

Enhanced WordNet: Extraktion kategorieübergreifender semantischer Relationen aus WordNets Glossen

Tonio Weidler

Department of Computational Linguistics
69120 Heidelberg, Germany
weidler@cl.uni-heidelberg.de

Abstract

WordNets Glossen enthalten semantische Informationen über die Synsets, die von ihnen definiert werden. Diese umfassen neben sehr speziellen Relationen auch allgemeine Eigenschaften wie Attribute und Funktionen von Substantiven. Trotzdem wurden sie bis dato noch nicht als Relationen in WordNet integriert. In dieser Arbeit untersuche ich die automatische Extrahierbarkeit solcher Informationen als Relationen und evaluiere ihren Nutzen am Beispiel der Koreferenzresolution.

1 Einleitung

WordNet (Miller u. a., 1993) ist eine der bekanntesten elektronischen Lexikographien in der Computerlinguistik. Ursprünglich als Werkzeug zur Überprüfung psycholinguistischer Theorien gedacht, fand WordNet schnell Verwendung in diversen NLP Anwendungen. Nicht zuletzt liegt dies an der hohen Qualität der komplett per Hand erstellten Ressource, die trotzdem einen großen Teil des englischen Wortschatzes abdeckt.

Im Gegensatz zu vielen anderen Lexikographien enthält WordNet keine Einträge zu Wörtern im eigentlichen Sinne, sondern vielmehr zu sogenannten *Synsets*. Diese repräsentieren ein *Konzept*, das von mehreren Wörtern (*Synonyme*) beschrieben werden kann. In WordNet werden diese Synsets durch semantische Relationen innerhalb ihrer Wortklassen (Substantive, Verben, Adjektive, Adverbien) verbunden. Jede Wortart hat dabei ein Set von möglichen Relationen, das sich in einigen Punkten von denen der anderen Klassen unterscheidet. Für Substantive finden sich unter anderem *Hypernym/Hyperonym* Beziehungen, *Antonyme* und *Meronyme*, Verben gliedern sich in flacheren *Hyperonymstrukturen* und Adjektive organisieren sich in Clustern aus ähnlichen Begriffen (*similar to*) um *Antonympaare*.

Auch wenn WordNet seit Version 1.7 einige Relationen zwischen den Wortklassen enthält, sind diese nicht sehr ergiebig. Sie umfassen bisher eine Relation der morphologischen Abstammung (z.B. *to walk* → *a walk*) an der Substantive und Verben teilhaben können sowie Relationen zur Domain von Synsets zwischen allen Wortklassen. Neben diesen sind viele weitere Relationen denkbar, die zwischen Verben, Substantiven und Adjektiven gefunden werden können. Bereits Miller (1993) merkte an, dass zur eindeutigen Unter-

scheidung von Nomen mindestens drei *Eigenschaften* des Konzepts definiert werden müssen, seine *Attribute* (z.B. *bird* → *warm-blooded*, *egg-laying*, ...), seine *Bestandteile (parts)* (z.B. *table* → *leg*, *tabletop*, *tableware*) und seine *Funktion* (z.B. *vehicle* → *transport*). Abgesehen von den Bestandteilen eines Konzepts, die als Meronym-Relation realisiert wurden, finden sich diese Eigenschaften bis heute nicht in WordNet Relationen wieder.

Vielmehr wurden sie im Nachhinein durch sogenannte *Glossen* teilweise verwirklicht, natürlichsprachliche Beschreibungen der Konzepte, die sie von anderen Konzepten unterscheiden sollen. Diese Glossen sind jedoch nicht mit den Synsets annotiert, die sie enthalten und weisen auch sonst keine Merkmale auf, die eine computerlinguistische Verwendung vereinfachen würden.

