



Korpuslinguistik

Annis₃-Korpussuchtool

Suchen in tief annotierten Korpora

Anke Lüdeling, Marc Reznicek, Amir Zeldes,
Hagen Hirschmann
hirschhx@hu-berlin.de

... und vielen anderen Mitarbeitern der HU-Korpuslinguistik

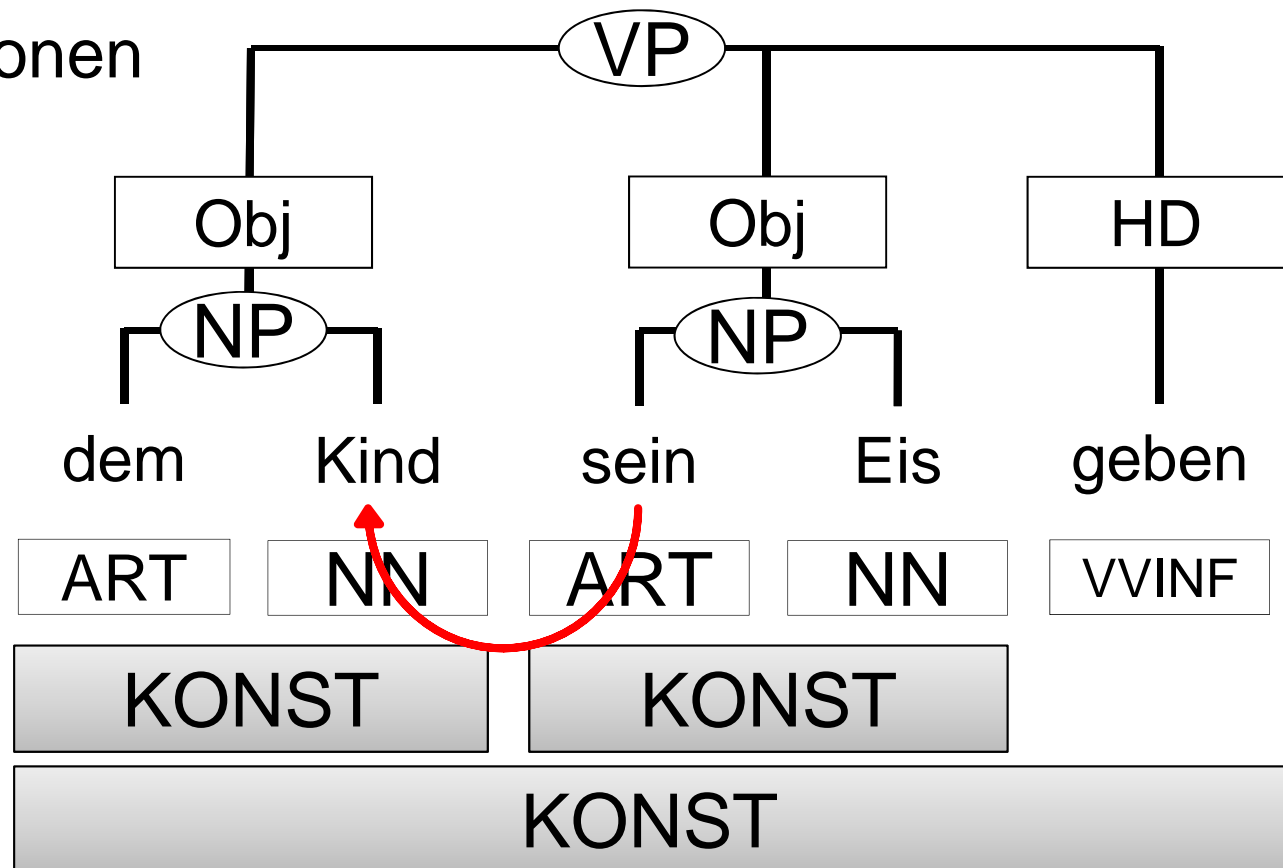
Ziele

- Wie/Was kann man in ANNIS suchen?
 - Wortformen?
 - Token-Annotationen (Lemmata/Wortarten)?
 - Linguistische Muster
(durch Token-Annotationen)?
- Wie sucht man nach mehreren/beliebigen Annotationen gleichzeitig?
- Wie filtert man nach Metadaten?

