# Approaches to Machine Translation: Rule-based, Statistical and Hybrid

*Scientific Assignment*

# Selection of papers


- *Forcada, M.L.,Tyers, F.M. & Ramírez, G.* **The Apertium Machine Translation Platform: five years on.** First International Workshop on Free/Open-Source Rule-Based Machine Translation, Alacant, Spain 2009.
- *Carbonell, J. Klein, St., Miller, D. Steinbaum, M., Grassiany, T. and Frey, J*. (2006) **Context-based Machine Translation**, Proc, 7th ACL in the Americas, Cambridge, Ms
- *Mariño, J.B., Banchs, R.E., Crego, J.M., de Gispert, A., Lambert, P., Fonollosa, J.A.R. And Costa-jussà, M.R*. **Bilingual N-gram Statistical Machine Translation** Proc. of the 10th Machine Translation Summit (MTsummit'05), pages 275-82. Pukhet (Thailand), Sep 2005.
- *Costa-jussà, M.R. and Fonollosa, J.A.R*. **Statistical Machine Reordering** Empirical Methods in Natural Language Processing, pages 70-76 (EMNLP06), Sydney, July 2006.
- *Koehn, P. and Hoang, H.* (2007) **Factored Translation Models**, Proc of EMNLP, Prague
- *Thumair G*. (2009) **Comparing different architectures of hybrid machine translation** Proc. Of Mt-Summit XII, Canada
- *Formiga, L., Hernández, A., Mariño, J.B., Monte E.* (2012). **Improving English to Spanish out-of-domain translations by morphology generalization and generation.** A: Monolingual Machine Translation Workshop. "Proceedings of the Monolingual Machine Translation Workshop". San Diego: 2012, p. 6-16.
- *Giménez, J. and Márquez, Ll.* (2010) **Asiya: An Open Toolkit for Automatic Machine Translation (Meta-)Evaluation** The Prague Bulletin of Mathematical Linguistics, No. 94.

# Task to do

- FOR ONE PAPER AMONG THE LIST, EXTRACT THE FOLLOWING INFORMATION (based solely on the info in the paper)

  - Main problem (2 lines)
  - Related state-of-the-art (5 lines)
  - Main contribution (5 lines)
  - Proposals to extend the work (3 lines)

# Evaluation Criteria

- (1) Adjustment to the content of the paper
- (2) Quality of the explanation

*Both (1-10) 10 being the top mark*

# Example

- *Banchs, R. E. and Costa-jussà M.R.* **A Semantic Feature for Statistical Machine Translation** *ACL HLT: 5th Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-5), June, Portland*

# Main problem

The main problem faced in the paper is about deciding the proper translation for a particular word (with multiple related or unrelated meanings) using contextual information from the source.

# Related state-of-the-art

- Carpuat & Wu (2007) introduce word sense disambiguation techniques into statistical machine translation; Carpuat & Wu (2008) dynamically build context dependent phrasal translation lexicons; Haque et al. (2009) built different syntactic and lexical features for incorporating information about the neighboring words; and España-Bonet et al. (2009) train local classifiers using linguistic and context information to translate a phrase.

# Main Contribution

The main contribution of the paper is an **approach based on statistical knowledg**e. This approach can take into account source context information and it is able to catch semantics from lexical words by using reduced representations. The idea is to introduce a **feature function** that is able to benefit the use of the translation units that have been extracted from the **most similar training sentence given the test sentence**. Similarity is computed by means of the **cosine distance** in the reduced space of **latent semantic analysis**.

# Proposals to extend the work

Ideas for further work can be related (1) to reduce the computational cost in time by better integrating the dynamic feature and (2) to extract further knowledge of the context by using other semantic representations by using neural network