Can ITGs generate hand alignments?

Anders Søgaard
(*but* presented by Michael Carl, CBS, Denmark)

Center for Language Technology
University of Copenhagen
Njalsgade 140-142
DK-2300 Copenhagen S
Email: soegaard@hum.ku.dk

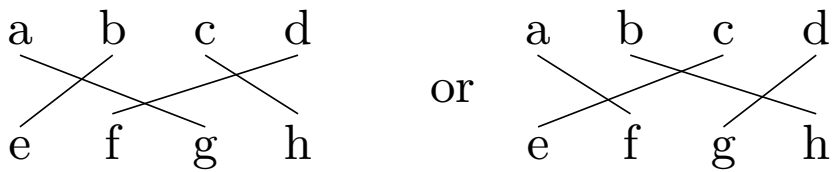
May 27 2009

Wu (1997)

Definition

An ITG is a SCFG in which any production rule with RHS $\langle \phi, \psi \rangle$ is such that the linear order of indeces in ϕ is either monotone or inverted in ψ.

- Wu (1997) show that ITGs, unlike SCFGs, have a Chomsky normal form (NF-ITG).
- The translations defined by ITGs can be recognized in $\mathcal{O}(|G|n^6)$.
- ITGs can't generate inside-out alignments



Background

1997 Wu introduces ITGs and NF-ITGs and notes that ITGs do not induce inside-out alignments.

2003 Zens and Ney test the adequacy of NF-ITGs against two (2) automatically aligned data (En-Fr and En-Ge).

2006 Wellington et al. test the adequacy of ITGs against five (5) hand-aligned data, but with some bias and non-standard assumptions:

- English target language in all datasets.
- Disjunctive interpretation of translation units.

2009 Søgaard and Kuhn test the adequacy of NF-ITG wrt. *specific* alignment configurations on 11 datasets.

2009 Søgaard and Wu test the adequacy of ITG and NF-ITG wrt. *specific* alignment configurations on 11 datasets.

Zens and Ney (2003)

- Zens and Ney (2003) use GIZA++ to *automatically* align parallel text from Canadian Hansards (En-Fr) and Verbmobil (En-Ge).
- They estimate parse failure rate for NF-ITG and their own extension of NF-ITG.
- They interpret TUs conjunctively.
- They report a coverage of 87.0-91.6% on Canadian Hansards and a coverage of 73.5-81.3% on Verbmobil.
- The result was generalized to Ja-En in Zens et al. (2004).

Wellington et al. (2006)

- Wellington et al. (2006) were the first to replicate the experiments of Zens and Ney (2003) on hand-aligned parallel text.
- They use five (5) (small) datasets compared to the two (2) datasets used in Zens and Ney (2003).
- They interpret TUs disjunctively.
- They report an average coverage of about 98% across the five datasets.

Søgaard and Kuhn (2009)

- Søgaard and Kuhn (2009) identify a number of alignment configurations, e.g. inside-out alignments and cross-serial discontinuous TUs (see below), that cannot be induced by NF-ITGs, binary SCFGs, binary STSGs and binary STAGs.
- They count the frequency of such configurations in 11 hand-aligned parallel texts and derive lower bounds on empirical adequacy of the above theories.
- *Caution:* The published results are, as noted by the authors in their presentation at NAACL-HLT 2009, not entirely correct due to an error when preprocessing the data.

Søgaard and Wu (2009)

- Søgaard and Wu (2009) first note that NF-ITGs cannot induce all alignments that can be induced by ITGs, e.g. they cannot induce discontinous TUs.
- They extent the work of Søgaard and Kuhn (2009) to ITG, but also include (corrected) results for NF-ITG.
- We compare our results to their configuration-based lower bounds below.

Motivation

(i) Lower bounds of empirical adequacy helps us *interpret* performance results in SMT.

- System comparison - different lower bounds?
- Error reduction vs. error reduction *modulo* lower bounds.

(ii) Empirical adequacy informs system development:

- The new Stanford system Phrasal is informed by earlier studies in Søgaard and Kuhn (2009) and Søgaard and Wu (2009); see Galley and Manning (2010).

Experiment/Data

	datasets	sent. pairs	disj.	ITG	NF-ITG
Wellington et al.	5	1,427	✓	✓	
Ours	12	3,215		✓	✓

	Sentences	Links
Da-De	266	1314
Da-It	26	1386
Da-Ru	33	833
Da-Sp	966	8944
En-Fr	100	1279
En-Ge	987	23243
En-Po	100	1198
En-Sp	100	1198
Po-Fr	100	1290
Po-Sp	100	1189
Sp-Fr	100	1303
Hansard	337	7418
Total	3215	50595

Algorithm

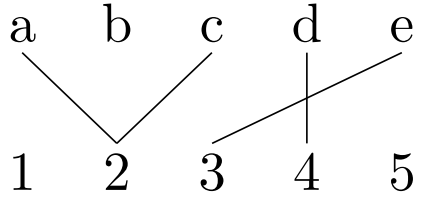
(a) The bichart is initialized by inserting all continuous translation units (possibly with empty gaps).

