

# Word Order Does NOT Differ Significantly Between Chinese and Japanese

*Chenchen Ding<sup>1,2</sup>*

*Masao Utiyama<sup>1</sup>*

*Eiichiro Sumita<sup>1</sup>*

*Mikio Yamamoto<sup>2</sup>*

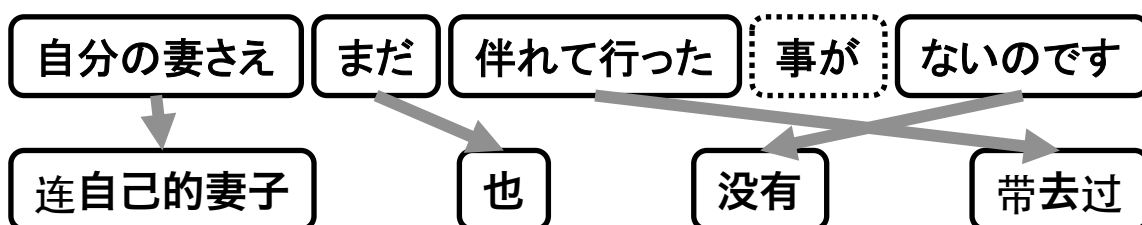
*<sup>1</sup>NICT, <sup>2</sup>Univ. of Tsukuba*

# Pre-reordering

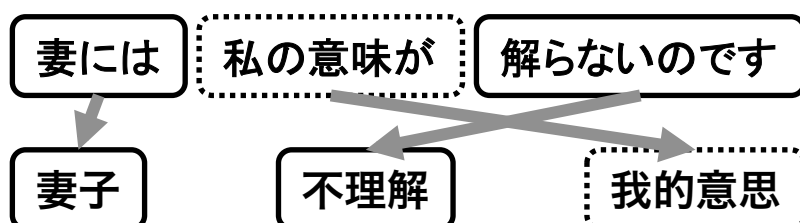
- For SVO-to-SOV
  - Dependency-based
    - English-to-SOV [Xu+ 2009]
  - HPSG-based head finalization
    - English-to-Japanese [Isozaki+ 2012]
    - Chinese-to-Japanese [Han+ 2012]
- For SOV-to-SVO
  - Japanese-to-English
    - [Komachi+ 2006]
    - [Katz-Brown & Collins 2008]
    - [Sudoh+ 2011]
    - [Hoshino+ 2013]
    - [Ding+ 2014]

# Chinese & Japanese

- Sharing more similar features
  - Than English-Japanese
- Conversely, Chinese ...
  - Avoids long attributes before nouns
    - Common in Japanese



