

# Strukturierte Modellierung von Affekt in Text

Roman Klinger

Institut für Maschinelle Sprachverarbeitung, University of Stuttgart, Germany

klinger@ims.uni-stuttgart.de

## Abstract

Emotionen, Stimmungen und Meinungen sind Affektzustände, welche nicht direkt von einer Person bei anderen Personen beobachtet werden können und somit als „privat“ angesehen werden können. Um diese individuellen Gefühlsregungen und Ansichten dennoch zu erraten sind wir in der alltäglichen Kommunikation gewohnt, Gesichtsausdrücke, Körperposen, Prosodie, und Redeinhalte zu interpretieren. Das Forschungsgebiet Affective Computing und die spezielleren Felder Emotionsanalyse und Sentimentanalyse entwickeln komputationelle Modelle, mit denen solche Abschätzungen automatisch möglich werden.

Diese Habilitationsschrift fällt in den Bereich des Affective Computings und liefert in diesem Feld Beiträge zur Betrachtung und Modellierung von Sentiment und Emotion in textuellen Beschreibungen. Wir behandeln hier unter anderem Literatur, soziale Medien und Produktbeurteilungen. Um angemessene Modelle für die jeweiligen Phänomene zu finden, gehen wir jeweils so vor, dass wir ein Korpus als Basis nutzen oder erstellen und damit bereits Hypothesen über die Formulierung des Modells treffen. Diese Hypothesen können dann auf verschiedenen Wegen untersucht werden, erstens, durch eine Analyse der Übereinstimmung der Annotatorinnen, zweitens, durch eine Adjudikation der Annotatorinnen gefolgt von einer komputationellen Modellierung, und drittens, durch eine qualitative Analyse der problematischen Fälle.

Wir diskutieren hier Sentiment und Emotion zunächst als Klassifikationsproblem. Für einige Fragestellungen ist dies allerdings nicht ausreichend, so dass wir strukturierte Modelle vorschlagen, welche auch Aspekte und Ursachen des jeweiligen Gefühls beziehungsweise

der Meinung extrahieren. In Fällen der Emotion extrahieren wir zusätzlich Nennungen des Fühlenden. In einem weiteren Schritt werden die Verfahren so erweitert, dass sie auch auf Sprachen angewendet werden können, welche nicht über ausreichende annotierte Ressourcen verfügen.

Die Beiträge der Habilitationsarbeit sind also verschiedene Ressourcen, für deren Erstellung auch zugrundeliegende Konzeptionsarbeit notwendig war. Wir tragen deutsche und englische Korpora für aspektbasierte Sentimentanalyse, Emotionsklassifikation und strukturierte Emotionsanalyse bei. Des Weiteren schlagen wir Modelle für die automatische Erkennung und Repräsentation von Sentiment, Emotion und verwandten Konzepten vor. Diese zeigen entweder bessere Ergebnisse, als bisherige Verfahren oder modellieren Phänomene erstmalig. Letzteres gilt insbesondere bei solchen Methoden, welche auf durch uns erstellte Korpora ermöglicht wurden.

In den verschiedenen Ansätzen werden wiederkehrend Konzepte gemeinsam modelliert, sei es auf der Repräsentations- oder der Inferenzebene. Solche Verfahren, welche Entscheidungen im Kontext treffen, zeigen in unserer Arbeit durchgängig bessere Ergebnisse, als solche, welche Phänomene getrennt betrachten. Dies gilt sowohl für den Einsatz künstlicher neuronaler Netze, als auch für die Verwendung probabilistischer graphische Modelle.

## 1 Introduction

## 2 Einleitung

### 2.1 Die Analyse von Affekt in Text und die Verarbeitung von Sprache

Die verschiedenen Aspekte, mit denen sich *Natural Language Processing*, also die Verarbeitung von natürlicher Sprache mit Hilfe von Computersystemen und mathematischen Modellen, beschäftigt, wird häufig in verschiedene Ebenen unterteilt. Dies

<sup>0</sup>Die vorliegende Version der zusammenfassenden Darstellung meiner Habilitationsleistung ist bis auf das Layout identisch mit der offiziellen Veröffentlichung auf <http://dx.doi.org/10.18419/opus-10994>.

sind üblicherweise Aspekte der Phonetik, welche unter anderem die Schnittstelle zwischen Audiosignal und einer Verschriftlichung untersucht, Aspekte der Morphologie (der Struktur von Worten), der Syntax (welche Regeln analysiert, nach denen einzelne Elemente von Sprache zusammengesetzt werden), der Semantik (die sich mit der Bedeutung von Sprache beschäftigt), des Diskurses und der Pragmatik (welche nicht-wörtliche und kontextabhängige Bedeutungen untersucht) (Mitkov, 2003).

Es kann weiterhin zwischen propositionaler und nichtpropositionaler Bedeutung unterschieden werden. In der Modellierung fällt die Erkennung von Entitätsnamen in Text zum Beispiel in die erste Kategorie, während die Interpretation von paralinguistischen Phänomenen typischerweise in die zweite Kategorie fällt. In dieser Habilitationsarbeit diskutieren wir solche letztgenannten Phänomene, welche erstmalig von Janyce Wiebe als private Zustände einer Person, eines Autors, oder einer Figur beschrieben wurden. Dies sind also solche Eigenschaften, die nicht direkt von außen beobachtet werden können (Bruce and Wiebe, 1999; Wilson and Wiebe, 2005; Wiebe and Riloff, 2005). Wilson and Wiebe (2005) beschreiben diese *private states* als „Attributionen sowie Meinungen, Stimmungen/Sentiments, Spekulationen, Evaluationen und andere“<sup>1</sup>. Dies umfasst also streng genommen alle Zustände einer Person, welche nicht direkt beobachtbar sind. In der Praxis wird der Begriff allerdings in der Regel im Kontext der Analyse subjektiver Sprache, der Sentimentanalyse, Emotionsanalyse und Meinungsanalyse genutzt. Insbesondere wird der Begriff des „privaten Zustands“ nicht im Sinne von „Privatsphäre“ verstanden.

## 2.2 Affekt und Emotion in der Psychologie

Wir verwenden in dieser Arbeit den Begriff „Affekt“ im unscharfen Sinne, wie er in dem Forschungsfeld „Affective Computing“ gebraucht wird: „Affective Computing is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects“ (Picard, 2000). Mit der Bildung von Affekt im Menschen gehen verschiedene Prozess einher, wie zum Beispiel die Bildung von Emotionen oder auch Meinungen, allerdings sind Affekt, Sentiment und Emotion nicht synonym. Feldman-Barrett (2018) schreibt hierzu:

<sup>1</sup>“attributions, as well as opinions, emotions, sentiments, speculations, evaluations and other”, Übersetzungen aus dem Englischen durch mich

“Affect is the general sense of feeling that you experience throughout each day. It is not emotion but a much simpler feeling with two features. The first is how pleasant or unpleasant you feel, which scientists call valence. The pleasantness of the sun on your skin, the deliciousness of your favorite food, and the discomfort of a stomachache or a pinch are all examples of affective valence. The second feature of affect is how calm or agitated you feel, which is called arousal.“

*Affekt* besteht also aus den Komponenten *Valenz* und *Arousal*. In der Analyse von natürlicher Sprache wird der Begriff *Sentiment* häufig für die Einordnung der Wahrnehmung eines Ereignisses, eines Produkts oder einer Person (allgemein: eines Auslösers oder einer Ursache) verwendet, wobei typischerweise die diskreten Werte oder eine Skala zwischen den Werten *positiv* und *negativ* genutzt werden. *Sentiment* in diesem Sinne ist also dem Begriff *Valenz* sehr ähnlich.

Eine Emotion ist ein im Verhältnis zum Affekt komplexeres Phänomen, welches von Scherer (2005) folgendermaßen beschrieben wird:

“In the framework of the component process model, emotion is defined as an episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism.”

Entscheidend ist also ein Stimulus, der eine Reaktion von organischen Teilsystemen verursacht. Diese beschreibt (Scherer, 2005, Tabelle 1) als *Information Processing, Support, Executive, Action* und *Monitor*, was sich in den folgenden Emotionskomponenten widerspiegelt: (1) *Cognitive Appraisal*, (2) *Körperreaktion*, (3) *Motivationskomponente, Handlungstendenzen*, (4) *Expression* (Gesicht und Sprache) und (5) *subjektives Gefühl* (Emotionswahrnehmung). Ferner grenzt Scherer den Begriff der Emotion von Vorlieben, Einstellungen (*attitude*), affektiven Dispositionen, und interpersönlichen Haltungen (*stance*) ab. Wir halten also fest, dass in der (textuellen) Sentimentanalyse Affekt in Form von Valenz sowie Einstellungen und Haltungen zusammenkommen, aber eben nicht identisch mit Emotion sind.



Figure 1: Basisemotionen nach Ekman (1999)

Für die Meinungsbildung wie auch die Emotionsbildung ist auch das *Cognitive Appraisal* relevant, insbesondere, wenn es um die Bewertung von Ereignissen geht. (Smith and Ellsworth, 1985, Tabelle 6) stellen fest, dass zur Diskriminierung der Emotionen *Freude, Traurigkeit, Wut, Langeweile, Herausforderung, Hoffnung, Angst, Interesse, Verachtung, Ekel, Frustration, Überraschung, Stolz, Scham* und *Schuld* die folgenden Dimensionen zur Bewertung eines Ereignisses ausreichend sind: Wie *angenehm* war die Situation, welche *Verantwortung* obliegt der Fühlenden, wie *sicher* ist sich die Fühlende bezüglich der Interpretation des Ereignisses, wieviel *Aufmerksamkeit* bedarf das Ereignis, wieviel *Aufwand* bringt es mit sich, und wieviel *situationelle Kontrolle* obliegt der Fühlenden?

Es stellt sich nun die Frage der Relevanz dieser psychologischen Erkenntnisse für die Erstellung von automatischen Systemen zur Emotionserkennung. Die Antwort hängt von den Inhalten der zu analysierenden Texte ab – zur Zuweisung einer Emotion (also der automatischen Klassifikation von Emotionen in Text) verlassen wir uns auf die Inhalte im Text. Diese können Beschreibungen der an einer Emotion partizipierenden Teilsysteme sein, also von Körperreaktionen, Handlungen und Handlungstendenzen, der Bewertung von Ereignissen oder des Ausdrucks einer Emotion. Weiterhin kann der Text eine direkte Nennung einer Emotion enthalten. Zu der Bewertung eines Ereignisses sollten Beschreibungen der jeweiligen Appraisal-Dimensionen mit die Schlussfolgerung, welche Emotion vorliegt, eingehen.

In der automatischen Klassifikation von Emotionen in Text wird die Erkennung und Interpretation dieser Aspekte sowie die Einschätzung ihrer Relevanz allerdings in der Regel (meiner Kenntnis nach bis auf Balahur et al. (2012) und Shaikh et al. (2009)) einem statistischen Modell überlassen. Ein Optimierungsverfahren schätzt auf Basis annotierter Daten, welche Textelemente mit den Emotionen

auf welche Art und Weise im Kontext stehen.

Als solches Referenzsystem werden üblicherweise eines von drei psychologischen Modellen genutzt. Das Modell von Ekman (1999) stellt eine Quelle diskreter Emotionsklassen dar. Ekman (1999) hat diese Basisemotionen so voneinander unterschieden, dass alleine eine Unterscheidung der Emotion im Gesicht, also der Mimik, möglich ist. Er fokussiert sich also bezüglich des Emotionsmodells von Scherer (2005) auf die Ausdruckskomponente. Abbildung 1 zeigt die Basisemotionen nach Ekman, nämlich Freude, Wut, Ekel, Angst, Traurigkeit und Überraschung.

Ein weiteres Modell von Basisemotionen wurde von Plutchik (2001) vorgeschlagen. Er organisiert diese auf einem Rad (siehe Abbildung 2), wobei auf den Blättern die Grundemotionen und zwischen diesen sekundäre Emotionen liegen. Je weiter eine Emotion innen liegt, desto stärker wird die Intensität ( $\approx$  Arousal) wahrgenommen. Entsprechend dieses Modells ist Ekel also eine starke Form von Langeweile, was entsprechend des Appraisalmodells von Smith and Ellsworth (1985) nicht von der Hand zu weisen ist, auch wenn es auf den ersten Blick wie eine unzulässige Vereinfachung erscheint. Des Weiteren sind gegensätzliche Emotionen auch auf dem Rad gegenüberliegend angeordnet.

Neben diesen Modellen der Basisemotionen, welche diskrete Klassen zur Einordnung von Text liefern, findet sich auch das Circumplex-Modell in der automatischen Analyse von Text zur Emotionsmodellierung (Posner et al., 2005, siehe Abbildung 3). Dieses Modell stellt allerdings streng genommen kein Emotionsmodell, sondern ein Modell des Affekts dar. Die Appraisaltheorien verknüpfen jedoch diese Dimensionen mit den im Kontext von Valenz und Arousal wahrgenommenen Emotionen. Unser Fokus liegt in dieser Arbeit auf der Modellierung von diskreten Emotionsklassen.

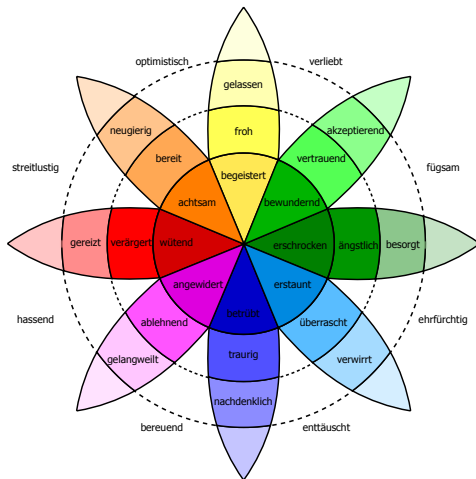


Figure 2: Emotionsmodell nach Plutchik (2001), gemeinfrei ([https://commons.wikimedia.org/w/index.php?title=File:Plutchik-wheel\\_de.svg&oldid=365781600](https://commons.wikimedia.org/w/index.php?title=File:Plutchik-wheel_de.svg&oldid=365781600))

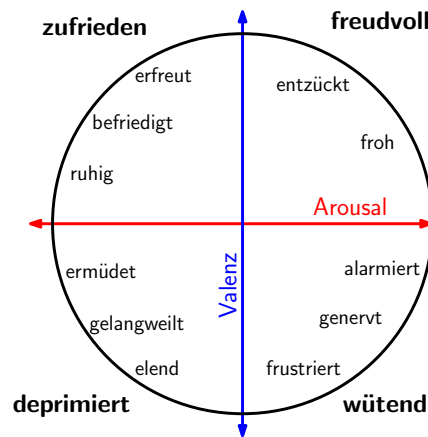


Figure 3: Das Circumplexmodell nach Posner et al. (2005), angefertigt nach Seo and Huh (2019), Übersetzung durch mich.

### 2.3 Ziel dieser Habilitation und Wissenschaftliches Vorgehen

Zum aktuellen Zeitpunkt funktionieren automatische Systeme zur Erkennung von Sentiment und Emotion in textuellen Ausdrücken noch nicht generell zufriedenstellend. Wir untersuchen hier nun Varianten der Modellierung dieser Phänomene, welche auf verschiedenen Korpora, Annotationen und informatischen Verfahren basieren, um letztendlich ein besseres Verständnis der maschinellen Lernaufgaben zu erhalten, was zu Verfahren führen soll, welche eine bessere Genauigkeit und Vollständigkeit liefern, als bisher. Für einige Probleme, insbesondere die strukturierte Emotionsanalyse, liegen bisher auch noch gar nicht hinreichende Ansätze und Korpora vor, so dass neben der komputationellen Modellierung zunächst auch die Entwicklung eines Verständnisses der Realisierungen der Phänomene im Text notwendig ist.

Das Ziel dieser Habilitation ist es, Phänomene des Affekts, von Emotionen und Sentiment entsprechend ihrer Realisierungen in Text zu modellieren. Die Struktur und das Verhältnis der Themen zueinander sind in Abbildung 4 visualisiert. Bei jedem Thema gehen wir in einem iterativen Prozess vor, welcher in Abbildung 5 skizziert ist. Dieser Prozess beginnt mit einer Hypothese. Diese umfasst Annahmen bezüglich des Modellierungsziels sowie, welche Dateneigenschaften genutzt werden. Das Modellierungsziel wird zum Beispiel in Form von einer Definition von Klassen festgehalten. Da-

tenannahmen finden sich später in der technischen Modellierung wieder, zum Beispiel über die Notwendigkeit der expliziten Formulierung von relationaler Information (zum Beispiel Sequenzinformation in Form von graphischen Modellen, rekurrenten neuronalen Netzen oder konvolutionalen neuronalen Netzen) oder auch konkrete Charakteristika, welche als Merkmale festgehalten werden.

Insbesondere Annahmen zu dem Modellierungsziel werden typischerweise in einer Annotationsrichtlinie festgehalten. Diese Richtlinie basiert zwar auf dem Modellierungsziel, kann aber auch bereits Vereinfachungen enthalten, von denen angenommen wird, dass sie komputationell notwendig sind, um eine akzeptable Performanz zu erreichen. Die Spezifikation der Annotation bewegt sich also im Spannungsfeld zwischen dem eigentlichen Modellierungsziel (zum Beispiel Emotion) und dem komputationell zu erreichenden, angenommenem Desiderat (zum Beispiel eine bestimmte Teilmenge von diskreten Emotionen zu repräsentieren).

Die Annotation wird nun von (idealerweise mehreren, unabhängigen) Annotatorinnen durchgeführt. Das Ergebnis ist eine Manifestation der durch die Annotationsrichtlinie formalisierten Hypothese, wie das Phänomen modelliert werden kann. Diese, insbesondere auf das Modellierungsziel ausgerichtete, Realisierung kann nun evaluiert werden, in dem die Zuverlässigkeit der Annotation evaluiert wird. Diese Qualitätsabschätzung dient zum einen der Verifikation, dass das Modellierungsziel tatsächlich in dem Korpus realisiert ist, zum an-

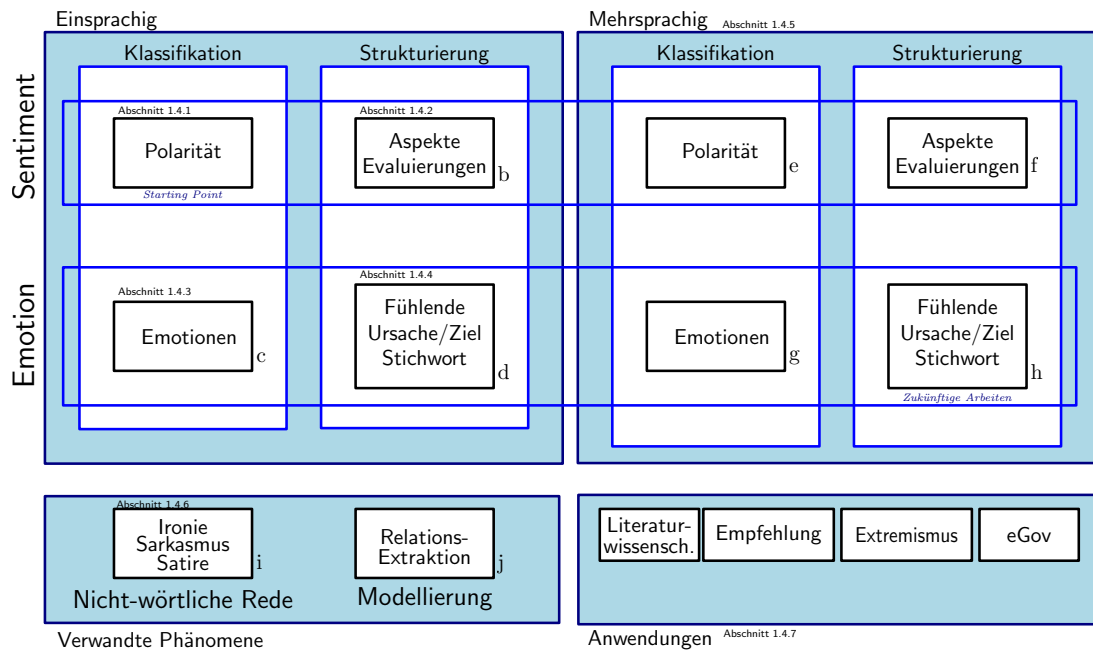


Figure 4: Struktur der Forschungsbeiträge in dieser Habilitationsarbeit

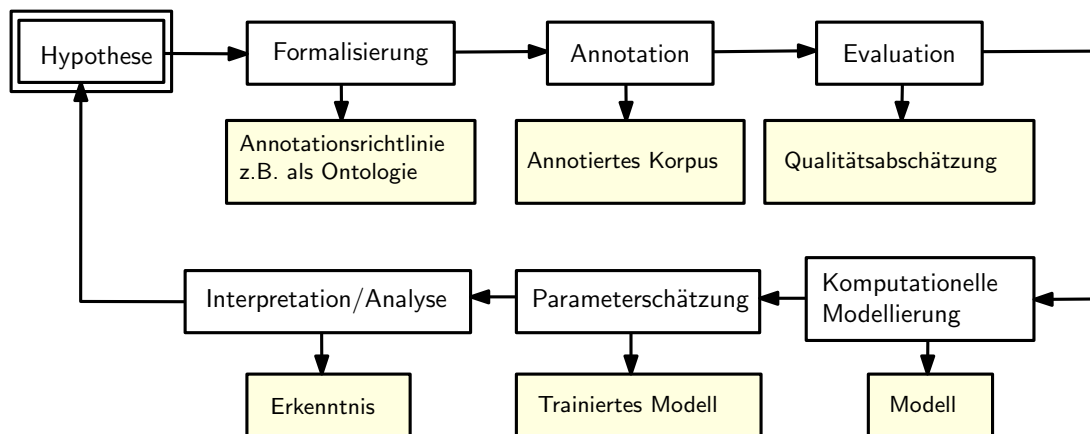


Figure 5: Wissenschaftliches Vorgehen

deren aber auch als Maß, wie herausfordernd die Aufgabe für Menschen ist.

Neben dem direkten Zweck der Analyse der Annotation dient das Korpus als Basis für die datengetriebene Modellierung. Zunächst werden Unabhängigkeitsannahmen und Merkmale implementiert, welche Annahmen über die Dateneigenschaften realisieren. Parameter, welche nicht direkt festgelegt werden können, werden in einem weiteren Schritt geschätzt, der zu einem trainierten Modell führt. Der Vergleich dieses Modells zu Variationen (zum Beispiel mit anderen Annahmen) führt zu Erkenntnissen, die die Hypothese verifizieren und falsifizieren. Dies kann zu neuer Hypothesenbildung führen.

Zur Erstellung eines komputationellen Modells

reicht es also nicht aus eine existierende Model-larchitektur aus einem etablierten Paradigma des maschinellen Lernens zu nutzen; zumindest ist dies nicht dem Erkenntnisgewinn dienlich. Vielmehr führt insbesondere die Kombination aus Datenerstellung, Problemformalisierung, Modellierung, Training und Evaluation von Korpus und Modell zu Erkenntnissen, wie ein Phänomen so repräsentiert werden kann, dass automatische Verfahren eine bessere Performanz zeigen und der Nutzen von Emotions- und Sentimentanalyse erhöht wird.

## 2.4 Varianten der Affektmodellierung

Die in dieser Habilitationsarbeit diskutierten Arbeiten betrachten verschiedene Phänomene aus dem Kontext der Affektmodellierung in Text. Dies ist

insbesondere Sentiment- und Emotionsklassifikation (Drückt ein Text etwas Positives oder Negatives aus? Wird in einem Text eine Emotion ausgedrückt oder wird sie bei einem Leser verursacht?) sowie die strukturierte Analyse (Auf welchen Aspekt bezieht sich die positive oder negative Aussage? Wer fühlt die Emotion und was hat sie ausgelöst?). Wir untersuchen hierbei (je nach Teilgebiet und Fragestellung) Aspekte der Korpuserstellung, der Auswahl geeigneter maschineller Lernmethoden sowie, in verschiedenen Fragestellungen wiederkehrend, inwieweit verschiedene Teilbereiche explizit gemeinsam modelliert werden können und dadurch jeweils besser repräsentiert werden. Dies betrifft die Formulierung im Modell und den Lernprozess oder die Inferenzprozedur.

Im Folgenden führen wir die verschiedenen Fragestellungen, welche für den monolingualen Fall an Hand eines einfachen Beispiels in Abbildung 6 illustriert sind, konkret ein, diskutieren jeweils kurz den Stand der Technik und Kunst, auf welchem wir aufbauen und welche unsere Arbeiten motivieren. Ab Abschnitt 3 betrachten wir dann konkret die neuen Beiträge dieser Habilitationsarbeit.

#### 2.4.1 Sentiment-Klassifikation

Den Startpunkt dieser Arbeit stellt die Sentiment-Klassifikation dar. Das Ziel ist es, eine Funktion zu finden, welche einer Texteinheit (einem Buch, einer Produktbeurteilung, einer Nachrichtenüberschrift, einem Blogpost, einem Tweet, einem Facebookpost oder einem Satz, seltener auch nur einem Wort) eine Klasse *positiv* oder *negativ* zuweist. Häufig wird das Problem als ein Ein-Klassen-Problem formuliert, so dass es auch üblich ist, zusätzlich die Klasse *gemischt* oder *neutral* hinzuzufügen. *Gemischt* bedeutet hierbei, dass beide Polaritäten in der Texteinheit auftreten. Zusätzlich kann auch die Intensität modelliert werden, entweder mit Hilfe einer kontinuierlichen Variable oder in dem die Klassen *sehr positiv* und *sehr negativ* mit eingebunden werden.

Ein Beispiel für eine solche Polaritätserkennung ist die Erkennung, dass der folgende Satz als *positiv* einzuordnen ist:

„Der Kaffee schmeckt mir gut.“

Im Gegensatz dazu ist der folgende Satz als *gemischt* zu beurteilen:

„Er ist ein netter Mensch, aber gestern war er schlecht gelaunt.“

Weitere Beispiele aus dem SemEval 2013 Task 2b Korpus stellt Tabelle 1 vor (Nakov et al., 2013).

Ein solches System zur Sentimentklassifikation findet zum Beispiel in der Aggregation von Nachrichten in sozialen Medien Anwendung. Es kann zum Beispiel ein Warnsystem für Produktentwickler geschaffen werden, welches auf eine Häufung von negativen Kommentaren hinweist, welche einen bestimmten Firmennamen nennen. Ebenso könnten in Produktbeurteilungen die Textstellen automatisch markiert werden, welche besonders positiv oder besonders negativ sind, so dass der Leser sich leichter einen Überblick verschaffen kann. Des Weiteren finden solche System ihren Einsatz in der Kaufsentscheidung von Aktien, in dem sie auf Nachrichtentexte angewendet werden.

