

# Concepts for Annotating Negation in Corpora

October 15th, 2024

Hagen Hirschmann

Humboldt-Universität zu Berlin

[hagen.hirschmann@hu-berlin.de](mailto:hagen.hirschmann@hu-berlin.de)

# Warning

- Please keep expectations low 😊
- No presentation about own research
- Expect a report on conceptual shortcomings in corpus linguistics (when dealing with negation and similarly complex linguistic phenomena)

# Plan

- Very brief introduction to basic concepts in corpus linguistics
- Annotating aspects of negation
  - Annotation cues
    - Categorizing them
    - Looking inside: word-internal negation triggers
    - Normalization and variation (example from historical corpus)
  - Describing syntactic and semantic aspects of negation
    - Introduction: syntax in corpora
    - Describing the negated expression: scope, focus association, syntactic relation as independent features
      - Examples, examples, examples
      - What do we do with the data that is already available
      - Conclusions
      - Suggestion for a more refined description

# "Corpus"

- Definition of "Corpus": Natural language in its full context (text data)
- Collections of examples could also be considered corpora; perhaps we need different definitions or types of corpora
  - For analyzing the versatility of negation, a collection of sentences containing negation, extracted from bigger data, may be ideal
- Annotation as a criterion for 'Corpus'?
  - Not strictly necessary, but widely used and arguably essential for usability
  - Trend: Corpora increasingly connecting multiple levels of linguistic description
- Aim for NegLab: construct a great, sophisticated, illustrative data base to investigate qualitative-quantitative aspects of negation!

# "Annotation"

- Any kind of analysis
- Transcription, normalization, and metadata must also be considered annotations
- Annotation of negation means making as many aspects of negation explicit as possible:
  - Lexical features
  - Morphological aspects
  - Semantics
  - Pragmatics
  - Syntax
  - Information structural references
    - This can only be achieved through many interconnected analyses and various data formats.

# Annotating Specific Aspects of Negation

## - Negation Cues

- Negation as a lexical feature – marking negation cues (operators)
  - Which layer of description?
    - Explicit marking of NEG-function is usable
      - *Ist das **nicht** toll? Isn't it great?*
    - Negation as a lexical category? "NEG" as a POS feature?
      - Hopefully not on a basic level (modern corpora can have different POS tagsets with different granularity)
      - STTS has PTKNEG as a POS category, but only for the lemma *nicht* (this is not beneficial, because as an effect, PTKNEG is just mechanically assigned to every instance of *nicht*)
      - Most other tagsets for word categories (PTB, Universal POS <https://universaldependencies.org/u/pos/>) do not cover the feature negation
      - Most tag sets for syntactic functions bypass negation and subsume it under modification

# Annotation example

- *Peter wird diese Probleme keinesfalls lösen.*  
*Peter will by no means solve these problems.*

	0	1	2	3	4	5	6
[word]	<b>Peter</b>	<b>wird</b>	<b>diese</b>	<b>Probleme</b>	<b>keinesfalls</b>	<b>lösen</b>	.
[S]	s1						
[pos]	NE	VAFIN	PDAT	NN	ADV	VWINF	\$.
[lemma]	Peter	werden	dies	Problem	keinesfalls	lösen	.
[Negation_cue]					keinesfalls		

# Annotating Specific Aspects of Negation

## - Negation Cues

- Negation as a multiword feature – marking negation cues (operators)
  - Another argument to mark negation operators on a separate level
- Different cases of more than one word expressing negation:
  - Multi-word lexemes (*nie und nimmer* – see below)
  - Phrasal operators (<sub>PP</sub>[*auf keinen Fall*] – <sub>PP</sub>[*by no means*])
  - Double (marked) negation (*ne ... pas*; *no quiero nada*)



# Annotating Specific Aspects of Negation

## - Negation Cues

- Negation as a multiword feature – marking negation cues (operators)
  - Another argument to mark negation operators on a separate level

	0	1	2	3	4	5	6
[word]	<b>Peter</b>	<b>ne</b>	<b>résoudra</b>	<b>pas</b>	<b>ces</b>	<b>problèmes</b>	<b>.</b>
[S]	s1						
[pos]	NAM	ADV	VER:futu	ADV	PRO:DEM	NOM	SENT
[lemma]	Peter	ne	résoudre	pas	ce	problème	.
[Negation_cue]		ne pas					

# Annotating Specific Aspects of Negation

## - Negation Cues

- Negation as a multiword feature – marking negation cues (operators)

	0	1	2	3	4	5	6	7	8
[word]	<b>Auf</b>	<b>keinen</b>	<b>Fall</b>	<b>möchte</b>	<b>Peter</b>	<b>gar</b>	<b>nicht</b>	<b>mitkommen</b>	<b>.</b>
[S]	s1								
[pos]	APPR	PIAT	NN	VMFIN	NE	PTK	ADV	VWINF	\$.
[lemma]	auf	kein	Fall	mögen	Peter	gar	nicht	mitkommen	.
[Phrase_cat]	PP					ADVP			
[Negation_cue]	auf keinen Fall						nicht		

# Annotating Specific Aspects of Negation

## - Negation Cues

- Negativity as a lexical semantic feature – negative sentiment
  - Automatic sentiment analysis has significant economic impact
    - Many well-funded projects that spent a lot of energy in making sentiment of words explicit
    - Outcome: lists with words, graded for polarity
    - Example: PolArt ("A manually specified, word-level polarity lexicon for German", comprising ~9.500 word lemmas, <https://aclanthology.org/W09-4635/>, Klenner et al. 2009)

# Annotating Specific Aspects of Negation

## - Negation Cues

	A	B	C	D	E
1	Lemma	Funktion	Grad	Wortart	Funktion_Wortart
2	nie	SHI	0	neg	SHI_neg
3	nix	SHI	0	neg	SHI_neg
4	keinsterweise	SHI	0	neg	SHI_neg
5	keinerweise	SHI	0	neg	SHI_neg
6	nichts	SHI	0	neg	SHI_neg
7	ohne	SHI	0	neg	SHI_neg
8	nicht	SHI	0	neg	SHI_neg
9	kein	SHI	0	neg	SHI_neg
10	niemals	SHI	0	adj	SHI_adj
11	kein	SHI	0	adj	SHI_adj
12	keinesfalls	SHI	0	adj	SHI_adj
13	weniger	SHI	0	adj	SHI_adj
14	kaum	SHI	0	adj	SHI_adj
15	ebensowenig	SHI	0	adj	SHI_adj
16	einzigartig	SHI	0	adj	SHI_adj
17	stinklangweilig	NEG	1	adj	NEG_adj
18	beeinträchtigt	NEG	1	adj	NEG_adj
19	Ekel_erregend	NEG	1	adj	NEG_adj
20	verlorengegangen	NEG	1	adj	NEG_adj
21	manisch	NEG	1	adi	NEG_adi

# Annotating Specific Aspects of Negation

## - Negation Cues

- Negativity as a lexical semantic feature – negative sentiment
  - With such lists, NLP applications are fed in order to return accurate gradings of blogs, comments, etc. in the web
  - Linguistic knowledge gain?
    - Maybe not directly, but data can be very valuable for further analyses
      - Example: Morphological analysis of negative adjectives – which morphemes are most frequent?
      - Among the most frequent morphemes not necessarily negative ones
      - Pejorative meaning results from negative or shifting affixes in combination with neutral or positive lexical stems (*kind-isch*, *erwerb-s-los*, *un-fein*) or the other way around (*stümper-haft*, *schmutz-ig*)
      - More fine-grained analysis would be interesting

Morphological segmentation of ~1000 negative adjectives processed with SMOR  
(<https://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/smor/>)

# Annotating Specific Aspects of Negation

## - Negation Cues

Rank	Freq	Word
1	129	ig
2	108	isch
3	98	un
4	74	lich
5	74	ver
6	64	er
7	60	los
8	32	be
9	28	haft
10	24	ist
11	18	bar
12	15	feind
13	15	krank
14	14	miss

(→ Militär-ist-isch)

- Most frequent morphemes in PoArt list:
- 990 most negative (-1) adjective forms were processed in the following way:
  - morphological component analysis by SMOR (see reference at bottom)
  - keep only one suggestion with most morphemes (errors in segmentation possible)
- count morphemes with AntConc (<https://www.laurenceanthony.net/software/antconc/>)
- Full list: <https://box.hu-berlin.de/f/cfd69801def34e6aabd4/?dl=1>

Morphological segmentation of ~1000 negative adjectives processed with SMOR (<https://www.ims.uni-stuttgart.de/forschung/ressourcen/werkzeuge/smor/>)

# Annotating Specific Aspects of Negation

## - Negation Cues

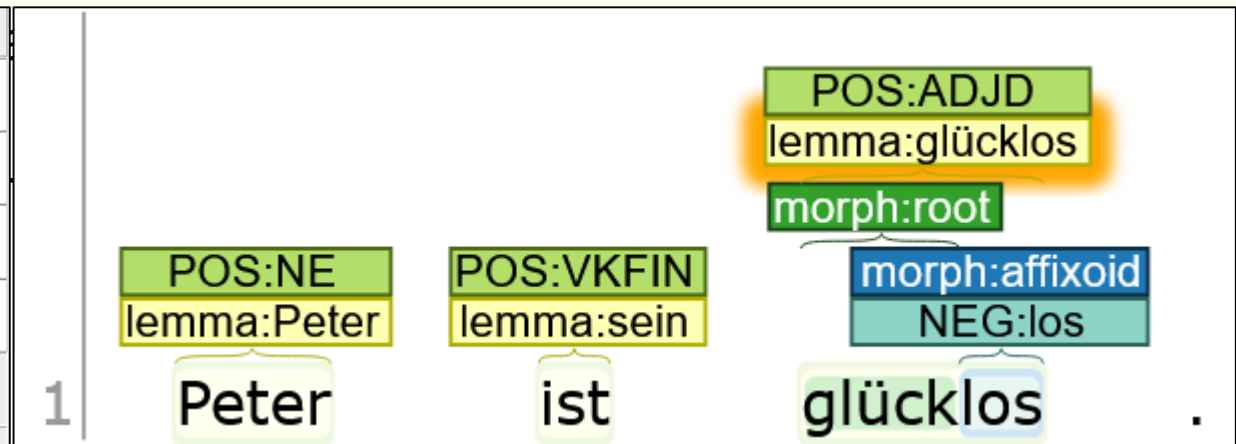
- Annotating morphological elements (e.g. morphemes expressing negation) in the corpus
  - Aim: Mark operators of negation in transparent, exocentric expressions (unehrlich – dishonest)
    - Example: annotating *Peter ist glücklos* as a variant of *Peter hat kein Glück*
    - To assign the source of the negation to *-los* we have to do the analysis within *glücklos*

