohamed Aly (2008). "Video Suggestion and Discovery for Youtube: Taking Random Walks through the View Graph." In: *Proceedings of the International Conference on the World Wide Web (WWW)*. Beijing, China: Association for Computing Machinery, pp. 895–904. DOI: 10/bb3xtr.

Baroni, Marco, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetti (2009). "The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora." In: *Language Resources and Evaluation* 43.3, pp. 209–226. DOI: 10.1007/s10579-009-9081-4.

Bird, Steven, Ewan Klein, and Edward Loper (2009). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, Inc. ISBN: 9780596516499.

Breck, Eric, Yejin Choi, and Claire Cardie (2007). "Identifying Expressions of Opinion in Context." In: *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*. Hyderabad, India, pp. 2683–2688. URL: www.aaai.org/Papers/IJCAI/2007/IJCAI07-431.pdf.

Brinton, Laurel J. (1985). "Verb Particles in English: Aspect or Aktionsart." In: *Studia Linguistica* 39.2, pp. 157–168. ISSN: 00393193, 14679582. DOI: 10/fb9g2q.

Burchardt, Aljoscha, 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)*. Genoa, Italy: European Language Resources Association, pp. 969–974. URL: www.lrec-conf.org/proceedings/lrec2006/pdf/339_pdf.pdf.

Chen, Danqi and Christopher D. Manning (2014). "A Fast and Accurate Dependency Parser using Neural Networks." In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, pp. 740–750. DOI: 10.3115/v1/D14-1082.

Choi, Yejin and Claire Cardie (2008). "Learning wiCth Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis." In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Honolulu, Hawaii: Association for Computational Linguistics, pp. 793–801. ACL ANTHOLOGY: D08-1083.

Choi, Yoonjung and Janyce Wiebe (2014). "+/-EffectWordNet: Sense-level Lexicon Acquisition for Opinion Inference." In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, pp. 1181–1191. DOI: 10.3115/v1/D14-1125.

Choi, Yoonjung, 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)*. Baltimore, Maryland, USA: Association for Computational Linguistics, pp. 107–112. DOI: 10.3115/v1/W14-2618.

Clematide, Simon 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)*. Lisbon, Portugal, pp. 7–13. DOI: 10.5167/uzh-45506.

Cohen, Jacob (1960). "A Coefficient of Agreement for Nominal Scales." In: *Educational and Psychological Measurement* 20.1, pp. 37–46. DOI: 10.1177/001316446002000104.

Danescu-Niculescu-Mizil, Cristian and Lillian Lee (2010). "Don't 'Have a Clue'?: Unsupervised Co-Learning of Downward-Entailing Operators." In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*. Uppsala, Sweden: Association for Computational Linguistics, pp. 247–252. ACL ANTHOLOGY: P10-2046.

Danescu-Niculescu-Mizil, Cristian, 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 Association for Computational Linguistics (HLT/NAACL)*. Boulder, Colorado, USA: Association for Computational Linguistics, pp. 137–145. ACL ANTHOLOGY: N09-1016.

Dang, Hoa Trang (2008). "Overview of the TAC 2008 Opinion Question Answering and Summarization Tasks." In: *Proceedings of the Text Analysis Conference (TAC)*. Gaithersburg, Maryland, USA: National Institute of Standards and Technology. URL: tac.nist.gov//publications/2008/papers.html.

Das, Dipanjan and Noah A. Smith (2011). "Semi-Supervised Frame-Semantic Parsing for Unknown Predicates." In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL/HLT)*. Portland, Oregon, USA: Association for Computational Linguistics, pp. 1435–1444. ACL ANTHOLOGY: P11-1144.

Das, Dipanjan, Nathan Schneider, Danqi Chen, and Noah A. Smith (2010). "Probabilistic Frame-Semantic Parsing." In: *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT/NAACL)*. Los Angeles, California, USA: Association for Computational Linguistics, pp. 948–956. ACL ANTHOLOGY: N10-1138.

Deng, Lingjia, 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)*. Sofia, Bulgaria: Association for Computational Linguistics, pp. 120–125. ACL ANTHOLOGY: P13-2022.