Beispiel für Annotationen in ANNIS



- Token
- Token-Annotationen
- Spannen
- Bäume
- Pointer



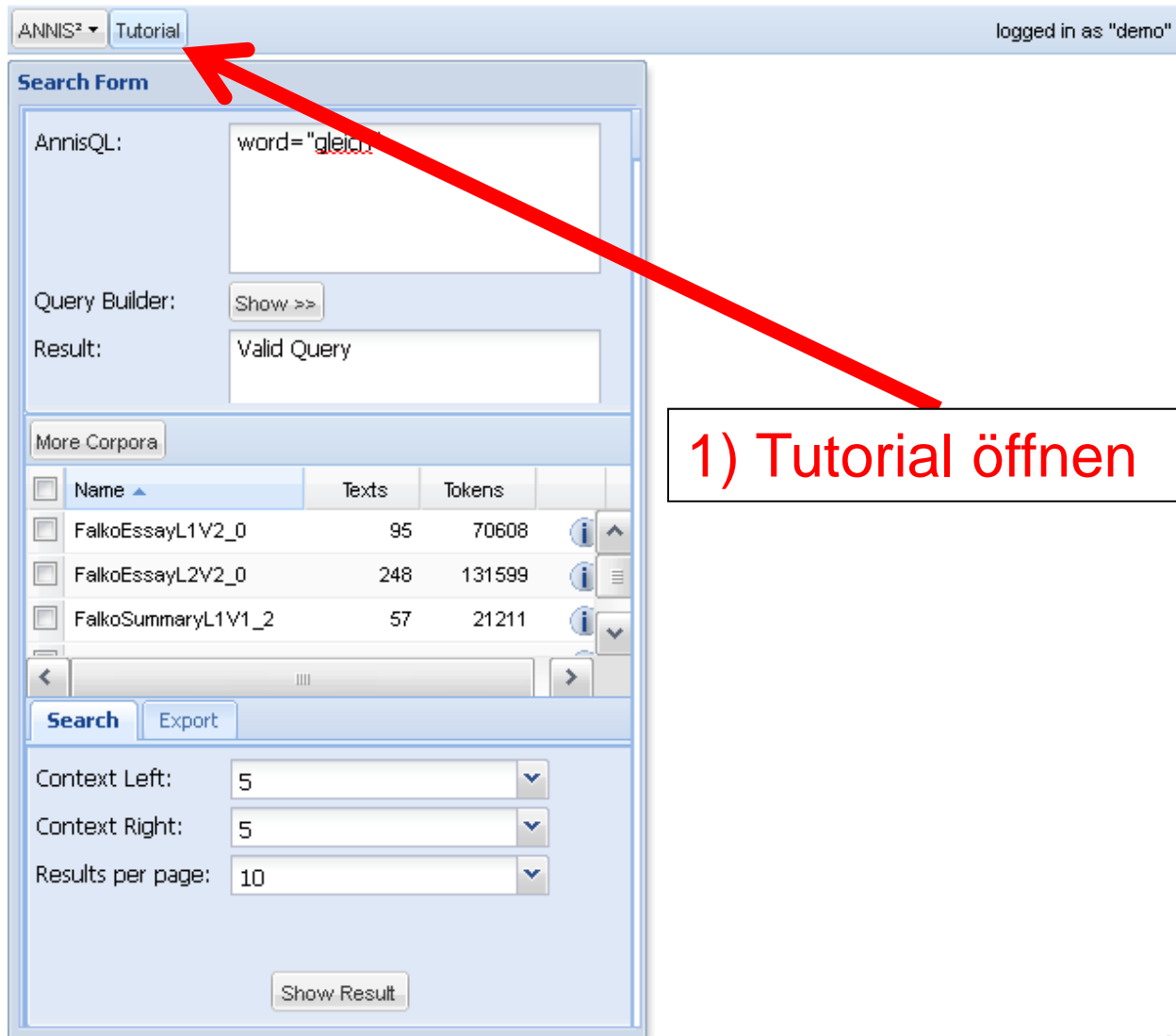
Annis₃



- **Annis-Portal** (öffentlich und log in)
korpling.org/annis3/
- → alle Korpora, die in ANNIS verfügbar sind

- **Annis-Portal speziell für Lernerkorpora**
(Falko, Whig, Bematac L2, Kobalt DaF)
korpling.org/falko-suche/

Das Web-Interface: Tutorial



ANNIS² Tutorial logged in as "demo"

Search Form

AnnisQL: word="gleich"

Query Builder: Show >>

Result: Valid Query

More Corpora

<input type="checkbox"/>	Name	Texts	Tokens	
<input type="checkbox"/>	FalkoEssayL1V2_0	95	70608	i ^
<input type="checkbox"/>	FalkoEssayL2V2_0	248	131599	i ≡
<input type="checkbox"/>	FalkoSummaryL1V1_2	57	21211	i v

Search Export

Context Left: 5

Context Right: 5

Results per page: 10

Show Result

1) Tutorial öffnen

Das Web-Interface: Tutorial

Get help about: Using the ANNIS2 interface

Using the ANNIS2 Interface

The ANNIS2 interface is comprised of several windows, the most important of which are the search form and the results window.

The Search Form

2) Thema wählen

Search Form

AnnisQL: "gleich"

Query Builder: Show >>

Result: Valid Query

More Corpora

<input type="checkbox"/>	Name	Texts
<input type="checkbox"/>	FalkoEssayL1V2_0	95
<input type="checkbox"/>	FalkoEssayL2V2_0	248
<input type="checkbox"/>	FalkoSummaryL1V1_2	57

Search Export

Context Left: 5

Context Right: 5

Results per page: 10

Show Result

Search Form

AnnisQL: cat="S" & node & #1
>secedge[func="SB"] #2

Query Builder: Show >>

Result: Valid Query

More Corpora

<input checked="" type="checkbox"/>	Name	Texts	Tokens	
<input checked="" type="checkbox"/>	GermanDiachronicTreeb...	35	9540	i
<input type="checkbox"/>	WikiNews	1	130	i
<input type="checkbox"/>	arabic.test	1	11	i

Das Web-Interface: Abfrage

<http://korpling.german.hu-berlin.de/falko-suche/search.html>

ANNIS? Tutorial logged in as "demo"

Search Form

AnnisQL: "gleich"

