

DEVELOPING AN ANNOTATION SCHEME AND A REFERENCE CORPUS FOR PREPOSITION SENSES IN GERMAN

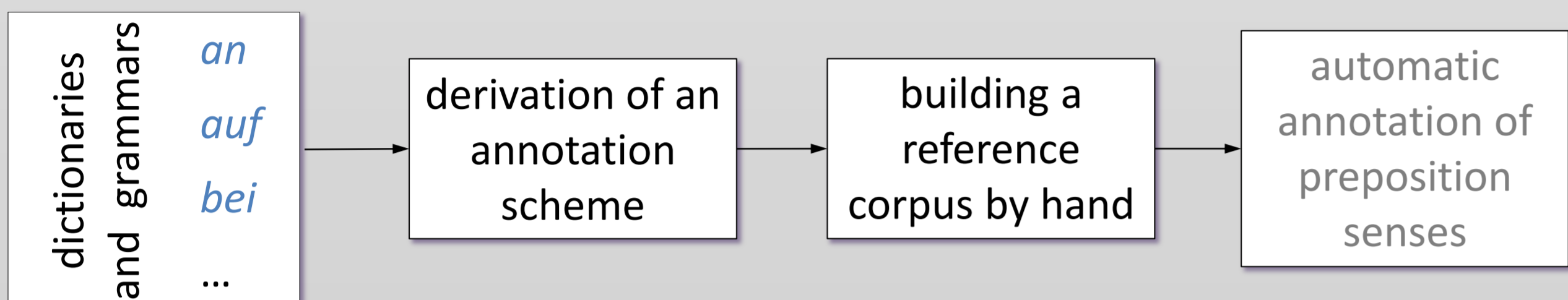
Antje Müller, Katja Keßelmeier, Claudia Roch, Jan Strunk, Tobias Stadtfeld, Tibor Kiss
{mueller, kesselmeier, roch, strunk, stadtfeld, tibor}@linguistics.rub.de

- Preposition-Noun-Combinations (PNCs) minimally consist of a preposition (P) and a singular count noun (N) without a determiner (D)
auf Anfrage (after being asked), *unter Voraussetzung* (under prerequisite), *unter Androhung* (under threat), *mit Vorbehalt* (with reservation)
- PNCs can be extended as long as the resulting projection does not contain a D
auf parlamentarische Anfrage (after being asked in parliament), *bei absolut klarer Zielsetzung* (given a clearly present aim), *mit schwer beladenem Rucksack* (with heavily loaded backpack)

- PNCs violate the rule that singular count nouns have to appear with a determiner:
„Substantive mit Merkmalkombination „zählbar“ plus Singular haben also grundsätzlich immer ein Artikelwort bei sich [...]“ (Dudenregel 442)
- PNCs have been treated as exceptions but they are indeed productive and no more idiomatic than other phrasal combinations (Dömges et al. 2007, Kiss 2007)

Licensing conditions for PNCs? → not every P can appear with every N in a PNC, probably depending on the meaning of the P

