



## Metaphor annotation

The poster presents a pilot study to develop an automatic system for metaphor recognition. The system is designed to recognize **deliberate metaphors** and to mark the metaphorically used word ("**MW**") along with a literally used word ("**anchor**") in the immediate context. The MW in its normal meaning comes from the **source domain** while the anchor's meaning indicates the **target domain**.

The data of the pilot study are German sermons (<https://www.sermon-online.de/>). An example is given in (1). The passage is about the ideas and thoughts that people have about their lives. Accordingly, the expression *Gedanken* 'thoughts' within the compound *Gedankengebäude* 'thought building' is meant literally and contrasts with the metaphorical expressions referring to construction (*Gebäude* 'edifice, *niederreißen* 'tear down').

(1) mit ihnen [= diesen Waffen] **reißen** wir alle hohen **Gedankengebäude nieder**  
'with them [= these weapons] we **tear down** all the high **edifices of thought**' (#2 Jürg Birnstiel 19.08.2001)

## Annotation guidelines

- ▶ **Deliberate metaphors:** We annotate:
  - 1 **Unconventional, novel metaphors** (2)
  - 2 **Revitalized conventional metaphors:** metaphors that are revitalized by the fact that the speaker repeatedly refers to the metaphor's image in the course of the speech (1)
  - 3 **Similes:** i.e., overtly marked metaphors that may be introduced by a **marker** like *like*, *similar to*, etc. (3)
- ▶ **Metaphor** → **anchor:** Usually, in the immediate context of the metaphorically used word (**MW**), there is a word that is used literally, showing the metaphor-typical domains switch. We call this word (which belongs to the target domain) an **anchor**.
  - ▶ Relate the MW to its anchor by a pointer
  - ▶ Note: MW and anchor can occur within one word (1), (2)
- ▶ **Direct dependency relation:** Typically, the MW and its anchor are in a fairly direct dependency relation, e.g., verb and its subject/object (*der Zorn beginnt zu kochen* 'anger begins to boil', cf. (2)) or relations between heads and modifiers (*Gedankengebäude* 'thought building', *Silberseele* 'silver soul', cf. (1) and (3)).
- ▶ **Pronouns:** If the metaphor word or its anchor is a pronoun, a coreference link to the (content word) antecedent must also be inserted (3)

(2) Als der Demetrius seinen Kollegen diese Konsequenzen schildert, da beginnt in ihrer **Silberseele** der **Zorn** über diese religiös verbrämte Sabotage der Souvenirindustrie zu **kochen**.  
'When Demetrius describes these consequences to his colleagues, anger begins to boil in their silver souls over this religiously dressed up sabotage of the souvenir industry.'  
(#6 Theo Lehmann 13.11.1983)

Note: *Demetrius is a silversmith. The neologism Silberseele 'silver soul' is meant to express that these people were very fixated on their income, which they saw as threatened.*

(3) Wir haben einen ganz heftigen Widerstand gegen Gott. [...] **Das** ist **wie** die **Bremsen blockieren** beim Auto und das Auto quersteht.  
'We have quite a fierce resistance to God. [...] It's like the brakes blocking on a car and the car stalls.'  
(#7 Winrich Scheffbuch 20.06.1976)

<b>Corpus</b>	<b>#Texts</b>	106	<b>#MWs</b>	927
<b>statistics</b>	<b>#Tokens</b>	324,240	<b>#Anchors</b>	697

### Resources and References:

Annotation: **WebAnno** (Eckart de Castilho et al. 2016)  
 Processing of vectors: **gensim.models** (Rehurek and Sojka 2010)  
 Clustering: **nlTK.KMeansClusterer** (Bird, Loper, and Klein 2009)  
 Pretrained **German word embeddings**: <https://deepset.ai/>  
 Coreference: **CorZu** (Tuggener 2016)  
 Dependency parsing: **ParZu** (Sennrich et al. 2009)

Bird, Steven, Edward Loper, and Ewan Klein (2009). *Natural Language Processing with Python*. O'Reilly.  
 Eckart de Castilho, Richard et al. (Dec. 2016). "A Web-based Tool for the Integrated Annotation of Semantic and Syntactic Structures". In: *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)*, pp. 76–84.  
 Rehurek, Radim and Petr Sojka (2010). "Software Framework for Topic Modelling with Large Corpora". In: *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pp. 45–50.  
 Sennrich, Rico et al. (2009). "A new hybrid dependency parser for German". In: *Proceedings of the GSCL Conference*.  
 Shutova, Ekaterina, Simone Teufel, and Anna Korhonen (2013). "Statistical Metaphor Processing". In: *Computational Linguistics* 39.2, pp. 301–353.  
 Tuggener, Don (2016). "Incremental Coreference Resolution for German". PhD thesis. University of Zurich.

