



# Corpuslinguistics

## Annis<sub>2</sub>-Corpus query tool searching in Falko

Marc Reznicek, Anke Lüdeling, Amir Zeldes, Hagen Hirschmann

[marc.reznicek@staff.hu-berlin.de](mailto:marc.reznicek@staff.hu-berlin.de)

... and many more members of the HU corpus team

# Annis2 SFB 632



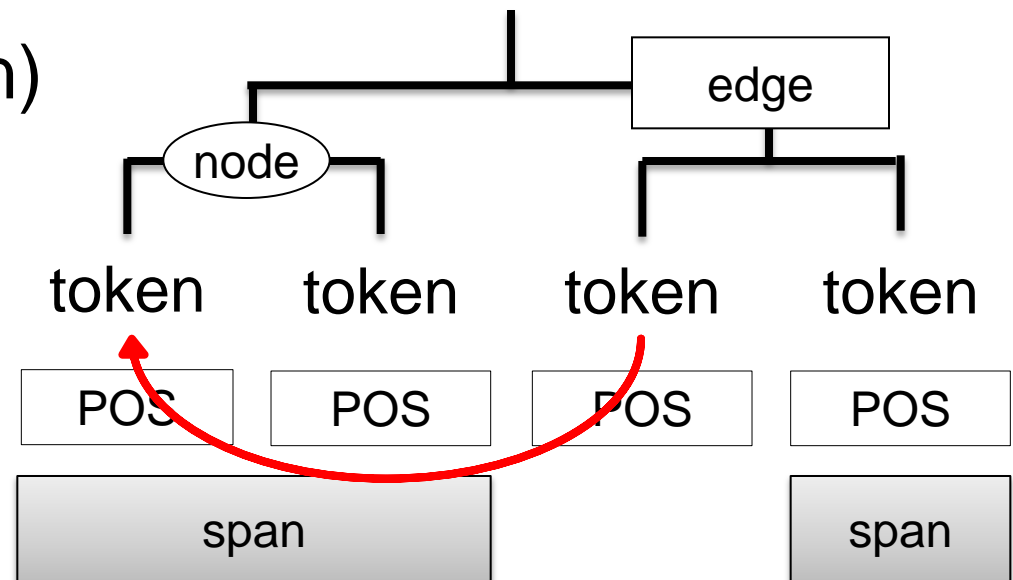
## ANNotation of Information Structure

(Dipper et al. 2004; Chiarcos et al. 2008; Zeldes et al. 2009)

<http://www.sfb632.uni-potsdam.de/d1/annis/>

### ■ search engine for deeply annotated linguistic corpora

- token (-annotation)
- spans
- trees
- pointer



# web interface: query

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

The screenshot shows the ANNIS search interface. At the top, there is a navigation bar with "ANNIS?" and "Tutorial" on the left, and "logged in as 'demo'" on the right. The main content area is titled "Search Form".

Annotations with red arrows point to the following elements:

- query window**: Points to the "AnnisQL:" text input field containing the query "gleich".
- query checker & hit count**: Points to the "Query Builder:" section, which includes a "Show >>" button and a "Result:" field displaying "Valid Query".
- selection of corpora (CTRL + click for multiple items)**: Points to a table of corpora under the "More Corpora" section.
- left and right context**: Points to the "Context Left:" and "Context Right:" input fields, both set to the value "5".
- start query**: Points to the "Show Result" button at the bottom of the form.

The "More Corpora" table contains the following data:

<input type="checkbox"/>	Name	Texts	Tokens	
<input type="checkbox"/>	FalkoEssayL1V2_0	95	70608	
<input checked="" type="checkbox"/>	FalkoEssayL2V2_0	248	131599	
<input type="checkbox"/>	FalkoSummaryL1V1_2	57	21211	
<input type="checkbox"/>	FalkoSummaryL2V1_2	107	40865	
<input type="checkbox"/>	FalkoSummaryVLV1_0	12	11114	

Below the table, there are "Search" and "Export" buttons. At the bottom of the form, there are input fields for "Context Left:" (5), "Context Right:" (5), and "Results per page:" (10).

# web interface:hits

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

Page 1 of 4 | Token Annotations | Show Citation URL | Displaying Results 1 - 10 of 39

Berufsleben sind Frauen und Männer gleich . Ist es aber gut  
 Berufsleben sein Frau und Mann gleich . sein es aber gut  
 NN VAFIN NN KON NN ADJD \$. VAFIN PPER ADV ADJD

falko (grid)

Select Displayed Annotation Levels ▾

lemma	Berufsleben	sein	Frau	und	Mann	gleich	.	sein	es	aber	gut
word	Berufsleben	sind	Frauen	und	Männer	gleich	.	Ist	es	aber	gut
pos	NN	VAFIN	NN	KON	NN	ADJD	\$.	VAFIN	PPER	ADV	ADJD
tok	Berufsleben	sind	Frauen	und	Männer	gleich	.	Ist	es	aber	gut

ZH1 (grid)

Select Displayed Annotation Levels ▾

ZH1lemma	Berufsleben	sein	Frau	und	Mann	gleich	.	sein	es	aber	gut
ZH1pos	NN	VAFIN	NN	KON	NN	PTKVZ	.	VAFIN	PPER	ADV	ADJD
ZH1	Berufsleben	sind	Frauen	und	Männer	gleich	.	Ist	es	aber	gut
tok	Berufsleben	sind	Frauen	und	Männer	gleich	.	Ist	es	aber	gut

