

# kobalt@ANNIS

## spezielle Suchanfragen

Marc Reznicek

Kobalt- Workshop

11.10.2012, Växjö (Schweden)

# Kurze Wiederholung:

## Prinzip 1: Variablen-Wert-Paare

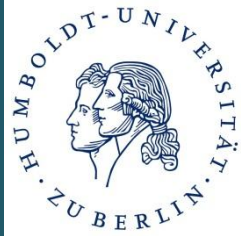
word = "das"

Variable1  
("Wortform")

Wert

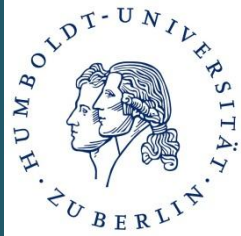
<b>word</b>	Sofern	<b>das</b>	System	herrscht
pos	KOUS	ART	NN	VVFIN
lemma	sofern	d	System	herrschen

# Kurze Wiederholung: Prinzip 2: Beziehungen



- Einzelne **Variable-Wert**-Paare werden durch "&" verbunden.

# Kurze Wiederholung: Prinzip 2: Beziehungen



- Einzelne **Variable-Wert**-Paare werden durch "&" verbunden.
- Zwischen den Paaren muss **IMMER eine Beziehung hergestellt** werden

# Kurze Wiederholung: Prinzip 2: Beziehungen

- Einzelne **Variable-Wert**-Paare werden durch "&" verbunden.
- Zwischen den Paaren muss **IMMER eine Beziehung hergestellt** werden
- Auf die VW-Paare bezieht man sich mit # der Reihe nach.

Variable<sub>1</sub> = Wert<sub>1</sub> & ← Ausdruck 1: #1  
Variable<sub>2</sub> = Wert<sub>2</sub> & ← Ausdruck 2: #2  
#1 "Beziehung" #2

# Variablen im Korpus






Search Form


AnnisQL:

Query Builder:

Result:

History:













<input type="checkbox"/>	Name ▲	Texts	Tokens	
<input type="checkbox"/>	KobaltL1v1.2	5	3458	
<input type="checkbox"/>	KobaltL2v1.2	45	30022	
<input type="checkbox"/>	KobaltL2v1.2BEL	5	4381	
<input type="checkbox"/>	KobaltL2v1.2CNM	5	3098	
<input checked="" type="checkbox"/>	KobaltL2v1.2SWE	5	2887	



Korpusmetadaten-  
Browser  
öffnen

# Variablen im Korpus

Meta Data for KobaltL2v1.2SWE

meta data		available annotations		
Name	Value	node annotations	example (click to use query)	url
projectName	KobaltL2v1.2	ZH1Const	ZH1Const="NX"	
projectURL	<a href="#">Kobalt project site</a>	lemma.	lemma.=","	
		pos.	pos.="NN"	
		ZH1Root	ZH1Root="SIMPX"	
		ZH1TopoFields	ZH1TopoFields="MF"	
		ZH1	ZH1=","	
		ZH1DepID	ZH1DepID="75,000000"	
		ZH1Diff	ZH1Diff="CHA"	
		ZH1S	ZH1S="s8"	
		ZH1lemma	ZH1lemma=","	
		edge types		
		edge annotations		

linke Spalte:

**alle vorhandenen Variablen**

mittlere Spalte:

**häufigster Vertreter**

rechte Spalte:

**fester Link**

# Komplexe Vorfelder: tokenarity

- **tokenarity** erlaubt es, die Menge der Token in einer Spanne zu spezifizieren.

```
ZH1TopoFields="VF"&  
#1:tokenarity=16
```



# Komplexe Vorfelder: tokenarity

- **tokenarity** erlaubt es, die Menge der Token in einer Spanne zu spezifizieren.

ZH1TopoFields="VF"&  
**#1:tokenarity=16**



Path: KobaltL2v1.2SWE > Kobalt\_SWE\_007\_2011\_12

irgendwo auf dem Weg , das heißt auf dem langen Weg von 1950 bis zu 2011 , fangten wir an  
 ADV APPR ART NN \$, PRELS VVFIN APPR ART ADJA NN APPR CARD APPR APPR CARD \$, VVFIN PPER PTKVZ  
 irgendwo auf d Weg , d heißen auf d lang Weg von @card@ bis zu @card@ , fangten wir an

ZH1TopoFields	VF																
ZH1TopoFields					C	VC			NF								
tok	irgendwo	auf	dem	Weg	,	das	heißt		auf	dem	langen	Weg	von	1950	bis	zu	2011

15 **Andere wichtige Aspekte, die dafür sprechen, dass es der Jugend heute besser geht**, ist, dass

14 **Wenn man sich eine höhere Ausbildung in der Universität oder Hochschule wollen**, muss man

# Konstituenten in Feldern:

# i #, # = #, # r #, # | #

- Konstituenten, die im Vorfeld stehen

→ #1 **beinhaltet** #2

```
ZH1TopoFields="VF"&  
ZH1Const= "ADVX"&  
#1_#2
```

# Konstituenten in Feldern:

# i #, # = #, # r #, # | #

- Konstituenten, die im Vorfeld stehen

→ #1 **beinhaltet** #2

ZH1TopoFields="VF"&

ZH1Const= "ADVX"&

#1\_#2

	<b>Heute</b>	besitzen fast alle ...
<b>Die Vielfalt von Fahrmittel</b>	<b>heute</b>	macht auch das Leben leichter
<b>Ein Jugendlicher von</b>	<b>heute</b>	ist ein Lebewesen
	<b>ADVX</b>	
<b>VF</b>		