## Metaphor recognition

Shutova, Teufel, and Korhonen (2013) propose the following system for metaphor recognition:

- ▶ **Seed set:** They manually annotate V+N pairs: <verb,subject> or <verb,object>, where the verb is the MW
- ▶ **Clusters:** Feature-based clusters of similar verbs and nouns are generated for each verb and noun from the seed set
- ▶ **Application:** New texts are parsed and each verb-subject/object pair is checked if it is similar to a pair from the seed set, i.e. if it is in the corresponding cluster pairs

Our approach modifies and extends this system:

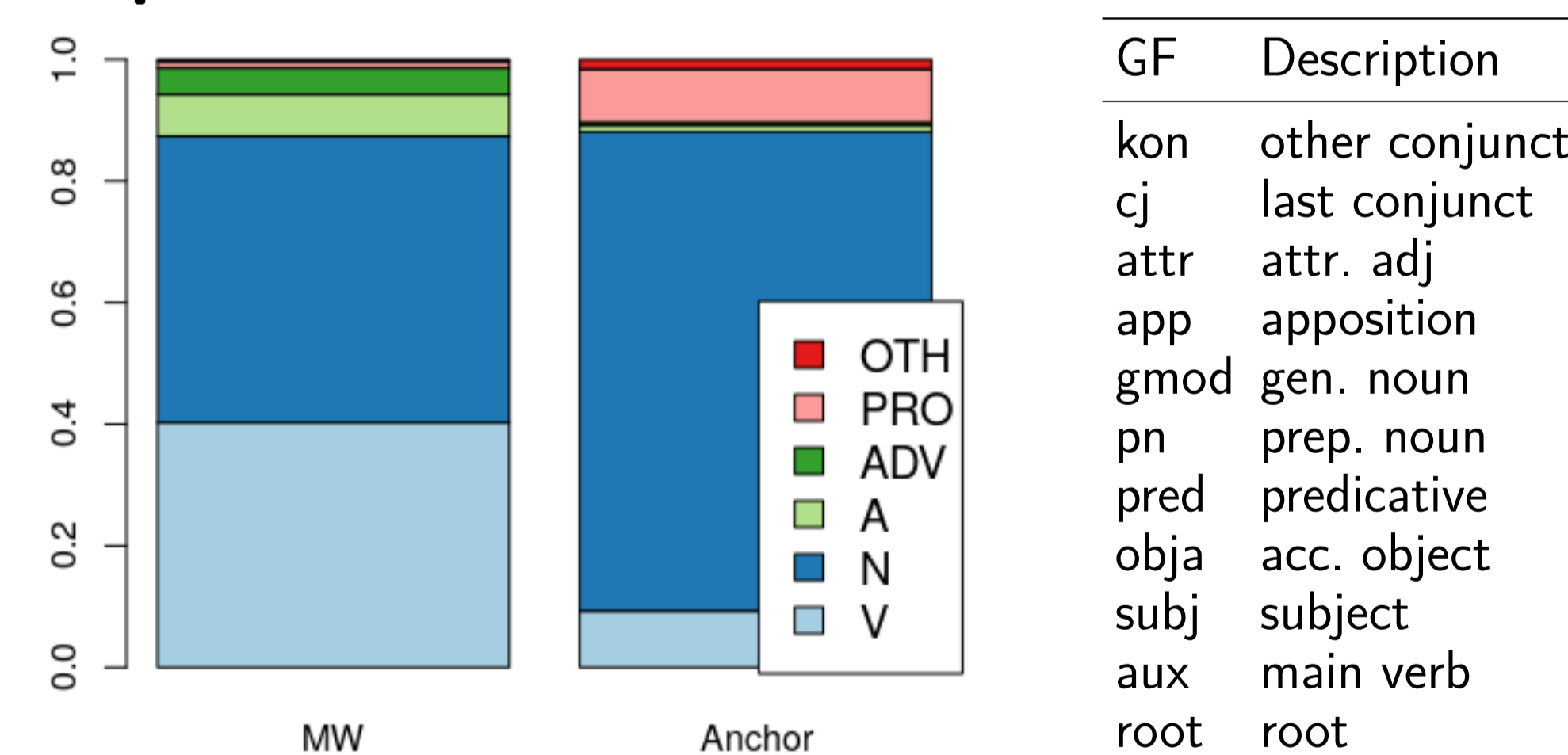
- ▶ **Seed set:** We annotate all kinds of POS and relations, e.g. <verb,subject>, <noun, adj.modifier>, etc.
- ▶ **Clusters:** We generate clusters based on word embeddings for verbs, nouns, and adjectives
- ▶ **Morphological analysis:** If a candidate compound is not contained in the cluster, the individual morphemes are considered (*Silberseele* → *Seele*)