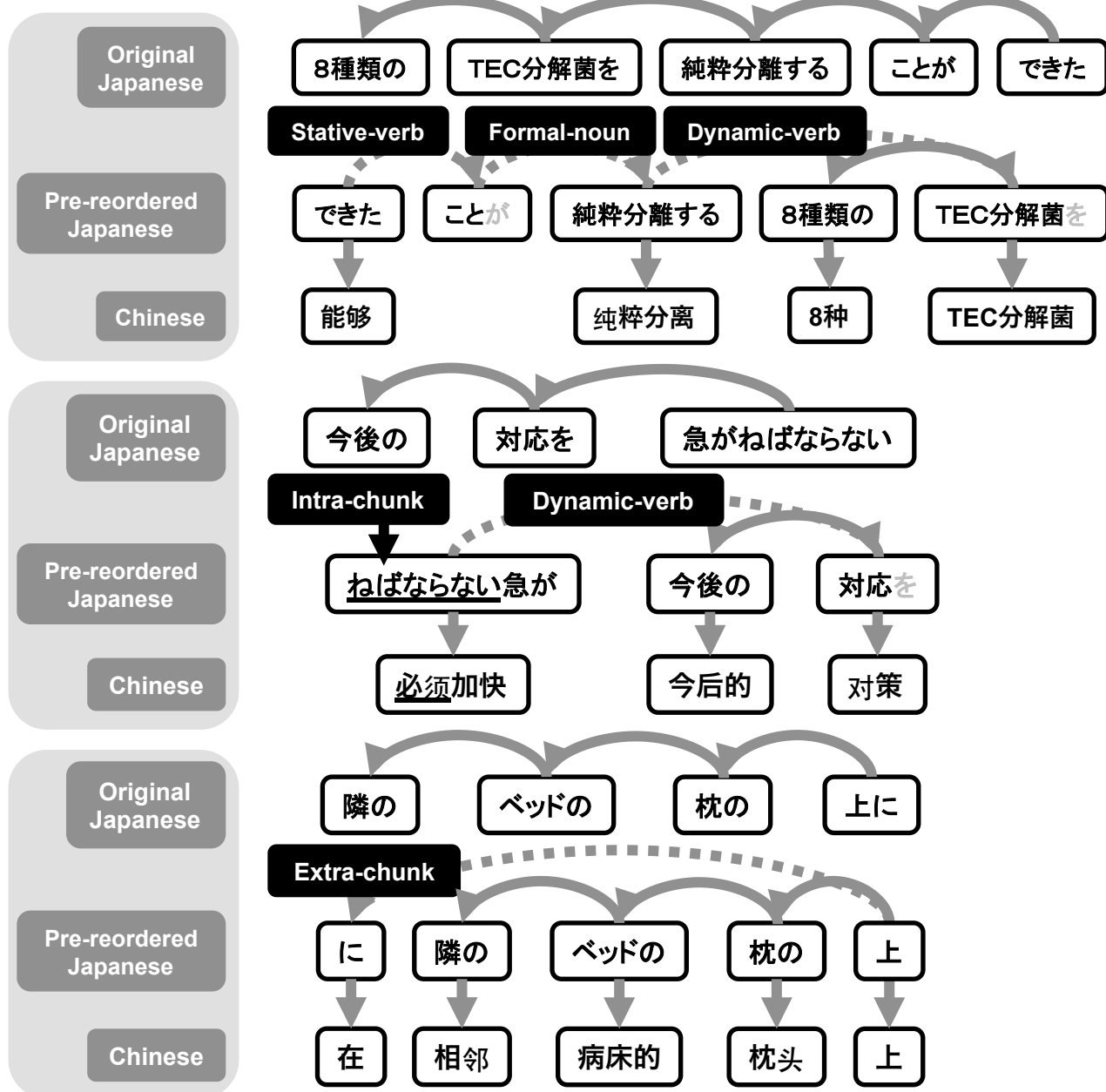
– Different case frame of Japanese



# Proposed Approach

- Three inter-chunk rules
  - Formal-noun rule
    - For “こと”
  - Stative-verb rule
    - For “いる”, “ある”, “できる”
  - Dynamic-verb rule
    - For the other verbs
- Intra-chunk rule
  - For auxiliary verbs, copula
- Extra-chunk rule
  - For noun suffixes (case marker, etc.)

# Pre-reordering Rules



# Experiment

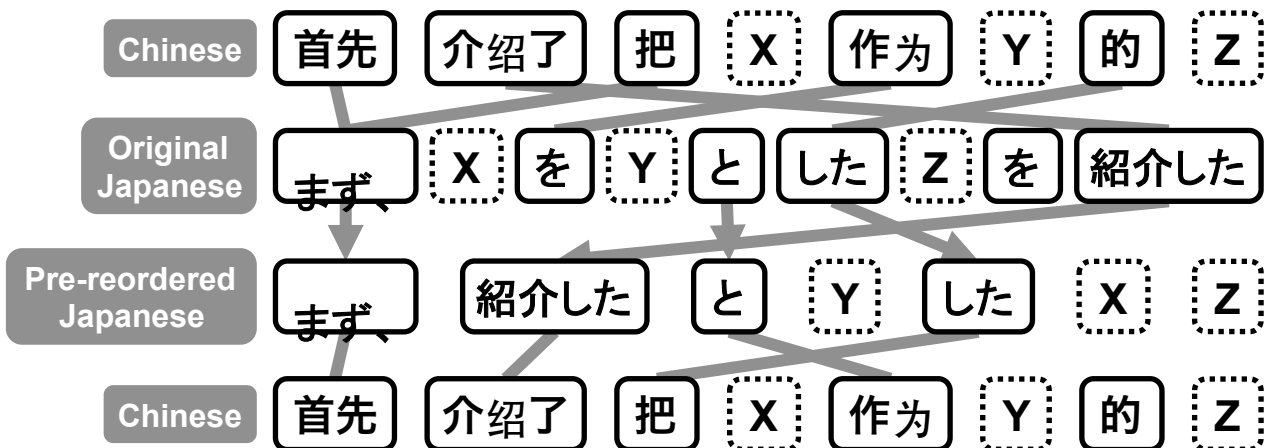
- ASPEC-JC [Nakazawa+ 2014]
- Japanese
  - MeCab + CaboCha (IPA)
- Chinese
  - Stanford segmenter (CTB)
- SMT system
  - PB SMT of Moses
- Results
  - Slight gain on BLEU (approx. +0.2)
  - Mediocre on RIBES (approx. +0.015)
  - +6.5 in human evaluation

# Discussion

- Chinese and Japanese
  - Relatively similar word order

→ Kendall's  $\tau$

Japanese	$\tau$
Original	.847
Pre-reordered	.865



- Proposed approach
  - Rigidity → more flexibility needed

# Conclusion & Future Work

- Chinese and Japanese, actually...
  - Have fairly similar word order

## → Reordering problem

- Not so serious as English-Japanese
- Difficult to handle by simple rules

## • Future work

- Investigate the case frame
  - More effective rules
- Alignment of function work
  - Not high quality in baseline