Ziel dieser Arbeit ist es deshalb, die Möglichkeit der automatisierten Extraktion von Relationen zwischen Wortklassen aus diesen Glossen zu untersuchen und den Nutzen derselben anhand einer beispielhaften NLP Anwendung zu evaluieren. Ich werde dazu zunächst verwandte Arbeiten vorstellen (Abschnitt 2) und dann auf die nötige Vorverarbeitung der Glossen eingehen (Abschnitt 3). Daraufhin stelle ich die neuen Relationen vor und vertiefe den Extraktionsvorgang (Abschnitt 4). Die Ergebnisse der Extraktion fasse ich in Abschnitt 5 zusammen und diskutiere sie auf Grundlage einer Evaluation am Beispiel der Koreferenzresolution (Abschnitt 6) in Abschnitt 7 um abschließend ein Fazit zu ziehen (Abschnitt 8).

2 Verwandte Arbeiten

Die Glossen in WordNet standen bereits in anderen Forschungsarbeiten im Mittelpunkt. Besonders bekannt geworden ist *Extended Wordnet* (Harabagiu, Miller und Moldovan, 1999). In diesem Projekt wurden die Glossen zunächst disambiguiert, um eine bessere Vernetzung zu gewähren. In einem anschließenden Schritt haben die Autoren die Glossen erst in eine logische und darauf aufbauend semantische Form transformiert. Die gewonnenen Informationen konnten in anderen Anwendungen verwandt werden, zum Beispiel für das Question Answering System von Moldovan und Novischi (2002).

In sogenannten Stand-Off Files werden neben WordNet noch weitere Dateien bereitgestellt. Diese

umfassen beispielsweise den "Princeton Annotated Gloss Corpus", eine teilweise Disambiguierung der Glossen in WordNet auf der auch das System dieser Arbeit aufbaut. Die Glossen sind zudem in einer an der USC/ISI (2006) transformierten logischen Form verfügbar.

3 Parsing

Das System dieser Arbeit, EhWoN (**E**nhanced **W**ord**N**et), adaptiert, integriert und reimplementiert Teile dieser Vorverarbeitung der Glossen. Zunächst wurden die Wörter jeder Glosse disambiguiert, damit die Teilnehmer der extrahierten Relationen im Nachhinein wieder den Synsets in WordNet zugeordnet werden konnten. Eine syntaktische und semantische Transformation der Glossen in eine logische Form hat zudem ermöglicht, komplexere Konstrukte zu nutzen, als sie mit z.B. simplen POS-Patterns realisierbar gewesen wären.

3.1 Disambiguierung der Glossen

Zwar wurden die Wörter der Glossen in der offiziellen Version von WordNet nicht mit den Synsets verlinkt, für die sie stehen. Allerdings wurde mit dem *Princeton Annotated Gloss Corpus* eine Sammlung an XML-Dateien bereitgestellt, in denen alle WordNet Glossen mit Part-of-Speech Tags annotiert und die in WordNet zu findenden Tokens als solche markiert wurden. Ein Teil dieser Tokens wurde zudem manuell oder automatisch disambiguiert (etwa 68%). Diese Annotationen habe ich für EhWoN extrahiert und mit einer simplen *most-common-sense* Disambiguierung der verbleibenden Tokens kombiniert.

3.2 Logic Form Transformation

Um die Glossen in eine logische Form zu übertragen bot sich die Verwendung von EasySRL (Lewis, He und Zettlemoyer, 2015) an, einem State-of-the-Art System, das aus englischen Sätzen logische Repräsentationen bildet. Hierzu bedient es sich des CCG Formalismus und kombiniert diesen mit der automatisierten Annotation von semantischen Rollen. Es entstehen logische Representationen, die Informationen zur Argumentstruktur der Prädikate enthalten. EhWoN verwendet EasySRL, um die definierenden Subglossen¹ jeder Glosse in eine logische Form zu transformieren, die daraufhin in ein weiterverarbeitbares Array umgewandelt werden kann. Die Glosse für eines der Synsets von *sparrow* lautet beispielsweise

small brownish European songbird (1)

und wird von EasySRL in die folgende logische Form übertragen:

sk(#x.(songbird(x) & european(x) & brownish(x) & small(x))) (2)

¹einige Glossen enthalten mehrere Teilglossen

Aus dieser Repräsentation lassen sich auch für komplexere Strukturen einfach Relationen ableiten.

