

# Evaluation der Qualität lexikalischer Ressourcen zur Stimmungserkennung in literarischen Texten

Dirk Reinel

Hochschule für Angewandte Wissenschaften Hof

Alfons-Goppel-Platz 1

95028 Hof, Deutschland

dreinel@iisys.de

## Zusammenfassung

In dieser Veröffentlichung wird die Qualität bzw. Effektivität von lexikalischen Ressourcen zur automatischen Stimmungserkennung in literarischen Texten evaluiert. Dazu werden die drei unterschiedlichen Wortlisten Sentiment Phrase List (SePL), NRC Emotion Lexicon und SentimentWortschatz (SentiWS) sowie ein manuell klassifiziertes Referenzset verwendet. Der Testkorpus besteht aus 20 ausgewählten Märchen aus der Sammlung der schönsten Kinder- und Hausmärchen der Brüder Grimm, online bezogen von der Website des Projekt Gutenberg-DE. Durch die Berechnung bestimmter Maßzahlen wird gezeigt, dass die Werte der automatisch klassifizierten Texte mit den Werten des Referenzkorpus korrelieren. Es wird jedoch deutlich, dass die berechneten Korrelationen, bedingt durch die begrenzte Anzahl der Texte des Testkorpus, sehr instabil sind. Zudem wird gezeigt, dass die manuelle Klassifikation (positiv / negativ) auf Satzebene durch zwei voneinander unabhängige Personen bei dieser Art von literarischen Texten zufriedenstellend funktioniert.

## 1 Einleitung

Literarische Texte im Allgemeinen und Märchen im Speziellen sind und waren schon immer Träger von verschiedenen Emotionen und Stimmungen. Durch gezielte Variationen und Schwankungen im Verlauf der Handlung, wie beispielsweise durch den Tod handelnder Figuren, erzeugen Autoren unterschiedliche Stimmungen, die die Leserschaft „fesseln“ und mitfühlen lassen sollen. Bisher beschränken sich Informationen über diese Texte meist auf verschiedene Metadaten (z.B. Autorennamen), quantitative Größen (z.B. Anzahl der Seiten), (semi-) professionelle Kritiken und, in den Zeiten des Web 2.0, Nutzerbewertungen.

Unsere Forschungsarbeit, die sich mit dem Thema „Analyse von literarischen Texten mit Methoden aus den Bereichen Sentiment Analysis und Opinion Mining“ beschäftigt, knüpft an diesem Punkt an. Durch verschiedene Untersuchungen und Experimente, z.B. mit Wortlisten oder Methoden des maschinellen Lernen, sollen Emotionen und Stimmungen in literarischen Werken automatisch identifiziert und entsprechend verarbeitet bzw. visualisiert werden. Somit werden die momentan verfügbaren Informationen zu Texten um eine neue Dimension erweitert, wodurch eine neue Möglichkeit zur Suche entsteht, eine Suche auf Basis von Stimmungen und Emotionen.

Ziel dieser laufenden Arbeit, die einen Teil der erwähnten Forschungsarbeit darstellt, ist einerseits die generelle Evaluierung bereits vorhandener Typen von Wortlisten für die Verwendung zur Stimmungserkennung in literarischen Texten. Andererseits soll eine Methode zur Erzeugung eines geeigneten Referenzkorpus getestet und dieser Korpus anschließend auf Plausibilität geprüft werden. Dazu wird ein Testkorpus, bestehend aus verschiedenen Märchen der Brüder Grimm, zusammengestellt, entsprechend vorverarbeitet und anschließend von zwei voneinander unabhängigen Personen manuell klassifiziert. Nach Prüfung der Plausibilität (Übereinstimmung), durch Berechnung des statistischen Maßes *Cohens Kappa*, wird einer der beiden manuell klassifizierten Korpora als Referenz ausgewählt. Anschließend wird der ursprüngliche Testkorpus von drei ausgewählten Wortlisten automatisch klassifiziert und die Ergebnisse werden mit dem Referenzkorpus verglichen.

## 2 Verwandte Arbeiten

In den vergangenen Jahren wurde in den Bereichen *Opinion Mining* und *Sentiment Analysis* intensive Forschungsarbeit geleistet. Eine gute Übersicht über die verschiedenen Themen und Forschungsbereiche gibt die Veröffentlichung von Liu und Zhang [Liu and Zhang, 2012]. Ergänzend sind die beiden, erst kürzlich erschienenen, Artikel von Feldman [Feldman, 2013] und Cambria und Kollegen [Cambria et al., 2013] zu nennen, die ebenfalls einen aktuellen Überblick über dieses Forschungsfeld liefern.

