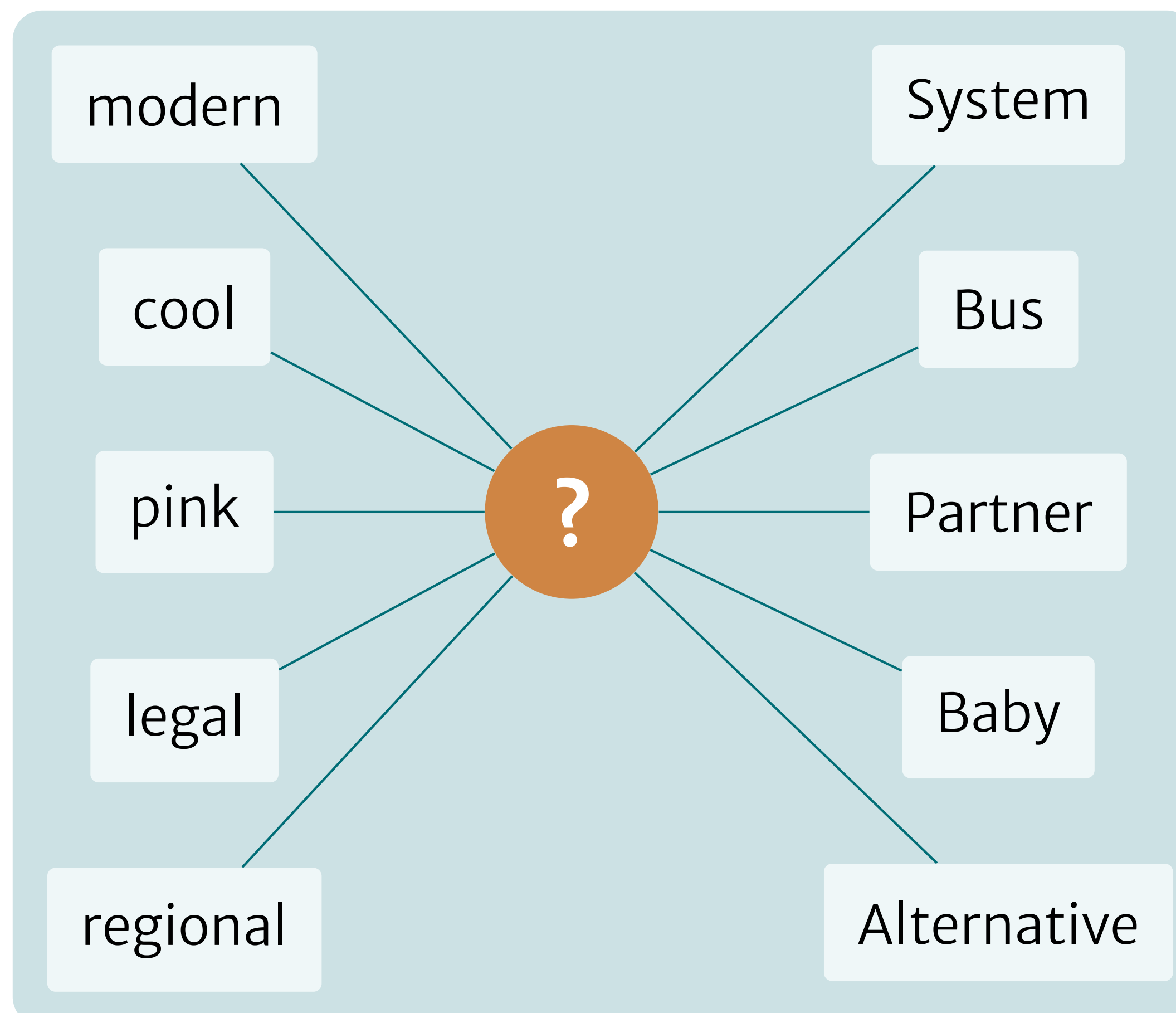


## Adjectives as modifiers

In German, adjectives can modify nouns, verbs, or other adjectives. Corpus evidence for **modifier-modifiee** patterns is useful for linguistic research in various areas, including:

- semantic compositionality
- synonymy and polysemy
- semantic clustering
- distributional properties of adjectives and adverbials
- ...

**AET** is a new web-based query tool that can be used to collect corpus evidence according to complex syntactic or morphological criteria.



## Our contribution: AET

AET is designed to provide information on modifier-modifiee pairs that goes beyond the mere surface forms found in a corpus. The types of filter criteria currently supported include:

- morphological
- syntactic
- lexical
- character-based
- statistical

Pairs are retrieved from a dependency-parsed corpus, so that modifiers and modifiees do not need to be adjacent in the text:

Der Kuchen **schmeckt**<sub>v</sub> wegen der verwendeten Zuckerart **gut**<sub>Adj</sub>.  
the cake **tastes**<sub>v</sub> due to the used sugar type **good**<sub>ADJ</sub>

About AET Query Documentation

modifier\_wordform\_ends(ende) AND  
modifiee\_derivationtype(deadjectival)

Submit

## How to use AET

- Form complex queries using the boolean operators AND, OR and NOT.
- More specific queries execute faster.
- Results are sorted by lemma pair frequency.
- Select a lemma pair to view surface forms in individual sentences.
- Export results to \*.csv to work with data offline.

first previous 1 of 5 pages. (Total Items: 470) next last

Export results

Occurrences of the Lemma Pair | Modifier Lemma | Modifiee Lemma | POS Tags

11 | gähmend | Leere | AN

10 | fehlend | Transparenz | AN

first previous 1 of 1 pages. (Total Items: 10) next last

Frequency of "fehlende Transparenz" as surface form: 6

Sentence containing this pair:

Keine ausreichende Kontrolle, **fehlende Transparenz**, über 1,3 Milliarden Euro Schulden und 9,5 Milliarden Euro auf Geheimkonten: Der Rechnungshof geht in seinem aktuellen Bericht zur finanziellen Lage Salzburgs mit dem Finanzmanagement des Landes hart ins Gericht.

