

# Integrating Case Frame into Japanese to Chinese Hierarchical Phrase-based Translation Model

---

Jin'an Xu, Jiangming Liu, YuFeng Chen, Yujie Zhang, Fang Ming, Shaotong Li

# Outlook

◆ Motivation

◆ Case Frame

◆ Method

◆ Experiment

◆ Conclusion

# Motivation

- Hierarchical phrase-based model limit
  - Rule length limit (10)
  - Variables limit (2)
  - Glue rules
  - Long-term reordering
- Linguistic features (Japanese)
  - subject – object – verb structure
  - auxiliary words

# Outlook

◆ Motivation

◆ Case Frame

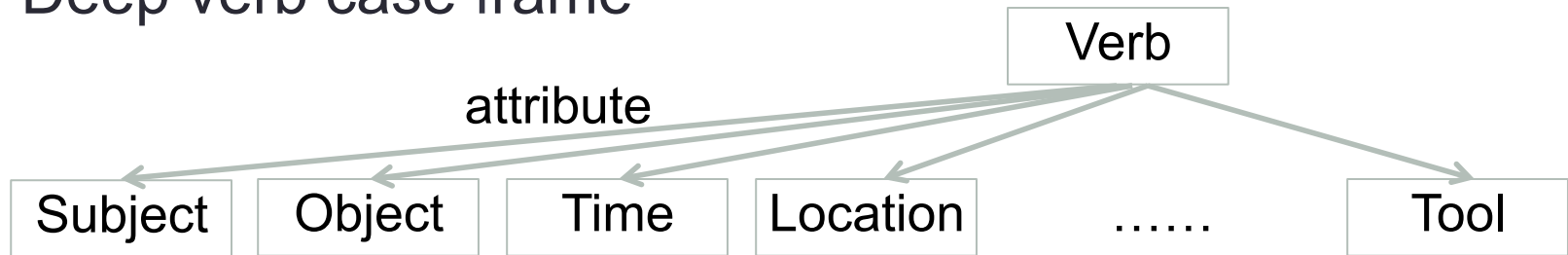
◆ Method

◆ Experiment

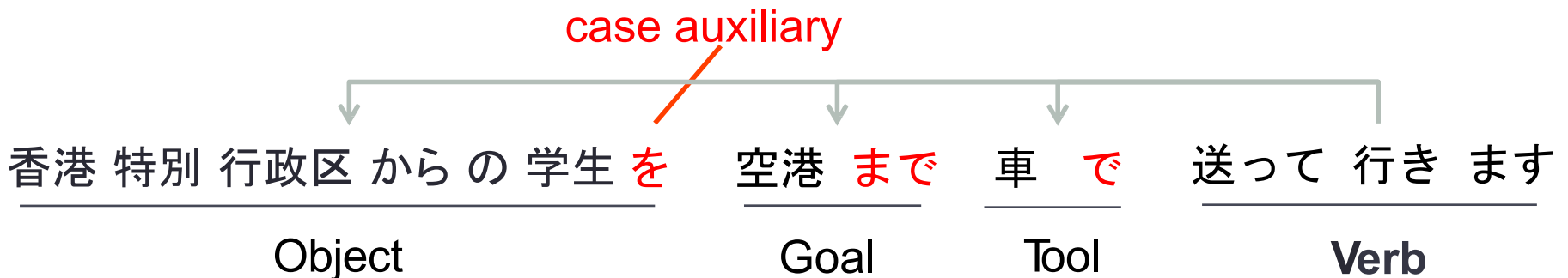
◆ Conclusion

# Verb Case Frame

- Deep verb case frame



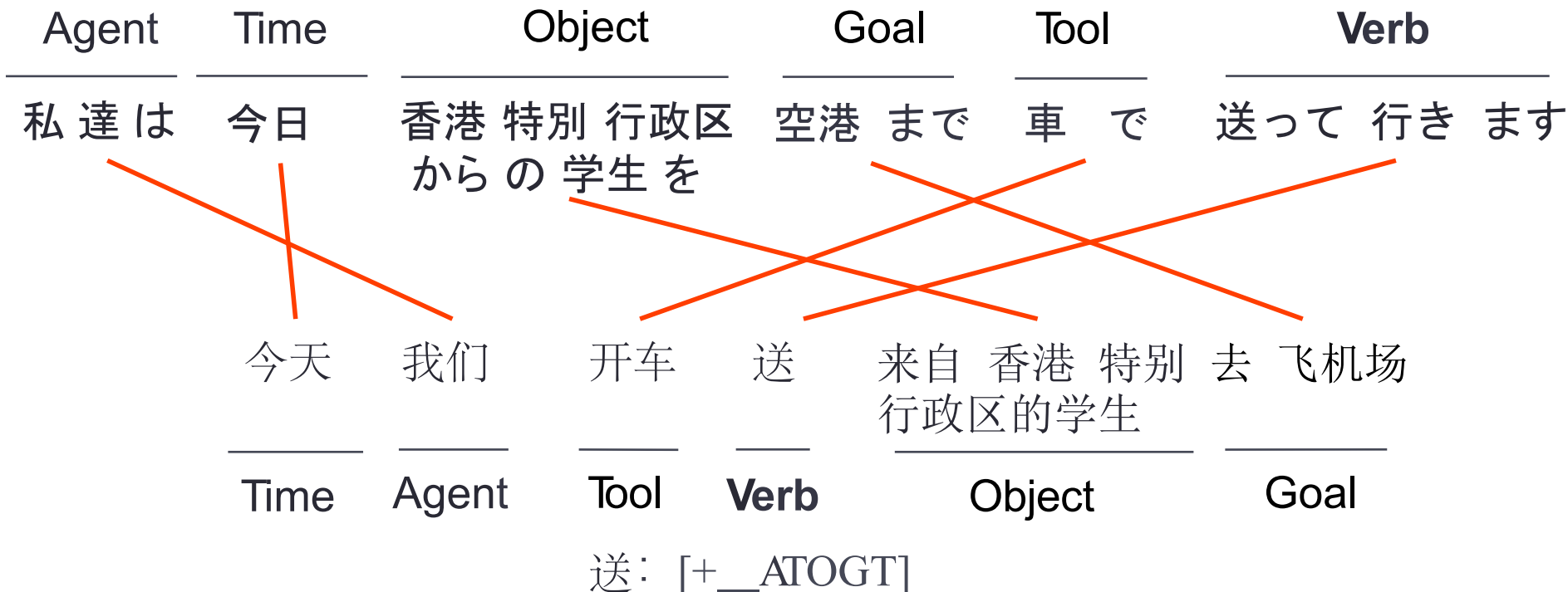
- Specific to Japanese explicit case frame



# Verb Case Frame

- Deep verb case frame between paralleled sentences in two languages

送って 行き ます: [+\_\_ATOOGT]



# Verb Case Frame

- Deep case frame to shallow case frame for Japanese

Agent	Time	Object	Goal	Tool	Verb