ZH2 (grid)  
 ZHverb (grid)  
 text (grid)  
 Volltext

" Der Feminismus hat den Interessen der Frauen mehr geschadet als genützt " Früher hatten die Frauen viel weniger Freiheit als heute . Es gab eine riesige Menge von gesellschaftlichen Verurteilen , wie sich die Frauen benehmen und sogar leben sollten . Nehmen wir zum Beispiel den bekannten Regel von " 3K " . Es ist verständlich , was von Frauen damals erwartet wurde und es ist auch klar , dass solch ein Leben nicht alle befriedigen konnte . Die Frauen hatten den Wunsch , am gesellschaftlichen Leben teilzunehmen , gleich wie Männer zu arbeiten und zu verdienen , aber nicht zu Hause mit den Kinder zu bleiben . Sie haben dafür gekämpft und den Kampf gewonnen . Heute gibt es keine Gesetze mehr , die den Frauen eine bestimmte gesellschaftliche Rolle vorschreiben oder ihre Möglichkeiten und Rechte begrenzen . Frauen und Männer haben gleiche Chancen im Politik , Beruf oder in der Gesellschaft . Natürlich ist es nicht überall so . Bis jetzt existieren nicht nur in den sogenannten Ländern der 3. Welt , sondern auch in den hochentwickelten Staaten verschiedene Unternehmen , in denen die Frauen weniger verdienen nur deswegen , weil sie Frauen sind , obwohl sie die gleiche Arbeit machen . Schon heute , wenn die Frauen das Gewünschte erreicht haben , taucht sofort ein neues Problem auf : der Feminismus scheint mehr ein Übel , als eine langerwartete Freiheit zu sein . Das lässt sich in allen Bereichen des Lebens merken . Im Berufsleben sind Frauen und Männer gleich . Ist es aber gut ? Dieser Regel von " 3K " hat aus dem Frauenleben nicht verschwunden Sie haben viel Hausarbeit zu tun . Ganz oft achtet man darauf erst dann , wenn sie nicht gemacht wird . Wenn es in der Familie kleine Kinder gibt

partiture:  
learner text

partiture TH1

more partitures

# principle 1: variable-value pairs

■ **word** = "System"

**variable1**  
("word form")

**value**

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

# principle 1: variable-value pairs

■ word = "System" the / this / that

...finds *System* (and nothing else)

# principle 1: variable-value pairs

■ **pos** = "NN"  
↑ **variable2** ("word form")      ↑ **value**

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

# principle 1: variable-value pairs

■ pos = "NN"

...finds *Riesen, Frauen, Student, ...*  
*giants, women, students*



# principle 1: variable-value pairs

■ lemma = "System"

variable3  
("lemma")

value

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

# principle 1: variable-value pairs

■ lemma = "System"

...finds *System, Systemen, Systems...*  
*system.nom/dat/gen*

# searching strings

- Find all word forms "*meinen*" in FalkoEssayL2V2.3:

```
word = "meinen"
```

- What do you find?

# lemma

- "base forms" of words
- **Find all forms of the verb "*meinen*":**

lemma = "meinen"

- → problem: lemmatization is not always intuitive.
- example: lemma of *sich*

# patterns (regular expressions)

- Annis allows pattern matching on all annotation layers
- for patterns use `//` instead of `" "`
- **Find all words including *"mein"***

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

# patterns: Joker ■

- an arbitrary character al. → *als*, *alt*, ...
- ■ two arbitrary characters al.. → *alle*, *alte*, *also*
- ■ ■ three arbitrary characters al... → *alles*, *altes*,  
*alias*, ...

# task

- What do you get for?

word = /g.b./

# patterns: ? and \* +

da*s?*

the last character is optional

→  $\phi$ , s → *da, das*

da*s\**

the last character occurs 0 to  $\infty$  times

→  $\phi$ , s, ss, ... → *da, das, dass, dassssssss*

da*s+*

the last character occurs 1 to  $\infty$  times

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



# task

- What happens, if you combine those operators?

word = /Frau.?**?**/

woman

word = /Frau.**\***/

word = /Frau.**+**/

# task

- Try to find all word forms associated with base forms ending in *-lang*

# task

- Try to find all word forms associated with base forms ending in *-lang*

→ lemma = `/.*lang/`

**hits e.g.:**

*bislang*

*lebenslang*

*jahrelang*

# task

- Try to find all word forms associated with base forms starting in *-lang*



lemma = /lang.+/

**hits e.g.:**

*lange*

*langsam*

*langweilig*

# alternatives: a or b = **(a|b)**

- parentheses and | ("or") allow for simultaneous search of different words:
- different forms:

```
word = /(Mann|Frau|Kind)/
```

- strings:

```
word = /(Mann|Mannes)/
```

- substrings

```
word=/bes(ser|t).?/
```

# task

- Find all forms of the verb "*meinen*" in present tense below and only those.

mein	e
mein	st
mein	t
mein	en
mein	t
mein	en

# solution

```
word =/mein(s?t|en?)/
```

or

```
word =/mein(e|st|t|en)/
```

→ Often there are different ways to get the same results

# problem?

w...e(s?t|en?)/  
*- infinitivs*

*- possessive pronouns*

or

word =/mein(e|st|t|en)/



# search for part-of-speech (POS)

- different types of POS-tag sets
- most German corpora use STTS
  - ADJA                   attributives Adjektiv
  - ADV                    Adverb
  - ART                    Artikel
  - NN                    normales Nomen
  - VVFIN                 finites Verb
  - ...