In der Veröffentlichung von Mohammad und Turney wird die Erstellung eines englischsprachigen Lexikons mit emotionstragenden Unigrammen (*NRC Emotion Lexicon*) beschrieben [Mohammad and Turney, 2010] und dessen Anwendung unter anderem anhand von Versuchen mit den ins Englisch übersetzten Märchen der Brüder Grimm gezeigt [Mohammad, 2011]. In einem aktuelleren Artikel [Mohammad, 2012] führt Mohammad dieses Thema weiter aus und beschreibt die Suche nach Emotionen in E-Mails und Büchern.

Klenner präsentiert in seiner Arbeit [Klenner, 2009] einen regelbasierten Ansatz zur Sentimentanalyse für die deutsche Sprache und evaluiert diesen anhand eines literarischen Textes.

Neben den in der Arbeit behandelten Lexika *Sentiment Phrase List (SePL)* [Rill et al., 2012a] und *SentimentWortschatz (SentiWS)* [Remus et al., 2010], existieren zudem noch weitere lexikalische Ressourcen für die deutsche Sprache. Dabei handelt es sich zum einen um das *Polarity Lexicon* [Clematide and Klenner, 2010] mit ca. 8.000 meinungstragenden Worten, zum anderen um *GermanPolar-*

tyClues [Waltinger, 2010] mit mehr als 10.000 Einträgen. Beide Ressourcen beinhalten Unigramme sowie einen dazugehörigen Meinungswert.

### 3 Testkorpus und lexikalische Ressourcen

#### 3.1 Erstellung eines geeigneten Testkorpus

Bei der Auswahl geeigneter Texte für die Durchführung der Versuche entschieden wir uns für die Märchen der Brüder Grimm<sup>1</sup>. Diese literarischen Texte kommen aus den folgenden Gründen zum Einsatz. Zum einen sind es kurze Texte, die schnell gelesen und klassifiziert werden können. Außerdem sind sie leicht verständlich und weisen eine klare Struktur auf, wodurch sie für erste Experimente bestens geeignet scheinen. Des Weiteren beinhalten sie eine Vielzahl verschiedener Emotionen und Figuren, die größtenteils stereotypisch handeln und somit in Klassen eingeteilt werden können. Ein letzter und wichtiger Punkt der für den Einsatz dieser Märchen spricht, ist die Verwendung in anderen Forschungsarbeiten (siehe z.B. [Mohammad, 2011]). Dies erlaubt es weitere, beispielsweise sprachübergreifende, Vergleiche durchzuführen.

Um die Texte in digitaler Form zu erhalten, wurde ein Werkzeug entwickelt, mit dessen Hilfe es möglich war, den gesamten Korpus automatisch von der Website des Projekt Gutenberg-DE<sup>2</sup> zu beziehen. Da der Märchenkorpus insgesamt aus 194 Texten besteht, eine Menge die ohne großen Zeitaufwand nicht gelesen und manuell klassifiziert werden kann, wurden daraus 20 Texte ausgewählt. Diese Texte wurden anschließend manuell so aufbereitet, dass jeder Satz bzw. jede zusammengehörende Konstellation auf einer eigenen Textzeile stand. Diese Vorarbeit erleichterte einerseits den Taggern die Identifikation und manuelle Klassifikation der Sätze und ermöglichte andererseits die reibungslose Durchführung der automatischen Klassifikation auf Basis der Wortlisten. Die Vorverarbeitung der 20 Texte wurde vollständig manuell durchgeführt, da erste Versuche mit automatischen Methoden zeigten, dass die strikte Satztrennung durch Pattern, wie z.B. Satzzeichen, nicht zufriedenstellend funktionierte. Ein Problem waren dabei fehlende Satzzeichen innerhalb der Texte. Des Weiteren wurden durch die automatische Satztrennung Textfragmente, die logisch zusammengehören, getrennt (z.B. „Soll die dumme Gans bei uns in der Stube sitzen!“ sprachen sie.“).

