Chao Su, Yu-hang Guo, He-yan Huang, Shu-min Shi, Chong Feng, 2017. Incorporating target language semantic roles into a string-to-tree translation model. *Frontiers of Information Technology & Electronic Engineering*, **18**(10):1534-1542. https://doi.org/10.1631/FITEE.1601349

Incorporating target language semantic roles into a string-to-tree translation model

Key words: Machine translation; Semantic role; Syntax tree; String-to-tree model

Corresponding author: He-yan Huang
E-mail: hhy63@bit.edu.cn

ORCID: http://orcid.org/0000-0002-0320-7520

Motivation

- The string-to-tree model is one of the most successful syntax-based statistical machine translation models.
- Because of a lack of semantic information, the string-to-tree model often produces translation with semantic role confusion.
- Semantic roles tend to agree better between two languages than syntactic structures and constitute the skeleton of a sentence.
- We try to merge semantic roles with syntax information into one string-to-tree model, so that it can learn and choose better translation rules.

Main idea

- In natural language sentences, a syntactic structure often undertakes a semantic role.
  - A noun phrase may be a role of ARG0 (agent) or ARG1 (patient).
  - Syntactic structures can be the child node of a semantic argument in a tree.
  - Translation systems can learn which syntactic structure often takes which semantic role, and how to order the semantic chunks.

Main idea

- Furthermore, semantic roles often show some hierarchical structure:
  - a predicate and its arguments become one argument of another predicate;
  - the predicate *get* and its arguments, *ARG0 you* and *ARG1 it*, constitute the *ARGM-TMP* argument of the predicate *run*;
  - This hierarchy inspires us to build a structure-like parse tree, which can help us understand the semantic relations among the chunks and improve translation performance.

(a): When [ARG0 you] [TARGET get] [ARG1 it], run at top speed to buy a SLR.

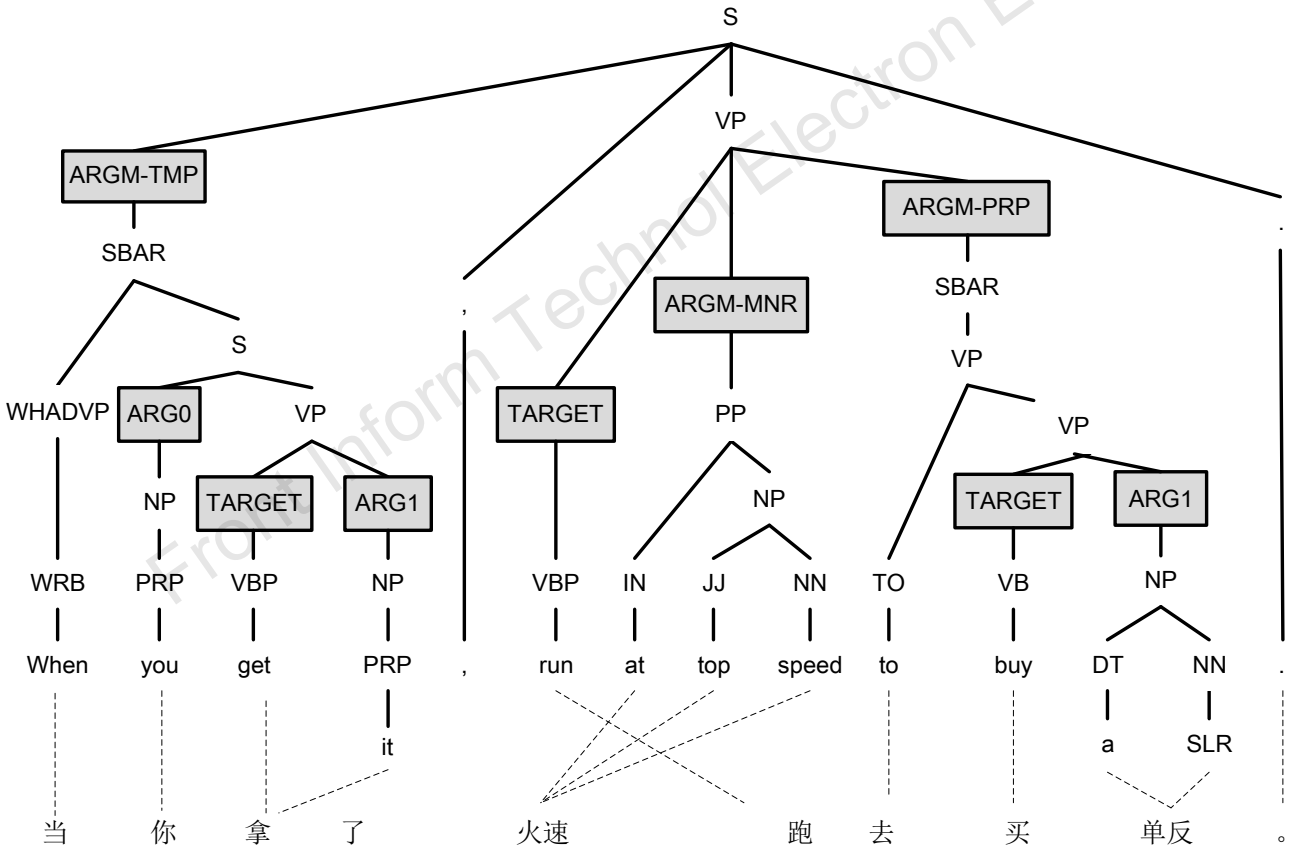
(b): [ARGM-TMP When you get it], [TARGET run] [ARGM-MNR at top speed] [ARGM-PRP to buy a SLR].

