Correlation of Translation Phenomena and Fidelity Measures

John White, Monika Forner
2.2.1.2 Accuracy

- 2.2.1.2.1 fidelity in text as a whole
- 2.2.1.2.2. Accuracy on indiv. sentence level – syntax – no valid measurements
- 2.2.1.2.3 Types of errors – syntax – no valid measurements
Procedure

- Sample DARPA scores
  - F-E, S-E
  - Every 20th text sorted by adequacy (approx. 35 ea.)
  - 4 worst, 4 middle, 4 best from those

- Develop list of translation issues
  - From general contrastive F/S – English
  - From observed translation glitches
  - Focused on 1 phenomenon: noun compounds
Noun Compounds

N1 de N2 de N3

N3 N2 N1
Or
N1 of N3 N2
Etc, etc.

... and don’t forget modifiers – N1 de N2 adj2 adj1 etc.
French – English results

French CMP NP by adequacy

Text ID

ade
n comp
Spanish – English Results

Spanish CMP-NP by adequacy

Text ID

0 0.2 0.4 0.6 0.8 1 1.2

3010P 3013P 3009PA 3067L 3044P 3034SY 3064NL 3046PA 3080NL 3034WG 3063NL 3099NL

Text ID

ade
ncomp
Issues / Next Steps

Good, but bad, compound handling
Sometimes English is more forgiving of Romance WO
How possible is it to automate n-comp scoring?
Lexical phenomena -- are the compounds idiomatic & in the dictionary?

Next Steps
- same exercise for larger sample
- same exercise with other potential indicators (adj-noun, concord, etc.)