#### 3.2 Verwendete Ressourcen

Für die Versuche verwenden wir drei lexikalische Ressourcen, die unterschiedliche Strukturen und Einträge aufweisen und somit auf verschiedene Weise auf die Daten des Testkorpus angewendet werden müssen (siehe Kapitel 4.2). Eine dieser Ressourcen enthält zudem ausschließlich Einträge in englischer Sprache, wodurch eine Vorverarbeitung in Form einer Übersetzung notwendig wird.

Die *Sentiment Phrase List* (SePL) enthält in der aktuellen Version 1.1 15.142 meinungstragende Adjektiv- und Substantiv-Phrasen mit einer Phrasenlänge bis zu  $N = 5$ . Jeder Eintrag der Liste enthält zudem einen Wert (Opinion Value), der die Meinung der Phrase auf einer stetigen Skala im Bereich  $[-1, +1]$  widerspiegelt. Die Liste beinhaltet ausschließlich Phrasen in deutscher Sprache, die Einträge sind zudem lemmatisiert. Der generische Ansatz zur sprachunabhängigen Erstellung einer solchen Liste wurde

<sup>1</sup>Brüder Grimm: Die schönsten Kinder- und Hausmärchen (<http://gutenberg.spiegel.de/buch/6248/1>)

<sup>2</sup><http://www.projekt.gutenberg.de/>

in der Veröffentlichung von Rill und Kollegen [Rill *et al.*, 2012b] vorgestellt.

Das *NRC Emotion Lexicon* enthält 141.820 Einträge mit englischsprachigen Unigrammen (Version 0.92). Dabei existieren jedoch pro eindeutigem Unigramm zehn Einträge, die binär die Zugehörigkeit zu [keiner, einer, mehreren] der 8 Basisemotionen [Plutchik, 1980] sowie die Zuordnung zu „positiv“ oder „negativ“ anzeigen. Dadurch verringert sich die tatsächliche Anzahl zunächst auf 14.182 Wörter, wobei nicht jedem Wort ein Wert zugeordnet ist.

Um diese Ressource für die Versuche verwenden zu können, musste sie entsprechend vorbereitet werden. Dazu wurden im ersten Schritt alle Wörter entfernt, die keine Zuordnung zu „positiv“ oder „negativ“ hatten. Im zweiten Schritt wurde die, nun deutlich kleinere, Liste automatisch mit Hilfe der Google Translate API<sup>3</sup> übersetzt. Dabei kam ein einfacher „first-best“ - Ansatz zum Einsatz, d.h. es wurde immer der jeweils erste Übersetzungsvorschlag übernommen. Einige Unigramme konnten nicht automatisch übersetzt werden, wodurch sich die Größe der Liste auf letztendlich 5.255 Einträge verringerte.

Tabelle 1 zeigt die schrittweise Selektion durch die verschiedenen Verarbeitungsschritte und die entsprechende Anzahl der Einträge, die den jeweiligen Schritt passieren konnten.

Selektionsschritte	Anzahl der Einträge	
	Absolut	Relativ
Vollständige Liste	141.820	-
Eindeutige Wörter	14.182	100%
Wörter mit Zuordnung	5.555	39,17%
Übersetzte Wörter	5.255	37,05%

Tabelle 1: Übersicht der Selektionsschritte beim *NRC Emotion Lexicon*.

Die Ressource *SentiWS* enthält in der verwendeten Version 1.8c 3.468 meinungstragende deutsche Unigramme in ihrer Grundform mit einem dazugehörigen Meinungswert auf einer stetigen Skala im Bereich  $[-1, +1]$ . Eine Besonderheit dieser Liste ist die generelle Separation positiver und negativer Wörter durch die Auflistung in zwei verschiedenen Sublisten. Zu allen Einträgen werden zudem Part-of-Speech Tags sowie, falls existent, verschiedene Beugungsformen gelistet. Diese Ressource enthält neben Adjektiven und Adverbien auch meinungstragende Verben und Substantive.

### 4 Experimente

Im Folgenden wird die manuelle Klassifikation von Sätzen zur Erzeugung des Referenzkorpus sowie die automatische Klassifikation auf Basis der Wortlisten beschrieben. Bei letzterem werden ebenfalls kurz die notwendigen Schritte der Vorverarbeitungen der Märchen (Testkorpus), die Annahmen für Wortlisten und die Algorithmen zur Erkennung bzw. Extraktion beschrieben. Am Ende des Kapitels wird schließlich der für die automatische Klassifikation der Sätze zuständige Algorithmus erläutert.

