

Proposer: John Lee, 504420, J.Lee@ed.ac.uk

Self-Proposed: No

Supervisor: Frank Keller, 504407, keller@inf.ed.ac.uk

Subject Areas: Natural Language Processing,

Suitable for the following degrees: BSc in Artificial Intelligence and Computer Science, BSc in Artificial Intelligence and Psychology, MA in Linguistics and Artificial Intelligence,

Principal goal of the project: Investigate automated comparison of raw machine translation output with post-edited output, to develop improved translations

Description of the project:

A local company, **Verbalis**, carries out German to English machine translation for a German client. The raw output is post-edited by another company engaged by the client. A substantial parallel corpus now exists of the raw and post-edited translations (and the original German). The aim of the project is to investigate methods of comparing raw and post-edited translations, to derive guidance for improving the raw translations.

The problem of comparing raw and post-edited translations can be couched as a paraphrasing task. A substantial literature on automatic paraphrasing exists, which this project can rely on (e.g., Barzilay and Lee 2003; see also ACL 2003 Paraphrasing Workshop). Also the literature on statistical machine translation is relevant (starting with Brown et al. 1993; see also NIST MT activity). Existing software tools will be usable or adaptable for much of this activity.

Resources Required: A range of natural language evaluation software

Degree of Difficulty: medium to hard

Background Needed: Experience of natural language software; ideally an interest in machine translation

References:

ACL 2003 Paraphrasing Workshop: <http://nlp.nagaokaut.ac.jp/IWP2003/>

Barzilay, Regina and Lillian Lee (2003). Learning to Paraphrase: An Unsupervised Approach Using Multiple-Sequence Alignment, Proceedings of HLT-NAACL 2003, pp. 16-23.

Brown, P., S. Della Pietra, V. Della Pietra, and R. Mercer (1993). The Mathematics of Statistical Machine Translation: Parameter Estimation, Computational Linguistics, 19(2).

NIST MT activity: <http://www.nist.gov/speech/tests/mt/index.htm>