# Annotating Specific Aspects of Negation

## - Negation Cues

- Annotating morphological elements (e.g. morphemes expressing negation) in the corpus
  - Aim: Mark operators of negation in transparent, exocentric expressions (unehrlich – dishonest)

	0	1	2	3	4
[word]	Peter	ist	glücklos	.	
[S]	s1				
[pos]	NE	VAFIN	ADJD		\$.
[lemma]	Peter	sein	glücklos	.	
[Morph_units]			Glück	-los	
[Morph_cat]			root	affixoid	
[Negation_cue]				los	



<https://inception-project.github.io/>

<https://exmaralda.org>



# Non-standard data

Gesundheitseinschränkungen auf keinster vweise eingegangen, mußte deshalb die Reha abbrechen!!!! Werde diese Klinik aufgrund der Ärzte und Psychotherapeuten nicht, nie wieder aufsuchen und kann sie **keinstenfalls** weiterempfehlen, es tut mir leid um die anderen Mitarbeiter, sie machen einen super Job!!Rezepte,Verordnungen und Überweisungen können und werden dort nicht ausgestellt. es muß alles von zu Hause

[https://www.klinikbewertungen.de/klinik-forum/erfahrung-mit-wicker-klinik-bad-wildungen?fac\\_id=psysom&allbew](https://www.klinikbewertungen.de/klinik-forum/erfahrung-mit-wicker-klinik-bad-wildungen?fac_id=psysom&allbew)

- This effect becomes more critical the less standardized the data we are examining
- Variants like *not-n't*, *keinstenfalls-keinesfalls*, *nichts-nix* etc. are form variants (also stylistic and register variants) of the same negation cue (rather than different operators)
  - We must be able to make this explicit
- Data normalization useful (or necessary) in almost any cases (very crucial: historical data, learner data, colloquial registers, ...)
  - Keep the original variant, assign a more abstract, normalized form on a different level of description

# Non-standard Data and Normalization

- Very positive example: reference corpus middle high German (Referenzkorpus Mittelhochdeutsch, <https://www.linguistics.rub.de/rem/>, Petran et al. 2016)
- Here, negation cues are very systematically marked and labelled
  - Positive effect: We can explore the variation instantly

# Non-standard Data and Normalization

The screenshot shows the ANNIS interface with the following components:

- Search Bar:** Query: `pos=/. *NEG.* / _=_ tok_dipl`
- Search Results:** 40839 matches in 190 documents.
- Document List:**

Name	Texts	Tokens
14_2-wobd_2	4	75.587
15_1-boehm	1	19.076
15_1-ofr	2	22.009
15_1-omd	5	88.065
15_1-oozd	5	63.199
- Result Detail (Path: 14\_2-boehm > F329 (tok\_dipl 1970 - 1980)):**

ere vnd lon der marter **nicht** benomen wann czweyerley marter ift

Annotations: `tok_dipl`
- Table of Linguistic Data:**

reference	003va,37					003va,38				
<b>tok_dipl</b>	hat	er	fich	mit	worten	<b>ny</b>	vorgeffen	vnd	nicht	auf
<b>tok_anno</b>	hat	er	sich	mit	worten	ny	vorgessen	vnd	nicht	auf
<b>lemma</b>	haben	er	sich	mit	wort	nie	vergessen	und	nicht	auf
<b>lemmald</b>	<a href="#">GD00610</a>	<a href="#">GE05989</a>	<a href="#">GS27271</a>	<a href="#">GM05669</a>	<a href="#">GW26936</a>	<a href="#">GN05069</a>	<a href="#">GV01423</a>	<a href="#">GU05758</a>	<a href="#">GN04912</a>	<a href="#">GA06002</a>
<b>pos</b>	VAFIN	PPER	PRF	APPR	NA	<b>AVNEG</b>	VVIN	KON	PTKNEG	APPR
<b>posLemma</b>	VA	PPER	PRF	AP	NA	AVD	VV	KO	PI	AP
<b>inflection</b>	3.Sg.Präs.Ind.Unr	*.Sg.Mask.Nom	3.Sg.Akk		*.Dat.Pl		*			

- Data annotation primarily processed using EXMARaLDA (<https://exmaralda.org/>, Schmidt 2012)
- Data publication in ANNIS (<https://corpus-tools.org/annis/>)

# Non-standard Data and Normalization

Help us make ANNIS better! not logged in [Logi](#)

pos=/. \*NEG.\* / \_ \_ tok\_dipl

Base text

2 / 4084 / 4084 Displaying Results 11 - 20 of 40839 Result for: pos=/. \*NEG.\* / \_ \_ tok\_dipl

selected corpora:  
 14\_2-wobd\_2, 17\_1-oobd, 16\_2-wmd, 14\_2-wmd, 15\_1-wobd, 16\_1-wobd, 15\_1-oobd, 16\_1-ofr, 16\_1-oobd, 16\_2-omd, 14\_2-ofr, 16\_2-wobd\_2, 17\_1-wobd, 16\_1-wmd, 14\_2-boehm, 16\_2-ofr, 16\_1-omd, 14\_2-omd, 16\_2-wobd, 17\_1-omd, 15\_2-omd, 15\_1-omd, 15\_1-wobd\_2, 15\_1-ofr, 17\_1-wmd, 15\_2-wobd\_2, 15\_2-wmd, 15\_2-oobd, 15\_2-ofr, 17\_1-ofr, 14\_2-wobd, 15\_1-boehm, 15\_2-wobd, 16\_2-oobd, 15\_1-wmd, 14\_2-oobd

query to analyze:  
 pos="PTKNEG" \_ \_ tok\_dipl \_ \_ lemma="nicht"

	Node number/name	Selected annotation of node	Comment
1	1	pos	automatically created from pos="PTKNEG"
2	2	tok_dipl	automatically created from tok_dipl
3	3	lemma	automatically created from lemma="nicht"

26140 matches in 189 documents

Export  
 Frequency Analysis

15_1-ofr	2	22.009									
15_1-omd	5	88.065									
15_1-oobd	5	63.199									
<b>pos</b>	VAFIN	PPER	PRF	APPR	NA	AVNEG	VVINP	KON	PTKNEG	APPR	
<b>posLemma</b>	VA	PPER	PRF	AP	NA	AVD	VV	KO	PI	AP	
<b>inflection</b>	3.Sg.Präs.Ind.Unr	*.Sg.Mask.Nom	3.Sg.Akk		*.Dat.Pl		*				

# Non-standard Data and Normalization

pos=/.\*NEG.\*/\_=\_ tok\_dipl

Help/Examples Query Result x Frequency Analysis x

Base text v

Query / 4084 S SI Displaying Results 11 - 20 of 40830 Result for pos=/.\*NEG.\*/\_=\_ tok\_dipl

- ~100 variants of *nicht*

nit (11402), nicht (8790), nít (1926), nitt (739), niht (670), neit (374), nícht (372), nu̇t (239), niet (225), nû̇t (166), Nicht (158), nyet (145), Nit (135), nv̇t (101), nyt (80), níht (72), nich (71), nichte (61), neít (52), Nyet (23), nîht (19), nichten (19), nichtes (19), nýet (18), Nitt (16), nichṫ (13), nichts (12), int (12), nytt (12), nüt (10), nu̇tt (10), Niht (8), nichtt (7), nut (7), níchte (6), neüt (6), nýt (6), [...] (6), nīcht (5), Nít (5), níd (5), nýet (5), nút (4), Nich (4), Niet (4), nv̇t (4), nu̇t (4), nýt (4), nichtew (4), nytt (3), níet (3), nítt (3), nûtt (3), nīcht (3), nichtē (3), nich[...] (3), n[...] (2), nīcht (2), Int (2), nid (2), nī (2), nît (2), Nu̇t (2), nit( (2), Nícht (2), nycht (2), nu̇ti (2), [...]icht (2), ni[...] (2), nichtev (2), nyetdae (1), [...]t (1), [...]cht (1), nichṫ (2) (1), nital= (1), nīcht (1), níchts (1), n (1), nir (1), nu̇ (1), níg (1), nī (1), ínt (1), Neit (1), Nu̇t (1), nith (1), nivt (1), ni̇ (1), nyth (1), nu̇st (1), nu̇t (1), nýt (1), nýt (1), ních (1), nichtenn (1), Nycht (1), nichh (1), niten (1), noitt (1), nýth (1), nýtt (1), níchten (1)

posLemma	VA	PPER	PRF	AP	NA	AVD	VV	KO	PI	AP
inflection	3.Sg.Präs.Ind.Unr	*.Sg.Mask.Nom	3.Sg.Akk		*.Dat.Pl		*			