(b) The algorithm returns *true* if both spans are complete.

(c) If not, it begins to check for possible rule applications that do not violate the alignment. If a rule is applied, the algorithm returns to (b).

The algorithm runs in $\mathcal{O}(|G|n^8)$ because derivation is constrained by a conjunctive interpretation of translation units and no normal form assumptions are made.

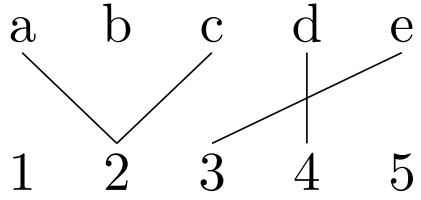
Example 1



	a	b	c	d	e
a					
b					
c	c_1				
d				c_2	
e					c_3

	1	2	3	4	5
1					
2		c_1			
3			c_3		
4				c_2	
5					

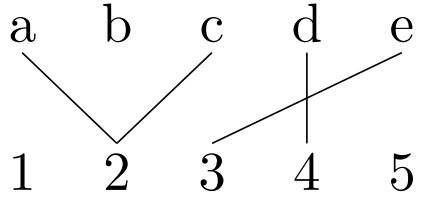
Example 1



	a	b	c	d	e
a					
b					
c	c_1				
d				c_2	
e				c_4	c_3

	1	2	3	4	5
1					
2		c_1			
3			c_3		
4				c_2	
5			c_4		

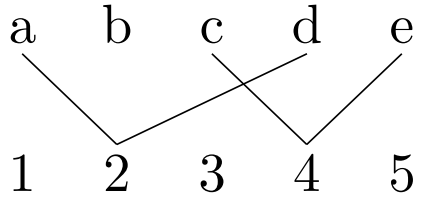
Example 1



	a	b	c	d	e
a					
b					
c	c_1				
d				c_2	
e	c_5			c_4	c_3

	1	2	3	4	5
1					
2		c_1			
3			c_3		
4				c_2	
5	c_5		c_4		

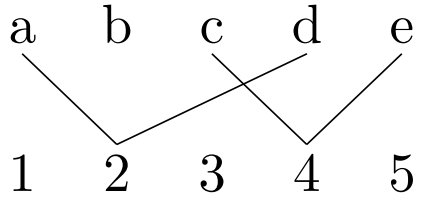
Example 2



	a	b	c	d	e
a					
b					
c					
d					
e					

	1	2	3	4	5
1					
2					
3					
4					
5					

Example 2



	a	b	c	d	e
a					
b					
c					
d					
e					

	1	2	3	4	5
1					
2					
3					
4					
5					

? No possible derivation steps.

Metric

- Our algorithm returns *true* if an alignment can be generated by a ITG, resp. NF-ITG.
- Our metric is $\frac{\#\text{ true}}{\text{sent. pairs}}$, i.e. 1-PFR.
- Lower bounds have been estimated for other metrics, but PFR (or 1-PFR) is most widely used.
  - BLEU (Dreyer et al., 2007)
  - AER/CPER/TUER (Søgaard and Kuhn, 2009; Søgaard and Wu, 2009)

Results

	NF-ITG	SW09(NF)	ITG	SW09	LR
En-Fr	65.00	78.00	68.00	94.00	32.00
En-Po	65.00	81.00	67.00	95.00	25.00
En-Sp	73.00	85.00	74.00	93.00	30.00
Po-Fr	63.00	76.00	63.00	91.00	44.00
Po-Sp	80.00	92.00	81.00	99.00	53.00
Sp-Fr	68.00	77.00	68.00	93.00	51.00
AV	69.00	81.50	70.17	94.17	39.16
Da-De(25)	47.62	-	49.35	-	43.29
Da-It(25)	60.00	-	60.00	-	60.00
Da-Ru(25)	47.05	-	47.05	-	29.41
Da-Sp(25)	30.68	*59.50	35.54	*89.63	29.59
En-Ge(15)	38.97	*30.70	45.13	*52.68	12.31
Hansard(15)	76.98	-	81.75	-	50.79

Conclusion

- The ITG is search space is more adequate than local reordering or IBM with $k = 4$ (Zens and Ney, 2003; Dreyer, 2007; this work).
- Coverage wrt. hand alignments is still only about 70%.
- This may be unimportant in practice (for SMT), but other alternatives exist, incl. RCG (Søgaard, 2009) and PB-SMT with discontinuous phrases (Galley and Manning, 2010).