Automatic Extraction of Semantic Competence Descriptions Using German Module Descriptions from Higher Education as an Example

Timo Raschke[1] and Johannes Konert[2]

Abstract: For managing competencies, competency descriptions in natural language are usually used. A conversion to semantic data records is advised, if cataloging and organization of these data is needed. An approach was developed to extract German semantic competence triples consisting of competence verb, object and context using the example of continuous text module descriptions from higher education (in the specific case of the Beuth University of Applied Sciences Berlin) using recent solutions from the research area of Natural Language Processing. The techniques *Constituency Parsing*, *Dependency Parsing* and *Semantic Role Labeling* are compared based on their suitability. First results are presented based on a developed software architecture and a resulting prototype. The results show that competence extraction with an F1 measure of up to 50.2 % is possible on the basis of dependency parsing. The results of this work create a basis for the further use of semantic competence descriptions for a variety of other areas of application, for example, similarity calculations of university courses or the creation of dependency graphs for courses in a course.

Keywords: Natural Language Processing, dependency parsing, competence extraction, semantic

[1] Beuth Hochschule für Technik Berlin, Fachbereich VI – Informatik und Medien, Luxemburger Str. 10, 13353 Berlin, research@timoraschke.de
[2] Hochschule Fulda, Fachbereich Angewandte Informatik, Luxemburger Straße 123, 36037 Fulda, johannes.konert@informatik.hochschule-fulda.de, https://orcid.org/0000-0003-0022-535X

doi:10.18420/delfi2020-ws-110

Automatisierte Extraktion semantischer Kompetenzbeschreibungen am Beispiel von deutschsprachigen Modulbeschreibungen aus der Hochschullehre

Timo Raschke[1] und Johannes Konert[2]

Abstract: Um Kompetenzen von Personen zu verwalten, werden für gewöhnlich Kompetenzbeschreibungen in menschlicher Sprache verwendet. Soll nun eine semantische Katalogisierung und Organisation mittels dieser Daten erfolgen, so empfiehlt sich eine Konvertierung in semantische Datensätze. Es, wurde ein Ansatz entwickelt, um deutschsprachige, semantische Kompetenztripel bestehend aus Kompetenzverb, Objekt und Kontext am Beispiel von Fließtext-Modulbeschreibungen aus der Hochschullehre (im konkreten Falle der Beuth Hochschule für Technik Berlin) mittels aktueller Lösungsansätze des Forschungsgebiets Natural Language Processing zu extrahieren. Es werden *Konstituenzparsing*, *Dependenzparsing* und *Sematinc Role Labelling* anhand ihrer Eignung verglichen und erste Ergebnisse anhand eines entwickelten Entwurfs und ein daraus resultierender Prototyp präsentiert. Die Ergebnisse zeigen, dass die Kompetenzextraktion mit einem F_1-Maß von bis zu 70,1 % auf Basis von Dependenzparsing möglich ist. Durch die Ergebnisse dieser Arbeit wird eine Grundlage für die weitere Verwendung von semantischen Kompetenzbeschreibungen für eine Vielzahl anderer Anwendungsgebiete geschaffen, zum Beispiel Ähnlichkeitsberechnungen von Hochschulkursen oder die Erstellung von Abhängigkeitsgraphen für Kurse eines Studiengangs.

Keywords: Natural Language Processing, Dependenzparsing, Kompetenzextraktion, Semantik

1 Einleitung

Im Laufe des Lebens gewinnt ein Mensch eine Vielzahl von Erfahrungen. Wenn aus diesen Erfahrungen Fähigkeiten oder Fertigkeiten hervorgehen, wird von einer Kompetenz gesprochen. Es gibt viele Möglichkeiten, Kompetenzen zu erlangen: Von der Grundschule über die Oberschule bis hin zum Studium oder Berufsleben. Es gilt einen Überblick zu erhalten, welche Fächer, Kurse, Module, Weiterbildungen oder Schulungen welche Kompetenzen vermitteln und welche diese gegebenenfalls voraussetzen. Weiterhin soll ermittelt werden, welche Kompetenzen wurden bereits erlangt wurden und wie diese nachgewiesen werden können.

