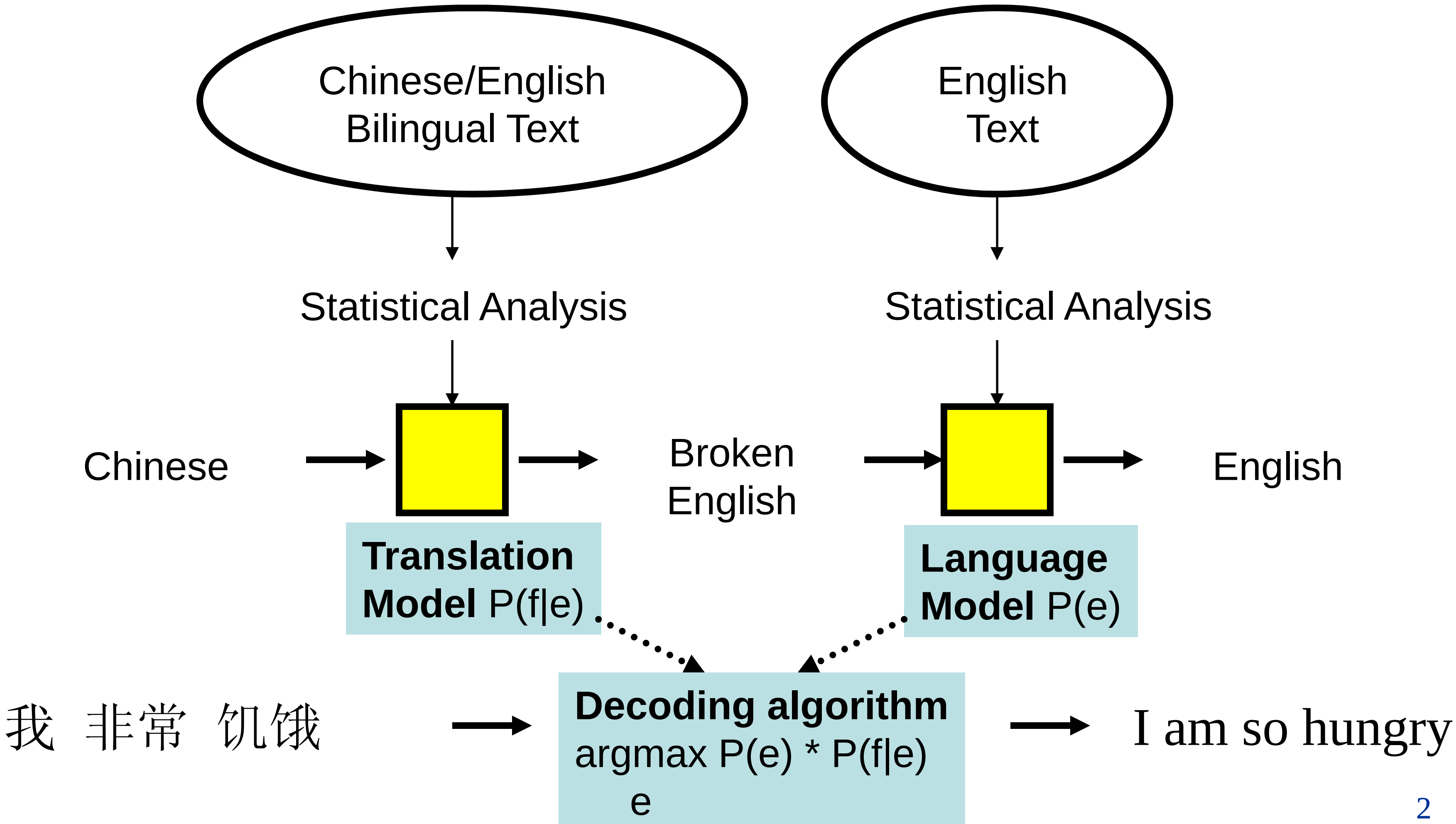
Syntactic structure alignment based Statistical machine translation

