



Das Falko-Korpus im ANNIS-Suchinterface der Humboldt-Universität zu Berlin

15.03.2022
Hagen Hirschmann

Online-Referenzen: Korpusdokumentation und -zugang

- Falko-Dokumentationshomepage:
<https://hu.berlin/falko>
- Falko-Handbuch:
www.linguistik.hu-berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko/FalkoHandbuchV2/
- Falko-Suchinterface (ANNIS):
<https://korpling.german.hu-berlin.de/falko-suche/>
- ANNIS-Homepage (hier auch lokale Version von ANNIS erhältlich):
www.corpus-tools.org/annis

Das Suchinterface ANNIS

- **Im Korpus-Suchinterface ANNIS wird Falko (mit allen Annotationen) durchsuchbar gemacht**
 - **ANNIS-Portal** → alle Korpora, die an der HU in ANNIS verfügbar sind:
<https://hu.berlin/annis>
 - **ANNIS-Portal speziell für Lernerkorpora**
→ Falko, WHiG, BeMaTaC L2, Kobalt-DaF, KanDel:
<https://hu.berlin/annis-falko>

Falko und andere Korpora durchsuchen...

- Was kann man in ANNIS suchen?
 - Wortformen
 - Token-Annotationen:
 - Lemmata
 - Wortarten
 - dadurch: Linguistische Muster
 - Sätze und Satztypen
 - Korpuspezifische Kategorien, z.B. im Falko-Korpus:
 - strukturelle Abweichungen, basierend auf Zielhypothesen
- Zusätzlich Filterung nach diversen Metadaten