# Stuttgart-Tübingen-Tagset (STTS)

<b>ADJ</b> ektive	<b>N</b> oun	<b>P</b> ronoun	<b>V</b> erb	<b>P</b> arTiKel	<b>KO</b> njunction
<b>ADJA</b>	<b>NN</b>	<b>PDS</b>	<b>VVFIN</b>	<b>PTKZU</b>	<b>KOUI</b>
<b>ADJD</b>	<b>NE</b>	<b>PDAT</b>	<b>VVIMP</b>	<b>PTKNEG</b>	<b>KOUS</b>
		<b>PIS</b>	<b>VVINFINF</b>	<b>PTKVZ</b>	<b>KON</b>
		<b>PIAT</b>	<b>VVIZU</b>	<b>PTKANT</b>	
		<b>PIDAT</b>	<b>VVPP</b>	<b>PTKA</b>	
		<b>PPER</b>	<b>VAFIN</b>		
		<b>PPOSS</b>	<b>VAIMP</b>		
		<b>PPOSAT</b>	<b>VAINFINF</b>		
		<b>PRELS</b>	<b>VAPP</b>		
		<b>PRELAT</b>	<b>VMFIN</b>		
		<b>PRF</b>	<b>VMINFINF</b>		
		<b>PWS</b>	<b>VMPP</b>		
		<b>PWAT</b>			
		<b>PWAV</b>			

# Stuttgart-Tübingen-Tagset (STTS)

<b>VERB</b>	<b>full verb</b>	auxiliar verb	modal verb
finite	<b>V</b> VFIN	VAFIN	VMFIN
imperativ	<b>V</b> VIMP	VAIMP	
infinite	<b>V</b> VINF	VAINF	VMINF
Infinitive with <i>zu</i>	<b>V</b> VIZU		
participle 2	<b>V</b> VPP	VAPP	VMPP

# task

- Find all possessive pronouns

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

# principle II: relations

- single variable-value pairs are connected by **&**
- VV-pairs may not stay unrelated
- You refer to VV-pairs via **#** plus **index**.

variable<sub>1</sub> = value<sub>1</sub> & ← expression1: #1  
variable<sub>2</sub> = value<sub>2</sub> & ← expression 2: #2  
#1 relation #2

# principle II: relations

$\text{variable}_1 = \text{value}_1$  & ← expression 1: #1  
 $\text{variable}_2 = \text{value}_2$  & ← expression 2: #2  
 #1 relation #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 annotation 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

search for order:  
e.g. noun following "zu"

word = "zu"

&

pos = "NN"

&

#1.#2

I am at home

Ich

bin

zu

Hause

PPER

VAFIN

APPR

NN

Satz

Subj

Adv

# task: token order

- **Find two successive adjectives**
- attention: there are two types of adjectives
  - ADJA & ADJD

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



# question

- **How can I make sure that**
  - **both adjectives belong to the same constituent?**
  - **One adjective modifies the other?**

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

# use target hypotheses

- differences between learner text (tok, ctok) and target hypotheses (ZH1,ZH2) are marked in (ZH1Diff, ZH2Diff).

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

<b>ZHDiff</b>	<b>operation on target hypothesis</b>
<b>INS</b>	<b>token inserted</b>
<b>DEL</b>	<b>token deleted</b>
<b>CHA</b>	<b>token changed</b>
<b>MERGE</b>	<b>multiple tokens merged</b>
<b>SPLIT</b>	<b>token split into multiple tokens</b>
<b>MOVS</b>	<b>token moved from here</b>
<b>MOVT</b>	<b>token moved here</b>

# task

- Find all reflexive pronouns missing in the learner text (`ctok`, `ctokpos`, `ctoklemma`)
- target hypotheses layers for TH1 are: `ZH1`, `ZH1pos`, `ZH1lemma` Differences are annotated on `ZH1Diff`, `ZH1posDiff`, `ZH1lemmaDiff`.

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

# task

- ...all **definite** articles missing in the learner text.

## solution:

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

# filter metadata

ANNIS? Tutorial

**Search Form**

AnnisQL: pos="NN"

Query Builder: Show >>

Result: 21842

More Corpora

<input type="checkbox"/>	Name ▲	Texts	Tokens	
<input type="checkbox"/>	FalkoEssayL1V2_0	95	70608	<a href="#">i</a>
<input checked="" type="checkbox"/>	FalkoEssayL2V2_0	248	131599	<a href="#">i</a>
<input type="checkbox"/>	FalkoSummaryL1V1_2	57	21211	<a href="#">i</a>
<input type="checkbox"/>	FalkoSummaryL2V1_2	107	40865	<a href="#">i</a>
<input type="checkbox"/>	FalkoSummaryVLV1_0	12	11114	<a href="#">i</a>
<input type="checkbox"/>	HUKonstruktionsgramma	22	19340	<a href="#">i</a>