私 達 <b>は</b>	今日	香港 特別 行政区 からの 学生 <b>を</b>	空港 <b>まで</b>	車 <b>で</b>	送って 行き ます
ガ( <b>は</b> )	時間(NULL)	ヲ( <b>を</b> )	マデ( <b>まで</b> )	デ( <b>で</b> )	

# Outlook

- ◆ Motivation
- ◆ Case Frame
- ◆ Method
- ◆ Experiment
- ◆ Conclusion



# Method

- Case Frame Rule extraction
  - Obtain case frame rules from paralleled sentences with word alignments
- Transforming to SCFG
  - Transform case frame rules into hiero rules.

# Method

- Case Frame Rule extraction

- examples of case frame rules

- Lexical rule

(ガ) 私達は, 我们

- Reordering rule

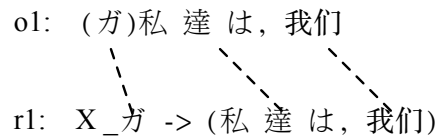
(x1:ガ) (x2:時間) (x3:ヲ) (x4:マデ) (x5:デ) 送って行きます (NULL), x2 x1 x5 送 x3 x4

# Method

- Transforming to SCFG

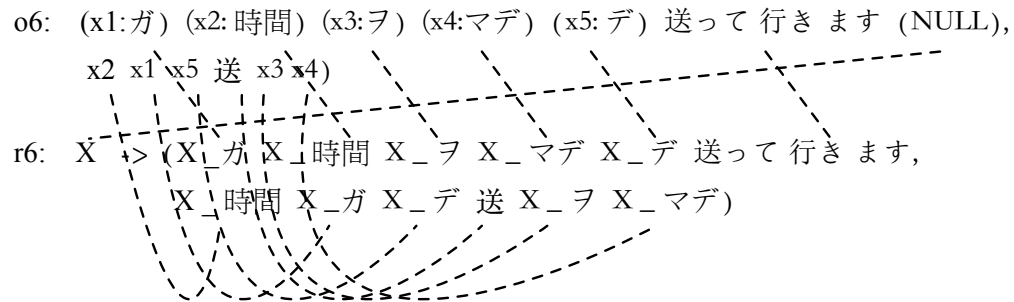
- examples

o1: (ガ)私 達は, 我们  
r1: X\_ガ -> (私 達は, 我们)



(a) the example of phrase rule transformation

o6: (x1:ガ) (x2:時間) (x3:ヲ) (x4:マデ) (x5:デ) 送って行きます (NULL),  
x2 x1 x5 送 x3 x4)  
r6: X\_ガ X\_時間 X\_ヲ X\_マデ X\_デ 送って行きます,  
X\_時間 X\_ガ X\_デ 送 X\_ヲ X\_マデ)