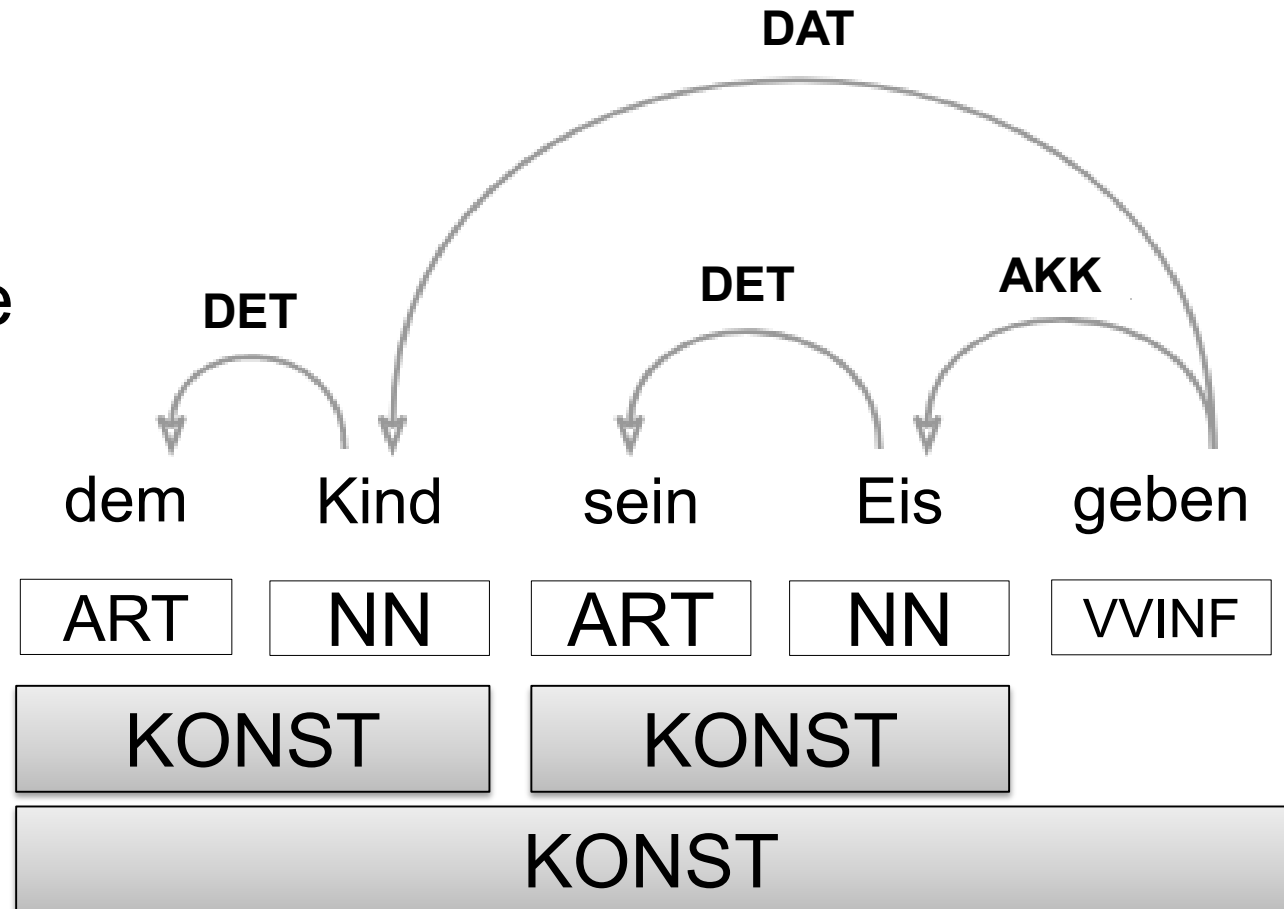
Beispiele für Annotationen in ANNIS

• syntakt. Bäume

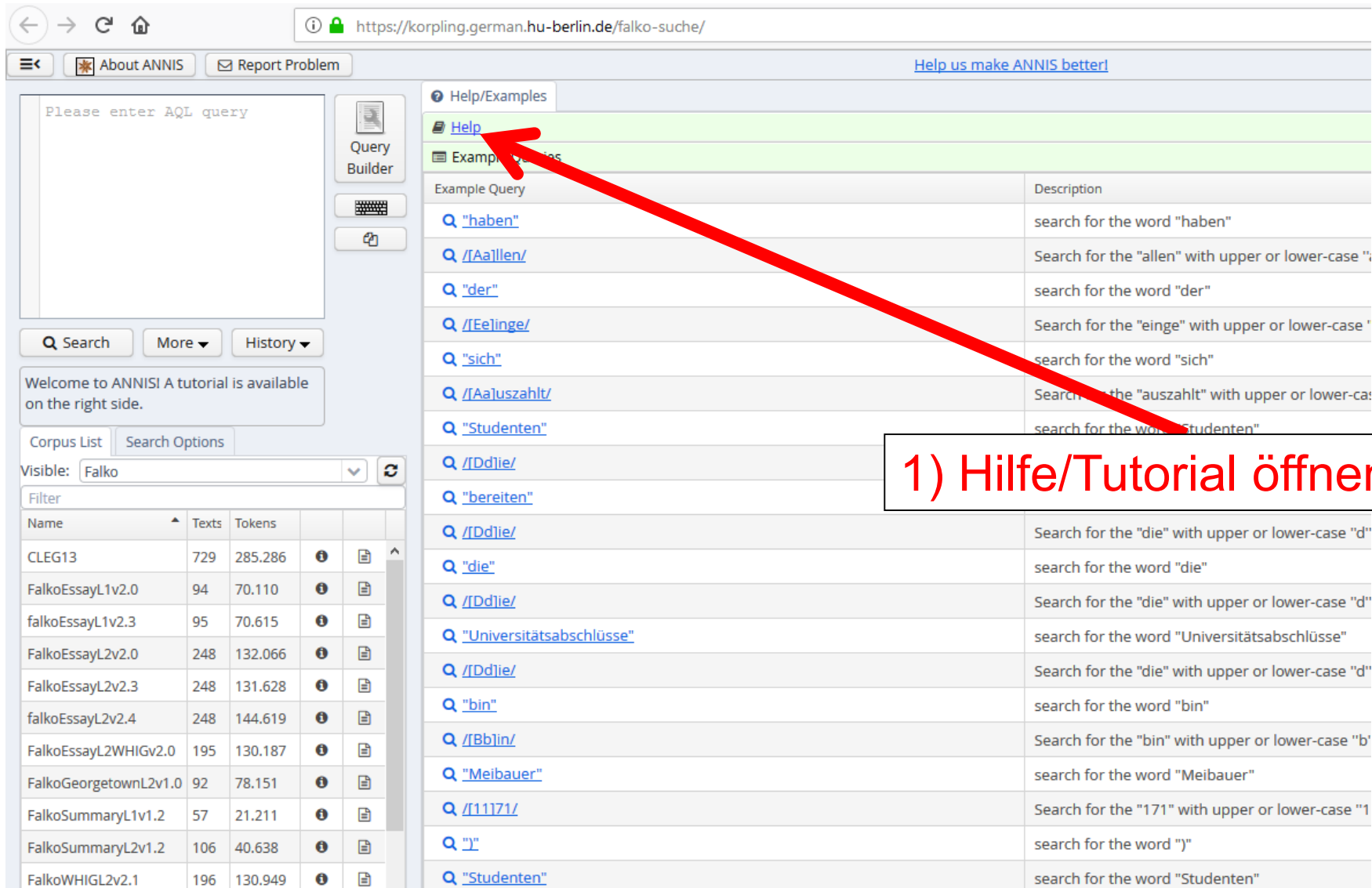
• Token

• Token-Annotationen

• Spannen



Das Web-Interface: Tutorial



The screenshot shows the ANNIS web interface. On the left, there is a 'Query Builder' section with a text input field for an AQL query and buttons for 'Search', 'More', and 'History'. Below this is a 'Corpus List' section with a 'Visible:' dropdown set to 'Falko' and a 'Filter' input field. A table lists various corpora with columns for Name, Texts, and Tokens. On the right, there is a 'Help/Examples' section with a 'Help' link highlighted in green. Below this is a table of 'Example Query' and 'Description' pairs. A red arrow points from the 'Help' link to a text box containing the instruction '1) Hilfe/Tutorial öffnen'.

Help/Examples

[Help](#)

Example Query

Example Query	Description
Q "haben"	search for the word "haben"
Q /fAa llen/	Search for the "allen" with upper or lower-case "
Q "der"	search for the word "der"
Q /fEe linge/	Search for the "einge" with upper or lower-case "
Q "sich"	search for the word "sich"
Q /fAa uszahlt/	Search for the "auszahlt" with upper or lower-ca
Q "Studenten"	search for the word "Studenten"
Q /fDd ie/	
Q "bereiten"	
Q /fDd ie/	Search for the "die" with upper or lower-case "d"
Q "die"	search for the word "die"
Q /fDd ie/	Search for the "die" with upper or lower-case "d"
Q "Universitätsabschlüsse"	search for the word "Universitätsabschlüsse"
Q /fDd ie/	Search for the "die" with upper or lower-case "d"
Q "bin"	search for the word "bin"
Q /fBb in/	Search for the "bin" with upper or lower-case "b"
Q "Meibauer"	search for the word "Meibauer"
Q /f11 71/	Search for the "171" with upper or lower-case "1"
Q "I"	search for the word "I"
Q "Studenten"	search for the word "Studenten"

1) Hilfe/Tutorial öffnen

Das Web-Interface: Tutorial

The screenshot shows the ANNIS web interface. On the left, there is a sidebar with a search input field, a 'Query Builder' button, and a table of corpus items. The main content area displays the 'ANNIS Help' page, specifically the 'Complete List of Operators' section. A red arrow points from the 'Complete List of Operators' link in the sidebar to the corresponding section in the main content area.

