

Profilierungsmodul Computerlinguistik I • Dr. Zangeneid

Konversion von Konstituentenstrukturen zu Abhängigkeitsstrukturen

Referentin: Lin Shui
30.01.2024





Inhalt

1. **Abhängigkeitsstrukturen**
2. **LTH Abhängigkeitspräsentationsschema (LTH)**
3. **Stanford Abhängigkeitspräsentationsschema (SD)**
4. **Computerlinguistische Forschungen**
5. **Zusammenfassung**



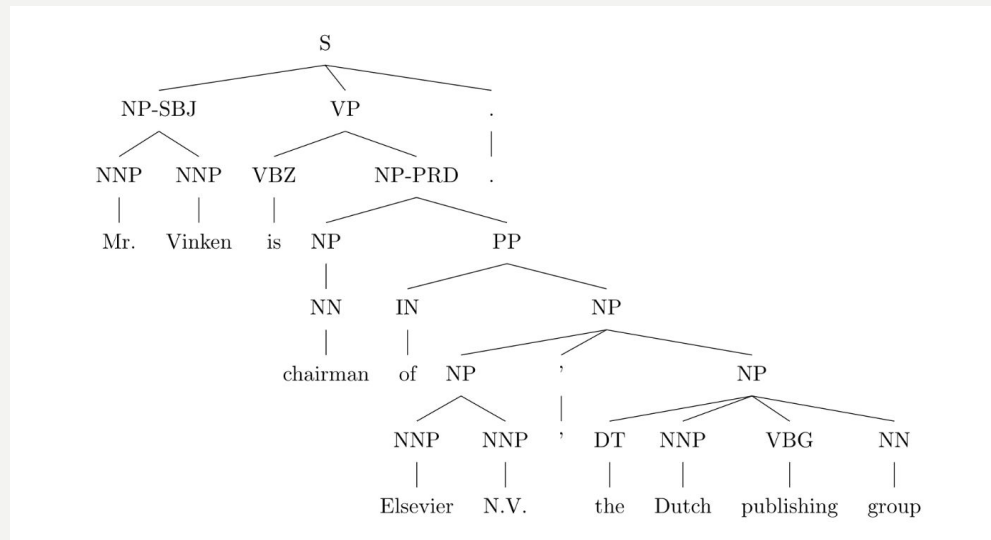
Dependenz vs Konstituenz:

Dependenz und Konstituenz sind **elementare Strukturaspekte** von Sätzen. Dependenzgrammatiken und Konstituentengrammatik sind daher zwei grundlegende Typen von Grammatiken. ("2 Dependenz und Konstituenz". Valenzgrammatik des Deutschen: Eine Einführung, 2011,S.21)



Dependenz vs Konstituenz:

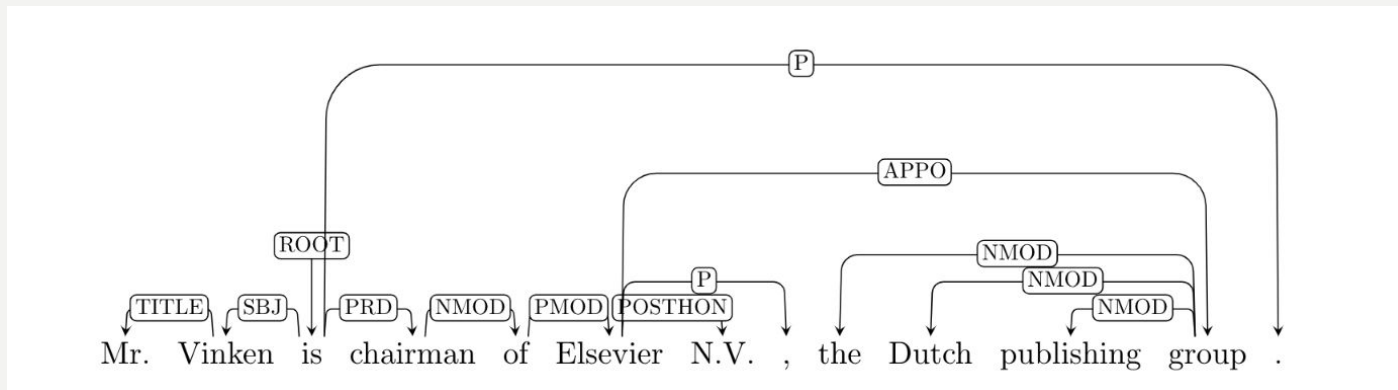
- Eine **Konstituentengrammatik** ist durch die Kombinatorik eines Satzes oder einer Wortgruppe mit der **Teil-Ganzes-Beziehung** charakterisiert.
- Der **Konstituentenbaum** des Satzes: "Mr. Vinken is chairman of Elsevier N. V., the Dutch publishing group." (Herr Vinken ist Vorsitzender der Elsevier N. V., der niederländischen Verlagsgruppe.)





