

Mitte – Links

Satzfunktionen im Lernermittelfeld

Marc Reznicek

Kobalt- Workshop

11.10.2012, Växjö (Schweden)

Übersicht

Hintergrund

- Mittelfeld & Wortstellung
- Betrachtungsebene

Methode

- topologische Felder
- Abhängigkeiten

Quantitativer Vergleich

Ausblick

Hintergrund: Mittelfeld & Variation

VF	LSK	MF	RSK
<i>die arbeitenden Menschen</i>	<i>sollen</i>	<i>[mehr und besser]</i> <i>[ihre Arbeit]</i>	<i>machen</i>

(BY003)

- Die Konstituentenabfolgen im deutschen Mittelfeld sind sehr variabel. (Haider/Rosengren 2003)
- Linearisierung hängt von vielen graduellen Faktoren gleichzeitig ab.
- Wie erwerben L2-Lerner die zielsprachliche Einflussstärke dieser Faktoren? (Kurz 2000, Heylen 2005, Bader/Häussler 2010)

- | | | |
|--|---|---|
| ▪ Grammatische Funktion
Subjekt., dir. Objekt., etc. | ▪ Konstituentengewicht
Wortanzahl, Silbenanzahl | ▪ Belebtheit
belebt, unbelebt, abstrakt |
| ▪ Kasus
Nominativ, Akkusativ ... | ▪ Satztyp
Hauptsatz, Nebensatz | ▪ Definitheit
definit, indefinit |
| ▪ Wortart
Personalpronomen, Vollnomen | ▪ Semantische Rolle
Agens, Patiens, Benefactor | |

Betrachtungsebene: explorativ

Finden L1-spezifische Formen im MF Verwendung?

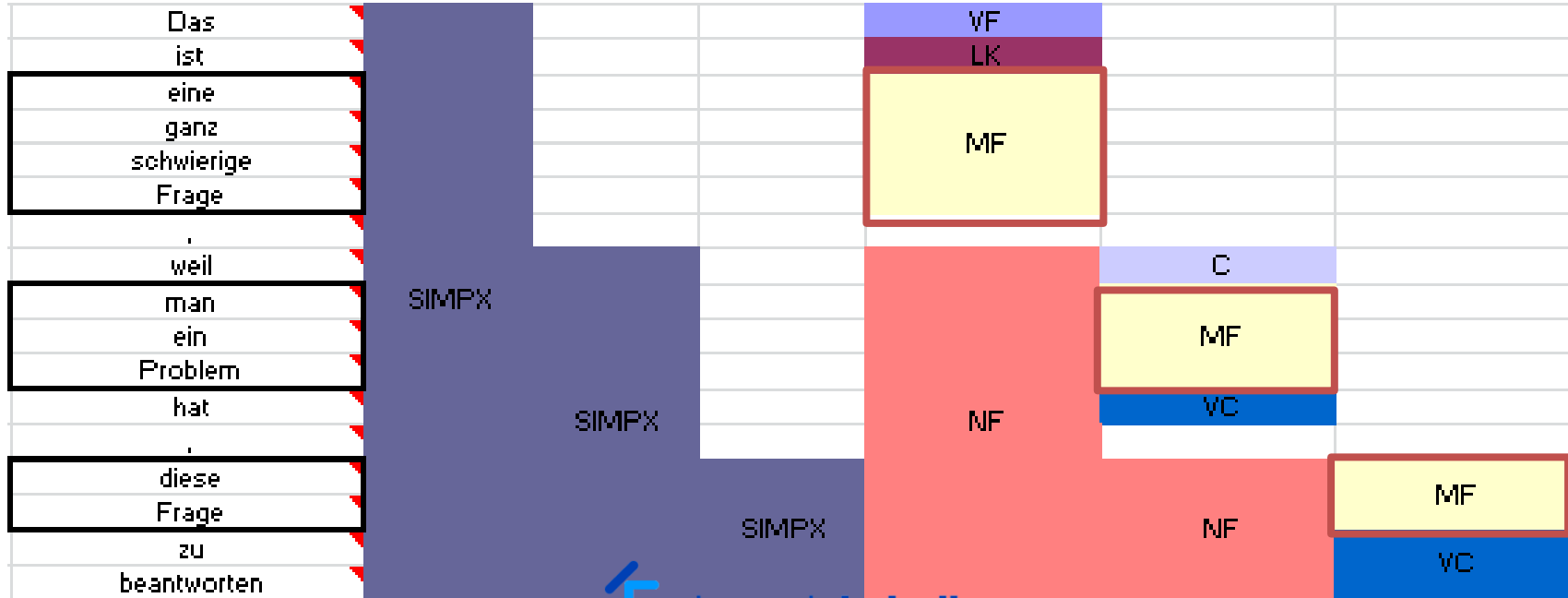
- POS-Tags
- Konstituenten

Werden im MF L1-spezifischen Funktionen realisiert?

- grammatische Funktion
- semantische Rollen

Felderannotation (ZH1)

- **automatische Annotation**
 - Berkeley Parser (Cheung & Penn 2009)
 - TüBa-Schema (Telljohann et al. 2012)
 - Qualität für Pilotstudie akzeptabel



Probleme Felderannotation

ZH1:ZH1	Fields:ZH1TopoFi
?	
Bäume	VF
sind	LK
auch	MF
heutzutage	
unterschätzt	
,	
auf	
eine	
Weise	
genauso	
wie	
das	
ganze	
Konzept	
von	
Stabilität	
Bäume	VF
sind	LK
stabil	MF
,	
wir	VF
nicht	MF
.	