(c): When you get it, run at top speed to [TARGET buy] [ARG1 a SLR].

Fig. 1 A sample SRL result for an English sentence

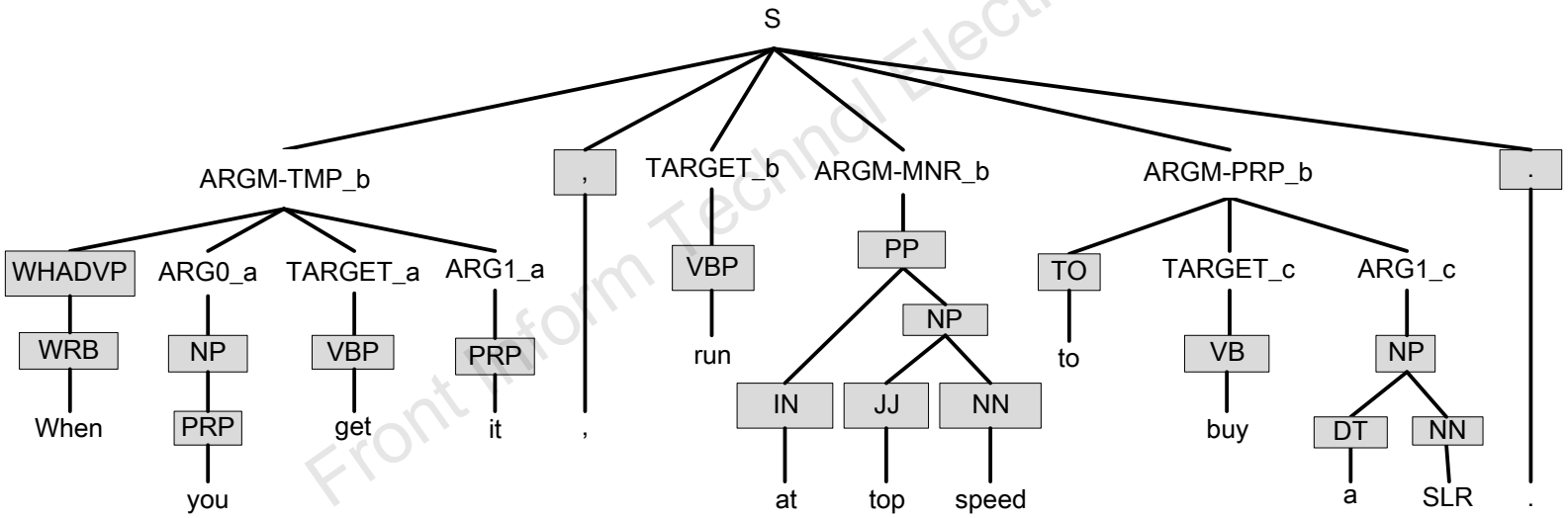
Method

1. Constructing the syntax-role tree



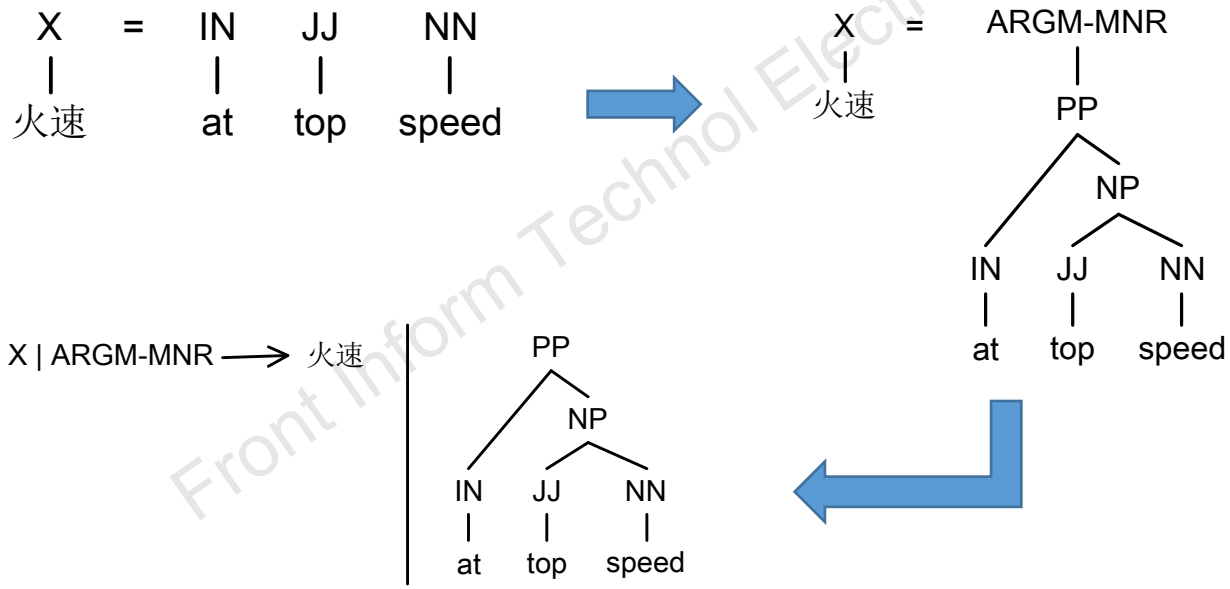
Method (Cont'd)

2. Constructing the role-syntax tree



Method (Cont'd)

3. Learning translation rules



Experimental setup

- Training corpora: BOLT, FBIS, and large-scale data
- Tuning set: NIST MT02
- Test sets: NIST MT03, MT04, MT05, MT06 (NIST), MT08, MT12 (general), and MT08-12 (progress)
- Parsing: Berkeley parser
- SRL tool: ASSERT
- Word-align tool: Mgiza
- LM: 5-gram

Major results

- Both SRT and RST outperform the baseline on the BOLT spoken corpus.

Table 2 Experimental results on the BOLT corpus

Method	BLEU score	Meteor score
Str2tr	13.29	22.52
SRT	13.85* (+0.56)	23.14* (+0.62)
RST	13.51 (+0.22)	22.73 (+0.21)
Phb	**14.17**	**23.25**

Major results (Cont'd)

- In terms of the FBIS news corpus, SRT performs better than RST and exhibits an improvement of 1.33 BLEU points.

