

Automatic alignment and aspects of using parallel corpora / Àâòîàòè÷ãñêîå âûðåâîèåâàîéå è ïàðåëëåâüüíûõ êîòîñíîñîå

Àâîîð Hanne Martine Eckhoff
04.10.2009 ä.
Iñēāāíāā íāñāēāíēå 28.10.2009 ä.

Ìàòåðèååü è äåòèè (Iðçåíòòèé)

For the study of languages such as OCS, that are mainly extant in translation, an aligned parallel corpus is invaluable. In the PROIEL corpus, all the translations are automatically aligned with the Greek original at token level, with a success rate of about 97 %. In this lecture we discuss the automatic token aligner and demonstrate how the token alignments can be used in combination with multiple layers of annotation to do sophisticated contrastive work on the translation languages.