#### 4.1 Manuelle Klassifikation von Sätzen

Um eine Bewertungsgrundlage für die Ergebnisse der späteren automatischen Klassifikation zu erhalten, mussten alle Texte des Testkorpus manuell klassifiziert werden.

<sup>3</sup><https://developers.google.com/translate/>

Dafür wurden zwei voneinander unabhängige Personen beauftragt. Diese Personen, im Folgenden Tagger 1 (T1) und Tagger 2 (T2) genannt, wurden nicht explizit über die Arbeitsweise der Algorithmen oder den Einsatz von Wortlisten informiert. Beide Tagger erhielten alle Texte des Testkorpus sowie den Auftrag emotions- bzw. stimmungstragende Sätze mit entsprechenden Tags zu markieren. Sätze mit positiven Emotionen sollten mit dem Tag „<p>“, Sätze mit negativen Emotionen mit Tag „<n>“ am Zeilenende versehen werden. Sätze, die keine Emotionen tragen oder in denen sich positive und negative Stimmungen gegenseitig „aufheben“, sollten nicht markiert werden. Insgesamt sollte jeder der beiden Tagger 1.604 Sätze lesen und klassifizieren.

Zur Überprüfung der Übereinstimmung der klassifizierten Korpora von Tagger 1 und Tagger 2 wurde für jeden Text das statistische Maß *Cohens Kappa* berechnet. Dabei stellten wir fest, dass die Übereinstimmung bei 70% der Texte zufriedenstellend war ( $\kappa = 0,57$ ). Im gesamten Durchschnitt lag die Übereinstimmung bei  $\kappa = 0,51$ . Dieser Wert erscheint für Texte dieser Art durchaus akzeptabel. Werden nun ausschließlich Sätze betrachtet die von beiden Personen als positiv oder negativ klassifiziert wurden, ergibt sich eine Übereinstimmung von 87%.

Für die Evaluation der verschiedenen Wortlisten wurden die Ergebnisse von Tagger 1 als Referenz ausgewählt.

## 4.2 Automatische Klassifikation von Sätzen

Um den erstellten Testkorpus (siehe Kapitel 3.1) für die automatische Klassifikation verwenden zu können, sind einige Vorverarbeitungsschritte notwendig. Diese Schritte sowie die verwendeten Algorithmen zur Extraktion stimmungstragender Wörter und Phrasen werden in Abbildung 1 schematisch dargestellt.

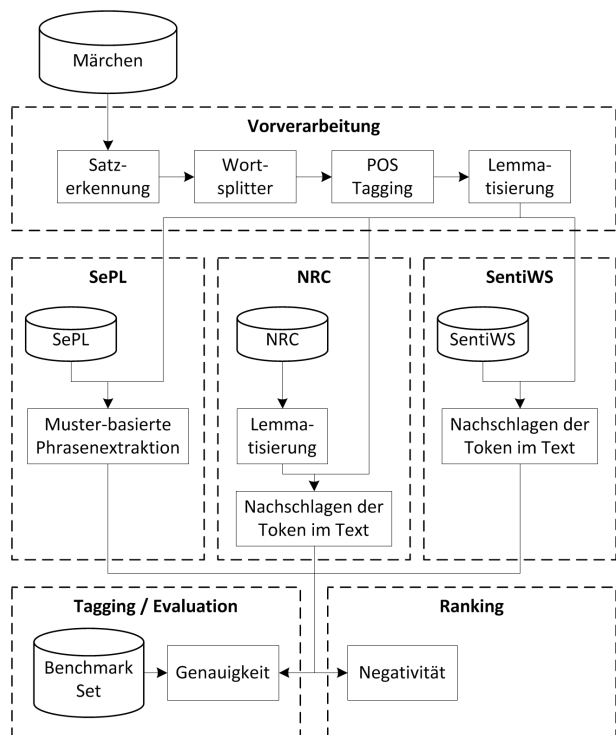


Abbildung 1: Schematische Darstellung des Ablaufs der Experimente.

## Vorverarbeitung

Im ersten Schritt müssen alle Sätze identifiziert und separiert werden. Dieser Schritt wurde bereits manuell durchgeführt, wodurch ein simpler Algorithmus, der das jeweilige Zeilenende in Texten erkennt an dieser Stelle ausreichend ist. Anschließend werden Apache OpenNLP<sup>4</sup> Tokenizer sowie Apache OpenNLP Part-of-Speech Tagger angewendet, um einzelne Wörter zu extrahieren und mit einem korrekten POS-Tag zu versehen. Zum Schluss werden alle Wörter lemmatisiert. Dazu wird das auf Morphy<sup>5</sup> [Lezius *et al.*, 1998] basierende deutsche Morphologie-Lexikon<sup>6</sup> verwendet.