Esuli, Andrea and Fabrizio Sebastiani (2005). "Determining the Semantic Orientation of Terms through Gloss Classification." In: *Proceedings of the Conference on Information and Knowledge Management (CIKM)*. Bremen, Germany: Association for Computing Machinery, pp. 617–624. ISBN: 1-59593-140-6. DOI: 10.1145/1099554.1099713.

Fancellu, Federico, 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)*. Berlin, Germany: Association for Computational Linguistics, pp. 495–504. ACL ANTHOLOGY: P16-1047.

Fillmore, Charles J. (1967). "The Case for Case." In: *Proceedings of the Texas Symposium on Language Universals*. Ed. by Emmon Bach and Robert Harms. New York City, New York, USA: Holt, Rinehart and Winston, pp. 1–90. URL: eric.ed.gov/?id=ED019631.

Firth, John R. (1957). "A Synopsis of Linguistic Theory, 1930-1955." In: *Studies in Linguistic Analysis*, pp. 1–32.

Flekova, Lucie and Iryna Gurevych (2016). "Supersense Embeddings: A Unified Model for Supersense Interpretation, Prediction, Utilization." In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*. Berlin, Germany: Association for Computational Linguistics, pp. 2029–2041. DOI: 10.18653/v1/P16-1191.

Giannakidou, Anastasia (2011). "Negative and Positive Polarity Items: Variation, Licensing, and Compositionality." In: *Semantics: An In-*

ternational Handbook of Natural Language Meaning. Ed. by Claudia Maienborn, Klaus von Heusinger, and Paul Portner. Vol. 2. Berlin, Germany: Mouton de Gruyter, pp. 1660–1712. ISBN: 978-3-11-018523-2.

Gouws, Stephan, Yoshua Bengio, and Greg Corrado (2015). "BilBOWA: Fast Bilingual Distributed Representations without Word Alignments." In: *International Conference on Machine Learning (ICML)*. Lille, France, pp. 748–756. URL: proceedings.mlr.press/v37/gouws15.pdf.

Gyamfi, Yaw, Janyce Wiebe, Rada Mihalcea, and Cem Akkaya (2009). "Integrating Knowledge for Subjectivity Sense Labeling." In: *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT/-NAACL)*. Boulder, Colorado, USA: Association for Computational Linguistics, pp. 10–18. ACL ANTHOLOGY: N09-1002.

Hamilton, William L., 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)*. Austin, Texas, USA: Association for Computational Linguistics, pp. 595–605. DOI: 10.18653/v1/D16-1057.

Hamp, Birgit 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*. Madrid, Spain: Association for Computational Linguistics, pp. 9–15. ACL ANTHOLOGY: W97-0802.

Harabagiu, Sanda, Andrew Hickl, and Finley Lacatusu (2006). "Negation, Contrast and Contradiction in Text Processing." In: *Proceedings of the National Conference on Artificial Intelligence (AAAI)*. Boston, Massachusetts, USA, pp. 755–762. CITESEER: 10.1.1.93.7652.

Hasan, Kazi Saidul and Vincent Ng (2013). "Frame Semantics for Stance Classification." In: *Proceedings of the Conference on Computational Natural Language Learning (CoNLL)*. Sofia, Bulgaria: Association for Computational Linguistics, pp. 124–132. ACL ANTHOLOGY: W13-3514.

Haveliwala, Taher H. (2002). "Topic-Sensitive PageRank." In: *Proceedings of the International Conference on the World Wide Web (WWW)*. Honolulu, Hawaii, USA: Association for Computing Machinery, pp. 517–526. ISBN: 1-58113-449-5. DOI: 10.1145/511446.511513.

Henrich, Verena, Erhard Hinrichs, and Tatiana Vodolazova (2014). "Aligning GermaNet Senses with Wiktionary Sense Definitions." In: *Human Language Technology Challenges for Computer Science and Linguistic (LTC)*. Ed. by Zygmunt Vetulani and Joseph Mariani. Springer, pp. 329–342. ISBN: 9783319089584. DOI: 10.1007/978-3-319-08958-4_27.