**Search Result - pos="NN" (5, 5)**

Page 1 of 2185

Token Annotations Show Citation URL

[i](#) Der **Feminismus** hat den Interessen der Frauen  
Feminismus haben d Interesse d [unknown]  
NN VAFIN ART NN ART NN

- [falko \(grid\)](#)
- [ZH1 \(grid\)](#)
- [ZHverb \(grid\)](#)
- [ZH2 \(grid\)](#)
- [text \(grid\)](#)
- [Volltext](#)

[i](#) Der Feminismus hat  
d Feminismus haben d Interesse d [unknown] mehr schaden als  
ART NN VAFIN ART NN ART NN ADV VVPP KOKOM

- [falko \(grid\)](#)
- [ZH1 \(grid\)](#)
- [ZHverb \(grid\)](#)
- [ZH2 \(grid\)](#)
- [text \(grid\)](#)
- [Volltext](#)

[i](#) Feminismus hat den Interessen der **Frauen** mehr geschadet als genützt Im  
Feminismus haben d Interesse d [unknown] mehr schaden als nützen im  
NN VAFIN ART NN ART NN ADV VVPP KOKOM VVPP APPRART

- [falko \(grid\)](#)
- [ZH1 \(grid\)](#)
- [ZHverb \(grid\)](#)
- [ZH2 \(grid\)](#)
- [text \(grid\)](#)
- [Volltext](#)

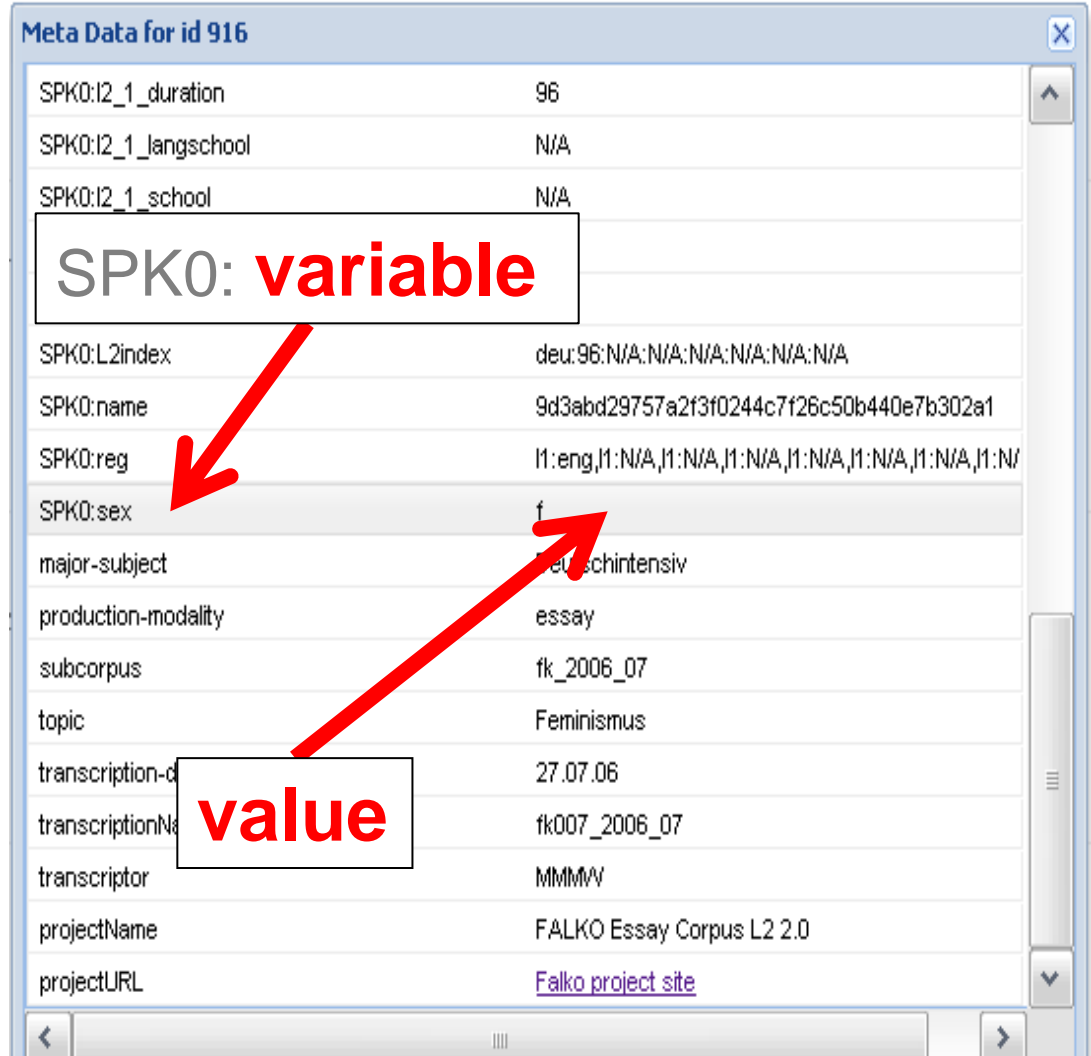
[i](#) mehr geschadet als genützt Im **Bereich** der Literatur ist viel über  
mehr schaden als nützen im Bereich d Literatur sein viel über  
ADV VVPP KOKOM VVPP APPRART NN ART NN VAFIN ADV APPR