Kompetenzen sind für die spätere Auswahl zum Beispiel für konsekutive Masterstudiengänge oder eine berufliche Tätigkeit relevant und müssen in vielen Fällen

[1] Beuth Hochschule für Technik Berlin, Fachbereich VI – Informatik und Medien, Luxemburger Str. 10, 13353 Berlin, research@timoraschke.de
[2] Hochschule Fulda, Fachbereich Angewandte Informatik, Luxemburger Straße 123, 36037 Fulda, johannes.konert@informatik.hochschule-fulda.de, https://orcid.org/0000-0003-0022-535X

doi:10.18420/delfi2020-ws-110

Modell	Sprache	Genauigkeit	Trefferquote	F_1
Klein und Manning (2003) [KM03]	Englisch	83,1	82,1	82,9
Rafferty und Manning (2008) [RM08]	Deutsch	-	-	79,7
Zhu, Zhang, Chen, Zhang und Zhu (2013) [Zh13]	Englisch	91,5	91,1	91,3
Joshi, Peters und Hopkins (2018) [JPH18]	Englisch	94,8	93,8	94,3

Tabelle 1 Vergleich zwischen den Ansätzen für Konstituenzparsing

mit Nachweisen belegt werden. Kompetenzen werden im Rahmen eines Studiums unter anderem mittels Kompetenzbeschreibungen beschrieben. Dabei handelt es sich in der Regel um Fließtexte, die gemäß ihrer Natur nicht besonders gut durch Maschinen semantisch interpretierbar sind. Eine Möglichkeit zur Vereinfachung dieses Ablaufs ist die Digitalisierung der Dokumentation von Kompetenzen mittels semantischer Datensätze. Durch eine semantische Digitalisierung können Kompetenzen digital als Zertifikate ausgestellt werden, wodurch maschinenlesbar dokumentiert werden kann, welche Kompetenzen im Rahmen zum Beispiel eines Studiums erworben wurden.

Dieser Beitrag basiert auf einer abgeschlossenen Masterarbeit und hat zum Ziel, folgende Forschungsfragen zu klären:

1. Welche Voraussetzungen und Daten der Modulbeschreibungen von Studiengängen an deutschen Hochschulen sind relevant für eine solche Extraktion?
2. Mit welchen Verfahren der maschinellen Verarbeitung natürlicher Sprache, im Englischen Natural Language Processing (NLP) genannt, lassen sich Kompetenzbeschreibung präzise aus deutschsprachigen Fließtexten extrahieren?
3. Welche speziellen Herausforderungen stellt die Verwendung der deutschen Sprache an eine solche Extraktion?

2 Aktueller Stand der Forschung und verwandte Arbeiten

Nach intensiver Recherche ist das Forschungsgebiet der automatisierten Extraktion von Kompetenzen aus *deutschsprachigen* Texten noch unerforscht oder zumindest nicht in den gesichteten wissenschaftlichen Publikationen thematisiert. Aufgrund dessen liegt das

Modell	UAS[1]	LAS[1]
Chen und Manning (2014) [CM14]	92,0	90,7
Kiperwasser und Goldberg (2016) [KG16]	93,9	91,9
Dozat und Manning (2017) [DM17]	95,7	94,1

Tabelle 2 Vergleich zwischen verschiedenen Ansätzen für Dependenzparsing

Hauptaugenmerkt darauf, verschiedene Methoden und Ansätze zur Beantwortung der Forschungsfragen (s.o.) zu erörtern. Das Forschungsgebiet des NLP konzentriert sich hauptsächlich auf die englische Sprache. Wo möglich werden Methoden erläutert, die für die Extraktion aus Texten deutscher Sprache anwendbar sind.

2.1 Konstituenzparsing, Dependenzparsing, Semantic Role Labeling

Konstituenzparsing versucht aus mittels Part-of-Speech Tagging[3] (POS) annotierten Texten deren Konstituenzgrammatiken zu bilden. Es wurden vier Ansätze des Konstituenzparsings betrachtet. Dazu zählen der auf selektiver Aufteilung der Elternknoten und einer sprachlich abgeleiteten Aufteilung von Merkmalen von [KM02] und eine Adaption des Ansatzes für die deutsche Sprache von [RM08]. Weiterhin das Shift-Reducing Modell von [Zh13] und der ELMo[4]-basierte Ansatz von [JPH18]. Die Ergebnisse aller vorgestellten Modelle sind in Tabelle 1 zu finden. Die Ergebnisse wurden auf unterschiedlichen Datensätzen erzielt. Bei - liegen keine detaillierten Ergebnisse vor.