Dieses Forschungsfeld wurde zu Beginn stark von Janyce Wiebe geprägt, welche zunächst das Klassifikationsproblem in die Klassen *subjektiv* und *objektiv* vorschlug (Wiebe and Riloff, 2005). Die Einordnung in *positiv* und *negativ* wird im Folgenden dann häufig nur bei subjektiven Texten als sinnvoll erachtet, da Sentimentanalyse eben den Schwerpunkt auf die Aggregation von subjektiven Meinungen legt. Eine weitere frühe Arbeit ist die auf distributioneller Semantik aufbauende Methode von Turney (2002), der die Polarität eines Texts durch automatisch bestimmte Ähnlichkeitswerte zu den Worten “poor” und “excellent” bestimmt. Wegweisend waren weiterhin verschiedene Arbeiten mit Hilfe von maschinellem Lernen auf Basis von verschiedenen Textkorpora (Täckström and McDonald, 2011; Sayeed et al., 2012; Shi and Li, 2011; Pang and Lee, 2004; Wiebe, 2000, i. a.). Mit dem Wiederaufkommen von künstlichen neuronalen Netzen in den 2010er-Jahren und dem damit einhergehenden Trend der Verwendung von Wortvektoren zur Eingaberepräsentation (Einbettungen, wie sie unter anderem mit Hilfe von Word2Vec (Mikolov et al., 2013a) oder GloVe (Pennington et al., 2014) erzeugt werden), wurde auch untersucht, inwieweit diese Vektoren zur Repräsentation von Sentiment verbessert werden können, zum einen durch eine Anpassung mit Hilfe von externem Wissen oder durch gemeinsame Anpassung in der Modelloptimierung (Yu et al., 2017; Tai et al., 2015; Tang et al., 2016; Maas et al., 2011; Tang et al., 2014; Faruqi et al., 2015).

Korpora zur Sentimentklassifikation umfassen zum Beispiel (Socher et al., 2013-10; Aggeri et al., 2013; Uryupina et al., 2014; Nakov et al., 2013).

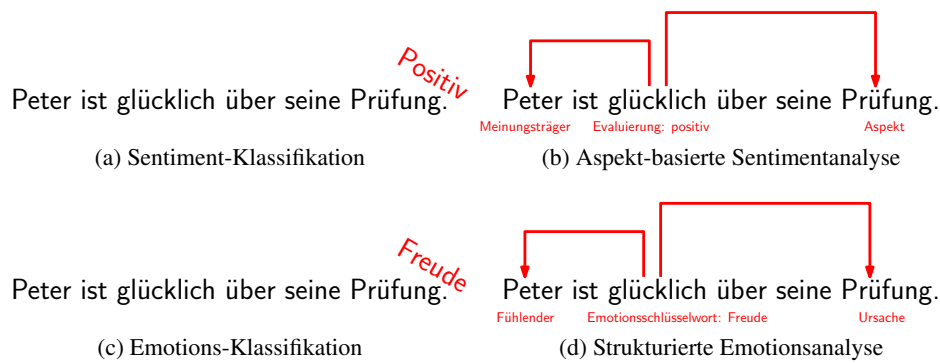


Figure 6: Beispiele für die in dieser Habilitationsarbeit diskutierten Fragestellungen im monolingualen Fall.

Klasse	Text
Negativ	And I never had a good time while sat by your bedside with papers and poetry about Estella.
Negativ	Displeased with Lion, I may take this opportunity to learn Windows 7. Or Linux. Or poke myself in the eye.
Neutral	Aged under 26 and keen to become a foreign correspondent? Apply for the Marie Colvin scholarship <a href="http://t.co/1NwsQSam">http://t.co/1NwsQSam</a>
Neutral	Because Sunday just screams Pancake Day
Neutral	AC Milan reportedly want Miralem Pjanic to join the San Siro in January #FootballFanatic
Positiv	Delicious warm sun...great to see you (if only briefly) after hurricane sandy whipped through here a couple hours ago.
Positiv	Never start working on your dreams and goals tomorrow.....tomorrow never comes....if it means anything to U, ACT NOW!

Table 1: Beispiele aus dem SemEval 2013 Task 2b Korpus für Sentimentanalyse in Twitter. Die möglichen Klassen sind *negativ*, *neutral-or-neutral*, *positiv* (Nakov et al., 2013).

Allerdings sind nicht all diese Methoden mit allen Korpora verglichen worden, was eine Motivation für unsere Arbeiten in Abschnitt 3.1 darstellt. Tabelle 2 zeigt, wie spärlich die Methoden-Korpusmatrix besetzt war.

### 2.4.2 Aspekt-basierte Sentimentanalyse

Eine Erweiterung der Sentimentklassifikation, welche im Allgemeinen nicht mehr sinnvoll als einfache Klassifikation modelliert werden kann, ist die Erkennung von Aspektennennungen zusammen mit deren Beurteilung, was gelegentlich in der Anwendung auch als Meinungsanalyse (*Opinion Mining*) bezeichnet wird (vgl. Abb. 4 auf Seite 5 und 6 auf Seite 7). Diese Aufgabe kann unterschiedlich ausgestaltet werden, sowohl auf Annotations- als auch auf komputationeller Seite. In der Literatur sind (mindestens) drei verschiedene Ansätze zu beobachten. Betrachten wir den Satz

„Das Essen schmeckt sehr gut, aber die Bedienung könnte aufmerksamer sein.“,

welcher so aus einem Korpus mit Restaurantbeurteilungen stammen könnte. Mit Hilfe der Sentimentklassifikation, wie sie in Abschnitt 3.1.1 ein-

geführt wurde, wäre eine Zuweisung von *positiv* oder *negativ* eine unangemessene Vereinfachung. Auch die Verwendung einer Klasse *gemischt* lässt nur eingeschränkt Rückschlüsse auf den Inhalt zu. Weitere Beispiele finden sich in Tabelle 3.

In der Formulierung im Sinne einer *Closed-World*-Annahme (Ansatz 1) werden die relevanten Aspekte vollständig vordefiniert. Ganu et al. (2009) verwenden in der Restaurant-Domäne zum Beispiel *Food, Service, Price, Ambience, Anecdotes, Miscellaneous*. In dem oben genannten Beispiel wäre die Ausgabe eines Systems somit „Food: *positive*“; „Service: *negative*“. Hiermit wäre aber keine automatische Zuweisung von Äußerungen zu einem möglichen Aspekt *Geographical Location* möglich (wenn zum Beispiel ein Satz „Das Restaurant liegt gut erreichbar in der Deutz-Kalker-Str.“ zu bewerten wäre).

Um diese Einschränkung zu umgehen wird (Ansatz 2) zunächst erkannt, welche Segmente im Text einen Aspekt beschreiben. Dies ist somit ein *Open-World*-Ansatz. Ein Sentimentanalyse-System hat dann die Aufgabe, aus einem gegebenen Aspekt und dem umgebenden Text die Polarität abzuleiten. In dem genannten Beispiel wäre dies für „Essen“

	BOW	AVE	RETROFIT	JOINT	LSTM	BiLSTM	CNN
<i>SST-fine</i>	-	-	-	-	+	+	+
<i>SST-binary</i>	-	+	+	-	+	+	+
<i>OpeNER</i>	+	-	-	-	-	-	-
<i>SenTube-A</i>	+	-	-	-	-	-	-
<i>SenTube-T</i>	+	-	-	-	-	-	-
<i>SemEval</i>	-	-	-	+	-	-	-

Table 2: Verfügbare Ergebnisse von State-of-the-Art-Sentiment-Modellen (Spalten) für etablierte Evaluationskorpora Mitte 2017 (Zeilen) (Barnes et al., 2017). Korpora: SST fine and binary (mehrere Klassen bzw. auf binäre Klassen abgebildet): Socher et al. (2013-10), OpeNER: Agerri et al. (2013), SenTube: Uryupina et al. (2014), SemEval: Nakov et al. (2013). Methoden: AVE: Durchschnittsbildung der Einbettungsvektoren (Faruqui et al., 2015), RETROFIT: Anpassung der Vektoren mit Hilfe eines semantischen Wörterbuchs (Faruqui et al., 2015), JOINT: Optimierung der Einbettungen zur Repräsentation von Sentiment gemeinsam mit syntaktischer und genereller semantischer Information (Tang et al., 2014). LSTM, BiLSTM, CNN beschreiben jeweilige Module innerhalb eines neuronalen Netzes (Tabelle aus Barnes et al., 2017).

*positiv* und für „Bedienung“ *negativ*. Eine Variante dieses Verfahrens ist es, nicht nur die Segmente, welche einen Aspekt beschreiben zu identifizieren, sondern auch die Phrasen, welche das Sentiment ausdrücken. Die Relation zwischen solchen evaluierenden Sentiment-Phrasen und Aspektennennungen können dann bezüglich ihrer Polarität weiter eingeordnet werden. Hier wäre in diesem Ansatz 3 die Ausgabe („Essen“, „sehr gut“, *positiv*) und („Bedienung“, „könnte aufmerksamer sein“, *negativ*).<sup>2</sup>

Solche aspektbezogenen Sentimentanalyseverfahren helfen zum Beispiel, automatisch Bewertungen von bestimmten Aspekten zu aggregieren. Hu and Liu (2004) erstellen zum Beispiel einen graphischen Überblick zu den Inhalten von Beurteilungstexten. Im Sinne von Information Retrieval helfen solche Systeme zum Beispiel auch, konverse Ansichten zu dem selben Aspekt zu finden, welche dann manuell verglichen werden können.

Wir tragen in Abschnitt 3.1.2 ein Modell bei, welches Aspektennennungen und evaluierende Phrasen entsprechend Ansatz 3 erkennt. Wir zeigen hierbei, dass gemeinsame Inferenz zu einer Perfor-

<sup>2</sup>Für eine weitere Aggregation und Analyse ist es sinnvoll die extrahierten Aspektennennungen zu gruppieren, wie zum Beispiel von Chen et al. (2016); Zhao et al. (2014); Xiong et al. (2016) vorgeschlagen.

manzverbesserung führt. In Abschnitt 3.2.2 zeigen wir eine bilinguale ressourceneffiziente Modellierung von Ansatz 2, wobei die Aspektennennungen als gegeben angenommen werden. Wir stellen ein System vor, welches besser funktioniert, als andere Ansätze, welche ähnliche Ressourcenannahmen machen.

Wir nutzen für einen weiten Teil unserer Experimente das auto- und kameraorientierte JPDA Korpus (Kessler et al., 2010) und vergleichbar annotierte Tweets von Spina et al. (2012a). Wir tragen des Weiteren zwei deutsche (Klinger and Cimiano, 2014; Sängler et al., 2016) und ein englisches Korpus (Klinger and Cimiano, 2014) bei.

### 2.4.3 Emotionsklassifikation

Die Sentimentanalyse ordnet Texte in positiv und negativ ein – dies ist für Anwendungen, in denen es konkret um Unterscheidungen von Emotionen geht, nicht ausreichend. Zum Beispiel würde die Erkennung, dass eine Menge von Twiternachrichten Angst vor einem Ereignis ausdrückt andere Reaktionen oder Aktionen, zum Beispiel auf Seiten einer Administration, verursachen, als Wut auf dasselbe Ereignis. Des Weiteren ist auch die Realisierung in Text und die unterschiedliche Verwendung in sich interessant.

In der Analyse von Emotionen in Text haben sowohl die Sentimentklassifikation, als auch die aspektbezogene Analyse eine Entsprechung. Im Klassifikationsfall werden statt der Einordnung in *positiv* und *negativ* typischerweise diskrete Klassen entsprechend der Basisemotionen, entweder in einer Multilabel- oder einer Singlelabelformulierung, genutzt. Beispiele für die Verwendung der von Ekman (1999) vorgeschlagenen Klassen sind die Arbeiten von Strapparava and Mihalcea (2007); Li et al. (2017); Alm et al. (2005). Die acht Emotionen des mittleren Rings des Emotionsrads nach Plutchik (2001) wurden zum Beispiel von Mohammad et al. (2015) genutzt. Abdul-Mageed and Ungar (2017) betrachtet unserer Kenntnis nach die größte Menge von Emotionen, welche bisher genutzt wurde, nämlich alle 24 primären Emotionen, die in Plutchiks Modell genannt werden.

Womöglich durch den Ursprung der Arbeiten zur Emotionsanalyse in Text in der Psychologie war lange Zeit ein Arbeitsschwerpunkt die Erstellung von entsprechenden Wörterbüchern, bei denen diese vom Kontext isoliert einer Emotion zugeordnet wurden. Beispiele sind das NRC Emotionswörterbuch (Mohammad and Turney, 2013)



	Text	Aspekt	Subjektive Phrase	Polarität
USAGE	I had no problems with the return.	return	no problems	positiv
	The washer itself is great, the included hose is junk.	washer hose	great junk	positiv negativ
	It looks very neat, like a storage container, and using it is very simple and easy.	looks using using	very neat very simple easy	positiv positiv positiv
SCARE	Update auf Samsung S2 gescheitert: Paketdatei ungueltig.	Update Paketdatei	gescheitert ungueltig	negative negativ
	Vorweg WICHTIGE ANMERKUNG: Wen die Vibration im Takt der Musik nervt, einfach in den Systemeinstellungen unter „Ton > Automatische Haptik“ den Haken bei Poweramp rausnehmen ;-)... Hätte nicht gedacht, dass der Player wirklich so gut ist! Klangbild ausgezeichnet, Handhabung super einfach, geile Einstellmöglichkeiten, viele Extras und ein superpraktischer Lockscreen. Genau das was ich beim Standart Player immer vermisst habe! Wirklich ein ausgezeichnete Player.	Player Lockscreen Einstellmöglichkeiten Handhabung Klangbild Player	ausgezeichneter superpraktischer geile super einfach ausgezeichnet wirklich so gut ist	positiv positiv positiv positiv positiv positiv

Table 3: Beispiele aus dem USAGE-Korpus (Klinger and Cimiano, 2014) und dem SCARE-Corpus (Sänger et al., 2016).

oder WordNetAffect (Strapparava and Valitutti, 2004), wobei auch eine bedeutende Methode und Ressource<sup>3</sup> aus der psychologischen Forschung stammt (Chung and Pennebaker, 2007).

Emotionskorpora existieren für verschiedene Domänen, von Nachrichtentexten (Strapparava and Mihalcea, 2007) über Blogs (Aman and Szpakowicz, 2007) und Dialogen (Li et al., 2017), und Literatur (Alm et al., 2005) bis hin zu Tweets (Mohammad and Bravo-Marquez, 2017; Liu et al., 2017; Mohammad, 2012). Eine herausfordernde Domäne bildet das ISEAR-Korpus (Scherer and Wallbott, 1997), bei dem Probanden gebeten wurden emotionale Ereignisse kurz zu schildern, ab (Beispiele sind in Tabelle 4 zu finden). Dieses Korpus zeigt eine Herausforderung aller Ansätze zur Emotionsanalyse: Implizite Beschreibungen sowie Ereignisbeschreibungen ohne konkrete Emotionsworte sind schwierig zu interpretieren und lassen sich daher nicht direkt mit Emotionen verknüpfen. Die Emotionstheorien der Psychologie bieten dazu zwar Ansätze, diese sind aber bisher nicht in der komputationellen Linguistik genutzt worden.

Wir tragen zu dieser Situation ein automatisch generiertes Korpus von impliziten Emotionsnennungen in Twitter, welcher in einem Shared-Task

<sup>3</sup>Linguistic Inquiry and Word Count (LIWC), <http://www.liwc.net/>

verwendet wurde, sowie ein deutsches und englisches Korpus bei, welcher ähnlich wie ISEAR erzeugt wurde, wobei aber Crowdsourcing zum Einsatz kam. Des Weiteren erstellen wir eine Ressource, in der die Art der Kommunikation der Emotion (van Meel, 1995, z.B. Vokal, Mimik, Körperposen) annotiert ist, um zu untersuchen, wie die Verteilung dieser Modalitäten im Verhältnis zu den verschiedenen Emotionen in Literatur realisiert ist (Kim and Klinger, 2019b).

Ähnlich wie Sentimentklassifikation hat auch Emotionsklassifikation bereits substantielle Aufmerksamkeit bekommen, wenn auch in geringerem Maße. Methodisch wurden verschiedene maschinelle Klassifikationsverfahren verwendet, zum Beispiel Gated-Recurrent-Units (Abdul-Mageed and Ungar, 2017), Stützvektorverfahren und Naive-Bayes (Aman and Szpakowicz, 2007) oder auch Transferlernen (Felbo et al., 2017). Allerdings wurden bisher nicht die Mächtigkeit von neuronalen Netzen und merkmalsbasierten Verfahren auf einem Korpus verglichen. Wir tragen dies ebenso bei wie einen Vergleich verschiedener Adjudikationsstrategien und eine Analyse des Einflusses auf die Modellperformanz.

Neben der Emotionsklassifikation wurde vor kurzem als Erweiterung die Vorhersage der Intensität einer gegebenen Emotion in einem Tweet als Auf-

Klasse	Text
Wut	A certain boy came into my room and started telling me that I was not pretty, that I had no manners and that I should learn some. I had just come to campus and he was in fourth year.
Wut	A colleague asked me for some advice and as he did not have enough confidence in me he asked a third person.
Ekel	The smell of garlic in rush-hour bus
Ekel	A dirty person, smelling bad and unknown laying on me.
Angst	A few weeks back my daughter had very high fever and I was scared that it might lead to brain damage or convulsions.
Angst	A terrible cloud-burst started when we were camping in a tent on a great bald field.
Schuld	A friend of mine was drunk and having an argument on the street with a woman, apparently his girlfriend. I was walking on the other side of the street and thought that I was not supposed to interfere. Later I heard that he had shot himself that evening.
Schuld	Accidentally dropped a friends camera flash, so that it was broken.
Freude	A friend which I hadn't seen for years, sent me a small present.
Freude	After one year of searching, my friend and I have found a flat we are able to pay for.
Traurigkeit	A friend of mine got pregnant and failed the first year at university. The boyfriend refused his responsibility and they ended up in court.
Traurigkeit	A good friend quitted me after he had behaved in a strange manner.
Scham	A few days back I had to ask a professor a ridiculous question because I had lost at a game of cards and had to pay a forfeit.
Scham	A girl I was with yelled something out to this group of guys I knew. It sounded really childish.

Table 4: Beispiele aus dem ISEAR Korpus, welcher häufig zur Emotionsanalyse eingesetzt wird (Scherer and Wallbott, 1997). Die möglichen Klassen folgen Ekmans Basiseemotionen mit der Erweiterung um *Ekel* und *Scham*.

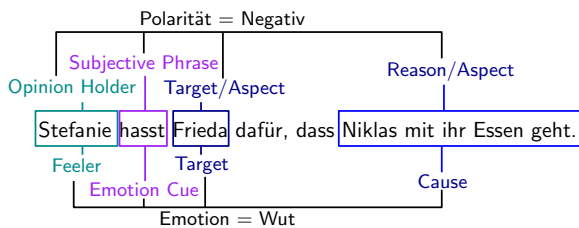


Figure 7: Vergleich der strukturierten Sentimentanalyse (oben) und der strukturierten Emotionsanalyse (unten).

gabe vorgeschlagen. In zwei Shared-Tasks wurden hierzu Arbeitsgruppen eingeladen, Systeme zu entwickeln (Mohammad and Bravo-Marquez, 2017; Mohammad et al., 2018). Wir tragen hier ein automatisches System bei, mit dem wir in der erstgenannten Iteration des Shared-Tasks den zweiten Platz belegten.

#### 2.4.4 Strukturierte Emotionsanalyse

Die Entsprechung der aspektbezogenen Sentimentanalyse in der Emotionsanalyse verlangt nach einer Generalisierung der verwendeten Rollen. Statt nur von einem *Aspekt* zu sprechen wird nun genereller eine *Ursache* untersucht und eine Emotion kann sich auch auf ein *Ziel* richten, welches nicht die Ursache darstellt. Des Weiteren sprechen wir von einer *Fühlenden* statt einem *Meinungsträger* (wobei wir letzteres in dieser Arbeit nicht diskutieren, wir verweisen die interessierte Leserin auf Wie-

gand et al. (2016); Ku et al. (2009); Wiegand et al. (2015); Wiegand and Klakow (2010)). Ein Beispiel des Vergleichs von strukturierter Sentiment- und Emotionsanalyse wird in Abbildung 7 gezeigt. Annotierte Beispiele finden sich in Tabelle 5.

Wie auch die aspektbezogene Sentimentanalyse ermöglicht diese strukturierte Emotionsanalyse weitere Fragestellungen, die automatisch beantwortet werden können. Zum Beispiel wäre die Erkennung, dass bestimmte Figuren in einem Text ängstlich sind, ohne eine solche Strukturierung nicht möglich.

Die Arbeiten in diesem Forschungsbereich sind spärlich. Das einzige uns bekannte Korpus, welches mit allen drei dieser Rollen annotiert ist, wurde von Mohammad et al. (2015) vorgestellt, wobei die Domäne, aus der die Texte stammen, recht eng definiert ist. Die Daten von Ghazi et al. (2015) enthalten Ursachenannotationen. Interessanterweise sind Emotionsursachenannotationen in anderen Sprachen als Englisch weiter verbreitet: Russo et al. (2011) erstellt ein Korpus in Italienisch, Mei et al. (2012); Gui et al. (2016) erstellen solche Korpora für Chinesisch. Die letzten beiden Ressourcen werden auch bereits für Modellierungsuntersuchungen genutzt. Chen et al. (2010) nutzt einen regelbasierten Ansatz und Gui et al. (2017) formuliert das Problem als Question-Answering-Aufgabe.

Wir tragen zu dieser Situation zunächst unser englisches Literaturkorpus REMAN (Relational

Emotion Annotation for Fiction) bei, bei dem all diese Rollen annotiert sind. Ein Modellierungsexperiment zeigt auf, dass die Vorhersage von Ereignisbeschreibungen in der Rolle der Ursache (also eine Segmentierungsaufgabe von potentiell langen Spannen) nicht hinreichend funktioniert, was uns zu der Vereinfachung als Charakterrelationsproblem führt (was wir als Klassifikationsproblem modellieren, bei dem zwei gegebene Charaktere bezüglich ihrer Emotionsrelation eingeordnet werden, siehe Abschnitt 3.3.2). Wir untersuchen weiterhin, inwieweit die gemeinsame Modellierung der verschiedenen Rollen zu einem Performanzgewinn führt.

#### 2.4.5 Mehrsprachige Sentiment- und Emotionsanalyse

Während wir bisher davon ausgegangen sind, dass eine Trainingsmenge zur Schätzung der Vorhersagefunktion aus derselben Sprache stammt wie die Texte, auf die die Funktion später angewendet werden soll, folgt hieraus eine Einschränkung: Annotierte Daten müssen für die jeweilige Zielsprache zur Verfügung stehen. Falls das nicht der Fall ist, müssen Verfahren entwickelt werden, welche ein Modell aus einer Sprache auf eine andere übertragen können, auch wenn keine Trainingsdaten in der Zielsprache existieren (hierbei handelt es sich um die Verwendung von Methoden der Domänenadaptation bzw. des Transferlernens (Pan and Yang, 2009)). Wir betrachten hier Ansätze für die Übertragung von Klassifikationsmodellen in Sentiment und aspektbezogener Sentimentanalyse und Emotionsklassifikation.

**Sentimentklassifikation.** Zu der Übertragung von Sentimentklassifikationsmodellen von einer Sprache auf eine andere wurden in der Vergangenheit verschiedene Ansätze untersucht. Diese umfassen die Verwendung von maschineller Übersetzung (Balahur and Turchi, 2012), um annotierte Daten in einer anderen Sprache als der, in der annotiert wurde, zu generieren, wie auch Verfahren, welche automatisch die Übertragung auf eine andere Sprache mit Hilfe von Domänenadaptation lernen (Pretenhofer and Stein, 2011), die Modelle delexikalisieren (Almeida et al., 2015) oder mehrsprachige Worteinbettungen nutzen oder generieren (Mikolov et al., 2013b; Hermann and Blunsom, 2014; Artetxe et al., 2016; Zhou et al., 2015). Unsere Hypothese in unserer Arbeit ist (siehe Abschnitt 3.2.1), dass maschinelle Übersetzung zu

teuer ist, um tatsächlich in ressourcenschwachen Sprachen angewendet zu werden und schlagen daher einen Ansatz vor, in dem zum einen die ursprünglichen Einbettungen nicht verändert werden und zum anderen eine Projektion in einen gemeinsamen Repräsentationsraum so optimiert werden kann, dass er Sentiment wie auch die sprachlichen Ähnlichkeiten angemessen abbildet. Wir werden sehen, dass unsere Verfahren besser funktionieren, als Methoden, welche vergleichbare Annahmen über die verfügbaren Ressourcen machen, allerdings nicht besser, als der Einsatz statistischer Übersetzungsverfahren.

**Aspektbasierte Sentimentanalyse.** Während einige Arbeiten im Sentimentklassifikationsfall existieren, sind vor unseren Studien nur wenige Ansätze zur aspektbasierten Übertragung von Sentimentanalyseverfahren zwischen Sprachen erschienen. Eine Ausnahme ist der CLOpinionMiner (Zhou et al., 2015), welcher englische Texte nach Chinesisch übersetzt und diese weiter mit Hilfe von Kotraining optimiert. Xu et al. (2013) verwenden Selftraining und Annotationsprojektion, um Meinungsträger auch in Chinesisch zu erkennen, wobei das annotierte Korpus in Englisch vorliegt.

Wir stellen in dieser Arbeit zwei Ansätze zu aspektbasierter Sentimentanalyse vor. Zum einen ist dies die Übertragung unseres Modells zur gemeinsamen Erkennung von Aspekten und evaluierenden Phrasen mittels maschineller Übersetzung und Annotationsprojektion. Hierbei untersuchen wir, inwieweit die Übersetzungsqualität Einfluss auf das System in der Zielsprache hat. Des Weiteren erweitern wir unser mehrsprachiges Sentimentklassifikationsmodell um die Verwendung von vordefinierten Aspekten. Beide Verfahren werden in Abschnitt 3.2.2 vorgestellt.

**Emotionsklassifikation.** Zu der mehrsprachigen Emotionsklassifikation existieren bisher nahezu keine Arbeiten. Es wurde zwar untersucht, inwieweit der Emotionsgehalt bei der maschinellen und der menschlichen Übersetzung erhalten bleibt (Öhman et al., 2016; Kajava and Öhman, 2019; Rabinovich et al., 2017; Öhman et al., 2018), allerdings wurden diese Erkenntnisse bisher nur im Ansatz in Modelle übertragen. Arbeiten zu der automatischen Übertragung von Emotionen zwischen Sprachen existieren fast gar nicht, mit der Ausnahme von Becker et al. (2017), welche verschiedene Klassifikationsverfahren auf eigens erstellten Nach-

Text	Segment	Role
And if this was a necessary preparation for what, should follow, I would be the very last to complain of it. We went to bed again, and the forsaken child of some half-animal mother, now perhaps asleep in some filthy lodging for tramps, lay in my Ethelwyn's bosom. I loved her the more for it; though, I confess, it would have been very painful to me had she shown it possible for her to treat the baby otherwise, especially after what we had been talking about that same evening.	I forsaken child half-animal mother filthy lodging for tramps	Feeler emotion expression, sadness target, experiencer emotion expression, disgust target, cause emotion expression, disgust target, cause
So we had another child in the house, and nobody knew anything about it but ourselves two. The household had never been disturbed by all the going and coming. After everything had been done for her, we had a good laugh over the whole matter, and then Ethelwyn fell a-crying.	The household disturbed all the going and coming	experiencer other emotion cause
What a provoking man you are! You know what I mean well enough.	provoking man I	emotion expression, anger cause experiencer

Table 5: Beispiele aus dem REMAN Korpus, Relationsnamen wurden zur Vereinfachten Darstellung auf die Sequenzen abgebildet (Kim and Klinger, 2018).

richtenkorpora vergleichen.

