

UNTO A FULL GROWN MAN

Does syntax-based SMT deserve its complexity?



周国栋

苏州大学自然语言处理实验室

<http://nlp.suda.edu.cn/~gdzhou/>

2012年9月20日





Complexity of syntax-based SMT

- Translation modeling
 - Wu Dekai
 - Simplified CFG
 - David Chang
 - Unlabeled CFG
 - Liu Qun/Zhang Min
 - Tree forest
- Language modeling
 - Just a dream!



- » Memory requirement
- » CPU requirement



How much can syntax-based SMT contribute?

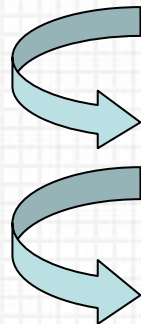
- Just a little!
 - Local syntax can be better (and has already been) captured by phrase-based SMT.
 - Long syntax is too tough for syntax-based SMT.





What contribute most to SMT?

- Morphology (40%): not understandable
 - Unknown word and new entity
- Syntax (20%): understandable
 - Local syntax: 10%(phrase-based)
 - Long syntax: 10%(syntax-based)
- Discourse(40%): not understandable
 - Cohesion: (Zero) anaphora
 - Coherence: discourse structure



•我昨天上街，看见一个人，打了一下招呼。

•I took to the streets yesterday and saw a man hit a bit of greeting.

•我昨天走上街头，看见一个人打一个问候的位。



谢谢

- 明年10月底11月初计算语言学年会苏州大学见!