ANNIS Help

Complete List of Operators

AQL currently includes the following operators:

Operator	Description	Illustration	Notes
.	direct precedence	AB	For non-terminating sequences, the maximum number of segments which can be specified with this operator is 3 and 4 tokens respectively.
.*	indirect precedence	A xyz B	For specific sequences, the maximum number of segments which can be specified with this operator is 3 and 4 tokens respectively. As above, the maximum number of segments which can be specified, e.g. 3.
^	directly near	AB or BA	Same as precedence. In corpora with multiple layers, the layer of the query may be specified.
^.*	indirectly near	A xyz B or B xyz A	Like indirect precedence. The form A xyz B is used when the layer of the query is between 3 and 4 tokens respectively. The default maximum number of segments which can be specified with this operator is 3 and 4 tokens respectively.

7

2) Thema wählen

Das Web-Interface: Abfrage

<https://hu.berlin/annis-falko>

Suchfenster
(Sucheingabe hier)

Abfrage starten

Statistik & Export

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

Name	Texts	Tokens		
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	
FalkoSummaryL2v1.2	106	40.638	i	
FalkoWHIGL2v2.1	196	130.949	i	
KanDeL_cross_cohort_v2	425	73.920	i	

Das Web-Interface: Such-Einstellungen



The screenshot shows a search interface with the following elements and annotations:

- Search Bar:** Contains the text **"lernen"**. A red arrow points from the **Suche starten** button to the search bar.
- Buttons:** **Search**, **More** (dropdown), and **History** (dropdown).
- Results Summary:** **147 matches in 64 documents**.
- Search Options Panel:**
 - Left Context:** Set to **5**. A red arrow points from the **linker Kontext** label to this field.
 - Right Context:** Set to **5**. A red arrow points from the **rechter Kontext** label to this field.
 - Show context in:** Set to **tokens (defa ...** (dropdown). A red arrow points from the **Results pro Seite** label to this field.
 - Results Per Page:** Set to **10**. A red arrow points from the **Results pro Seite** label to this field.
 - Order:** Set to **Ascending** (dropdown).
- Query Builder:** A button with a magnifying glass icon and the text **Query Builder**.

Prinzip I: Variablen-Wert-Paare

- tok = "das"
Variable1 ("tok" = Wortform) Wert

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

"tok" = alle Wörter und Satzzeichen im Original-Lernertext

Prinzip I: Variablen-Wert-Paare

- tok = "das"

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

Prinzip I: Variablen-Wert-Paare

- pos = "ART"
 - ↑ Variable2 ("pos" = Wortart)
 - ↑ Wert

tok	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"
Variable3 ("Lemma") Wert

tok	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, ...*

(Vorausgesetzt, die Lemmatisierung des best. Artikels ist entsprechend erfolgt)

beliebig erweiterbar...

• ZH1S = "S"

Variable4

("Satz auf ZH1-Ebene")

Wert

tok	Sofern	das	System	herrscht
pos	KOUS	ART	NN	VVFIN
lemma	sofern	d	System	herrschen
ZH1S	S			

beliebig erweiterbar...

- ZH1S = "S"

...findet alle Sätze wie
Sofern das System herrscht

auf der Suche Ebene "ZH1S"

Suche nach Strings (Zeichenketten)