4 Relationen

Der Aufbau von WordNets Glossen folgt keiner einheitlichen Struktur. Deshalb ließen sich eine Vielzahl verschiedener Relationen extrahieren, die teils sehr speziell wären und deshalb auch nur sehr spärlich extrahiert werden könnten. Die für EhWoN konzipierten Relationen fokussieren sich deshalb zum einen auf Relationen zwischen Synsets *verschiedener* Wortklassen. Zum anderen habe ich vermieden, zu spezielle Relationen zu formulieren, für die kaum Beispiele in den Glossen zu finden gewesen wären.

Zunächst wurden die bereits von Miller (1993) vorgeschlagenen *Eigenschaften* der Substantive extrahiert, die *Funktion* eines Konzepts und seine *Attribute*. Eine dritte Relation umfasst adjektivische *Spezifizierungen von Verben*.

Bei allen extrahierten neuen Relationen handelt es sich um einseitige Beziehungen. Die Relationen sind meist relative Spezifizierungen mit Bezug auf ein Hyperonym, weshalb die definierten Synsets nicht unbedingt gute Beispiele für ihre Spezifikatoren (wie *Attribute* und *Funktionen*) sind, wenn der Kontext des Hyperonyms fehlt. Somit sind sie im Gegensatz zu den meisten Relationen des originalen WordNets auch nicht vererbbar. Ein *Baum* mag eine *große* Pflanze sein, dieses Attribut scheint aber nicht mehr angebracht, wenn von einem *Bonsai* (Hyponym von Baum) die Rede ist.

Da es sich bei den Glossen vielmehr um Satzfragmente handelt als um ganze Sätze (siehe Beispiel (1) im Kontrast zu "A sparrow is a small..."), muss für jede Glosse zunächst die Hauptentität in der logischen Form festgestellt werden. Aufgrund des in dieser Hinsicht relativ einheitlichen Aufbaus der Glossen genügt hier die Heuristik, jene Konstante als Repräsentator des Synsets zu wählen, die als Erstes deklariert wird.

Im Folgenden sollen alle extrahierten Relationen und die Patterns, nach denen sie gefunden wurden, detaillierter diskutiert werden.

4.1 Attribute (Substantiv-Adjektiv)

Substantive werden in zahlreichen Glossen zunächst von ihren Hyperonymen abgegrenzt, indem sie mithilfe von Adjektiven spezifiziert werden. Diese Adjektive gehen eine *Attribut-Relation* mit dem Substantiv ein. In Beispiel (1) finden sich drei solcher Attribute, *brown*, *small* und *european*.

Pattern. Sie lassen sich aus den logischen Formen extrahieren, indem all jene einstelligen Adjektiv-Prädikate als Attribut verwendet werden, deren Leerstelle durch die Hauptentität ausgefüllt wird.

4.2 Funktion (Substantiv-Verb)

Die Funktion eines Substantives ist eine Aktion, die von ihm typischerweise ausgeführt wird, für die es benutzt wird oder zu deren Zweck es gedacht ist. Eine klare Definition ist jedoch häufig schwierig. Während die Funktion eines Fahrzeuges relativ klar als *transportieren* oder *fahren* bezeichnet werden kann, ist es zumindest diskutabel, ob man bei Vögeln von *Singen* als deren Funktion sprechen kann. Für die Definition dieser Relation soll dies jedoch so gesehen werden. Ein weiteres Problem der Relation ist die relativ offene Formulierung, die die semantische Rolle des Substantivs im Verb nicht festlegt. Um derartige Subrelationen der Funktion zu extrahieren, wäre allerdings zum einen eine zuverlässigere logische Transformation der Glossen von Nöten. Außerdem sind die Glossen nicht ergiebig genug, um speziellere Relationen in großen Mengen zu extrahieren.