Dependenzparsing versucht anhand von POS annotierten Texten deren Dependenzgrammatiken zu bilden. Diese können durch Dependenzgraphen abgebildet werden. Dazu wurden vier Ansätze des Dependenzparsings behandelt. Es handelt sich dabei um den für die deutsche Sprache entwickelten *Zurich Dependency Parser for German (ParZu)* von [Se09] und [SS13], den auf neuronalen Netzwerken basierenden Parser von [CM14], der unter anderem auch für Deutsch adaptiert wurde und ein auf bidirektionalen LSTMs basierenden Parser von [KG16], der wiederum von [DM17] optimiert wurde. Die Ergebnisse der vorgestellten Modelle sind in Tabelle 2 zu finden.

[3] Englisch für *Annotation von Wortarten*
[4] Embeddings from Language Models, Englisch für *Einbettungen aus Sprachmodellen*

Modell	Genauigkeit	Trefferquote	F_1
He, Lee, Levy und Zettlemoyer (2018) [He18]	81,9	84,0	82,9
Ouchi, Shindo, und Matsumoto (2018) [OSM18]	88,5	85,5	87,0
Li et al. (2019) [Li19]	85,7	86,3	86,0
Shi und Lin (2019) [Sh19]	85,9	87,0	86,5

Tabelle 3 Vergleich zwischen verschiedenen SRL Ansätzen auf dem CoNLL-2012 Datensatz

Semantic Role Labeling[5] (SRL) versucht aus Texten natürlicher Sprache automatisiert deren Semantik zu ermitteln. Im Folgenden werden vier SRL-Modelle vorgestellt. [He18] entwickelten einen BIO[6]-basierten Ansatz mittel neuronaler Netzwerke und vortrainierten ELMo-Modellen. [OSM18] entwickelten einen spannenbasierten mit Hilfe eines neuronalen Netzwerkes und vortrainierten ELMo-Modellen. Eine Kombination dieser beiden Ansätze setzt das Ende-zu-Ende-Modell von [Li19] ein. Ein auf vortrainierten BERT[7]-Modellen basierender Ansatz wurde von [SL19] entwickelt. Die Ergebnisse aller vorgestellten Ansätze sind in Tabelle 3 zu finden.

2.2 Zusammenfassung

Die drei vorgestellten Techniken Konstituenzparsing, Dependenzparsing und SRL können grundsätzlich alle für die Extraktion von Kompetenztripeln aus Kompetenzbeschreibungen in natürlicher Sprache verwendet werden, wobei POS die Grundlage aller drei Techniken bildet. Konstituenzparsing und Dependenzparsing stellen dabei eine erste Voranalyse dar der Sprache dar, durch die mittels weiterer Logik ein Tripel aus Kompetenz, Objekt und Kontext erstellt werden kann. SRL stellt durch das Hinzufügen von Semantik eine interessante Alternative zu den beiden erstgenannten Techniken dar, da dadurch bereits ein Teil der notwendigen Weiterverarbeitung durch Algorithmen reduziert werden kann. Leider existiert zum Zeitpunkt der Ausarbeitung kein geeignetes Modell für SRL in deutscher Sprache.

[5] Englisch für *Annotation von semantischen Rollen*
[6] Beginning-Inside-Outside, Englisch für *Beginn-Innen-Außen*
[7] Bidirectional Encoder Representations from Transformers, Englisch für *Bidirektionale Encoder-Darstellungen von Transformatoren*

3 Ansatz für deutschsprachige Kompetenzextraktion

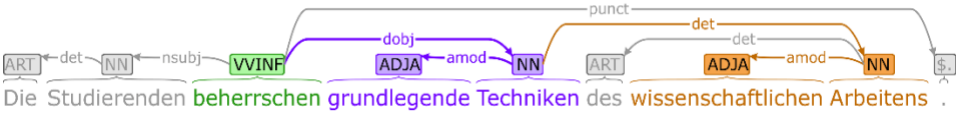

Abbildung 6 Beispiel der Extraktion eines Kompetenztripels einer beispielhaften deutschsprachigen Kompetenzbeschreibung

Auf Grundlage der durchgeführten Recherche wurde ein auf Dependenzparsing basierender Entwurf zur automatisierten Extraktion von Kompetenzbeschreibungen (als Kompetenztripel) mit dem Fokus auf deutschsprachige Texte entwickelt. Ein Kompetenztripel wird als ein Tripel definiert, bestehend aus einem Kompetenzverb, dem zum Kompetenzverb zugehörigen Objekt und gegebenenfalls einem Kontext, in dem die Kompetenz erbracht wird. Ziel ist es, dass der Entwurf Kompetenztripel innerhalb eines Textes (also einer Kompetenzbeschreibung) in deutscher Sprache erkennen und extrahieren kann.