Query Builder: Show >>

Result: Valid Query

More Corpora

<input type="checkbox"/>	Name	Texts	Tokens	
<input type="checkbox"/>	FalkoSummaryL1V2_0	95	70608	
<input checked="" type="checkbox"/>	FalkoSummaryL2V2_0	248	131599	
<input type="checkbox"/>	FalkoSummaryL1V1_2	57	21211	
<input type="checkbox"/>	FalkoSummaryL2V1_2	107	40865	
<input type="checkbox"/>	FalkoSummaryLV1_0	12	11114	

Search Export

Context Left: 5

Context Right: 5

Results per page: 10

Show Result

Suchefenster

Auswahl der Korpora für die Suche (STRG+Klick für Mehrfachauswahl)

Abfrage starten

Das Web-Interface: Suchfenster

Search Form

AnnisQL: word = "gleich"

Query Builder: Hide <<

Result: 39

More Corpora

1) Suchanfragefenster

2) Trefferanzahl

The image shows a screenshot of a search interface titled "Search Form". It is divided into three main sections: "AnnisQL:", "Query Builder:", and "Result:". The "AnnisQL:" section contains the text "word = 'gleich'", where the word "gleich" is underlined in red. A red arrow points from a white box containing the text "1) Suchanfragefenster" to this underlined word. The "Query Builder:" section contains a button labeled "Hide <<". The "Result:" section contains the number "39". A red arrow points from a white box containing the text "2) Trefferanzahl" to the number "39". At the bottom of the interface, there is a button labeled "More Corpora".

Das Web-Interface: Such-Einstellungen

The image shows a web interface for search settings. At the top, there are two buttons: "Search" and "Export". The "Search" button is circled in red. Below these buttons are three input fields: "Context Left:" with the value "5", "Context Right:" with the value "5", and "Results per page:" with the value "10". Each of these three fields has a red arrow pointing to it from a text box on the right. The text boxes contain the following labels: "linker Kontext" (pointing to Context Left), "rechter Kontext" (pointing to Context Right), and "Resultate pro Seite" (pointing to Results per page). At the bottom of the interface is a button labeled "ShowResult", with a red arrow pointing to it from a text box labeled "Suche starten".

Context Left:	5	linker Kontext
Context Right:	5	rechter Kontext
Results per page:	10	Resultate pro Seite

ShowResult

Suche starten

Das Web-Interface: Export-Einstellungen

The screenshot shows a web interface with two tabs: 'Search' and 'Export'. The 'Export' tab is selected and circled in red. Below the tabs, there are four sections: 'Exporter:', 'Context Left:', 'Context Right:', and 'Parameters:'. The 'Exporter:' section has a dropdown menu with 'WekaExporter' selected. A red arrow points from a box labeled 'Format wählen' to the dropdown menu. Below the dropdown menu, there is a 'Perform Export' button. A red arrow points from a box labeled 'Export starten' to this button.

Search Export

Exporter: WekaExporter

Context Left: SimpleTextExporter

Context Right: TextExporter

Parameters: WekaExporter

GridExporter

Perform Export

Format wählen

Export starten

Das Web-Interface: Treffer

Textname

word, POS, lemma

Weitere Ebenen

Partitur

Volltext

Textmetadaten

Korpusmetadaten

ANNIS² Tutorial

Search Form

AnnisQL: word = "deutsch"

Query Builder: Show >>

Result: 8

History: Query History

More Corpora

Name	Text	Tokens
<input type="checkbox"/> FalkoEssayL1v2.2		95 70648
<input type="checkbox"/> FalkoEssayL2v2.0		248 132066
<input checked="" type="checkbox"/> FalkoEssayL2v2.2		248 132069
<input type="checkbox"/> FalkoSummaryL1v1_2		57 2
<input type="checkbox"/> FalkoSummaryL2v1_2		107 40
<input type="checkbox"/> FalkoSummaryVLv1_0		12 114
<input type="checkbox"/> FalkoWHIGL2v1.0		92 63496

Search Result - word = "deutsch" (5, 5)

Page 1 of 1

Token Annotations

Citation URL

Document Path

Displaying Results 1 - 8 of 8

Path: FalkoEssayL2v2.2 > fk022_2006_08_L2v2.2

der wirkliche Welt , oft **deutsch** zu sehen ist . Während

ART ADJA NN \$, ADV ADJD PTKZU VWINF VAFIN \$, KOUS

d wirklich Welt , oft deutsch zu sehen sein . während

ZH2 (grid)

ZH0 (grid)

ZH1 (grid)

Select Displayed Annotation Levels

ZH1	der	wirklichen	Welt	,	oft	deutsch	zu	sehen	ist	.	Während
ZH1Diff		CHA		DEL							
ZH1S		s7									s8
ZH1gpos	ART	ADJA	NN		ADV	ADJD	PTKZU	VWINF	VAFIN	\$.	ZH1:ZH1S = s7
ZH1gposDiff				DEL							
ZH1lemma	d	wirklich	Welt	,	oft	deutsch	zu	sehen	sein	.	während
ZH1lemmaDiff				DEL							
ZH1pos	ART	ADJA	NN		ADV	ADJD	PTKZU	VWINF	VAFIN	\$.	Art
ZH1posDiff				DEL							
tok	der	wirkliche	Welt	,	oft	deutsch	zu	sehen	ist	.	Während

ZHverb (grid)

falko (grid)

leamer (grid)

text (discourse)

