Surface Realisation Using Full Delexicalisation

Anastasia Shimorina[α] **Claire Gardent**[ϕ]

LORIA[αϕ] University of Lorraine[α] CNRS[ϕ]

Surface Realisation

Surface realisation is a task of generating sentences from meaning representation structures.

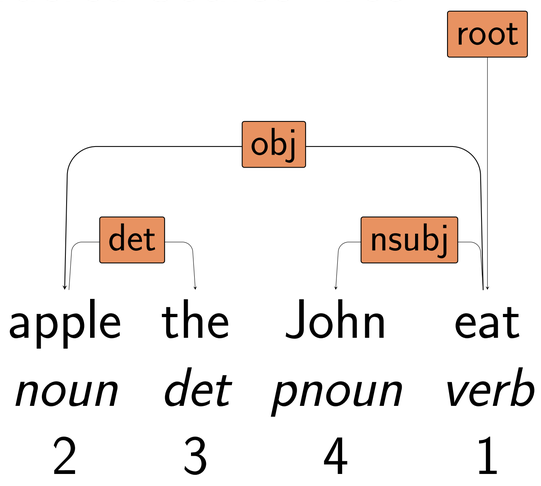
Data: shallow track of the SR'19 shared task, i.e., UD structures with word order removed and tokens lemmatised.

11 languages: Arabic (ar), English (en), Spanish (es), French (fr), Hindi (hi), Indonesian (id), Japanese (ja), Korean (ko), Portuguese (pt), Russian (ru), Chinese (zh).

Pipeline Approach

Word Ordering (WO)

Unordered Source Tree



Factored biLSTM model

Input:

2:noun:obj:1 3:det:det:2

4:pnoun:nsubj:1 1:verb:root:0

Output: 4 1 3 2

Morphological Realisation (MR)

Relexicalise and inflect:

4 1 3 2

↓

<John; N, sing> <eat; V, pres>

<the; det> <apple; N, sing>

↓

John eats the apple

- Character-based neural morphological inflection
- dictionary-based MR: ja, ko, zh

Contraction Generation (C)

Two approaches:

* rule-based
* neural

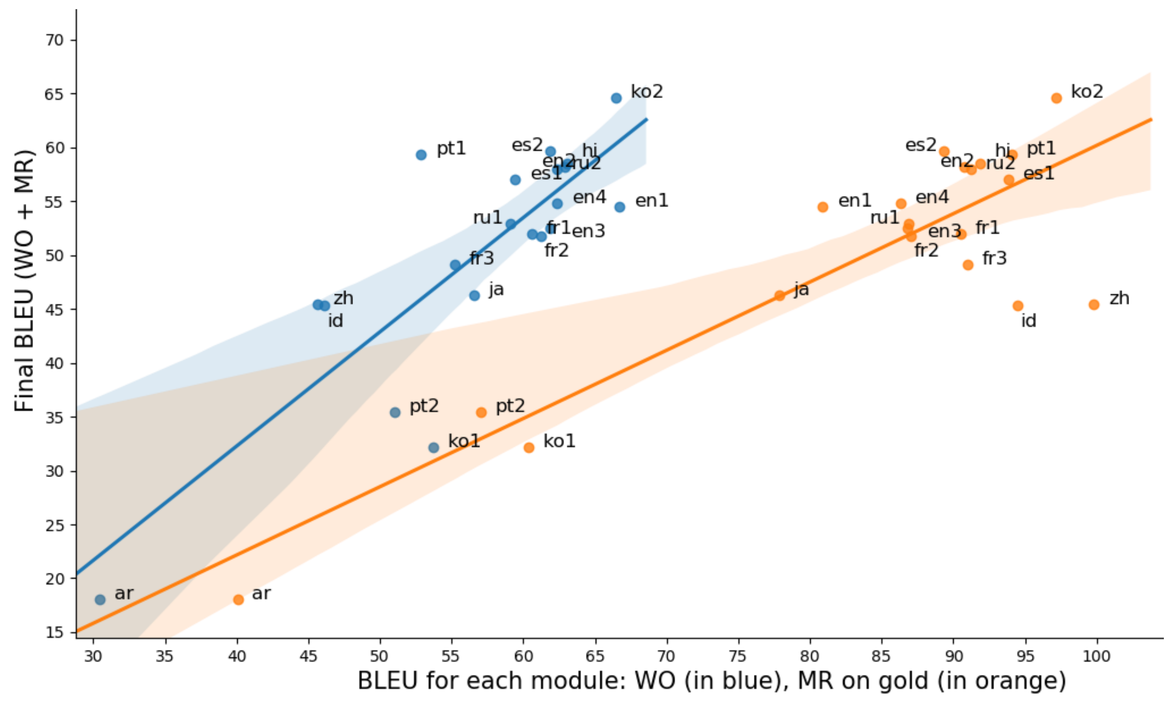
Examples:

- fr: *Le chat* / *L'alouette chante*
  Elision: *Le* → *L'*
- pt: **Eis lo* → *Ei-lo*
  Clitic attachment: *Eis lo* → *Ei-lo*

Results

– High accuracies of MR do not guarantee good performance while evaluating in context.

	Acc.	Amb. %	Amb. count
ar	90.87	7.29	1,815
en	96.35	0.84	226
es	98.85	0.85	418
fr	98.40	1.48	430
hi	89.95	6.46	569
id	98.52	0.55	47
ja	NA	3.62	800
ko	NA	0.86	945
pt	98.95	0.85	233
ru	97.25	0.72	933
zh	NA	0	0



– [Modular] evaluation is important.

- test WO against lemma sequence, rather than references
- morphological inflection in context ≠ morphological inflection accuracy
- evaluate MR on gold sequence of lemmas