Abhängigkeit vs Konstituierung:

- Eine **Abhängigkeitsgrammatik** beschreibt Beziehung/**Abhängigkeiten** zw. Wörtern und stellt die verborgenen strukturellen Beziehungen zwischen den einzelnen Elementen im Satz dar.
- Der **Abhängigkeitsbaum** mit der LTH Repräsentation des Satzes: "Mr. Vinken is chairman of Elsevier N. V., the Dutch publishing group." (Herr Vinken ist Vorsitzender der Elsevier N. V., der niederländischen Verlagsgruppe.)





Motivation:

1. Die meisten aktuellen Spitzen-**Abhängigkeitsparser** verwenden überwachtes Lernen. Für Englisch gibt es einige manuell annotierte Abhängigkeits-Treebanks. Dennoch sind konstituentenbasierte Treebanks wie der Penn Treebank dominanter. (Lin, 2019)
2. Die **Abhängigkeitssyntax** bietet aus praktischer Sicht eine Reihe von Vorteilen, wie die Verfügbarkeit effizienter Parsing-Algorithmen, die Sätze in linearer Zeit zu analysieren (de Marneffe & Manning, 2008).
3. Die Abhängigkeitsgrammatik ist universell, während Konstituenten eher **englisch zentriert** wären. (Mel'čuk, 1988).



Konversion Methode:

Konvertierung basiert auf der Zuweisung eines **Kopfs (syntaktischer Herr)** zu jeder Konstituente des Syntaxbaums und mithilfe des Kopfs Dependenzbäumen zu erstellen.

“The cat sat on the mat.” (Die Katze saß auf der Matte.)