Pattern. Die Glossen enthalten Informationen zur *Funktion* eines Substantives nach verschiedenen Mustern. Die meisten Funktionsverben sind die Prädikate einfacher *Events* in der logischen Form, an denen die Hauptentität als *ARG0* (agens) oder *ARG1* (patiens) teilnimmt. Einige Substantive werden auch mit *past* und *present participles* modifiziert, wie z.B. in "*a singing bird*". Auch die Verben dieser Partizipien werden dem Synset als Funktion hinzugefügt. Zuletzt finden sich noch Funktionen in Gerundiumkonstruktionen wie "*a whip for controlling horses*".

4.3 Verb Spezifikationen (Verb-Adjektiv)

Verb Spezifikationen umfassen alle Adverbien, die ein mögliches Hyperonym eines Verbs spezifizieren, um es von anderen Hyponymen abzugrenzen. Die Glosse zu *sprint* beispielsweise lautet *run very fast*. Das Hyperonym von *sprint*, *run*, wird durch die Verb Spezifikation (*very*) *fast* modifiziert um *sprint* eindeutig abzugrenzen.

Pattern. Aus den Transformationen lässt sich diese Relation auf zwei Arten extrahieren. Im ersten Fall treten sie schlicht als Prädikat des Events auf, das das Verb repräsentiert (ähnlich den Attributen). Die restlichen Verb Spezifikationen finden sich als Argument sogenannter *Modifier*. Diese können z.B. MNR (manner), TMP (temporal) oder DIR (directionals) sein und beschreiben die Art, wie das Verb modifiziert wird, genauer.

5 Extraktionsergebnisse

Bei der Extraktion dieser Relationen wurden für insgesamt 47 732 unterschiedliche Synsets mindestens eine der Relationen gefunden. Der Rest der insgesamt 82 103 Substantiv- und 13 764 Verbglossen, die transformierbar waren, enthält meist nur eine natürlichsprachliche Version der bereits bestehenden Relationen. Hierbei werden im Fall von Substantiven

Table 1: Extraktionsergebnisse der drei Relationen für 82 103 Substantive und 13 764 Verben

	Synsets mit Relation	Relation Gesamt	Durchschnitt pro Synset
Attribute	26096 (31,78%)	35948	0,44
Funktion	26843 (32,69%)	32966	0,4
Verb Spez.	1067 (7,75%)	1240	0,09

häufig statt des zu definierenden Synsets dessen Meronyme spezifiziert. Verbglossen sind leider häufig auf sehr verschiedenen Weisen konstruiert, die von EasySRL und den darauffolgenden Extraktionsalgorithmen nur schwer erkennbar sind.

Genauere Extraktionsergebnisse finden sich in Tabelle 1. *Attribute* und *Funktionen* ließen sich bei je etwa 30% der Substantivglossen finden. In Anbetracht der zuvor erwähnten Einschränkungen ist das ein relativ hoher Anteil. Verbspezifikationen finden sich etwas seltener (10%), was zu erwarten war. Für Verben ließen sich zahlreiche weitere ähnliche Relationen finden, die aber alle nicht viele Vertreter fänden.

Bei einer stichprobenartigen manuellen Überprüfung der Relationsergebnisse hat sich insbesondere die Attributrelation als sehr präzise herausgestellt, nicht zuletzt wegen des sehr einfachen Patterns. Selbiges gilt für die ähnlichen Verbspezifikationen. Die Extraktion der Funktionen von Substantiven hat wie zu erwarten Funktionen sehr verschiedener Art produziert. Neben den in der Formulierung der Relation erwarteten Unterschieden in der semantischen Rolle des Substantives für das Verb, wurden auch Verben bestimmter Konstrukte extrahiert, die sich eher nicht als Funktion beschreiben lassen. Im Besonderen sind dies Verben in Definitionen nach dem Muster *the action of ...* oder *the process of ...*.