## Pilot study

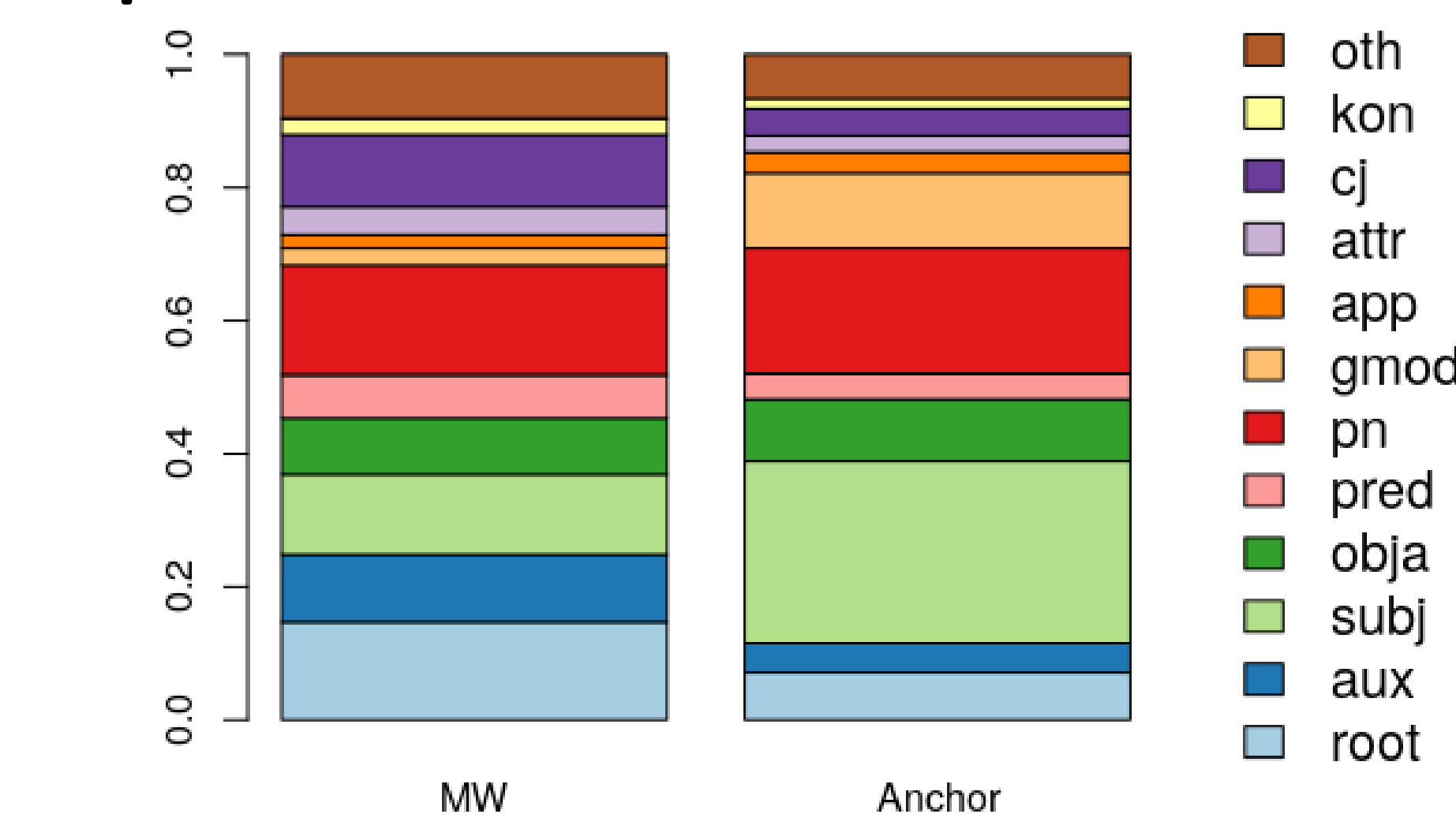
First results from 571 MW–Anchor pairs:

Top MWs		Top Anchors	
Lemma	Freq	Lemma	Freq
Licht 'light'	55	Gott 'god'	29
Finsternis 'darkness'	17	Wort 'word'	25
gebären 'give birth'	9	Leben 'live'	23
leuchten 'shine'	7	Blut 'blood'	20
dunkel 'dark'	7	Jesus	18
machen 'make'	7	sie 'she/they'	18