MF vs. NF

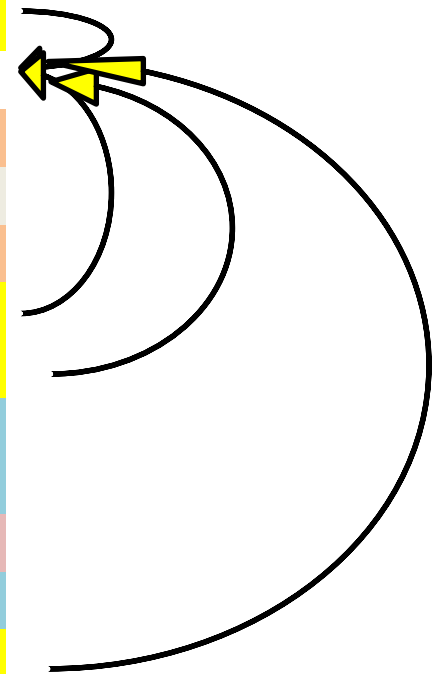
Satz-
klammern
im MF

ZH1:ZH1	Fields:ZH1TopoFi
Arbeit	
.	
Jede	VF
Generation	
hat	LK
ihre	MF
Leiden	
,	
für	VF
die	
heutigen	
Jugendlichen	
ist	LK
das	MF
Leben	
zwar	
besser	
,	
aber	
sie	VF
müssen	MF
noch	
viel	
Schwierigkeit	
durch	
eigene	
Mühe	
auslösen	
.	

Abhängigkeiten

- hierarchische Abhängigkeiten

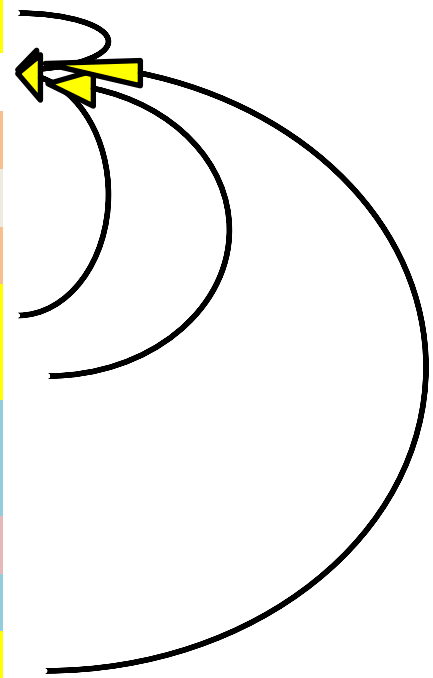
Index	tok	DepEdge
1	Das	2
2	ist	0
3	eine	6
4	ganz	5
5	schwierige	6
6	Frage	2
7	,	2
8	weil	12
9	man	12
10	ein	11
11	Problem	12
12	hat	2
13	,	12
14	diese	15
15	Frage	17
16	zu	17
17	beantworten	11



Abhängigkeiten

- hierarchische Abhängigkeiten
- grammatische Funktionen

Index	tok	DepEdge	DepFunc
1	Das	2	SB
2	ist	0	
3	eine	6	NK
4	ganz	5	MO
5	schwierige	6	NK
6	Frage	2	PD
7	,	2	PUNC
8	weil	12	CP
9	man	12	SB
10	ein	11	NK
11	Problem	12	OA
12	hat	2	MO
13	,	12	PUNC
14	diese	15	NK
15	Frage	17	OA
16	zu	17	PM
17	beantworten	11	OC



MATE (Bohnet 2019)

TiGer (Albert et al. 2003)

Abhängigkeiten

- hierarchische Abhängigkeiten

- grammatische Funktionen

MATE (Bohnet 2019)

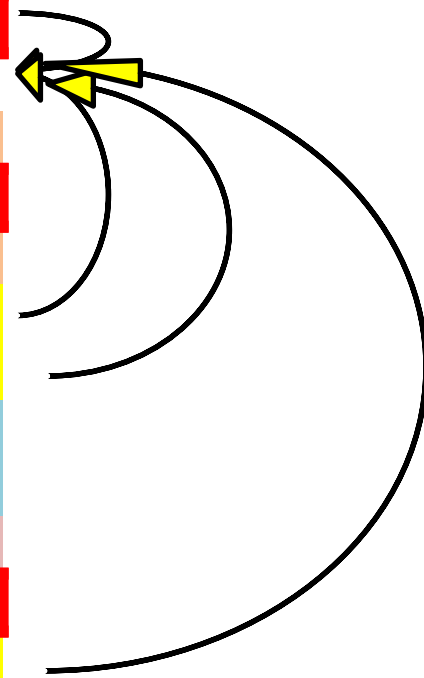
TiGer (Albert et al. 2003)

- **SB** = Subjekt

- **OA** = Akkusativobjekt

- **MO** = Modifikator

Index	tok	DepEdge	DepFunc
1	Das	2	SB
2	ist	0	
3	eine	6	NK
4	ganz	5	MO
5	schwierige	6	NK
6	Frage	2	PD
7	,	2	PUNC
8	weil	12	CP
9	man	12	SB
10	ein	11	NK
11	Problem	12	OA
12	hat	2	MO
13	,	12	PUNC
14	diese	15	NK
15	Frage	17	OA
16	zu	17	PM
17	beantworten	11	OC