Praxisorientierte Studienabschlüsse , nein danke Die meisten Universitätsabschlüsse sind nicht praxisorientiert , gerade deswegen sind sie von besonderem Wert ! Die Universitätsstudienwege geben den Studenten Rahmen und Richtungen , aber nicht das ganze Inhalt . Das Praxis lernt man einfach am besten später , wenn man schon eine Stelle habe , oder ein Praktikum macht . Aber wenn man schon berufstätig ist , kann man kaum nicht so viele Theorien und hat weniger Zeit um zu experimentieren . Die Frage , Praxisorientierung oder nicht , wird an meiner Uni häufig diskutiert . Ich studiere ein Fach , Architektur , wo die Trennung zwischen Theorie und Praxis , zwischen dem Studium und der wirkliche Welt , oft **deutsch** zu sehen ist . Während meines Studiums mache ich mehrere Luftschlösser , ja nicht selten unmögliche Entwürfe , die wahrscheinlich nicht einem Statiker gefallen wurden . Aber meine Entwürfe sind ja noch nicht für die Realität geeignet , sondern sind sie Übungen um so sehen wie man vielleicht etwas machen könnte . Falls ich mich , während des Studiums , am besten für eine Stelle und ein Berufstätigkeit , vorbereiten wollte , sollte ich mich viel mehr mit Bauhandlungen , Vermessung , Bauökonomie und Präsentationen von Projekten beschäftigen . Überhaupt nicht mit den Raumtheorien . Aber so lernt man nicht die Architektur und ich möchte Achitektur studieren , nicht nur lernen Architektin zu werden . Um ein Fach im Griff zu bekommen braucht man oft zahlreiche Theorien , Texte , Bilder und Übungen - und gerade dies sind die Stärken einer Uni . Wenn man ein Fach schon kennt und am liebsten auch von dem Fach verführt worden ist , kann man ziemlich einfach das Praxisorientierte Arbeit lernen . Umgekehrt ist es schwieriger , aber ich denke nicht unmöglich . Mir gefällt es , an einer Uni zu sein . Ich will mein Studium benutzen um zu übertreiben und überprüfen , dazu möchte ich viel lesen und Zeichnen und auf diese Weise für das Berufsleben vorbereitet zu werden , obwohl es mir während meines Praktikums ein bisschen peinlich war , dass ich z. B noch nicht wusste welche Massen man auf einem Plan schreiben sollte . Die Universität bietet Bildung an . Bildung ist was man übrig hat , wenn man alles man lernte vergessen hat . Deswegen sind Die Universitätsabschlüsse von besonderem Wert .

Prinzip I: Variablen-Wert-Paare

■ word = "das"

Variable1
("Wortform")

Wert

word	Sofern	das	System	herrscht
pos	KOUS	ART	NN	VVFIN
lemma	sofern	d	System	herrschen

Prinzip I: Variablen-Wert-Paare

■ word = "das"

...findet *das* (und nichts anderes)

Prinzip I: Variablen-Wert-Paare

■ pos = "ART"

Variable2
("Wortart")

Wert

word	Sofern	das	System	herrscht
pos	KOUS	ART	NN	VVFIN
lemma	sofern	d	System	herrschen

Prinzip I: Variablen-Wert-Paare

■ pos = "NN"

...findet *Riesen, Frauen, Student, ...*

Prinzip I: Variablen-Wert-Paare

■ lemma = "d"

Variable
("Lemma")

Wert

word	Sofern	das	System	herrscht
pos	KOUS	ART	NN	VVFIN
lemma	sofern	d	System	herrschen

Prinzip I: Variablen-Wert-Paare

■ lemma = "d"

...findet *die, dem, den, ...*

beliebig erweiterbar...

■ **satz** = "NS"

Variable4
("Satztyp")

Wert

word	Sofern	das	System	herrscht
pos	KOUS	ART	NN	VVFIN
lemma	sofern	d	System	herrschen
satz	NS			

beliebig erweiterbar...

- **satz = "NS"**

...findet alle Nebensätze wie
Sofern das System herrscht

(sofern die Daten wie gezeigt annotiert sind)

Suche nach Strings

- **Suchen Sie nach allen Vorkommen der Wortform "meinen" in FalkoEssayL2V2.4:**

```
word = "meinen"
```

- Was wird gefunden?
- Ist das interessant?
- Was wird nicht gefunden, was interessant sein könnte?

Lemmata

- "Basisformen" von Wörtern
- **Suchen Sie nach allen Vorkommen der Formen des Verbs *meinen*:**

lemma = "meinen"

- → Problem: Lemmatisierung ist willkürlich; man muss wissen, wie lemmatisiert wurde.
- Beispiel: Lemma von *sich*

Lemmata

- "Basisformen" von Wörtern
- **Suchen Sie nach allen Vorkommen der Formen des Possessivartikels:**

lemma = "mein"

Mustersuche (reguläre Ausdrücke)

- Annis₂ erlaubt Mustersuchen auf allen Annotationsebenen
- Mustersuchen werden statt in " " in // eingefügt
- Z. B. kann man damit nach allen Wörtern suchen, die *...mein...* enthalten.

```
word = /.*mein.*/
```

Mustersuche: **Joker** .

- ein beliebiges Zeichen al. → *als*, *alt*, ...
- ■ zwei beliebige Zeichen al.. → *alle*, *alte*, *also*
- ■ ■ drei beliebige Zeichen al... → *alles*, *altes*,
alias, ...