Huang, Yang and Henry J. Lowe (2007). "A Novel Hybrid Approach to Automated Negation Detection in Clinical Radiology Reports." In: *Journal of the American Medical Informatics Association* 14, pp. 304–311. DOI: 10/bttjs6.

Ikeda, Daisuke, Hiroya Takamura, Lev-Arie Ratinov, and Manabu Okumura (2008). "Learning to Shift the Polarity of Words for Sentiment Classification." In: *Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP)*. Hyderabad, India, pp. 296–303. ACL ANTHOLOGY: I08-1039.

Jiang, Tianyu and Ellen Riloff (2018). "Learning Prototypical Goal Activities for Locations." In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*. Melbourne, Australia: Association for Computational Linguistics, pp. 1297–1307. ACL ANTHOLOGY: P18-1120.

Jindal, Nitin and Bing Liu (2008). "Opinion Spam and Analysis." In: *Proceedings of the International Conference on Web Search and Data Mining (WSDM)*. Palo Alto, California, USA: Association for Computing Machinery, pp. 219–230. ISBN: 978-1-59593-927-2. DOI: 10.1145/1341531.1341560.

Joachims, Thorsten (1999). "Making Large-Scale SVM Learning Practical." In: *Advances in Kernel Methods - Support Vector Learning*. Ed. by Bernhard Schölkopf, Christopher J. C. Burges, and Alexander J. Smola. The MIT Press, pp. 169–184. ISBN: 9780262194167.

Kang, Jun Seok, Song Feng, Leman Akoglu, and Yejin Choi (2014). "ConnotationWordNet: Learning Connotation over the Word+Sense Network." In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*. Baltimore, Maryland: Association for Computational Linguistics, pp. 1544–1554. DOI: 10.3115/v1/P14-1145.

Kessler, W. and Hinrich Schütze (2012). "Classification of Inconsistent Sentiment Words using Syntactic Constructions." In: *Proceedings of the International Conference on Computational Linguistics (COLING)*. Mumbai, India, pp. 569–578. ACL ANTHOLOGY: C12-2056.

Kim, Seongsoon, Hyeokyoon Chang, Seongwoon Lee, Minhwan Yu, and Jaewoo Kang (2015). "Deep Semantic Frame-Based Deceptive Opinion Spam Analysis." In: *Proceedings of the International Conference on Information and Knowledge Management (CIKM)*. Melbourne, Australia: Association for Computing Machinery, pp. 1131–1140. ISBN: 978-1-4503-3794-6. DOI: 10/gftjs4.

Kim, Soo-Min and Eduard Hovy (2006). "Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text." In: *Proceedings of the Workshop on Sentiment and Subjectivity in Text (SST@ACL)*. Sydney, Australia: Association for Computational Linguistics, pp. 1–8. ACL ANTHOLOGY: W06-0301.

Kiritchenko, Svetlana and Saif M. Mohammad (2016). "The Effect of Negators, Modals, and Degree Adverbs on Sentiment Composition." In: *Proceedings of the Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA@NAACL-HLT)*. San Diego, California, USA: Association for Computational Linguistics, pp. 43–52. DOI: 10.18653/v1/W16-0410.

Klenner, Manfred, Angela Fahrni, and Stefanos Petrakis (2009). "PolArt: A Robust Tool for Sentiment Analysis." In: *Proceedings of the Nordic Conference on Computational Linguistics (NoDaLiDa)*. Odense, Denmark, pp. 235–238. ACL ANTHOLOGY: W09-4635.

Krifka, Manfred (1991). "Some Remarks on Polarity Items." In: *Semantic Universals and Universal Semantics*. Ed. by Dietmar Zaefferer. Foris Publications, pp. 150–189. ISBN: 978-3-11-013391-2.

Ladusaw, William (1980). *Polarity Sensitivity as Inherent Scope Relations*. Outstanding Dissertations in Linguistics. New York: Garland Publishing. ISBN: 978-0-8240-4555-5.

Landis, J. Richard and Gary G. Koch (1977). "The Measurement of Observer Agreement for Categorical Data." In: *Biometrics* 33.1, pp. 159–174. DOI: 10.2307/2529310.

Linebarger, Marcia Christine (1980). "The Grammar of Negative Polarity." PhD thesis. Cambridge, Massachusetts, United States: Massachusetts Institute of Technology. URL: hdl.handle.net/1721.1/16067.