Mittelfelder in Pilotkorpus

- Anzahl an Mittelfeldern im Pilotkorpus
 - Weißrussen haben deutlich mehr MF pro Text

DEU	SWE	CNM	BEL	Σ
78	57	47	107	
69	49	83	82	
71	51	62	87	
57	52	65	150	
57	65	59	87	
332	274	316	513	1435

Z.B.: ZH1TopoFields="MF" & ZH1Diff = /MOV./& #1_i_#2 & meta::transcriptionName=/.+_BEL.+/

Abfolgevarianz im Mittelfeld



▪ Extraktion von Abfolgen grammatischer Funktionen

→ sehr hohe Variation

→ Vergleichbarkeit bei Datenmenge problematisch

OC	29	MO-NK-NK-NK-OA	3	NK-OA-MNR-NK-NK
CJ	21	MO-MO	3	SB-MO-NK-OA
NK-OA	17	OA-MO-MO	2	SB-MO-PD
SB	15	MO-MO-MO	2	SB-MO-OA
MO	12	PD-CD-CJ	2	SB-MO-MO-PNC-NK-PUNC-NK-NK-APP
RE	11	MO-NK-NK-NK	2	OA-MO-NK-MO-MO-NK-SB-CD-NK-CJ-NK-AG
RC	9	NK-SB-NK-NK-OA	2	MO-NK-NK-NK-AG
SB-MO	9	MO-MO-MO-NK	2	MO-NK-PD-NK-NK-AG
OA	8	OP-NK-NK	2	OP-NK-NK-NK-NK-AG-CD-CJ-OC
MO-PD	4	MO-NK-NK-OA-MNR-NK-NK	2	MO-NK-PUNC-CJ-PUNC-CJ-CD-CJ
SB-NK-DA-MO-MO-CM-NK-CC	4	SB-MO-NG	2	EP-DA-MO-MO-MO-MO
NK-NK-OA	4	NK-MO-NK-PD	2	MO-NK-SB
MO-NK-OA	4	NK-SB-MO-NG-PD	2	OA-CD-CJ-MO-NK-NK

Abfolgevarianz im Mittelfeld



- **Extraktion von Abfolgen grammatischer Funktionen**
 - sehr hohe Variation
 - Vergleichbarkeit bei Datenmenge problematisch
- **Reduktion auf ein stärker eingegrenztes Phänomen**
 - **Besetzung der ersten Stelle* im MF**

OC	29	MO-NK-NK-NK-OA	3	NK-OA-MNR-NK-NK
CJ	21	MO-MO	3	SB-MO-NK-OA
NK-OA	17	OA-MO-MO	2	SB-MO-PD
SB	15	MO-MO-MO	2	SB-MO-OA
MO	12	PD-CD-CJ	2	SB-MO-MO-PNC-NK-PUNC-NK-NK-APP
RE	11	MO-NK-NK-NK	2	OA-MO-NK-MO-MO-NK-SB-CD-NK-CJ-NK-AG
RC	9	NK-SB-NK-NK-OA	2	MO-NK-NK-NK-AG
SB-MO	9	MO-MO-MO-NK	2	MO-NK-PD-NK-NK-AG
OA	8	OP-NK-NK	2	OP-NK-NK-NK-NK-AG-CD-CJ-OC
MO-PD	4	MO-NK-NK-OA-MNR-NK-NK	2	MO-NK-PUNC-CJ-PUNC-CJ-CD-CJ
SB-NK-DA-MO-MO-CM-NK-CC	4	SB-MO-NG	2	EP-DA-MO-MO-MO-MO
NK-NK-OA	4	NK-MO-NK-PD	2	MO-NK-SB
MO-NK-OA	4	NK-SB-MO-NG-PD	2	OA-CD-CJ-MO-NK-NK

*Es wurden nur die vom teilsatzbestimmenden Verb abhängigen GF betrachtet

Suchen nach linken MF-Rand

ZH1:function& ZH1TopoFields="MF"& #1_I_#2

keine SMS) , sammelte sich auf den Straßen und nicht in Facebook ,
 PIAT NE \$(\$, VVFIN PRF APPR ART NN KON PTKNEG APPR NE \$,
 kein SMS) , sammeln er|es|sie auf d Straße und nicht in Facebook ,

- ctok (grid)
- Root (grid)
- learner (grid)
- TopoFields (grid)

Select Displayed Annotation Levels ▾

ZH1TopoFields	FKONJ	LK	FKOORD
ZH1TopoFields	MF	FKONJ	
ZH1TopoFields	FKOORD		
ZH1TopoFields		VC	MF

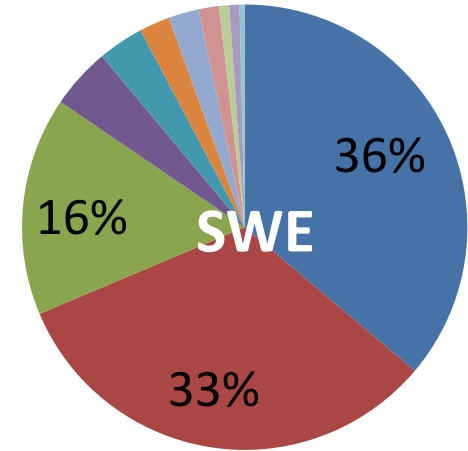
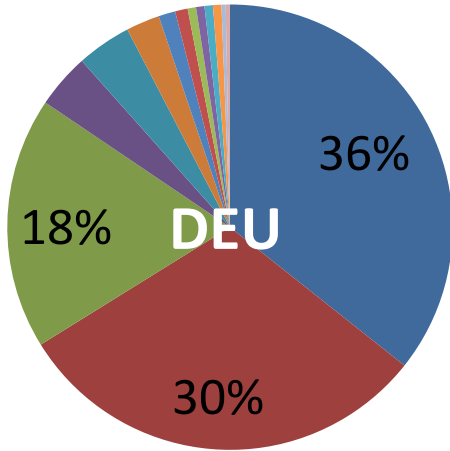
tok keine SMS) , sammelte sich auf den Straßen