6 Evaluation

Eine komplette intrinsische Evaluation der extrahierten Relationen würde eine sehr aufwendige manuelle Annotation der Glossen mit den Zielrelationen voraussetzen. Besser geeignet ist deshalb eine extrinsische Evaluation der gewonnenen Daten, die den Nutzen der Relationen für eine computerlinguistische Anwendung misst. Damit lässt sich zwar keine Aussage darüber machen, wie gut die extrahierten Daten die Informationen der Glossen wirklich widerspiegeln. Bei einem positiven Effekt auf die Performance eines anderen Systems durch die Integration der neuen Relationen ließe sich aber zumindest eine gute Qualität vermuten.

Eine mögliche Anwendung der *Attribut*- und *Funktion*-Relation ist die Koreferenzresolution. Während aktuelle Systeme hauptsächlich syntaktische und lexikalische sowie einige semantische Features verwenden (Durrett und Klein, 2013), möchte ich mit

den folgenden Beispielen darlegen, wie die hier extrahierten Relationen bei bestimmten Phänomenen in der Koreferenzresolution zusätzlicher Indikator für ein System sein können.

Die Beispielsätze (3) und (4) illustrieren, wie ein Koreferenzresolutionssystem die Attributrelation nutzen kann, um die markierten Anaphern zu ihren korrekten Antezedenten zuzuordnen.

The hunter and his dog track down **the elk**.
The large majestic animal runs as fast as possible. (3)

The jeweler showed her a ruby and **an emerald**. She chose **the green gemstone**. (4)

In Beispiel (3) ließe sich *The large majestic animal* allen drei Entitäten aus dem ersten Satz zuordnen. Selbst wenn das Koreferenzsystem bereits Hyperonym Relationen einbezieht, wären noch immer sowohl der Hunde (*dog*) als auch der Elch (*elk*) ein Tier (*animal*). Das System könnte nun aber die Attributrelation nutzen, um zu erkennen, dass ein Elch für sein Hyperonym eher groß ist. Dies wäre ein Indikator dafür, dass die Anapher dem möglichen Antezedenten *the elk* zugeordnet werden sollte.

Ebenso verhält es sich in Beispiel (4). Die Information, dass ein Smaragd (*emerald*) grün ist und ein Rubin (*ruby*) rot, kann dem System helfen, zu entscheiden, dass der grüne Edelstein (*green gemstone*) auf *emerald* verweist und nicht auf *ruby*.

In Beispiel (5) und (6) hingegen könnte das Koreferenzresolutionssystem mithilfe der Funktionsrelation eine bessere Entscheidung treffen.

A cat watches **a sparrow**. **It** began to **sing**. (5)

The cook placed **the meal** on the plate. (6)
 Then the customers **ate it**.

Die Information, dass ein Vogel typischerweise *singt*, im Gegensatz zu einer Katze, und eine Mahlzeit eher gegessen wird als ein Teller wäre ein guter Indikator, die Anaphern jeweils korrekt zuzordnen.

6.1 Cort

Um den Nutzen dieser beiden Relationen (Funktion und Attribute) für Koreferenzresolutionssysteme zu evaluieren, habe ich sie in **Cort** (Martschat und Strube, 2015) integriert. Cort ist ein Koreferenzresolutionstoolkit, das darauf ausgelegt ist, schnell Theorien zu überprüfen, indem es die Integration neuer Features erleichtert. Es basiert auf dem mention-pair Ansatz, nach dem jede Entscheidung, eine Anapher einem Antezedenten zuzuordnen, unabhängig von allen anderen Mentions betrachtet wird, und trainiert einen *Latent*

Structured Perceptron.

6.2 EhWoN Features

Insgesamt habe ich je eine Featuregruppe pro Relation entworfen, die je zwei Features enthält. Die beiden Features betrachten jeweils den selben Gesichtspunkt, einmal aus der Perspektive der Anapher, einmal aus der Perspektive des Antezedenten. Bei den folgenden Erläuterungen verwende ich deshalb die Begriffe Phrase und Paraphrase um die Mentions zu benennen. Dabei ist die Paraphrase eine potentielle Umschreibung der Phrase. Da nicht immer zu gewährleisten ist, dass die Texte, in denen Koreferenzresolution betrieben werden soll, mit WordNet Senses annotiert sind, verwenden die Features alle möglichen Synsets zu einem Lemma im Text.