Table 3 Experimental results on the FBIS corpus

Test set	BLEU score (%)				Meteor score (%)			
	Str2tr	SRT	RST	Phb	Str2tr	SRT	RST	Phb
MT03	28.83	**30.35*** (+1.52)	29.29 (+0.46)	29.57	28.20	28.71* (+0.51)	28.01 (–0.19)	**28.84**
MT04	31.33	**33.15**$^{*\#}$ (+1.82)	31.87* (+0.54)	31.46	29.31	**30.11**$^{*\#}$ (+0.80)	29.53* (+0.22)	29.97
MT05	28.27	**29.92**$^{*\#}$ (+1.65)	28.69 (+0.42)	28.33	28.66	29.26* (+0.60)	28.62 (–0.04)	**29.47**
MT06	27.76	**29.27**$^{*\#}$ (+1.51)	28.32* (+0.56)	28.72	27.03	27.97* (+0.94)	27.20 (+0.17)	**28.10**
MT08	22.27	**22.71**$^{*\#}$ (+0.44)	21.76 (–0.51)	22.11	24.09	**24.40*** (+0.31)	23.59 (–0.50)	24.37
MT08-12	21.12	**22.39**$^{*\#}$ (+1.27)	21.00 (–0.12)	21.84	23.82	24.35* (+0.53)	23.26 (–0.56)	**24.57**
MT12	20.57	**21.68**$^{*\#}$ (+1.11)	20.79 (+0.22)	21.00	22.90	23.34 (+0.44)	22.81 (–0.09)	**23.45**
Average	25.74	**27.07** (+1.33)	25.96 (+0.22)	26.14	26.29	26.88 (+0.59)	26.15 (–0.14)	**26.97**

* and # denote that the result is significantly better than those of Str2tr and Phb, respectively (at significance level $p<0.01$)

Major results (Cont'd)

- In terms of large-scale data, both SRT and RST beat the baseline. RST shows an improvement of 1.13 BLEU points.

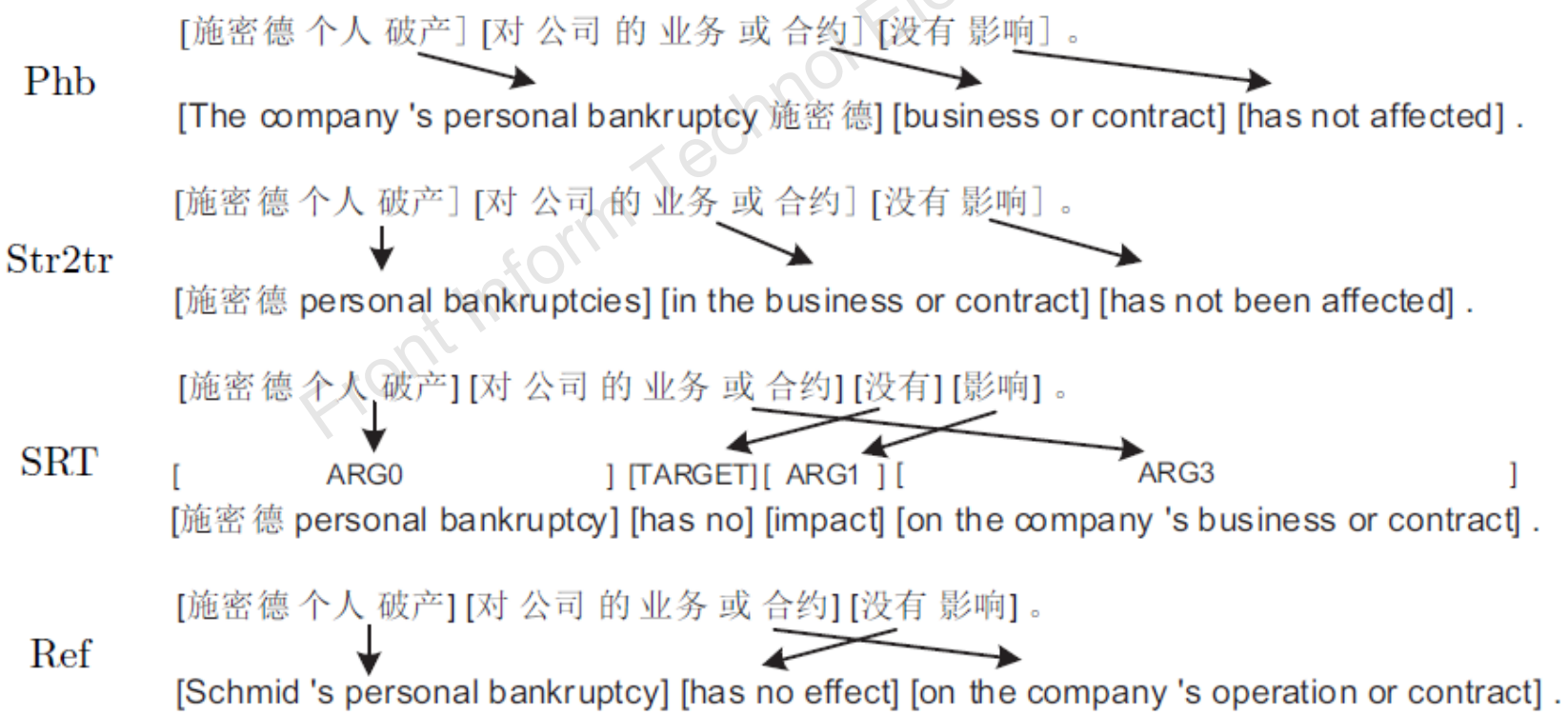
Table 4 Experimental results on the large-scale corpus