Wir tragen hierzu ein vergleichbares Korpus in Deutsch und Englisch bei, welches im Stil der ISEAR-Studie erstellt wurde und untersuchen die Übertragung eines Modells mit Hilfe von maschineller Übersetzung (Troiano et al., 2019).

**Strukturierte Emotionsklassifikation.** Zu der Übertragung von Systemen zwischen Sprachen, welche Rollen im Kontext von Emotionen erkennen, gibt es bisher keine Vorarbeiten. Weiterhin haben die (wenigen) Arbeiten zur monolingualen Rollenerkennung noch nicht zu zufriedenstellenden Ergebnissen geführt (wie wir in Abschnitt 2.4.4 schon angesprochen haben), so dass wir uns zunächst auf den monolingualen Fall konzentrieren und die Übertragung zwischen verschiedenen Sprachen als ein Thema sehen, welches außerhalb dieser Habilitationsarbeit angesiedelt ist.

#### 2.4.6 Ironie, Sarkasmus und Satire

Ironie, Sarkasmus und Satire verwenden häufig besonders emotionsgeladene Formulierungen und erzeugen Erwartungen, welche dann im Anschluss, insbesondere mit einer humoristischen Intention, verletzt werden. Somit sind diese Themen mit der Modellierung von Affekt verwandt, aber nicht mit ihnen identisch. Komputationelle Ironie und Sar-

kasmuserkennung verfolgt zwei Ziele. Das erste ist es, die Interpretation von Texten (also auch von Sentiment und Emotion) korrekt ihrer Bedeutung entsprechend durchzuführen und nicht von der wörtlichen Aussage irritiert zu werden. Es ist bekannt, dass die spezifische Behandlung von Ironie und Sarkasmus die Sentimentanalyse verbessern können (Maynard and Greenwood, 2014; Ghosh et al., 2015; Van Hee et al., 2015). Das zweite Ziel ist es, mit Hilfe der komputationellen Modellierung ein besser Verständnis zu erlangen und Theorien zu hinterfragen.

Die Begriffe *Ironie* und *Sarkasmus* bezeichnen beide Stilmittel, bei denen die Erwartung eines Lesers und Betrachters verletzt wird. In beiden Fällen wird häufig eine humoristische Komponente erwartet. Diese Formulierung wurde so von Utsumi (1996) als *Implicit Display Theory of Verbal Irony* vorgeschlagen. Ironie kann auch nicht-verbal auftreten, während dies bei Sarkasmus vermutlich nicht beobachtet werden kann – hier ist eine verletzende, auf eine Person, eine Gruppe oder ein Objekt gerichtete Komponente notwendigerweise vorhanden und stellt die spezifische Eigenschaft von Sarkasmus im Vergleich zu dem generelleren Stilmittel der Ironie dar (Reyes et al., 2013; Rajadesingan et al., 2015; Kreuz and Glucksberg,

1989).

Häufig werden die Begriffe Ironie und Sarkasmus allerdings austauschbar verwendet (Clift, 1999), und es wird in der Regel von einer hohen Ähnlichkeit der Begriffe Ironie, Sarkasmus und Satire ausgegangen (Ptáček et al., 2014).

Filatova (2012) hat eines der ersten frei verwendbaren Korpora, welcher bezüglich des Auftretens von Ironie und Sarkasmus (ohne eine Grenze zwischen den beiden Konzepten zu ziehen) annotiert wurde, veröffentlicht (Beispiele solcher Reviews sind in Abbildung 8 und 9 zu sehen). Es handelte sich hierbei um Amazonbeurteilungen, wobei auch Metadaten mit zur Verfügung gestellt sind. Dies unterstützte erstmalig die Möglichkeit, zu untersuchen, welche linguistischen Charakteristika wie auch bestimmte Muster bezüglich des Sentiments tatsächlich eine Rolle spielen, um Ironie/Sarkasmus zu identifizieren. Wir haben hierzu eine umfangreiche Merkmalsanalyse von bisher in der Literatur vorgeschlagenen Aspekten durchgeführt, um erstmalig ein Verständnis der Bedeutung von solchen manuell entwickelten Charakteristika von Ironie und Sarkasmus in einem prädiktiven Modell zu erlangen.

Weiterhin untersuchen wir die tatsächliche Austauschbarkeit der Begriffe Ironie und Sarkasmus, wobei wir uns auf durch Twitternutzer selbst markierte Beiträge (mit Hilfe von Hashtags) verlassen. Wir untersuchen, ob solche Beiträge komputational unterschieden werden können. Dies ist unter anderem durch die weitestgehend manuelle Korpusanalyse von Wang (2013) motiviert, welche zwar interessante Ergebnisse bezüglich der Verwendung von Sarkasmus und Ironie vorstellt, diese allerdings nicht weiter mit Hilfe von automatischen Klassifikationsverfahren untersucht.

Ironie, Sarkasmus und ein recht hoher Gehalt an Emotionen finden sich in *Satire* wieder, einem Stilmittel, bei dem mit Hilfe von Humor und häufig auch Sarkasmus und Ironie auf gesellschaftliche Probleme hingewiesen werden soll (cf. Sulzer, 1771, p. 995ff.). Die automatische Erkennung von Satire in deutschen Nachrichtentexten behandeln wir in McHardy et al. (2019).<sup>4</sup> Hier modellieren wir Satire allerdings im Gegensatz zu Buschmeier et al. (2014) und Ling and Klinger (2016) nicht auf Basis von ausgewählten Merkmalen, welche wir dann analysieren, um ein besseres Verständnis

des Phänomens zu erlangen, sondern gehen einen anderen Weg, der mit den aktuellen Veränderungen von modernen Klassifikationsverfahren ermöglicht wurde: Wir erstellen ein großes annotiertes Korpus auf Basis der Publikationsquelle (satirische Internetseite oder reguläre Internetseite) und trainieren ein Deep-Learning-Modell mit einem Attention-Mechanismus, den wir im Anschluss untersuchen, um ein besseres Verständnis zu erlangen, was das Modell gelernt hat. Die Forschungsfrage ist, ob das Satiremodell tatsächlich Satiremerkmale repräsentiert oder nur lernt, aus welcher Publikationsquelle ein Artikel stammt. Unter der Annahme, dass das Modell tatsächlich Satireeigenschaften lernt, ermöglicht dies dann in einem weiteren Schritt die Modellintrospektion, um ein tieferes Verständnis der Verwendung von Satire zu erlangen.

## 2.5 Anwendungen und Methodische Generalisierung

Die bisher genannten Themen haben wir zwar mit der Notwendigkeit, Sentiment und Emotionen in Text zu modellieren, motiviert, allerdings bisher noch wenige konkrete Anwendungen vorgestellt, welche potentiell von solchen Methoden profitieren. Im Folgenden sprechen wir Anwendungen aus dem Bereich der Politikwissenschaften, der Marktanalyse sowie der Literaturanalyse an. Des Weiteren weisen wir auf unseren grundlegenden Ansatz zur Relationserkennung hin, welcher über die Vorhersage von Strukturen im Kontext von Affektive Computing hinausgeht.

**Rechtsradikalismuserkennung.** Zunächst behandeln wir die Erkennung von möglicherweise rechtsradikal eingestellten Personen in sozialen Medien. Rechtsradikalität tritt in der Regel mit aggressivem Verhalten auf und geht mit zugrundeliegenden Einstellungen von Fremdenfeindlichkeit, Rassismus, Antisemitismus, Sozialdarwinismus und Nationalchauvinismus sowie der Verherrlichung des historischen Nationalsozialismus und der Unterstützung von Diktaturen einher (Stöss, 2010). Bisherige verwandte Arbeiten haben zum einen entsprechende Gruppen in Facebook identifiziert (Ting et al., 2013) oder bestimmte Nachrichten erkannt, welche vordefinierten Kriterien genügen. Ashcroft et al. (2015); Wei et al. (2016) sind zum Beispiel an dschihadistischen Inhalten interessiert, Scanlon and Gerber (2014) an Recruitmentaktivitäten extremistischer Vereinigungen. Es wurden

<sup>4</sup>Die Forschungsfrage und Methode wurde von mir entwickelt. Die Arbeit wurde von mir angeleitet.

## Customer Review

 Alan E. Schmidt

★★★★★ **Deeply Moved**

July 7, 2009

Format: Audio CD

I can not believe my eyes, or my ears.... The Jonas Brothers have once again blessed the world. These tunes are exactly what my melody starved soul needed. I was this close to losing my sanity and being unable to perform life's basic functions of walking, talking, and blinking. Thank heavens the release date was not pushed back.

I am concerned that the stereo in Mom's 86 Yugo can do this CD justice. Maybe if I jerry rig the home speakers and put them in the back seat I can gain some respect. Some Jonas Brothers RESPECT... It is essential that the music be loud for me... It has to penetrate the court mandated padded helmet I must wear when I venture outdoors. It does not look stupid however, I hand drew little skulls all over it to give it the "cool factor".


Just yesterday I personally experienced the Jonas Brothers touch two more lives. I was on my way to pick up Mom from the colon cleansing clinic when I was stopped at a red light. Of course I had my new Jonas Bros music at maximum level with the Yugo's windows completely lowered. The two girls in the car next to me were crying. I did not see them at first since my peripheral vision is not the best in my padded helmet. The Jonas Bros had not only awakened their deep seated emotions but it had also stimulated their dormant humor senses. They were laughing uncontrollably. I know it was the Jonas Bros because the crying laughing ladies were looking directly toward me in Mom's Yugo. I waved and winked but only generated more tears from these lovely ladies.

I believe that a wise career move for the Jonas Bros would be a multi state penitentiary concert tour. Kind of like Johnny Cash did way back in the day. Their music would definitely help rehabilitate some of those thugs. Don't you worry about their safety... the Jonas Bros are some pretty tough looking dudes. Look at the CD cover. I would not want to meet them in a dark alley, even with my padded helmet on. I know in my heart that the Jonas Bros will be very popular among the prison system population. Like female entertainers at a bachelor party.... Everyone cheering them on and wishing for some alone time with them... to "talk about life".

I will need to order myself a Jonas Bros purity ring soon. I will wear it proudly just like my idols. Afterall, I have been getting a lot of female attention lately. (mostly stares). I need to send out a message. Back off you laughing flirtatious ladies. I am saving myself for someone special.

Oh my lovely Jonas Brothers, Please don't make me wait this long again for another CD. Look for me at all your concerts. I will be the one in the padded helmet. The one with all the little skulls drawn on it.

Figure 8: Beispiel 1 einer ironischen Beurteilung, von <https://www.amazon.com/gp/customer-reviews/R107RLDLVICKLM>

 M. Anderson

★★★★★ **Cash your paycheck and flush it**

March 30, 2005

Take your friend out for beers before you waste your money on this thing. If you don't use it, it could last a very long time.

Figure 9: Beispiel 2 einer ironischen Beurteilung, von <https://www.amazon.com/gp/customer-reviews/R29EKCU0TMWWUG>

weiterhin Methoden entwickelt um ISIS- (Ferrara et al., 2016) oder Dschihadismus-Unterstützer (Kaati et al., 2015; Wei and Singh, 2017) zu identifizieren.

Wir behandeln die Erkennung von rechtsextremistisch eingestellten Personen in Twitter, wobei wir zum einen wortbasierte Verfahren einsetzen, diese aber mit Methoden vergleichen, welche den durchschnittlichen Emotionsgehalt von Nachrichten mit als Signal nutzen. Dieser Ansatz erlaubt uns, unsere Hypothese zu untersuchen, dass die Verteilung von Emotionen in Nachrichten rechtsradikal eingestellter Person anders ist, als im Durchschnitt der Bevölkerung.

**Marktanalyse.** Die Analyse von Sentiment in Produktbeschreibungen und -bewertungen hat bereits einen Wert, um besonders positive oder negative Bewertungen oder solche von bestimmten Aspekten zur Entscheidungsfindung herauszusuchen. Ein Beispiel einer solchen Aggregation ist die Arbeit von Hu and Liu (2004), welche Zusammenfassungen aus großen Mengen von Produktbeurteilungen erstellen.

Ein weiterer Aggregationsschritt wäre die Erstellung von Ranglisten von Produkten, welche ebenfalls für potentielle Kunden und Werbetreibende informativ wäre. Wir untersuchen, ob solche Listen mit Hilfe von aspektbasierter Sentimentanalyse erstellt werden können und analysieren den Informationsgehalt der verschiedenen Aspektennennungen. Wir formulieren diese Aufgabe, in dem wir Verkaufsrangordnungen mit Sentimentinformation approximieren.

**Literaturwissenschaften.** In dem vorherigen Absatz haben wir Produktrangordnungen auch zur Evaluation von Sentimentanalysesystemen beurteilt. Ähnlich nutzen wir für die Literaturanalyse die außertextuelle Variable des Genres, um deren Zusammenhang mit dem Verlauf von Emotionen zu untersuchen. Unsere Hypothese ist hierbei, dass Texte verschiedener Genres unterschiedliche Emotionsverläufe zeigen.

Bisherige Arbeiten unterschieden zwischen stilbasierten und inhaltsbasierten Ansätzen. Stilistische Verfahren nutzen zum Beispiel Frequenzen

von Funktionswörtern, Interpunktion und Wortarten (Karlgrén and Cutting, 1994; Kessler et al., 1997; Stamatos et al., 2000; Feldman et al., 2009). Inhaltsbasierte Verfahren nutzen lexikalische Informationen, zum Beispiel durch die Verwendung von Topic-Models (Karlgrén and Cutting, 1994; Hettlinger et al., 2015, 2016). Die einzige vorhergehende Arbeit, welche Emotionssequenzen im Kontext von Genre betrachtet ist Samothrakis and Fasli (2015). In Gegensatz zu dieser zeigen wir zum einen auch in einem Bag-of-Words-Ansatz, dass Emotionswörter (in Form einer Schnittmenge der auftretenden Wörter mit einem Emotionswörterbuch) ähnliche Vorhersagequalität liefern, wie ein Ansatz, der alle Wörter betrachtet. Des Weiteren untersuchen wir die Emotionsverläufe in größerem Detail und verknüpfen die Erkenntnisse mit einer qualitativen Analyse typischer und untypischer Texte.

**Generelle Deklarative Relationserkennung.** Während die drei oben genannten Arbeiten Anwendungen sind, welche durch Affective Computing verbessert werden, stellt noch eine weitere Methode einen Anteil dieser Habilitationsarbeit dar, welcher eine methodische Generalisierung ist. Die Motivation dieser Arbeit war, dass es auf der einen Seite verschiedene Werkzeuge gibt, welche die Entwicklung genereller maschineller Lernverfahren unterstützen, wie zum Beispiel Deep-Learning-Bibliotheken (zum Beispiel Theano<sup>5</sup>, PyTorch<sup>6</sup>, Keras<sup>7</sup> oder TensorFlow<sup>8</sup>) oder Bibliotheken zur Modellierung von probabilistischen Modellen, wie Alchemy<sup>9</sup> oder Factorie<sup>10</sup>. Diese Methoden sind allerdings, je nach Anwendungsfall, nicht leicht zugänglich und in der Regel von Domänenexperten ohne spezifische technische Ausbildung nur schwer zugänglich.

Auf der anderen Seite existieren nur wenige Bibliotheken, welche relationale Informationsextraktion spezifisch behandeln. Dies führt dazu, dass Modelle für die relationale Informationsextraktion, wie sie auch die strukturierte Sentiment- und Emotionsanalyse darstellt, häufig sehr spezifische Entwicklungen darstellen, welche schwierig auf andere Domänen übertragbar sind. Diese Situation

<sup>5</sup><http://deeplearning.net/software/theano/>

<sup>6</sup><https://pytorch.org/>

<sup>7</sup><https://keras.io/>

<sup>8</sup><https://www.tensorflow.org/>

<sup>9</sup><http://alchemy.cs.washington.edu/>

<sup>10</sup><http://factorie.cs.umass.edu/>

möchten wir mit DeRE (Adel et al., 2018) verbessern, welches die Aufgabenspezifikation und Modellspezifikation trennt. Dies ermöglicht dann zum Beispiel eine einfachere Verwendung eines Modells, welches für eine Problemstellung entwickelt wurde, auf einem anderen Problem. Dies wiederum unterstützt den Transfer von Methodenwissen zwischen verschiedenen Anwendungsdomänen.

### 3 Modellierung

Im Folgenden stelle ich nun meine jeweiligen Beiträge der verschiedenen Aspekte vor (vgl. auch Abbildung 4). In Abschnitt 3.1.1 diskutieren wir den Ausgangspunkt der Arbeiten sowie die Beiträge zu monolingualer Sentimentklassifikation zusammen mit Domänenadaptation. In Abschnitt 3.1.2 erweitern wir diese Arbeiten zu aspektbezogener Sentimentanalyse mit Hilfe von probabilistischen Modellen sowie Deep-Learning.

In den Abschnitten 3.2.1 und 3.2.2 werden die Arbeiten auf mehrere Sprachen übertragen, zum einen mit Hilfe einer Projektion von semantischen Repräsentationen (*Einbettungen*) und mit Hilfe von maschineller Übersetzung von Annotationsprojektion zur Erstellung von annotierten Daten, welche dann mit probabilistischen Modellen genutzt werden.

Danach erweitern wir die Fragestellungen auf Emotionsanalyse, zunächst auf monolinguale Emotionsklassifikation (und Regression) in Abschnitt 3.3.1. Dies wird in Abschnitt 3.3.2 auf strukturierte Emotionsanalyse übertragen. Erste Ansätze zu mehrsprachiger Emotionsanalyse finden sich in Abschnitt 3.4.

In Abschnitt 3.5.1 besprechen wir Ironie, Sarkasmus und Satireerkennung, in Abschnitt 3.5.2 eine deklarative Modellierungssprache für Relationserkennungsmodelle und in Abschnitt 3.5.3 noch verschiedene Anwendungen im Kontext von Sentiment- und Emotionsanalyse.

Abschnitt 4 fasst die Beiträge dieser Arbeit zusammen und gibt einen kurzen Ausblick.

#### 3.1 Monolinguale Sentimentanalyse

##### 3.1.1 Sentimentklassifikation

Sentimentklassifikation ist vielfältig bearbeitet worden, so dass wir den Schwerpunkt in dieser Habilitationsarbeit nicht auf diese Problemstellung legen. Allerdings stellte sich die Situation mit dem Aufkommen von neuen Deep-Learning Verfahren so dar, dass verschiedene Korpora etabliert wa-

ren, welche zur Evaluierung von Sentimentklassifikationsverfahren genutzt wurden. Weiterhin wurde aber von jeweiligen Autoren neuer Modelle nur eine Auswahl dieser Korpora genutzt, so dass eine Vergleichbarkeit nur eingeschränkt gegeben war. Tabelle 2 auf Seite 8 zeigt, wie spärlich die Methoden-Korpusmatrix besetzt war.

Auf Basis dieses Standes der Technik reimplementierten wir eine Auswahl von Verfahren, welche vielversprechende Ergebnisse liefern, und zwar die Durchschnittsbildung der Einbettungen (Faruqui et al., 2015), Retrofitting (Faruqui et al., 2015), die Optimierung der Einbettungen zur Repräsentation von Sentiment gemeinsam mit syntaktischer und genereller semantischer Information (Tang et al., 2014), sowie LSTM, BiLSTM, CNN Modelle (Barnes et al., 2017)<sup>11</sup>. Die zu untersuchende Forschungsfrage war, ob bestimmte Verfahren auf bestimmten Korpora durchgehend bessere Ergebnisse liefern. Des Weiteren betrachten wir den Einfluss der Einbettungsdimension.

Im Durchschnitt zeigen BiLSTM-Modelle die beste Performanz. Ein weiteres Ergebnis ist, dass Modelle mit höherer Einbettungsdimension in der Regel zu besseren Ergebnissen führen. Im Gegensatz zu den zum Beispiel recht häufig genutzten 300-dimensionalen Einbettungen führen 600 Dimensionen durchgehend zu besseren Ergebnissen. Wir schließen daraus, dass eine Erstellung von vortrainierten Einbettungen in Zukunft höhere Dimensionen betrachten sollte. Für Details verweisen wir auf Tabelle 3 in Barnes et al. (2017).

Neben der Untersuchung von existierenden Verfahren auf verschiedenen Korpora ergibt sich daraus auch direkt die Frage, ob ein Modell insbesondere bei stark unterschiedlichen Domänen außerhalb der Anwendungsdomäne geschätzt werden kann, oder ob dies zu einem prohibitiven Performanzverlust führt (Barnes et al., 2018b).<sup>12</sup> Dieses sogenannte Problem der Domänenadaptation wird in den letzten Jahren insbesondere mit Transferlernen bearbeitet, in deren Klasse auch unser Verfahren eingeordnet werden kann, welches gemeinsam die Wortähnlichkeit zwischen bekanntermaßen vergleichbaren Worten in der Quelldomäne

<sup>11</sup>Mein Anteil an dieser Arbeit war die Motivation der Studie, das Studiendesign sowie in Teilen die Auswahl der Daten und Methoden. Die praktische Durchführung der Experimente lag bei Jeremy Barnes.

<sup>12</sup>Ich habe die Forschungsfrage definiert, vorgeschlagen die entwickelte Methode zur Domänenadaptation zu verwenden, war bei der Entwicklung der Methode maßgeblich beteiligt und habe die Experimente mitgestaltet.

und der Zieldomäne so weit wie möglich beibehält, aber dabei die Wortrepräsentation so optimiert, dass Polaritäten (also positive und negativ konnotierte Worte) unterscheidbar werden.<sup>13</sup> In einigen Domänenkombinationen stellt unser Verfahren einen neuen Stand der Technik dar (Barnes et al., 2018b, Abbildungen 2 und 3). Dies ist insbesondere bei Domänenpaaren der Fall, die von einer höheren Unähnlichkeit geprägt sind.

### 3.1.2 Aspektbasierte Sentimentanalyse

Die meisten Arbeiten zu aspektbasierter Sentimentanalyse nutzen Korpora im Englischen, da die Menge an annotierten Ressourcen in anderen Sprachen nur gering ist. Um diese Situation zu verbessern haben wir zwei weitere Korpora entwickelt. Dies ist zunächst das USAGE-Korpus (Bielefeld University Sentiment Analysis Corpus for German and English), in dem wir evaluierende, subjektive Phrasen inklusive ihrer Polarität im Kontext mit einer Aspektkennung annotierten (Klinger and Cimiano, 2014).<sup>14</sup> Hier wurden Amazon-Beurteilungen als Basis genutzt, welche aus verschiedenen Produktkategorien stammen. Der SCARE-Korpus (The Sentiment Corpus of App Reviews with Fine-grained Annotations in German) folgt derselben Annotationsrichtlinie und enthält deutsche Beurteilungen aus dem Google-Play-App-Store (Sänger et al., 2016).<sup>15</sup> Beispiele aus dem USAGE-Korpus und aus dem SCARE-Korpus finden sich in Tabelle 3. Das USAGE-Korpus und das SCARE-Korpus stellen damit eines der wenigen auf diese Weise annotierten Ressourcen im Deutschen dar (das einzige weitere vergleichbare Korpus ist das MLSA-Korpus, Clematide et al. (2012)). Eine Erweiterung des USAGE-Korpus wurde auch in einem *Shared-Task* verwendet (Ruppenhofer et al., 2014)<sup>16</sup>. USAGE und SCARE-Korpus wurden in verschiedenen weiteren Arbeiten verwendet (Shah et al., 2019; Deng et al., 2016; Sun et al., 2017; Jebbara and Cimiano, 2016; Sonntag and Stede, 2014; Monett and Stolte, 2016; Vilares et al., 2017, i. a.).

Auf den genannten Korpora (Kessler et al., 2010;

<sup>13</sup>Einbettungen neigen durch den ähnlichen Kontext positiver und negativer Worte ohne besondere Behandlung dazu, dass sich solche Begriffe in den Repräsentationen sehr ähnlich sind.

<sup>14</sup>Die gesamte Arbeit wurde von mir durchgeführt. Die Annotation wurde durch zwei Hilfskräfte unterstützt.

<sup>15</sup>Die Annotation wurde durch einen von mir angeleiteten Studenten angeleitet und durchgeführt.

<sup>16</sup>Der Teil des Shared-Tasks der sich mit dieser Aufgabe beschäftigte wurde vollständig durch mich organisiert. Die Datenannotation wurde durch mich angeleitet.



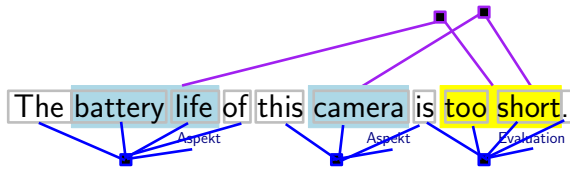


Figure 10: Beispielhaft ausgerollter Faktorgraph des in Klinger and Cimiano (2013a) vorgeschlagenen Modells.

Spina et al., 2012b) war zu dem Entwicklungszeitpunkt die Arbeit von Jakob and Gurevych (2010) der aktuelle Stand der Forschung zur komputationalen Modellierung. Diese verwenden ein Linear-Chain-Conditional-Random-Field (Lafferty et al., 2001) zur Identifikation von Aspekten. In dieser Arbeit wird von der bereits vorhandenen Kenntnis von evaluierenden Phrasen ausgegangen. Die Ergebnisse suggerieren, dass diese Modellierung im Vergleich zur isolierten Erkennung vorteilhaft ist, also die grammatikalische Relation zwischen Aspekten und evaluierenden Phrasen hilfreich ist.

Im Vergleich zu dieser Arbeit entwickelten wir ein komplexeres probabilistisches graphisches Modell, welches nicht nur lineare Abhängigkeiten direkt aufeinanderfolgender vorherzusagender Variablen in der Wortsequenz betrachten kann. Stattdessen werden auch Relationen zwischen Aspekten und evaluierenden Phrasen sowie deren Polarität explizit über größere Entfernungen modelliert (Klinger and Cimiano, 2013a). Diese Erweiterung über eine lineare Kette hinaus ist notwendig, um zum einen entfernte Zusammenhänge, die über die Beschränkung der Markovbedingung hinaus gehen, zu betrachten. Zum anderen werden so auch Eigenschaften innerhalb einer Entität messbar, wie zum Beispiel den Beginn und das Ende einer evaluierenden Phrase oder eines Aspekts, unabhängig von der Anzahl der enthaltenen Worte.

