General theme: How does the CLARIN infrastructure contribute to making micro and macro-comparative linguistic research better (esp. broadening the empirical basis) while keeping it doable for a linguist and perhaps even bring it to a new level.

Concrete example: adjectives in many Germanic languages (Dutch, German, Swedish) have other inflection in definite NPs than in indefinite NPs. Some research questions that come to mind are:

* How should these phenomena be described?
* Can these descriptions be related to general principles of language?
* What accounts for the fact that some languages have these phenomena (e.g. the Germanic languages mentioned) but others not (most if not all Romance languages)
* What about dialectical variation within each of these languages?
* Are there other languages from completely different language families with similar phenomena?
* In Dutch, there are several exceptions to the main rule of adjectival inflection in definite NPs, some of them quite systematic (e.g. het bijvoeglijk naamwoord, het medisch onderzoek, de medisch onderzoeker) .
  + How should these exceptions be accounted for?
  + Do German and Swedish also have exceptions to the main rule, and are they similar in kind or completely different?

A researcher wants investigate these matters with a solid empirical basis. She therefore wants to search for relevant occurrences of these phenomena in large richly annotated corpora and other linguistic resources. But it must be possible to do this in a very user friendly way:

* Search in large PoS-tagged corpora e.g. CGN, SONAR, DPC and corresponding corpora for other languages e.g for r.e.s over PoS-tags e.g. def.det – adjective (infltype) + - noun
* Search in large fully parsed corpora, e.g.
  + LASSY-Small
    - DACT
    - ETKS interface
    - GreTEL interface
  + LASSY Large
  + Similarly for German and Scandinavian languages (maybe some researchers from there can help us, linguists and/or people active in CLARIN )
* Taalportaal should set most of the common queries ready!
* Maybe search in Cornetto / local Wordnets for adjectives and semantically related words (synonyms, antonyms, hyponyms, co-hyponyms) to assist in determining test whether certain phenomena are related to / driven by semantic properties of adjectives.
* Typological databases such as TDS, WALS, etc; in multiple micro-comparative databases via MIMORE (or perhaps even via the generic content search facilities)
* Historical perspective, historical data, relation with *Nederlab*
* Maybe postag, parse some unparsed corpora using web services offered in the CLARIN inmfrastructure. (TTNWW project) . Make these results available again in the CLARIN infrastructure – facilities to do this, to generate metadata in a maximally automated manner, provenance data, etc.