Doctoral Thesis Proposal

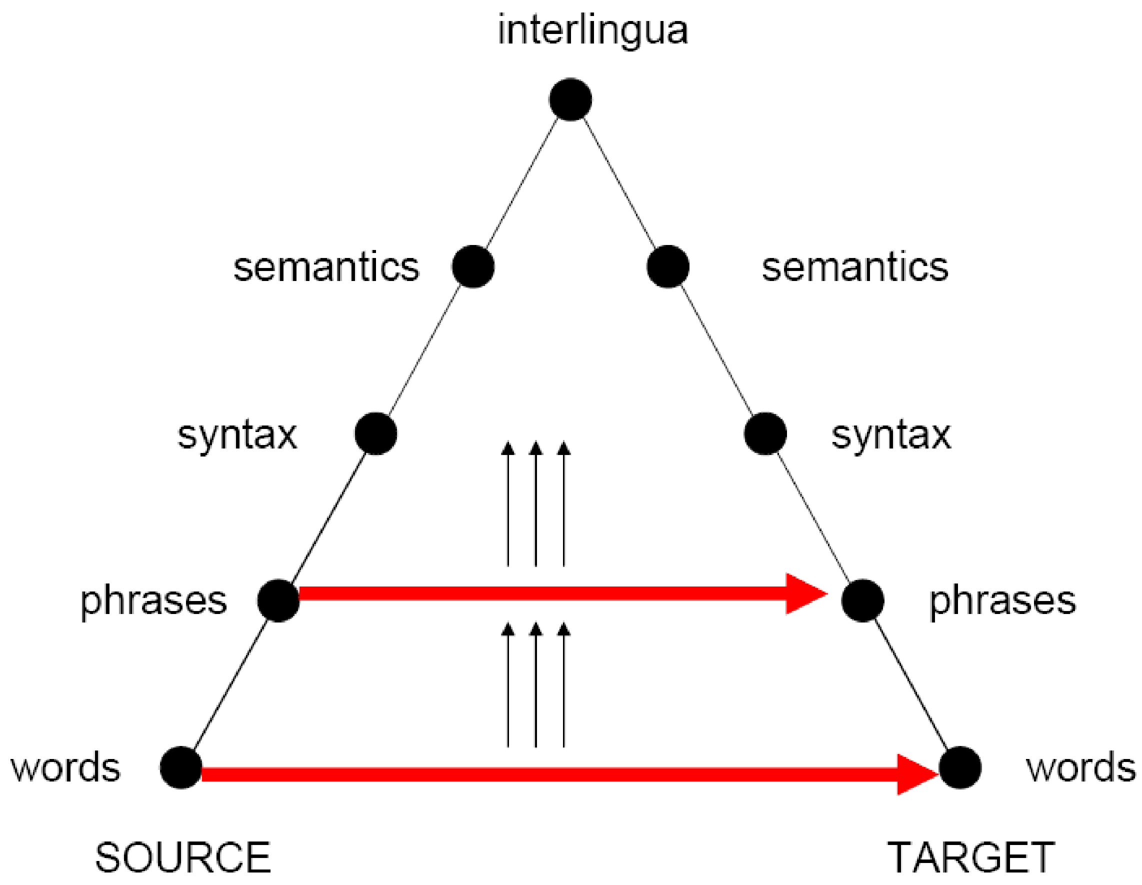
Sun Jun

National University of Singapore

SMT in nutshell



SMT in nutshell (cont.)



Syntax based----Why Syntax?

- Much more grammatical output
- Accurate control over re-ordering
- Appropriate insertion of function words
- Word translations need to depend on grammatically-related words

Syntax based

- String-to-tree (Galley et al., 2006; Marcu et al., 2006; Yamada and Knight, 2001):
  - Source sentence: a word sequence; Target sentence: a tree
  - Decoding as parsing (CYK)
- Tree-to-string (Quirk et al., 2005; Liu et al., 2006):
  - Source sentence: a tree; Target sentence: a word sequence
  - Tree2string rules
  - Decoding more easier (translate each internal node in a bottom-up manner)
- Tree-to-tree (Eisner, 2003; Graehl and Knight, 2004; Zhang et al., 2007):
  - Source sentence: a tree; Target sentence: a tree
  - Tree2tree rules
  - Decoding as Tree2string (bilingual parsing)

Motivations

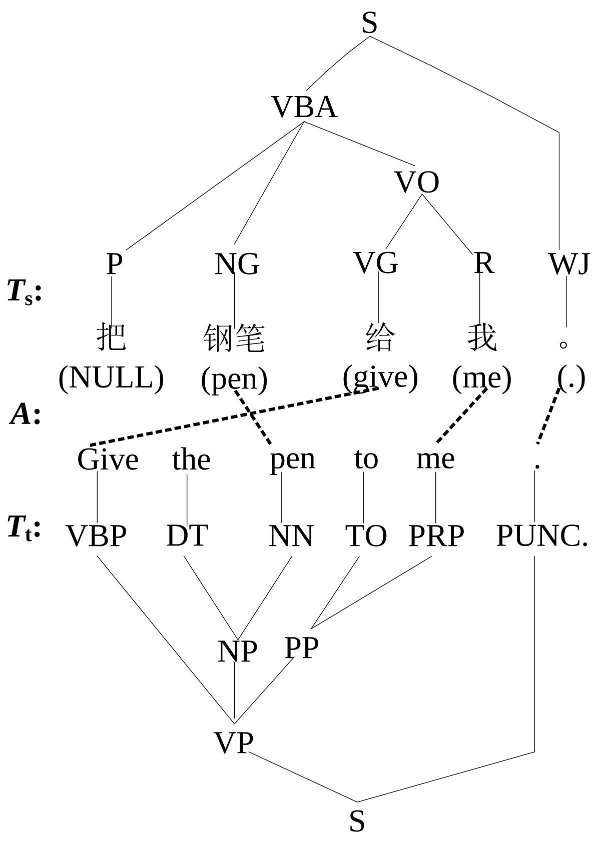
- More & Variant Syntactic translational equivalences (TEs)
  - Continuous phrases vs. discontinuous phrases
  - Syntactic phrases vs. non-syntactic phrases
- Comparison of TEs are still not well studied
  - Local/Global Reordering rules
  - Continuous/discontinuous phrasal rules
  - Lexical/partial-lexical/non-lexical rules
- TEs acquired by word alignment
  - Only flat features adopted

Motivations (cont.)