Liu, Bing (2012). "Sentiment Analysis and Opinion Mining." In: *Synthesis Lectures on Human Language Technologies* 5.1, pp. 1–167. DOI: 10.2200/S00416ED1V01Y201204HLT016.

Liu, Yang, Xiaohui Yu, Bing Liu, and Zhongshuai Chen (2014). "Sentence-Level Sentiment Analysis in the Presence of Modalities." In: *Proceedings of the International Conference on Computational Linguistics and Intelligent Text Processing (CICLing)*. CICLing 2014. Berlin, Heidelberg: Springer, pp. 1–16. ISBN: 978-3-642-54902-1. DOI: 10.1007/978-3-642-54903-8_1.

Macleod, Catherine, Ralph Grishman, Adam Meyers, Leslie Barrett, and Ruth Reeves (1998). "NOMLEX: A Lexicon of Nominalizations." In: *Proceedings of the International Congress of the European Association for Lexicography (EURALEX)*. Liège, Belgium: Euralex, pp. 187–193. ISBN: 978-2-87233-091-1. URL: euralex.org/publications/nomlex-a-lexicon-of-nominalizations.

Manning, Christopher D., Raghavan Raghavan, and Hinrich Schütze (2008). *Introduction to Information Retrieval*. Cambridge, United Kingdom: Cambridge University Press. ISBN: 0521865719.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013). "Efficient Estimation of Word Representations in Vector Space." In: *Proceedings of the Workshop at the International Conference on Learning Representations (ICLR)*. Scottsdale, Arizona, USA. arXiv: 1301.3781.

Miller, George A., Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller (1990). "Introduction to WordNet: An

On-line Lexical Database." In: *International Journal of Lexicography* 3, pp. 235–244. DOI: 10.1093/ijl/3.4.235.

Mitchell, Margaret (2013). "Overview of the TAC2013 Knowledge Base Population Evaluation: English Sentiment Slot Filling." In: *Proceedings of the Text Analysis Conference (TAC)*. National Institute of Standards and Technology. URL: www.nist.gov/tac/publications/2013/additional.papers/KBP2013_Sentiment_Slot_Filling_overview.TAC2013.proceedings.pdf.

Moilanen, Karo and Stephen Pulman (2007). "Sentiment Composition." In: *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP)*. Borovets, Bulgaria: Association for Computational Linguistics. ISBN: 978-954-91743-7-3. URL: www.clg.ox.ac.uk/_media/people:karo:sentcompranlp07final.pdf.

Morante, Roser (2010). "Descriptive Analysis of Negation Cues in Biomedical Texts." In: *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*. Valletta, Malta: European Language Resources Association, pp. 1429–1436. ACL ANTHOLOGY: L10-1229.

Morante, Roser and Walter Daelemans (2009). "A Metalearning Approach to Processing the Scope of Negation." In: *Proceedings of the Conference on Computational Natural Language Learning (CoNLL)*. Boulder, Colorado, USA: Association for Computational Linguistics, pp. 21–29. ACL ANTHOLOGY: W09-1105.

Paradis, Carita and Caroline Willners (2006). "Antonymy and Negation: The Boundedness Hypothesis." In: *Journal of Pragmatics* 38.7, pp. 1051–1080. ISSN: 0378-2166. DOI: 10.1016/j.pragma.2005.11.009.

Pedersen, Ted and Varada Kolhatkar (2009). "WordNet::SenseRelate::AllWords: A Broad Coverage Word Sense Tagger That Maximizes Semantic Relatedness." In: *Proceedings of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies*. Boulder, Colorado, USA: Association for Computational Linguistics, pp. 17–20. ACL ANTHOLOGY: N09-5005.

Polanyi, Livia and Annie Zaenen (2006). "Contextual Valence Shifters." In: *Computing Attitude and Affect in Text: Theory and Applications*. Ed. by James G. Shanahan, Yan Qu, and Janyce Wiebe. Vol. 20. The Information Retrieval Series. Dordrecht, Netherlands: Springer Netherlands, pp. 1–10. ISBN: 978-1-4020-4102-0. DOI: 10.1007/1-4020-4102-0_1.

