Can the Internet help improve Machine Translation?

Ariadna Font Llitjós
aria@cs.cmu.edu

Introducing a largely automated method that uses online post-editor feedback to automatically improve translation rules. Our approach attacks the problem at its core and uses local fixes to detect incorrect grammar rules that need to be refined and lexical entries that need to be added, thus generalizing beyond post-editing feedback.

1. Correction Instance Extraction
SL: Gaudí was a great artist
TL: Gaudi era un artista grande

2. Finding Triggering Features
The lexical entries for gran and grande are identical
δ(error word, corrected word) = δ(grande, gran) = ∅

3. Blame Assignment
Need to postulate a new feature, feat1, say

4. Rule Refinement

5. Output of Refined System