Eingabeparameter des Extraktionsprozesses sind Sätze deutscher Sprache. Auf einem eingegebenen Satz werden zuerst mittels POS die Wortarten der Wörter des Satzes ermittelt. Auf Grundlage dessen kann nun auf dem Satz ein Dependenzparsing durchgeführt werden, dessen Ausgabe der mit Wortarten und Dependenzen annotierte Eingabesatz ist. Dependenzparsing wurde als Hauptbestandteil des Prozesses gewählt, da es im Gegensatz zu Konstituenzparsing ähnlich wie der Ausgabeparameter Prädikatzentriert ist und durch die gebildeten Dependenzbeziehungen zwischen den einzelnen Wörtern Objekte und Kontexte einfacher ermittelt werden können. SRL, als möglicher weiterer vielversprechender Ansatz, wurde vorerst nicht in Betracht gezogen, da aktuell kein deutsches Modell existiert und das Anlegen eines deutschen Datensatzes inklusive Ressourcen für Training eines Modells für diesen ersten Ansatz im Rahmen der vorliegenden wissenschaftlichen Arbeit nicht möglich war. Auf dem gebildeten Dependenzgraphen werden nun weitere Schritte durchgeführt um Kompetenztripel bestehend aus Kompetenzverben, deren Objekte und Kontexte zu ermitteln. Der im Folgenden dargestellte Extraktionsprozess verwendet für die Bezeichnung der Wortarten das TIGER Annotationsschema ([Pr03]) und für Dependenzen Universal Dependencies 2.0 ([MDS15]). Das Ergebnis einer beispielhaften Extraktion ist in Abbildung 1 als Dependenzgraph dargestellt, wobei grün gefärbte Wörter Kompetenzverben, lilafarbene Objekte, orangefarbene Kontexte und grau gefärbte Wörter Elemente ohne Funktion darstellen.

Algorithmische Schritte der Extraktion von Kompetenztripeln:

1) **Kompetenzverben ermitteln:** Finde infinitive Vollverben (VVINF), infinitive Vollverben mit zu (VVIZU) oder finite Verben (VVFIN) innerhalb des Satzes.
2) (optional) **Prüfen gegen Positivliste:** Prüfe für jedes gefundene Verb, ob es in einer Positivliste (zum Beispiel Taxonomiestufen nach Bloom) steht. Verwerfe nicht gefundene Verben.

3) Prüfe für jedes gefundene Kompetenzverb:
   a) **Objekte ermitteln:** Finde direkte Objekte (dobj) des Verbs, die normale Nomen (NN) sind
      i) Finde gegebenenfalls das Adjektiv (ADJA) des gefundenen Objekt-Nomens (NN) mittels des adjektiven Modifikators (amod)
      ii) **Kontext ermitteln:** Finde Nomen (NN) die eine nominale Modifikator (nmod) oder Determinator (det) Dependenz zum Nomen (NN) des Objektes besitzen
         (1) Finde gegebenenfalls das Adjektiv (ADJA) des Kontext-Nomens (NN) mittels des adjektiven Modifikators (amod)

Gemäß des zuvor beschrieben Entwurfs wurde ein Prototyp (in weiteren Verlauf CompEx[8] genannt) implementiert. Er wurde in Python implementiert und ermöglicht eine Extraktion von Kompetenztripeln für deutschsprachige Sätze auf Basis von Stanford CoreNLP[9] (POS, Dependenzparsing). Die Extraktion des Kompetenzverbs kann mit Hilfe eines optionalen Taxonomiewörterbuchs gesteuert werden. Der Dependenzgraph wird im Anschluss mit Hilfe eines Semgrex-Ausdrucks[10] analysiert. Weiterhin kann der Prototyp seine Ergebnisse gegen einen annotierten Datensatz prüfen und gängige Kennwerte berechnen. Der Prototyp ist unter der MIT-Lizenz lizensiert und auf GitHub[11] abrufbar.