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

```
tok = "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"

auch möglich: Suche nach Wortkategorie:

pos = "PPOSAT"

Mustersuche (reguläre Ausdrücke)

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

Mustersuche: Joker .

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

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 wie gezeigt kombinieren?

tok = /Frau.?!/

tok = /Frau.* /

tok = /Frau.+ /

Anwendungsbeispiele

lemma = `/.+lang/`

- Findet alle Wörter (Grundformen), die auf *-lang* enden (*bislang, lebenslang, jahrelang, ...*)

tok = `/.*mein.*`

- Findet alle Wörter (Grundformen), die irgendwo *-mein-* enthalten (*mein, vermeintlich, meinetwegen, ...*)

Anwendungsbeispiele

lemma = /ver.*/

- Findet alle Wörter (Grundformen), die auf *ver-* beginnen (*veralbern, verschieben, versus, ...*)

tok = /.*lichem/

- Findet mit hoher Wahrscheinlichkeit nur *-lich*-Derivationen mit starker Adjektivendung (Dativ)

Gruppieren mit () und |

- Frau(en)? findet *Frau* und *Frauen*
- (e|E)r findet *er* und *Er*
→ wichtig bei Satzanfängen
- lemma=/(Frau|Mann|Kind)/
findet diese drei Lemmata
mit einer Suche

Aufgabe

- Finden Sie mit der Wortformsuche Wörter, deren Grundform auf *–lichen*, *-licher* oder *–lichem* endet

→ tok = /.*liche(n|r|m) /

- ... und zusätzlich noch komparativisch sein kann?

→ tok = /.*liche(re)?(n|r|m) /

Suche nach Wortart

- Es gibt unterschiedliche Wortartensysteme (→ Tagsets) für Korpora
- allgemein in der Linguistik unterschiedliche Wortartensysteme
- Die meisten deutschen Korpora benutzen das Tagset STTS
 - ADJA attributives Adjektiv
 - ADV Adverb
 - ART Artikel
 - NN normales Nomen
 - VVFIN finites Verb
 - ...

komplette Liste online verfügbar:

<https://homepage.ruhr-uni-bochum.de/stephen.berman/Korpuslinguistik/Tagsets-STTS.html>

Stuttgart-Tübingen-Tagset (STTS):

hierarchische Ordnung der Wortklassenkürzel

ADJ ektiv	N omen	P ronomen	V erb	P ar Ti kel	KO njunktion
ADJA	NN	PDS	VFIN	PTKZU	KOUI
ADJD	NE	PDAT	VVIMP	PTKNEG	KOUS
		PIS	VVINFINF	PTKVZ	KON
		PIAT	VVIZU	PTKANT	
		PIDAT	VVPP	PTKA	
		PPER	VAFIN		
		PPOSS	VAIMP		
		PPOSAT	VAINFINF		
		PRELS	VAPP		
		PRELAT	VMFIN		
		PRF	VMINFINF		
		PWS	VMPP		
		PWAT			
		PWAV			

Stuttgart-Tübingen-Tagset (STTS): Bereich Verben

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 allen possessiven Pronomen („substituierend“) und Artikeln („attribuierend“)

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


Prinzip II: Relationen

- Einzelne **Variable-Wert**-Paare nacheinander geschrieben
- Zwischen den Paaren muss **IMMER eine Beziehung hergestellt** werden
- Die Beziehung wird mittels entsprechender Operatoren ausgedrückt (nächste Folie)

Variable₁ = Wert₁

"Beziehung"

Variable₂ = Wert₂



Ausdruck 1



Beziehungsoperator



Ausdruck 2

Liste möglicher Beziehungen ...

The screenshot shows the ANNIS interface with the 'Help' tab selected. The main content area displays the 'Complete List of Operators' page. The page includes a table with the following columns: Operator, Description, Illustration, and Notes.

Operator	Description	Illustration	Notes
.	direct precedence	AB	For non-terminating sequences of multiple segments which consist of specified words.
.*	indirect precedence	A xyz B	For specific sequences of words. n, m can be 3 and 4 tokens maximum. As above, specified, e.g.
^	directly near	AB or BA	Same as precedence. In corpora with the layer or may be specified.
^*	indirectly near	A xyz B or ...	Like indirect precedence. The form ^ between 3 and default maximum.

The interface also shows a sidebar with a 'Query Builder' button and a 'Corpus List' table. The 'Corpus List' table has the following data:

Name	Texts	Tokens		
CLEG13	729	285.286	ⓘ	📄
FalkoEssayL1v2.0	94	70.110	ⓘ	📄
falkoEssayL1v2.3	95	70.615	ⓘ	📄
FalkoEssayL2v2.0	248	132.066	ⓘ	📄
FalkoEssayL2v2.3	248	131.628	ⓘ	📄
falkoEssayL2v2.4	248	144.619	ⓘ	📄
FalkoEssayL2WHIGv2.0	195	130.187	ⓘ	📄
FalkoGeorgetownL2v1.0	92	78.151	ⓘ	📄
FalkoSummaryL1v1.2	57	21.211	ⓘ	📄
FalkoSummaryL2v1.2	106	40.638	ⓘ	📄
FalkoWHIGL2v2.1	196	130.949	ⓘ	📄
KanDeL_cross_cohort_v2	425	73.920	ⓘ	📄

Prinzip II: Relationstypen

pos = /PPOS(S|AT)/

← Ausdruck 1

=

← Beziehungsoperator: "Abdeckung"

tok = "meinen"

← Ausdruck 2

				ANNIS Help
1. Introduction				
2. Using the ANNIS interface				
2.1. Search Form				
2.2. Result Window				
2.3. Query Builder				
3. ANNIS Query Language (AQL)				
3.1. Searching for Word Forms				
3.2. Searching for Annotations				
3.3. Searching using Regular Expressions				
3.4. Searching for Trees				
3.5. Searching for Pointing Relations				
3.6. Exporting Results				
3.7. Frequency Analysis				
	>	direct dominance	A B	A specific edge type may be specified, e.g. <code>>secedge</code> to find secondary edges. Edge labels are specified in brackets, e.g. <code>>[func="0A"]</code> for an edge with the function 'object, accusative'
	>*	indirect dominance	A ... B	For specific distance of dominance, <code>>n,m</code> can be used, e.g. <code>>3,4</code> - dominates with 3 to 4 edges 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
	o	overlap	AAA BBB	For overlap only on the left or right side, use <code>o_l</code> and <code>o_r</code> respectively

Aufgabe

- Finden Sie möglichst viele Vorkommen von Präfixverben mit der Lemmasuche und der Wortartensuche verknüpft