- A more powerful syntax-based translation model
  - Global reordering … (as former syntax-based models)
  - Improve grammaticality of target translation results (tree2tree)
  - Well model all kinds of TEs
    - *Continuous phrases vs. non-continuous phrases*
    - *Syntactic phrases vs. non-syntactic phrases*
- To acquire more effective TEs (syntactic structure alignment)
  - Overcome easy but non-accurate word alignment based methods
  - A structure alignment framework
  - Design more effective syntactic features
- Specified decoding algorithm & parameterization methods

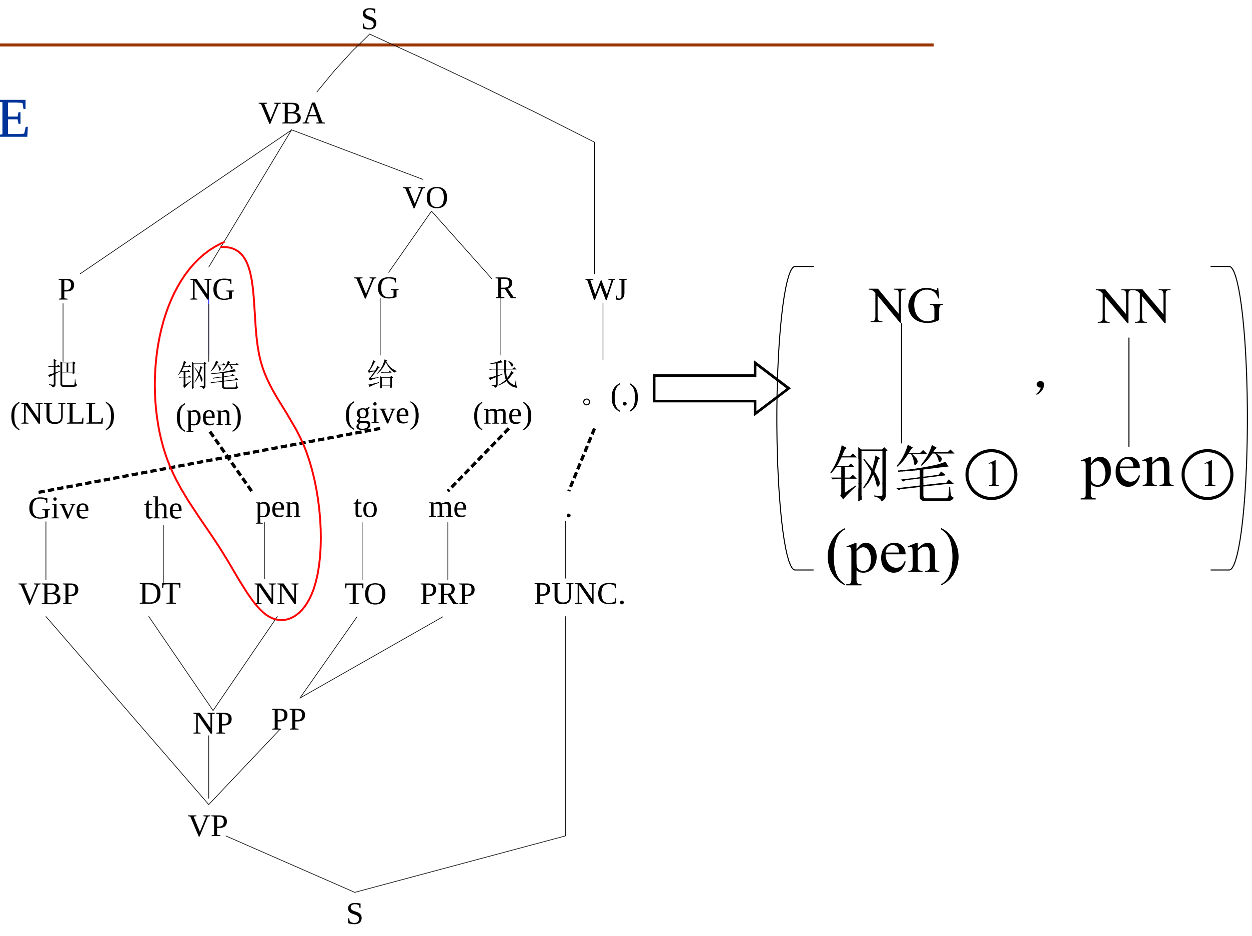
Pre Work/TM

- Synchronous Tree Substitution Grammar
- Synchronous Tree sequence Substitution Grammar
- Synchronous discontinuous Tree sequence Substitution Grammar