Test set	BLEU score (%)				Meteor score (%)			
	Str2tr	SRT	RST	Phb	Str2tr	SRT	RST	Phb
MT03	31.80	33.74* (+1.94)	**34.26**$^{*\#}$ (+2.46)	33.50	30.55	30.92$^{\#}$ (+0.37)	**31.50**$^{*\#}$ (+0.95)	30.64
MT04	34.29	35.60$^{*\#}$ (+1.31)	**36.28**$^{*\#}$ (+1.99)	34.79	31.20	31.57$^{\#}$ (+0.37)	**32.13**$^{*\#}$ (+0.93)	31.09
MT05	33.19	34.24* (+1.05)	**34.73**$^{*\#}$ (+1.54)	33.92	31.55	31.51 (–0.04)	**32.41**$^{*\#}$ (+0.86)	31.56
MT06	32.85	32.89$^{\#}$ (+0.04)	**33.28**$^{\#}$ (+0.43)	31.93	29.76	29.75$^{\#}$ (–0.01)	**30.28**$^{*\#}$ (+0.52)	29.15
MT08	26.11	**26.32**$^{\#}$ (+0.21)	26.29$^{\#}$ (+0.18)	23.99	26.48	26.61$^{\#}$ (+0.13)	**26.84**$^{\#}$ (+0.36)	25.39
MT08-12	25.34	25.13$^{\#}$ (–0.21)	**25.39**$^{\#}$ (+0.05)	24.06	26.32	26.20$^{\#}$ (–0.12)	**26.60**$^{\#}$ (+0.28)	25.70
MT12	22.88	23.72* (+0.84)	**24.08**$^{*\#}$ (+1.20)	23.48	24.89	25.15$^{\#}$ (+0.26)	**25.44**$^{*\#}$ (+0.55)	24.94
Average	29.49	30.23 (+0.74)	**30.62** (+1.13)	29.38	28.68	28.82 (+0.14)	**29.31** (+0.63)	28.35

* and # denote that the result is significantly better than those of Str2tr and Phb, respectively (at significance level $p<0.01$)

Example analysis

- SRT successfully recognizes the prepositional phrase *on the company's business or contract* as an *ARG3* argument and moves it to the end of sentence.



Conclusions

- We have presented an effort that uses semantic information for SMT.
- Our methods improve the translation performance in the following aspects:

  (1) They use the semantic roles in the target language to reorder the semantic chunks in output;

  (2) They use the hierarchy of predicate-argument structure to generate semantic chunks and gather them together.
- In future work, we will explore in detail the situations in which our models work and how to use the overlapped semantic roles dropped in this work.