```
lemma = /(ver|be|ent|zer|.)*/  
& pos=/V.*/
```

Negation !=

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

```
tok =/mein(e|st|t|en)/    _=_  
lemma!="mein"
```

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

tok = "zu"

pos = "NN"

Ich

bin

zu

Hause

PPER

VAFIN

APPR

NN

Satz

Subj

Adv

Tokenfolgen - Aufgabe

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

pos = /ADJ./ ■

pos = /ADJ./

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	,
tok	weil	sie	entdeckt	eine	Aspekte	der	Gesellschaft		,

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

Aufgabe

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

Lösung:

```
ZH1lemma="ein,,  
=  
ZH1Diff="INS"
```

Hinweis

- Mit sämtlichen Informationen des Korpus kann man sehr akkurat bestimmte Klassen finden, die nicht explizit annotiert sind
 - Beispiel: alle Flexionsfehler bei dem bestimmten oder unbestimmten Artikel finden
 - Strategie: Wir suchen die Wortart ART (laut STTS-Tagset für den bestimmten und unbestimmten Artikel); die Lemmaform der Lernertextebene (Variable: lemma) soll der Lemmaform der ZH1-Ebene (Variable: ZH1lemma) entsprechen; die ZH1-Form soll gegenüber der Lernertextform aber geändert worden sein

Lösung:

