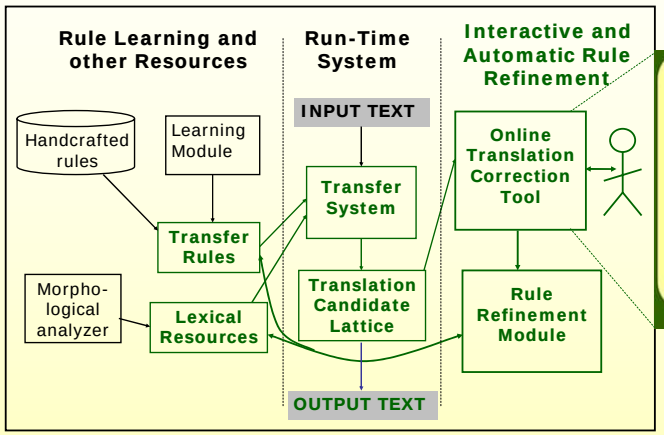
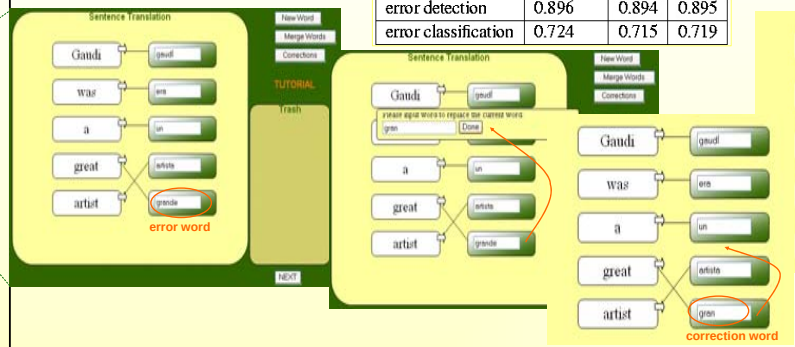
Can the Internet help improve Machine Translation?

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.





	precision	recall	F1
error detection	0.896	0.894	0.895
error classification	0.724	0.715	0.719

1. Correction Instance Extraction

SL: *Gaudí was a great artist*
TL: *Gaudí era un artista grande*
AL: ((1,1),(2,2),(3,3),(4,5),(5,4))

C_Action 1: edit (*grande→ gran*)

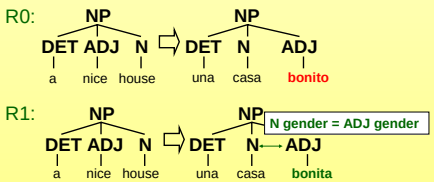
CTL: *Gaudi era un artista gran*

C_Action 2: change word order (*gran artista*)

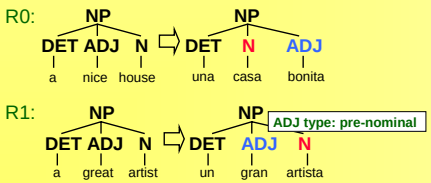
CTL: *Gaudí era un gran artista*
AL: ((1,1),(2,2),(3,3),(4,4),(5,5))

Types of Refinement Operations

REFINE a translation rule:
R0 → R1 (change R0 to make it more specific or more general)



BIFURCATE a translation rule:
R0 → R0 (same, general rule)
→ R1 (add a new more specific rule)

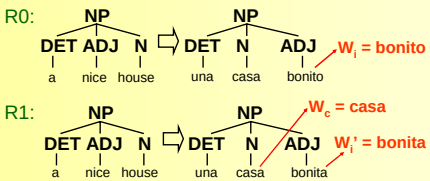


Formalizing Error Information

W_i = error
W_i' = correction
W_c = clue word

$$\overrightarrow{TL_m}=(W_1,...W_i...,W_n)$$

$$\overrightarrow{TL'_m}=(W_1,...,W_i',...,W_c,...,W'_n)$$



Comparison at the feature level to detect triggering features

Delta function: $\delta(W_i,W_i')$

Examples: gen = masc, gen = fem
δ(bonito,bonita) = {gender}
δ(comíamos,comía) = {person,number}
gen = fem, pers = 3, num = sg
δ(mujer,guitarra) = {∅}

If δ set is empty, this means that the feature language is not expressive enough
→ postulate a new binary feature (*feat_0*, *feat_1*, etc.)

2. Finding Triggering Features

The lexical entries for *gran* and *grande* are identical

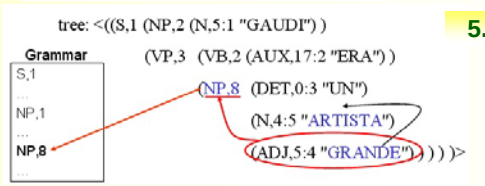
δ(error word, corrected word) = δ(*grande*, *gran*) = ∅

Need to postulate a new feature, **feat1**, say

```
ADJ::ADJ |: [great] -> [grande]
((X1::Y1)
((x0 form) = great)
((y0 agr num) = sg)
((y0 agr gen) = masc)
((y0 feat1) = -))

ADJ::ADJ |: [great] -> [gran]
((X1::Y1)
((x0 form) = great)
((y0 agr num) = sg)
((y0 agr gen) = masc)
((y0 feat1) = +))
```

3. Blame Assignment



4. Rule Refinement

5. Output of Refined System

Gaudi era un artista grande
Gaudi era un gran artista
**Gaudi era un grande artista*