# Konstituenten in Feldern:

# i #, # = #, # r #, # | #

- Konstituenten, die alleine im Vorfeld stehen  
→ #1 überspannt gleiche Token wie #2

```
ZH1TopoFields="VF"&  
ZH1Const= "ADVX"&  
#1_=_#2
```

# Konstituenten in Feldern:

# i #, # = #, # r #, # | #

- Konstituenten, die alleine im Vorfeld stehen

→ #1 überspannt gleiche Token wie #2

ZH1TopoFields="VF"&

ZH1Const= "ADVX"&

#1\_=\_#2

Heute	gibt es ..
so	scheint es
Leider	ist das
ADVX	
VF	

Heute	gibt	es
Heute	muss	man
Aber	seien	wir
Früher	mußte	man
heute	geht	es
Heute	sind	Bücher
Heute	wählt	man
Dazu	gibt's	
Hier	kann	man
Heutzutage	gibt	es
so	scheint	es
Vielleicht	geht	dieses
dann	beginnen	
Leider	ist	das
Hier	tauchen	andere
So	ist	es
dann	bekomme	ich
Früher	freute	man
dann	ist	die
Einerseits	bin	ich
Heutzutage	entstehen	die

# Konstituenten in Feldern:

# i #, # = #, # r #, # | #

- Konstituenten, am rechten Vorfeldrand

→ #1 teilt rechten Rand mit #2

```
ZH1TopoFields="VF"&
```

```
ZH1Const= "ADVX"&
```

```
#1_r_#2
```

# Konstituenten in Feldern:

# i #, # = #, # r #, # | #

- Konstituenten, am rechten Vorfeldrand

→ #1 teilt rechten Rand mit #2

ZH1TopoFields="VF"&

ZH1Const= "ADVX"&

#1\_r\_#2

	Fast	immer	
	Jetzt	aber	kommt die Frage
Von	da	oben	könnte man sich
		ADVX	
VF			

# eingebettete Felder: #\_i\_#

- **Mittelfelder in Vorfeldern**

```
ZH1TopoFields="VF"&  
ZH1TopoFields="MF"&  
#1_i_#2
```



# eingebettete Felder: #\_i\_#

- **Mittelfelder in Vorfeldern**

```
ZH1TopoFields="VF"&
ZH1TopoFields="MF"&
#1_i_#2
```

Will	<b>man auf sie nicht</b>	verzichten,	kostet ...
Noch eine Sache, ohne die	<b>wir bestimmt nicht</b>	auskommen können,	ist ...
Das Einzige, was	<b>hier vielleicht ein bisschen kompliziert</b>	ist,	ist ...
	MF		
VF			

# mehrfach gefüllte Felder: ZH1Const

- **Zwei Konstituenten im Vorfeld\***

\*bisher haben wir keine hierarchischen Konstituenten

```
ZH1TopoFields="VF"& ZH1Const & ZH1Const &  
#1_i_#2 & #1_i_#3 &  
#2.#3
```

# mehrfach gefüllte Felder: ZH1Const

- **Zwei Konstituenten im Vorfeld\***

\*bisher haben wir keine hierarchischen Konstituenten

ZH1TopoFields="VF"& **ZH1Const** & **ZH1Const** &  
 #1\_i\_#2 & #1\_i\_#3 &  
 #2.#3

Einige	von uns	sind
Das Wort	Depression	gebrauchte man
NX	PX/NX	
VF		

# Nicht-pronominaler NP-Kern

- Token ist ein Appellativ (NN) oder Eigennamen (NE)

ZH1pos=/N./

# Artikelfehler

- fehlender grammatisch obligatorischer Artikel

ZH1pos=/ART/ &  
ZH1Diff = "INS" & #1\_=\_#2

- grammatisch unzulässiger Artikel

ctokpos=/ART/ &  
ZH1Diff = "DEL" & #1\_=\_#2

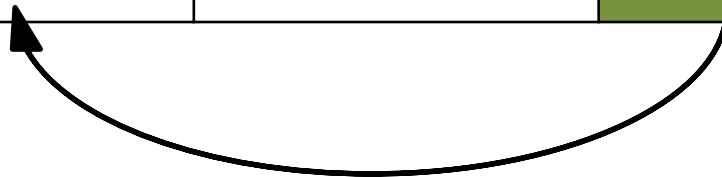
# Bezugswort von Relativsätzen

- Token, von denen Relativsätze abhängen

ZH1 & ZH1 &

#1->dep[func="REL"] #2

Es gibt	<b>Sachen</b>	die an den Gegenteil	deuten
Das Internet ermöglicht auch	<b>die Kommunikation</b>	die alle diese Reisen	fordern
	S		REL



# Bezugswort von Relativsätzen

- **Bezugswort mit Relativsatz im Mittelfeld**

ZH1 & ZH1 &  
 ZH1TopoFields="MF" &  
 #1->dep[func="REL"] #2 &  
 #3\_i\_#2

Wir können	natürlich nicht	<b>der jugendliche Leute,</b>	<b>die in dem 19. Jahrhundert, lebten</b>	fragen, ob sie ...
	MF			
		OA		REL



- Koreferenz (sobald es erste Annotationen gibt)
- Hierarchien in Konstituenten
- Hierarchien in Feldern



# 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](http://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>.