- [falko \(grid\)](#)
- [ZH1 \(grid\)](#)
- [ZHverb \(grid\)](#)

text- & learner based metadata

# filter metadata

- metadata:
  - variables and values for
    - text
    - lerner



Meta Data for id 916

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:N/A
SPK0:name	9d3abd29757a213f0244c7126c50b440e7b302a1
SPK0:reg	l1:eng,l1:N/A,l1:N/A,l1:N/A,l1:N/A,l1:N/A,l1:N/A,l1:N/A
SPK0:sex	f
major-subject	deutschintensiv
production-modality	essay
subcorpus	fk_2006_07
topic	Feminismus
transcription-date	27.07.06
transcriptionName	fk007_2006_07
transcripator	MMM/V
projectName	FALKO Essay Corpus L2 2.0
projectURL	<a href="#">Falko project site</a>

# filter metadata

- metadata queries are formulated as `meta::variable = "value"`

Find all word forms of lemma "Mann"  
written by females: **sex** .

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



# filter metadata

- filter via L1
- `meta::reg=/l1:country code/`
- **Find all form of the adjective "deutsch" in english texts: `code= eng`**

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

# filter metadata

- For a more specific selection of language acquisition orders, you need a `.*` between each VV-pair.
- `meta::reg=/variable1:value1.*variable2:value2/`
- **Find all word forms of "deutsch" in texts written by danish learners with L2 English before German.**

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

# language codes in Falko (selection)

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

# example

- compare missing articles for spanish (spa) and danish (dan) speakers
  - Important: how many articles are there in total?
- How many **possible** articles are missing?

total amount of articles

```
ZH1pos= "ART"&  
meta::reg=/l1:dan/
```

```
ZH1pos= "ART"&  
meta::reg=/l1:ita/
```

# counting tokens

- Suppressing the value in a query includes all items annotated for a variable

word



amount of tokens

lemma



amount of lemmas

- This way you can find the amount of tokens for a special L1 group

word &

**meta::reg=/l1:ita/**

# counting tokens

- How many tokens of japanese speakers are included in Falko?

word &  
meta::reg=/l1:jpn/

# counting texts

- Each text starts with `TXTstructure = "start "` und ends mit `TXTstructure = "end"`.
- **How many texts of italian learners are included in Falko?**

```
TXTstructure="start" &  
meta::reg=/l1:ita/
```

# syntax

## (authentic) study

- question: Is acquisition of relatives independent of the grammatical function of the relative pronoun?
- hypotheses:
  - 1. all syntactical functions for relatives are acquired simultaneously
  - 2. Initially only prototypical functions for relatives are acquired.



# most simple structure (subject)

- relative pronoun in subject position:

relative pronoun

```
POS=/N.*/ & POS=/PREL.*/ &  
#1 ->dep[deprel="SUBJ"] #2
```

subject

- accusative object (=OBJA) or dative object (=OBJD)

```
ZH1pos = /N.*/ & ZH1pos=/PREL.*/ &  
ZH1 & ZH1 & #1_=_#3 & #2_=_#4 &  
#3 ->dep[deprel="OBJA"] #4
```

- dependencies relate only TH-layers directly

# character sets [ ]

- With [ ] you can describe sets of allowed characters:
  - [aeiou] – simple vocals
  - [A-Z] – all capital letters between A and Z
  - [0-9]+ - all hnumbers
- Beispiele:
  - **What kind of words do you find?**

**lemma = /[A-Z][a-z]+-[A-Z][a-z]+/**

# character sets [ ]: Umlaute

- [A-Za-zÄÖÜß] – Umlaute special characters have to be mentioned expressively
- [A-Za-zÄÖÜäöüß0-9-] – all characters that can occur in a word