(b) the example of reordering rule transformation

# Outlook

- ◆ Motivation
- ◆ Case Frame
- ◆ Method
- ◆ Experiment
- ◆ Conclusion

# Experiment

- Experiment data
  - CWMT 2011 Japanese-Chinese Corpus (sentence pairs)
    - Training data: 280 thousand
    - Development data: 500
    - Testing data: 900
  - ASPEC-JC Corpus (sentence pairs)
    - Training data: 680 thousand
    - Development data: 2090
    - Testing data: 1800

# Experiment

- System
  - exp1: Strong hierarchical phrase-based system (baseline)
  - exp2: exp1 with case frame rules
  - exp3: exp1 with manually case frame rules
- Variables in rule are without distinction during decoding

# Experiment

- Experiment result

system	system	CWMT		ASPEC	
		# of Rules	BLEU(%)	# of Rules	BLEU(%)
exp1	baseline	14.7M	28.05	215.67M	27.17
exp2	exp1+cfr	14.7M+0.71M	28.36	215.67+7.21M	27.48
exp3	exp1+cfr	/	28.65	/	27.63

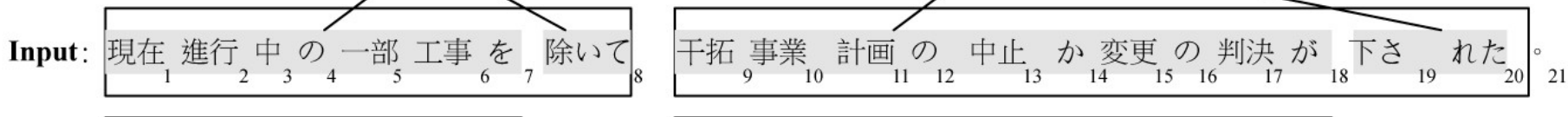
# Experiment

## • Examples

**Reference:** … 除了现在正在进行中的部分工程外 法院判决中止或变更开垦事业计划。

**Baseline:** … 现在进行中的中止工程外开垦事业计划或的部分变更的判决下达的。

**Hiero+CaseRule:** … 除了 现在进行中的中止工程 下达 开垦事业计划或的部分变更的判决。



<sup>ガ</sup>  
transformed case rule X -> ( X 除いて, 除了 X ) is applied in span [1, 8]

<sup>ヲ</sup>  
transformed case rule X -> ( X 下された, 下达 X ) is applied in span [9, 20]



# Experiment

## • Examples

**Reference** 北京市进一步通过依据郊区县县的医院的区域医疗中心的建设提高医疗服务的能力和水平

**Baseline** 北京市政府进一步郊区区县医院施政的区域医疗中心通过建设增加医疗服务的能力和水平

**Hiero+CFR** 北京市政府通过郊区区县医院依靠区域医疗中心的建设进一步提高了医疗服务的能力和水平

北京<sub>1</sub>市<sub>2</sub>は<sub>3</sub>さら<sub>4</sub>に<sub>5</sub>郊<sub>6</sub>外<sub>7</sub>区<sub>8</sub>県<sub>9</sub>の<sub>10</sub>病<sub>11</sub>院<sub>12</sub>に<sub>13</sub>依<sub>14</sub>拠<sub>15</sub>す<sub>16</sub>る<sub>17</sub>区<sub>18</sub>域<sub>19</sub>医<sub>20</sub>療<sub>21</sub>セ<sub>22</sub>ン<sub>23</sub>タ<sub>24</sub>ー<sub>25</sub>の<sub>26</sub>建<sub>27</sub>設<sub>28</sub>を<sub>29</sub>通<sub>30</sub>じ<sub>31</sub>て<sub>32</sub>医<sub>33</sub>療<sub>34</sub>サ<sub>35</sub>ー<sub>36</sub>ビ<sub>37</sub>ス<sub>38</sub>の<sub>39</sub>能<sub>40</sub>力<sub>41</sub>と<sub>42</sub>水<sub>43</sub>準<sub>44</sub>を<sub>45</sub>高<sub>46</sub>め<sub>47</sub>た<sub>48</sub>

transformed case rule  $X \rightarrow (X1 X2 X3 \text{通じて}, X1 \text{通过} X3 X2)$  is applied in span [1, 17]

transformed case rule  $X \rightarrow (X \text{高めた}, \text{提高了} X)$  is applied in span [18, 23]

# Outlook

- ◆ Motivation
- ◆ Case Frame
- ◆ Method
- ◆ Experiment
- ◆ Conclusion

# Conclusion

- ◆ This paper presented an approach to improve HPB model systems by augmenting the SCFG with Japanese CFRs.
- ◆ The CF are used to introduce new linguistically sensible hypotheses into the translation search space while maintaining the Hiero robustness qualities and avoiding computational explosions.
- ◆ We obtain significant improvements over a strong HPB baseline in the Japanese-to-Chinese task.

# Future work

- Soft/hard constraints on case frame rule matching
- Challenge to resolve the problem of tense and aspect etc.

Q & A