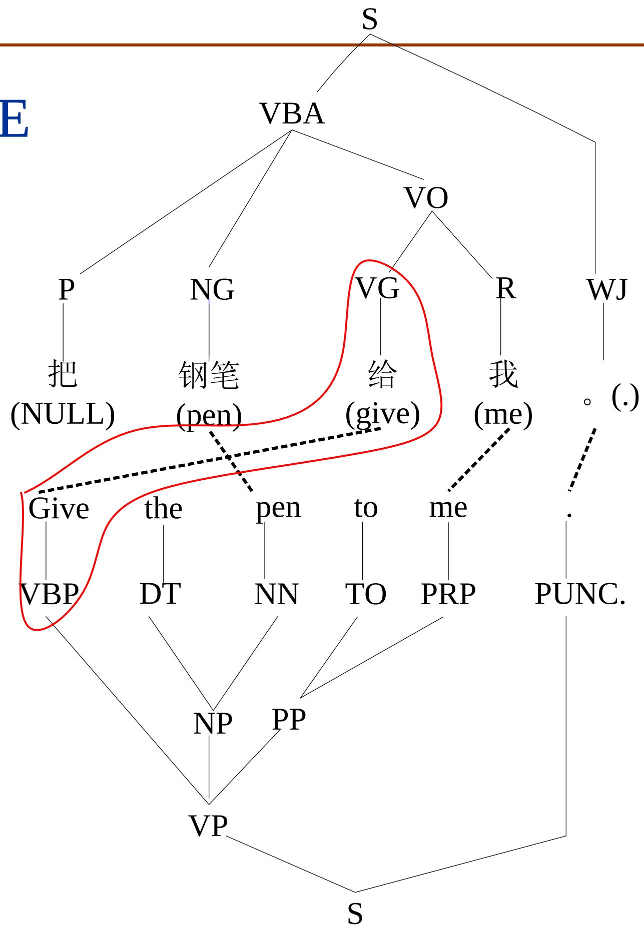
Pre Work/TM/STSG

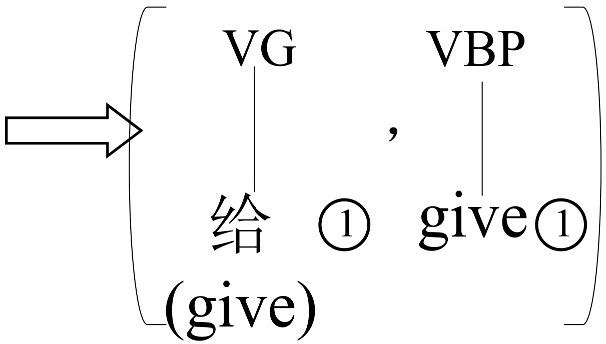
- TE



Pre Work/TM/STSG(cont.)

