Elements of Language Processing and Learning

Khalil Sima'an and Gideon Maillette de Buy Wenniger
Institute for Logic, Language and Computation (ILLC)
University of Amsterdam, Amsterdam, The Netherlands
k dot simaan the-at-sign uva dot nl

December 11, 2012

1 Project II: Statistical Machine Translation

1.1 Introduction

In this project you will write a *review article* of a selected set of articles each describing a different approach to Statistical Machine Translation.

1.2 AIMS and DIMS

The review should have three aims (in the sequel referred to as AIMS):

- Explain how these models differ from each other. Analyze and compare these approaches along different dimensions (later referred to as DIMS):
  - what does the training data contain?
  - what are the units extracted from the parallel corpus for the translation model?
  - what probability model is used for the translation model?
  - how is the model trained (estimated/tuned) on the training data?
  - what is the (formal apparatus for) the *reordering model* used?
- Explain how these models depend on each other in the articles you have read.
- Point at the possible merits of each approach relative to the other approaches: you can show examples where the one will do better than the other.

1.3 Articles

The following articles are on the menu:

Word-based "The IBM gang". The mathematics of statistical machine translation: parameter estimation. Brown, Della Pietra, Della Pietra and Mercer. DOWNLOAD *study only model 1 and model 2*!!!

Phrase-based P. Koehn, F.J. Och, and D. Marcu (2003). Statistical phrase based translation. In Proceedings of the HLT/NAACL 2003. Download from DOWNLOAD

Hierarchical David Chiang. A hierarchical phrase-based model for statistical machine translation. 2005. In Proc. ACL, pages 263270. DOWNLOAD

1.4 Article structure and tasks

Write a scientific review article (max. 2000 words) about *at least two* of these three models with the following structure:

Abstract Summarize what your article is about (max 200 words).

Background Describe the two models of your choice (max 600 words in total).

Comparison Compare the two models as in AIMS (see Introduction) along the dimensions DIMS listed in the Introduction (max. 600 words).

Analysis What assumptions do these models make and what is your own analysis of the strength/validity of these assumptions relative to one another? Try to hypothesize the kind of phenomena where the one model might translate better than the other? (max. 400 words)

Conclusion Conclusion and a bird-eye's view of the findings (max 200 words).

1.5 Guidelines and Deadline

- You will start reading and working on this project **before the lecture of 6 Dec**. This means that there will be overlap with the time given for the first project. This is because we gave you *extra time* to finish Project I!
- To make it more fun, you will work in groups of **three students**. Students who worked together in Project I **cannot** work together in Project II.
- Submit your report by: **23 December 2012 night.**