

# Extraction of Multilingual MWEs from Aligned Corpora

Eric Wehrli<sup>1</sup> and Aline Villavicencio<sup>2</sup>

<sup>1</sup>LATL-CUI, University of Geneva

<sup>2</sup>Institute of Informatics, Federal University of Rio Grande do Sul

- Multilingual repository of MWEs extracted from aligned corpora

- Languages:

- English, French and Portuguese

- MWE types:

- compounds, idioms, collocations

- Mono and multilingual methods:

- symbolic parser: FIPS

- pre-defined structural configurations and log-likelihood

- statistical methods: mwetoolkit

- pre-defined POS patterns and combination of association measures