TopoFields:ZH1TopoFields = MF

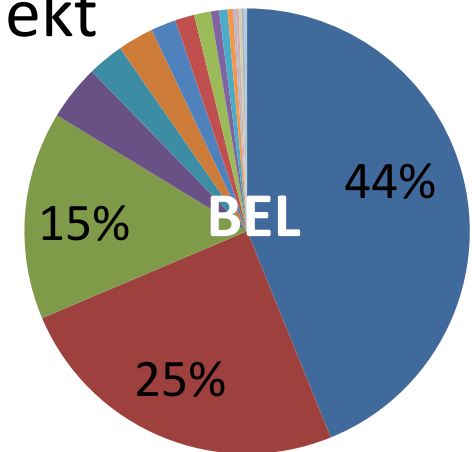
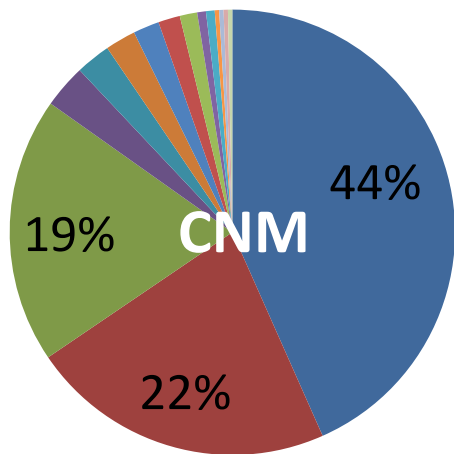
- ZH2 (grid)
- ZH1 (grid)

SMS)	,	sammelte	sich	auf	den	Straßen
68,000000	69,000000	70,000000	71,000000	72,000000	73,000000	74,000000	75,000000
SMS)	,	sammeln	er es sie	auf	d	Straße
NE	\$(,\$	VVFIN	PRF	APPR	ART	NN
Name.*2.*3.*4	Paren.Right	Pun.Comma	Full.3.Sg.Past.Ind	Refl.Subst.3.Acc.Sg.*6	APPR	Def.Dat.Pl.Fem	Reg.Dat.Pl.Fem
NE	\$(,\$	VVFIN	PRF	APPR	ART	NN
MO		PUNC	CJ	OA	MO	NK	NK
SMS)	,	sammelte	sich	auf	den	Straßen

Linke Mittelfeldperipherie: L1/L2



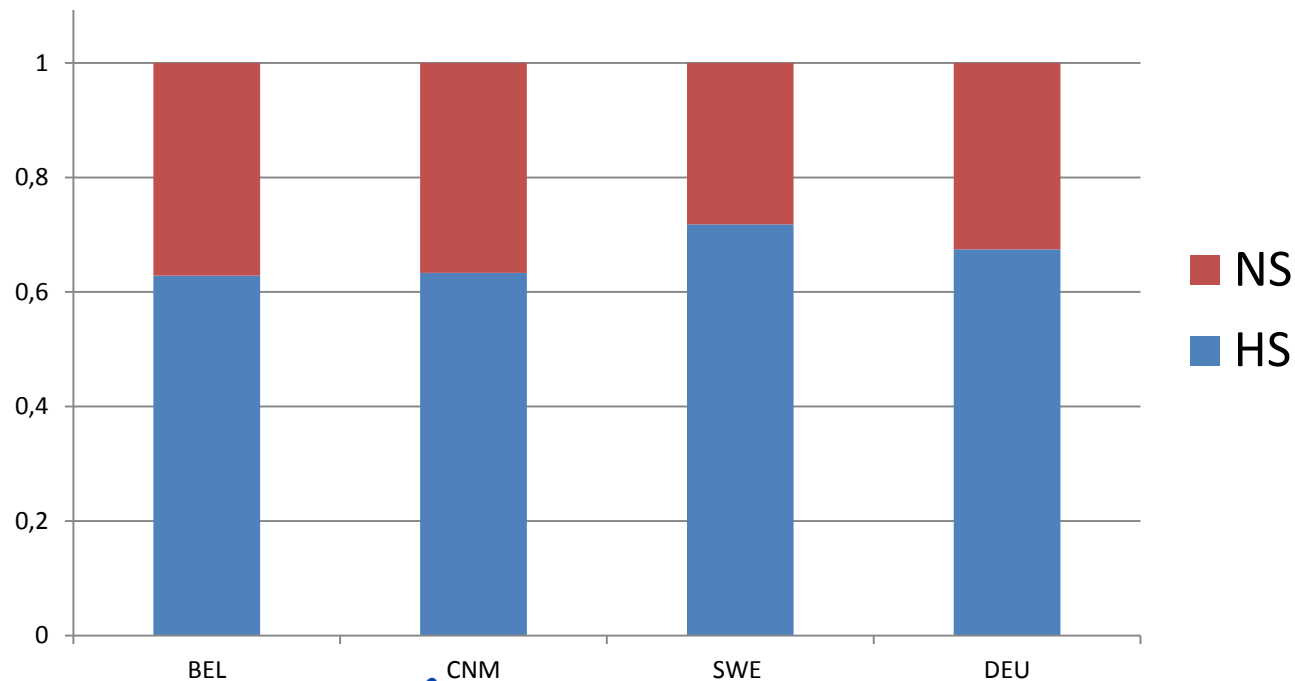
- SB - Subjekt
- MO - Modifikator
- OA - Akkusativobjekt
- DA - Dativobjekt