```
lemma _ = _ ZH1lemma &  
#1 == #2 &  
#1 _ = _ pos="ART" &  
#1 _ = _ ZH1Diff="CHA"
```

Metadaten (Informationen über den jeweiligen Text) finden

The screenshot displays the ANNIS interface for the 'falkoEssayL2v2.4' corpus. On the left, a 'Corpus List' table shows various documents, with 'falkoEssayL2v2.4' selected. The main area shows 'Corpus information for falkoEssayL2v2.4 (ID: 7121)' with a 'Metadata' table and an 'Available annotations' table. Two red callout boxes with arrows point to the 'falkoEssayL2v2.4' row in the corpus list and the 'Available annotations' table.

Alle Annotationen und Metadaten anzeigen

text- & lernerbezogene Metadaten anzeigen

Name	Value
projectName	Falko Essay L2 v2.4

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

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

Link to corpus: https://korpling.german.hu-berlin.de/falko-suche/#_c=ZmFsa29Fc3NheUwydjluNA

Metadaten finden

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

Meta Data for id 916

SPK0:I2_1_duration	96
SPK0:I2_1_langschool	N/A
SPK0:I2_1_school	N/A
SPK0: Variable	
SPK0:L2index	deu:96:N/A:N/A:N/A:N/A:N/A
SPK0:name	9d3abd29757a2f3f0244c7f26c50b440e7b302a1
SPK0:reg	1:eng,1:N/A,1:N/A,1:N/A,1:N/A,1:N/A,1:N/A,1:N/A
SPK0:sex	f
major-subject	ber_schintensiv
production-modality	essay
subcorpus	fk_2006_07
topic	Feminismus
transcription-d	27.07.06
transcriptionN	fk007_2006_07
transcriptor	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]`).

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

Nach Metadaten filtern

- Nach der Muttersprache von Lernern sucht man mit dem Zusatz (zu einer gegebenen Suchanfrage

& meta::L1index=/.*LÄNDERCODE.*/

- Finden Sie alle

Formen des Adjektivs *deutsch*

in den Texten

englischer Mutter-

sprachler (Code= **eng)**

```
lemma="deutsch"&  
meta::L1index=  
/.*eng.*/
```


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		

(Vollst. Auswahl im Falko-Handbuch)

Metadaten: C-Test-Ergebnisse der Falko-Lerner nutzen

- Hinweise:
 - Der geringste Wert im Korpus ist 60 (von 100; dies gilt für 5 Lernende bzw. Texte), der höchste Werte ist 100 (2 Lernende bzw. Texte)
- Man kann das Spektrum beliebig aufsplitten (stratifizieren):
 - Jeden Wert berücksichtigen
 - Zehnerschritte
 - ...

```
tok&  
meta::ctest=/6[0-9]/
```

...Findet alle Token (Wortformen) bei Lernenden mit einem Score von 60-69

```
pos= "PTKVZ" &  
meta::ctest=/7[0-9]/
```

...Findet alle Verbpartikeln (Wortformen) bei Lernenden mit einem Score von 70-79

In ANNIS Frequenzen einsehen

- Die Gesamttrefferzahl pro Suchanfrage wird unterhalb des Sucheingabefensters angezeigt
- Unter „More“ > „Frequency Analysis“ kann man Frequenzverhältnisse anzeigen lassen:
 - Szenario: Wir wollen wissen, welche Formen die verschiedenen abgetrennten Verpartikeln haben und wie häufig sie sind



pos= "PTKVZ"

- „Frequency Analysis“
- Anstelle von pos wollen wir aber Lemmata ausgegeben bekommen
- Wir schreiben statt „pos“ „lemma“ in die Zelle („Selected annotation of node“)
- „Perform frequency analysis“

Zusammenfassung: Operatoren (Reguläre Ausdrücke)

.	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 $[\wedge abc]$ = alles <i>außer</i> die Menge)
(a b)	a oder b
a{2,3}	a 2 bis 3 mal

Zusammenfassung: Operatoren für Tokenrelationen in ANNIS

- #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!

Einführungsliteratur Korpuslinguistik

- Hirschmann, Hagen (2019): Korpuslinguistik. Eine Einführung. Stuttgart; Metzler.
- Lemnitzer, Lothar; Zinsmeister, Heike (2006): Korpuslinguistik – Eine Einführung
Tübingen; Gunter Narr Verlag.
- Lüdeling, Anke & Kytö, Merja (2009) (Hg.): Corpus Linguistics. An International Handbook. Vol 1 und 2. (Reihe Handbücher zur Sprach- und Kommunikationswissenschaft) Berlin; Mouton de Gruyter.