# Annotating Specific Aspects of Negation

## - Syntax Parses in Corpora

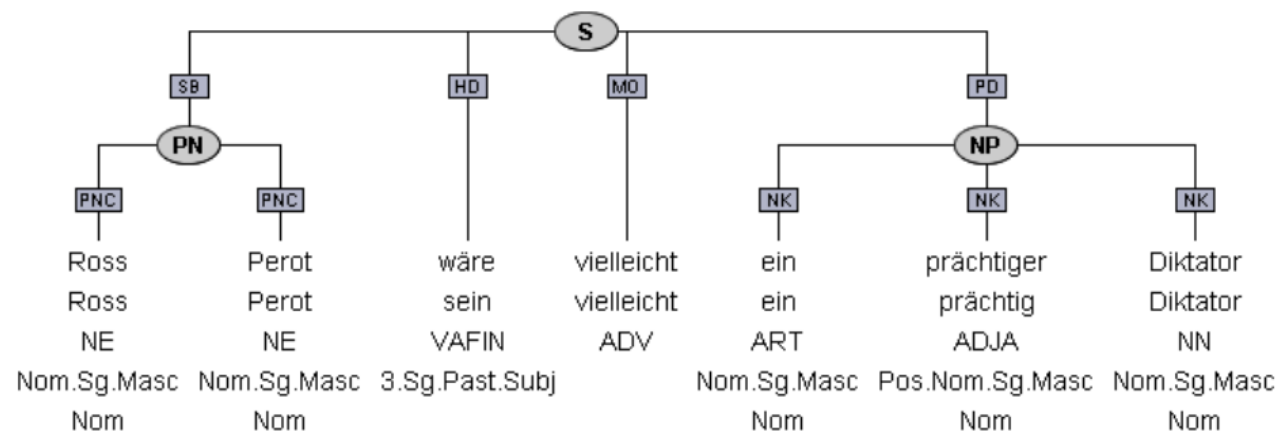
- Aim: Analyze negation in its syntactic environment
- Warning: Syntax descriptions in corpora are never fully theoretized syntax models, because
  - Many rules and restrictions lead to rather "bulky" structures
  - "Uncontrolled" natural language often shows variants that are not covered by restricted models
- As a result, common syntax "models" for corpora do without
  - movement (no traces, deep structure, only surface positions)
  - strictly binary structures (flat, shortened phrases)
  - null elements
  - functional phrases
  - ...
- Crossing branches are often allowed (in dependency annotations always)
- Extensive edge labelling depicts functional relations (traditional terminology)
- As an effect, corpus-based syntax modelling only seems to "somehow reconstruct the meaning of the sentences"



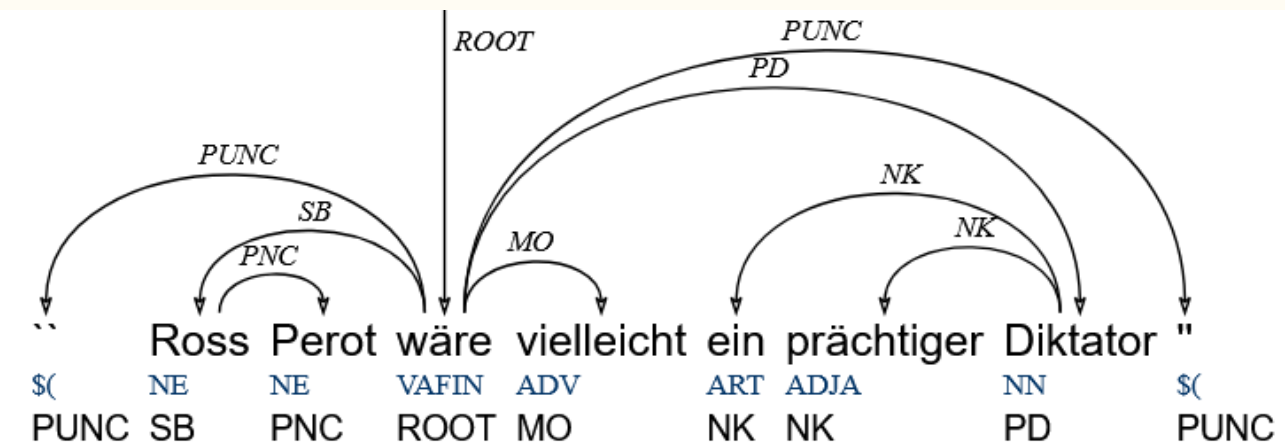
# Annotating Specific Aspects of Negation

## - Syntax Parses in Corpora

- To different approaches to capture basic features of syntactical surface
  - Flat constituency parses



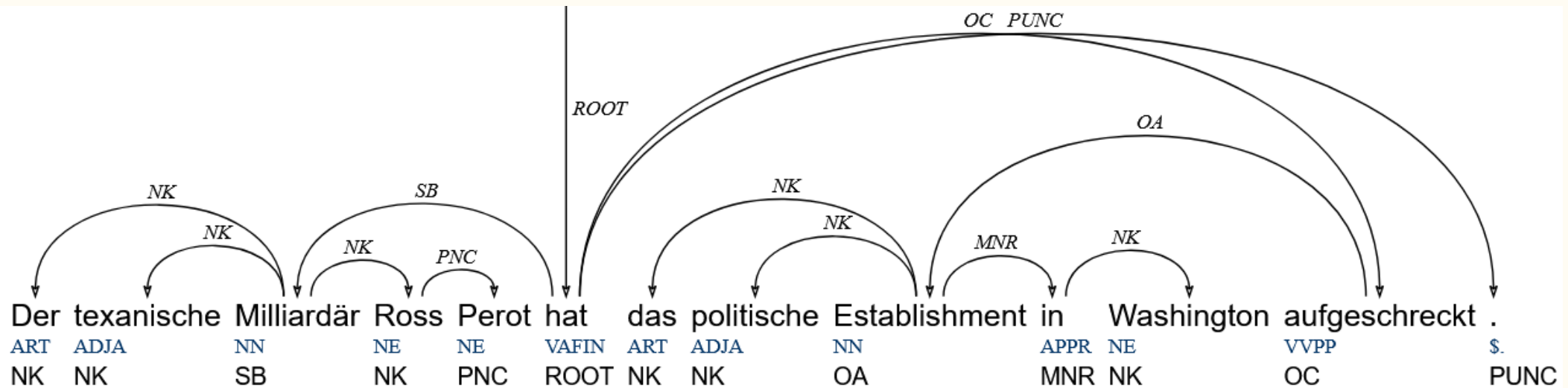
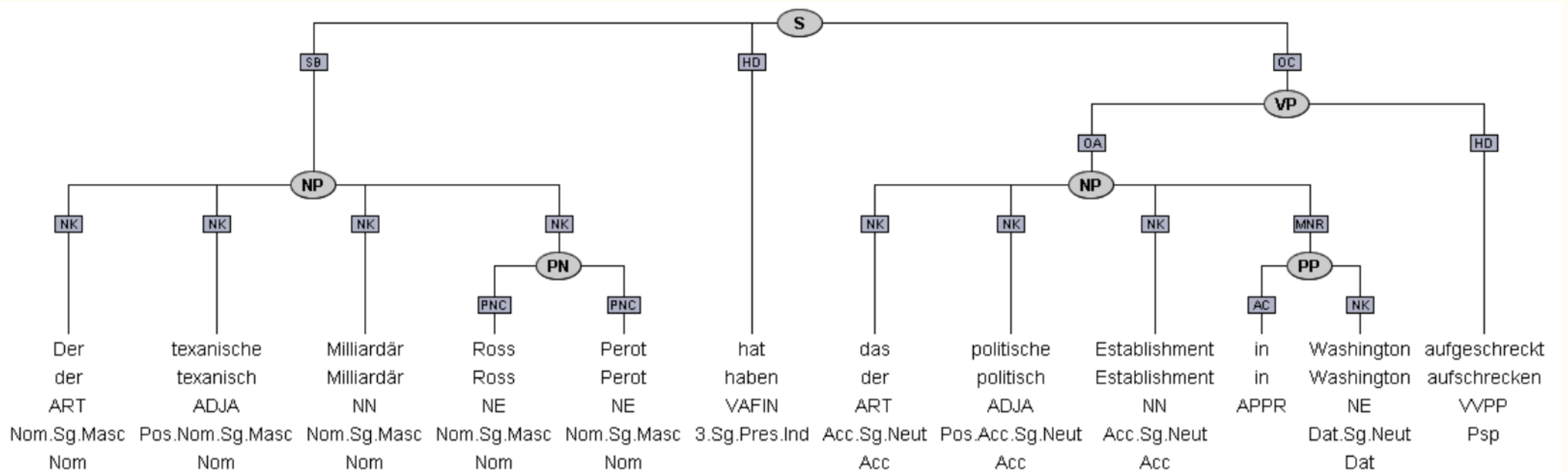
- Dependency parses



Tiger corpus data (constituency and dependency structure) and analysis tool TigerSearch available at [www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/tiger/](http://www.ims.uni-stuttgart.de/forschung/ressourcen/korpora/tiger/)

# Annotating Specific Aspects of Negation

## - Syntax





# Annotating Specific Aspects of Negation

## - "Negated Expression"

- When analyzing the negated part, we have to clarify what grammatical system we want to describe
- Three notions in corpus annotations
  - (1) Syntactic mother constituent (for constituency annotations; ~syntactic range ["Bereich"], Blühdorn 2012, p. 76, according to Jacobs 1982, 1991; also "Wirtskonstituente"/"host constituent") or reference word (for dependency annotations)
  - (2) Negation scope
  - (3) Negation-associated focus
- In practice of syntactic annotation in corpora, it often remains unclear if (1), (2), or (3) is annotated

# Annotating Specific Aspects of Negation

## - "Negated Expression"

### Negation scope

- "negation is generally (...) a unique propositional operator" (NegLab introductory website [https://www.uni-frankfurt.de/149292001/Negation\\_in\\_Language\\_and\\_Beyond](https://www.uni-frankfurt.de/149292001/Negation_in_Language_and_Beyond)) (for the semantic structure of sentences, Blühdorn 2012, pp. 252-262 claims a three-part differentiation of the NEG-scope in sentences)
  - What about cases of "Sondernegation" or partial/special negation, involving focus particles?