## Sentiment Phrase List

Für die Anwendung dieser Liste, die größtenteils Phrasen enthält, wird ein musterbasierter Algorithmus für die Extraktion meinungstragender Wörter und Phrasen verwendet. Dazu werden die Texte nach bestimmten Adjektivphrasen ([adverbiales Adjektiv | Adverb | Partikel | Verb im Partizip Perfekt | unbestimmtes Pronomen] & Adjektiv) und Substantivphrasen ([Adjektiv | Adverb | unbestimmtes Pronomen] & Substantiv) durchsucht und anschließend mit den 15.142 Einträgen der Wortliste verglichen. Für die anschließende Klassifizierung der Sätze in „positiv“ und „negativ“ wurden bestimmte Schwellenwerte für Phrasen festgelegt. Phrasen mit einem Opinion Value ( $OV \leq -0,3$ ) werden negativ gewertet. Phrasen mit einem  $OV \geq 0,3$  werden positiv gewertet. Alle Phrasen mit einem Opinion Value zwischen diesen beiden Schwellenwerten werden als neutral angesehen.

## NRC Emotion Lexicon

Bei dieser übersetzten Wortliste, die größtenteils Unigramme enthält, kommt ein einfacher „Lookup - Algorithmus“ zum Einsatz. Das bedeutet, dass im Text nach einzelnen NRC Token gesucht wird und diese entsprechend ihrer binären Zuordnung bewertet werden. Um die Trefferquote zu erhöhen werden die Wörter dieser Liste zudem noch lemmatisiert.

## SentimentWortschatz

Da diese Liste ebenfalls ausschließlich Unigramme enthält, die zudem bereits in ihrer Grundform vorhanden sind, wird wiederum ein „Lookup - Algorithmus“ verwendet. Für die Einteilung der Unigramme in die Klassen „positiv“ und „negativ“ wird die bereits vorhandene Vorgabe durch die Liste verwendet (siehe Kapitel 3.2).

## Algorithmus zur Klassifizierung der Sätze

Durch die entsprechenden Vorarbeiten und Festlegungen liefern alle Listen für stimmungstragende Wörter oder Phrasen eine Klassifizierung in „positiv“ und „negativ“. Um damit Sätze des Testkorpus klassifizieren zu können, wird ein einfacher Algorithmus verwendet, der die Anzahl gefundener positiver und negativer Wörter und Phrasen pro Satz zählt und auf Basis der Mehrheit eine Entscheidung trifft. Ist beispielsweise die Anzahl der negativ klassifizierten Wörter in einem Satz größer als die Anzahl positiv klassifizierter Wörter, wird dieser Satz „negativ“ klassifiziert. Sollten in einem Satz gleich viele positive und negative Phrasen vorkommen wird dieser neutral eingestuft. Durch diese simple Methode wird die Qualität der Listen, im Bezug auf die Erkennung stimmungstragender Wörter in Texten dieser Art, direkt miteinander vergleichbar.

<sup>4</sup><http://opennlp.apache.org/>

<sup>5</sup><http://www.wolfganglezius.de/doku.php?id=cl:morphy>

<sup>6</sup><http://www.danielnaber.de/morphologie/>

## 5 Erste Ergebnisse

Tabelle 2 zeigt einen direkten Vergleich der relativen Häufigkeit von positiv oder negativ klassifizierten Sätzen im Verhältnis zu allen Sätzen pro Text. Verglichen werden die Ergebnisse von Tagger 1 mit den Ergebnissen der drei Ressourcen, wobei die Tabelle nach den Ergebnissen von Tagger 1 absteigend sortiert ist. Es zeigt sich, dass die beiden Wortlisten *Sentiment Phrase List* (mit 30%) und *SentimentWortschatz* (mit 37%) deutlich weniger Sätze eindeutig zuordnen können als Tagger 1. Die Ergebnisse des *NRC Emotion Lexicon* liegen mit durchschnittlich 60% eindeutig klassifizierter Sätze nur knapp hinter denen von Tagger 1 (68%). Auch wird deutlich, dass die Ergebnisse dieser Ressource erstaunlich stabil sind.