Aufgabe

- Welche Wortformen bekommen Sie mit?

word = /g.b./

Mustersuche: ? und * +

das[↩]s?

das vorherige Zeichen ist optional

→ ϕ , s → da, das

das[↩]s*

das vorh. Zeichen kommt 0- bis ∞ mal vor

→ ϕ , s, ss, ... → da, das, dass, dassssssssss

das[↩]s+

das vorh. Zeichen kommt 1- bis ∞ mal vor

→ s, ss, ... → das, dass, dassssssssssss

Aufgabe

- Was passiert, wenn Sie die Operatoren kombinieren?

```
word = /Frau.??/
```

```
word = /Frau.* /
```

```
word = /Frau.+ /
```

Aufgabe

- **Versuchen Sie alle Wörter (Grundformen) zu finden, die auf *-lang* enden.**

Aufgabe

- Versuchen Sie alle Wörter (Grundformen) zu finden, die auf *-lang* enden.



lemma = `/.+lang/`

Treffer z.B.:

bislang

lebenslang

jahrelang

Aufgabe

- Versuchen Sie alle Wörter (Grundformen) zu finden, die mit *lang-* beginnen.

→ lemma = /lang.+/

Treffer z.B.:

lange

langsam

langweilig

Suche nach Wortart

- Es gibt unterschiedliche Wortartensysteme (→ Tagsets) für Korpora
- allgemein in der Linguistik unterschiedliche Wortartensysteme
- Die meisten deutschen Korpora benutzen das Tagset STTS

<input type="checkbox"/> ADJA	attributives Adjektiv
<input type="checkbox"/> ADV	Adverb
<input type="checkbox"/> ART	Artikel
<input type="checkbox"/> NN	normales Nomen
<input type="checkbox"/> VVFIN	finites Verb

...

<http://www.ims.uni-stuttgart.de/projekte/corplex/TagSets/stts-table.html>

Stuttgart-Tübingen-Tagset (STTS)

ADJektiv	Nomen	Pronomen	Verb	Partikel	Konjunktion
ADJA	NN	PDS	VFIN	PTKZU	KOUI
ADJD	NE	PDAT	VVIMP	PTKNEG	KOUS
		PIS	VVINF	PTKVZ	KON
		PIAT	VVIZU	PTKANT	
		PIDAT	VVPP	PTKA	
		PPER	VAFIN		
		PPOSS	VAIMP		
		PPOSAT	VAINF		
		PRELS	VAPP		
		PRELAT	VMFIN		
		PRF	VMINF		
		PWS	VMPP		
		PWAT			
		PWAV			

Stuttgart-Tübingen-Tagset (STTS)

VERB	Vollverb	Auxiliar	Modalverb
finit	VV FIN	V AFIN	V MFIN
Imperativ	VV IMP	V AIMP	
infinit	VV INF	V AINF	V MINF
Infinitiv mit <i>zu</i>	VV IZU		
Partizip 2	VV PP	V APP	V MPP

Aufgabe

- Suchen Sie nach Possesivpronomen.



```
pos =/PPOS(S|AT)/
```

Prinzip II: Relationen

- 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.

Variable₁ = Wert₁ & ← Ausdruck 1: #1
Variable₂ = Wert₂ & ← Ausdruck 2: #2
#1 "Beziehung" #2

Prinzip II: Relationstypen

$\text{Variable}_1 = \text{Wert}_1$ &  Ausdruck 1: #1
 $\text{Variable}_2 = \text{Wert}_2$ &  Ausdruck 2: #2
 #1 "Beziehung" #2

Operator	Description	Illustration	Notes
.	direct precedence	A B	For non-terminal nodes, precedence is determined by the right most and left most terminal children
.*	indirect precedence	A x y z B	For specific sizes of precedence spans, .n,m can be used, e.g. .3,4 - between 3 and 4 token distance
=	identical coverage	A B	Applies when two annotations cover the exact same span of tokens
i	inclusion	AAA B	Applies when one annotation covers a span identical to or larger than another
>	direct dominance	A B	A specific edge type may be specified, e.g.: >secedge to find secondary edges. Edges labels are specified in brackets, e.g. >[func="OA"] for an edge with the

Aufgabe

- Finden Sie nun Vorkommen von `word =/mein(e|st|t|en)/`, die ausschließlich finite Vollverben sind.

