Automatic Construction of Morpho-Syntactic Descriptions of Verbal MWEs from Crowdsourcable Input

Simon Clematide and Mark Luder
**Vision**
Let
- humans
- NLP tools (PoS taggers, morphological analyzers, dependency parsers)
- corpora and automated corpus queries
do what they can do best.

**Problem A**
*Manual* creation of exact and detailed morpho-syntactic descriptions of MWEs is tedious.

**Solution A**
Use automatic linguistic analyzers such as morphological analyzer, PoS tagger, and parsers.
Problem B 1
Traditional lexicographic entries typically use an infinitive form to represent a construction.

Problem B 2
Real-world sentences contain more than just the construction.

Solution B
A non-expert (“the crowd”) is able to simplify real-world sentences into a canonical inflected form with special tokens for fully variable constituents.

Problem C
The variability/fixedness of a construction should be recorded on the basis of empirical corpus-linguistic data.

Solution C
Corpus queries that are automatically derived from the morpho-syntactic descriptions (=“constructions”) can test for modifiability, optionality, slot filler collocations, etc.