

**Title: Semi-Automatic Online Glossing**  
**Category: TPI poster presentation**

Interlinear morphemic glosses (see *The Leipzig Glossing Rules*<sup>1</sup>) provide the reader with important information on the meanings and grammatical features of words or parts of words and therefore contribute to a better understanding of the structure of sentences in less well-known languages. Generating interlinear glosses for longer passages of text, however, is a laborious task. Thus, for researchers working with original texts in large scale typological studies, it would be convenient to have a glossing tool at hand which helps to (semi-) automatically get information on the morphological structure of the words by suggesting potential glossings, from which they can then choose the appropriate candidate in view of the translation.

This paper describes the prototype of an online glossing tool which tries to fill this gap. The tool makes use of the finite-state command-line application `xfst` (Beesley & Karttunen, 2003) and the PHP scripting language. Finite-state morphology assumes that both morphotactics and the morpho-phonological alternation rules can be modeled as finite-state transducers. The advantage of finite-state transducers is that they are bidirectional, i.e., the same transducer can be used both for the analysis and generation of words, differing only with respect to the choice of the input side. This means that morpho-phonological rules as well as the translations of morphological categories and lexical material have to be stated only once in the grammar.

The glossing tool can be subdivided into an analysis part and a generation part. During the analysis of the input sentence(s) all possible morphological structures for each word are looked up in the morphology. Since the system only works on the level of words and does not take context information into account, the user can choose that respective gloss from a drop down list of possible morphological analyses (see Figure 1) which seems to be most appropriate for the given context.

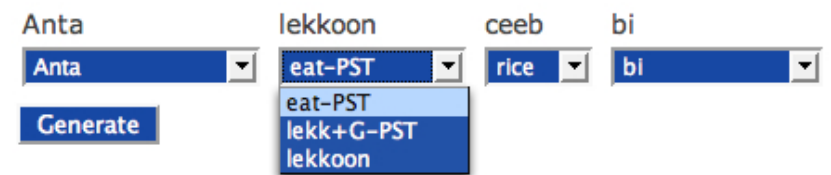


Figure 1: Drop down list of possible morphological analyses for each word

Depending on the user's choices the corresponding segmentations of the words are generated from the gloss by applying the rules backwards. The result is then displayed in the typical format with glosses being left-aligned vertically, word by word, with the example:

<sup>1</sup><http://www.eva.mpg.de/lingua/resources/glossing-rules.php>

Anta	lekk-oon	ceeb	b-i
Anta	eat-PST	rice	CLASS.B-PROX

Figure 2: Semi-automatically generated interlinear morphemic glosses

The tool is exemplified from the Niger-Congo language Wolof, for which a finite-state morphology has been written and implemented. Its morphology is exclusively suffixing, making occasional reference to phonological context information and therefore nicely illustrates the opportunities of the tool.

## References

Beesley, Kenneth R., and Lauri Karttunen. 2003. *Finite State Morphology*. CSLI Studies in Computational Linguistics. Stanford, CA: CSLI Publications.