Märchen	Relative Häufigkeiten			
	T1	SePL	NRC	SentiWS
Drei Spinnerinnen	0,89	0,43	0,68	0,43
Wasser des Lebens	0,85	0,22	0,66	0,40
Hund und Sperling	0,84	0,24	0,65	0,20
Geschenke d. kl. V.	0,83	0,60	0,87	0,50
Hans mein Igel	0,79	0,31	0,78	0,31
Aschenputtel	0,79	0,22	0,73	0,38
Zwei Brüder	0,74	0,20	0,76	0,31
Zaunkönig und Bär	0,74	0,21	0,74	0,39
Zwölf Brüder	0,70	0,29	0,78	0,46
Nixe im Teich	0,69	0,39	0,78	0,43
Schneew. und RR	0,68	0,31	0,75	0,46
Das Rätsel	0,68	0,23	0,77	0,38
Eisenofen	0,68	0,37	0,67	0,42
Geist im Glas	0,57	0,23	0,80	0,25
Gevatter Tod	0,55	0,27	0,58	0,32
Jorinde / Joringel	0,55	0,30	0,64	0,43
Hans im Glück	0,53	0,38	0,75	0,29
Rotkäppchen	0,51	0,23	0,59	0,38
Spindel, Webers. ...	0,48	0,33	0,74	0,33
Die klugen Leute	0,47	0,17	0,68	0,29

Tabelle 2: Vergleich der relativen Häufigkeit positiv / negativ klassifizierter Sätze im Vergleich zu allen Sätzen.

Um die Ergebnisse der Listen vergleichen und bewerten zu können, wurden pro Text drei Negativitätswerte berechnet. Der erste Wert beschreibt das Verhältnis von negativ klassifizierten Sätzen im Vergleich zu allen Sätzen ( $neg/all$ ), der zweite Wert das Verhältnis von negativ klassifizierten Sätzen im Vergleich zu positiv klassifizierten Sätzen ( $neg/pos$ ). Der letzte Wert gibt schließlich das Verhältnis von negativ klassifizierten Sätzen im Vergleich zu allen positiv oder negativ klassifizierten Sätzen an ( $neg/(neg + pos)$ ).

Tabelle 3 zeigt die berechneten Korrelationskoeffizienten der Ergebnisse der verschiedenen Negativitätswerte. Dabei wurde jeweils die Korrelation zwischen den Ergebnissen des Referenzkorpus (Tagger 1) mit denen der einzelnen Listen untersucht. Es wird deutlich, dass die Ergebnisse aller drei Listen mit denen des Referenzkorpus für alle Negativitätswerte korrelieren.

Durch den Einsatz von lediglich 20 Referenztexten ist diese Korrelation allerdings instabil. Werden die Ergebnisse eines einzigen Textes aus der Korrelationsberechnung entfernt, verändern sich die Korrelationskoeffizienten für alle Negativitätswerte bereits sehr stark. Besonders das Fehlen von extrem unilateral klassifizierten Texten führt

Kandidaten	Korrelationskoeffizient		
	$neg/all$	$neg/pos$	$neg/(neg + pos)$
T1-SePL	0,49	0,88	0,52
T1-NRC	0,63	0,77	0,76
T1-SentiWS	0,45	0,83	0,56

Tabelle 3: Korrelationskoeffizienten der Ergebnisse.

dabei zu einer deutlichen Verschlechterung der Korrelation. Eine Korrelation der Ergebnisse der *Sentiment Phrase List* mit dem Referenzkorpus ist in diesem Fall nicht mehr erkennbar und auch die Korrelation zwischen den Ergebnissen des *SentimentWortschatz* und der Referenz nimmt stark ab (mit Ausnahme der Ergebnisse von  $neg/all$ ). Lediglich die Ergebnisse des *NRC Emotion Lexicon* bleiben stabil und zeigen weiterhin eine Korrelation.

## 6 Zusammenfassung und Ausblick

In dieser laufenden Arbeit wurden Experimente mit drei verschiedenen lexikalischen Ressourcen mit dem Ziel durchgeführt, Aussagen über deren Qualität beim Einsatz zur Stimmungserkennung in literarischen Texten zu treffen. Dazu wurde ein Testkorpus erstellt, der 20 ausgewählte Märchen der Brüder Grimm beinhaltet. Alle Sätze in diesem Korpus wurden manuell von zwei voneinander unabhängigen Personen klassifiziert. Anschließend wurde ein Referenzkorpus erstellt. Zudem wurden alle Sätze des Testkorpus, basierend auf den Daten der drei Wortlisten, automatisch klassifiziert. Die Resultate der automatischen Klassifikation wurden mit denen des Referenzkorpus verglichen und erste Ergebnisse präsentiert.