Satzarten: HS vs. NS

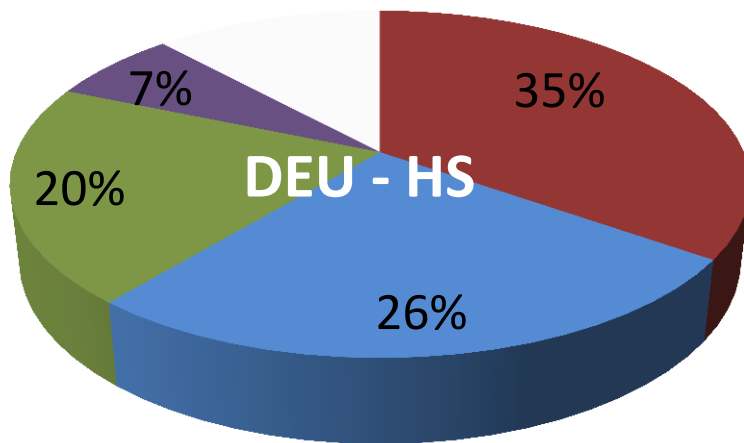
- Satzart ist ein bekannter wichtiger Faktor für die WS im Mittelfeld

Satzartenverteilung im Pilotkorpus

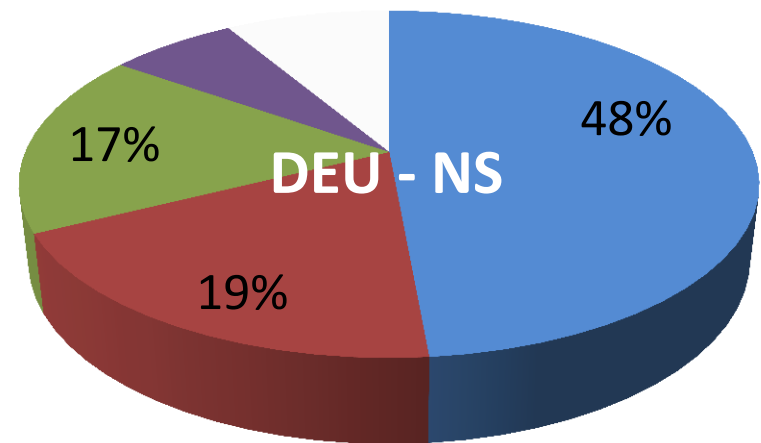


Satztypen: L1

Hauptsatz (HS)



Nebensatz (NS)



- SB - Subjekt
- MO - Modifikator
- OA - Akkusativobjekt
- DA - Dativobjekt

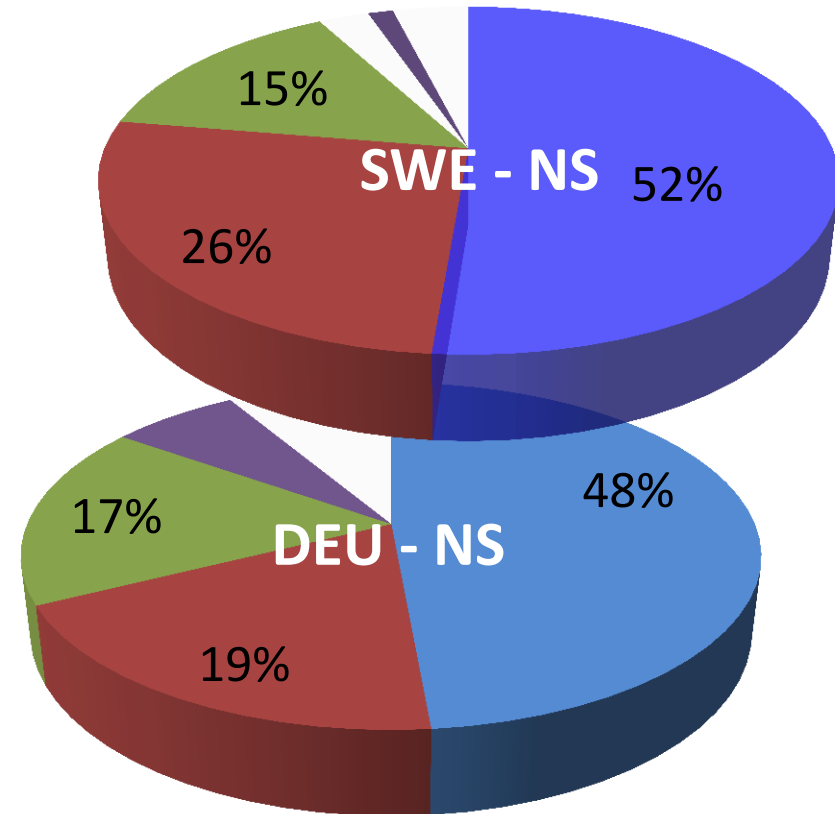
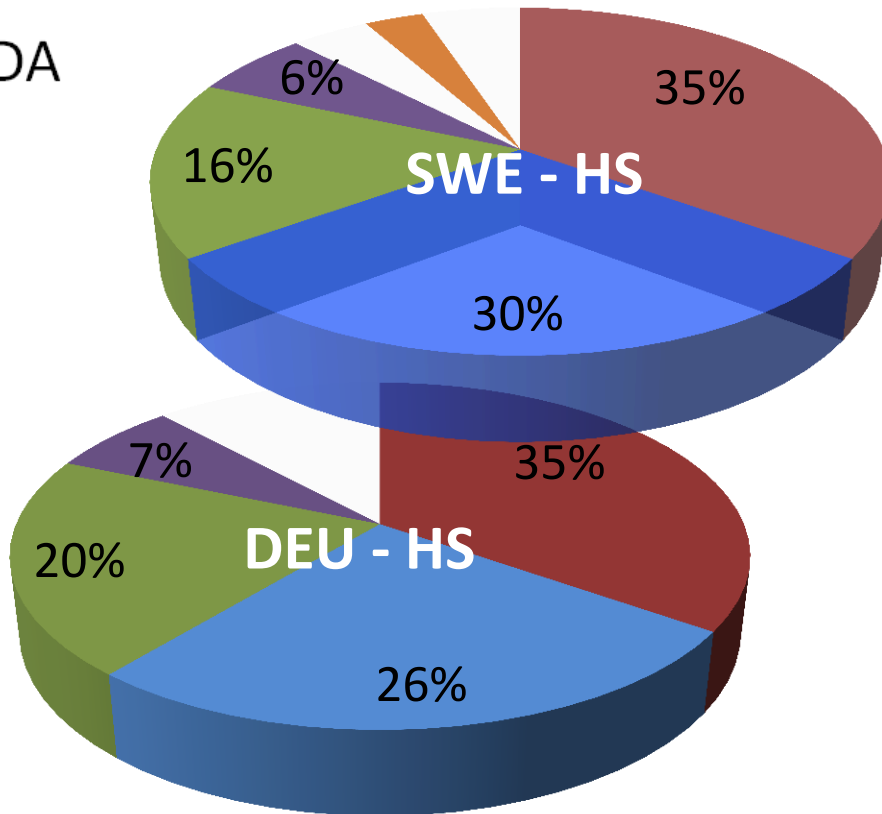
Kobalt_BEL_005_2011_03