Grammatik

$S \rightarrow NP VP^*$

$NP \rightarrow DT NN^*$

$VP \rightarrow VBD^* PP$

$PP \rightarrow IN^* NP$

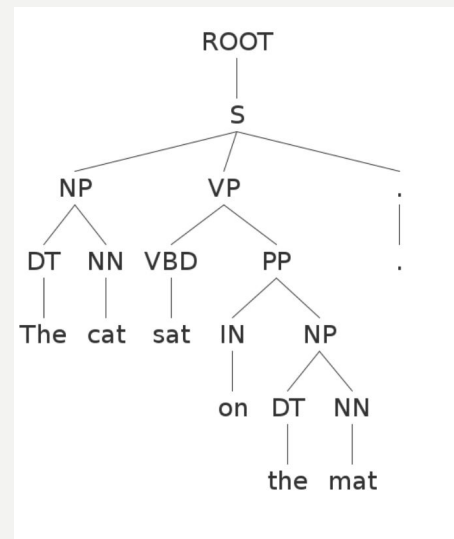
$DT \rightarrow \text{'The' | 'the'}$

$NN \rightarrow \text{'cat' | 'mat'}$

$VBD \rightarrow \text{'sat'}$

$IN \rightarrow \text{'on'}$

Konstituentenbaum

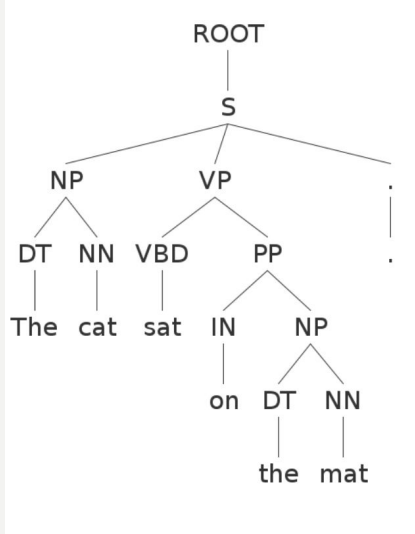




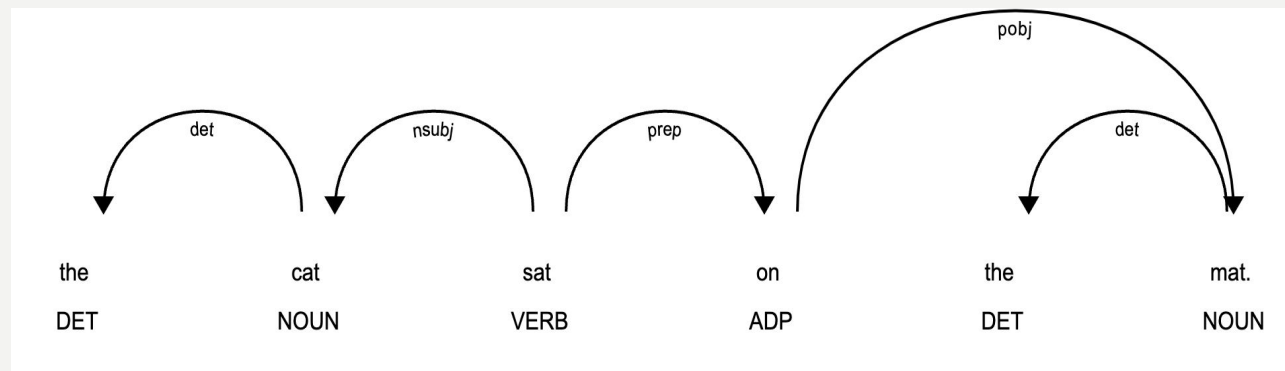
Konversion Methode:

“The cat sat on the mat.” (Die Katze saß auf der Matte.)

Konstituentenbaum



Abhängigkeitsbaum(SpaCy)





Konversion Methode:

am allerwichtigsten: Köpfe zu finden

1. Magerman (1994) erstellte eine **Kopfperkolationstabelle**, eine Reihe von Prioritätslisten, um die Köpfe von Konstituenten zu finden.
2. Yamada und Matsumoto (2003) **modifizierten die Tabelle** weiter, und ihr Verfahren ist bis heute das beliebteste geworden.
3. PENN2MALT (Nivre, 2006) ist eine Neuimplementierung der Methode von Yamada und Matsumoto und definiert auch eine Reihe von **Heuristiken**, um **Kanten im Abhängigkeitsbaum zu erschließen**.



LTH

*Lunds Tekniska Högskola,
Universitet Lund, Fakultät för Ingenieurvetenskaperna*

1. Ein neues Verfahren zur Umwandlung von **englischen Konstituentenbäumen**, die im Penn-Treebank-Annotationsstil vorliegen, in Dependenzbäume.
2. Das Ziel ist, eine bessere Methode für semantische Verarbeitung als bisherige Methoden zu entwickeln.
3. Es wurde ein reichhaltiges Set an Kanten verwendet und Pfeile eingeführt.



Modifikation von Dependenzpfeilen:

DE: warum, so fragen sie sich, sollte es der EG gehören?

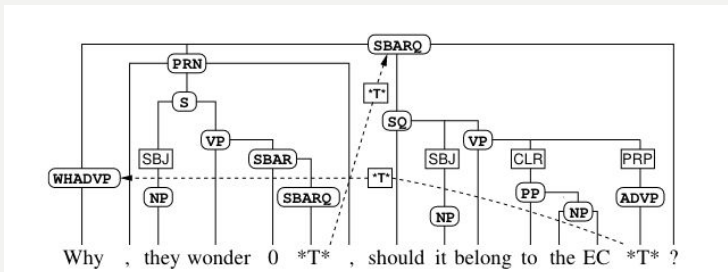


Figure 1: A constituent tree from the Penn Treebank.