# The Notion of Scope (of Negation)

- Accordingly, in "simple" sentences, the scope of negation should always be the whole sentence
  - More interesting in complex sentences with additional non-sentential propositions (deverbal noun phrases, participle constructions, modal verb constructions (?), ...)
- Scope of negation is a controversial notion
  - A well-thought-out definition is inevitable
  - Some illustrating examples from annotation guidelines for syntactically parsed corpora will follow

# The Notion of Scope (of Negation) in Corpus Annotation Guidelines

- ... urgently needs improvement

## 26.4 Negation

The negative element *not* is always left unlabeled and is attached in accordance with the policy governing the attachment of all adverbials. See reduced relatives sections in section 13 [Gerunds and Participles] and in section 8 [Shared Complements and Modifiers] for more information about negation in those cases.

```
( (S (NP-SBJ I)
  (VP do
    not
    (VP understand))
  .))
```

(82) Du mußt nicht kommen

In solchen Fällen sollte die Anbindung von Adjunkten und der Negationspartikel *nicht* ihrem Skopus entsprechen (wenn möglich). Um den Skopus genau zu bestimmen, empfiehlt sich der folgende Test:

- (83) du mußt nicht kommen →
- Es ist **nicht** notwendig, daß du kommst
  - \*Es ist notwendig, daß du **nicht** kommst

Penn Treebank Guidelines,  
Bies et al. 1995, p. 311

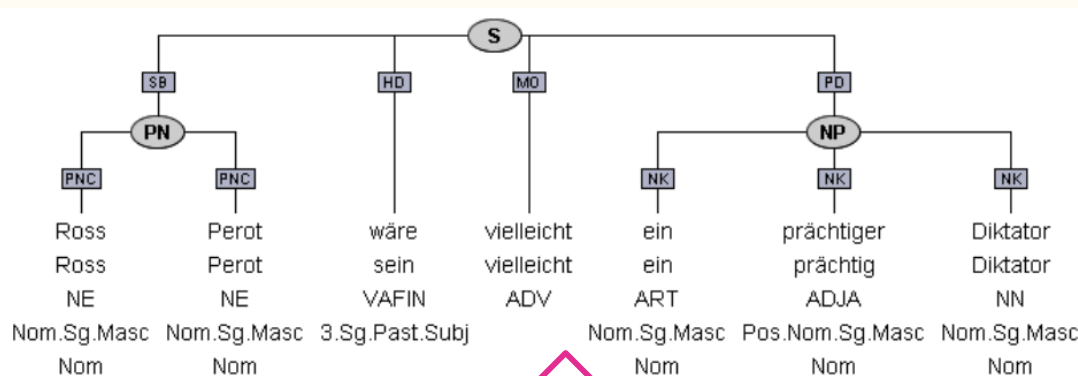
Tiger Treebank Guidelines,  
Albert et al. 2003, p. 91

# The Notion of Scope (of Negation)

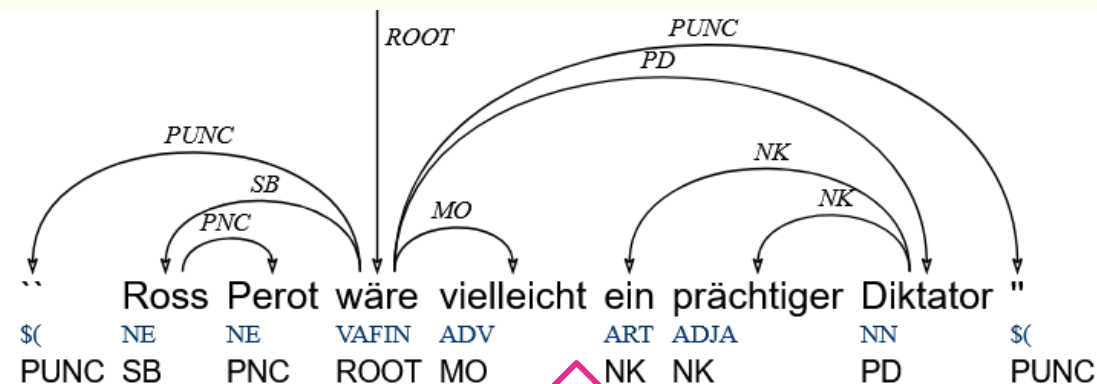
- "Scope: part of the sentence affected by the negation cue (Vincze et al. 2008), that is, all elements whose individual falsity would make the negated statement strictly true (Blanco and Moldovan 2011d)  
(Jiménez-Zafra et al. 2019, p. 13)
  - Good summary of corpora annotated for negation in English and other languages
  - Unfortunately, these datasets were designed for testing automatic negation detection and are not well-suited for general linguistic research

# The Notion of Scope (of Negation)

- Accordingly, in "simple" sentences, negation should always refer to the whole sentence
  - In these cases, *nicht* would (according to Tiger constituency and dependency treebank guidelines) be attached to highest node
  - Scope differences between *vielleicht*, *nicht*, and other operators are not expressed in the formalism (can only be concluded from word order tendencies)



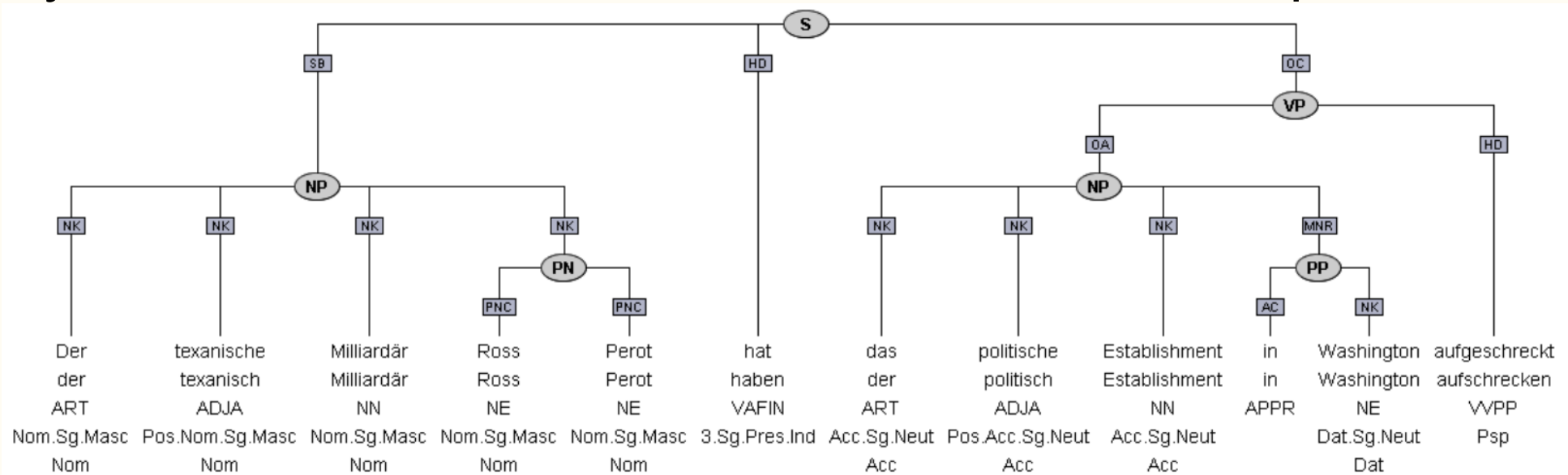
*nicht*



*nicht*

# The Notion of Scope (of Negation)

- Same scope in every position?
- Even  
*[Not the Texan billionaire] has shocked the political establishment should have propositional scope...*
- Syntactic treatment conflicts with semantic assumptions