## Properties of word forms

word form	vertretbaren
lemma	vertretbar
word type	adjective
derivation type	deverbal
derivation scheme	Vx
prefix	ver
suffix	bar
derivation tree	(((ver)[V].[V]),(tret)[V])[V], (bar)[A].[V].[A]
prev. derivation step	vertret+bar
composition	non-compositum
gradation	positive
frequency in AET	62
number	plural
case	dative
gender	feminine
never inflected	no
occurs sentence-final	yes

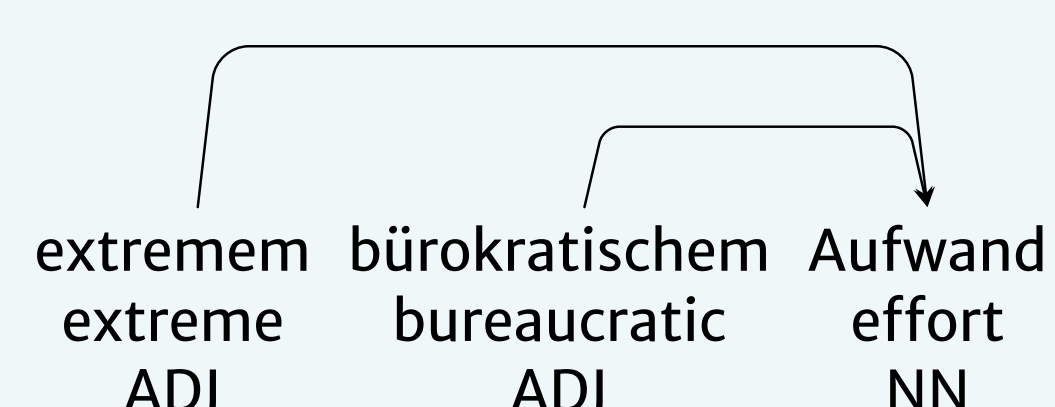
Tab. 1: Searchable DB fields in AET for word forms, here: *vertretbaren*

## AET facts

- Relational MySQL database
- Website written in CakePHP
- Search language parser written in Python
- Data source: Corpora from the Workshop on Machine Translation 2014 (WMT14) shared tasks
- Sentences parsed with MATE
- Morphological analyses added with DerivBase and CELEX
- Demo version: 1,248,869 modifier-modifiee token pairs, 616,447 type pairs, 1,250,284 sentences
- Full database: 13,274,424 token pairs, 3,738,414 type pairs, 8,206,827 sentences

## Future work

Adjectives sometimes occur in chains, where two or more adjectives modify the same token. So far, AET represents each instance of modification in such cases as an individual modifier-modifiee pair. We plan to add search terms for filters on the chain level.



AET separates functionality from data: Other languages or other corpora can be introduced with minimal effort.

The architecture can also accommodate other types of modification, such as nominal compounds or prepositions and their complements.

Additional fields can be included in the database with little effort. The search language can easily be updated or extended as necessary.

## Selection of search terms

```
modifier_lemma(blau)
modifiee_lemma(Haus)
modifier_wordform(rotem)
modifiee_wordform(Kreuzen)
modifier_lemma_starts(ab)
modifiee_lemma_starts(be)
modifier_wordform_starts(ver)
modifiee_wordform_starts(ent)
modifier_lemma_ends(lich)
modifiee_lemma_ends(ung)
modifier_wordform_ends(ende)
modifiee_wordform_ends(ungen)
modifier_pos(a)
modifiee_pos(v)
modifier_pos_stts(ADJA)
modifiee_pos_stts(VVINFL)
modifier_derivationtype(deverbal)
modifiee_derivationtype(denominal)
modifier_never_inflects(true)
modifier_always_positive(true)
pair_type_frequency_greater(100)
pair_type_frequency_lower(100)
```

## Properties of modifier-modifiee pairs

modifier word form	benötigten Mittel
modifiee word form	Mittel
modifier lemma	benötigt
modifiee lemma	Mittel
modifier POS tag	ADJA
modifiee POS tag	NN
POS category pair	AN
pair frequency	56
precedence	yes

Tab. 2: Searchable DB fields in AET for modifier-modifiee pairs, here: *benötigten Mittel*

## References

- Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). The CELEX Lexical Database.
- Björkelund, A., Bohnet, B., Hafdel, L., and Nugues, P. (2010). A High-performance Syntactic and Semantic Dependency Parser.
- Bojar, O., Buck, C., Federmann, C., Haddow, B., Koehn, P., Leveling, J., Monz, C., Pecina, P., Post, M., Saint-Amand, H., Soricut, R., Specia, L., and Tamchyna, A. (2014). Findings of the 2014 Workshop on Statistical Machine Translation.
- Zeller, B., Šnajder, J., and Padó, S. (2013). DERivBase: Inducing and Evaluating a Derivational Morphology Resource for German.

See proceedings for full bibliography.