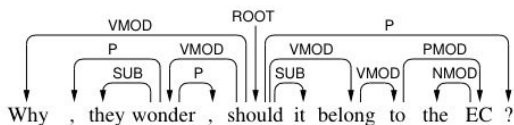


Figure 2: Dependency tree by PENN2MALT.

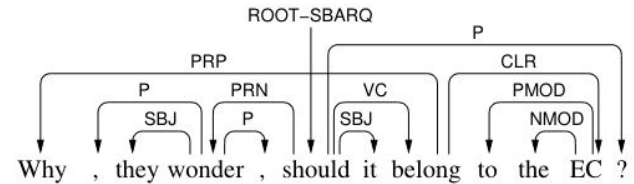


Figure 3: Dependency tree by the new procedure.

Bemerkung:

1. *T* steht für Spur/Trace
2. Das Wort „Why“ bezieht sich auf den **Zweck des Verbs** „belong“. (PRP: purpose or reason)
3. Beim Penn2Malt fehlt es die Relation zwischen Frage und Antwort: It should belong to the EC because. . .
4. In der neuen Methode verknüpfen wir Wh-Wörter mit ihren **semantischen Köpfen**



Ein reiches Set von Kanten:

1. verwendet direkten Objekten (**OBJ**) und indirekten Objekten (**IOBJ**).
2. Das Funktions-Label auf dem Wurzel-Token wurde verwendet, um die Art des Satzes auszudrücken: **ROOT-S (simple declarative clause)**, wenn die Wurzelkonstituente S oder SINV war, **ROOT-SBARQ** (Direct question introduced by a wh-word or a wh-phrase) und **ROOT-SQ** (Inverted yes/no question) für SBARQ bzw. SQ, und **ROOT-FRAG** für alles andere.
3. verwendet strukturelle Bezeichnungen **EXP** (Expletivum/expletive), **CLF** (Spaltsatz/cleft), and **GAP** (Lücke/gapping).



LTH - Experiments

1. Parsing Performance (MaltParser, MSTParser)

	MALTPARSER		MSTPARSER	
	Labeled	Unlabeled	Labeled	Unlabeled
PENN2MALT	90.30%	91.36%	92.04%	93.06%
New conversion	87.63% (28%)	90.54% (9%)	86.92% (64%)	91.64% (20%)

Table 3: Parsing accuracy. Relative error increase in brackets.

- Die neue Konvertierungsmethode führt zu **komplexeren** Dependenzbäumen und wie erwartet, ist das neue Format für Parser schwieriger.
- Die Parsing-Genauigkeit ist in der neuen Konvertierung **geringer als bei PENN2MALT**.



LTH - Experiments

2. Semantic Role Classification (FrameNet)

Method	Accuracy
PENN2MALT	64.3%
New conversion	72.5% (23%)

Table 5: Semantic role classification results.

- Alle Sätze wurden mit dem **MXPOST-Tagger** getaggt und **MaltParser** ausgeführt.
- Das neue Format führt zu einer Fehlerreduktion von 23% bei der Klassifikation.



SD - Versionen

Universität Stanford

- Eine Abhängigkeitsrepräsentation für Englisch wurde **seit 2005** entwickelt, die als **Stanford Dependencies** bekannt wurde.
- Ab 2013 begann man an **Universal Dependencies** zu arbeiten, einer ähnlichen Abhängigkeitsrepräsentation, die für **alle Sprachen** geeignet ist.
- **Seit der Version 3.5.2** ist die Standardausgabe des Stanford Parsers und Stanford CoreNLP die neue Universal Dependencies (UD)-Repräsentation, und die ursprüngliche Stanford **Abhängigkeitsrepräsentation** wurden nicht mehr gepflegt.



SD - 2008

Universität Stanford

- Ziel ist, dass eine **unkomplizierte** Beschreibung grammatischer Beziehungen für jeden Benutzer verfügbar sein sollte.
- Einem einfachen Design folgen und **semantisch gehaltvolle Informationen** bereitstellen, sowie ein automatisches Verfahren zur Extraktion der Beziehungen anbieten.
- **Nicht** für den Zweck der Parser-Bewertung entworfen.