6.2.1 Feature der Attributsrelation

Die erste Featuregruppe gibt an, ob eine Paraphrase die Phrase umschreibt, indem sie ein Hypernym (bis zu drei Ebenen tief) des Kopfes der Phrase mit mindestens einem Adjektiv modifiziert, das ein Attribut des Phrasenkopfes ist. Dieses Hypernym selbst muss wiederum der Kopf der Paraphrase sein. Dabei werden nicht nur die Attribute selbst betrachtet, sondern auch alle Synsets, die in einer *similar to* Beziehung zu den Synsets der Attribute stehen.²

6.2.2 Feature der Funktionsrelation

Die zweite Featuregruppe ermittelt, ob der Kopf der Phrase laut der Abhängigkeitsstruktur des Satzes in direkter Abhängigkeit zu einem Verb als Subjekt oder Objekt steht, das die Funktion der Paraphrase ist.

6.3 Resultate

Als Trainings-, Development- und Testdaten wurden von mir die entsprechenden Datensätze aus der CoNLL-2012 Shared Task (Pradhan u. a., 2012) verwendet. Insgesamt habe ich vier verschiedene Modelle mit Cort trainiert, ein Baseline Modell mit den Features, die für die Ergebnisse in Martschat und Strube (2015) verwendet wurden, ein Modell aus Baseline- und Attributfeatures, ein Modell aus Baseline- und Funktionsfeatures und ein Modell mit allen Features. Die von den Modellen produzierten Koreferenzketten für die Testdaten habe ich anhand der gängigsten drei Maße für Koreferenzketten evaluiert, *MUC*, *BCUBED* und *CAEFE*. Diese drei Maße wurden auch in der CoNLL-2012 Shared Task für den Mittelwert verwendet, nach dem die eingereichten Systeme verglichen wurden. Die Ergebnisse sind in Tabelle 2 aufgeführt.

Im Durchschnitt am besten schneidet die Baseline in Kombination mit den Funktionsfeatures ab, allerdings nur numerisch besser als die Baselinefeatures allein.

²Interessanterweise hatte diese Erweiterung der möglichen Adjektive um das bis zu 50-fache allerdings keinen Effekt auf die Modelle zu den verwendeten Daten.

Table 2: F1 Score verschiedener Maße für verschiedene Featurekombinationen

	MUC	BCUBED	CAEFE	Average
Baseline	73.95	61.02	57.15	64.04
+ Attribute	73.92	61.07	56.77	63.92
+ Funktion	74.08	61.22	57.3	64.20
+ Alle	74.14	61.22	56.96	64.10

Die Attributfeatures verschlechtern das Ergebnis der Modelle sogar, wenn auch nur geringfügig.

7 Auswertung

Es gelingt den Features, die die extrahierten Relationen in das Koreferenzsystem integrieren nicht, dessen Leistung nennenswert zu verbessern. Zwar scheinen die Funktionsfeatures geringfügig effektiver als die Attributfeatures, allerdings bewegen sich diese Unterschiede in Bereichen, die kaum statistisch signifikant sind.

Dass die Features fast keinen Effekt haben, kann auf diverse Gründe zurückgeführt werden. Zum einen sind wohl schlicht nicht genügend solcher Relationen aus den Glossen extrahierbar. Als Definition enthalten diese zwar genügend Information, um ein Synset klar von anderen abzutrennen, erschöpfen aber bei weitem nicht die Vielfalt an Möglichkeiten es zu beschreiben. Attribute und Funktionen finden sich jeweils bloß in etwa einem Drittel der Glossen und selbst dort wo sie zu finden sind meist nicht in ausreichender Menge für diese Anwendung. Dies gilt im besonderen für Funktionen, die bei weitem nicht alle typischen Aktionen der Synsets umfassen können.