Das probabilistische graphische Modell basiert auf zwei verschiedenen Arten von Faktoren: (1) Ein Faktortyp misst die Güte einer subjektiven evaluierenden Phrase mit ihrer Polarität sowie die Güte einer Aspektphrase. (2) Ein weiterer Faktortyp misst die Wahrscheinlichkeit, dass zwischen einer subjektiven Phrase und einer Aspektphrase eine Relation besteht. Diese Faktoren sind in Abbildung 10 illustriert. Faktor 1 (in violett in der Abbildung dargestellt) nutzt die Zeichenfolge, die Wortklasse und den Kontext als Merkmal, Faktor 2 (blau) den Abstand der Phrasen voneinander sowie Eigenschaften des kürzesten Dependenzpfades. Durch ei-

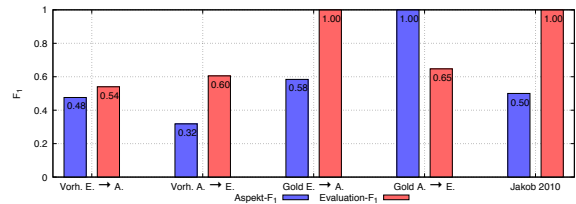


Figure 11: Ergebnisse des in Klinger and Cimiano (2013a) vorgestellten Modells auf dem Auto-Teilkorpus von Kessler et al. (2010).

ne Markov-chain-Monte-Carlo-Inferenz werden sowohl für die Vorhersage der Phrasen sowie der Relationen beide Faktoren genutzt. Hierbei ist entscheidend, dass Aspektphrasen immer zusammen mit einer Relation vorhergesagt werden und in einem Samplingschritt nicht isoliert existieren können.

Diese Inferenz stellt eine Instanz der gemeinsamen Modellierung von Teilproblemen dar. Abbildung 11 zeigt die Ergebnisse für den Teil des Korpus, der sich auf Automobile fokussiert (die Ergebnisse für Kameras sind vergleichbar). Wir sehen, dass die Vorhersage von Aspekten ohne Kenntnis der evaluierenden subjektiven Phrasen bei 32 %  $F_1$  liegt (blaue Balken im zweiten Block). Unter Kenntnis der korrekten evaluierenden Phrasen steigt dieser Wert auf 58 %  $F_1$  (dritter Block). Andersherum ist die Vorhersage von evaluierenden Phrasen ohne Kenntnis von Aspekten mit 54 %  $F_1$  möglich (erster Block), wobei die Kenntnis der annotierten Aspekte 65 %  $F_1$  ermöglicht (vierter Block). Somit ist die gemeinsame Inferenz der Aspektvorhersage und der evaluierenden Phrasen sinnvoll. Wenn statt der korrekten Annotation eine Vorhersage eingesetzt wird, ist der Effekt weniger deutlich, kann aber weiterhin beobachtet werden. Neben dieser Erkenntnis ist das vorgeschlagene Modell besser, als das bisher beste Modell: Der dritte Block in der Abbildung ist bezüglich des experimentellen Designs vergleichbar zu dem fünften Block. Das neue Modell erreicht also 58 %  $F_1$  statt 50 %  $F_1$  von Jakob and Gurevych (2010).

In dieser Arbeit wurde nur die Qualität der Phrasenvorhersagen gemessen, auch wenn die Relation zwischen diesen mitbetrachtet wurde. Die explizite Modellierung und ihr Beitrag wurde in einer darauffolgenden Publikation untersucht (Klinger and Cimiano, 2013b).<sup>17</sup> Hier stellt sich heraus, dass ein zusätzlicher Vorteil in der Aspektvorhersage beobachtet werden kann. Eine gemeinsame Modellierung trägt allerdings nicht positiv zu der Rela-

<sup>17</sup>Die gesamte Arbeit wurde von mir durchgeführt.

tionsvorhersage bei. Diese Arbeiten stellen auch nach aktuellem Forschungsstand eine Ausnahme dar, da nach wie vor die meisten Ansätze von der Erkennung von Aspekten in einem Vorverarbeitungsschritt ausgehen. Wir haben jedoch hier gezeigt, dass es durchaus sinnvoll ist, beide Probleme gemeinsam zu modellieren.

Eine weitere Arbeit zu dem Thema umfasst Vorhersagen auf unserem SCARE-Korpus (Sänger et al., 2017). Die Forschungsfrage war hier, inwieweit aus *Einbettungen* eine Merkmalsmenge erstellt werden kann, welche die Vorhersagequalität in einem einfachen linearen Segmentierungsmodell verbessert.<sup>18</sup> Unsere Ergebnisse zeigen, dass eine Verbesserung mit generellen Wikipedia-Einbettungen nur für subjektive Phrasen beobachtet wird. Um auch eine Verbesserung der Erkennung von Aspektphrasen zu erreichen, mussten domänenspezifische Einbettungen auf Google-Play-Reviews erstellt werden. Dies ist schlüssig, sind Aspekte doch viel eher domänenspezifisch (z.B. „Motor“, „Lautstärke“, „Wasserverbrauch“ bei Waschmaschinen, oder Spielspaß, Graphik bei Spielen), als evaluierende, eine subjektive Meinung ausdrückende Wörter (meistens Adjektive, zum Beispiel „gut“, „schnell“, „zufriedenstellend“). Auch diese Erkenntnis führt dazu, dass zukünftige Arbeiten Aspekterkennung nicht als Vorverarbeitungsschritt behandeln sollten.

### 3.2 Mehrsprachige Sentimentanalyse

In den meisten Sprachen, auch solchen, welche von vielen Sprechern genutzt werden, stehen keine ausreichenden Sentimentanalyseressourcen zur Verfügung. Selbst für die meisten europäischen Sprachen ist die Situation unbefriedigend. Daraus ergibt sich die Frage, inwieweit man die Verwandtschaft von verschiedenen Sprachen nutzen kann, um auf einem Korpus mit verfügbaren Daten ein System zu entwickeln, welches auch auf Daten in einer anderen Sprache entwickelt werden kann. Solche Ansätze nutzen entweder einen Transfer von einer Sprache auf eine andere oder lernen Sprachverwandtschaften gemeinsam mit der Klassifikationsaufgabe zu repräsentieren. Wir diskutieren im Folgenden Verfahren aus beiden methodischen Gruppen.

<sup>18</sup>Hierbei handelt es sich um eine von mir angeleitete studentische Arbeit.

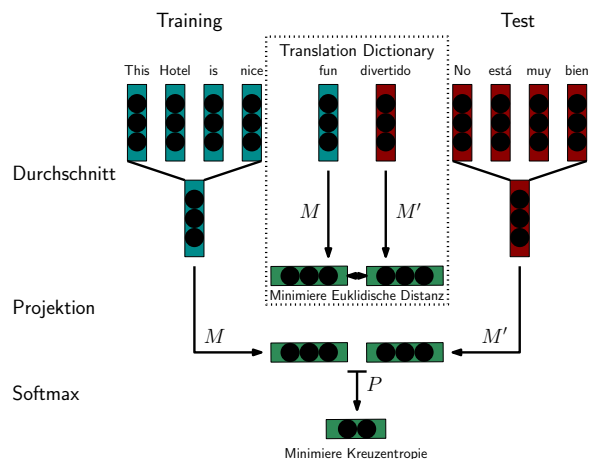


Figure 12: Bilingual Sentiment Embedding Model (BLSE), vorgestellt in Barnes et al. (2018a)

#### 3.2.1 Mehrsprachige Sentimentklassifikation

Das Ziel der mehrsprachigen Sentimentklassifikation ist es, ein annotiertes Korpus einer Sprache  $A$  zu nutzen, um einen Klassifikator zu schätzen, welcher auf unannotierten Daten einer anderen Sprache  $B$  angewendet wird, ohne dass annotierte Daten zur Verbesserung des Systems in der Zielsprache zur Verfügung stehen. Für diese Übertragung können weitere unannotierte Daten  $U_A$  und  $U_B$  zum Einsatz kommen. Die meisten Systeme und Ansätze, um diese Aufgabe zu lösen, haben aber dennoch einen vergleichsweise hohen Bedarf an Ressourcen in der Zielsprache  $B$ , zum Beispiel parallele Korpora der Quell- und Zielsprache oder ein statistisches Übersetzungssystem. Das Ziel unserer Arbeit war vielmehr nun, einen Ansatz zu entwickeln, der vergleichbar gute Ergebnisse liefert und ohne diesen Bedarf auskommt.

Das Verfahren, welches wir in Barnes et al. (2018a) vorstellen (BLSE), benötigt nur eine große Zahl unannotierter Texte in Quell- und Zielsprache (oder vortrainierte Einbettungen) sowie annotierte Daten in der Quellsprache. Die Grundidee, welche auch in Abschnitt 3.1.1 zu tragen kommt, ist die folgende (Abbildung 12 zeigt das Modell graphisch): Auf Basis einer Vektorrepräsentation der Worte in der annotierten Trainingsmenge sowie einem (verhältnismäßig kleinen) Übersetzungswörterbuch sowie der Vektorrepräsentation der korrespondierenden Worte in der Zielsprache (soweit die Übersetzung in dem Wörterbuch vorhanden ist) optimieren wir zwei Projektionsmatrizen  $M$  für die Quellsprache und  $M'$  für die Zielsprache so, dass die Projektion eine Klassifikation mit hoher Qualität erlaubt, so-

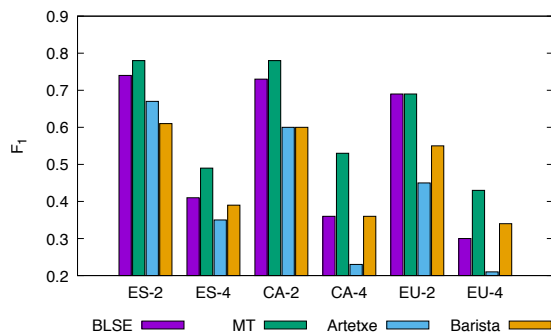


Figure 13: Ergebnisse des BLSE Modells (Abbildung generiert nach Barnes et al., 2018a). Die Zahl hinter dem Sprachkürzel bezeichnet die Anzahl der verwendeten Klassen (sehr positiv, sehr negativ, positiv, negativ vs. positiv, negativ).

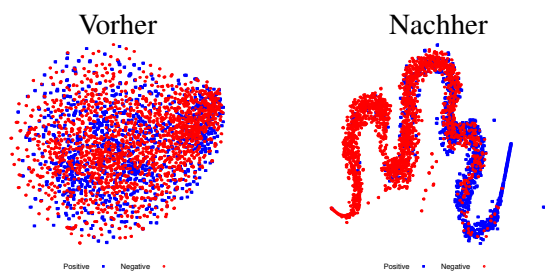


Figure 14: t-SNE-Visualisierung des spanischen Wortvektorraums vor und nach der Projektion mit BLSE. Eine klare Separierung von positiven und negativen Worten ist erkennbar, obwohl keine annotierten Daten für Spanisch genutzt werden (Abbildung generiert nach Barnes et al., 2018a).

wie die Projektionen der korrespondierenden Worte ähnlich zueinander sind. Die Ergebnisse dieses Ansatzes sind in Abbildung 13 dargestellt und werden mit anderen Ansätzen für mehrsprachige Klassifikation verglichen. Wir sehen hier, dass bei unterschiedlichen Korpora ressourcenaufwendige maschinelle Übersetzung immer das beste Ergebnis liefert (MT). Allerdings zeigt BLSE bessere Ergebnisse als ARTETXE (Artetxe et al., 2016, 2017) und BARISTA (Gouws and Søgaard, 2015), welche ähnliche Datenbedarfe haben wie BLSE. Eine qualitative Analyse zeigt auch, dass die entstehenden projizierten Einbettungsräume der Zielsprache Sentiment gut repräsentieren, obwohl keine Trainingsdaten in der Zielsprache verwendet wurden (Abbildung 14).

Mit diesem Verfahren ist es nun mit besserer Performanz als bisher möglich, ohne jegliche annotierten Daten einer ressourcenschwachen Zielsprache ein kompetitives Sentimentklassifikationsmodell zu schätzen.

### 3.2.2 Mehrsprachige Aspektbasierte Sentimentanalyse

In Abschnitt 3.1.1 haben wir monolinguale Sentimentklassifikation vorgestellt, welche wir in Abschnitt 3.2.1 um eine multilinguale Komponente erweitert haben. Dieser Abschnitt stellt nun eine vergleichbare multilinguale Erweiterung der aspektbezogenen Sentimentanalyse dar, wie sie in Abschnitt 3.1.2 vorgestellt wurde.

In dem Klassifikationsfall ist der Verlust der Information, welche Worte welche Information tragen sowie die Reihenfolge und grammatikalische Struktur bei dem Wechsel der Sprache, akzeptabel. Somit kann, bei entsprechender Verwendung von Modellen, welche sich nicht auf diese Charakteristika verlassen, bei einer Übersetzung oder Projektion auf diese Aspekte verzichtet werden. Dies ist zum Beispiel bei BLSE der Fall, welches einen Maximum-Entropy-Klassifikator nutzt.

Diese Informationen sind bei Segmentierungs- und Relationsvorhersageaufgaben wie der Aspekterkennung, der Erkennung von evaluierenden Phrasen sowie der Relation zwischen diesen aber unbedingt notwendig und müssen daher zumindest in einem gewissen Maße erhalten bleiben. Wir schlagen zwei grundlegend verschiedene Modelle vor, die dieser Bedingung genügen und überprüfen, ob diese zusätzliche Komplexität zu einer Verringerung der Performanz führt.

**Statistische Übersetzung, Projektion und Instanzselektion.** Der erste vorgeschlagene Ansatz ist grundsätzlich modellagnostisch bezüglich des verwendeten Sentimentanalysemodells, da die Übertragung auf der Datenebene durchgeführt wird (Klinger and Cimiano, 2015). Unter der Annahme der Verfügbarkeit eines maschinellen Übersetzungssystems werden die Instanzen zunächst von der Quellsprache in die Zielsprache übersetzt. Des Weiteren werden die Phrasenannotationen mit Hilfe der Alinierung zwischen den Sprachen projiziert, ähnlich, wie es für Wortarten von Yarowsky et al. (2001); Yarowsky and Ngai (2001) und für semantisches und syntaktisches Parsing von Padó and Lapata (2009); Agić et al. (2014) bereits vorgeschlagen wurde.<sup>19</sup> Die Annotationsprojektion ist in Abbildung 15 beispielhaft dargestellt.

Die Forschungsfrage, die wir hier behandeln ist, ob dieses Vorgehen von Übersetzung und Projektion zu einer qualitativ ausreichenden Ressource in

<sup>19</sup>Die gesamte Arbeit wurde von mir durchgeführt.

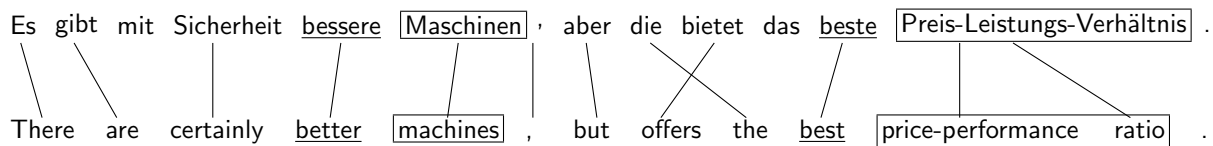


Figure 15: Beispiel für eine Projektion der Annotation eines deutschen Textes auf eine englische maschinell erzeugte Übersetzung. Die Projektion von “die” auf “the” ist ein Fehler, der durch die fehlerhafte Übersetzung ausgelöst wurde. Allerdings wirkt sich dieser hier nicht auf die Projektion der Annotation aus (Klinger and Cimiano, 2015).

der Zielsprache führt. Des Weiteren untersuchen wir, ob ein Entfernen von qualitativ schlechten Übersetzungen oder Projektionen zu einer Verbesserung des Modells führt.

Es stellt sich in Experimenten auf dem USAGE-Korpus, welches in Abschnitt 3.1.2 vorgestellt wurde, heraus, dass Modelle, welche man auf der automatischen Übersetzung gefolgt von einer Projektion trainiert, eine deutlich geringere Vorhersagequalität liefern, als Modelle, welche direkt in der Zielsprache trainiert wurden. Daraus ergibt sich die Frage, wie man die möglicherweise ursächlichen Qualitätsmängel in Übersetzung und Projektion korrigieren kann beziehungsweise Instanzen geringerer Qualität von der Trainingsmenge ausschließen kann. Hierzu diskutieren wir drei Möglichkeiten, welche alle auf der Abschätzung der Übersetzungsqualität basieren. Sie werden mit einer manuellen Beurteilung der Übersetzungsqualität verglichen. Die Verfahren wurden von Specia et al. (2013); Shah and Specia (2014) inspiriert.

Das erste Verfahren nutzt ein Sprachmodell, welches auf einer großen Zahl von Amazon-Bewertungen und dem Werkzeug SRILM (Stolcke, 2002; Stolcke et al., 2011) basiert. Dieses wird dann genutzt um dem Ursprungssatz eine Wahrscheinlichkeit zuzuweisen. Die Intuition ist hier, dass möglicherweise bestimmte Sätze der Quellsprache untypisch oder ungewöhnlich sind und nicht geeignet sind, um ein Modell zu erstellen. In dem zweiten Verfahren wird analog mit Hilfe eines Sprachmodells die Wahrscheinlichkeit des Zielsprachsatzes bewertet. Dieses Verfahren kann intuitiv mögliche Fehler in der Übersetzung aufdecken und ebenfalls wenig typische Instanzen identifizieren. Der dritte Ansatz basiert auf der Alinierung des Originalsatzes mit der Übersetzung und der daraus abzuleitenden Perplexität. Hier wurde FastAlign (Dyer et al., 2013) als Nachverarbeitung zur Abschätzung genutzt, da die Werte direkt aus Google Translate nicht zugreifbar sind.

Abbildung 16 zeigt den Einfluss der Instanzse-

lektion basierend auf den Werten aus diesen drei Ansätzen in einer Domäne des USAGE-Korpus. Hier wurde das Modell eingesetzt, welches in Abschnitt 3.1.2 vorgestellt wurde (Klinger and Cimiano, 2013a). Wir sehen, dass beide Sprachmodelle (für die Zielsprache und die Quellsprache) weniger gute Ergebnisse liefern, als die Alinierungswahrscheinlichkeiten. Dieses Verfahren zeigt allerdings Ergebnisse einer ähnlichen Güte wie die manuelle Selektion qualitativ hochwertiger Instanzen. Bei der Verfügbarkeit eines Übersetzungssystems ist also mit diesem Verfahren durchaus die Erstellung von Modellen in einer Zielsprache möglich, für die keine annotierte Daten vorliegen.

Auf Basis der Bestimmung des Grenzwerts zur Auswahl der Instanzen auf einer unabhängigen Menge und der Anwendung des Systems zur Vorhersage von Aspekten und evaluierenden Phrasen in einer anderen Zielsprache als der Quellsprache, getestet auf dem Sprachpaar Deutsch und Englisch, zeigt sich, dass der Einfluss auf die Aspekterkennung erheblicher ist, als auf die Erkennung von evaluierenden Phrasen. In der Summe ist dieses Verfahren für den Transfer in einer Teilmenge von Sprachpaaren tauglich, um eine akzeptable Qualität auch ohne annotierte Daten in der Zielsprache zur Verfügung zu haben. Allerdings bliebe hier zu zeigen, dass dies auch tatsächlich für eine größere Menge von Sprachpaaren funktioniert.

**Aspektspezifische Übertragung mit Hilfe von Repräsentationsprojektion.** Das in dem vorherigen Abschnitt vorgestellte Verfahren hat den Vorteil, dass die Übertragung bereits auf Datenebene stattfindet. Somit lassen sich für die automatische Klassifikation beliebige maschinelle Lernverfahren nutzen. Der Nachteil ist allerdings, dass diese Verfahren eben gar keinen Zugriff mehr auf die Beziehungen zwischen den verwendeten Sprachen haben. Das Modell, welches in Abschnitt 3.2.1 vorgestellt wurde, modelliert die Sprachverwandtschaft und die Klassifikationsaufgabe simultan. Wir stellen nun vor, wie wir dieses Modell um

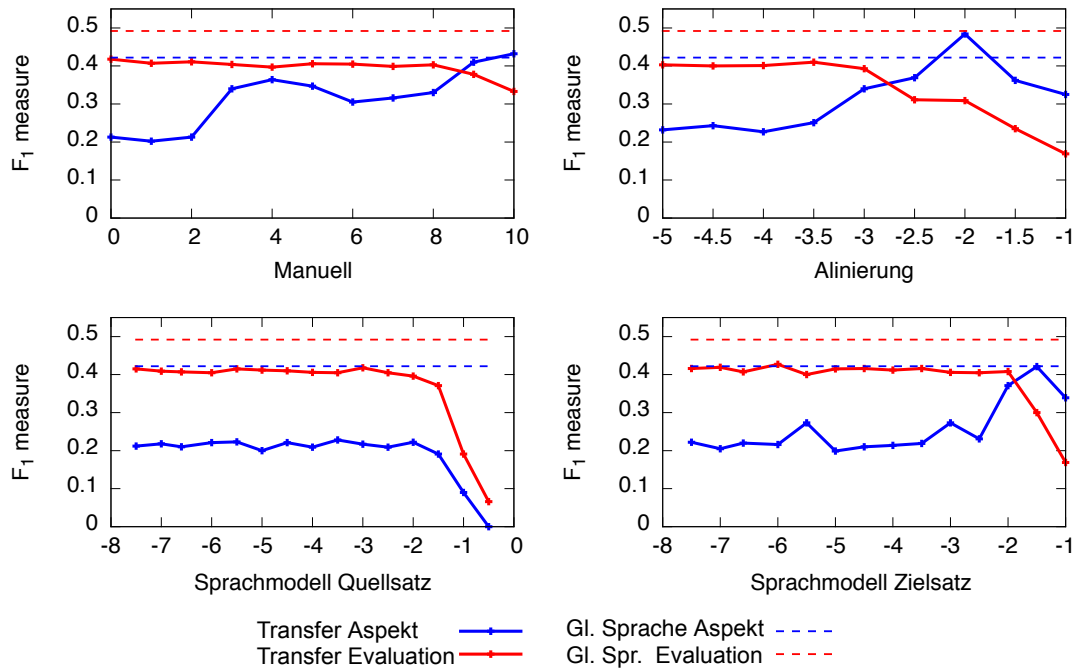


Figure 16: Vollständige Ergebnisse für die Beurteilungen von Kaffeemaschinen für die Projektion von Deutsch nach Englisch (Klinger and Cimiano, 2015).

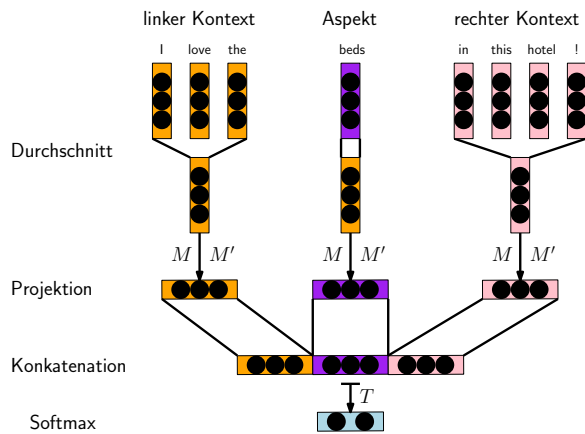


Figure 17: Modell zur aspektbezogenen Sentimentklassifikation (Barnes and Klinger, 2019)

die Möglichkeit, Aspekte einzubeziehen, erweitern (Barnes and Klinger, 2019).<sup>20</sup> Die Klassifikation wird nun auf Basis von drei Teilen durchgeführt, wie in Abbildung 17 gezeigt ist. Die Projektionsmatrizen  $M$  zum Trainingszeitpunkt auf der Quellsprache und  $M'$  zur Vorhersage auf der Zielsprache werden nun dreimal angewendet, und zwar auf den linken Kontext neben einem Aspekt, den Aspekt selbst, sowie den rechten Kontext eines Aspekts. Jede dieser Anwendungen führt zu einer

<sup>20</sup>Ich war hier bei dem Modellentwurf sowie dem Experimentdesign beteiligt. Die konkreten Experimente wurden durch einen von mir betreuten Doktoranden durchgeführt.

Repräsentation des jeweiligen Satzteils. Die Konkatenation dieser Teile wird dann entsprechend des Sentiments klassifiziert. In dieser Studie nehmen wir an, dass die Bestimmung des Aspekts durch ein Orakel erfolgt. Zukünftige Arbeiten werden die Integration dieser Vorhersage in die Klassifikation behandeln (siehe Abschnitt 3.1.2).

Die Ergebnisse in Tabelle 6 zeigen auf verschiedenen Korpora sowie mit einer unterschiedlichen Granularität der vorherzusagenden Klassen (*positiv, negativ* vs. *sehr negativ, negativ, positiv, sehr positiv*), dass das aspektbezogene BLSE bei binärer Vorhersage besser, als bestehende Ansätze funktioniert. Der Ansatz MUSE (Lample et al., 2018) funktioniert allerdings bei vier vorherzusagenden Klassen deutlich besser. Eine umfangreichere Variante der Ergebnistabelle in (Barnes and Klinger, 2019, Tabelle 7) zeigt zusätzlich noch den Vergleich zu einer rein satzbasierten Klassifikation und der Isolation von Aspekt und Kontext. Es zeigt sich, dass Kontext wie auch Aspekt in Isolation nicht ausreichen, um die Vorhersage einer Polarität vorzunehmen. Die Interaktion beider Teile ist notwendig, wie man bei dem Performanzverlust von durchgängig zwischen 10 und 20 Prozentpunkten sehen kann. Im Ergebnis ist unser Verfahren geeignet, um auch aspektbasiert Sentimentanalyse auf Sprachen ohne annotierte Daten zu übertragen.

Allerdings ist der Vorsprung von ressourcenhungriger maschineller Übersetzung hier stärker ausgeprägt, als bei der reinen Sentimentklassifikation auf Sätzen im Ganzen.

### 3.3 Monolinguale Emotionsanalyse

Die bisherigen Abschnitte schilderten die Erweiterung von monolingualer Sentimentklassifikation hin zu multilingualer Sentimentklassifikation, strukturierter monolingualer Modellierung sowie zu strukturierter multilingualer Modellierung. Hierbei sind die involvierten Polaritätswerte weitestgehend auf *positiv* und *negativ*, optional mit unterschiedlichen Abstufungen (*sehr positiv* und *sehr negativ*), verblieben. Wir erweitern nun im Folgenden diese Ansätze um mehr als diese Polaritätsklassen, zunächst im Sinne von Klassifikationsaufgaben (in Abschnitt 3.3.1), bei denen die Klassen weiterhin einer Textinstanz als Ganzes zugewiesen werden. In Abschnitt 3.3.2 wird die Emotionsanalyse in die strukturierte Modellierung überführt.

