

## Language, learning and usage-based theory: tackling nominal and verbal morphology in Slavic

Dagmar Divjak <[d.divjak@sheffield.ac.uk](mailto:d.divjak@sheffield.ac.uk)>

The University of Sheffield

Usage-based theories of language are built on the assumption that our ability to extract and entrench the distributional patterns available in the input enables learners to build a grammar from the ground up. This circumvents the needs for an innate universal grammar. But it does not tell us which patterns are relevant. And it remains customary for linguists to approach the data using linguistic categories - such as Case or Tense, Aspect and Mood - categories that were never intended to reflect the workings of the mind. In this talk, I will argue that it might be better to take the input as starting point and derive categories that resemble those native speakers might derive. Models from Learning Theory can help with this. I will present two case studies that capitalize on a merger of cognitive linguistics and cognitive psychology, and aim to infuse Usage-Based linguistics with insights from Learning Theory ... with a little help from computational engineering.

The first case study uses insights from Learning Theory to challenge the idea that theoretical linguistic constructs such as tense, aspect and mood (TAM) predict best how native speakers of Russian read sentences containing verbs meaning *to try* in real time. Discrimination learning, as implemented in the NDL algorithm, proposes simple 3-letter usage-patterns and predicts the time it takes subjects to read and integrate these verbs into a sentence significantly better than all TAM markers combined.

|             |        | <i>poprobuet (s/he will try)</i> | <i>pytajsja (you try!)</i> | <i>starajutsja (they are trying)</i>  |
|-------------|--------|----------------------------------|----------------------------|---------------------------------------|
| TAM marking | Tense  | future                           | present                    | present                               |
|             | Aspect | perfective                       | imperfective               | imperfective                          |
|             | Mood   | indicative                       | imperative                 | indicative                            |
| Trigraphs   | ----   | -bue-                            | -taj-                      | -raju- (-ju- is 1 letter in Cyrillic) |

Contrary to what mainstream (psycho)linguistic models assume, speakers do not (and do not need to) analyse verb forms in terms of abstract linguistic concepts such as tense, aspect and mood when they process language. Instead, they can rely on simple letter sequences that are linked directly to an experience and embed crucial information about that experience (i.e., is it over, ongoing, or coming up; was it something that they completed, or simply did for a while; was it an order). This demonstrates that honouring parsimony (naivety and simplicity) in the structures that are hypothesized to exist, and in the way in which behaviour is explained, is a powerful research stance, in particular for designing cognitively realistic accounts of language knowledge and representation.

The second case study demonstrates how biologically inspired machine learning techniques can pinpoint the essence of native speaker intuitions. Polish boasts fascinating examples of seemingly unmotivated allomorphy, and the genitive singular of masculine inanimate nouns (which can be *-a* or *-u*) is its prime example. Criteria for choice have been proposed that are semantic, morphological or phonological in nature, but most of these are unreliable, yielding conflicting predictions (Dąbrowska 2005). Furthermore, although *-u* occurs with at least twice as many nouns, *-a* is the default ending for new words entering the language. The NDL algorithm, that implements discrimination learning, predicts the choice between *-a* and *-u* better using simple sequences of 3 letters (letter triplets or trigraphs) than models running on richly annotated corpus data. In addition, it explains the unexpected preference of *-a* as genitive ending for new words in terms of the learnability of words taking the *-a* ending, their phonological predictability and their contextual (semantic) typicality.

On their own, linguists and psychologists would have approached these questions rather differently and, from within their disciplinary cages, would have arrived at answers that would necessarily have remained partial. Integrative interdisciplinarity, on the other hand, relies on a simultaneous, interspersed methodological endeavour to arrive at more encompassing answers that combine depth of analysis with breadth of explanation. It presupposes mutually complementary theories, shared testable hypotheses as well as compatibility of research methodologies. But what wins the game is a good dose of willingness to question your customary ways of doing things.