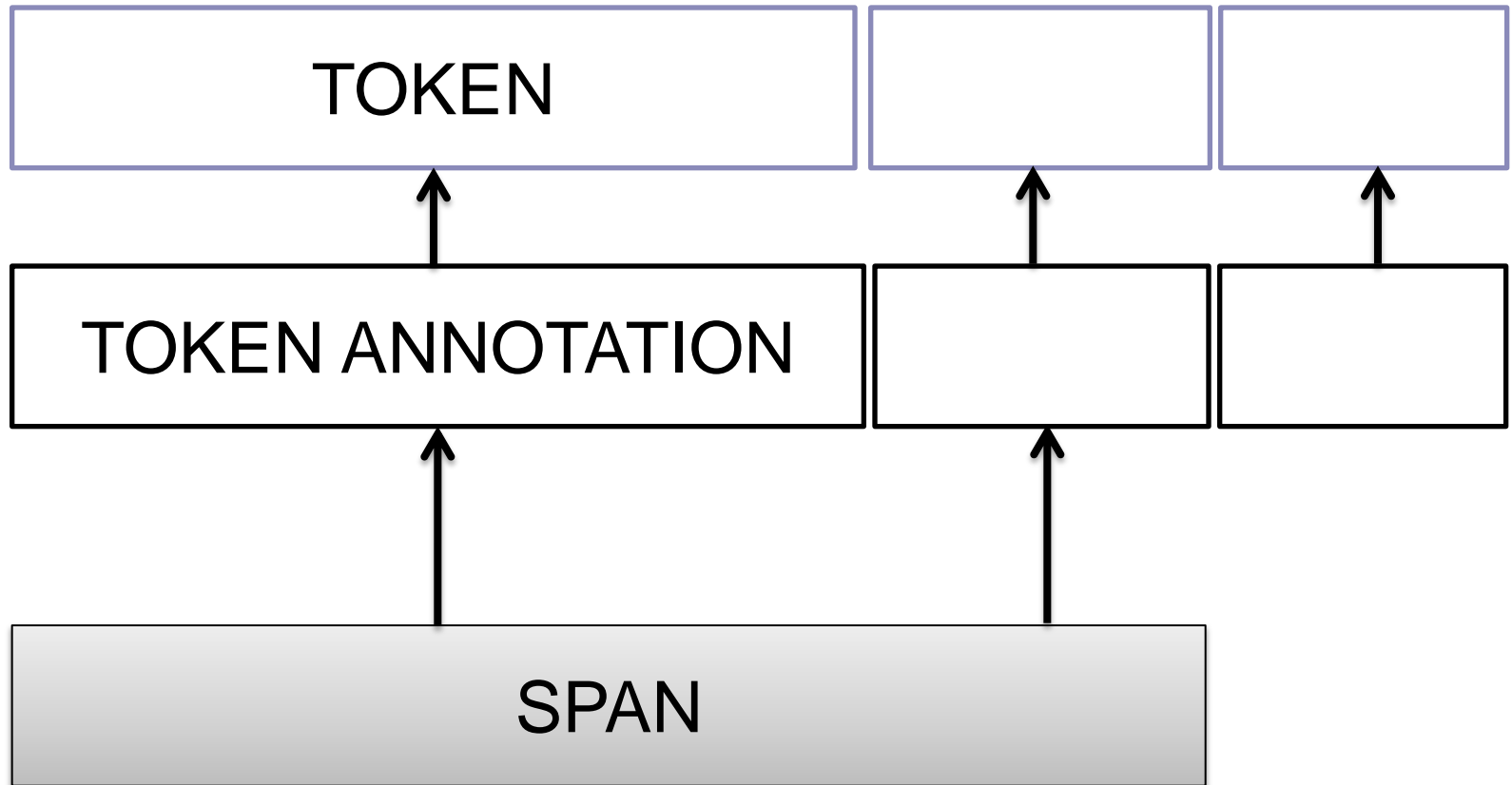
lemma = /[A-ZÄÖÜ][a-zäöüß]+-[A-ZÄÖÜ][a-zäöüß]+/

# excluded character sets [ ^ ]

- [ ^ ] defines a set of characters which aren't allow to occur at the given position
- [ ^aeiouäöüAEIOUÄÖÜ ] – no vocals
- [ ^äöüÄÖÜ ] – no Umlaute
- example:
  - Find all word without "ß"!

```
word=/[^ß]+/
```