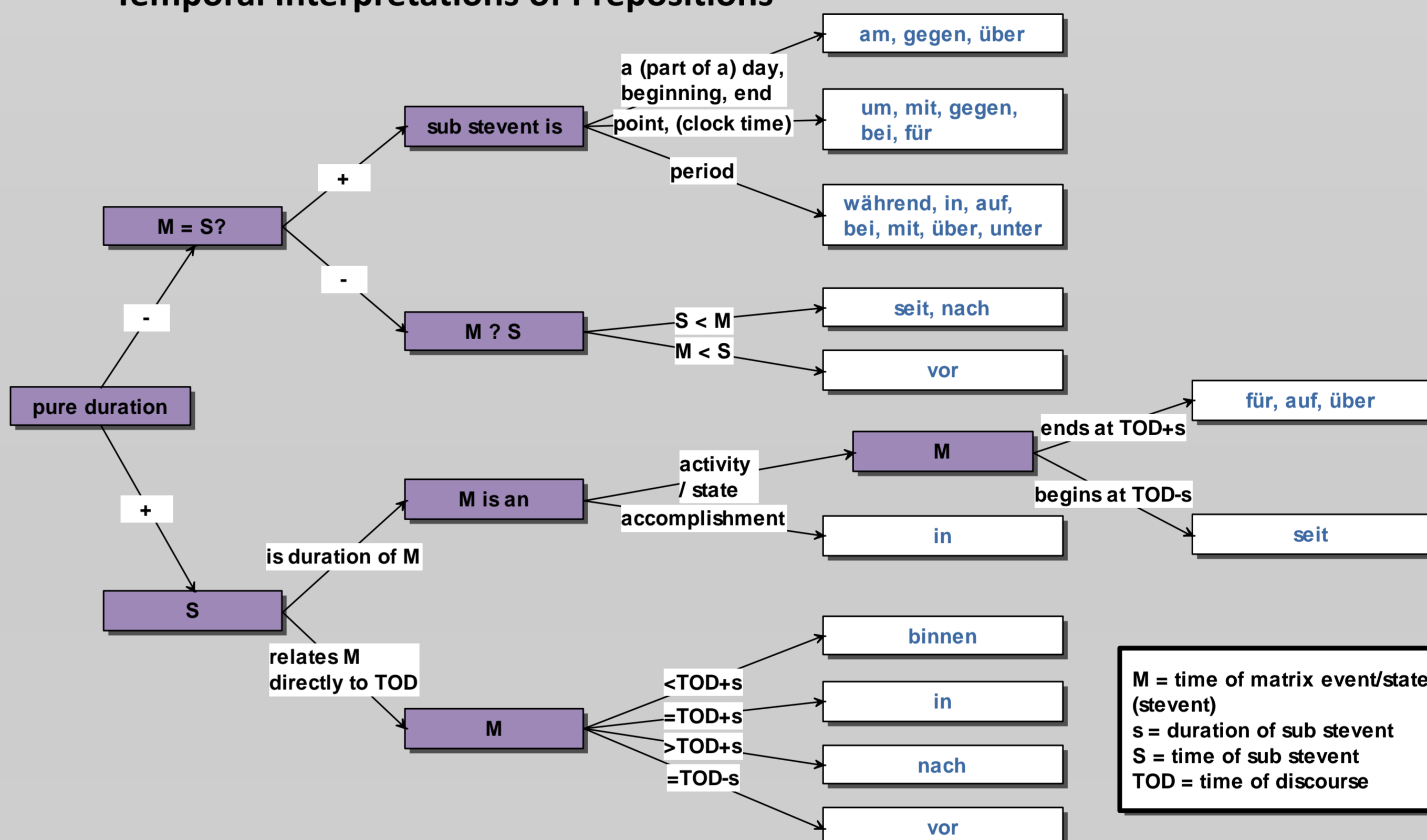
We limit our analysis to the following 23 prepositions: *an, auf, bei, binnen, dank, durch, für, gegen, gemäß, hinter, in, mit, mittels, nach, neben, ohne, seit, über, um, unter, vor, während, wegen*



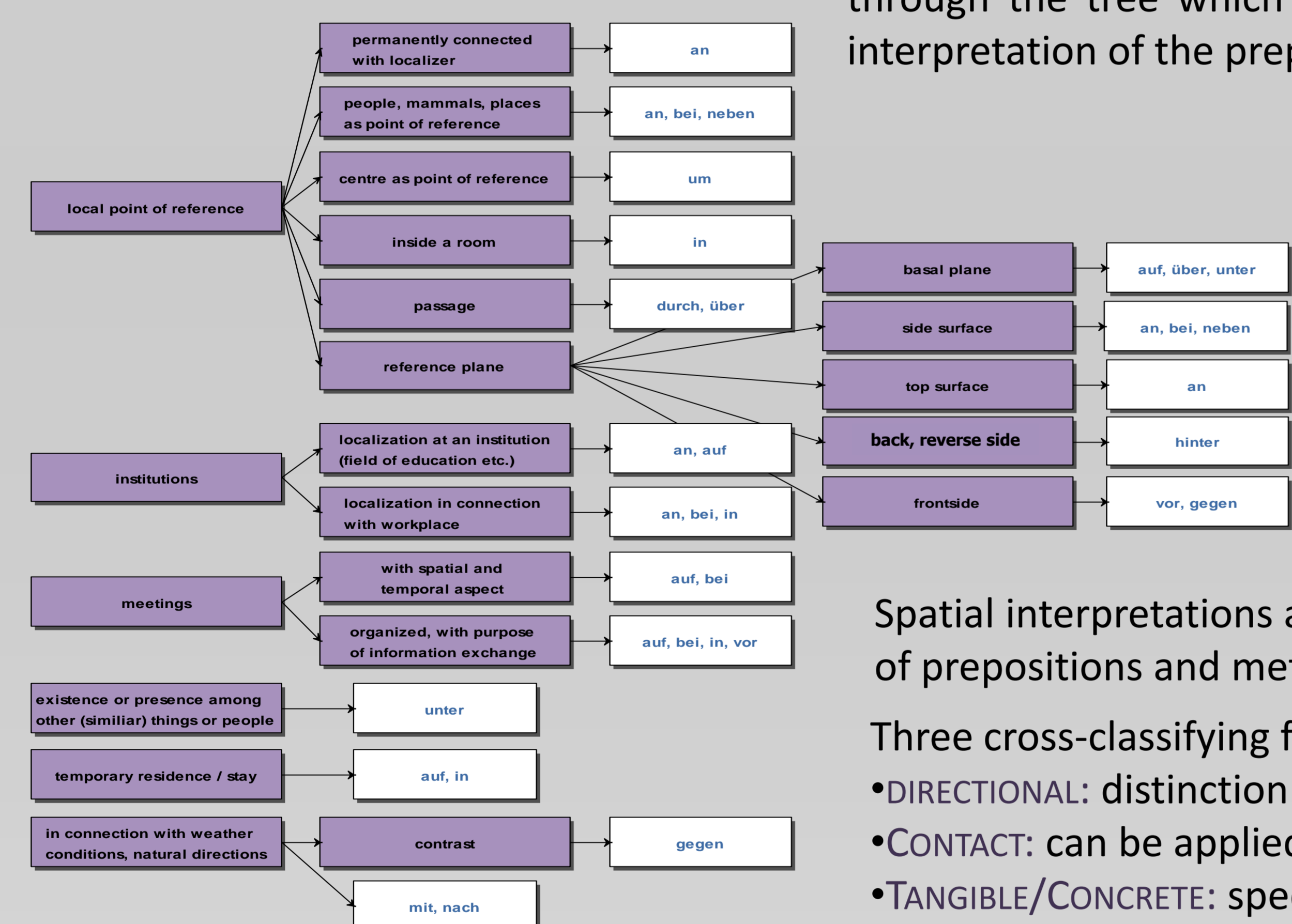
- Combining details of several reference works to define a tagset of preposition senses:
- Deutsche Grammatik (Helbig & Buscha, 2001): about 54 different categories, 1-6 senses per preposition
 - DDaF – Duden Deutsch als Fremdsprache (Duden, 2002): about 40 categories, 1-18 senses per preposition
 - Lexikon deutscher Präpositionen (Schröder, 1986): feature based analysis, about 200 features