Porter, Martin F. (1980). "An Algorithm for Suffix Stripping." In: *Program* 14.3, pp. 130–137. DOI: 10.1108/eb046814.

Prettenhofer, Peter 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). Uppsala, Sweden: Association for Computational Linguistics, pp. 1118–1127. ACL ANTHOLOGY: P10-1114.

Riloff, Ellen, 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)*. Seattle, Washington, USA: Association for Computational Linguistics, pp. 704–714. ACL ANTHOLOGY: D13-1066.

Ruppenhofer, Josef 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)*. Lisboa, Portugal, pp. 67–76. DOI: 10/gfvq2k.

Sanchez-Graillet, Olivia and Massimo Poesio (2007). "Negation of protein–protein interactions: analysis and extraction." In: *Bioinformatics* 23.13, pp. 424–432. DOI: 10/cmkgxf.

Schmid, Helmut (1994). "Probabilistic Part-of-Speech Tagging using Decision Trees." In: *Proceedings of the International Conference on New Methods in Language Processing (NeMLaP)*. Manchester, United Kingdom, pp. 44–49. CITESEER: 10.1.1.28.1139.

Schneider, Nathan, 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)*. San Diego, California, USA, pp. 546–559. DOI: 10.18653/v1/S16-1084.

Schouten, Kim and Flavius Frasincar (2016). "Survey on Aspect-Level Sentiment Analysis." In: *IEEE Transactions on Knowledge and Data Engineering* 28.3, pp. 813–830. DOI: 10/gft2gn.

Schulder, Marc, 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)*. Taipei, Taiwan: Asian Federation of Natural Language Processing, pp. 624–633. ACL ANTHOLOGY: I17-1063.

Schulder, Marc, Michael Wiegand, and Josef Ruppenhofer (2018a). "Automatically Creating a Lexicon of Verbal Polarity Shifters: Mono- and Cross-lingual Methods for German." In: *Proceedings of the International Conference on Computational Linguistics (COLING)*. Santa Fe, New Mexico, USA: International Committee on Computational Linguistics, pp. 2516–2528. ACL ANTHOLOGY: C18-1213.

Schulder, Marc, Michael Wiegand, Josef Ruppenhofer, and Stephanie Köser (2018b). "Introducing a Lexicon of Verbal Polarity Shifters for English." In: *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*. Miyazaki, Japan: European Language Resources Association, pp. 1393–1397. ISBN: 979-10-95546-00-9. ACL ANTHOLOGY: L18-1222.

Schulder, Marc, Michael Wiegand, and Josef Ruppenhofer (under review). "Bootstrapped Creation of a Lexicon of Sentiment Polarity Shifters." In: *Journal of Natural Language Engineering* Special Issue on Processing Negation.

Sennrich, Rico, 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)*. Potsdam, Germany, pp. 115–124. URL: files.ifi.uzh.ch/cl/volk/papers/Sennrich_Schneider_Volk_Warin_Pro3Gres_GSCL.pdf.

Shwartz, Vered, Yoav Goldberg, and Ido Dagan (2016). "Improving Hypernymy Detection with an Integrated Path-Based and Distributional Method." In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*. Berlin, Germany: Association for Computational Linguistics, pp. 2389–2398. ACL ANTHOLOGY: P16-1226.

Socher, Richard, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher 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)*. Seattle, Washington, USA: Association for Computational Linguistics, pp. 1631–1642. ACL ANTHOLOGY: D13-1170.

Stoyanov, Veselin and Claire Cardie (2011). "Automatically Creating General-Purpose Opinion Summaries from Text." In: *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP)*. Hissar, Bulgaria: Association for Computational Linguistics. ACL ANTHOLOGY: R11-1028.

Szarvas, György, Veronika Vincze, Richárd Farkas, and János 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)*. Columbus, Ohio, USA: Association for Computational Linguistics, pp. 38–45. DOI: 10/fh6mg4.

Taboada, Maite, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede (2011). "Lexicon-Based Methods for Sentiment Analysis." In: *Computational Linguistics* 37.2, pp. 267–307. DOI: 10/cdb62c.