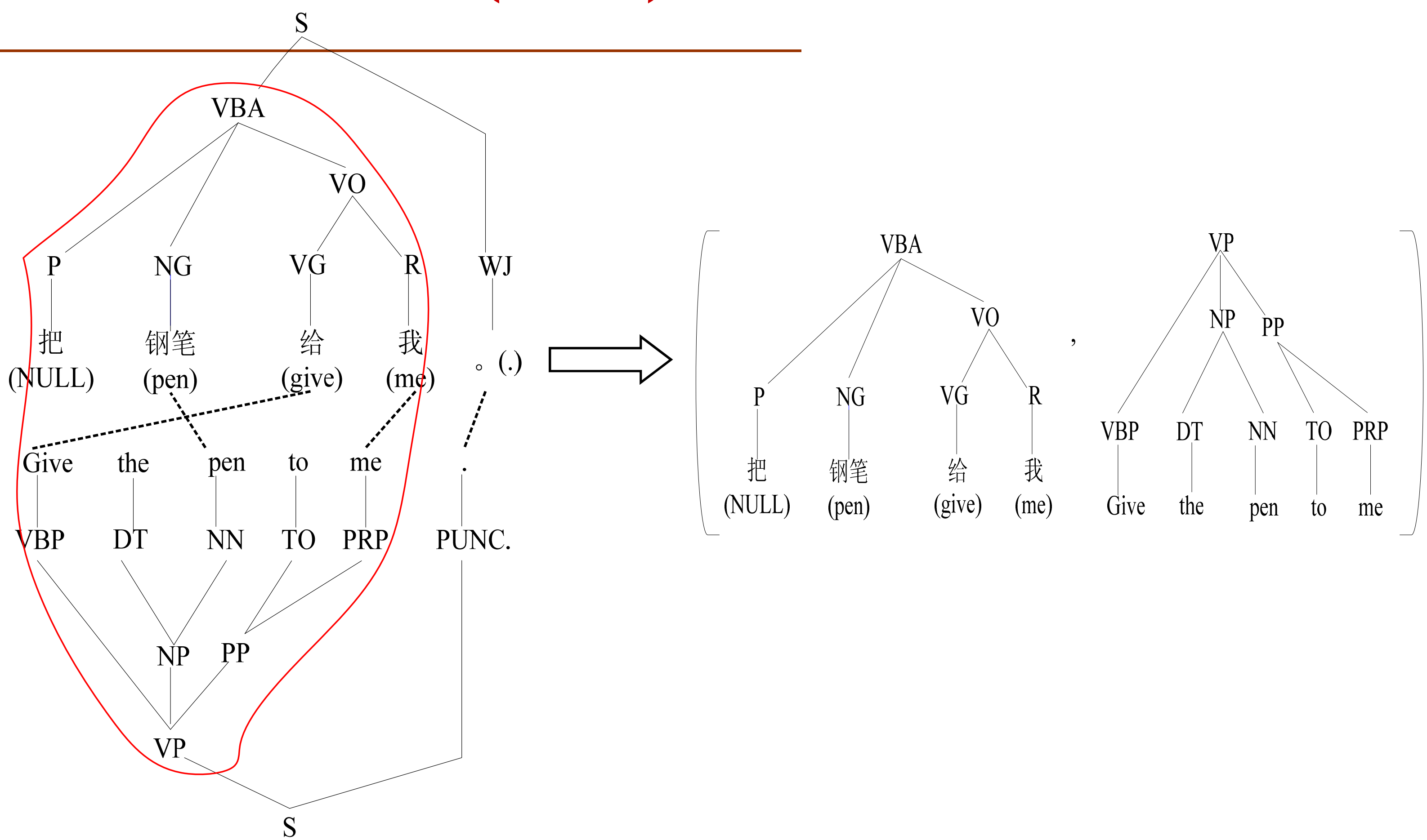
- TE



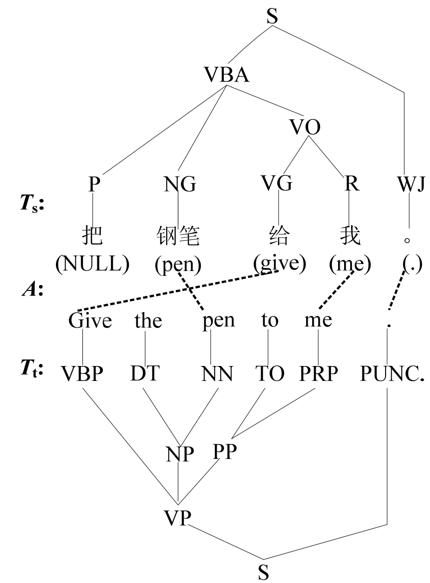


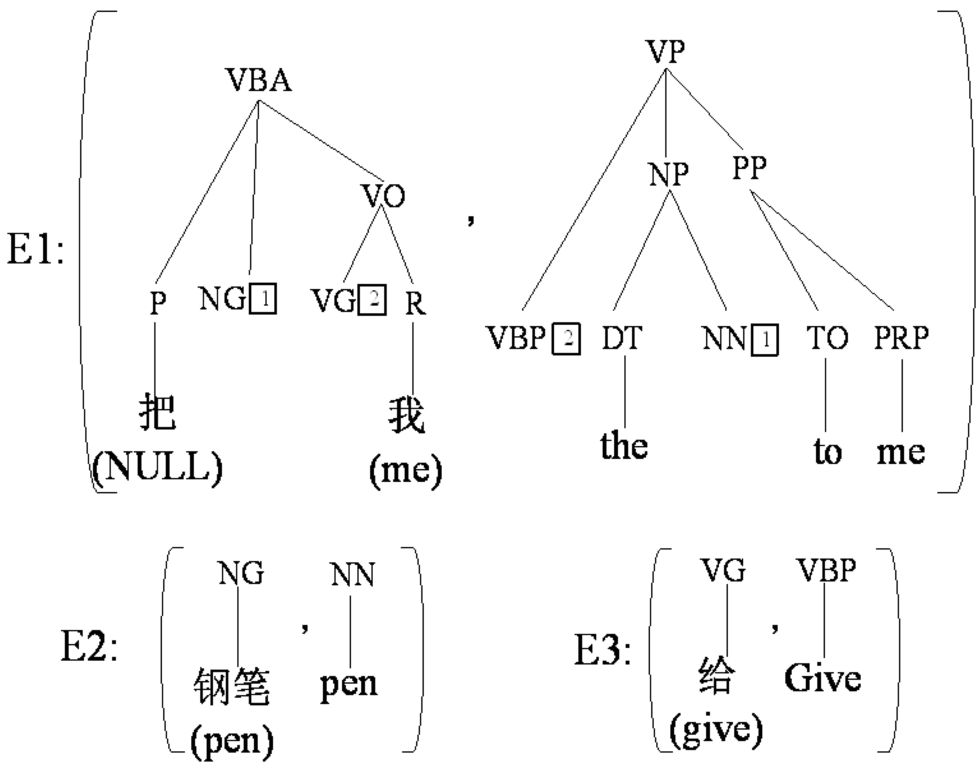
Pre Work/TM/STSG(cont.)

- TE



Pre Work/TM/STSG(cont.)



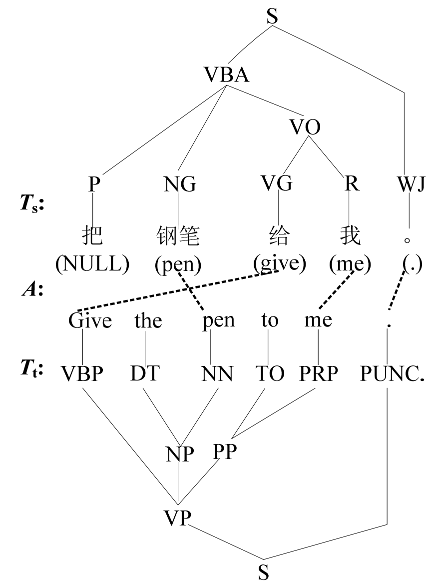


Pre Work/TM/STSG(cont.)

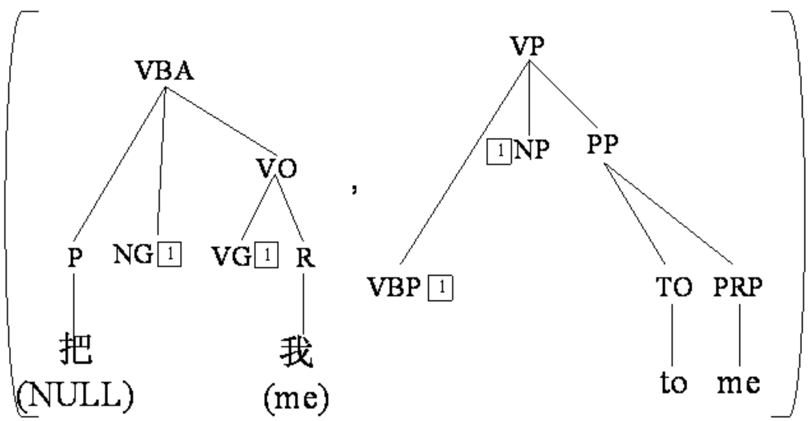
- Pros
  - Non-isomorphic tree alignment
  - Multi-level global structure distortion
  - Discontinuous phrases
- Cons
  - Non-syntactic phrases