`word =/meinen?/ &`

`pos= "VVFIN" &`

`#1 _ = _ #2`



Ausdruck 1: #1



Ausdruck 2: #2



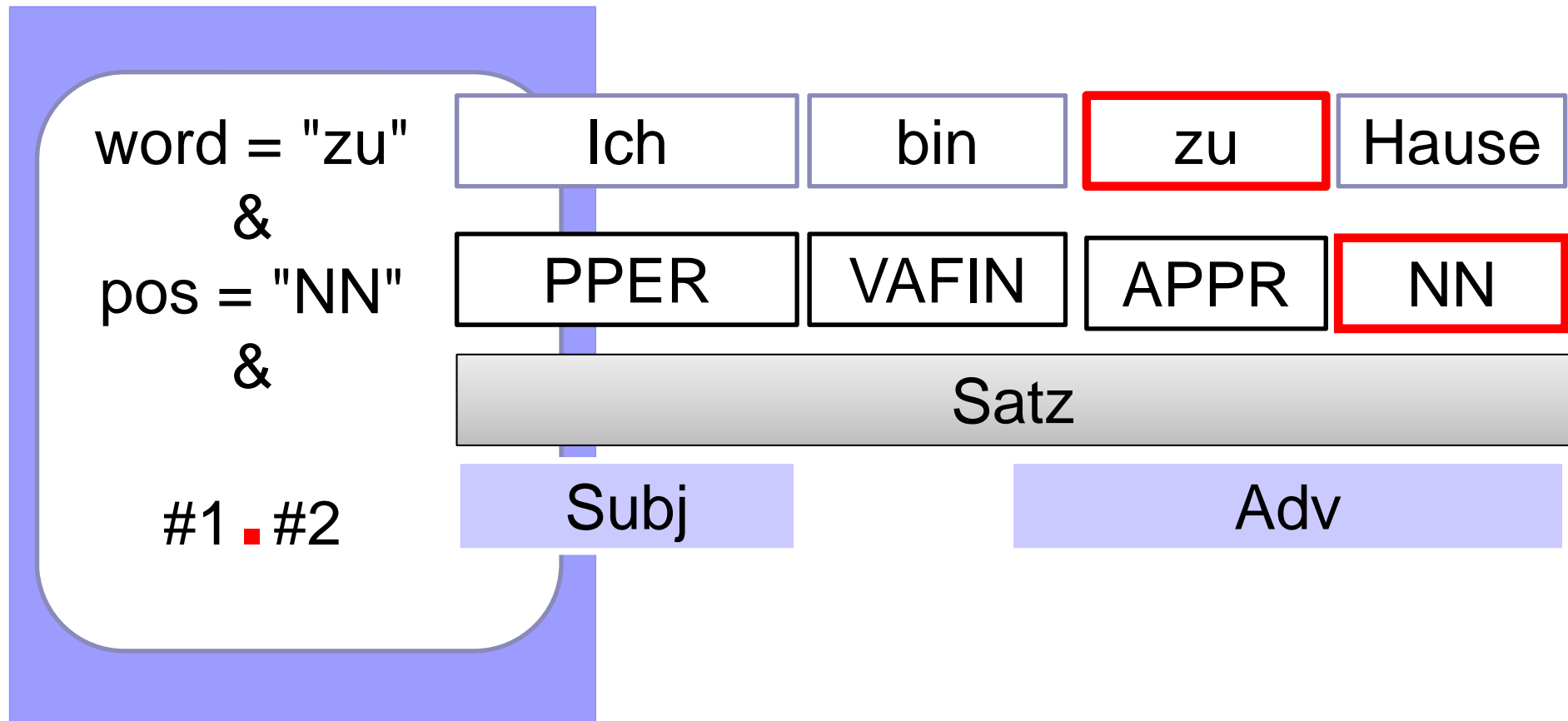
Relation: Ab-
deckung

Negation !=

- **!** bedeutet Negation
 - Der Operator wird vor dem "="-Zeichen eingefügt.
 - **Finden Sie in alle Vorkommen von**
word =/mein(e|st|t|en)/ , die nicht das Lemma
"mein" haben.

```
word =/mein(e|st|t|en)/ &  
lemma!="mein" &  
#1 _=_ #2
```

Suche nach Abfolgen: z.B. Nomen folgt auf "zu"



Tokenfolgen - Aufgabe

- **Suchen Sie nach zwei aufeinanderfolgenden Adjektiven.**
- **Achtung: Es gibt zwei Typen von Adjektiven**
 - ADJA & ADJD

```
pos = /ADJ./ &  
pos = /ADJ./ &  
#1.#2
```


Zielhypothesen

Unterschiede zwischen Zielhypothese und Originaltext sind durch "edit tags" auf der Diff-Ebene markiert.

ZH1lemma	weil	sie		ein	Aspekt	d	Gesellschaft	entdecken	,
ZH1Diff			MOVS	CHA	CHA			MOVT	
ZH1pos	KOUS	PPER		ART	NN	ART	NN	VVPP	,\$
ZH1	weil	sie		einen	Aspekt	der	Gesellschaft	entdeckt	,
word	weil	sie	entdeckt	eine	Aspekte	der	Gesellschaft		,

ZH1lemma	wie	d	ander	Frau
ZH1Diff			CHA	
ZH1pos	KOKOM	ART	ADJA	NN
ZH1	wie	die	anderen	Frauen
word	wie	die	andere	Frauen

Edit Tags

ZHDiff	Operation in Zielhypothese
INS	Token eingefügt
DEL	Token gelöscht
CHA	Token geändert
MERGE	mehrere Token verbunden
SPLIT	Token in mehrere aufgespalten
MOVS	Token von hier bewegt
MOVT	Token hierhin bewegt

Aufgabe

- Finden Sie alle Reflexivpronomen, die in den Lernertexten fehlen (erst einmal theoretisch).
- Die Ebene für die ZH1-Wortart heißt ZH1pos. Die Ebene für die Edit Tags heißt ZH1Diff.

```
ZH1pos="PRF" &  
ZH1Diff="INS"&  
#1_=_#2
```

Aufgabe

- ...alle indefiniten Artikel, die in den Lernertexten fehlen.