Talukdar, Partha Pratim, Joseph Reisinger, Marius Paşca, Deepak Ravichandran, Rahul Bhagat, and Fernando Pereira (2008). "Weakly-Supervised Acquisition of Labeled Class Instances Using Graph Random Walks." In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Waikiki, Honolulu, USA: Association for Computational Linguistics, pp. 582–590. ACL ANTHOLOGY: D08-1061.

Tsochantaridis, Ioannis, Thorsten Joachims, Thomas Hofmann, and Yasemin Altun (2005). "Large Margin Methods for Structured and

Interdependent Output Variables." In: *Journal of Machine Learning Research (JMLR)* 6.Sep, pp. 1453–1484. URL: jmlr.csail.mit.edu/papers/volume6/tsochantaridis05a/tsochantaridis05a.pdf.

Turney, Peter (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)*. Philadelphia, Pennsylvania, USA: Association for Computational Linguistics, pp. 417–424. DOI: 10/fphr6f.

Upadhyay, Shyam, 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)*. Berlin, Germany: Association for Computational Linguistics, pp. 1661–1670. DOI: 10.18653/v1/P16-1157.

Velikovich, Leonid, 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 Association for Computational Linguistics (HLT/NAACL)*. Los Angeles, California, USA: Association for Computational Linguistics, pp. 777–785. ISBN: 978-1-932432-65-7. ACL ANTHOLOGY: N10-1119.

Wiegand, Michael and Josef Ruppenhofer (2015). "Opinion Holder and Target Extraction based on the Induction of Verbal Categories." In: *Proceedings of the Conference on Computational Natural Language Learning (CoNLL)*. Beijing, China: Association for Computational Linguistics, pp. 215–225. DOI: 10.18653/v1/W15-2921.

Wiegand, Michael, 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)*. Uppsala, Sweden, pp. 60–68. ACL ANTHOLOGY: W10-3111.

Wiegand, Michael, Sylvette Loda, and Josef Ruppenhofer (2018a). "Disambiguation of Verbal Shifters." In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC)*. Miyazaki, Japan: European Language Resources Association, pp. 608–612. ACL ANTHOLOGY: L18-1097.

Wiegand, Michael, Maximilian Wolf, and Josef Ruppenhofer (2018b). "Negation Modeling for German Polarity Classification." In: *Language Technologies for the Challenges of the Digital Age*. Ed. by Georg Rehm and Thierry Declerck. Lecture Notes in Computer Science. Springer International Publishing, pp. 95–111. ISBN: 978-3-319-73706-5. DOI: 10.1007/978-3-319-73706-5_9.

Wilson, Theresa, Paul Hoffmann, Swapna Somasundaran, Jason Kessler, Janyce Wiebe, Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan (2005a). "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*. Vancouver, British Columbia, Canada, pp. 34–35. ACL ANTHOLOGY: H05-2018.

Wilson, Theresa, Janyce Wiebe, and Paul Hoffmann (2005b). "Recognizing Contextual Polarity in Phrase-level Sentiment Analysis." In: *Proceedings of the Joint Conferences on Human Language Technology and on Empirical Methods in Natural Language Processing (HLT/EMNLP)*. Vancouver, British Columbia, Canada: Association for Computational Linguistics, pp. 347–354. DOI: 10.3115/1220575.1220619.

van der Wouden, Ton (1997). *Negative Contexts: Collocation, Polarity and Multiple Negation*. Routledge. ISBN: 978-1-134-77333-6.

Yu, Hongkun, Jingbo Shang, Meichun Hsu, Malu Castellanos, and Jiawei Han (2016). "Data-Driven Contextual Valence Shifter Quantification for Multi-Theme Sentiment Analysis." In: *Proceedings of the Conference on Information and Knowledge Management (CIKM)*. Indianapolis, Indiana, USA: Association for Computing Machinery, pp. 939–948. ISBN: 978-1-4503-4073-1. DOI: 10.1145/2983323.2983793.

Zou, Bowei, Guodong Zhou, and Qiaoming 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)*. Seattle, Washington, USA: Association for Computational Linguistics, pp. 968–976. ACL ANTHOLOGY: D13-1099.