Pre Work/TM/STSSG

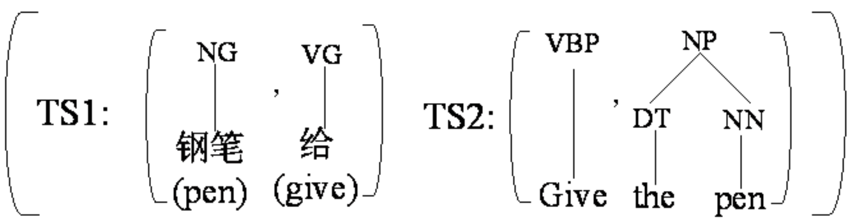
➢ Synchronous Tree sequence Substitution Grammar



ETS1:



ETS2:

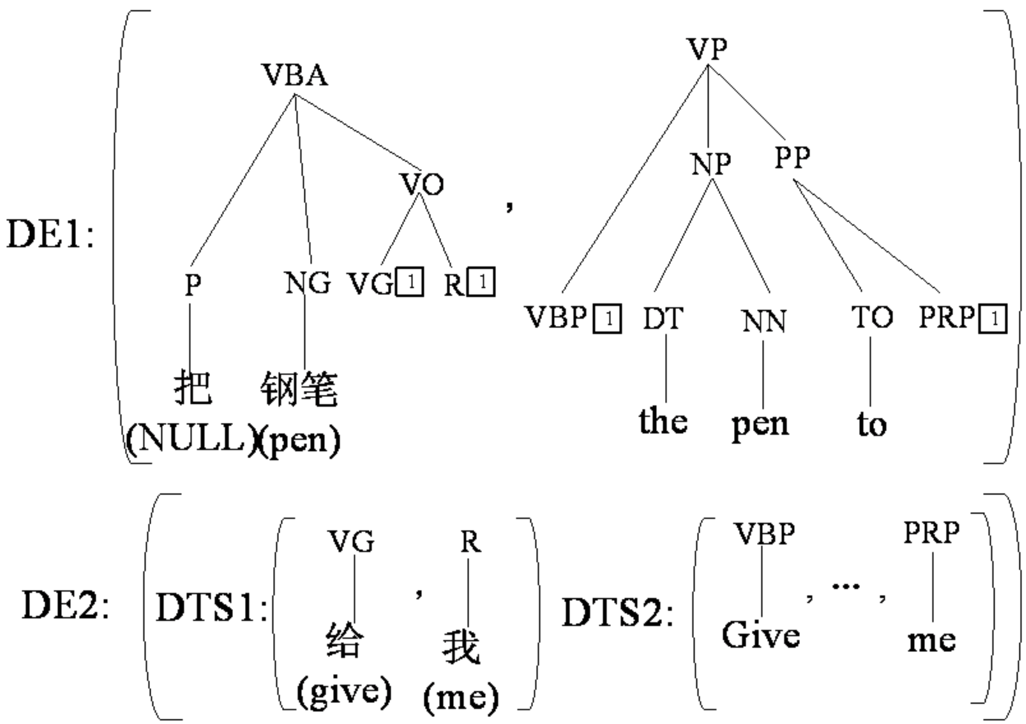


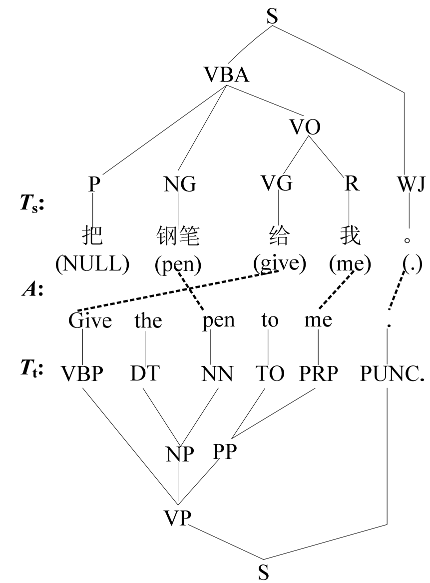
Pre Work/TM/STSSG(cont.)

- Pros
  - Fully cover STSG
  - Non-syntactic phrases
- Cons
  - Discontinuous phrases only acquired from continuous TE

Pre Work/TM/SDTSSG

➤Synchronous discontinuous Tree sequence Substitution Grammar





Pre Work/TM/SDTSSG(cont.)

- Pros
  - Fully cover STSG & STSSG
  - Discontinuous phrases obtained from both continuous TEs and discontinuous TEs
- Cons
  - Solve a few language phenomena but not much
  - Noisy rules obtained from wrong word alignments

Pre Work/TM/SDTSSG(cont.)

- Why SDTSSG works?

  E.g. 两 人 就 对簿公堂 。

  - STSSG:
    - *the two candidates would 对簿公堂*
  - SDTSSG:
    - *the two people can confront other countries at court leisurely manner*
  - Reasons: more effective TEs obtained
    - *VV(对簿公堂) ||| VB(confront) NP(JJ(other),NNS(countries)) IN(at) NN(court) … JJ(leisurely) NN(manner)*

Pre Work/TM/SDTSSG(cont.)

- Why SDTSSG works?

