

Empirical lower bounds on translation unit error rate for the full class of inversion transduction grammars

Anders Søgaard and Dekai Wu

The 11th International Conference on Parsing Technology (IWPT 2009)

October 7 2009

Alignments and normal forms

Related work

Configurations

Data

Results

Alignments and normal forms

Theorem (Wu, 1997)

For any inversion transduction grammar G , there exists an equivalent inversion transduction grammar G' in which every production takes one of the following forms:

$$\begin{array}{lll} S \rightarrow \epsilon|\epsilon & A \rightarrow \rightarrow x|\epsilon & A \rightarrow [BC] \\ A \rightarrow x|y & A \rightarrow \epsilon|y & A \rightarrow \langle BC \rangle \end{array}$$

Definition

If two words are recognized simultaneously by an inversion transduction grammar they are said to be *aligned*.

...word alignment becomes simply the special
or: case of phrasal alignment at the parse tree leaves
 (Wu, 1997, p. 398).

However, $\{S \rightarrow \langle a|a B c|c \rangle, B \rightarrow b|b\}$ and
 $\{S \rightarrow \langle AS' \rangle, S' \rightarrow \langle BC \rangle, A \rightarrow a|a, B \rightarrow b|b, C \rightarrow c|c\}$ do not induce
 the same alignments on abc and cba .

Related work

Measures: PRF = parse failure rate, AER, CPER, TUER = translation unit error rate.

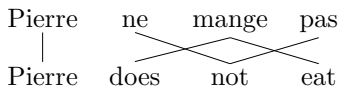
Zens and Ney (2003)	PRF	NF-ITGs
Wellington et al. (2006)	PRF	ITGs
Søgaard and Kuhn (2009)	TUER	NF-ITGs
This work	TUER+PRF	ITGs

$$\text{TUER} = 1 - \frac{2|S_U \cap G_U|}{|S_U| + |G_U|}$$

Configurations

- Configurations beyond NF-ITGs *and* ITGs: IOAs (Wu, 1997), CDTUs (Søgaard and Kuhn, 2009), bonbons (see below) and MIX-DTUs.
- Configurations beyond NF-ITGs: DTUs.

Bonbons:



Data

	Snts	TUs	IOAs	DTUs	CDTUs	Bonbons	MIX-DTUs
Da-Sp	926	6441	0.56	9.16	0.81	0.16	0.23
En-Fr	100	869	0.23	2.99	0.12	0.23	0.23
En-Ge	987	17354	1.75	5.55	0.45	0.05	0.79
En-Po	100	783	0.26	2.17	0.00	0.00	0.38
En-Sp	100	831	0.48	1.32	0.00	0.00	0.36
Po-Fr	100	862	0.23	3.13	0.58	0.00	0.46
Po-Sp	100	882	0.11	0.90	0.00	0.00	0.00
Sp-Fr	100	914	0.11	2.95	0.55	0.00	0.22

Results

	ITGs		NF-ITGs	
	LB-TUER	LB-PFR	LB-TUER	PFR
Da-Sp	1.58	10.37	8.54	40.50
En-Fr	0.69	6.00	2.88	22.00
En-Ge	2.75	47.32	5.24	69.30
En-Po	0.64	5.00	2.43	19.00
En-Sp	0.84	7.00	1.80	15.00
Po-Fr	1.04	9.00	3.36	24.00
Po-Sp	0.11	1.00	0.90	8.00
Sp-Fr	0.77	7.00	3.06	23.00
AV	1.05	11.59	3.53	27.60