4 Evaluation

Für eine repräsentative Evaluation ist ein geeigneter Testdatensatz für deutschsprachige Kompetenztripel aus Kompetenzbeschreibungen erforderlich. Da zum Zeitpunkt dieser Forschungsarbeit kein geeigneter Datensatz existierte, erfolgte eine manuelle Erstellung eines Testdatensatzes auf Basis der Modulbeschreibungen des Fachbereich VI der Beuth Hochschule für Technik Berlin für *Medieninformatik Bachelor* [Be16] und *Medieninformatik Master* [Be17].

Die Evaluation wurde schrittweise mit ansteigender Komplexität durchgeführt: Zu Beginn erfolgt lediglich die Erkennung der Kompetenzverbs, anschließend die Erkennung des Kompetenzverbs inklusive seiner Objekte und schlussendlich die Erkennung des kompletten Kompetenztripels. Alle Schritte werden jeweils einmal mit und ohne Taxonomiewörterbuch nach den Taxonomiestufen von Bloom durchgeführt. Die Berechnungen wurden auf Wortebene durchgeführt, d.h. ein falsches Wort eines drei

[8] Akronym für *Competency Extractor*
[9] Weitere Infos über CoreNLP siehe https://stanfordnlp.github.io/CoreNLP/
[10] Werkzeug zum Finden von Mustern in Dependenzgraphen, siehe https://nlp.stanford.edu/software/tregex.html
[11] CompEx ist verfügbar unter https://github.com/traschke/bht-compex

Durchlauf	Genauigkeit	Trefferquote	F_1-Maß
Verb	65,1 | 84,0	76,0 | 46,0	70,1 | 59,4
Verb + Objekt	53,3 | 66,8	52,7 | 33,3	53,0 | 44,4
Verb + Objekt + Kontext	52,9 | 66,0	47,8 | 31,0	50,2 | 42,1

Tabelle 4 Ergebnisse der Evaluation des Prototyps (ohne | mit Taxonomiewörterbuch)

Wörter umfassenden Objekts führt zu einem 66 % korrekten Objekt, statt das gesamte Objekt als falsch anzusehen. Die Ergebnisse der Evaluation sind in **Error! Reference source not found.** zu finden. Die Ergebnisse zeigen, dass dieser erste Ansatz für die Extraktion von Kompetenztripeln aus deutschsprachigen Kompetenzbeschreibungen verglichen mit den Ergebnissen der themenverwandten Arbeiten noch keine perfekten, geschweige denn zufriedenstellende Ergebnisse liefert.

Die Erkennung des Kompetenzverbs innerhalb eines Satzes einer Kompetenzbeschreibung verläuft dabei am zuverlässigsten, ist aber verbesserungsbedürftig. Dies kann zum einen auf Schwachstellen des Extraktionsalgorithmus, zum anderen aber auch auf die Qualität der Daten (Inhaltlich wie aber auch in der manuellen Annotation) zurückzuführen sein. Auffällig ist die voranschreitende Verschlechterung der Erkennung, je komplexer und verschachtelter Eingabesätze sind, desto schlechter ist die Erkennung. Dies ist zum einen auf den Entwurf zurückzuführen, der eher einfache Satzstrukturen erwartet, aber auch auf fehlerhafte Annotationen durch den deutschsprachigen Dependenzparser von CoreNLP, wie während der Anwendung aufgefallen sind.

5 Fazit und Ausblick

Dieser Beitrag zeigt einen ersten Ansatz der Extraktion von Kompetenztripeln aus deutschsprachigen Kompetenzbeschreibungen. Wie die Ergebnisse zeigen, ist dieser erste Ansatz noch nicht fähig, verlässliche Ergebnisse zu liefern. Wie in der Einleitung erwähnt, besteht die Herausforderung bei der Entwicklung eines Extraktionsverfahrens darin, dieses auf die deutsche Sprache anzuwenden.

Mit den Ergebnissen eines solchen Extraktionssystem, wie hier vorgestellt, könnten Ähnlichkeiten zwischen zwei oder mehreren Kompetenzbeschreibungen berechnet werden. Es ließe sich so zum Beispiel auf Hochschulebene bestimmen, welche bereits absolvierten Kurse auf Grund ihrer vermittelten Kompetenzen als Voraussetzungen für andere Kurse gelten. Es ließe sich daraus ein Abhängigkeitsbaum für Module eines Studiengangsaufbauen oder redundant vermittelte Kompetenzen ermitteln. Weiterhin kann über eine vollständige Individualisierung des Studienverlaufs auf Basis von Kompetenzen nachgedacht werden. Studierenden werden auf Basis des jeweiligen Studiengangs und ihrer persönlichen Interessen entsprechend Module angeboten.