Lösung:

```
ZH1lemma="d"&  
ZH1Diff="INS"&  
#1_=_#2
```

Metadaten (Informationen über den jeweiligen Text) finden

The screenshot displays the ANNIS interface for the document 'falkoEssayL2v2.4'. On the left, a 'Corpus List' table shows various documents with their respective text and token counts. The document 'falkoEssayL2v2.4' is highlighted in blue. A red arrow points from this row to a central panel titled 'Corpus information for falkoEssayL2v2.4 (ID: 7121)'. This panel is divided into two sections: 'Metadata' and 'Available annotations'. The 'Metadata' section shows the document name 'Falko Essay L2 v2.4'. The 'Available annotations' section lists various annotations such as 'ctok', 'ctoklemma', 'ctokpos', 'fm', 'lemma', 'macro', 'pos', 'topo1', 'topo2', 'TXTstructure', 'verbfehlertyp', 'Edge types', and 'Meta Annotations'. A second red arrow points from the 'Available annotations' section to a text box that says 'text- & lernerbezogene Metadaten anzeigen'. At the bottom of the interface, there is a link to the corpus: https://korpling.german.hu-berlin.de/falko-suche/#_c=ZmFsa29Fc3NheUwydjuNA.

Name	Texts	Tokens		
CLEG13	729	285.286	i	
FalkoEssayL1v2.0	94	70.110	i	
falkoEssayL1v2.3	95	70.615	i	
FalkoEssayL2v2.0	248	132.066	i	
FalkoEssayL2v2.3	248	131.628	i	
falkoEssayL2v2.4	248	144.619	i	
FalkoEssayL2WHIGv2.0	195	130.187	i	
FalkoGeorgetownL2v1.0	92	78.151	i	
FalkoSummaryL1v1.2	57	21.211	i	

Name	Value
projectName	Falko Essay L2 v2.4

Name	Example (click to use query)	URL
ctok	ctok=","	
ctoklemma	ctoklemma="d"	
ctokpos	ctokpos="NN"	
fm	fm="fm:eng"	
lemma	lemma="d"	
lemma.	lemma.="d"	
macro	macro="title"	
pos	pos="NN"	
pos.	pos.="NN"	
topo1	topo1="MF"	
topo2	topo2="VCE"	
TXTstructure	TXTstructure="PAR"	
verbfehlertyp	verbfehlertyp="flex"	

Alle Annotationen und Metadaten anzeigen

text- & lernerbezogene Metadaten anzeigen

Link to corpus: https://korpling.german.hu-berlin.de/falko-suche/#_c=ZmFsa29Fc3NheUwydjuNA

Metadaten finden

- Metadaten:
 - Variablen und Werte für
 - Text
 - Lerner
- Klicken Sie Beispiele an, um sie ins Suchfenster zu überführen

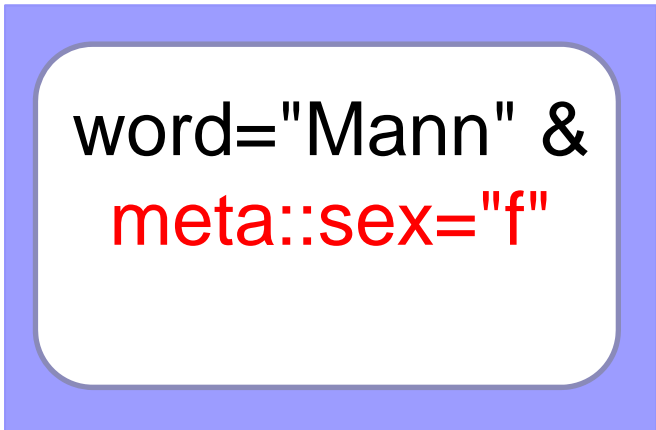
Field	Value
SPK0:l2_1_duration	96
SPK0:l2_1_langschool	N/A
SPK0:l2_1_school	N/A
SPK0:L2index	deu:96:N/A:N/A:N/A:N/A:N/A
SPK0:name	9d3abd29757a2f3f0244c7f26c50b440e7b302a1
SPK0:reg	l1:eng,l1:N/A,l1:N/A,l1:N/A,l1:N/A,l1:N/A,l1:N/A,l1:N/
SPK0:sex	f
major-subject	deutschintensiv
production-modality	essay
subcorpus	fk_2006_07
topic	Feminismus
transcription-d	27.07.06
transcriptionNa	fk007_2006_07
transcripator	MMM/V
projectName	FALKO Essay Corpus L2 2.0
projectURL	Falko project site

Nach Metadaten filtern

- Nach Metadaten sucht man mit `meta::Variable = "Wert"`

Finden Sie alle Wortformen von "Mann", die von weiblichen Lernern geschrieben wurden.

(Die Variable für Geschlecht ist `sex [m,f]`).



```
word="Mann" &  
meta::sex="f"
```

Nach Metadaten filtern

- Nach der Muttersprache von Lernern sucht man mit

- `meta::reg=/l1:LÄNDERCODE.*`

- Finden Sie alle

Formen des Adjektivs *deutsch*

in den Texten

englischer Mutter-

sprachler (Code= `eng`)

```
lemma="deutsch"&  
meta::reg=/l1:eng.*
```


Nach Metadaten filtern

- Soll die Sprachbiographie genauer beschrieben werden, muss zwischen beiden Informationen ein `.` stehen.
- `meta::reg=/Variable1:Wert1.*Variable2:Wert2/`
- Finden Sie alle Formen von "deutsch" in den Texten dänischer Muttersprachler mit L2 Englisch.

```
lemma="deutsch"&  
meta::reg=/l1:dan.*l2:eng./
```

Ein erster Vergleich

- Vergleichen Sie die Häufigkeit der "ung"-Derivationen bei Dänen und Italienern
- Wichtig: Wieviele Tokens gibt es pro Sprachgruppe
- Wieviele Ungs pro Token gibt es also?