nicht

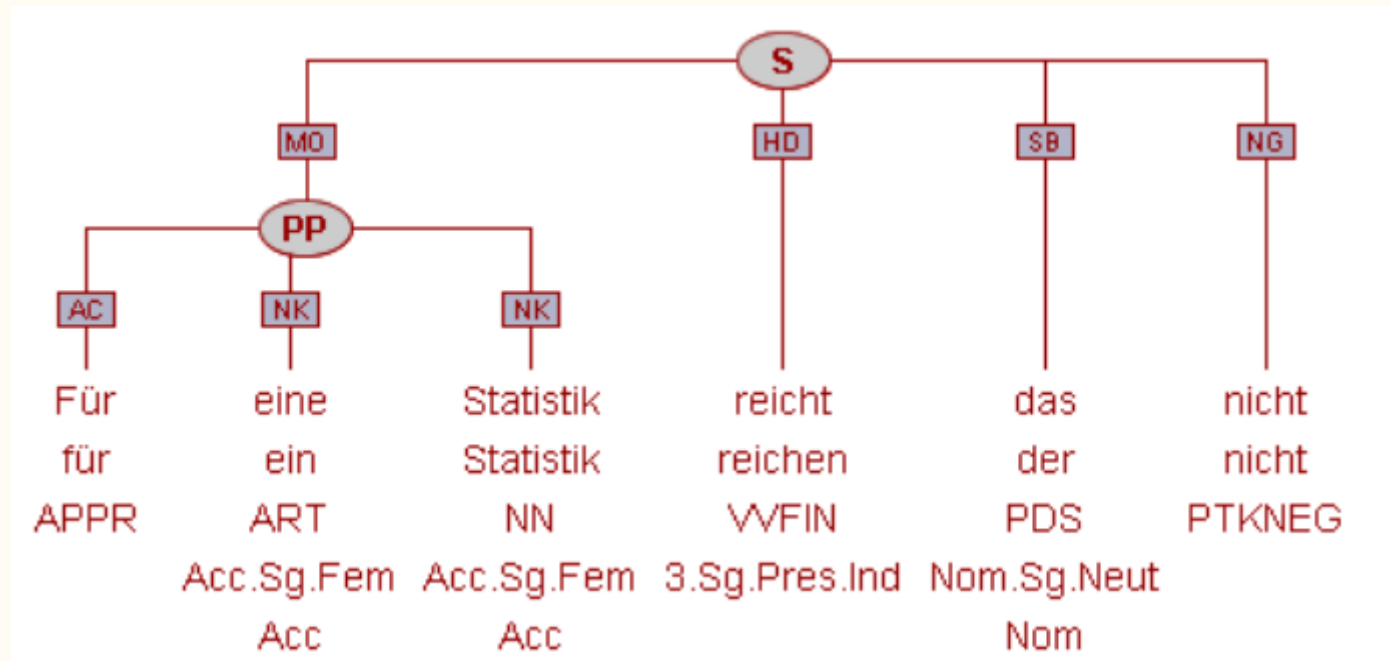
nicht

nicht

# Annotating Specific Aspects of Negation

## - Negated Expression

- ✓ scope of negation
- ✓ syntactical properties of *nicht*
- focus association?

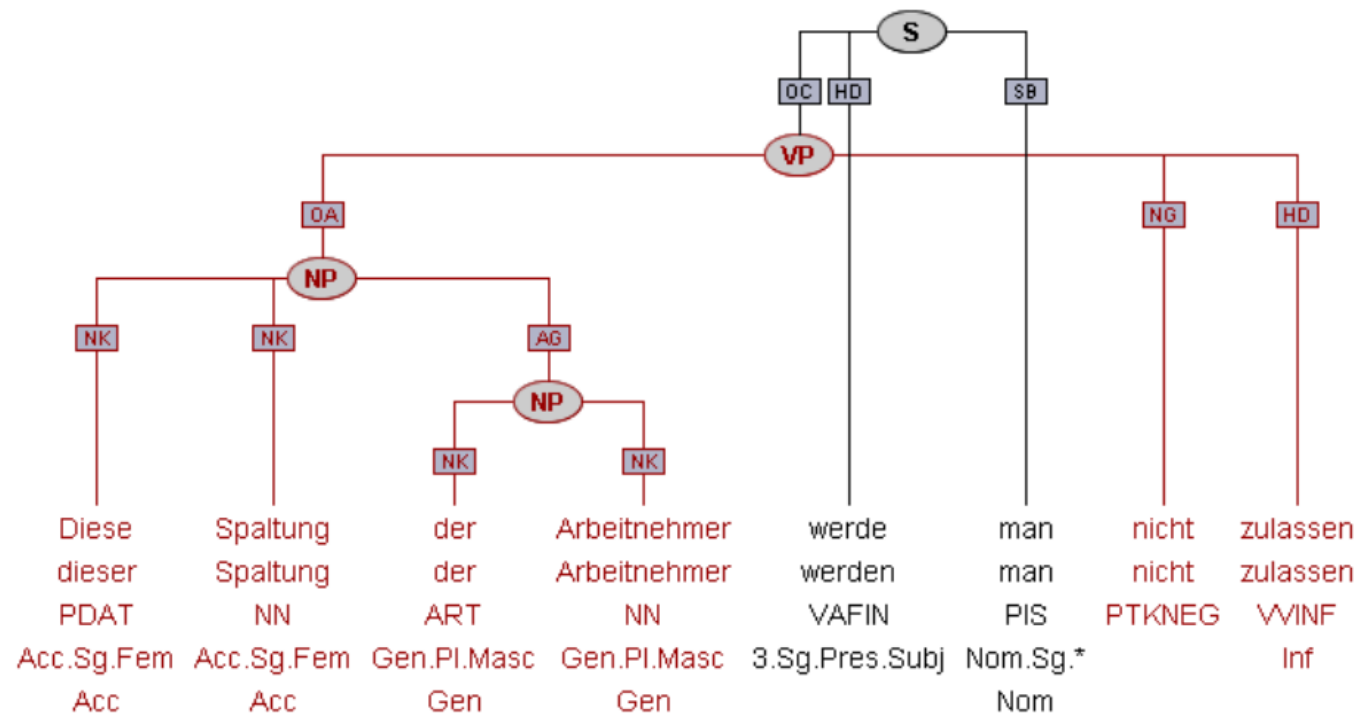




# Annotating Specific Aspects of Negation

## - Negated Expression

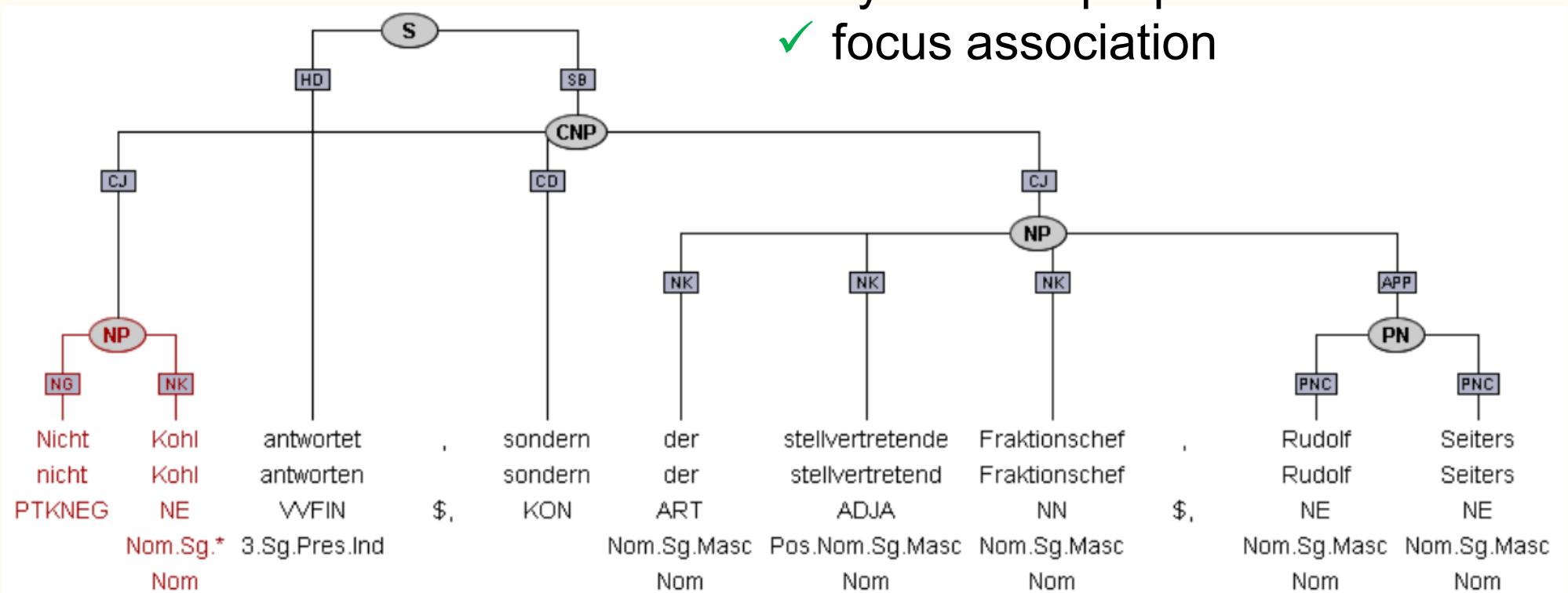
- scope of negation not represented?



# Annotating Specific Aspects of Negation

## - Negated Expression

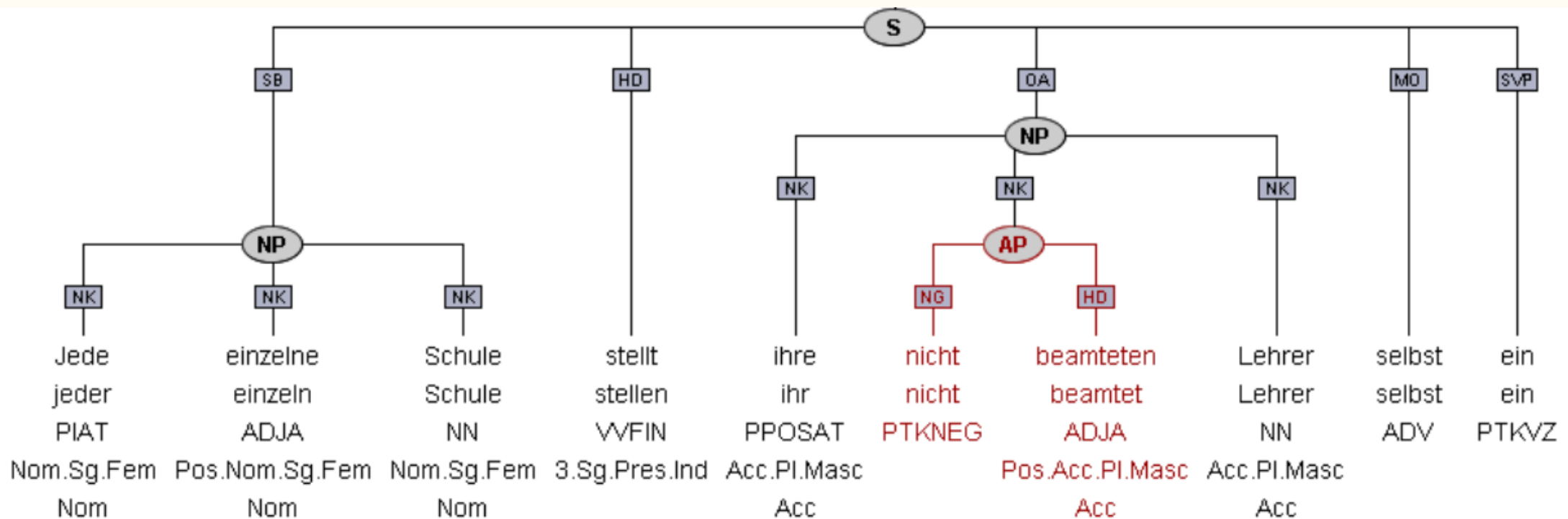
- scope of negation not represented?
- ✓ syntactical properties of *nicht*
- ✓ focus association