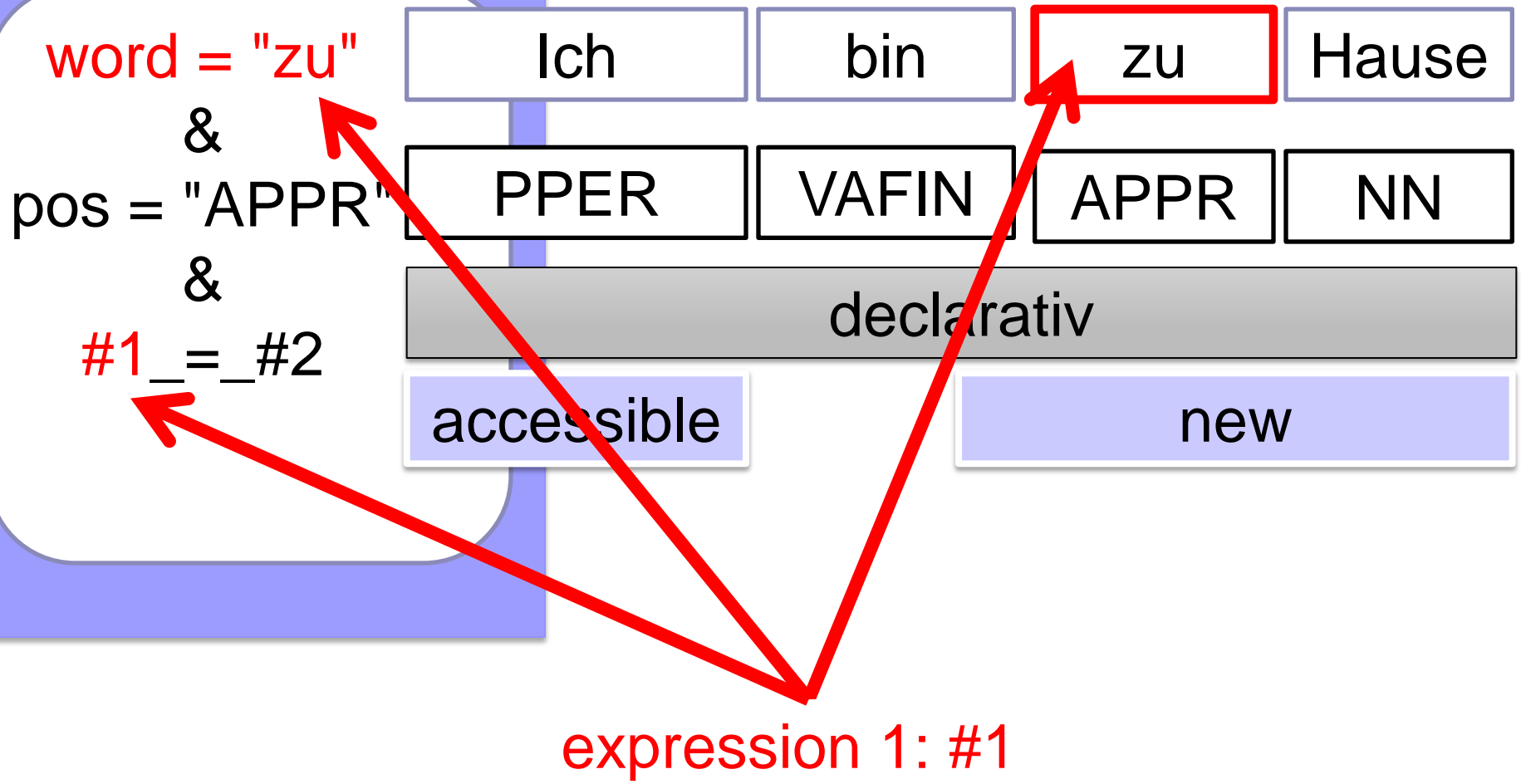
# Search on multiple layers



# Search on multiple layers

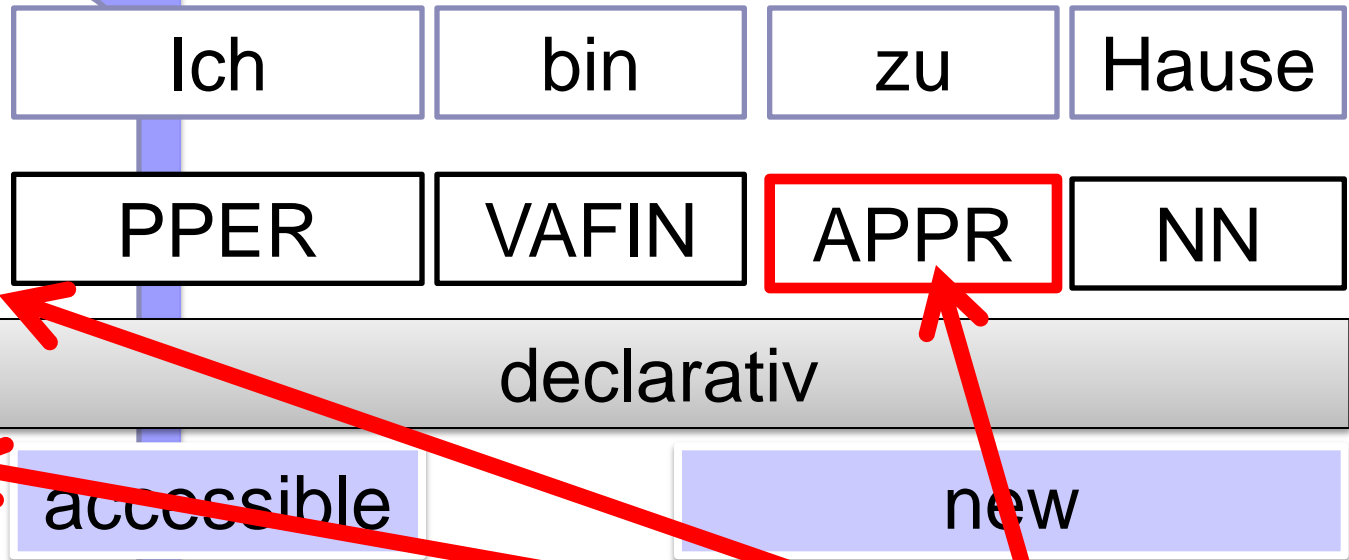
<b>word</b>	Ich	bin	zu	Hause
<b>pos</b>	PPER	VAFIN	APPR	NN
<b>mode</b>	declarativ			
<b>infostat</b>	accessible		new	