Gesamtanzahl der Token für die Muttersprachen

```
word&  
meta::reg=/l1:dan.*/
```

```
word&  
meta::reg=/l1:ita.*/
```

Syntax

(authentische) Studienarbeit

- Fragestellung: Werden relativische Anschlüsse unabhängig von der syntaktischen Funktion des Relativpronomens erworben?
- Hypothesen:
 - 1. Die syntaktischen Funktionen, die relativische Anschlüsse haben können, werden alle simultan erworben.
 - 2. Der Erwerb von Relativsätzen beschränkt sich lange auf die Verwendung des Prototyps, bei dem das Relativpronomen Subjektfunktion hat.

Theoretische Vorarbeit

- Welche syntaktischen Funktionen kann das Relativpronomen einnehmen?
 - Subjekt
 - Objekt (AKK)
 - Objekt (DAT)
 - innerhalb PP (Objekt)
 - innerhalb PP (Adverbial)
 - innerhalb *von*-PP (Passiv, log. Subj)
 - Genitivattribut (zu Subjekt, Objekt, ...)

Einfachste Struktur (Subjekt)

- Relativpronomen finden...
- Wie formuliert man aber eine Suchanfrage, die Relativpronomen findet, welche innerhalb des Relativsatzes in Subjektposition stehen?
- Relativsätze findet man verhältnismäßig einfach mit der Kantenbeziehung "RC" zwischen dem Kopf (V.FIN) des übergeordneten und dem Kopf des untergeordneten (V.FIN) Satzes.
- Es geht aber viel einfacher: PRELS/PRELAT genügt als Indikator für "RC" (Redundanz)

Einfachste Struktur (Subjekt)

- Relativpronomen in Subjektposition:
- POS=/V.* / & POS=/PREL.* / &
#1 ->dep[deprel="SB"] #2
- Statt Subjektposition ist auch Objektposition als
Akkusativobjekt (=OA) oder Dativobjekt (=DA)
möglich.
- POS=/V.* / & POS=/PREL.* / &
#1 ->dep[deprel="OA"] #2
- POS=/V.* / & POS=/PREL.* / &
#1 ->dep[deprel="DA"] #2

Weiteres Vorgehen

- Auf diese Weise (mit Abhängigkeitsbeziehungen) können alle Funktionen des Relativpronomens ausgedrückt werden.
- Diese werden jeweils zwischen Falko-L2 und Falko-L1 verglichen.
- Annahme: Die Diskrepanz (schon allgemein beim Vergleich von PRELS/PRELAT sichtbar) steigt, je tiefer die Einbettung des Pronomens.

Zusammenfassung - Operatoren

.	Ein beliebiges Zeichen
*	Beliebig viel (0 bis unendlich vom vorherigen Element)
+	Mindestens einmal (vorheriges Element)
?	Optional (vorheriges Element)
\	wörtlich (folgendes Zeichen)
!	nicht
[abc]	Menge (oder $[^abc]$ = alles <i>außer</i> die Menge)
(a b)	a oder b
a{2,3}	a 2 bis 3 mal

Zusammenfassung

- Mit ANNIS kann man:
 - in unterschiedlichen Korpora (auch gleichzeitig) suchen
 - die Ergebnisse quantifizieren
 - die Ergebnisse exportieren
- Man kann auch nach Metadaten filtern (Tutorial)

Sprachkürzel in Falko (Auswahl)

afr	afrikaans	nor	norwegisch
dan	dänisch	pol	polnisch
deu	deutsch	rus	russisch
ell	neugriechisch	spa	spanisch
eng	englisch	swe	schwedisch
fin	finnisch	tur	türkisch
fra	französisch	ukr	ukrainisch
heb	hebräisch	uzb	usbekisch
hun	ungarisch	xho	xhosa
isl	isländisch	yid	jiddisch
ita	italienisch	zho	zulu
jpn	japanisch		
lat	lateinisch		



Zusammenfassung

- Operatoren zu Tokenrelationen:

#1.#2

#1 wird direkt gefolgt von #2.

#1.*#2

#1 wird indirekt gefolgt von #2.

#1_=_#2

#1 und #2 beziehen sich auf die gleichen Token.

#1_i_#2

#1 ist in #2 enthalten.

Herzlichen Dank!

Literatur

- **Lüdeling, Anke; Doolittle, Seanna; Hirschmann, Hagen; Schmidt, Karin; Walter, Maik (2008):** Das Lernerkorpus Falko. In: *Deutsch als Fremdsprache* 45 (2), S. 67–73.
- **Reznicek, Marc; Walter, Maik; Schmidt, Karin; Lüdeling, Anke; Hirschmann, Hagen; Krummes, Cedric; Andreas, Thorsten (2010):** Das Falko-Handbuch. Korpusaufbau und Annotationen. Version 1.0. Berlin: Institut für deutsche Sprache und Linguistik, Humboldt-Universität zu Berlin. URL: <http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falko> [Stand: 12. Oktober 2010].
- **Zeldes, Amir; Ritz, Julia; Lüdeling, Anke; Chiarcos, Christian (2009):** ANNIS. A Search Tool for Multi-Layer Annotated Corpora. In: *Proceedings of Corpus Linguistics 2009, Liverpool, July 20-23, 2009*.