# Annotating Specific Aspects of Negation

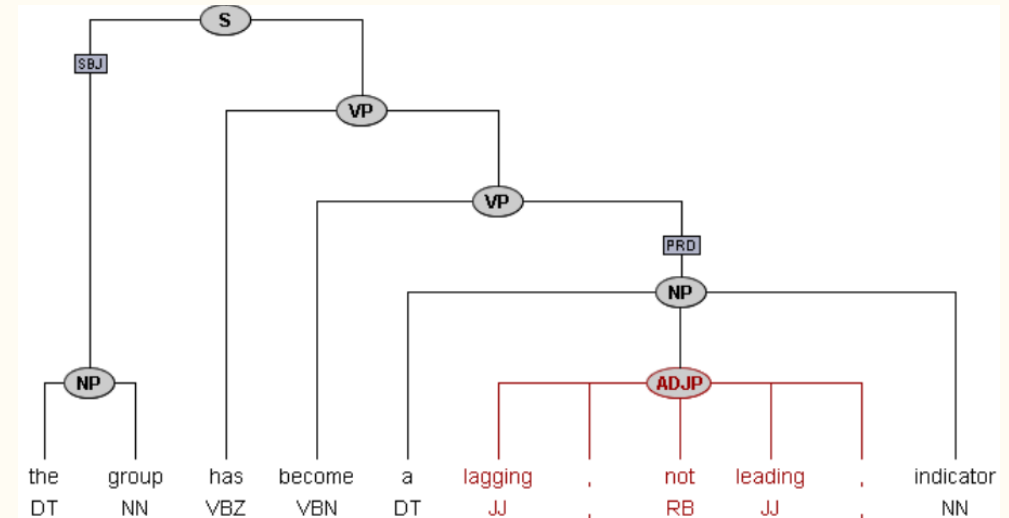
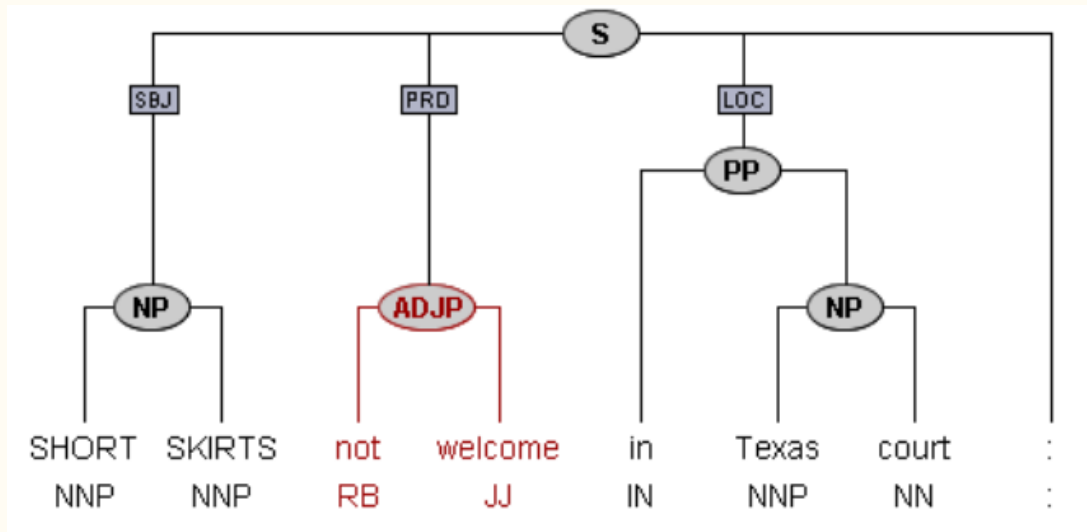
## - Negated Expression

- scope of negation (NP) not represented?
- ✓ syntactical properties of *nicht*



# Annotating Specific Aspects of Negation - Negated Expression

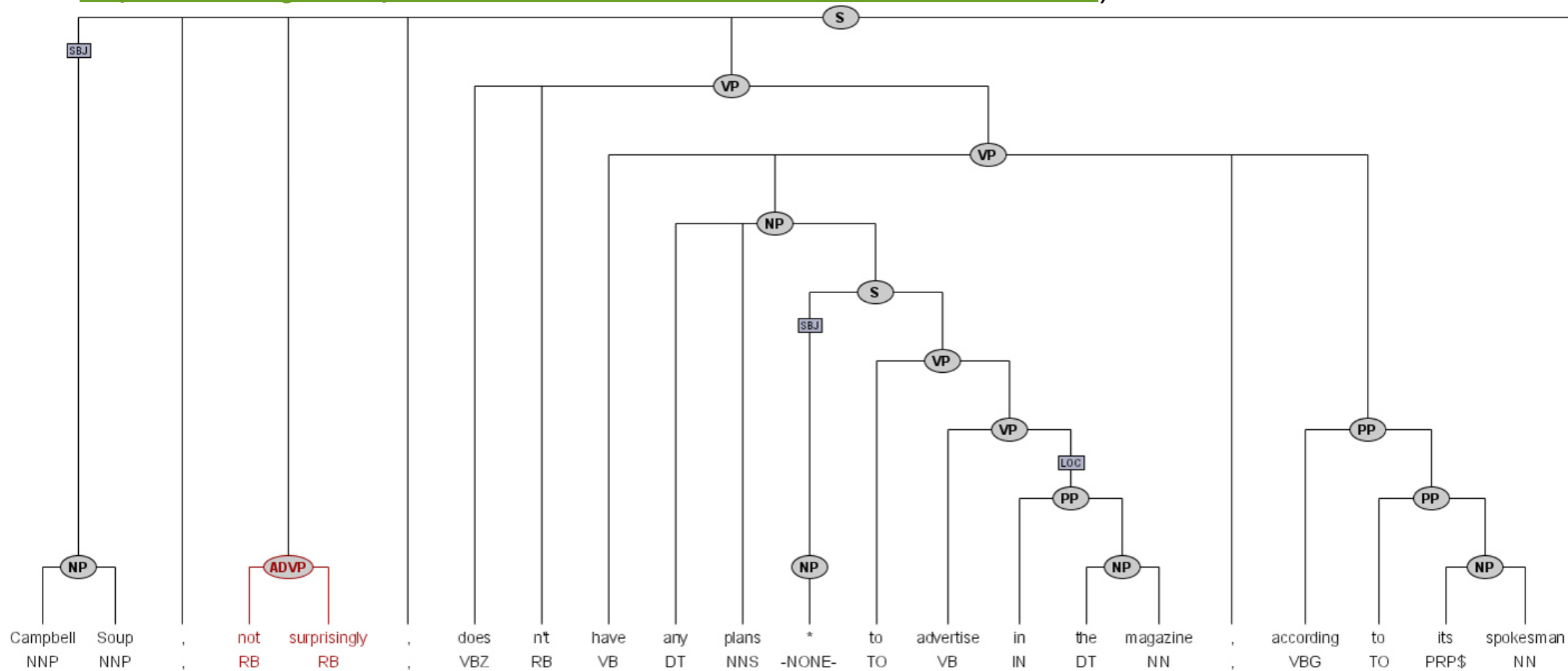
(Similar English examples from Penn Treebank Corpus;  
<https://catalog.ldc.upenn.edu/docs/LDC95T7/treebank2.index.html>)



# Annotating Specific Aspects of Negation

## - Negated Expression

(Similar English examples from Penn Treebank (PTB) Corpus;  
<https://catalog.ldc.upenn.edu/docs/LDC95T7/treebank2.index.html>)



# Conclusions

- Apparently, these descriptions do not reflect any state-of-the-art assumptions in syntax or semantics theory
- Negation is not described according to well-defined guidelines
- Moreover, the very superficial rules for binding can conflict with expressing certain semantic relations

# Conclusions

- In summary, we need to differentiate our criticism in:
  - inadequate syntactic interpretation
  - inadequate semantic interpretation
  - inconsistent annotations, resulting from
    - underspecified guidelines
    - inaccurate analyses by annotators