  E.g. 另一方面 英国 的 资料 则 预期 *将 显示 工业 活动 增加* 。

  - STSSG:
    - *data while Britain is expected to show that the number of industrial activities .*
  - SDTSSG:
    - *British data on the other hand , predicted that industrial activities will increase .*
  - Reasons: more effective TEs obtained
    - *AD(将) ... VV(增加) ||| MD(will) VB(increase)*

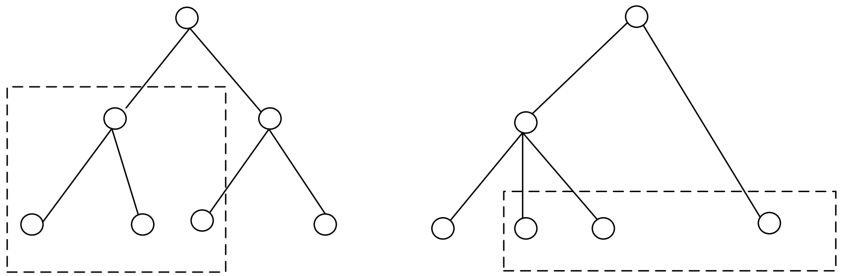
Pre Work/TM(cont)

- Experiment setting (basic setting's same for all models)
  - Training data: FBIS (7.2M (Chinese)+9.2M(English) words)
  - Develop set: NIST-02
  - Test set: NIST-05
- Performance

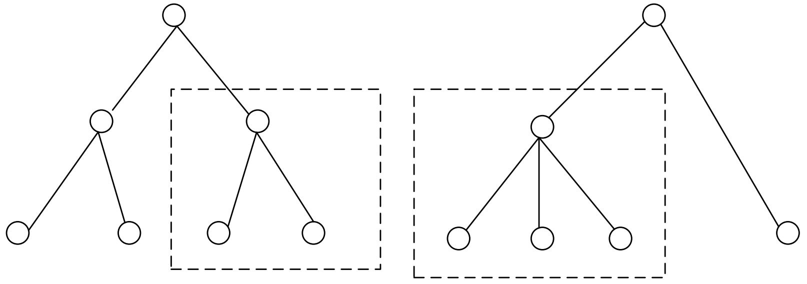
Model	BLEU
Moses	23.86
STSG	24.71
STSSG	25.66
SDTSSG	26.17

Pre Work/Structure alignment

- Problem definition
  - Tree sequence alignment



  - Sub-tree alignment



Pre Work/Structure alignment/Sub-tree alignment

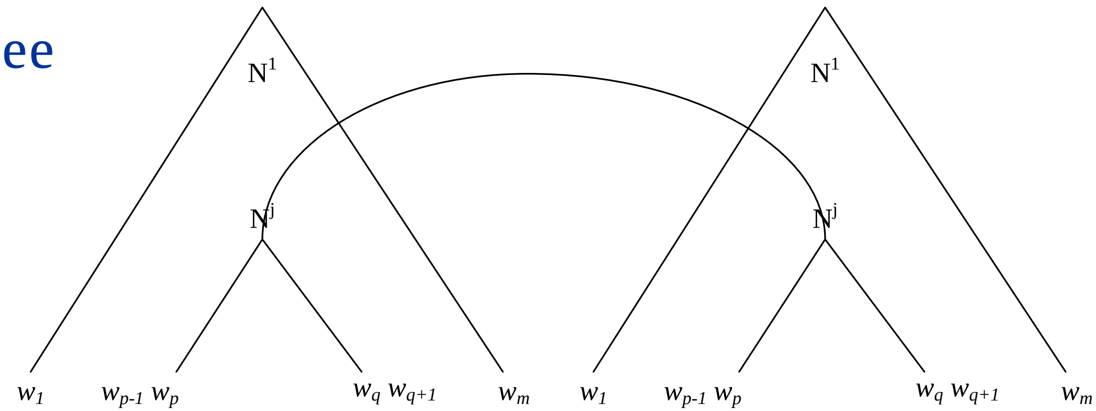
- Previous Work
  - Heuristics
  - Modify the original aligned structures
  - Most work on dependency constraint
  - Ask for word alignment as a premise

Pre Work/Structure alignment/Sub-tree alignment (cont.)

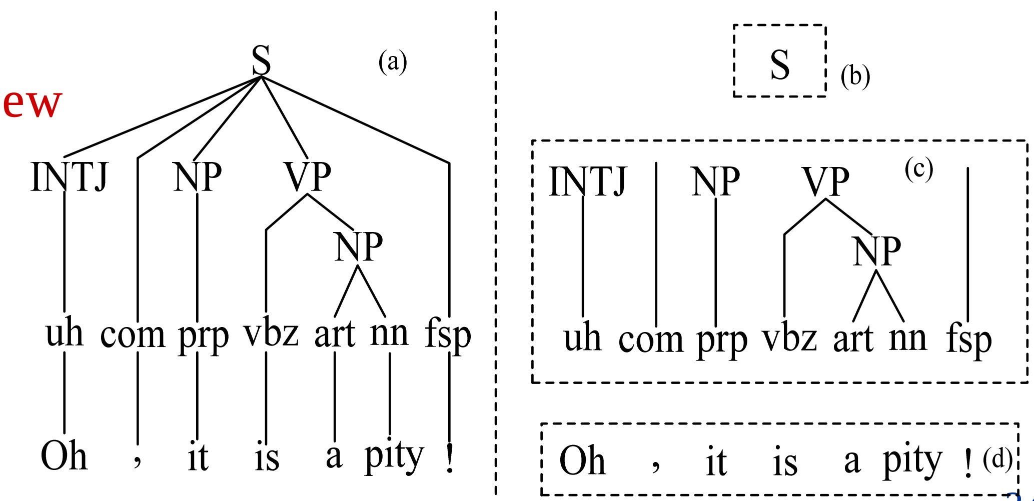
- preservation of the given tree structures;
- word-level alignments not fixed a priori
- ***structure features***
- ***statistical approach***

Pre Work/Structure alignment/Sub-tree alignment (cont.)

