Computerlinguistik

E06: Satzebene (2)

IDH INSTITUT FÜR DIGITAL HUMANITIES UNIVERSITÄT ZU KÖLN

Das CONLL-U Datenformat

- CONLL (Conference on Computational Natural Language Learning) ist ein Standard-Format für die Annotation von natürlicher Sprache.
- Tabellarisches Format mit Zeilen für einzelne Tokens und Spalten mit zusätzlichen Informationen:

ID	Form	Lemma	POS-Tag	Features	Head	Relation
1	Peter	Peter	PROPN	Masc|Nom|Sg	2	nsubj
2	sitzt	sitzen	VERB	3|Sg|Pres|Ind	3	root
3	.	.	$.	_	0	

Dependenz-Relationen (Auswahl)

- **root** – vom Satzknoten auf finites Verb.
- **aux** – vom Hilfsverb auf Vollverb
- **nsubj** – vom finiten Verb auf das (Nominativ-)Subjekt
- **obj** – vom Vollverb auf das direkte (Akkusativ-)Objekt
- **iobj** – vom Vollverb auf das indirekte (Dativ-)Objekt
- **obl** – vom Vollverb auf weitere (oblique) Objekte
- **case** – vom Nomen auf dessen Präposition
- **det, amod, num, nmod** – vom Nomen auf Determinierer, Adjektive, Numerale, modifizierende Nomen

Ein Beispielsatz

1	Der	die	DET	Def|Masc|Nom|Sg	2	det
2	Dozent	Dozent	NOUN	Masc|Nom|Sg	3	nsubj
3	ist	sein	AUX	3|Sg|Pres|Ind	13	root
4	mit	mit	APOS	Dat	7	case
5	einem	eine	DET	Indef|Masc|Dat|Sg	7	det
6	einfachen	einfach	ADJ	Pos|Masc|Dat|Sg|Wk|	7	amod
7	Satz	Satz	NOUN	Masc|Dat|Sg	12	obl
8	in	in	APOS	Acc	10	case
9	das	die	DET	Def|Neut|_|Sg	10	det
10	Thema	Thema	NOUN	Masc|Acc|_	12	obl
11	Dependenzparser	Dependenz-Parser	NOUN	Masc|Acc|_	10	nmod
12	gestartet	starten	VERB	_	3	aux
13	.	.	$.	_	0	

Literatur / Hausaufgabe

- **Zur Nachbereitung**:
  - Lesen Sie: Ramers (2007): Kapitel 4 (77-88)
  - Bearbeiten Sie die schriftlichen Aufgaben in ILIAS (folgen im Laufe der Woche).
- Die Texte (bzw. Links) finden Sie im Ilias-Seminarordner.