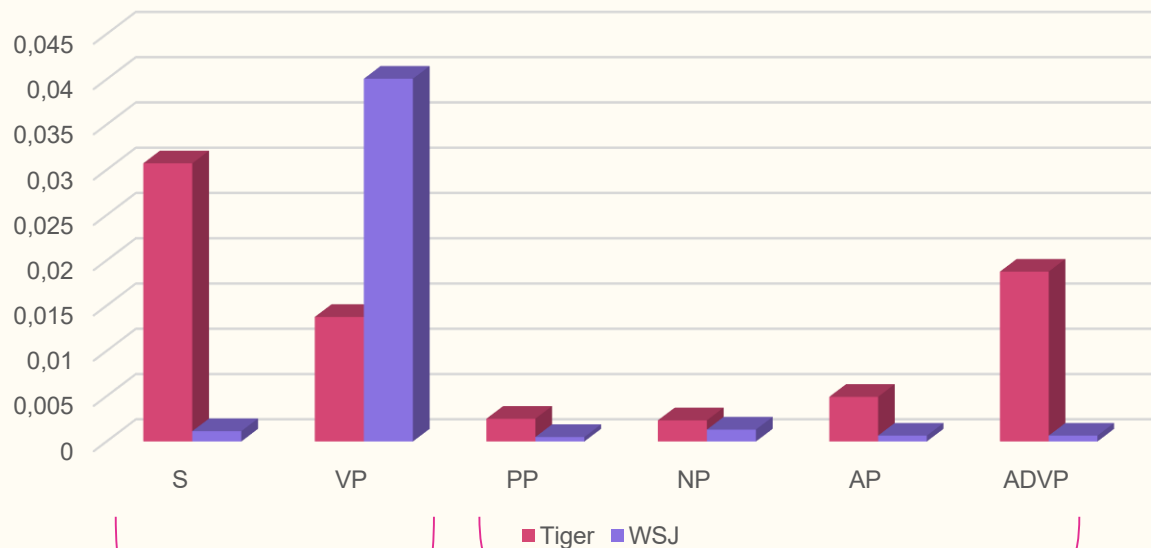
# Conclusions

- We would like to use this data to investigate quantitative tendencies of negation in different languages, varieties, registers, etc.
- This is only reasonable if we know
  - what the differences we measure mean and
  - if the corpus annotations are systematic
- In the following, quantitative differences in negation attachment between the two corpora we saw are measured



# "Negation Targets" of *nicht* and *not* in German Tiger Corpus vs. English PTB Corpus

Tiger-WSJ Negation Types



"Special negation" ("Sondernegation")

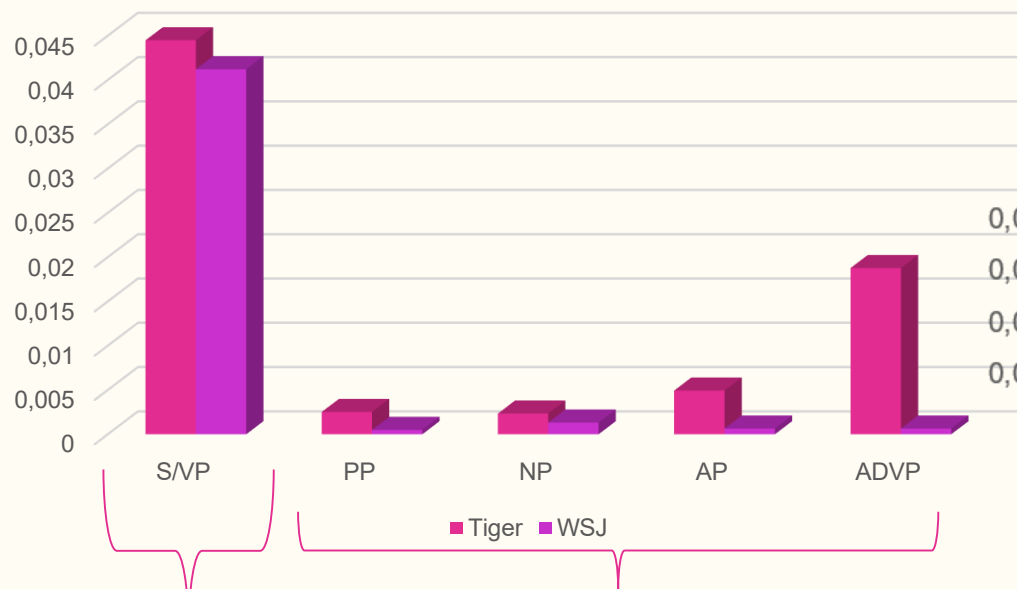
"Sentence negation" ("Satznegation")

- Differences in "special negation" reflect
  - different tendencies in syntactic attachment domain ("special negation" seems to be way more frequent in German)
  - strong tendency in German to lexicalize combinations such as *nicht mehr*, *nicht nur (... sondern auch)*, *noch nicht*, *gar nicht*, etc. (→ADVP)

Difference S-attachment/VP-attachments result from different grammatical systems and/or differences in the grammatical interpretation (annotation guidelines); they do not reflect differences in scope (e.g. propositional versus event negation)

# "Negation Targets" of *nicht* and *not* in German Tiger Corpus vs. English PTB Corpus

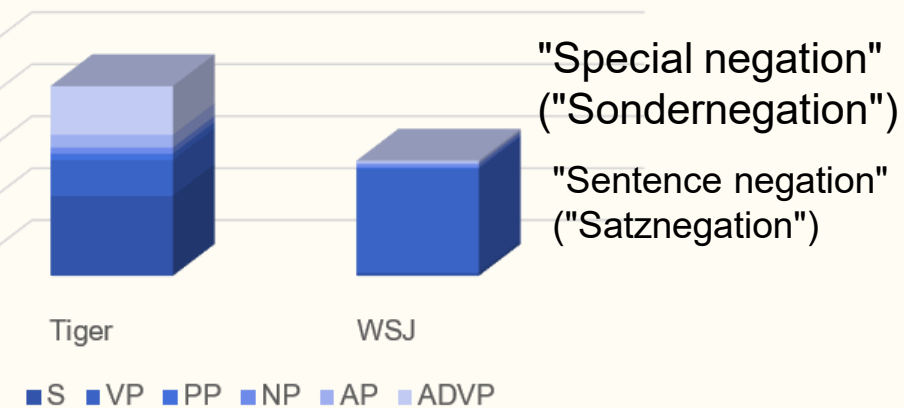
S/VP Conflated



"Sentence negation" ("Satznegation")

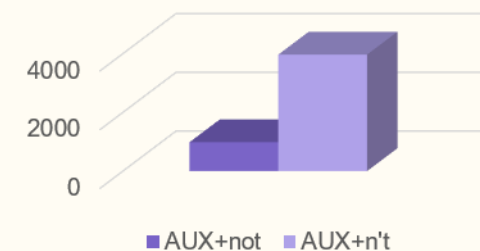
"Special negation" ("Sondernegation")

Tiger-WSJ Negation Types



Side note:

*not* vs. *n't* in WSJ



# Conclusions

- Clearly significant differences in relative frequency of negation in the two newspaper corpora
  - It results from the relatively high usage of "Sondernegation" and complex expressions involving *nicht*
- We cannot investigate scope differences and differences in focus association of negation
- So, at the end, I like to suggest a way of how these notions could possibly be treated in corpora

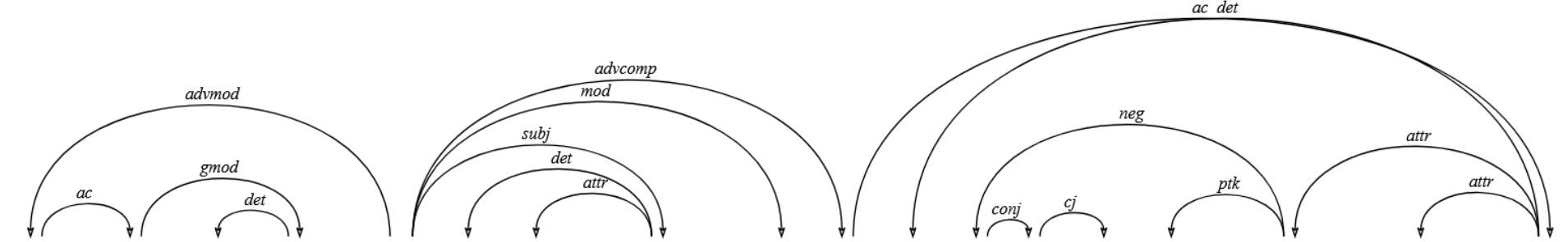
# Annotation Suggestion

- Different structurally **independent** layers of description
  - Negation cue
    - Separate layer with normalized forms
  - Syntactic embedding of negation cue
    - Probably dependency parses
    - Probably only partial description via phrase chunks
  - Focus association (if given in cases of "special negation")
    - Pointing relation between focus exponent and negation cue
  - Negation scope
    - Well-defined notion of scope, annotation realized as span

# Annotation Suggestion

[word]	Dem	ist	nix	hinzuzufügen	!
[S]	s1				
[pos]	PDS	VMFIN	PIS	VVIZU	\$.
[lemma]	d	sein	nix	hinzufügen	!
[NegCue]			nichts		
[NegScope]	S				