Satztypen: schwedisch - deutsch

- SB
- MO
- OA
- DA

Hauptsatz (HS)

Nebensatz (NS)

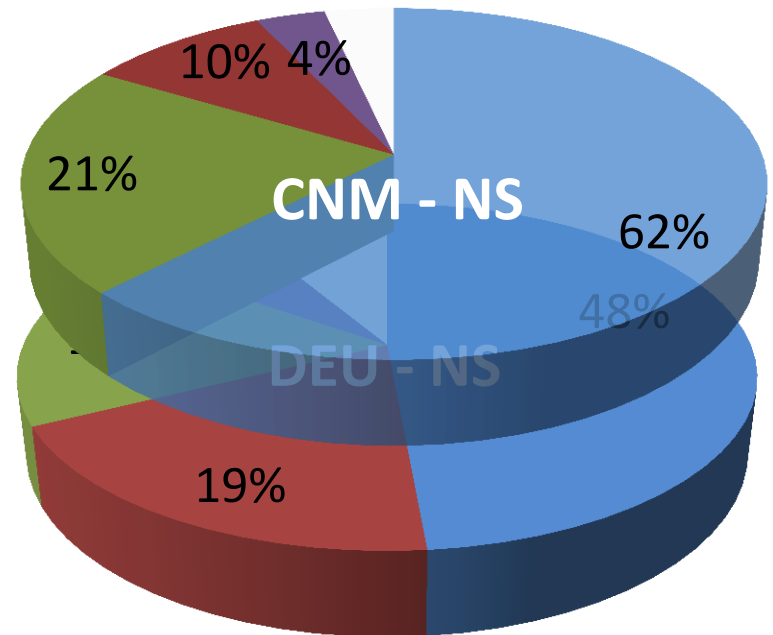
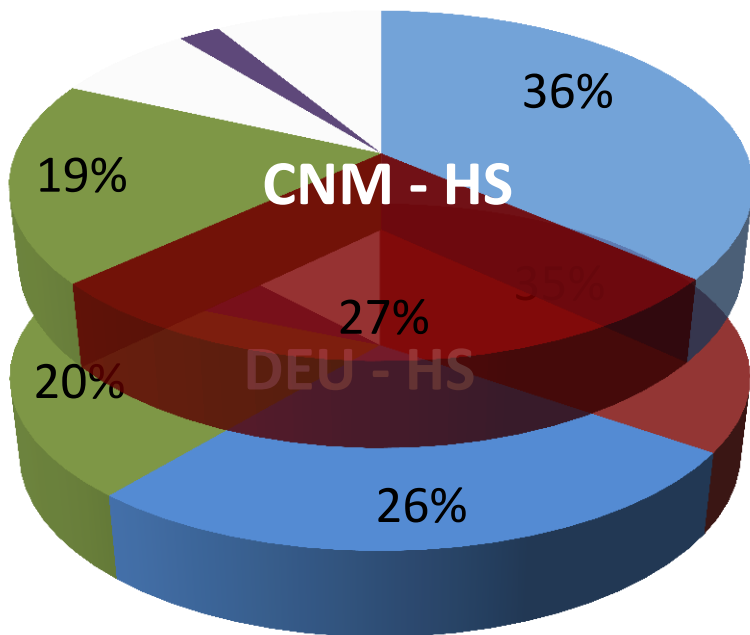


