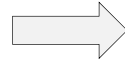
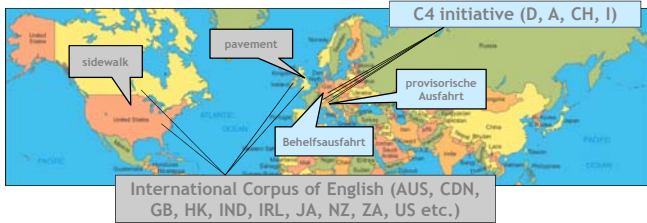


a System for the Comparison of Linguistic Varieties on the Basis of Corpora

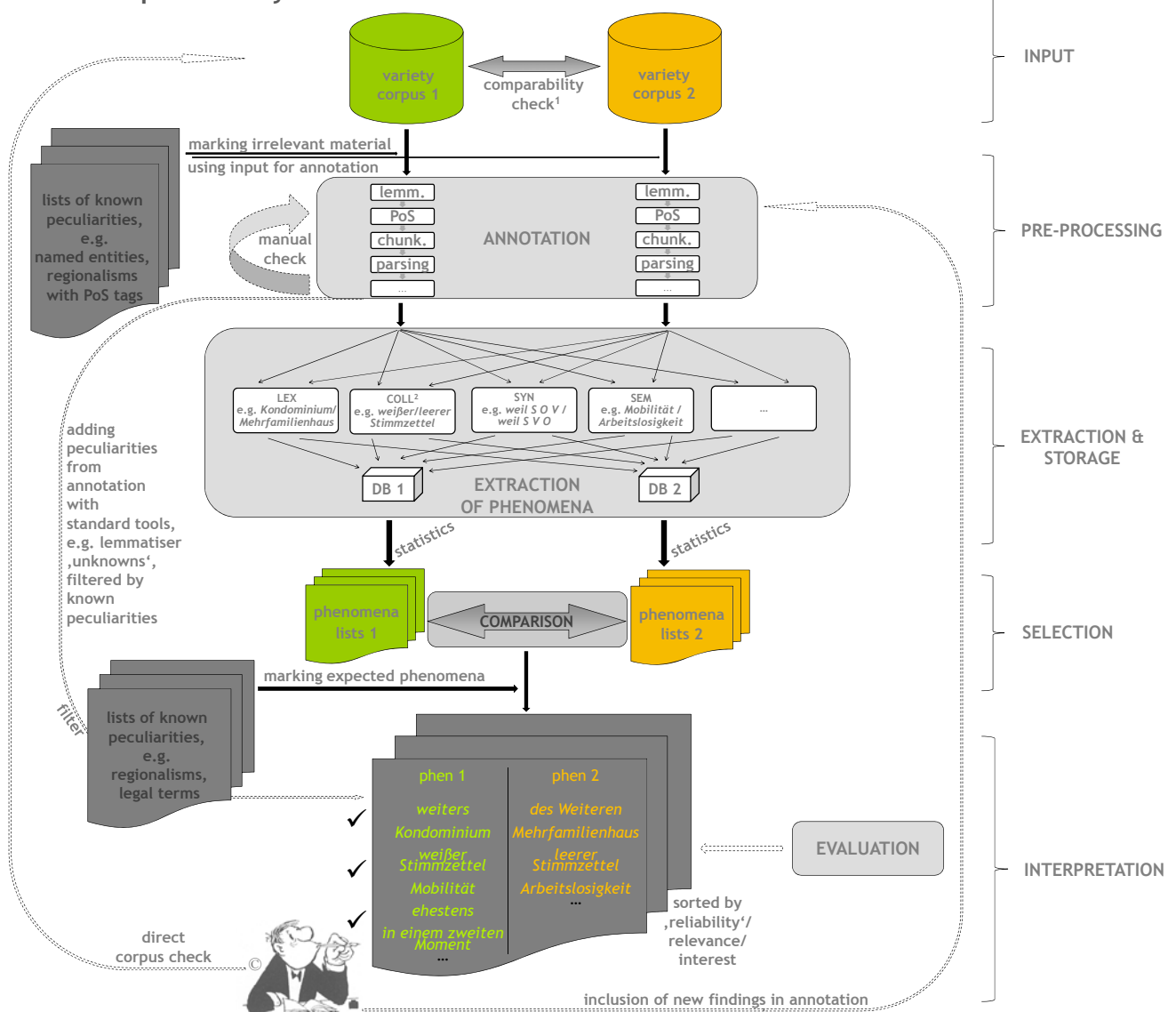
Stefanie Anstein - stefanie.anstein@eurac.edu

1. Varieties with different NLP coverage



- corpus comparison for
- variety documentation
 - variant lexicography
 - standardisation of lesser-used varieties
 - language teaching and learning
 - etc.

2. VIS-À-VIS preliminary architecture



3. Future Work

- implement detailed workflow
- enhance framework
- integrate established and original tools
- collect expected phenomena from single studies on South Tyrolean German
- test system for German varieties on the basis of C4 corpora

¹ Gries, S. T. (2007a): Exploring variability within and between corpora: Some methodological considerations. *Corpora*, 109-151.

Kilgarriff, A. (2001): Comparing corpora. *International Journal of Corpus Linguistics*, 6(1), 1-37.

² e.g. Ritz, J. & Heid, U. (2006): Extraction tools for collocations and their morphosyntactic specificities. In *Proceedings of LREC, 1925-1930*.