### Top POS:



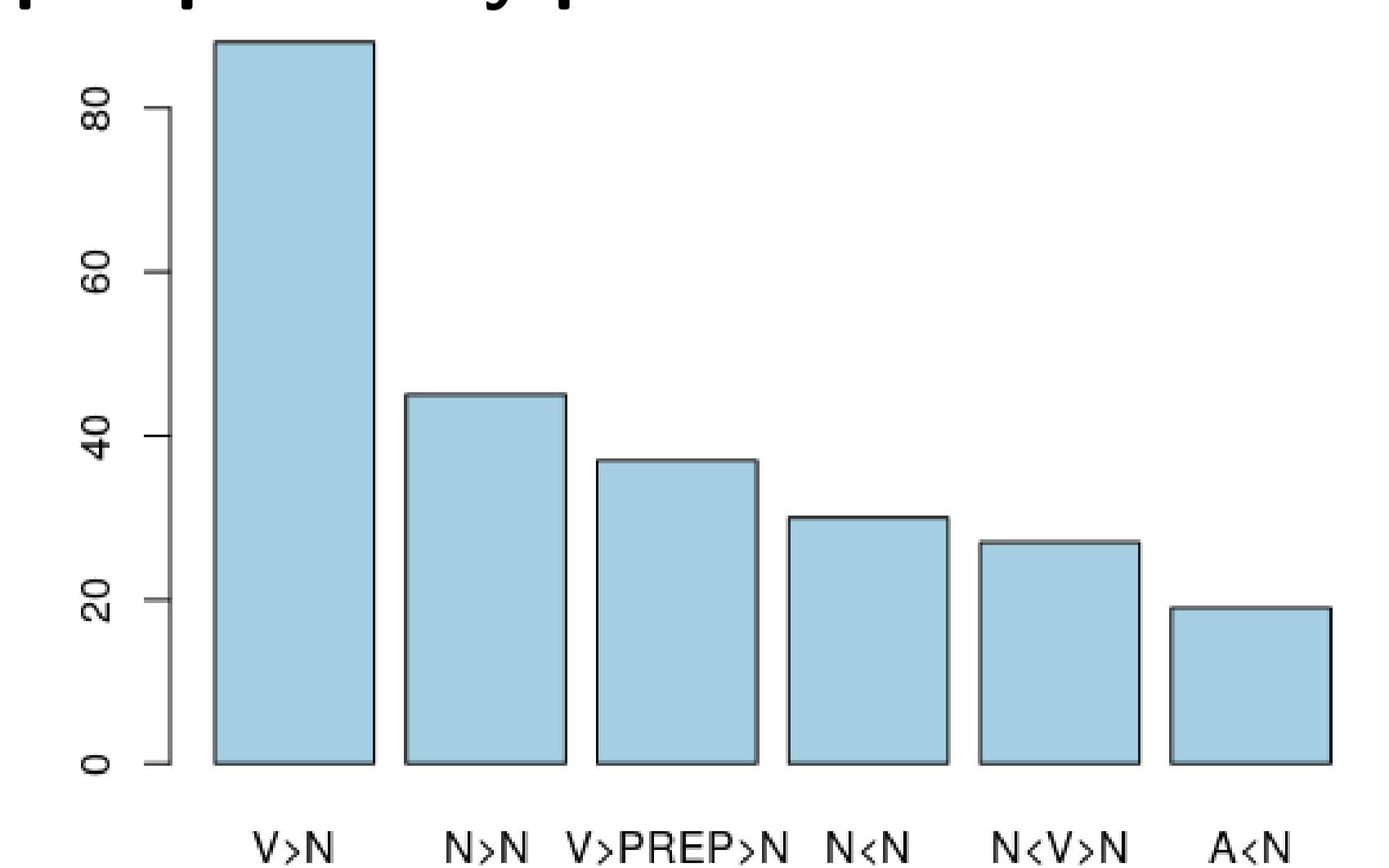
### Top functions:



Example: seed *Weg des Glaubens* 'way of faith'

- ▶ **Cluster with *Weg*:** [ grund, mitte, schliessen, feld, verlauf, weg, übergang, tief, gang, wenden, bruch, lücke, riss, pfad, hinweg, begrenzung, winden, füllen, überwinden, streichen, barriere, hebung ]
- ▶ **Cluster with *Glaube*:** [ mensch, freiheit, wille, wahrheit, seele, wesen, glauben, glaube, persönlichkeit, moral, vernunft, sünde, ehre, tugend, tapferkeit, gnade, güte, frömmigkeit, dankbarkeit ]
- ▶ The system would be able to recognize (novel) metaphors like *das Feld der Sünde* 'the field of sin', *der Bruch der Ehre* 'the breach of honor', *die Mitte der Seele* 'the center of the soul', etc.

### Top dependency paths:



Example: extracted features of *kochen* – *Silberseele*

- ▶ **POS tags:** V – N
- ▶ **Functions:** obji (inf. object) – pn (prep. compl.)
- ▶ **Dependency path:** V<V>PREP>N  
(*kochen* < *beginnt* > *in* > *Silberseele*)