- Annotation of temporal senses of Ps is based on a decision tree for temporal Ps of German developed by Durell and Brée (1993)
- Decisions in the tree are based on the distinction between a matrix and a subordinate event, on the characteristics of these events and on the identification of the temporal relationship between them

Temporal Interpretations of Prepositions



Spatial Interpretations of Prepositions



Annotators have to find a valid path through the tree which reflects the given interpretation of the preposition.

Spatial interpretations are organized in a decision tree based on Schröder (1986). It includes original spatial interpretations of prepositions and metaphorical extensions of these interpretations.

Three cross-classifying features are added:

- DIRECTIONAL**: distinction between local and directional interpretations, can be added to all leaves of the decision tree
- CONTACT**: can be applied to all reference planes, signifying whether or not a contact has been established
- TANGIBLE/CONCRETE**: specifies the concreteness of the planes

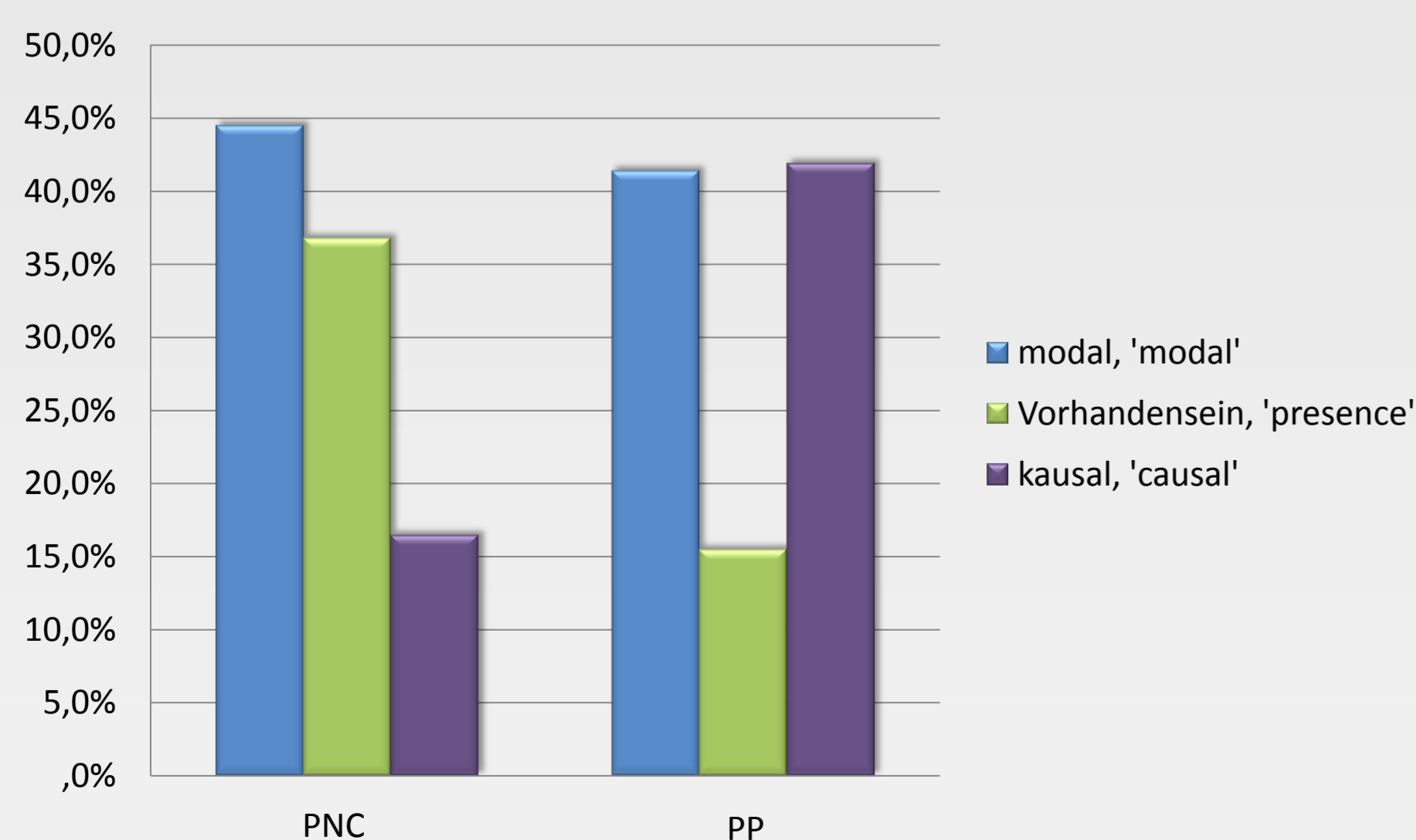
Further interpretations: MODAL (ART&WEISE, MEDIAL, INSTRUMENTAL, ENTSPRECHUNG, BEGLEITUMSTAND), KAUSAL (KAUSAL, KONDITONAL, KONZESSIV, FINAL, KONSEKUTIV), STELLUNGNAHME, ZUSTAND, INKLUSIV, QUANTITÄTSDIFFERENZ, UNTERORDNUNG, ZUORDNUNG, WECHSELBEZIEHUNG, RANGFOLGE, ÜBER-/UNTERSCHREITUNG, AGENS, EMPFÄNGER, BEZUGSPUNKT, STELLVERTRETUNG, AUSTAUSCH, BETEILIGUNG, KOMITATIV, RESTRIKTIV, ADVERSATIV, DISTRIBUTIV, VORHANDENSEIN, KOMPARATIV, PARTITIV, KOPULATIV, THEMA

Annotation on multiple levels:

- POS, morphology & chunks: POS-tagging with TreeTagger and Regression Forest Tagger; morphological information from morphological component SMOR; chunks added by TreeTagger
- countability: added by a countability classifier
- preposition meaning: annotated manually on the basis of our scheme

		die	Forderung	nach	Stilllegung	einer	Verbrennungsanlage
lexical	POS	ART	NN	APPR	NN	ART	NN
	morphology	def.acc.sg.fem	reg.acc.sg.fem	dat	reg.dat.sg.fem	indef.gen.sg.fem	reg.gen.sg.fem
	countability				countable		
	prep. meaning			final.governed			
syntactic	chunks	NC		PC		NC	
	relational dependency	ext-depend.	head	dependent.PNC			
global	hit				head	dependent.post-gen	
	sentence	title: no		idiom: no			

- noun meaning: information from HaGenLex and GermaNet
- external and internal dependencies: annotated manually
- hit: includes PNC or PP; possibility to add information about idiomaticity, telegraphic style ...
- sentence: sentences automatically tagged as headlines whenever appropriate



Some first results concerning *ohne* 'without':

- ohne* appears about ten times more often in PNCs than in PPs
- causal interpretation: significantly more often in PPs
- presence interpretation: significantly more often in PNCs
- significantly more internal dependents in PPs than in PNCs

Vorhandensein: Jugendliche ohne Ausbildungsplatz ("presence") ("Teenagers without an apprenticeship position")
modal: Eine Mofalenkerin, die ohne Helm unterwegs war, ... ("modal") ("A motorcyclist, riding without a helmet, ...")
kausal: Ohne Basiswissen bleibt die Innovation auf der Strecke. ("causal") ("Without basic knowledge, innovation falls by the wayside.")