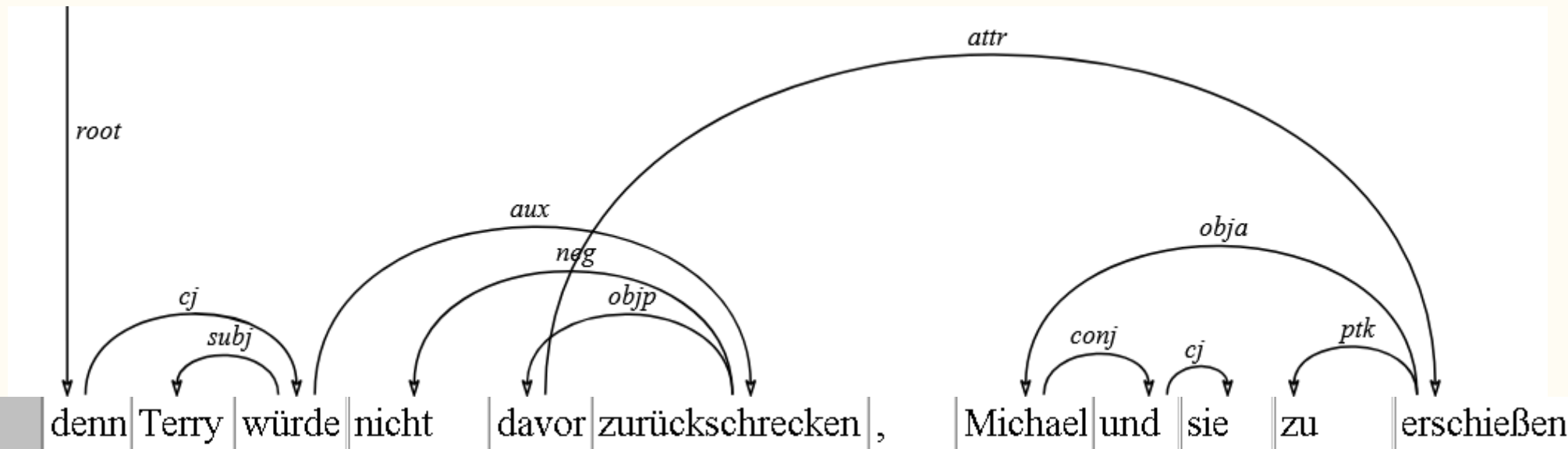
# Annotation Suggestion



[word]	Nach	Abschluss	der	Vorrunde	stehen	die	jungen	Rosenheimer	damit	auf	einem	nie	und	nimmer	zu	erwartenden	zweiten	Tabellenplatz	.
[S]	s1																		
[pos]	APPR	NN	ART	NN	VVFIN	ART	ADJA	NN	PAV	APPR	ART	ADV	KON	ADV	APPR	ADJA	ADJA	NN	\$.
[lemma]	nach	Abschluss	d	Vorrunde	stehen	d	jung	Rosenheimer	damit	auf	ein	nie	und	nimmer	zu	erwartend	zweit	Tabellenplatz	.
[NegCue]												nie und nimmer							
[NegScope]											NP								

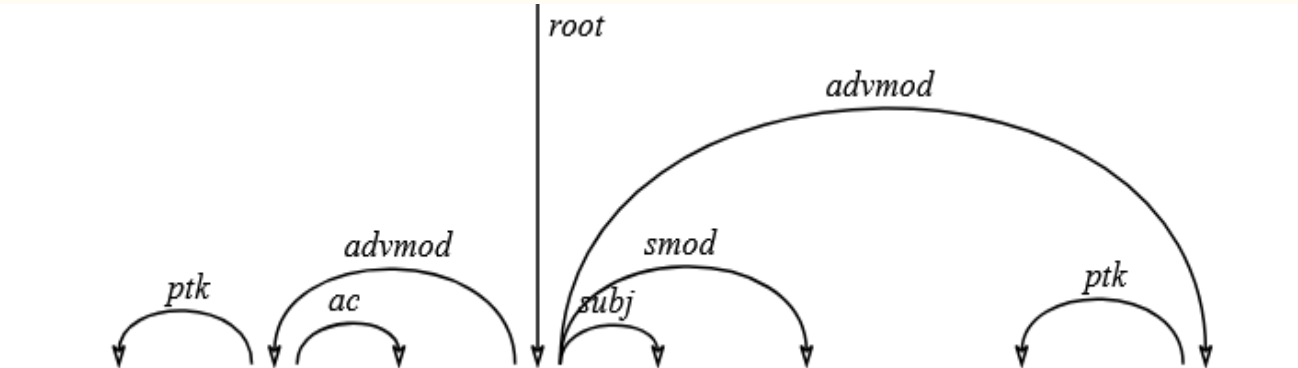
<https://shorturl.at/iGWIQ>

# Annotation Suggestion



[word]	denn	Terry	würde	nicht	davor	zurückschrecken	,	Michael	und	sie	zu	erschießen
[S]	s1											
[pos]	KON	NE	VAFIN	PTKNEG	PAV	VVINF	,	NE	KON	PPER	PTKZU	VVINF
[lemma]	denn	Terry	werden	nicht	davor	zurückschrecken	,	Michael	und	sie	zu	erschießen
[NegCue]				nicht								
[NegScope]		S										

# Annotation Suggestion



[word]	Nichtmal	mit	Raten	kam	man	hierbei	sonderlich	weit	.
[S]	s1								
[pos]	ADV	APPR	NN	VVFIN	PIS	PAV	ADJD	ADJD	\$.
[lemma]	nichtmal	mit	raten	kommen	man	hierbei	sonderlich	weit	.
[NegCue]	nicht einma:								
[NegScope]	S								
[Focus_Ass]	foc-ass		foc						



# Is It Worth the Effort?

- Annotation (formulating guidelines for analysis, measuring agreement between annotators, discussing difficult and unforeseen cases, etc.) greatly helps in understanding the phenomenon being explored
- Splitting the highly complex task (syntactic interpretation, scope specification, focus specification, etc.) into clearly defined parts significantly facilitates the project management
- No data collection can only roughly indicate the extent of diversion that negation has to offer
- Having independent annotation systems in the data will provide a high degree of flexibility in data analysis

# Tool Chain

- There is no all-in-one solution for processing corpus data
  - annotation  $\neq$  corpus search
  - different data formats for different purposes (editors processing xml data formats in background in multi-layer annotation procedures, editors for syntax, processing conll formats)
- No standard procedure to process corpus data

# Tool Chain (Suggestions)

- Annotation
  - EXMARaLDA (<https://exmaralda.org>, Schmidt 2012) as a easy-to-use annotation tool with unlimited, freely definable token and span annotation layers and increasing possibilities to implement automatic annotation steps
  - INCEpTION (<https://inception-project.github.io/>, Klie et al. 2018) as a tool providing the most flexibility in combining different annotation concepts (manual analyses, server installation, multi-annotator management, ...)
  - WebLicht ([https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main\\_Page](https://weblicht.sfs.uni-tuebingen.de/weblichtwiki/index.php/Main_Page)) for automatic NLP processing
- Converters for data merging and transformation
  - Pepper (<https://corpus-tools.org/pepper>)
  - Integrated import and export functions of programs above
- Search and statistics tools
  - ANNIS (<https://corpus-tools.org/annis>, Krause/Zeldes 2016)
  - R (<https://www.r-project.org/>, Kohl 2015), RStudio (<https://posit.co/download/rstudio-desktop/>)

# Tool Chain (Suggestions)

- Annotation

- EXMARaLDA (<https://exmaralda.org>, Schmidt 2012) as a easy-to-use annotation tool with unlimited, freely definable token and span annotation layers and increasing possibilities to implement automatic annotation steps
- INCEpTION (<https://inception-project.github.io/>, ... 2018) as a tool providing the most flexibility in combining annotation concepts (manual analyses, server installation, multi-...)
- WebLicht (<https://weblicht.sf.net/>) for automatic NLP processing (in Page) for

- Converters

- Pepper ([Pepper](#))
- Integrated [functions](#) of programs above

- Search and search tools

- ANNIS (<https://corpus-tools.org/annis>, Krause/Zeldes 2016)
- R (<https://www.r-project.org/>, Kohl 2015), RStudio (<https://posit.co/download/rstudio-desktop/>)

Workshop tomorrow: We will try most of them!  
 Thanks for listening and productive discussion!  
 The HU korpling-team is happy about cooperation!

# Thanks For Listening!

## • Reference List

- Blühdorn, Hardarik (2012) *Negation im Deutschen*. Syntax, Informationsstruktur, Semantik. Stuttgart; Narr. <https://ids-pub.bsz-bw.de/frontdoor/index/index/year/2023/docId/12348>
- Jacobs, Joachim (1982) *Syntax und Semantik der Negation im Deutschen*. München; Wilhelm Fink Verlag. [https://digi20.digitale-sammlungen.de/de/fs1/object/display/bsb00042700\\_00001.html](https://digi20.digitale-sammlungen.de/de/fs1/object/display/bsb00042700_00001.html)
- Jiménez-Zafra, Salud María; Morante, Roser; Martín-Valdivia, María Teresa; Ureña-López, L. Alfonso (2020) Corpora Annotated with Negation: An Overview. *Computational Linguistics*, 46(1):1–52. <https://aclanthology.org/2020.cl-1.5/>
- Klenner, Manfred; Fahrni, Angela; Petrakis (2009) PolArt: A Robust Tool for Sentiment Analysis. *Proceedings of the 17th Nordic Conference of Computational Linguistics (NODALIDA 2009)*. Odense, Denmark, 235–238. <https://aclanthology.org/W09-4635/>
- Klie, Jan-Christoph; Bugert, Michael; Boullosa, Beto; Eckart de Castilho, Richard; Gurevych, Iryna (2018) The INCEpTION Platform: MachineAssisted and Knowledge-Oriented Interactive Annotation. In *Proceedings of the 27th International Conference on Computational Linguistics - COLING 2018*. New-Mexico, USA 5-9. <https://aclanthology.org/C18-2002/>
- Kohl, Matthias (2015) *Introduction to statistical data analysis with R*. London: bookboon.
- Krause, Thomas/Zeldes, Amir (2016) ANNIS3: A new architecture for generic corpus query and visualization. In: *Digital Scholarship in the Humanities 2016* (31). (<https://bit.ly/2JumGIR>)
- Petran, F., Bollmann, M., Dipper, S., & Klein, T. (2016) ReM: A reference corpus of Middle High German – corpus compilation, annotation, and access. *Journal for Language Technology and Computational Linguistics*, 31(2), 1–15. <https://jllcl.org/article/view/208>
- Schmidt, Thomas (2012): EXMARaLDA and the FOLK tools – two toolsets for transcribing and annotating spoken language. In: *Proceedings of LREC 2010*. Valletta, Malta, [http://www.lrec-conf.org/proceedings/lrec2012/pdf/529\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2012/pdf/529_Paper.pdf)