Es zeigte sich, dass lexikalische Ressourcen durchaus dazu geeignet sind, Emotionen und Stimmungen in literarischen Texten zu erkennen. Die Erkennungsraten variieren jedoch von Liste zu Liste, sodass genauere Analysen erforderlich sind. Es wurde ebenfalls deutlich, dass der verwendete Testkorpus mit 20 Märchen deutlich zu klein dimensioniert war, um verlässliche Ergebnisse zu erhalten.

Die nächsten Schritte werden somit zum einen die Generierung eines größeren Referenzsets und zum anderen eine detaillierte Analyse der automatisch klassifizierten Daten sein. Dabei stehen vor allem Analysen zur Güte der automatischen Erkennung im Vordergrund (Precision und Recall) und basierend darauf die Suche und anschließende Auswertung von möglichen Problemen und Fehlerquellen.

## Literatur

- [Cambria *et al.*, 2013] Erik Cambria, Bjorn Schuller, Yunqing Xia, and Catherine Havasi. New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Intelligent Systems*, 28(2):15–21, 2013.
- [Clematide and Klenner, 2010] Simon Clematide and Manfred Klenner. Evaluation and Extension of a Polarity Lexicon for German. In *Proceedings of the 1st Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, pages 7–13, 2010.
- [Feldman, 2013] Ronen Feldman. Techniques and Applications for Sentiment Analysis. *Communications of the ACM*, 56(4):82–89, 2013.
- [Klenner, 2009] M Klenner. Süsse Beklommenheit und schmerzvolle Ekstase: Automatische Sentimentanalyse in den Werken von Eduard von Keyserling. In *Von*

*der Form zur Bedeutung: Texte automatisch verarbeiten (Proceedings of the Biennial GSCL Conference 2009)*, pages 91–97. Narr Francke Attempto Verlag GmbH, 2009.

- [Lezius *et al.*, 1998] Wolfgang Lezius, Reinhard Rapp, and Manfred Wettler. A Freely Available Morphological Analyzer, Disambiguator and Context Sensitive Lemmatizer for German. In *Proceedings of the 17th international conference on Computational linguistics - Volume 2*, pages 743–748, 1998.
- [Liu and Zhang, 2012] Bing Liu and Lei Zhang. A Survey of Opinion Mining and Sentiment Analysis. In Charu C. Aggarwal and ChengXiang Zhai, editors, *Mining Text Data*, pages 415–463. Springer US, 2012.
- [Mohammad and Turney, 2010] Saif M. Mohammad and Peter D. Turney. Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, CAAGET '10, pages 26–34, 2010.
- [Mohammad, 2011] Saif Mohammad. From Once Upon a Time to Happily Ever After: Tracking Emotions in Novels and Fairy Tales. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 105–114, 2011.
- [Mohammad, 2012] Saif M. Mohammad. From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems*, 53(4):730 – 741, 2012.
- [Plutchik, 1980] R. Plutchik. *A general psychoevolutionary theory of emotion*, pages 3–33. Academic press, 1980.
- [Remus *et al.*, 2010] Robert Remus, Uwe Quasthoff, and Gerhard Heyer. SentiWS – a Publicly Available German-language Resource for Sentiment Analysis. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pages 1168–1171, 2010.
- [Rill *et al.*, 2012a] Sven Rill, Sven Adolph, Johannes Drescher, Dirk Reinel, Jörg Scheidt, Oliver Schütz, Florian Wogenstein, Roberto V. Zicari, and Nikolaos Korfiatis. A Phrase-Based Opinion List for the German Language. In *Proceedings of KONVENS 2012*, pages 305–313, 2012.
- [Rill *et al.*, 2012b] Sven Rill, Johannes Drescher, Dirk Reinel, Jörg Scheidt, Oliver Schütz, Florian Wogenstein, and Daniel Simon. A Generic Approach to Generate Opinion Lists of Phrases for Opinion Mining Applications. In *Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM)*, 2012.
- [Waltinger, 2010] Ulli Waltinger. GermanPolarityClues: A Lexical Resource for German Sentiment Analysis. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, pages 1638–1642, 2010.