Einfluss von LF Grammatik

1. **Relation /xcomp/** steht für ein offenes satzwertiges Komplement ohne eigenes Subjekt. Das Subjekt des xcomp wird durch **ein anderes Element** im Satz bestimmt. (Fokkens, 2009, Sulger, 2010)
2. Diese Komplemente sind nicht-finite Formen, also Infinitive oder Partizipien.
3. **/xcomp/** stammt aus der Lexikalisch-Funktionalen Grammatik.
 - “He says that you like to swim.” **xcomp**(like, swim)
DE: "Er sagt, dass du gerne schwimmst."
 - “I consider him honest.” **xcomp**(consider, honest)
DE: "Ich halte ihn für ehrlich" - xcomp(halten, ehrlich)



Zusammengefasste Dependenzstrukturen (Collapsed dependencies)

Die **Präpositionen**, **Konjunktionen** sowie Informationen über die Bezugnahme von **Relativsätzen** werden zusammengeführt, um direkte Dependenz zwischen Wörtern zu erhalten.

“based in LA”

prep(based-7, in-8)
pobj(in-8, LA-9) → prep_in(based-7, LA-9)

“makes and distributes electronic, computer”

cc(makes-11, and-12)
conj(makes-11, distributes-13) → conj_and(makes-11, distributes-13)



Zusammengefasste Dependenzstrukturen (Collapsed dependencies)

according to	as per	compared to	instead of	preparatory to
across from	as to	compared with	irrespective of	previous to
ahead of	aside from	due to	next to	prior to
along with	away from	depending on	near to	pursuant to
alongside of	based on	except for	off of	regardless of
apart from	because of	exclusive of	out of	subsequent to
as for	close by	contrary to	outside of	such as
as from	close to	followed by	owing to	thanks to
as of	contrary to	inside of	preliminary to	together with

Table 1: List of two-word prepositions that the system can collapse.

by means of	in case of	in place of	on behalf of	with respect to
in accordance with	in front of	in spite of	on top of	
in addition to	in lieu of	on account of	with regard to	

Table 2: List of three-word prepositions that the system can collapse.



Anwendungen in verschiedene Domänen

NLP: Bessere Performance bei der Aufgabe Textual Entailment (RTE) und Repräsentation zur Extraktion von Meinungen über Merkmale.

Bioinformatik: Ein auf dem SD-Schema basierendes System bei der Extraktion von Beziehungen zwischen Genen und Proteinen das bisher beste System deutlich übertraf.



Down-stream effects of tree-to-dependency conversions

**Jakob Elming, Anders Johannsen, Sigrid Klerke, Emanuele Lapponi[†],
Hector Martinez, Anders Søgaard**

Center for Language Technology, University of Copenhagen

[†]Institute for Informatics, University of Oslo

1. 4 verschiedenen **Dependenzrepräsentationen**: das Yamada-Matsumoto Schema, das CoNLL 2007, das Schema EWT, das im English Web Treebank verwendet wird und LTH Schema.
2. Dependenzparser auf **fünf verschiedene Aufgaben** der natürlichen Sprachverarbeitung (NLP):
 - Negation Resolution
 - Semantic role labeling
 - Statistical machine translation
 - Sentence compression
 - Perspective classification
3. Vergleich des Systems mit Baseline-System, die nicht auf syntaktischen Merkmalen beruhen.



	bl	yamada	conll07	ewt	lth
DEPRELS	-	12	21	47	41
PTB-23 (LAS)	-	88.99	88.52	81.36*	87.52
PTB-23 (UAS)	-	90.21	90.12	84.22*	90.29
Neg: scope F_1	-	81.27	80.43	78.70	79.57
Neg: event F_1	-	76.19	72.90	73.15	76.24
Neg: full negation F_1	-	67.94	63.24	61.60	64.31
SentComp F_1	68.47	72.07	64.29	71.56	71.56
SMT-dev-Meteor	35.80	36.06	36.06	36.16	36.08
SMT-test-Meteor	37.25	37.48	37.50	37.58	37.51
SMT-dev-BLEU	13.66	14.14	14.09	14.04	14.06
SMT-test-BLEU	14.67	15.04	15.04	14.96	15.11
SRL-22-gold	-	81.35	83.22	84.72	84.01
SRL-23-gold	-	79.09	80.85	80.39	82.01
SRL-22-pred	-	74.41	76.22	78.29	66.32
SRL-23-pred	-	73.42	74.34	75.80	64.06
bitterlemons.org	96.08	97.06	95.58	96.08	96.57