# Search on multiple layers



# Search on multiple layers

word = "zu"  
&  
pos = "APPR"  
&  
#1\_=#2

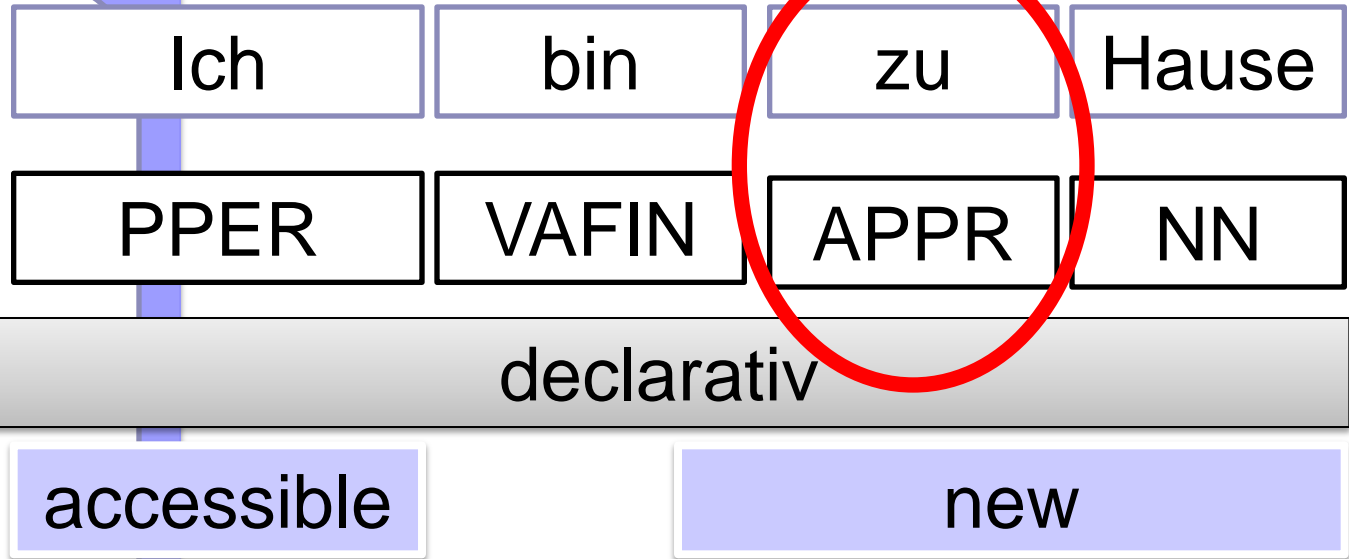


expression 2:  
#2



# Search on multiple layers

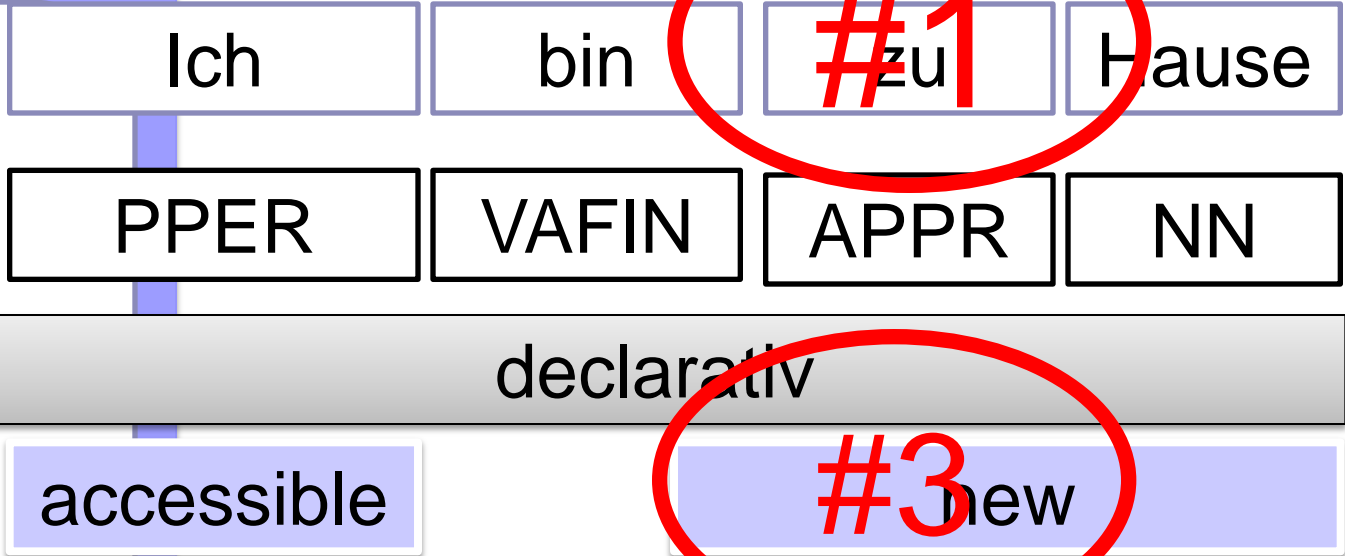
word = "zu"  
&  
pos = "APPR"  
&  
#1 == #2



Search for a token with the value "zu" for the variable "word", with the value "APPR" for the variable "pos". Both annotations shall refer to the same tokens.

# Search on multiple layers

```
word = "zu"  
&  
pos = "APPR"  
&  
istat = "new"  
&  
#1_ = _#2  
&  
#3_ i _#1
```



The token shall also be included in the span "new" on the infostatus layer.

# summary - operators

.	arbitrary character
*	an arbitrary amount of the last character before
+	at least one instance of the character before
?	last character before is optional
\	interpret next character as literary
!	not
[abc]	one element of the set
[^abc]	element except the ones in the set
(a b)	a or b
a{2,3}	2 to 3 times "a"

# summary relations

- operators on token relations

#1.#2

#1 directly followed by #2.

#1.\*#2

#1 indirectly followed by #2.

#1\_=#2

#1 and #2 refer to the same token/span

#1\_i\_#2

#1 is included in #2.

# summary ANNIS

- ANNIS allows:
  - (simultaneous) query of different corpora
  - quantify results
  - export results
  - filter metadata

# attention!

- corpora are always just samples of a language variety
- different corpora (may) give different results
- (sometimes) corpora include errors
- corpora can still help support or reject hypotheses

Thanks!  
Danke!

# 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*.