

HW 3

Languages: English & Ukrainian

Dataset: GNOME v1 from OPUS - a parallel corpus of localization files

Alignment: sentence level (phrases for application buttons/menus)

Number of word length mismatches: 208 of 619

Number of critical mismatches (difference above 1): 31

Reasons for the word length difference:

- “-” separated words in English are translated into separate words in Ukrainian
- Notes for alternative translation in square brackets for Ukrainian
- Longer translation of specific terms/phrases without short equivalents in Ukrainian
- English prepositions translated into morphologically modified (inflection) Ukrainian phrases
- Dataset formatting special characters like “%s” counted as words
- Translations providing more/less specific terms