#### 3.3.1 Emotionsklassifikation und -Regression

**Emotionskorpora und Modelltransfer.** Sentimentanalyse hat sich als eine Standardaufgabe zur Evaluation von Textklassifikationsmodellen entwickelt und so sind auch viele Beiträge von Wissenschaftlern zu dem Feld entstanden, die sich im Kern nicht mit dem Phänomen, sondern mit der Methode beschäftigen. Um diesen Vorteil auch der Emotionsanalyse zukommen zu lassen, haben wir die verschiedenen verfügbaren Emotionskorpora zusammengetragen und in einem angeglichenen Format verfügbar gemacht (Bostan and Klinger, 2018).<sup>21</sup> Diese Vorbereitung für die Arbeiten anderer Wissenschaftler erleichtert die Beantwortung der Fragestellung, welche Domänen, Annotationsverfahren und Emotionen verwendet werden und die Beantwortung der Forschungsfrage, inwieweit eine Übertragung eines Klassifikators von einem Korpus auf einen anderen möglich ist. Tabelle 7 stellt eine Übersicht der Korpora dar. Zu beachten ist hierbei, dass es sich zumindest zum Teil nicht um eine triviale Dateiformatkonvertierung handelt, da verschiedene Namen für Emotionen in den verschiedenen Daten aneinander angeglichen werden mussten. Tabelle 4 im Anhang von Bostan and Klinger (2018) stellt diese Aggregation im Detail dar.

<sup>21</sup>Die Arbeit wurde von einer von mir angeleiteten Doktorandin durchgeführt.



Amanda  
@Euringer

Folgen

That moment when Canadians realised global warming doesn't equal a tropical vacation #BCwildfire #Canadaburns #globalwarming

Original (Englisch) übersetzen

17:59 - 7. Juli 2015

Figure 18: Beispiel eines Tweets aus Schuff et al. (2017). Er wurde von Mohammad et al. (2017) als *negativ* annotiert. Je eine von drei unserer Annotatorinnen markierte hier *Wut*, *Traurigkeit* und *Überraschung*.

In ersten Experimenten zu der Übertragbarkeit von Modellen von einem Korpus auf einen anderen Korpus konnten wir zeigen, dass die Performanz auf den Zieldomänen tatsächlich sinkt. Zum aktuellen Stand der Wissenschaft ist aber noch unklar, ob Domänenadaptationsverfahren hier helfen würden (Bostan and Klinger, 2018, Abbildungen 1 und 2). Allerdings wird das Repository mit den Daten aktuell von 75 Personen beobachtet, so dass wir optimistisch sind, dass weitere Ansätze zu diesem Thema in Entwicklung sind.

Zu dieser Menge von Korpora haben wir die Ressource SSEC (Stance Sentiment Emotion Corpus) beigetragen, welche auch in der Tabelle gelistet ist (Schuff et al., 2017).<sup>22</sup> Hier haben wir uns als Ziel gesetzt, eine frei verfügbare Ressource zu schaffen, bei der alle Annotationen aller Annotatorinnen verfügbar sind. Wir stellten uns auch die Forschungsfrage, welche Adjudikationsstrategie zu einer guten Modellierbarkeit führt und welche Modellarchitektur am besten geeignet ist. Gerade bei der Annotation von subjektiven Phänomenen ist der Grad der Übereinstimmung der verschiedenen Bewertungen verhältnismäßig niedrig.

Ein Beispiel aus diesem Korpus ist in Abbildung 18 dargestellt. Die Ergebnisse der Anwendung einer Reihe von Klassifikationsmethoden zeigt, dass Deep-Learning (LSTM, BiLSTM, CNN, mit Einbettungen) etwas bessere Ergebnisse liefert, als andere Verfahren (SVM, MaxEnt, welche nur auf einem Bag-of-Words basieren). Dies ist insbesondere auf eine deutlich erhöhte Vollständigkeit zurückzuführen, allerdings sinkt die Genauigkeit (Schuff et al., 2017, Tabelle 6). Je nach Anwendungsfall dürfte aber dennoch das Deep-Learning-Verfahren zu bevorzugen sein.

<sup>22</sup>Die Arbeit wurde von mir mit angeleitet. Ein Teil der Experimente und Analysen wurden von mir durchgeführt.

		EN-ES OpeNER	EN-CA MultiBooked	EN-EU	EN-ES SemEval	EN-DE USAGE	Average	
Binary	SPLIT	BLSE	<b>66.8</b>	<b>69.8</b>	<b>66.3</b>	<b>62.2</b>	50.0	<b>63.0</b>
		VECMAP	65.8	64.4	65.1	60.0	39.9	59.0
		MUSE	58.3	64.3	50.2	59.8	<b>57.0</b>	57.9
		BARISTA	61.9	59.0	56.1	44.5	35.3	51.4
		MT	67.3	77.8	74.8	73.2	69.4	72.5
		UNSUP	71.6	73.5	64.0	<b>77.1</b>	-	-
Multiclass	SPLIT	BLSE	18.5	14.3	15.7	<b>40.6</b>	29.5	23.7
		VECMAP	29.2	30.9	28.0	38.9	27.9	31.0
		MUSE	<b>32.9</b>	<b>33.5</b>	<b>27.3</b>	27.4	39.7	<b>32.2</b>
		BARISTA	27.9	35.1	27.3	27.4	33.4	28.1
		MT	24.7	29.2	27.0	33.8	33.2	29.6
		UNSUP	28.9	26.9	23.9	31.7	-	-

Table 6: Ergebnisse des aspektbezogenen BLSE-Modells. Das beste projektionsbasierte Verfahren je Spalte wird mit einer *blauen Box* markiert. Die beste Methode wird mit einer *grünen Box markiert* (Barnes and Klinger, 2019).

Korpus	Granularität	Klassen	Größe	Thema	Referenz
AffectiveText	Überschriften	E + V	1,250	Nachrichten	Strapparava 2007
Blogs	Sätze	E + ne + me	5,025	Blogs	Aman 2007
CrowdFlower	Tweets	E + CF	40,000	Generell	Crowdflower 2016
DailyDialogs	Dialog	E	13,118	mehrere	Li 2017
Electoral-Tweets	tweets	P	4,058	Wahlen	Mohammad (2015)
EmoBank	Sätze	V+A+D	10,548	mehrere	Buechel 2017
EmoInt	Tweets	E – DS	7,097	Generell	Mohammad 2017
Emotion-Stimulus	Sätze	E + shame	2,414	Generell	Ghazi 2015
FB-Val.-arous.	Facebook	V+A	2,895	Fragebogen	Preoțiu 2016
Grounded-Emot.	Tweets	HS	2,585	Wetter/Ereignisse	Liu 2017
ISEAR	Beschreibungen	E + SG	7,665	Ereignisse	Scherer 1994
Tales	Sätze	E	15,302	Märchen	Ovesdotter 2005
SSEC	Tweets	P	4,868	Generell	Schuff 2017
TEC	Tweets	E ±S	21,051	Generell	Mohammad 2012

Table 7: Verfügbare Emotionskorpora, wie sie von Bostan and Klinger (2018) ermittelt und aggregiert wurden. [E] Ekman: Wut, Ekel, Angst, Freude, Traurigkeit, Überraschung, [P] Plutchik: Wut, Ekel, Angst, Freude, Traurigkeit, Überraschung, Vertrauen, Antizipation, [CF] Enthusiasmus, Spaß, Hass, Neutral, Liebe, Langeweile, Entspannung, leer, [DS] Ekel, Überraschung, [JS] Freude, Traurigkeit, [V] Valenz, [A] Arousal, [D] Dominanz, [SG] Scham, Schuld, [±S] Positive Überraschung, Negative Überraschung, [ne] keine Emotion [me] gemischte Emotion.

Des Weiteren konnten wir feststellen, dass eine Adjudikation durch eine Veroderung aller Annotationen zu einem besseren Ergebnis führt, als ein Majoritätsvotum oder eine Verundung. Dies ist vermutlich zum einen der Tatsache geschuldet, dass eine Veroderung zu mehr Beispielen je Klasse führt. Dennoch zeigt dies auch, dass die vermeintliche Erhöhung des Rauschens in den Annotationen nicht zu einer Modellierung mit geringerer Performanz führt. Somit ist dies ein Indiz, dass bei Emotionsannotation die Modellierung in dem Sinne von hoher Abdeckung sinnvoller ist, als die Zielsetzung von hoher Genauigkeit.

Eine weitere Ressource, welche wir für die Emotionsklassifikation erstellt haben ist im Rahmen der Organisation eines Shared-Tasks entstanden (Klin-

ger et al., 2018).<sup>23</sup> Diese Daten vereinen den populären Ansatz der Verwendung von schwacher Überwachung in sozialen Medien mit einer Ereigniszentrierung, wie sie in dem ISEAR-Korpus (Scherer and Wallbott, 1997) existiert. Die hier automatisch generierten Daten enthalten weitestgehend keine expliziten Emotionswörter und beschreiben Ereignisse. Dies ist insbesondere relevant, da solche Ereignisbeschreibungen durch das Fehlen von Emotionswörtern herausfordernder für eine generalisierende Emotionsklassifikation sind.

In Cevher et al. (2019) untersuchen wir die Übertragung von Modellen aus den Domänen, in

<sup>23</sup>Ich war der Hauptorganisator dieser Veranstaltung. Die Koautoren haben insbesondere bei der Organisation beigetragen. Die Daten und initialen Modelle, welche den Teilnehmern zur Verfügung gestellt wurden, wurden von mir erstellt.

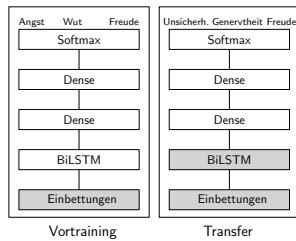


Figure 19: Model zum Transferlernen von Emotionen aus Text. Graue Boxen enthalten eingefrorene Parameter in dem jeweiligen Lernschritt (Cevher et al., 2019).

denen Emotionsklassifikationskorpora verfügbar sind, in den Bereich des Autofahrens und den damit verbundenen Emotionsäußerungen.<sup>24</sup> Es handelt sich somit um eine Anwendung in einer Realweltumgebung in Kombination mit einer bisher im Kontext von Emotionsanalyse kaum verwendeten Methodik. Frühere Arbeiten (Rozental et al., 2018; Chronopoulou et al., 2018), welche Transferlernen im Bereich der Emotionsklassifikation eingesetzt haben, haben vorgelernte Teilmodelle verwendet, welche nicht spezifisch für diese Anwendung angepasst sind, wie zum Beispiel ULMFit (Howard and Ruder, 2018). Eine Ausnahme ist DeepMoji, welches eine große Menge an schwach annotierten Daten nutzt (Felbo et al., 2017). Im Gegensatz dazu untersuchten wir, ob man ein Modell, welches auf einem qualitativ hochwertigen und etablierten Emotionskorpora erstellt wurde, auf eine bestimmte Domäne übertragen kann. Hierbei findet auch eine Anpassung des Emotionsinventars statt, da die Konzepte *Unsicherheit*, *Genervtheit* und *Freude* sinnvoller erscheinen als *Angst*, *Wut* und *Freude* (Dittrich and Zepf, 2019).

Das Übertragungsverfahren ist in Abbildung 19 gezeigt. Die Ergebnisse zeigen, dass alle verwendeten Korpora außer ISEAR (Scherer and Wallbott, 1997) in etwa gleich gut geeignet sind, um das Modell vorzutrainieren. Transferlernen führt zu einer Verbesserung von 2–10 Prozentpunkten, je nach verwendeten Korpora (Cevher et al., 2019, Tabelle 6). In der praktischen Anwendung ist also nicht in jedem Fall eine umfangreiche Neuerstellung von Ressourcen notwendig, zumindest sollten existierende Ressourcen mitgenutzt werden.

**Multimodalität.** Da wir uns bei der Emotionserkennung in der Regel auf Ereignisinterpretationen oder Emotionsausdrücke verlassen, die nicht nur

<sup>24</sup>Diese studentische Arbeit wurde durch mich angeleitet, wobei die Datenerhebung zusätzlich durch einen Mitarbeiter bei der Daimler AG unterstützt wurde.



Figure 20: Beispiel für Kurznachrichten auf Twitter mit zusätzlicher Bildinformation (Klinger, 2017).

in Textform vorliegen, besteht grundsätzlich die Frage, ob andere Modalitäten in den verschiedenen Anwendungen einen zusätzlichen Beitrag leisten können, also komplementäre Informationen liefern. Die Tweets, welche in Abbildung 20 zu sehen, zeigen zwei Beispiele. In dem linken Beispiel wird die Information, dass es sich um ein glückliches Ereignis handelt, von Bild und Text gestützt. In dem rechten Bild dagegen ist die Information komplementär.

In allen Arbeiten, welche in diese Habilitationssarbeit eingehen, wurde die Sentiment- und Emotionsvorhersage auf geschriebenen Text beschränkt. Von dieser Regel sind zwei Arbeiten auszunehmen, zum einen Cevher et al. (2019), bei der Audiosignale wie auch Videosignale mit Hilfe von existierenden kommerziellen Werkzeugen ausgewertet wurden. Die Ergebnisse sind durchgehend schlechter, als bei der Analyse einer manuellen Transkription der Daten und der Anwendung von Transferlernen, wie sie im vorherigen Abschnitt vorgestellt wurde. Diese Arbeit ist Teil einer Masterarbeit, bei der weiterhin das (nicht veröffentlichte) Ergebnis entstand, dass auch eine Kombination der Modalitäten keine Verbesserung der Vorhersagequalität mit sich zieht (Cevher et al., 2019).

Eine weitere Untersuchung des Zusammenspiels verschiedener Modalitäten stellt Klinger (2017) dar.<sup>25</sup> Hier wurden Kurznachrichten auf Twitter untersucht, und zwar insbesondere solche, welche nicht nur Text sondern auch Bildinformationen enthalten. Die Forschungsfrage war, inwieweit die Bildinformationen für eine Verbesserung der Emotionsvorhersagequalität genutzt werden können. Für die Bildanalyse wurde hier angenommen, dass sich zwei grundlegend unterschiedliche Arten von Tweets mit Bildern finden lassen: Zum einen solche, welche Text enthalten, zum anderen

<sup>25</sup>Diese Arbeit wurde durch mich alleine durchgeführt.



solche, auf denen tatsächliche graphische Inhalte vorhanden sind, wie zum Beispiel Fotos. Zwei Beispiele für solche Tweets finden sich in Abbildung 20. Diese Beispiele illustrieren auch die verwendeten Methoden: Zum einen wurde ein Verfahren zur automatischen Generierung von Bildunterschriften eingesetzt (Krizhevsky et al., 2012; Vinyals et al., 2015; Karpathy and Fei-Fei, 2015), zum anderen wurde mit Hilfe von Optical Character Recognition der Textinhalt extrahiert (Smith, 2007). In einer (unveröffentlichten) Masterarbeit haben wir darüber hinaus untersucht, ob Klassifikationsmodelle, welche Gesichter erkennen, deren Emotion erkennen, sowie Objekte und Szenen identifizieren, einen zusätzlichen Beitrag liefern. Dies ist für Objektklassifikation der Fall.

Die Ergebnisse (Klinger, 2017, Tab. 2) zeigen, dass die Qualität der Klassifikation auf Basis des Textes nicht mit Hilfe der Information der anderen beiden Modalitäten erreicht werden kann. Während die OCR-Information isoliert zumindest nur zu einem Verlust von 18 Prozentpunkten im  $F_1$ -Wert führt, zeigen die Bildunterschriften keinen relevanten Informationsgehalt. Eine Kombination der Merkmale führt nicht zu einer Verbesserung im Vergleich zur alleinigen Nutzung von Textmerkmalen. Dies zeigt, dass in dieser Modellierung die Informationen der verschiedenen Modalitäten nicht komplementär zueinander sind. Die Information im Text ist zur Klassifikation ausreichend.

**Emotionsintensitäten.** Neben der Zuweisung von Emotionsklassen kann auch die weitergehende Untersuchung der jeweils durch einen Autor ausgedrückten Intensitäten relevant sein. Zum Beispiel könnte eine Aggregation aller Tweets, welches ein bestimmtes Thema behandeln, durch Durchschnittsbildung diese Intensität mitnutzen. Zum anderen ist in Annotationsaufgaben von Emotionen erkennbar, dass intensivere Emotionsäußerungen eine höhere Übereinstimmung zwischen Annotatorinnen nach sich ziehen. Dies ist ein Indiz dafür, dass die Erkennung von intensiveren Emotionen eventuell auch von höherer Konfidenz des automatischen Systems geprägt ist (diese Hypothese muss allerdings noch untersucht werden).

Die Aufgabe der Intensitätserkennung wurde erstmals von Mohammad and Bravo-Marquez (2017) als Regressionsproblem formuliert und in einem Shared-Task realisiert. Wir untersuchten den Einfluss von automatisch generierten Wörterbüchern, welche Intensitätswerte und Ab-

straktheitswerte sowie den Zusammenhang mit verschiedenen Emotionen darstellen (Köper et al., 2017).<sup>26</sup> Hierzu realisierten wir diese Listen als Eingabe in einem neuronalen Netz, welches zusätzlich wortbasierte Informationen nutzt. Unsere Hypothese, dass solche Wortlisten einen positiven Beitrag leisten, konnte bestätigt werden, allerdings ist der absolute Performanzgewinn limitiert.

Eine weitergehende Forschungsfrage war, inwieweit die Wahrscheinlichkeit des Auftretens verschiedener Emotionen durch den Ausdruck einer intensivierten, abgeschwächten oder negierten bestimmten Emotion verändert wird (Strohm and Klinger, 2018, Ergebnisse in Abb. 5). Weitergehend wurde darauf aufbauend untersucht, wie die Repräsentation von modifizierten Emotionsworten verbessert werden kann (Bostan and Klinger, 2019, Ergebnisse in Tab. 3).<sup>27</sup> Wir konnten hier zeigen, dass eine explizite Repräsentation von Phrasen wie “nicht glücklich” oder “ein wenig traurig” in den Einbettungen zu einer Verbesserung der Intensitätsvorhersage bei Instanzen führt, die solche Formulierungen enthalten. Die Eigenschaft des verwendeten CNN-LSTM, auch Sequenzinformationen zu repräsentieren, scheint also nicht auszureichen. Wir konnten also zeigen, dass die Kompositionalität solcher Ausdrücke mit modelliert werden muss.

**Anwendungen.** Eine besondere Anwendung der Emotionsanalyse findet sich in den digitalen Geisteswissenschaften und hier insbesondere in der komputationellen Literaturanalyse. Einen Überblick über die Verwendung von Sentiment- und Emotionsanalyseverfahren in diesem Feld haben wir in Kim and Klinger (2019a) zusammengestellt.<sup>28</sup>

In Kim et al. (2017b,a) haben wir untersucht, inwieweit ein Zusammenhang zwischen dem Genre fiktionaler Texte und dem Emotionsverlauf des Textes beobachtet werden kann.<sup>29</sup> Unsere Annahme war, dass der Emotionsgehalt ein prädiktives Merkmal zur Genreklassifikation darstellt. Hier

<sup>26</sup>Diese Arbeit wurde durch zwei von mir in diesem Projekt betreuten Doktoranden durchgeführt.

<sup>27</sup>Diese Arbeiten wurden von einem von mir betreuten Studenten und einer von mir angeleiteten Doktoranden durchgeführt.

<sup>28</sup>Der Artikel wurde von mir und einem von mir betreuten Doktoranden etwa gleichberechtigt verfasst. Die Literaturrecherche oblag hauptsächlich dem Studenten, während ich die Struktur und den Aufbau verantwortete.

<sup>29</sup>Die Arbeit wurde von mir mitbetreut. Die wissenschaftliche Fragestellung wurde von mir definiert.

wurden drei Modelle zur Vorhersage des Genres verglichen, zum einen ein konvolutionales neuronales Netzwerk, welches nur auf die durchschnittlichen Emotionswerte im zeitlichen Verlauf zugreifen konnte, und zwei wortbasierte Modelle, eines welches im Sinne der Stilometrie die häufigsten Worte nutzt und eines, welches emotionstragende Worte nutzt. Während die beiden wortbasierten Modelle ähnliche Ergebnisse lieferten ( $\approx 0.8 F_1$ ), zeigt der Emotionsverlauf als Informationsquelle weniger gute Ergebnisse ( $\approx 0.59 F_1$ ). Dennoch liegen diese weit über einem zufälligen Ergebnis ( $0.12 F_1$ ). Details zu den Ergebnissen finden sich in (Kim et al., 2017a, Tab. 2). Es ist also festzuhalten, dass Emotion tatsächlich mit der Klasse des Genres eines Texts interagiert.

In Ehrlicher et al. (2019) haben wir untersucht, inwieweit der Emotionsgehalt von spanischen Reiseberichten im zeitlichen Verlauf der Jahrhunderte einer Veränderung unterliegt.<sup>30</sup> Dies geschah hauptsächlich auf einer Annotationsstudie, da die erstellten Daten nicht umfangreich genug sind, um einen detaillierten Klassifikator zu erstellen. Ein Modell, welches lediglich nichtemotionale von emotionalen Sätzen unterscheiden kann, konnte mit akzeptabler Performanz geschätzt werden. In zukünftigen Arbeiten werden hier mehr Daten erzeugt werden müssen, um Klassifikationsmodelle für die verschiedenen Emotionen zu entwickeln. Auf Basis der annotierten Daten konnten wir aber bereits erkennen, dass der Emotionsgehalt solcher Reiseberichte im 19. Jahrhundert mit 34 % höher liegt, als im 18. Jahrhundert (29 %). Dies ist insbesondere einem Zuwachs an Wut, Freude und Überraschung zuzurechnen, wobei Angst als einzige Emotion abnimmt.

Des Weiteren haben wir zwei Arbeiten durchgeführt, die die Nutzbarkeit der Ergebnisse von Emotionserkennungssystemen durch Domänenexperten als Ziel hatten. Eine Analyse der Emotionsverläufe in Franz Kafkas Texten „Das Schloss“ und „Amerika“ haben wir in Klinger et al. (2016) durchgeführt.<sup>31</sup> Mit einem ähnlich Schwerpunkt entwickelten wir ein Visualisierungswerkzeug, welches soziale Netzwerke mit Emotionsanalysen vereint (Barth et al., 2018).<sup>32</sup>

<sup>30</sup>Ich habe hier die Annotationsumgebung verwaltet sowie das Klassifikationsmodell auf den Daten geschätzt.

<sup>31</sup>Hierbei handelte es sich um eine von mir betreute Studienarbeit.

<sup>32</sup>Die Arbeit wurde durch einen Doktoranden und einen Studenten durchgeführt und von mir betreut.

### 3.3.2 Strukturierte Emotionsmodellierung

Wie in Abschnitt 2.4.4 besprochen, existierte nahezu kein Korpus, welches die genannten Rollen, welche in der Emotionsanalyse relevant sind, annotiert enthalten. Diese Situation verbessern wir in Kim and Klinger (2018).<sup>33</sup> Zu beantworten war die Frage, ob eine Annotation dieser Rollen in Literatur mit zufriedenstellender Qualität möglich ist und ob diese auch zu performanten Modellen führen kann.

Im Kontext des Digital Humanities Centers CRETA<sup>34</sup> haben wir ein Literaturkorpus erstellt, bei dem emotionale Ausdrücke in einem Satz im Kontext von einem davorstehenden und darauffolgenden Satz mit den genannten Rollen annotiert wurden. Hierbei wurde zwischen dem Ziel und der Ursache zwischen auftretenden Figuren und Ereignissen unterschieden. Diese Arbeit führte zu der Resource, welche in Tabellen 8 und 9 zusammengefasst wird. Die Übereinstimmung zwischen Annotatoren ist insbesondere bei den Emotionen eher gering (mit Werten bis zu 55  $F_1$ , wobei ein Überlapp eines Tokens bereits als korrekte Übereinstimmung zählt). Dieses Ergebnis ist allerdings im Kontext unserer vorherigen Arbeiten nicht überraschend (siehe Schuff et al. (2017) und Abschnitt 3.3.1). Die Übereinstimmung bei Figurnennungen ist mit bis zu 68  $F_1$  besser.

In einer Modellierungsstudie haben wir untersucht, inwieweit eine gemeinsame Repräsentation hilfreich ist, und ob die Kenntnis eines Teilnehmers an einer Relation einen positiven Einfluss auf andere Teilnehmer hat. Dies ist der Fall: Die Erkennung der Emotionsworte profitiert um 9 Prozentpunkte von der Information, welche Wörter den Fühlenden beschreiben. Die Erkennung der Fühlenden wird sogar um 22 Prozentpunkte besser, wenn die Phrase, welche die Emotion beschreibt, bekannt ist. Details zu diesen Ergebnissen sind in Tabelle 4 in Kim and Klinger (2018) zu sehen. Wir können festhalten, dass ohne eine gemeinsame Repräsentation die Modellierung nahezu unmöglich erscheint. Dennoch sind die Ergebnisse mit  $F_1$ -Werten für die Emotionsworte von 48 %, für Fühlende von 30 % und des Ziels von 6 % nicht zufriedenstellend. Die Ursachenmodellierung führt sogar zu einem  $F_1$ -Wert von 0.

Diese Herausforderung wurde auf zweierlei Wei-

<sup>33</sup>Die Arbeit wurde durch einen von mir angeleiteten Doktoranden durchgeführt.

<sup>34</sup><https://creta.uni-stuttgart.de>

Klasse	Insg.	Adjudik.	Grad			Annotationslänge				
			stark	schwach	neg.	1 Token		≥ 2 Token		
Emotions	Wut	192	156	5	12	7	106	68%	50	32%
	Antizipation	248	201	5	3	11	161	80%	40	20%
	Ekel	242	190	2	7	14	144	76%	46	24%
	Angst	254	183	11	16	17	145	79%	38	21%
	Freude	434	336	31	20	28	289	86%	47	14%
	Traurigkeit	307	224	10	2	13	168	75%	56	25%
	Überraschung	243	196	12	4	7	156	80%	40	20%
	Vertrauen	264	232	3	3	33	191	82%	41	18%
	Andere	432	207	4	4	4	133	64%	41	36%
Entities	Figur	2072	1715				1288	75%	427	25%
	Ereignis	858	615				38	6%	577	94%
	Andere	771	485				114	24%	371	76%

Table 8: Korpusstatistik für Emotionsannotationen. Spalten zeigen die Anzahl jeder Emotionsannotation (Kim and Klinger, 2018).

