

Recent Enhancements in the Bonito Interface

Vojtěch Kovář

Lexical Computing

`{vojtech.kovar}@sketchengine.co.uk`

SKEW 2, March 17 2011

Outline

- 1 Small additions to the interface
- 2 New wordlist form
- 3 Sketch diff enhancements

Small additions to the interface

- Character search field
 - for languages where tokenisation is hard
 - X translates to $[word = ". * X. *"]$
- Original concordance
 - link in the left panel
- *FIXORDER directive in sketch grammars
 - finally works
- Hierarchical header fields
 - hierarchy of the text type values
- Subcorpus hypotheses testing
 - indication of a subcorpus that is typical for the headword

New wordlist form

- One form instead of 4 on one page
 - more combinations – filtering, subcorpora, ...
 - non-words filtering
 - blacklist and whitelist filtering
 - text types listing with subcorpora
 - document counts as well as token counts and ARF
 - keyword lists on any pair of (sub)corpus and reference (sub)corpus
- Three parts of the form
 - what we search for
 - filtering options
 - output format

Sketch diff enhancements

■ Sketch diff by subcorpora

- word sketches for a word in a subcorpus
- compared with sketches for the same word in another subcorpus

■ Sketch diff by word form

- contrast of sketches filtered for two forms of the same lemma

■ Sketch diff by version of the grammar

- for grammar development purposes
- in the near future