Satztypen: chinesisch - deutsch

- SB
- MO
- OA
- DA

Hauptsatz (HS)

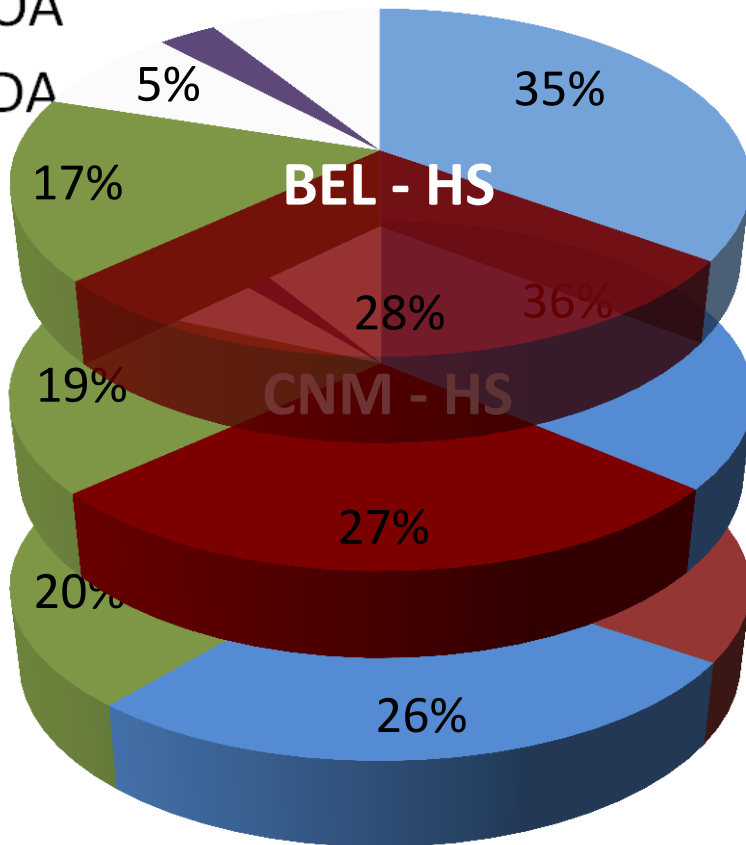
Nebensatz (NS)



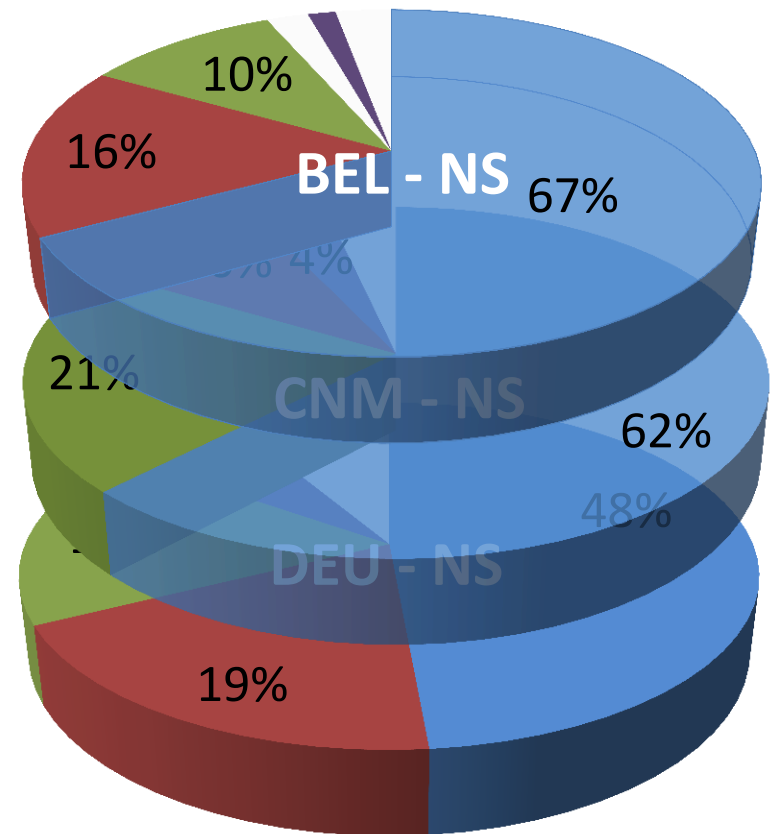
Satztypen: weißrussisch - chinesisch - deutsch

- SB
- MO
- OA
- DA

Hauptsatz (HS)