Relation	Insg.	Adjudik.	Emotion, welche Relation ausl.									Involv. Entitäten		
			Wut	Antizip.	Ekel	Angst	Freude	Andere	Traurigkeit	Überraschung	Vertrauen	Figur	Ereignis	Andere
Fühlende	2113	1717 48%	137	164	130	173	309	210	216	171	207	1704		
Ursache	1261	840 24%	48	45	70	95	174	74	134	125	75	87	398	343
Ziel	1244	1017 28%	106	129	125	96	135	121	62	80	163	444	315	257
$\sum$	4618	3574 77%	291	338	325	364	618	405	412	376	445	2238	717	601

Table 9: Korpusstatistik für Relationsannotationen (Kim and Klinger, 2018).

se bearbeitet: (1) Zum einen ergab sich daraus die Erstellung von weiteren, aktuell in Entwicklung befindlichen Korpora in weiteren Domänen, um die Schwierigkeiten besser abschätzen zu können, sowie mit mehr Daten der Identifikation der textuell realisierten Ursachen von Emotionen auf den Grund zu gehen. Hierbei wird aktuell untersucht, welche Modellierungsansätze am sinnvollsten sind. Dies geschieht im Kontext des DFG-finanzierten Projekts SEAT.<sup>35</sup> (2) Zum anderen haben wir die Aufgabe der Erkennung der Ursache einer Emotion im Kontext des Fühlenden auf eine typisierte Relationserkennung zwischen zwei Figuren im Text reduziert. In dem Beispiel in Abbildung 7 auf Seite 10 wäre also vorherzusagen, dass eine gerichtete Relation zwischen Stefanie und Frieda mit der Emotion *Wut* existiert. Diese Modellierung unterstützt im Folgenden auch die weitergehende Aggregation von literarischen Texten zu mit Emotionsklassen typisierten sozialen Netzwerken. Die Korpuserstellung und Evaluation wurde in Kim and

<sup>35</sup><https://gepris.dfg.de/gepris/projekt/380093645>

Klinger (2019c) publiziert.<sup>36</sup> Der Ablauf des implementierten Systems ist in Abbildung 21 zu sehen. Die wichtigste Forschungsfrage, die hier untersucht wird, ist, welches Relationserkennungsverfahren wie gut geeignet ist. Hierzu werden insbesondere verschiedene Varianten von rekurrenten neuronalen Netzen mit verschiedenen Arten, die Entitäten zu markieren (*positional indicators*), untersucht. Die Ergebnisse sind in Tabelle 4 in Kim and Klinger (2019c) zu sehen, und mit 65 %  $F_1$  zufriedenstellender. Diese vereinfachte Modellierung scheint also einer vielversprechenderen Formulierung zu folgen.

### 3.4 Mehrsprachige Emotionsklassifikation

Wie in Abschnitt 2.4.5 dargelegt, existieren nahezu keine Arbeiten, welche die Übertragung von Emotionsklassifikationssystemen zwischen Sprachen untersuchen. Wir tragen deutsche und englische Korpora bei, welche insbesondere implizite Emotionsäußerungen im Kontext von Ereignisbe-

<sup>36</sup>Die Arbeit wurde von einem von mir angeleiteten Doktoranden durchgeführt.

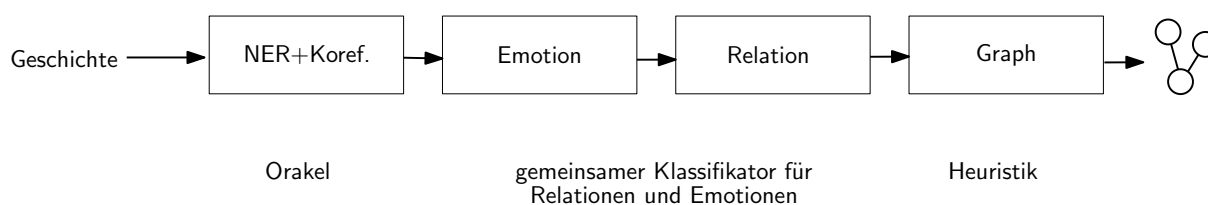


Figure 21: Ablauf der Emotionsanalyse zur Extraktion eines mit Emotionen typisierten sozialen Netzwerks aus fiktionalen Texten (Kim and Klinger, 2019c).

schreibungen enthalten (Troiano et al., 2019).<sup>37</sup> Somit steht diese Forschung im Kontext des Korpus ISEAR (Scherer and Wallbott, 1997) und ist auch verwandt mit dem Shared-Task zu impliziten Emotionsäußerungen (Klinger et al., 2018).

In allen drei Arbeiten (Scherer and Wallbott, 1994; Klinger et al., 2018; Troiano et al., 2019) besteht das Korpus aus Ereignisbeschreibungen der Form „Ich fühle [Emotionsname] als/wenn/dass. . .“ (bzw. I felt [emotion name], if/that/when. . .). Bei Scherer and Wallbott (1994) wurden Fragebögen in verschiedenen Sprachen gesammelt und ins Englische übersetzt. Bei Klinger et al. (2018) wurden Nachrichten mit entsprechenden Phrasen in dem Kurznachrichtendienst Twitter gesammelt. Im Gegensatz dazu haben wir bei Troiano et al. (2019) zum einen untersucht, wie man eine solche Datensammlung qualitativ hochwertig mit Hilfe eines Crowdsourcingdienstes durchführen kann, ob das entstehende Korpus mit ISEAR vergleichbar ist, und ob eine Übertragung des Vorgehens von Englisch nach Deutsch zu Schwierigkeiten führt.

Hierbei wurde ein zweistufiges Verfahren durchgeführt, in Phase 1 wurden Teilnehmer gebeten, die oben genannte Phrase mit einer emotionalen Ereignisbeschreibung zu vervollständigen, was zu Texten der Autor-/Fühlerperspektive führt, welche eine typische Länge von einem Satz haben. In Phase 2 wurde um die Beurteilung durch Lesende/Außenstehende gebeten. Insgesamt konnte eine gute Übereinstimmung beobachtet werden (siehe Abbildung 22). Ein weiterhin interessantes Ergebnis ist, dass englische Beschreibungen erheblich länger sind als deutsche Beschreibungen. Außerdem wird im Englischen der Begriff „Disgust“ weiter gefasst als der Begriff „Ekel“ im Deutschen, so dass das Spektrum der beschriebenen Ereignisse im Deutschen geringer ist. Dies mag die Hypothese motivieren, dass eine Übertragung zwischen

Sprachen schwierig sein kann und sogar eine Konzeptänderung (*concept drift*) mit behandelt werden muss.

Zur Analyse, ob eine Übertragung zwischen Crowdsourcing und Expertenannotation sowie zwischen Sprachen möglich ist, haben wir Bag-of-Words Maximum-Entropy-Klassifikator auf dem englischen original-ISEAR-Korpus sowie der von uns erstellten deutschen (deISEAR) und englischen Variante (enISEAR) trainiert. Hierbei zeigt sich, dass die Ergebnisse sowohl bei der Übertragung von Expertenannotation auf Crowdsourcingannotation (original-ISEAR nach enISEAR) wie auch von deISEAR nach enISEAR mittels maschineller Übersetzung durch Google Translate ähnlich sind (siehe Tabelle 10). Somit scheint bei maschineller Übersetzung von mit Emotionen behafteten Ereignisbeschreibungen der Verlust der emotionalen Konnotation zu vernachlässigen sein.

Weiterhin haben wir (unveröffentlicht) ein Modell auf IEST (Klinger et al., 2018) trainiert und auf enISEAR angewendet. Auch hier zeigt sich kein deutlicher Qualitätsverlust. Somit scheint diese Variante einer schwachen Überwachung für die besondere Form von Emotionskorpora geeignet zu sein. Es sei allerdings zu beachten, dass IEST ein Vielfaches größer ist, als die verschiedenen ISEAR-Korpusvarianten. Dies motiviert zukünftige Arbeiten, die die Zusammenhänge, möglicherweise mit Instanzselektion (siehe Abschnitt 3.2.2) oder Transferlernen (siehe Abschnitte 3.3.1 und 3.2.1), weiter untersuchen. Auch eine Übertragung von IEST auf andere Sprache erscheint naheliegend, da keine manuelle Annotation notwendig war.

### 3.5 Verwandte Phänomene

#### 3.5.1 Ironie, Sarkasmus und Satire

Auf Basis eines der ersten frei verwendbaren Korpora, welcher bezüglich des Auftretens von Ironie und Sarkasmus (ohne eine Grenze zwischen den beiden Konzepten zu ziehen) annotiert wurde (Filatova, 2012), haben wir zunächst die Rele-

<sup>37</sup>Ich habe die Forschungsfrage definiert sowie die Arbeit mit angeleitet. Des Weiteren habe ich die Modellierungsexperimente durchgeführt.

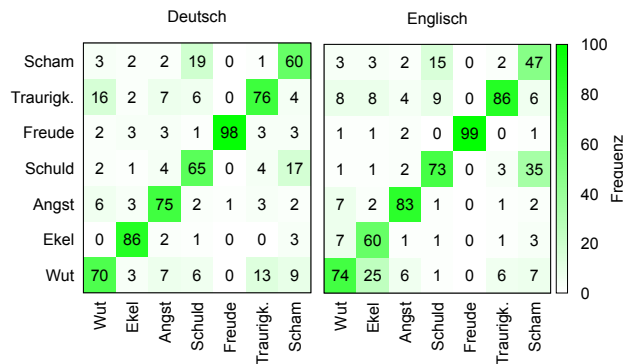


Figure 22: Konfusionsmatrix für die Erstellung der Korpora in Troiano et al. (2019). Spalten: Phase 1, Zeilen: Phase 2 des Annotationsprozesses.

Emotion	deISEAR						enISEAR					
	TP	FP	FN	P	R	F1	TP	FP	FN	P	R	F1
Wut	29	30	114	.49	.20	.29	27	32	116	.46	.19	.27
Ekel	65	57	78	.53	.45	.49	67	85	76	.44	.47	.45
Angst	70	77	73	.48	.49	.48	85	69	58	.55	.59	.57
Schuld	75	140	68	.35	.52	.42	79	161	64	.33	.55	.41
Freude	106	61	37	.63	.74	.68	94	43	49	.69	.66	.67
Traurigkeit	63	31	80	.67	.44	.53	70	29	73	.71	.49	.58
Scham	66	131	77	.34	.46	.39	49	111	94	.31	.34	.32
<i>Micro</i>	474	527	527	.47	.47	.47	471	530	530	.47	.47	.47

Table 10: Klassifikationsergebnisse der Übertragung eines auf ISEAR trainierten Modells (Troiano et al., 2019).

vanz der bis zu dem Studienzeitpunkt publizierten Merkmale untersucht (Buschmeier et al., 2014, Ergebnisse in Tabelle 1).<sup>38</sup> Dies stellt die erstmalige Untersuchung linguistischer Charakteristika wie auch bestimmter Muster bezüglich des Sentiments in diesem Umfang dar. Zusammengefasst ist festzustellen, dass Wortmerkmale einen unschlagbaren Ansatz darstellen, sowie Ironie und Sarkasmus nicht über die Gesamtbeurteilung zwischen Werten 1–5, welche als strukturiertes Metadatum mit Sternchen bereitgestellt sind, gleichverteilt sind. Die Sentimentmerkmale stellen keinen über die Wortinformationen hinausgehenden Beitrag, allerdings ist alleine mit der Veränderung des Sentiment im Verlauf des Texts eine Erkennung von Ironie/Sarkasmus mit 81 %-Genauigkeit möglich. Es ist somit festzuhalten, dass Sentiment und Wortmerkmale vermutlich auch in anderen Domänen ausreichen dürften, um Ironie und Sarkasmus zu erkennen. Dies ist, bezogen auf die Menge von vorgeschlagenen spezifischen Merkmalen zwar ein negatives Ergebnis, allerdings erlaubt die manuelle Integration von Ironie/Sarkasmus-Theorien einen stärkeren Erkenntnisgewinn als reine wortbasierte Merkmale,

<sup>38</sup>Das Forschungsthema wurde von mir vorgeschlagen und der durchführende Student von mir betreut.

so dass auch in Zukunft solche Arbeiten eine Berechtigung haben. Unsere Merkmalsanalyse hat einige darauf aufbauende Arbeiten motiviert, welche weitere Merkmale und ihre Bedeutung untersuchen (Taslioglu and Karagoz, 2017; Mladenović et al., 2017; Hernández-Farías et al., 2015; Joshi et al., 2016; Farías et al., 2016; Poria et al., 2016, *i. a.*).

Die deutlich meisten Arbeiten unterscheiden nicht zwischen Ironie und Sarkasmus, somit bleibt, aus empirischer Sicht, unklar, ob Autoren, zum Beispiel in sozialen Netzwerken diesen Unterschied überhaupt machen und ob dieser mit automatischen Systemen messbar ist (Tsur et al., 2010; Barbieri and Saggion, 2014; Riloff et al., 2013). Dieser Frage sind wir mit Ling and Klinger (2016) nachgegangen.<sup>39</sup>

Wir basieren unsere Analyse auf 99000 englischen Tweets, welche automatisch mit den Hashtags #irony/#ironic und #sarcasm/#sarcastic annotiert wurden. Mit Hilfe eines (zum Entwicklungszeitraum) typischen Modells für Sentimentanalyse separieren wir diese beiden Untertypen von Ironie und führen eine Merkmalsanalyse durch. Beispiele finden sich in Tabelle 11. Tabelle 12 zeigt

<sup>39</sup>Die Arbeit wurde von mir konzipiert und angeleitet.

Klasse	Tweet
sarkastisch	You're kind of wasting my time & energy. Thanks for that. #Sarcastic
sarkastisch	@USERNAM @USERNAME You probably just missed the text. #sarcastic
ironisch	The #rebel flag goes down. The #Cuban #flag goes up. #Confederacy #irony #communism #castro #cubanembassy
ironisch	Get to work to realize oatmeal packet I grabbed was 'Dino-oatmeal' – #monday #irony

Table 11: Beispiele für Tweets welche sarkastisch oder ironisch sind (Ling and Klinger, 2016).

die Klassifikationsergebnisse. Wir sehen in diesen *Accuracy*-Werten auf einem balancierten Korpus, dass die Unterscheidung von Ironie und Sarkasmus schwieriger ist, als Ironie und Sarkasmus gemeinsam von regulären Tweets zu unterscheiden (etwa 10 Prozentpunkte, vergleiche Spalte 2 und 5). Die Erkennung von Ironie und Sarkasmus ist isoliert voneinander etwa gleich herausfordernd (Spalte 3 und 4).

Eine weitergehende Analyse (Ling and Klinger, 2016, Tab. 2) zeigt die Bedeutung der verschiedenen Merkmale in den verschiedenen Klassifikationsaufgaben. Insbesondere unterscheidet sich die Länge der Tweets, die Polarität der Texte sowie die Verwendung von URLs, Hashtags und der Nennung von Personen.

Zusammengefasst ist also die Polarität eines Texts und deren Verlauf ein wichtiges Merkmal für die Erkennung von Ironie und Sarkasmus. Dies ist schlüssig, da Ironie und Sarkasmus häufig eine zu der gemeinten Polarität inverse Polarität ausdrückt. Bei Sarkasmus beobachten wir häufiger positive Ausdrücke, während bei Ironie diese eher gleichverteilt sind. Ironie wird in längeren Texten ausgedrückt als Sarkasmus, da die Beschreibungen situationeller Ironie tendenziell komplizierter zu fassen sind, als kurze sarkastische Kommentare.

Zum Abschluss dieses Abschnitts stellen wir einen methodisch alternativ aufgestellten Ansatz dar, der ebenfalls das Ziel hat einen Einblick in die notwendigen Eigenschaften eines lernenden Modells zu liefern. Im folgenden Fall der Satireklassifikation nutzen wir allerdings keinen merkmalsbasierten

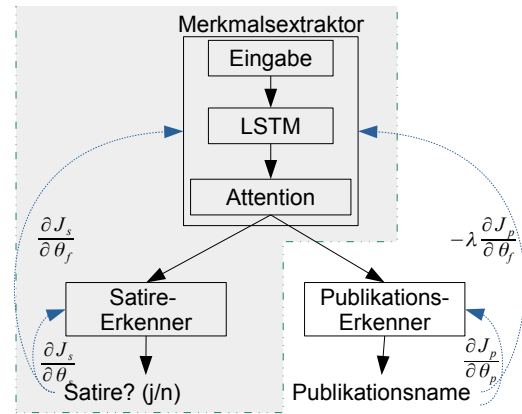


Figure 23: Architektur unseres Satireerkennungsmodells. Der graue Bereich auf der linken Seite zeigt den Satireerkenner, der weiße Bereich auf der rechten Seite ist die adversariale Komponente, welche verhindert, dass Publikationscharakteristika genutzt werden (McHardy et al., 2019).

Ansatz, sondern ein neuronales Modell.

Diese Arbeit basiert auf vorgehenden Arbeiten von Yang et al. (2017); De Sarkar et al. (2018), welche einen ähnlichen Ansatz verfolgen. Allerdings kann bei deren Verfahren nicht ausgeschlossen werden, dass das Modell Charakteristika der Publikationsquelle lernt und *nicht* spezifische Eigenschaften von Satire. Dieses Problem lösen wir mit Hilfe von adversarialem Training wie in Abb. 23 dargestellt. Wir lernen hier mit dem Satire- und dem Publikationserkenner auf Basis einer gemeinsamen Repräsentation. Die Zielfunktionen  $J_P$  (für die Publikationsvorhersage) und  $J_S$  (für die Satirevorhersage) nutzen Parameter  $\theta_P$  und  $\theta_S$ , wobei wir jeweils die Kreuzentropie optimieren. Um die Repräsentation allerdings davon abzuhalten etwas über die Publikationen zu lernen wird der Gradient bei der Optimierung dieser Komponente umgedreht. Dies ist die adversariale Komponente des Modells. Die Publikationsquelle soll ununterscheidbar werden, wobei der Klassifikator weiterhin versucht ein positives Ergebnis zu erreichen. In Abhängigkeit von der Gewichtung  $\lambda$  der adversarialen Komponente sehen wir in Abbildung 24, dass die Publikationserkennungsqualität auf die Güte zufälliger Vorhersage zurückgeht, ohne dass die Satireerkennung substantiell verschlechtert wird. Die Ergebnisse zeigen sich auch in der Analyse des Attention-Mechanismus, wie ein Beispiel in Abb. 25 exemplarisch illustriert. Die rote Unterlegung der Wörter korrespondiert hier mit den Attention-Gewichten. Bei der Verwendung der ad-

Feature Set	Ironie vs. Sarkasmus	Ironie vs. Regulär	Sarkasmus vs. Regulär	Figurativ vs. Regulär
Baseline	0.50	0.50	0.50	0.50
All-BoW	0.64	0.81	0.83	0.82
BoW Unigram	0.76	0.87	0.89	0.87
BoW Uni+Bigram	0.78	0.88	0.90	0.88
All	0.79	0.89	0.90	0.88

Table 12: Ergebnisse des Klassifikationsexperiments Satire vs. Ironie (Ling and Klinger, 2016).

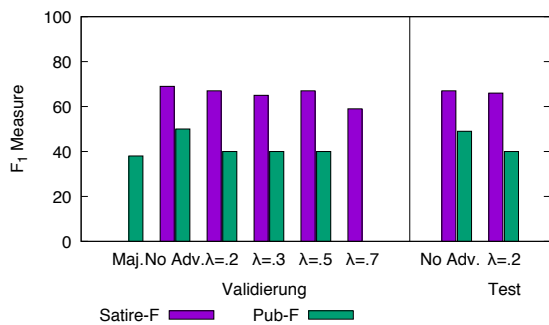


Figure 24: Ergebnisse der adversarialen Satireerkennung.  $\lambda$  stellt die Gewichtung zwischen den Teilen der Zielfunktion dar, bei  $\lambda = 0$  wird die adversariale Komponente nicht zum Training des Merkmalsextraktors genutzt (McHardy et al., 2019).

versarialen Komponente sind diese beiden Teile, welche tatsächlich auf Satire hinweisen, ausgeprägter als wenn diese Komponente nicht genutzt wird, um den Merkmalsextraktor zu korrigieren.

### 3.5.2 Deklarative Relationserkennung

Bisher haben wir uns dem Thema von der Anwendungsseite genähert, wobei wir auch strukturelle Unterschiede in den verschiedenen Aufgaben diskutiert haben. Nun analysieren wir das Problem von der methodischen Seite. Wir stellen fest, dass sich alle bisher vorgestellten Aufgaben entweder als Segmentierung (in Aspekte und subjektive Phrasen sowie Ursachen, Ziele und Fühler einer Emotion) oder  $k$ -äre Relationserkennung formulieren lassen, wobei letztere als das Füllen einer Schablone betrachtet werden kann. Es gibt also eine Schablone eines Sentimentereignisses oder Emotionereignisses, und jede dieser Schablonen wird mit Variablen befüllt, welche (typisierten) Segmenten entsprechen. Diese Formulierung ist weitreichender anwendbar als in der strukturierten Emotions- und Sentimentanalyse, zum Beispiel bei Semantic Role Labeling (Palmer et al., 2005; Pradhan et al.,

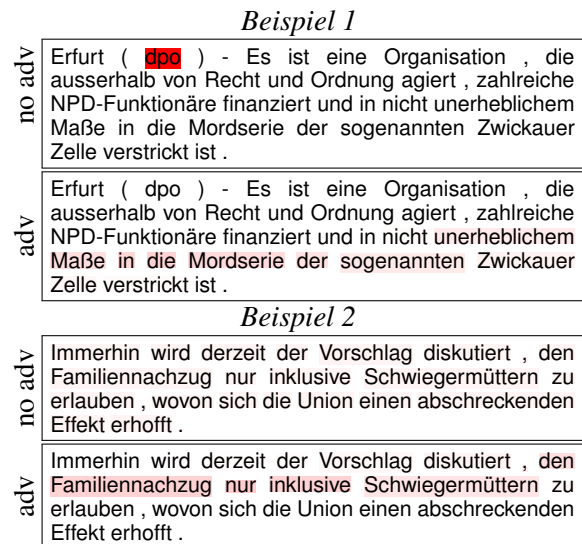


Figure 25: Ein Beispiel der Gewichte des Attention-Mechanismus mit und ohne adversariale Komponente (McHardy et al., 2019).

2007), temporaler Informationsextraktion, bei der Ereignisbeschreibungen sortiert werden (Pustejovsky et al., 2003a,b), der Ereigniserkennung in Rezepten (Abend et al., 2015) oder der biomedizinischen Ereigniserkennung (Nédellec et al., 2013) wie auch der Erkennung von Redner und Redewiedergabe (Scheible et al., 2016).<sup>40</sup>

Mit Adel et al. (2018) haben wir dazu einen Werkzeugkasten vorgestellt, welcher in einer deklarativen Beschreibungssprache die Aufgabe spezifiziert, eine Modellkompilation ermöglicht und im Folgenden verschiedene Modelle zur Verfügung stellen wird, deren Parameter auf jeweiligen annotierten Daten geschätzt werden. Hiermit leisten wir einen Beitrag zu einem vereinfachten Modellvergleich über verschiedene Subdisziplinen der Computerlinguistik hinweg.<sup>41</sup> Ein vereinfachtes Bei-

<sup>40</sup>Bei dem letztgenannten Papier war ich an der Modellierung beteiligt.

<sup>41</sup>Die Motivation und der Ansatz einer solchen Modellierung wurde von mir angestoßen. Ich leite dieses Projekt als

```

<deRESchema name="BioNLP-ST 2009" ver="0.01" auth="Klinger">
  <spantypes>
    <span name="Protein" predict="False"/>
    <span name="Gene_expression" anchors="Gene_expression"
      predict="True"/>
    <span name="Binding" anchors="Binding" predict="True"/>
  </spantypes>
  <frames>
    <frame name="Gene_expression">
      <slot name="Theme" types="Protein" cardinality="1"/>
    </frame>
    <frame name="Binding">
      <slot name="Theme" types="Protein" mincardinality="0"/>
    </frame>
  </frames>
</deRESchema>

```

Figure 26: Ein kleines aber vollständiges Schema für einen Teil des BioNLP-Shared-Tasks (Adel et al., 2018).

```

<deRESchema name="StructuredEmotion" ver="0.01"
  auth="Klinger">
  <spantypes>
    <span name="EmotionCue"
      anchors="AngerEmotion,SadEmotion,JoyEmotion"
      predict="True"/>
    <span name="Person" predict="True"/>
    <span name="Event" predict="True"/>
  </spantypes>
  <frames>
    <frame name="AngerEmotion">
      <slot name="Cue" types="EmotionCue" maxcardinality="1"/>
      <slot name="Feeler" types="Person" maxcardinality="1"/>
      <slot name="Target" types="Person,Event"/>
      <slot name="Cause" types="Person,Event"/>
    </frame>
  </frames>
</deRESchema>

```

Figure 27: Ein Schema zur Spezifikation der strukturierten Emotionserkennung. Zur Vereinfachung wird angenommen, dass nur drei verschiedene Emotionen modelliert werden (Adel et al., 2018).

spiel zur Deklaration der Aufgabe der biomedizinischen Informationsextraktion findet sich in Abbildung 26.

Mit diesem Verfahren können Modelle über verschiedene Phänomene hinweg eingesetzt werden, wenn sie strukturell vergleichbar sind. Einsetzbare Modelle sind zum Beispiel die oben eingeführten Modelle (Klinger and Cimiano, 2013a,b; Kim and Klinger, 2019c, 2018), eine Variante von Scheible et al. (2016) oder auch das in Klinger et al. (2011) vorgestellte Vorgehen.<sup>42</sup>

### 3.5.3 Weitere Anwendungen

Die folgenden Arbeiten untersuchen nicht *per se* Emotion oder Sentiment als Forschungsgegenstand, bauen aber auf diesen Konzepten zumindest zum Teil auf und sollen daher der Vollständigkeit halber kurz Erwähnung finden. Sie zeigen jeweils die Verwendbarkeit von Modellen in praktischen Anwen-

Zusammenschluß verschiedener ansonsten getrennter Projekte in der Arbeitsgruppe an.

<sup>42</sup>Das Modell in Klinger et al. (2011) wurde von mir unter Anleitung entwickelt.

dungen und tragen so zu einer Realweltevaluierung bei.

In Klinger et al. (2012) haben wir aspektbasierte Sentimentanalyse als Kombination von Topic-Modelling (Blei et al., 2003) formuliert und die Ergebnisse so aufbereitet, dass sie von Entscheidungsträgern in der Politik so genutzt werden können, dass Ansichten der Bevölkerung leichter einbezogen werden. In einer Nutzerstudie konnten wir zeigen, dass diese Informationsaggregation die Entscheidungsfindungen erleichtert.<sup>43</sup>

Eine weitere Anwendung von aspektbezogener Sentimentanalyse ist die Arbeit von Kessler et al. (2015).<sup>44</sup> Hier untersuchen wir, ob sich eine Präferenzliste von Produkten durch die Rekonstruktion des Verkaufsrangs nur auf Basis der aus Beurteilungen extrahierten bewerteten Merkmalen erstellen lässt. Wir konnten zeigen, dass einige Aspekte einen stärkeren Beitrag leisten als andere. So ist in der Kameradomäne die allgemeine Performanz und die Videofunktion wichtiger, als der Bildschirm und der Blitz. Dieses Verfahren kann für Produkthersteller genutzt werden, um ihren Arbeitsschwerpunkt auf die Merkmale eines Produkts zu legen, welche tatsächlich einen Einfluss auf den Verkaufsrang (und damit vermutlich auch auf die Zufriedenheit der Nutzer) haben.