1. Die Ergebnisse der Negation-Resolution sind mit syntaktischen Merkmalen in der Yamada-Annotation deutlich besser.
2. Die case-sensitive BLEU-Auswertung der SMT-Systeme zeigt, dass die Wahl der Dependenzrepräsentation keinen signifikanten Einfluss auf die Performance hat.
3. Die Dependenzrepräsentation EWT und LTH führen zu einer besseren Performance bei der Semantic Role Labeling (SRL) als conll07 und yamada.
4. Die Ergebnisse der Sentence Compression sind im Allgemeinen viel besser als die der Baseline.
5. In der Perspective Classification sehen wir, dass syntaktische Merkmale basierend auf Yamada- und lth-Annotationen zu Verbesserungen führen.



The Effect of Dependency Representation Scheme on Syntactic Language Modelling

Sunghwan Mac Kim

Department of Computing
Macquarie University

Sydney, NSW 2109, Australia
sunghwan.kim@mq.edu.au

John K Pate

Department of Computing
Macquarie University

Sydney, NSW 2109, Australia
john.pate@mq.edu.au

Mark Johnson

Department of Computing
Macquarie University

Sydney, NSW 2109, Australia
mark.johnson@mq.edu.au

1. **Ziel:** Die Studie untersucht den Einfluss verschiedener **Dependenzrepräsentationenschemas** auf die Leistung von Sprachmodellen.
2. 3 verschiedenen Dependenzrepräsentationen: **LTH, Penn2Malt und SD**
3. Keine der untersuchten Dependenzrepräsentationen war immer die beste und die Wahl der Dependenzrepräsentation hatte **keinen signifikanten Einfluss** auf die Gesamtleistung.



- 1. Abhängigkeitsstrukturen vs Konstituentenstrukturen**
- 2. Konversion Methode**
- 3. LTH und SD**
- 4. Weitere Forschungen und Anwendungen**



Referenz

- David M. Magerman (1994). Natural language parsing as statistical pattern recognition. Ph.D. thesis, Stanford University.
- Lin, Z. (2019). Constituent-to-dependency conversion. Zi Lin. <https://zi-lin.com/constituent-to-dependency/>
- De Marneffe, M.-C., & Manning, C. D. (2008). The Stanford typed dependencies representation. In Proceedings of the Language Resources and Evaluation Conference (LREC 2008).
- "2 Dependenz und Konstituenz". Valenzgrammatik des Deutschen: Eine Einführung, Berlin, New York: De Gruyter, 2011, pp. 21-43. <https://doi.org/10.1515/9783110254198.21>
- Igor Mel'čuk (1988). Dependency Syntax: Theory and Practice. State University Press of New York, Albany.
- Elming, J., Johannsen, A., Klerke, S., Lapponi, E., Martinez, H., & Søgaard, A. (2013). Down-stream effects of tree-to-dependency conversions.



Referenz

- Kim, S., Pate, J., & Johnson, M. (2014). The Effect of Dependency Representation Scheme on Syntactic Language Modelling. In G. Ferraro & S. Wan (Eds.), Proceedings of the Australasian Language Technology Association Workshop 2014 (pp. 4-13). Melbourne, Australia.
- Johansson, R., & Nugues, P. (2007). Extended Constituent-to-Dependency Conversion for English. In J. Nivre, H.-J. Kaalep, K. Muischnek, & M. Koit (Eds.), Proceedings of the 16th Nordic Conference of Computational Linguistics (NODALIDA 2007) (pp. 105–112).
- Fokkens, Antske (2009). LFG: Syntactic Theory.
<https://www.coli.uni-saarland.de/courses/syntactic-theory-09/slides/LFG4.pdf>
- Sulger, Sebastian (2010).. Kontrolle und finite Nebensätze in LFG.
<https://ling.sprachwiss.uni-konstanz.de/pages/home/butt/main/material/lfg-xcomp.pdf>