Document-Level Decoding in Moses

Nicola Bertoldi
Robert Grabowski
Liane Guillou
Michal Novak
Sorin Slavescu
Jose de Souza

MT Marathon 2012

Wednesday 5th September, 2012
Aim

- Produce a "generic" document-level decoding framework
- Apply to a problem, e.g. lexical consistency, anaphora
- Evaluate
Problem

Inter Sentence:
- S1: The ball is red.
- S2: The boy plays with the ball / it.

Intra Sentence:
- S1: The boy has a red ball and the girl has a blue ball.

In Italian, ball is translated as "palla" or "pallone"
Progress

- Discussion of problem, previous work and evaluation
- Design of approach
- Coding to begin soon...
Approach 1 - Cache-Based Model

- Translate sentence by sentence using a Moses phrase-based system
- Cache phrases/unigrams from N-best list of previously translated sentences
- Filter cache to include only content words
- At decoding, if phrase is found in the cache, add reward using LM or TM
- Applicable to LM or TM
Approach 2 - Dependency Links

- Identify dependencies in the source (words/phrases)
- Translate sentence by sentence using a Moses phrase-based system
- Extract features from target translation
- Add penalties / rewards based on feature(s) to influence translation of the second item in dependency link
- Intra sentence dependencies more complex case - may need to impose a translation order e.g. for anaphora
Evaluation: Changes may be small and not reflected by automated metrics, e.g. BLEU