Für die Funktionen entsteht aufgrund der gleichzeitig sehr offenen Formulierung in Hinblick auf die semantische Rolle des Substantivs auch viel Rauschen, da z.B. ein Objekt, mit dem typischerweise etwas bestimmtes getan wird, diese Aktion nicht unbedingt typischerweise selbst ausführt. Ein Fahrzeug transportiert, wird aber selten selbst transportiert.

Zuletzt sind die CoNLL Daten nicht unbedingt charakteristisch für Konstrukte wie die der Beispielsätzen (3) bis (6). Der Korpus besteht zum größten Teil aus Zeitungsartikeln, Nachrichten und Telefongesprächen, kaum aus Prosa Texten. Letztere enthalten aber bedeutend häufiger umschreibende Konstrukte, wie sie von den Attributfeatures erkannt werden könnten.

8 Schlussfolgerung

Die oben aufgeführten Gründe zeigen, warum die extrahierten Relationen für Koreferenzresolution, insbesondere in der Nachrichtendomäne, eher ungeeignet sind. Nichtsdestotrotz könnten die Relationen in anderen Anwendungen hilfreicher sein. Systeme, die Pfadähnlichkeiten als Maße benutzen, wie beispiels-

weise WSD Anwendungen, könnten großen Nutzen aus den neuen Verbindungen zwischen verschiedenen Wortklassen ziehen, da nun auch Ähnlichkeiten zwischen Synsets aus unterschiedlichen Klassen berechnet werden können. Natürlich müsste hierfür die Disambiguierung der Glossen bedeutend präziser werden.

Die automatisierte Extraktion von Relationen aus den WordNet Glossen war letzten Endes zwar möglich, doch selbst wenn mit verbesserten Mechanismen zur Disambiguierung und Transformation der Glossen noch weitere und insgesamt präzisere Relationspaare extrahiert werden könnten, werden die Glossen immer ein sehr begrenztes Potential haben. Vielversprechender könnte hingegen eine Extraktion der Relationen aus anderen Lexika oder sogar ganzen Texten nach bestimmten Patterns sein. Die extrahierte Information könnte in WordNet reintegriert werden und so ein bedeutend dichteres semantisches Netz entstehen. Größere Informationsquellen würden auch spezifischere und somit besser verwendbare Relationen erlauben. Zu untersuchen, ob dieses halbautomatisch erstellte semantische Netz Vorteile gegenüber vollautomatisch erstellten Ontologien bieten könnte, sei zukünftiger Forschung überlassen.

Literatur

- Durrett, Greg und Dan Klein (2013). "Easy Victories and Uphill Battles in Coreference Resolution". In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Seattle, Washington: Association for Computational Linguistics.
- Harabagiu, Sanda M., George A. Miller und Dan I. Moldovan (1999). "WordNet 2 – A Morphologically and Semantically Enhanced Resource". In: *Proc. SIGLEX 1999*.
- Lewis, Mike, Luheng He und Luke Zettlemoyer (2015). "Joint A* CCG Parsing and Semantic Role Labelling". In: *Empirical Methods in Natural Language Processing*.
- Martschat, Sebastian und Michael Strube (2015). "Latent Structures for Coreference Resolution". In: *Transactions of the Association for Computational Linguistics* 3, S. 405–418. ISSN: 2307-387X.
- Miller, George Armitage (1993). "Noun in WordNet: A Lexical Inheritance System". In: *5 Papers on WordNet*. Hrsg. von George Armitage Miller. Princeton University, S. 10–25.
- Miller, George Armitage u. a. (1993). "Introduction to WordNet: AN On-line Lexical Database". In: *5 Papers on WordNet*. Hrsg. von George Armitage Miller. Princeton University, S. 1–9.
- Moldovan, Dan und Adrian Novischi (2002). "Lexical Chains for Question Answering". In: *Proc. COLING 2002*.
- Pradhan, Sameer u. a. (2012). "CoNLL-2012 Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes". In: *Proceedings of the Sixteenth Conference on Computational Natural Language Learning (CoNLL 2012)*. Jeju, Korea.