In Hartung et al. (2017a,b) haben wir die Frage untersucht, welche Merkmale (insbesondere auch welche Emotionsmerkmale) einen Beitrag zur Erkennung von Personen liefern, welche dem rechts-extremistischen Spektrum oder einer solchen Gruppe zugeordnet werden können. Emotionsmerkmale leisten allerdings keinen zusätzlichen signifikanten Beitrag in der Gesamtheit aller Merkmale. Allerdings ist festzustellen, dass zwei Emotionsmerkmale insbesondere zur Vollständigkeit beitragen. Dies ist zum einen der erhöhte Anteil von emotionstragenden Wörtern bei Autoren, welche als rechtsradikal eingestuft sind, im Vergleich zu den anderen. Zum anderen ist die einzige einzelne Emotion, deren Verteilung unterschiedlich ist, die des Ekels.<sup>45</sup>

<sup>43</sup>Ich habe in dem EU-Projekt den Teilbereich der Datenanalyse geleitet, in dem dieses Publikation entstanden ist, war allerdings an der Nutzerstudie nur am Rande beteiligt.

<sup>44</sup>Die Arbeit wurde in Kollaboration mit einer Doktorandin durchgeführt.

<sup>45</sup>In diesen Arbeiten war ich nur zum Teil an der Konzeptualisierung beteiligt, habe aber die automatischen Emotionsanalysen durchgeführt.



## 4 Zusammenfassung und Ausblick

Das Ziel dieser Habilitationsarbeit war es, Formalisierungen von Affektmodellierungen in Text zu finden, welche komputationelle Klassifikationen und Strukturvorhersagen ermöglichen. Die Beiträge finden sich zum einen in Modellierungsvorschlägen, in annotierten Korpora, in komputationellen Modellen, sowie in Interpretationen und Analysen dieser verschiedenen Artefakte. Die Korpora stellen hierbei durch ihre Manifestation von Konzeptionalisierungen sowie durch die Vorbereitung statistischer Arbeiten einen Wert in sich dar. Sie erlauben auch weitergehende Analysen durch die Vergleiche verschiedener Annotationen verschiedener Annotatorinnen. Die darauf aufbauenden Modellierungen sind jeweils Implementierungen von Annahmen, welche Informationen zur Repräsentation des Phänomens wichtig sind. Hierbei ist zum Teil ein Abwägen zwischen der Akkuratheit des Modells und der erreichbaren Performanz bei der Vorhersage auf unbekanntem Daten notwendig.

Auf Seiten der beigetragenen Korpora ist unser vergleichbares Korpus von deutschen und englischen Amazonproduktbeurteilungen hervorzuheben. Dieses ist mit Aspektnennungen und evaluierenden Phrasen annotiert, sowie der Relation und der Polarität zwischen diesen (Klinger and Cimiano, 2014; Ruppenhofer et al., 2014). Derselben Annotationsrichtlinie folgt eine Ressource, welche App-Reviews als Basis nimmt (Sänger et al., 2016). Diese Korpora finden mittlerweile auch durch andere Arbeitsgruppen Verwendung, um die Methodenentwicklung voranzutreiben.

Für die Emotionsanalyse haben wir zwei multimodale Korpora erstellt, zum einen für Emotionserkennung in Autos (Cevher et al., 2019), zum anderen für Tweets, welche mit Bildern versehen sind (Klinger, 2017). Manuell annotierte Emotionstextkorpora haben wir auf Basis von spanischen Reiseberichten (Ehrlicher et al., 2019), für Tweets (Schuff et al., 2017; Bostan and Klinger, 2019), und Berichten emotional konnotierter Ereignisse mit Crowdsourcing (Troiano et al., 2019) und schwacher Überwachung (Klinger et al., 2018) erstellt. Unsere Aggregation der Mehrzahl verfügbarer Emotionskorpora erleichtert zukünftige Forschungsaufgaben in diesem Gebiet (Bostan and Klinger, 2018).

Korpora für strukturierte Emotionsanalyse stehen nun erstmalig für die Rollen des Fühlers, des Emotionsausdrucks, der Ursache und des

Ziels (Kim and Klinger, 2018) in Literatur sowie für emotionale Figurrelationen in Fan-Fiction zur Verfügung (Kim and Klinger, 2019c). Eine weitere neuartige Annotation erlaubt die Analyse der Art der Emotionskommunikation, also ob Körperpose, Sprache, Mimik, Gestik oder Textinhalt für den Ausdruck der Figur beschrieben wird (Kim and Klinger, 2019b).

Für die Analyse von nicht-wörtlicher Rede im Stilmittel der Satire kompilierten wir den ersten deutschen Korpus sowie die größte aktuell verfügbare Nachrichtentextressource zu diesem Thema. Für die Abgrenzung von Ironie zu Sarkasmus erstellten wir einen Twitterkorpus.

Trotz dieser substantiellen Ressourcenbeiträge bleibt auch jetzt noch festzuhalten, dass die Menge an annotierten Daten in den meisten Sprachen äußerst gering ist. Insbesondere zu den Emotionsphänomenen im Kontext von Rollen existieren kaum Korpora. Zukünftige Arbeiten, welche hier einen Beitrag leisten, werden nach wie vor einen wichtigen Einfluss haben.

Unsere manuell und zum Teil schwach annotierten Korpora realisieren Modellierungsansätze, welche auch für prädiktive Modelle eingesetzt wurden und somit weitergehend ihre Schlüssigkeit unter Beweis stellten. Dies sind Modelle zur Sentimentklassifikation (Barnes et al., 2017) und aspektbasierter Sentimentanalyse (Sänger et al., 2017; Kessler et al., 2015; Klinger and Cimiano, 2013b,a; Sänger et al., 2016; Sänger et al., 2017), auch über Sprachgrenzen hinweg (Barnes and Klinger, 2019; Klinger and Cimiano, 2015; Klinger et al., 2012), sowie Emotionsklassifikationsmodelle für Englisch und andere Sprachen (Troiano et al., 2019; Ehrlicher et al., 2019; Schuff et al., 2017; Klinger, 2017; Cevher et al., 2019), zum Teil mit einem Schwerpunkt auf Ereignisbeschreibungen (Klinger et al., 2018; Troiano et al., 2019) sowie Regressionsmodelle zur Vorhersage der Emotionsintensität (Bostan and Klinger, 2019; Köper et al., 2017). Wir haben weiterhin die ersten Modelle für strukturierte Emotionsanalyse in Literatur erstellt (Kim and Klinger, 2019c, 2018).

Ein Thema, welches in mehreren Modellierungsansätzen wiederkehrt, ist die Nutzung von gemeinsamer Repräsentation oder Inferenz. In McHardy et al. (2019) ist dies die gemeinsame Repräsentation von Satire und Publikationsquelle, in Buschmeier et al. (2014); Ling and Klinger (2016) von Sentiment und Ironie, von allgemeiner Wort-

bedeutung und Sentiment (Barnes et al., 2018b; Barnes and Klinger, 2019; Barnes et al., 2018a) in verschiedenen Sprachen und Domänen sowie zwischen Rollen in der Emotionsanalyse (Kim and Klinger, 2018) und Aspekt und Evaluierung (Klinger and Cimiano, 2013a,b, 2015).

Wir haben weiterhin gezeigt, dass Emotion und Sentiment als Merkmal verschiedene andere Aufgaben unterstützt, so die Erstellung von Bewertungsranglisten (Kessler et al., 2015), die Analyse von Genre in Literatur Kim et al. (2017b,a) und die Erkennung von Rechtsextremismus (Hartung et al., 2017a).

Offene Themen bestehen insbesondere in der Emotionsanalyse, bei der bereits in der Sentimentanalyse untersuchte Forschungsfragen noch weiter betrachtet werden müssen. So müssen strukturierte Analyseansätze in weiteren Domänen getestet werden und damit robuster gemacht werden. Diese sollten dann über verschiedene Sprachen hinweg analysiert werden. Hier ist im Moment noch die Vorhersage von mehr als zwei Sentimentklassen herausfordernd. Wir erwarten daher, dass Methoden, welche Polarität binär unterscheiden können, nicht ohne Anpassungen auf Emotionsanalyse übertragen werden können.

Es fehlt weitestgehend auch noch ein Verständnis, auf welche Weise Emotion in den verschiedenen Domänen kommuniziert wird. Wir schlagen vor, diese Forschungsfragen anzugehen, da die Situation im Vergleich zur Sentimentanalyse komfortabel ist: Emotionsforschung hat eine lange Historie in der Psychologie, und es steht aus, diese Erkenntnisse in die algorithmische Analyse von natürlicher Sprache zu übertragen. Es ist noch nicht abzusehen, welche Anstöße hier zum Beispiel die neuesten Emotionstheorien des Konstruktivismus, zum Beispiel von Feldman-Barrett (2018), für die Informatik und Computerlinguistik liefern werden. Wir erwarten einen substantiellen Vorteil durch interdisziplinäre Anstrengungen, da ihre Theorien neue Erkenntnisse in der Interaktion von Affekt und der Entwicklung einer Emotion darstellen. Diese dürften gerade bei Modellen, welche Affekt, Sentiment und Emotion gemeinsam modellieren, von Nutzen sein.

## Danksagung

Arbeiten, welche im Kontext mit dieser Habilitation standen, wurden bereits während meinen Beschäftigung als wissenschaftlicher Mitarbeiter

nach meiner Promotion bei dem Fraunhofer Institut für Wissenschaftliches Rechnen (SCAI) in Sankt Augustin begonnen. Hier habe ich von Christoph Friedrich und Juliane Fluck die Freiheit bekommen unsere Beiträge in dem EU-Projekt +Spaces nach meinen Interessen zu gestalten, so dass ich mich hier erstmals mit Sentimentanalyse beschäftigen konnte. Ausserdem habe ich hier erste Erfahrungen mit der Anleitung von Projekten und Mitarbeitern sammeln können. Vielen Dank für das Vertrauen. Vielen Dank auch an Antje Wolf, Philipp Senger, Alexander Klenner, Sumit Madan und Christian Ebeling. Mit Euch war die Zeit eine Party.

Als Postdoc bei Philipp Cimiano an der Universität Bielefeld (Arbeitsgruppe Semantische Datenbanken) konnte ich dieses Thema weiter hin zu aspektbasierter Sentimentanalyse im Projekt „Intelligent Technical Systems OstwestfalenLippe“ fortsetzen. Vielen Dank an Philipp sowie alle Gruppenmitglieder der Arbeitsgruppe, insbesondere an John McCrae und Matthias Hartung, von denen ich in dieser Zeit und darüber hinaus viel gelernt habe. Kurz nach meinem Fortgehen aus Bielefeld haben wir auch die Semalytix GmbH gegründet, eine Tätigkeit, welche mir geholfen hat meine wissenschaftlichen Beiträge zu hinterfragen und zum Teil neu auszurichten. Vielen Dank hier auch an Janik Jaskolski, der nicht müde wird den Sinn und die Verwendbarkeit von Forschungsergebnissen zu analysieren.

Während meiner Promotion und wiederholt während meiner Zeit in Bielefeld habe ich Andrew McCallums Arbeitsgruppe an der University of Massachusetts Amherst für mehrere Monate besuchen können. Er und seine Mitarbeiter haben mich bei der Entwicklung von probabilistischen Modellen unterstützt. Vielen Dank insbesondere an Andrew McCallum, Limin Yao, Michael Wick, Sebastian Riedel, Laura Dietz, Sameer Singh, Greg Druck, David Mimno und Jason Naradowsky. Ohne Euch hätte ich ein eingeschränkteres Verständnis von maschinellem Lernen.

Seit 2014 durfte ich zunächst als Vertretungsprofessor und im Anschluss daran als Akademischer (Ober)Rat in Verbeamtung an der Universität Stuttgart frei über meine Forschungsrichtungen entscheiden. Vielen Dank an Sebastian Padó für dieses Vertrauen. Ohne Dich wäre mir die Entwicklung hin zu einem selbständigen Wissenschaftler schwerer gefallen. Vielen Dank auch an Jonas Kuhn für Motivation, gute Laune und klare Kritik, Wiltrud

Kessler für die gute Zusammenarbeit und viel Spaß in Stuttgart und am Bodensee und Christian Scheible für gute Zusammenarbeit und viel Spaß in Stuttgart und Köln. Und natürlich vielen Dank an Maximilian Köper, Laura Bostan, Evgeny Kim, Enrica Troiano und Jeremy Barnes für Eure Geduld und tolle Zusammenarbeit.

Des Weiteren bedanke ich mich bei den verschiedenen Finanzierungsquellen, welche meine Forschung ermöglicht haben. Dies sind neben der Grundfinanzierung der Fraunhofer-Gesellschaft und dem EU-Projekt +Spaces das Projekt It's OWL des BMBF an der Universität Bielefeld (<https://www.its-owl.de>), das DFG-Projekt SEAT (KL 2869/1-1) und das Centrum für reflektierte Textanalyse (CRETA, <http://www.creta.uni-stuttgart.de/>). Weiterhin bedanke ich mich bei dem Wissenschaftscampus Tübingen (<https://www.wissenschaftscampus-tuebingen.de/>) und der Leibnizgesellschaft für die Einbindung, sowie der Universität Stuttgart für die Möglichkeit, meine Arbeit in einer sicheren Anstellung auszuüben.

Zu guter Letzt, vielen Dank Wiebke, für Deine Geduld und Unterstützung.

## References

- Muhammad Abdul-Mageed and Lyle Ungar. 2017. **EmoNet: Fine-grained emotion detection with gated recurrent neural networks**. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 718–728. Association for Computational Linguistics.
- Omri Abend, Shay B. Cohen, and Mark Steedman. 2015. **Lexical event ordering with an edge-factored model**. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1161–1171.
- Heike Adel, Laura Ana Maria Bostan, Sean Papay, Sebastian Padó, and Roman Klinger. 2018. **DERE: A task and domain-independent slot filling framework for declarative relation extraction**. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Brussels, Belgium. Association for Computational Linguistics.
- Rodrigo Agerri, Montse Cuadros, Sean Gaines, and German Rigau. 2013. **OpeNER: Open polarity enhanced named entity recognition**. In *Sociedad Española para el Procesamiento del Lenguaje Natural*, volume 51, pages 215–218.
- Željko Agić, Jörg Tiedemann, Danijela Merkle, Simon Krek, Kaja Dobrovoljc, and Sara Može. 2014. **Cross-lingual dependency parsing of related languages with rich morphosyntactic tagsets**. In *Proceedings of the EMNLP'2014 Workshop on Language Technology for Closely Related Languages and Language Variants*, pages 13–24. Association for Computational Linguistics.
- Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. 2005. **Emotions from text: Machine learning for text-based emotion prediction**. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 579–586. Association for Computational Linguistics.
- Mariana S. C. Almeida, Cláudia Pinto, Helena Figueira, Pedro Mendes, and André F. T. Martins. 2015. **Ali-gning opinions: Cross-lingual opinion mining with dependencies**. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 408–418. Association for Computational Linguistics.
- Saima Aman and Stan Szpakowicz. 2007. **Identifying expressions of emotion in text**. In *Text, Speech and Dialogue*, pages 196–205. Springer Berlin Heidelberg.
- Mikel Artetxe, Gorra Labaka, and Eneko Agirre. 2016. **Learning principled bilingual mappings of word embeddings while preserving monolingual invariance**. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2289–2294. Association for Computational Linguistics.
- Mikel Artetxe, Gorra Labaka, and Eneko Agirre. 2017. **Learning bilingual word embeddings with (almost) no bilingual data**. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 451–462. Association for Computational Linguistics.
- M. Ashcroft, A. Fisher, L. Kaati, E. Omer, and N. Prucha. 2015. **Detecting jihadist messages on twitter**. In *2015 European Intelligence and Security Informatics Conference*, pages 161–164.
- Alexandra Balahur, Jesús M. Hermida, and Andrés Montoyo. 2012. **Detecting implicit expressions of emotion in text: A comparative analysis**. *Decision Support Systems*, 53(4):742–753.
- Alexandra Balahur and Marco Turchi. 2012. **Multilingual sentiment analysis using machine translation?** In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*, pages 52–60. Association for Computational Linguistics.

- Francesco Barbieri and Horacio Saggion. 2014. [Modelling irony in twitter](#). In *Proceedings of the Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 56–64. Association for Computational Linguistics.
- Jeremy Barnes and Roman Klinger. 2019. [Embedding projection for targeted cross-lingual sentiment: Model comparisons and a real-world study](#). *Journal of Artificial Intelligence Research*, 66:691–742.
- Jeremy Barnes, Roman Klinger, and Sabine Schulte im Walde. 2017. [Assessing state-of-the-art sentiment models on state-of-the-art sentiment datasets](#). In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 2–12, Copenhagen, Denmark. Association for Computational Linguistics.
- Jeremy Barnes, Roman Klinger, and Sabine Schulte im Walde. 2018a. [Bilingual sentiment embeddings: Joint projection of sentiment across languages](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2483–2493, Melbourne, Australia. Association for Computational Linguistics.
- Jeremy Barnes, Roman Klinger, and Sabine Schulte im Walde. 2018b. [Projecting embeddings for domain adaption: Joint modeling of sentiment analysis in diverse domains](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 818–830. Association for Computational Linguistics.
- Florian Barth, Evgeny Kim, Sandra Murr, and Roman Klinger. 2018. [A reporting tool for relational visualization and analysis of character mentions in literature](#). In *Book of Abstracts – Digital Humanities im deutschsprachigen Raum*, Cologne, Germany.
- Karin Becker, Viviane P. Moreira, and Aline G.L. dos Santos. 2017. [Multilingual emotion classification using supervised learning: Comparative experiments](#). *Information Processing & Management*, 53(3):684–704.
- David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. [Latent Dirichlet Allocation](#). *Journal of Machine Learning Research*, 3:993–1022.
- Laura Ana Maria Bostan and Roman Klinger. 2018. [An analysis of annotated corpora for emotion classification in text](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2104–2119. Association for Computational Linguistics.
- Laura Ana Maria Bostan and Roman Klinger. 2019. [Exploring fine-tuned embeddings that model intensifiers for emotion analysis](#). In *Proceedings of the Tenth Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Minneapolis, USA. Association for Computational Linguistics.
- Rebecca F. Bruce and Janyce Wiebe. 1999. [Recognizing subjectivity: A case study in manual tagging](#). *Natural Language Engineering*, 5(2):187–205.
- Sven Buechel and Udo Hahn. 2017. [Emobank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 578–585. Association for Computational Linguistics.
- Konstantin Buschmeier, Philipp Cimiano, and Roman Klinger. 2014. [An impact analysis of features in a classification approach to irony detection in product reviews](#). In *Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 42–49, Baltimore, Maryland. Association for Computational Linguistics.
- Deniz Cevher, Sebastian Zepf, and Roman Klinger. 2019. [Towards multimodal emotion recognition in german speech events in cars using transfer learning](#). In *Conference on Natural Language Processing (KONVENS)*, pages 79–90, Erlangen/Nürnberg, Germany.
- Lu Chen, Justin Martineau, Doreen Cheng, and Amit Sheth. 2016. [Clustering for simultaneous extraction of aspects and features from reviews](#). In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 789–799. Association for Computational Linguistics.
- Yufeng Chen, Chengqing Zong, and Keh-Yih Su. 2010. [On jointly recognizing and aligning bilingual named entities](#). In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 631–639. Association for Computational Linguistics.
- Alexandra Chronopoulou, Aikaterini Margatina, Christos Baziotis, and Alexandros Potamianos. 2018. [NTUA-SLP at IEST 2018: Ensemble of neural transfer methods for implicit emotion classification](#). In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 57–64. Association for Computational Linguistics.
- Cindy Chung and James Pennebaker. 2007. *The psychological function of function words*, chapter 12. Psychology Press, New York.
- Simon Clematide, Stefan Gindl, Manfred Klenner, Stefanos Petrakis, Robert Remus, Josef Ruppenhofer, Ulli Waltinger, and Michael Wiegand. 2012. [ML-SA — a multi-layered reference corpus for German sentiment analysis](#). In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, pages 3551–3556. European Language Resources Association (ELRA).

- Rebecca Clift. 1999. [Irony in conversation](#). *Language in Society*, 28(4).
- Sohan De Sarkar, Fan Yang, and Arjun Mukherjee. 2018. [Attending Sentences to detect Satirical Fake News](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3371–3380. Association for Computational Linguistics.
- Yihan Deng, Thierry Declerck, Piroska Lendvai, and Kerstin Denecke. 2016. [The generation of a corpus for clinical sentiment analysis](#). In *The Semantic Web*, pages 311–324, Cham. Springer International Publishing.
- Monique Dittrich and Sebastian Zepf. 2019. [Exploring the validity of methods to track emotions behind the wheel](#). In *Persuasive Technology: Development of Persuasive and Behavior Change Support Systems*, pages 115–127. Springer International Publishing.
- Chris Dyer, Victor Chahuneau, and Noah A. Smith. 2013. [A simple, fast, and effective reparameterization of IBM model 2](#). In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 644–648. Association for Computational Linguistics.
- Hanno Ehrlicher, Roman Klinger, Jörg Lehmann, and Sebastian Padó. 2019. [Measuring historical emotions and their evolution: An interdisciplinary endeavour to investigate the ‘emotions of encounter’](#). *Laboratório Interdisciplinar sobre Informação e Conhecimento em revista (Liinc em revista)*, 15(1).
- Paul Ekman. 1999. Basic emotions. In M Dalglish, T; Power, editor, *Handbook of Cognition and Emotion*. John Wiley & Sons.
- Delia Irazú Hernández Farías, Viviana Patti, and Paolo Rosso. 2016. [Irony detection in twitter: The role of affective content](#). *ACM Trans. Internet Technol.*, 16(3):19:1–19:24.
- Manaal Faruqui, Jesse Dodge, Sujay Kumar Jauhar, Chris Dyer, Eduard Hovy, and Noah A. Smith. 2015. [Retrofitting word vectors to semantic lexicons](#). In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1606–1615, Denver, Colorado. Association for Computational Linguistics.
- Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. 2017. [Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1615–1625. Association for Computational Linguistics.
- Sergey Feldman, Marius Marin, Julie Medero, and Mari Ostendorf. 2009. [Classifying factored genres with part-of-speech histograms](#). In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Short Papers*, pages 173–176, Boulder, Colorado. Association for Computational Linguistics.
- Lisa Feldman-Barrett. 2018. *How Emotions Are Made: The Secret Life of the Brain*. Pan.
- Emilio Ferrara, Wen-Qiang Wang, Onur Varol, Alessandro Flammini, and Aram Galstyan. 2016. [Predicting online extremism, content adopters, and interaction reciprocity](#). In *Social Informatics*, pages 22–39, Cham. Springer International Publishing.
- Elena Filatova. 2012. [Irony and sarcasm: Corpus generation and analysis using crowdsourcing](#). In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, pages 392–398. European Language Resources Association (ELRA).
- Gayatree Ganu, Noemie Elhadad, and Amélie Marian. 2009. [Beyond the stars: Improving rating predictions using review text content](#). In *Twelfth International Workshop on the Web and Databases (WebDB 2009)*.
- Diman Ghazi, Diana Inkpen, and Stan Szpakowicz. 2015. [Detecting emotion stimuli in emotion-bearing sentences](#). In *Computational Linguistics and Intelligent Text Processing*, pages 152–165. Springer International Publishing.
- Aniruddha Ghosh, Guofu Li, Tony Veale, Paolo Rosso, Ekaterina Shutova, John Barnden, and Antonio Reyes. 2015. [SemEval-2015 Task 11: Sentiment Analysis of Figurative Language in Twitter](#). In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 470–478. Association for Computational Linguistics.
- Stephan Gouws and Anders Søgaard. 2015. [Simple task-specific bilingual word embeddings](#). In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1386–1390. Association for Computational Linguistics.
- Lin Gui, Jiannan Hu, Yulan He, Ruifeng Xu, Lu Qin, and Jiachen Du. 2017. [A question answering approach for emotion cause extraction](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1593–1602. Association for Computational Linguistics.
- Lin Gui, Ruifeng Xu, Qin Lu, Dongyin Wu, and Yu Zhou. 2016. [Emotion cause extraction, a challenging task with corpus construction](#). In *Social Media Processing*, pages 98–109. Springer Singapore.
- Matthias Hartung, Roman Klinger, Franziska Schmidtke, and Lars Vogel. 2017a. [Identifying right-wing](#)

- extremism in german twitter profiles: A classification approach. In *Natural Language Processing and Information Systems: 22nd International Conference on Applications of Natural Language to Information Systems, NLDB 2017, Liège, Belgium, June 21-23, 2017, Proceedings*, pages 320–325, Cham. Springer International Publishing.
- Matthias Hartung, Roman Klinger, Franziska Schmidtke, and Lars Vogel. 2017b. [Ranking right-wing extremist social media profiles by similarity to democratic and extremist groups](#). In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 24–33, Copenhagen, Denmark. Association for Computational Linguistics.
- Karl Moritz Hermann and Phil Blunsom. 2014. [Multilingual models for compositional distributed semantics](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 58–68. Association for Computational Linguistics.
- Irazú Hernández-Farías, José-Miguel Benedí, and Paolo Rosso. 2015. [Applying basic features from sentiment analysis for automatic irony detection](#). In *Pattern Recognition and Image Analysis*, pages 337–344, Cham. Springer International Publishing.
- Lena Hettinger, Martin Becker, Isabella Reger, Fotis Jannidis, and Andreas Hotho. 2015. [Genre classification on german novels](#). In *Proceedings of the 26th International Workshop on Database and Expert Systems Applications*, pages 249–253.
- Lena Hettinger, Fotis Jannidis, Isabella Reger, and Andreas Hotho. 2016. [Classification of literary subgenres](#). In *Proceedings of DHd 2016*, Leipzig, Germany.
- Jeremy Howard and Sebastian Ruder. 2018. [Universal language model fine-tuning for text classification](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 328–339. Association for Computational Linguistics.
- Minqing Hu and Bing Liu. 2004. [Mining and summarizing customer reviews](#). In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '04*, pages 168–177. ACM.
- Niklas Jakob and Iryna Gurevych. 2010. [Extracting opinion targets in a single and cross-domain setting with conditional random fields](#). In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1035–1045. Association for Computational Linguistics.
- Soufian Jebbara and Philipp Cimiano. 2016. [Aspect-based relational sentiment analysis using a stacked neural network architecture](#). In *Proceedings of the Twenty-second European Conference on Artificial Intelligence, ECAI'16*, pages 1123–1131, Amsterdam, The Netherlands, The Netherlands. IOS Press.
- Aditya Joshi, Vaibhav Tripathi, Kevin Patel, Pushpak Bhattacharyya, and Mark Carman. 2016. [Are word embedding-based features useful for sarcasm detection?](#) In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1006–1011, Austin, Texas. Association for Computational Linguistics.
- L. Kaati, E. Omer, N. Prucha, and A. Shrestha. 2015. [Detecting multipliers of jihadism on twitter](#). In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 954–960.
- Kaisla Kajava and Emily Sofi Öhman. 2019. [Are emotions and sentiments preserved in translation?](#) In *NoDaLiDa*.
- Jussi Karlgren and Douglass Cutting. 1994. [Recognizing text genres with simple metrics using discriminant analysis](#). In *COLING 1994 Volume 2: The 15th International Conference on Computational Linguistics*.
- Andrej Karpathy and Li Fei-Fei. 2015. [Deep visual-semantic alignments for generating image descriptions](#). In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- Brett Kessler, Geoffrey Nunberg, and Hinrich Schütze. 1997. [Automatic detection of text genre](#). In *35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*, pages 32–38, Madrid, Spain. Association for Computational Linguistics.
- Jason S. Kessler, Miriam Eckert, Lyndsie Clark, and Nicolas Nicolov. 2010. [The 2010 ICWSM JDPa Sentiment Corpus for the Automotive Domain](#). In *4th International AAAI Conference on Weblogs and Social Media Data Workshop Challenge (ICWSM-DWC 2010)*.
- Wiltrud Kessler, Roman Klinger, and Jonas Kuhn. 2015. [Towards opinion mining from reviews for the prediction of product rankings](#). In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 51–57, Lisboa, Portugal. Association for Computational Linguistics.
- Evgeny Kim and Roman Klinger. 2018. [Who feels what and why? annotation of a literature corpus with semantic roles of emotions](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1345–1359. Association for Computational Linguistics.
- Evgeny Kim and Roman Klinger. 2019a. [A Survey on Sentiment and Emotion Analysis for Computational Literary Studies](#). *Zeitschrift für Digitale Geisteswissenschaften*. Accepted.