- Two perspective of sub-tree structure
  - Inside-Outside View

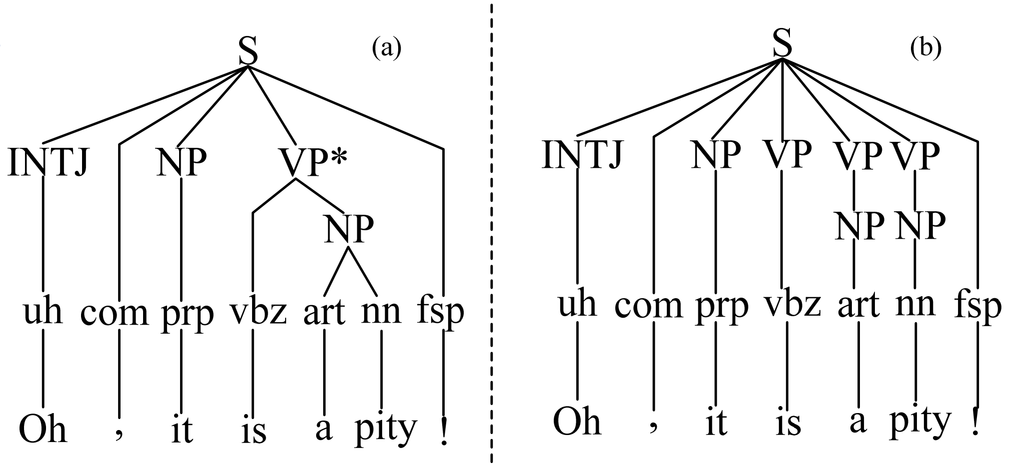


  - Top-Down generative View



Pre Work/Structure alignment/Sub-tree alignment (cont.)

- Sub-Structure Extraction
  - Head nodes
  - Tokens
  - Trunk
    - *POS tags*
    - *CFG rules*
    - *Path Generated Element (PGE)*
      - Inside PGE: VP-vbz; VP-NP-art; VP-NP-nn;
      - Ourside PGE: S-INTJ-uh; S-com; S-NP-prp; S-fsp



Pre Work/Structure alignment/Sub-tree alignment (cont.)

- Metrics

$$\text{precision} = \frac{|A \cap S|}{|A|} \qquad \text{recall} = \frac{|A \cap S|}{|S|}$$

$$\text{AER*} = 1 - \frac{|A \cap S| + |A \cap P|}{|A| + |S|} \qquad \text{AER} = 1 - \frac{2|A \cap S|}{|A| + |S|}$$

Pre Work/Structure alignment/Sub-tree alignment/Results

Exp ID	sub-model combination	Prior	sub-structure model			precision	recall	AER
			word	root	trunk			
~~1~~	~~Prior_word_root_POS~~	~~Y~~	~~Y~~	~~Y~~	~~POS~~	0.5824	0.2323	0.6679
2	Prior_word_root_CFG	Y	Y	Y	CFG	0.6004	0.2187	0.6794
3	word	--	Y	--	--	0.4488	0.4485	0.5513
4	root	--	--	Y	--	0.4862	0.2534	0.6668
5	PGE	--	--	--	PGE	0.4683	0.4869	0.5226
6	word_root	--	Y	Y	--	0.5838	0.5949	0.4106
7	word_PGE	--	Y	--	PGE	**0.5930**	**0.6136**	**0.3968**
8	root_PGE	--	--	Y	PGE	0.4905	0.5100	0.4999
9	word_root_PGE	--	Y	Y	PGE	0.5796	0.6039	0.4085
10	Prior_word	Y	Y	--	--	0.4751	0.4811	0.5219
11	Prior_root	Y	--	Y	--	0.4915	0.4888	0.5098
12	Prior_PGE	Y	--	--	PGE	0.5378	0.5624	0.4501
13	Prior_word_root	Y	Y	Y	--	0.6045	0.6247	0.3855
14	Prior_word_PGE	Y	Y	--	PGE	**0.6138**	**0.6383**	**0.3742**
15	Prior_root_PGE	Y	--	Y	PGE	0.5361	0.5600	0.4521
16	Prior_word_root_PGE	Y	Y	Y	PGE	0.6066	0.6311	0.3813
17	skip1_s1_span1	Baseline1				0.5354	**0.5724**	0.4467
18	skip2_s1_span1	Baseline2				0.5720	0.5520	**0.4382**

Some more to propose

- For syntax-based models
  - Specific decoding & training algorithm
  - Better parameterization methodology
- For structure alignment
  - More effective syntactic features
  - Discriminative training

Potential Contributions

- Formally define the syntactic TE from different granularities
- Explore the applicability and effectiveness of different TE to syntax-based models
- A series of structure alignment based model with varied modeling ability based on different grammars
- Comparative studies among these synchronous grammars to access their ability in describing parallel data
- An unsupervised framework for syntactic structure alignment
- Benefit syntactic structure alignment to MT

The End