Literaturverzeichnis

[Be16] Beuth Hochschule für Technik Berlin - Fachbereich VI: Modulhandbuch für den Bachelor-Studiengang Medieninformatik, 2016.

[Be18] Beuth Hochschule für Technik Berlin - Fachbereich VI: Modulhandbuch Masterstudiengang Medieninformatik, 2018.

[CM14] Chen, D., Manning, C.D.: A Fast and Accurate Dependency Parser Using Neural Networks. In: Proceedings of the 2014 Conference on Empirical Methods in Nataural Language Processing (EMNLP). Association for Computational Linguistics, Stroudsburg, PA, USA, S. 740-750, 2014.

[DM17] Dozat, T., Manning, C.D.: Deep Biaffine Attention for Neural Dependency Parsing. 5th International Conference on Learning Represent ICLR 2017 - Conference Track Proceedings, S. 1-8, 2017.

[He18] He, L., Lee, K., Levy, O., Zettlemoyer, L.: Jointly Predicting Predicates and Arguments in Neural Semantic Role Labeling. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Volume 2: Short Papers. Association for Computational Linguistics, Stroudsburg, PA, USA, S. 364-369, 2018.

[Jo18] Joshi, V., Peters, M., Hopkins, M.: Extending a Parser to Distant Domains Using a Few Dozen Partially Annotated Examples. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Stroudsburg, PA, USA, S. 1190-1199, 2018.

[KG16] Kiperwasser, E., Goldberg, Y.: Simple and Accurate Dependency Parsing Using Bidirectional LSTM Feature Representations. In: Transactions of the Association for Computational Linguistics, Volume 4. Association for Computational Linguistics, Stroudsburg, PA, USA, S. 313-327, 2016.

[KM02] Klein, D., Manning, C.D.: Fast exact inference with a factored model for natural language parsing. In: Proceedings of the 15th International Conference on Neural Information Processing Systems. MIT Press, Cambridge, MA, USA, S. 3-10, 2002.

[Li19] Li, Z., He, S., Zhao, H., Zhang, Y., Zhang, Z., Zhou, X., Zhou, X.: Dependency or span, end-to-end uniform semantic role labeling. In: Proceedings of the AAAI Conference on Artificial Intelligence, Volume 33. AAAI Press, Palo Alto, California USA, S. 6730-6737, 2019.

[Ma15] de Marneffe, M.-C., Dozat, T., Silveira, N., Haverinen, K., Ginter, F., Nivre, J., Manning, C.D.: Universal Dependencies: A cross-linguistic typology, 2015.

[OSM18] Ouchi, H., Shindo, H., Matsumoto, Y.: A Span Selection Model for Semantic Role Labeling. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Stroudsburg, PA, USA, S. 1630-1642, 2018.

[Pr03] Projekt TIGER: TIGER Annotationsschema. 2003.

[RM08] Rafferty, A.N., Manning, C.D.: Parsing Three German Treebanks: Lexicalized and Unlexicalized Baselines. In: Proceedings of the Workshop on Parsing German.

Association for Computational Linguistics, Stroudsburg, PA, USA, S. 40-46, 2008.

[Se09] Sennrich, R., Schneider, G., Volk, M., Warin, M.: A New Hybrid Dependency Parser for German. In: Proceedings of the German Society for Computational Linguistics and Language Technology. German Society for Computational Linguistics, Tübingen, Germany, S. 115-124, 2009.

[Sh19] Shi, P., Lin, J.: Simple BERT Models for Relation Extraction and Semantic Role Labeling. arXiv preprint arXiv:1904.05255, 2019.

[SVS13] Sennrich, R., Volk, M., Schneider, G.: Exploiting Synergies Between Open Resources for German Dependency Parsing, POS-tagging, and Morphological Analysis. In: Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013. INCOMA Ltd., Shoumen, Bulgaria, S. 601-609, 2013.

[Zh13] Zhu, M., Zhang, Y., Chen, W., Zhang, M., Zhu, J.: Fast and Accurate Shift-Reduce Constituent Parsing. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Stroudsburg, PA, USA, S. 434-443, 2013.