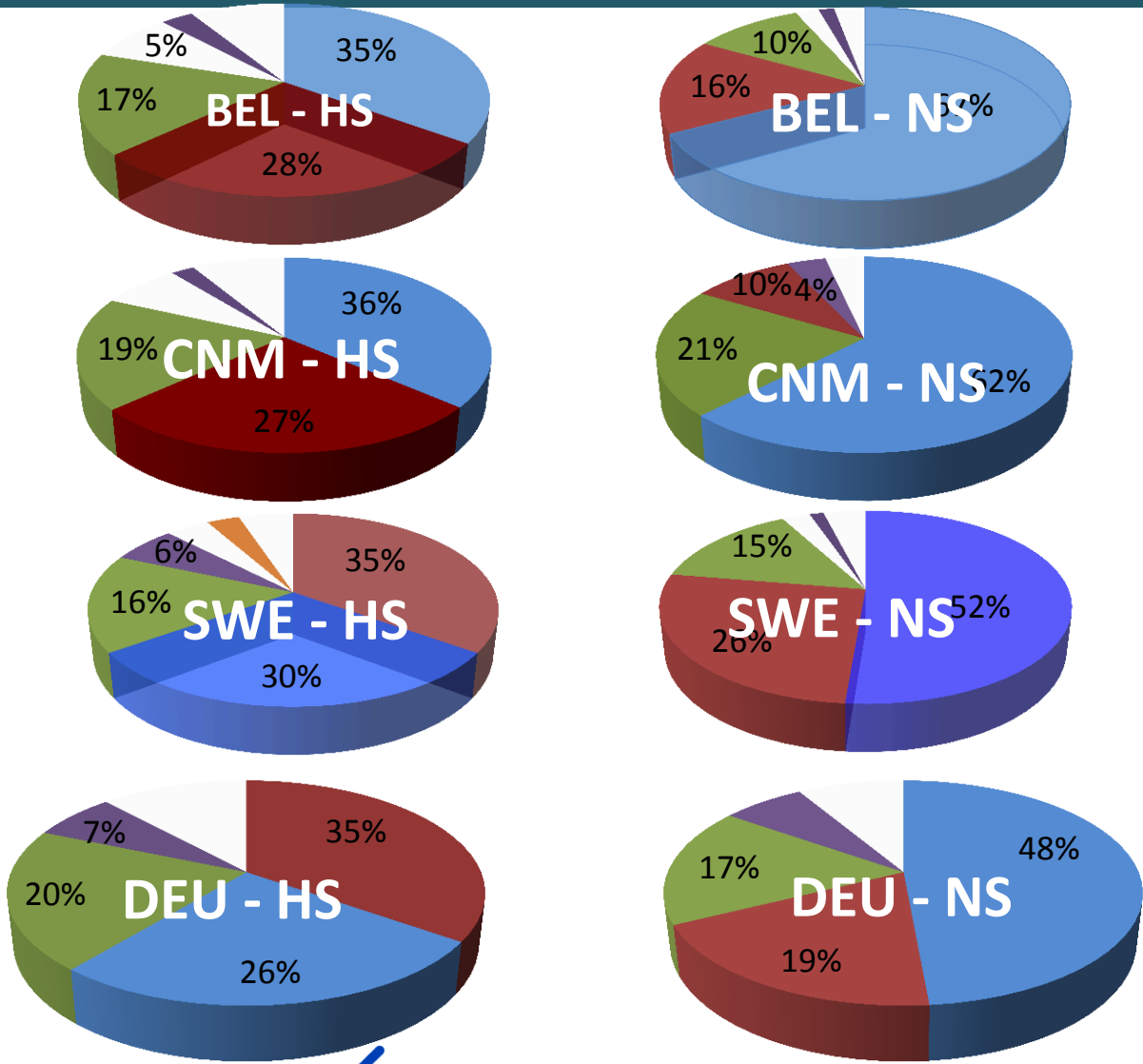
Nebensatz (NS)



Satztypen: weißrussisch – chinesisch – schwedisch - deutsch



- SB
- MO
- OA
- DA



- **ZH0 vergleichen (ZH1 mit Originalwortstellung)!**
- Tiefe der Einbettung miteinbeziehen
- form- und funktionsbasierte Perspektive kombinieren
- graduelle Beschreibung unterschiedlicher Einflüsse auf Linearisierungsabfolgen prüfen
- Wackernagelposition (Form) beachten
- Fehleranalyse durchführen

Vielen Dank!

- Bader, Markus; Häussler, Jana (2010):** Word Order in German. A Corpus Study. Exploring the Left Periphery. In: *Lingua* 120 (3), 717–762. <http://www.sciencedirect.com/science/article/B6V6H-4WN8HF6-1/2/b35ce0bed6dc57349da74641874d9e05>.
- Bohnet, Berndt (2010):** Top Accuracy and Fast Dependency Parsing is not a Contradiction. In: The 23rd International Conference on Computational Linguistics. (COLING 2010). Beijing, China.
- Cheung, Jackie Chi Kit; Penn, Gerald (2009):** Topological Field Parsing of German. In: ACL-IJCNLP '09. Proceedings of the Third Linguistic Annotation Workshop. Morristown, NJ, USA: Association for Computational Linguistics (ACL). www.aclweb.org/anthology/P/P09/P09-1008.pdf.
- Haider, Hubert; Rosengren, Inger (2003):** Scrambling. Non-triggered Chain Formation in OV Languages. In: *Journal of Germanic Linguistics* 15 (03), 203–267. <http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=180383&fulltextType=RA&fileId=S1470542703000291>.
- Heylen, Kris (2005):** A Quantitative Corpus Study of German Word Order Variation. In: Stephan Kepser, Marga Reis (Eds.): *Linguistic Evidence. Empirical, Theoretical and Computational Perspectives*. Berlin, New York: Mouton de Gruyter (85), 241–263.
- Kurz, Daniela (2000):** Wortstellungspräferenzen im Deutschen. Master Thesis. Universität des Saarlandes, Saarbrücken. Computerlinguistik.
- Telljohann, Heike; Hinrichs, Erhard W.; Kübler, Sandra; Zinsmeister, Heike; Beck (2012):** Stylebook for the Tübingen Treebank of Written German (TüBa-D/Z). Seminar für Sprachwissenschaft. Universität Tübingen, Germany. <http://www.sfs.uni-tuebingen.de/resources/tuebadz-stylebook-1201.pdf>.