

Werner Wegstein

Baseline Encoding for a Parallel Corpus: A Proposal.

1. Motivation

2. Previous Work

3. Sample Text

4. Actions

2. Previous Work

Laurent Romary / Nathalie Mehl / David Woolls. The Lingua Parallel Concordancing Project: Managing Multilingual Texts for Educational Purpose, in: Text Technology 5, 3 (1994) 206-220.

Jean Veronis (Ed.), Parallel Text Processing, Dordrecht 2000..

Stig Johansson / Jarle Ebeling / Signe Oksefjell: English-Norwegian Parallel Corpus. Manual. Department of British and American Studies. University of Oslo, 1999/2002.

<http://www.hf.uio.no/ilos/english/services/omc/enpc/ENPCmanual.pdf>

von Waldenfels: Regensburger Korpusinitiative Parallel Corpus (RPC)

TextGrid's Baseline Encoding for Text Data in TEI P5, June 2007

<http://http://www.textgrid.de/fileadmin/TextGrid/reports/baseline-all-en.pdf>

Werner Wegstein: Vorüberlegungen zu einem parallelen polnisch-deutschen Textkorpus, in: Waldemar Czachur / Marta Czyżewska, Vom Wort zum Text. Studien zur deutschen Sprache und Kultur. Festschrift für Professor Józef Wiktorowicz, Warszawa 2008, 145 – 152.

3. Sample Text (with sentence alignment)

Günter Grass:

„Unkenrufe“ (1999)

- (1) Der Zufall stellte den Witwer neben die Witwe.
- (2) Oder spielte kein Zufall mit, weil ihre Geschichte auf Allerseelen begann?
- (3) Jedenfalls war die Witwe schon zur Stelle, als der Witwer anstieß, stolperte, doch nicht zu Fall kam.
- (4) Er stellte sich neben sie.
- (5) Schuhgröße dreiundvierzig neben Schuhgröße siebenunddreißig. ...

„Wróżby kumaka“ (Grass 1999)

- (1) Przypadek postawił wdowca obok wdowy.
- (2) A może przypadek nie miał nic do rzeczy, ponieważ ich historia zaczęła się w Zaduszki?
- (3) W każdym razie wdowa była już na miejscu, kiedy wdowiec potknął się, stracił na chwilę równowagę, ale nie upadł.
- (4) Stał obok niej.
- (5) Numer butów czterdzieści trzy obok numeru butów trzydzieści siedem. ...

4. Actions