

## Shuly Wintner:

### Automatic Identification of Translationese: Highlighting the nature of translation

Translated texts differ from texts originally written in the same (target) language. Several Translation Studies hypotheses aim at explaining these differences. We use computational methodology, specifically supervised and unsupervised classification, to distinguish between translated and original texts. This facilitates a close inspection of the specific features along which the two types of texts differ.

This enterprise yields several findings:

1. Some Translation Studies hypotheses, especially those purporting to the universality of translationese features, are questionable;
2. Interference, namely the 'fingerprints' of the original text on the product of the translation process, is by far the dominating feature of translationese;
3. Interference is so powerful that by looking only at translations from several languages, the source language can be identified;
4. Translationese features are overshadowed by more salient features of the text, including genre, register, domain, etc.

We show that the import of these results is not only theoretical; they have implications for natural language processing applications, in particular statistical machine translation.