- Evgeny Kim and Roman Klinger. 2019b. [An analysis of emotion communication channels in fan-fiction: Towards emotional storytelling](#). In *Proceedings of the Second Workshop on Storytelling*, pages 56–64, Florence, Italy. Association for Computational Linguistics.
- Evgeny Kim and Roman Klinger. 2019c. [Frowning Frodo, wincing Leia, and a seriously great friendship: Learning to classify emotional relationships of fictional characters](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 647–653, Minneapolis, Minnesota. Association for Computational Linguistics.
- Evgeny Kim, Sebastian Padó, and Roman Klinger. 2017a. [Investigating the relationship between literary genres and emotional plot development](#). In *Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pages 17–26. Association for Computational Linguistics.
- Evgeny Kim, Sebastian Padó, and Roman Klinger. 2017b. [Prototypical Emotion Developments in Literary Genres](#). In *Digital Humanities 2017: Conference Abstracts*. McGill University and Université de Montréal.
- Roman Klinger. 2017. [Does optical character recognition and caption generation improve emotion detection in microblog posts?](#) In *Natural Language Processing and Information Systems: 22nd International Conference on Applications of Natural Language to Information Systems, NLDB 2017, Liège, Belgium, June 21-23, 2017, Proceedings*, pages 313–319, Cham. Springer International Publishing.
- Roman Klinger and Philipp Cimiano. 2013a. [Bi-directional inter-dependencies of subjective expressions and targets and their value for a joint model](#). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 848–854, Sofia, Bulgaria. Association for Computational Linguistics.
- Roman Klinger and Philipp Cimiano. 2013b. [Joint and pipeline probabilistic models for fine-grained sentiment analysis: Extracting aspects, subjective phrases and their relations](#). In *2013 IEEE 13th International Conference on Data Mining Workshops (ICDMW)*, pages 937–944.
- Roman Klinger and Philipp Cimiano. 2014. [The USA-GE review corpus for fine grained multi lingual opinion analysis](#). In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 2211–2218, Reykjavik, Iceland. European Language Resources Association (ELRA). ACL Anthology Identifier: L14-1656.
- Roman Klinger and Philipp Cimiano. 2015. [Instance selection improves cross-lingual model training for fine-grained sentiment analysis](#). In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, pages 153–163, Beijing, China. Association for Computational Linguistics.
- Roman Klinger, Orphée De Clercq, Saif Mohammad, and Alexandra Balahur. 2018. [IEST: WASSA-2018 implicit emotions shared task](#). In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 31–42, Brussels, Belgium. Association for Computational Linguistics.
- Roman Klinger, Sebastian Riedel, and Andrew McCallum. 2011. [Inter-event dependencies support event extraction from biomedical literature](#). In *Mining Complex Entities from Network and Biomedical Data (MIND), European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*.
- Roman Klinger, Philipp Senger, Sumit Madan, and Michal Jacovi. 2012. [Online communities support policy-making: The need for data analysis](#). In *Electronic Participation*, volume 7444 of *Lecture Notes in Computer Science*, pages 132–143. Springer Berlin Heidelberg.
- Roman Klinger, Surayya Samat Suliya, and Nils Reiter. 2016. [Automatic Emotion Detection for Quantitative Literary Studies – A case study based on Franz Kafka’s “Das Schloss” and “Amerika”](#). In *Digital Humanities 2016: Conference Abstracts*, pages 826–828, Kraków, Poland. Jagiellonian University and Pedagogical University.
- Maximilian Köper, Evgeny Kim, and Roman Klinger. 2017. [IMS at EmoInt-2017: Emotion intensity prediction with affective norms, automatically extended resources and deep learning](#). In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 50–57, Copenhagen, Denmark. Association for Computational Linguistics.
- Roger J. Kreuz and Sam Glucksberg. 1989. [How to be sarcastic: The echoic reminder theory of verbal irony](#). *Journal of Experimental Psychology: General*, 118(4).
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. [Imagenet classification with deep convolutional neural networks](#). In *Advances in neural information processing systems*, pages 1097–1105.
- Lun-Wei Ku, Chia-Ying Lee, and Hsin-Hsi Chen. 2009. [Identification of opinion holders](#). In *International Journal of Computational Linguistics & Chinese Language Processing, Volume 14, Number 4, December 2009*.
- John Lafferty, Andrew McCallum, and Fernando Pereira. 2001. [Conditional random fields: Probabilistic models for segmenting and labeling sequence data](#). In *International Conference on Machine Learning*, pages 282–289.

- Guillaume Lample, Alexis Conneau, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. 2018. [Word translation without parallel data](#). In *International Conference on Learning Representations*.
- Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. [DailyDialog: A manually labelled multi-turn dialogue dataset](#). In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 986–995. Asian Federation of Natural Language Processing.
- Jennifer Ling and Roman Klinger. 2016. [An empirical, quantitative analysis of the differences between sarcasm and irony](#). In *The Semantic Web: ESWC 2016 Satellite Events, Heraklion, Crete, Greece, May 29 – June 2, 2016, Revised Selected Papers*, pages 203–216. Springer International Publishing.
- Vicki Liu, Carmen Banea, and Rada Mihalcea. 2017. [Grounded emotions](#). In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 477–483.
- Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. [Learning word vectors for sentiment analysis](#). In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 142–150. Association for Computational Linguistics.
- Diana Maynard and Mark Greenwood. 2014. [Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis](#). In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 4238–4243. European Language Resources Association (ELRA).
- Robert McHardy, Heike Adel, and Roman Klinger. 2019. [Adversarial training for satire detection: Controlling for confounding variables](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 660–665, Minneapolis, Minnesota. Association for Computational Linguistics.
- Jacques M. van Meel. 1995. [Representing emotions in literature and paintings: A comparative analysis](#). *Poetics*, 23(1):159–176. Emotions and Cultural Products.
- Lee Sophia Yat Mei, Chen Ying, Huang Chu-Ren, and Li Shoushan. 2012. [Detecting emotion causes with a linguistic rule-based approach](#). *Computational Intelligence*, 29(3):390–416.
- Tomas Mikolov, Kai Chen, Greg S. Corrado, and Jeffrey Dean. 2013a. [Efficient estimation of word representations in vector space](#).
- Tomas Mikolov, Quoc V. Le, and Ilya Sutskever. 2013b. [Exploiting similarities among languages for machine translation](#).
- Ruslan Mitkov, editor. 2003. *The Oxford Handbook Computational Linguistics*. Oxford University Press.
- Miljana Mladenović, Cvetana Krstev, Jelena Mitrović, and Ranka Stanković. 2017. [Using lexical resources for irony and sarcasm classification](#). In *Proceedings of the 8th Balkan Conference in Informatics, BCI '17*, pages 13:1–13:8, New York, NY, USA. ACM.
- Saif Mohammad. 2012. [#emotional tweets](#). In *\*SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pages 246–255. Association for Computational Linguistics.
- Saif Mohammad and Felipe Bravo-Marquez. 2017. [Wassa-2017 shared task on emotion intensity](#). In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 34–49. Association for Computational Linguistics.
- Saif Mohammad, Felipe Bravo-Marquez, Mohammad Salameh, and Svetlana Kiritchenko. 2018. [SemEval-2018 task 1: Affect in tweets](#). In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 1–17. Association for Computational Linguistics.
- Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. 2017. [Stance and sentiment in tweets](#). *Special Section of the ACM Transactions on Internet Technology on Argumentation in Social Media*, 17(3).
- Saif M Mohammad and Peter D Turney. 2013. [Crowdsourcing a word–emotion association lexicon](#). *Computational Intelligence*, 29(3):436–465.
- Saif M Mohammad, Xiaodan Zhu, Svetlana Kiritchenko, and Joel Martin. 2015. [Sentiment, emotion, purpose, and style in electoral tweets](#). *Information Processing & Management*, 51(4):480–499.
- D. Monett and H. Stolte. 2016. [Predicting star ratings based on annotated reviews of mobile apps](#). In *2016 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 421–428.
- Preslav Nakov, Sara Rosenthal, Zornitsa Kozareva, Veselin Stoyanov, Alan Ritter, and Theresa Wilson. 2013. [SemEval-2013 task 2: Sentiment analysis in twitter](#). In *Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 312–320. Association for Computational Linguistics.



- Claire Nédellec, Robert Bossy, Jin-Dong Kim, Jung-Jae Kim, Tomoko Ohta, Sampo Pyysalo, and Pierre Zweigenbaum. 2013. [Overview of bionlp shared task 2013](#). In *Proceedings of the BioNLP Shared Task 2013 Workshop*, pages 1–7.
- Sebastian Padó and Mirella Lapata. 2009. [Cross-lingual annotation projection for semantic roles](#). *Journal of Artificial Intelligence Research (JAIR)*, 36:307–340.
- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. [The proposition bank: An annotated corpus of semantic roles](#). *Computational Linguistics*, 31(1):71–106.
- Sinno Jialin Pan and Qiang Yang. 2009. [A survey on transfer learning](#). *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359.
- Bo Pang and Lillian Lee. 2004. [A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts](#). In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics, Main Volume*, pages 271–278.
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. [Glove: Global vectors for word representation](#). In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.
- Rosalind W. Picard. 2000. *Affective Computing*. MIT Press.
- Robert Plutchik. 2001. [The nature of emotions human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice](#). *American Scientist*, 89(4):344–350.
- Soujanya Poria, Erik Cambria, Devamanyu Hazarika, and Prateek Vij. 2016. [A deeper look into sarcastic tweets using deep convolutional neural networks](#). *CoRR*, abs/1610.08815.
- Jonathan Posner, James A. Russell, and Bradley S. Peterson. 2005. [The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology](#). *Dev Psychopathol*, 17(3):715–734.
- Sameer Pradhan, Eduard H Hovy, Mitch Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2007. [OntoNotes: A Unified Relational Semantic Representation](#). In *Proceedings of the First IEEE International Conference on Semantic Computing*, pages 517–526.
- Daniel Preoțiuc-Pietro, H. Andrew Schwartz, Gregory Park, Johannes Eichstaedt, Margaret Kern, Lyle Ungar, and Elisabeth Shulman. 2016. [Modelling valence and arousal in facebook posts](#). In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 9–15. Association for Computational Linguistics.
- Peter Prettenhofer and Benno Stein. 2011. [Cross-lingual adaptation using structural correspondence learning](#). *ACM Transactions on Intelligent Systems and Technology*, 3(1):1–22.
- Tomáš Ptáček, Ivan Habernal, and Jun Hong. 2014. [Sarcasm Detection on Czech and English Twitter](#). In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 213–223. Dublin City University and Association for Computational Linguistics.
- James Pustejovsky, José M Castano, Robert Ingria, Roser Sauri, Robert J Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir R Radev. 2003a. [TimeML: Robust specification of event and temporal expressions in text](#). In *Proceedings of the 5th International Workshop on Computational Semantics (IWCS-5)*.
- James Pustejovsky, Patrick Hanks, Roser Sauri, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, et al. 2003b. [The TIMEBANK corpus](#). In *Proceedings of Corpus Linguistics 2003 Conference*, pages 647–656.
- Ella Rabinovich, Raj Nath Patel, Shachar Mirkin, Lucia Specia, and Shuly Wintner. 2017. [Personalized machine translation: Preserving original author traits](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 1074–1084. Association for Computational Linguistics.
- Ashwin Rajadesingan, Reza Zafarani, and Huan Liu. 2015. [Sarcasm detection on twitter: A behavioral modeling approach](#). In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, WSDM '15*, pages 97–106. ACM.
- Antonio Reyes, Paolo Rosso, and Tony Veale. 2013. [A multidimensional approach for detecting irony in Twitter](#). *Language Resources and Evaluation*, 47(1).
- Ellen Riloff, Ashequl Qadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert, and Ruihong Huang. 2013. [Sarcasm as contrast between a positive sentiment and negative situation](#). In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 704–714. Association for Computational Linguistics.
- Alon Rozental, Daniel Fleischer, and Zohar Kelrich. 2018. [Amobee at IEST 2018: Transfer learning from language models](#). In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 43–49. Association for Computational Linguistics.
- Josef Ruppenhofer, Roman Klinger, Julia Maria Struß, Jonathan Sonntag, and Michael Wiegand. 2014. [IGGSA Shared Tasks on German Sentiment Analysis](#). In *Workshop Proceedings of the 12th Edition of the KONVENS Conference*, pages 164–173, Hildesheim, Germany. University of Hildesheim.

- Irene Russo, Tommaso Caselli, Francesco Rubino, Ester Boldrini, and Patricio Martinez-Barco. 2011. [EMOCause: An easy-adaptable approach to extract emotion cause contexts](#). In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, pages 153–160. Association for Computational Linguistics.
- Spyridon Samothrakis and Maria Fasli. 2015. [Emotional sentence annotation helps predict fiction genre](#). *PloS one*, 10(11):e0141922.
- Mario Sanger, Ulf Leser, and Roman Klinger. 2017. [Fine-grained opinion mining from mobile app reviews with word embedding features](#). In *Natural Language Processing and Information Systems: 22nd International Conference on Applications of Natural Language to Information Systems, NLDB 2017, Liège, Belgium, June 21-23, 2017, Proceedings*, pages 3–14, Cham. Springer International Publishing.
- Asad Sayeed, Jordan Boyd-Graber, Bryan Rusk, and Amy Weinberg. 2012. [Grammatical structures for word-level sentiment detection](#). In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 667–676. Association for Computational Linguistics.
- Jacob R. Scanlon and Matthew S. Gerber. 2014. [Automatic detection of cyber-recruitment by violent extremists](#). *Security Informatics*, 3(1).
- Christian Scheible, Roman Klinger, and Sebastian Padó. 2016. [Model architectures for quotation detection](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1736–1745, Berlin, Germany. Association for Computational Linguistics.
- Klaus R. Scherer. 2005. [What are emotions? and how can they be measured?](#) *Social Science Information*, 44(4):695–729.
- Klaus R Scherer and Harald G Wallbott. 1994. [Evidence for universality and cultural variation of differential emotion response patterning](#). *Journal of personality and social psychology*, 66(2):310.
- Klaus R Scherer and Harald G Wallbott. 1997. [The isear questionnaire and codebook](#). Geneva Emotion Research Group.
- Hendrik Schuff, Jeremy Barnes, Julian Mohme, Sebastian Padó, and Roman Klinger. 2017. [Annotation, modelling and analysis of fine-grained emotions on a stance and sentiment detection corpus](#). In *Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 13–23, Copenhagen, Denmark. Workshop at Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics.
- Yeong-Seok Seo and Jun-Ho Huh. 2019. [Automatic emotion-based music classification for supporting intelligent iot applications](#). *Electronics*, 8(2):164.
- Faiz Ali Shah, Kairit Sirts, and Dietmar Pfahl. 2019. [Is the safe approach too simple for app feature extraction? a replication study](#). In *Requirements Engineering: Foundation for Software Quality*, pages 21–36, Cham. Springer International Publishing.
- Kashif Shah and Lucia Specia. 2014. [Quality estimation for translation selection](#). In *Conference of the European Association for Machine Translation, EAMT*.
- Mostafa Al Masum Shaikh, Helmut Prendinger, and Mitsuru Ishizuka. 2009. [A Linguistic Interpretation of the OCC Emotion Model for Affect Sensing from Text](#), pages 45–73. Springer London, London.
- Han-Xiao Shi and Xiao-Jun Li. 2011. [A sentiment analysis model for hotel reviews based on supervised learning](#). In *Proceedings of the International Conference on Machine Learning and Cybernetics*.
- C. A. Smith and P. C. Ellsworth. 1985. [Patterns of cognitive appraisal in emotion](#). *Journal of Personality and Social Psychology*, 48(4):813–838.
- Ray Smith. 2007. [An overview of the tesseract ocr engine](#). In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, volume 2, pages 629–633. IEEE.
- Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D. Manning, Andrew Ng, and Christopher Potts. 2013-10. [Recursive deep models for semantic compositionality over a sentiment treebank](#). In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1631–1642. Association for Computational Linguistics.
- Jonathan Sonntag and Manfred Stede. 2014. [Sentiment Analysis: What’s Your Opinion?](#), pages 177–199. Springer International Publishing, Cham.
- Lucia Specia, Kashif Shah, Jose G.C. de Souza, and Trevor Cohn. 2013. [QuEst - a translation quality estimation framework](#). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 79–84. Association for Computational Linguistics.
- D. Spina, E. Meij, A. Oghina, M. T. Bui, M. Breuss, and M. de Rijke. 2012a. [A Corpus for Entity Profiling in Microblog Posts](#). In *LREC Workshop on Information Access Technologies for Online Reputation Management*.
- Damiano Spina, Edgar Meij, Maarten de Rijke, Andrei Oghina, Minh Thuong Bui, and Mathias Breuss. 2012b. [Identifying entity aspects in microblog posts](#). In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, SIGIR ’12*, pages 1089–1090. ACM.

- Efstathios Stamatatos, Nikos Fakotakis, and George Kokkinakis. 2000. [Automatic text categorization in terms of genre and author](#). *Computational Linguistics*, 26(4):471–495.
- Andreas Stolcke. 2002. [Srilm-an extensible language modeling toolkit](#). In *Proceedings International Conference on Spoken Language Processing*, pages 257–286.
- Andreas Stolcke, Jing Zheng, Wen Wang, and Victor Abrash. 2011. [Srilm at sixteen: Update and outlook](#). In *Proc. IEEE Automatic Speech Recognition and Understanding Workshop*. IEEE SPS.
- R. Stöss. 2010. *Rechtsextremismus im Wandel*. Friedrich-Ebert-Stiftung, Berlin.
- Carlo Strapparava and Rada Mihalcea. 2007. [Semeval-2007 task 14: Affective text](#). In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 70–74. Association for Computational Linguistics.
- Carlo Strapparava and Alessandro Valitutti. 2004. [WordNet affect: an affective extension of WordNet](#). In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, pages 1083–1086. European Language Resources Association (ELRA).
- Florian Strohm and Roman Klinger. 2018. [An empirical analysis of the role of amplifiers, downtoners, and negations in emotion classification in microblogs](#). In *The 5th IEEE International Conference on Data Science and Advanced Analytics, Special Track on Sentiment, Emotion, and Credibility of Information in Social Data*, DSAA, Turin, Italy. IEEE.
- Johann Georg Sulzer. 1771. *Allgemeine Theorie der Schönen Künste*, 1. edition. Weidmann; Reich. In German.
- Changzhi Sun, Yuanbin Wu, Man Lan, Shiliang Sun, and Qi Zhang. 2017. [Large-scale opinion relation extraction with distantly supervised neural network](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 1033–1043, Valencia, Spain. Association for Computational Linguistics.
- Mario Sängler, Ulf Leser, Steffen Kemmerer, Peter Adolphs, and Roman Klinger. 2016. [SCARE — The Sentiment Corpus of App Reviews with Fine-grained Annotations in German](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 1114–1121, Paris, France. European Language Resources Association (ELRA).
- Kai Sheng Tai, Richard Socher, and Christopher D. Manning. 2015. [Improved semantic representations from tree-structured long short-term memory networks](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1556–1566. Association for Computational Linguistics.
- Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu. 2016. [Effective LSTMs for target-dependent sentiment classification](#). In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3298–3307. The COLING 2016 Organizing Committee.
- Duyu Tang, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, and Bing Qin. 2014. [Learning sentiment-specific word embedding for twitter sentiment classification](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1555–1565. Association for Computational Linguistics.
- Hande Taslioglu and Pinar Karagoz. 2017. [Irony detection on microposts with limited set of features](#). In *Proceedings of the Symposium on Applied Computing, SAC '17*, pages 1076–1081, New York, NY, USA. ACM.
- I-Hsien Ting, Hsing-Miao Chi, Jyun-Sing Wu, and Shyue-Liang Wang. 2013. [An approach for hate groups detection in facebook](#). In *The 3rd International Workshop on Intelligent Data Analysis and Management*, pages 101–106, Dordrecht. Springer Netherlands.
- Enrica Troiano, Sebastian Padó, and Roman Klinger. 2019. [Crowdsourcing and validating event-focused emotion corpora for German and English](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4005–4011, Florence, Italy. Association for Computational Linguistics.
- Oren Tsur, Dmitry Davidov, and Ari Rappoport. 2010. [ICWSM – A Great Catchy Name: Semi-Supervised Recognition of Sarcastic Sentences in Online Product Reviews](#). In *Proceedings of the Fourth International AAI Conference on Weblogs and Social Media*, pages 162–169. The AAI Press.
- Peter Turney. 2002. [Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews](#). In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 417–424. Association for Computational Linguistics.
- Oscar Täckström and Ryan McDonald. 2011. [Semi-supervised latent variable models for sentence-level sentiment analysis](#). In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 569–574. Association for Computational Linguistics.
- Olga Uryupina, Barbara Plank, Aliaksei Severyn, Agata Rotondi, and Alessandro Moschitti. 2014. [SenTube: A corpus for sentiment analysis on YouTube social media](#). In *Proceedings of the Ninth International*

- Conference on Language Resources and Evaluation (LREC'14), pages 4244–4249. European Language Resources Association (ELRA).
- Akira Utsumi. 1996. [A unified theory of irony and its computational formalization](#). In *Proceedings of the 16th conference on Computational linguistics - Volume 2, COLING '96*, pages 962–967. Association for Computational Linguistics.
- Cynthia Van Hee, Els Lefever, and Veronique Hoste. 2015. [LT3: Sentiment Analysis of Figurative Tweets: piece of cake #NotReally](#). In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 684–688. Association for Computational Linguistics.
- David Vilares, Miguel A. Alonso, and Carlos Gomez-Rodriguez. 2017. [Supervised sentiment analysis in multilingual environments](#). *Information Processing & Management*, 53(3):595 – 607.
- O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. 2015. [Show and tell: A neural image caption generator](#). In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3156–3164.
- Po-Ya Angela Wang. 2013. [#irony or #sarcasm — a quantitative and qualitative study based on twitter](#). In *Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC 27)*, pages 349–356. Department of English, National Chengchi University.
- Y. Wei, L. Singh, and S. Martin. 2016. [Identification of extremism on twitter](#). In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1251–1255.
- Yifang Wei and Lisa Singh. 2017. [Using network flows to identify users sharing extremist content on social media](#). In *Advances in Knowledge Discovery and Data Mining*, pages 330–342, Cham. Springer International Publishing.
- Janyce Wiebe. 2000. [Learning subjective adjectives from corpora](#). In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pages 735–740. AAAI Press.
- Janyce Wiebe and Ellen Riloff. 2005. [Creating subjective and objective sentence classifiers from unannotated texts](#). In *International conference on intelligent text processing and computational linguistics*, pages 486–497. Springer.
- Michael Wiegand, Christine Bocionek, and Josef Ruppenhofer. 2016. [Opinion holder and target extraction on opinion compounds – a linguistic approach](#). In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 800–810. Association for Computational Linguistics.
- Michael Wiegand and Dietrich Klakow. 2010. [Convolution kernels for opinion holder extraction](#). In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 795–803. Association for Computational Linguistics.
- Michael Wiegand, Marc Schulder, and Josef Ruppenhofer. 2015. [Opinion holder and target extraction for verb-based opinion predicates – the problem is not solved](#). In *Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 148–155. Association for Computational Linguistics.
- Theresa Wilson and Janyce Wiebe. 2005. [Annotating attributions and private states](#). In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, pages 53–60. Association for Computational Linguistics.
- Shufeng Xiong, Yue Zhang, Donghong Ji, and Yinxia Lou. 2016. [Distance metric learning for aspect phrase grouping](#). In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 2492–2502. The COLING 2016 Organizing Committee.
- Ruifeng Xu, Lin Gui, Jun Xu, Qin Lu, and Kam-Fai Wong. 2013. [Cross lingual opinion holder extraction based on multi-kernel svms and transfer learning](#). *World Wide Web Journal*, pages 299–316.
- Fan Yang, Arjun Mukherjee, and Eduard Dragut. 2017. [Satirical News Detection and Analysis using Attention Mechanism and Linguistic Features](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1979–1989. Association for Computational Linguistics.
- David Yarowsky and Grace Ngai. 2001. [Inducing multilingual POS taggers and NP bracketers via robust projection across aligned corpora](#). In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*.
- David Yarowsky, Grace Ngai, and Richard Wicentowski. 2001. [Inducing multilingual text analysis tools via robust projection across aligned corpora](#). In *Proceedings of the First International Conference on Human Language Technology Research*.
- Liang-Chih Yu, Jin Wang, K. Robert Lai, and Xuejie Zhang. 2017. [Refining word embeddings for sentiment analysis](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 534–539. Association for Computational Linguistics.
- Li Zhao, Minlie Huang, Haiqiang Chen, Junjun Cheng, and Xiaoyan Zhu. 2014. [Clustering aspect-related phrases by leveraging sentiment distribution consistency](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1614–1623. Association for Computational Linguistics.

Xinjie Zhou, Xiaojun Wan, and Jianguo Xiao. 2015. Clopinionminer: Opinion target extraction in a cross-language scenario. *IEEE/ACM Transactions on Audio, Speech & Language Processing*, 23(4):619–630.

Emily Öhman, Timo Honkela, and Jörg Tiedemann. 2016. The challenges of multi-dimensional sentiment analysis across languages. In *Proceedings of the Workshop on Computational Modeling of People’s Opinions, Personality, and Emotions in Social Media (PEOPLES)*, pages 138–142. The COLING 2016 Organizing Committee.

Emily Öhman, Kaisla Kajava, Jörg Tiedemann, and Timo Honkela. 2018. Creating a dataset for multilingual fine-grained emotion-detection using gamification-based annotation. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 24–30. Association for Computational Linguistics.