

Advancing Analogical Theory in Morphology (AdvAnTheMorph)

Sabine Arndt-Lappe, Mirjam Ernestus, Louis ten Bosch & Katharina Schuch

Trier University, RU Nijmegen

arndtlappe@uni-trier.de

DFG project, funding period: April 2025 – March 2028

Starting point

- Paradigmatic effects: Analogy is well-known to influence
 - affix choice among competing, synonymous affixes;
 - the acoustic realisation of complex words.
- Paradigmatic effects can be operationalised by means of computational analogical models: e.g. AML (Skousen, 1989)

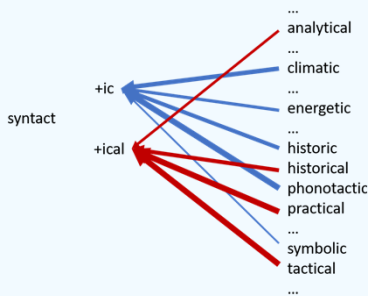


Figure 1: syntact+ic or syntact+ical? Paradigmatic effects as modelled in traditional AML. More similar words have a larger influence (as indicated by the thickness of lines)

Hypotheses to be tested in this project

If morphology is lexicon-based and is co-determined by usage,

a. People should use different affixes in different contexts: (a) linguistic context, (b) register

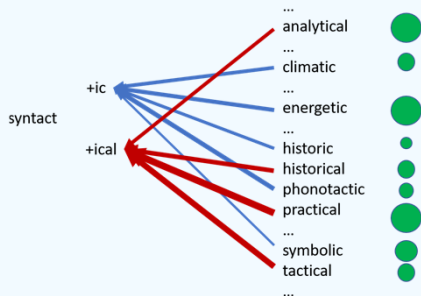


Figure 2: CAML: Context-sensitive AML. The thicker the dots, the higher the words' (context dependent) frequencies and the higher the influence on the target word. (Compare with Figure 1, where 'syntact+ic' receives more support.)

b. People should use different affixes depending on their individual vocabulary knowledge

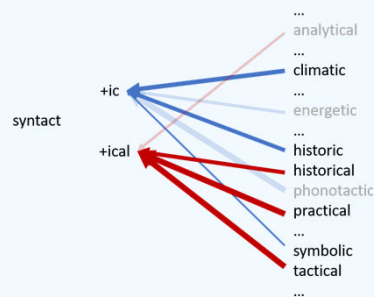


Figure 3: AML based on a different lexicon (compare with Figure 1, where 'syntact+ic' receives more support.)

Relevant preliminary work

Arndt-Lappe (2023) has shown that small differences in the contents of the lexicon lead to substantial differences in paradigmatic support.

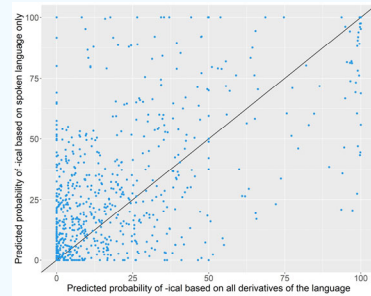


Figure 4: Predicted probability of -ical (vs. -ic) in two AML simulations, one of which is based on the full set of all -ic and -ical derivatives in the language, and one of which is based only on the derivatives that occur in the Spoken section of the British National Corpus. Correlation between the predictions of the two simulations: 0.49

Methods and goals

Developing a context-sensitive computational model (CAML, based on AML / Analogical Modeling)

- Extending the AML algorithm (Skousen, 1989) to incorporate context-dependent frequencies
- Evaluation and triangulation of CAML



Production experiments, testing for the role of

- the frequencies of analogical words
- register (formal, vs. casual)
- differences between individual speakers' lexicons, as assessed by means of questionnaires

Experimental paradigm: Participants hear a dialogue where a specific 'target' word is introduced, followed by an inflected or derived variant of that word. Part of this variant will be blurred with noise. Participants are asked to repeat the sentence, including the complete inflected/derived target word.

Example:

Audio input: Faulkner could string together four, five adjectives and make it sound real **syntax**. His command of **syntacxxx** structures pushed the language forward at least seventy-five years....

Participant repeats: His command of syntactic/syntactical/... structures

Corpus studies, testing for the role of linguistic context

- Speech corpora containing authentic, naturalistic speech

Examples:

I'm sure if you look round you can get a lot of **electrical** goodies cheap, or just as cheap in other places. (BNC, KSV 4519, spoken conversation)

Thomas Edison, the man who eventually invented the **electric** light bulb, said 'failure is essential to success ... you cannot succeed without failing ... double your failure rate if you want to succeed.' (BNC EB1 1197, academic prose)

Testbed

- Dutch regular plurals (e.g. aardappel-s, aardappel-en)
- English adjectival suffixes (e.g. microb-ical, microb-ic, microb-ial, microb-ian, microb-al)