Integrating Empty Category Detection into Preordering Machine Translation

Shunsuke TAKENO [†], Masaaki NAGATA [‡], Kazuhide YAMAMOTO [†]

[†] Nagaoka University of Technology

[‡] NTT Communication Science Laboratory

NUT

NTT

PBSMT suffers from dropped words

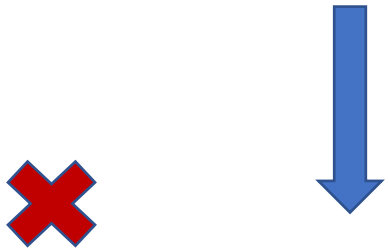
Iso-i de iru nn desu .

急い で いる ん です 。('m in a hurry .)

PBSMT suffers from dropped words

Iso-I de iru nn desu .

急い で いる ん です 。('m in a hurry .)



are in a hurry .

PBSMT suffers from dropped words

THIS WORK

Iso-I de iru nn desu .

pro 急い で いる ん です 。('m in a hurry .)

✖ are in a hurry .

⭕ I’m in a hurry .

PBSMT suffers from dropped words

THIS WORK

Iso-I de iru nn desu .

pro 急い で いる ん です 。('m in a hurry .)

Empty category

✖ are in a hurry .

⭕ I’m in a hurry .

Empty category detection (Takeno+2015)

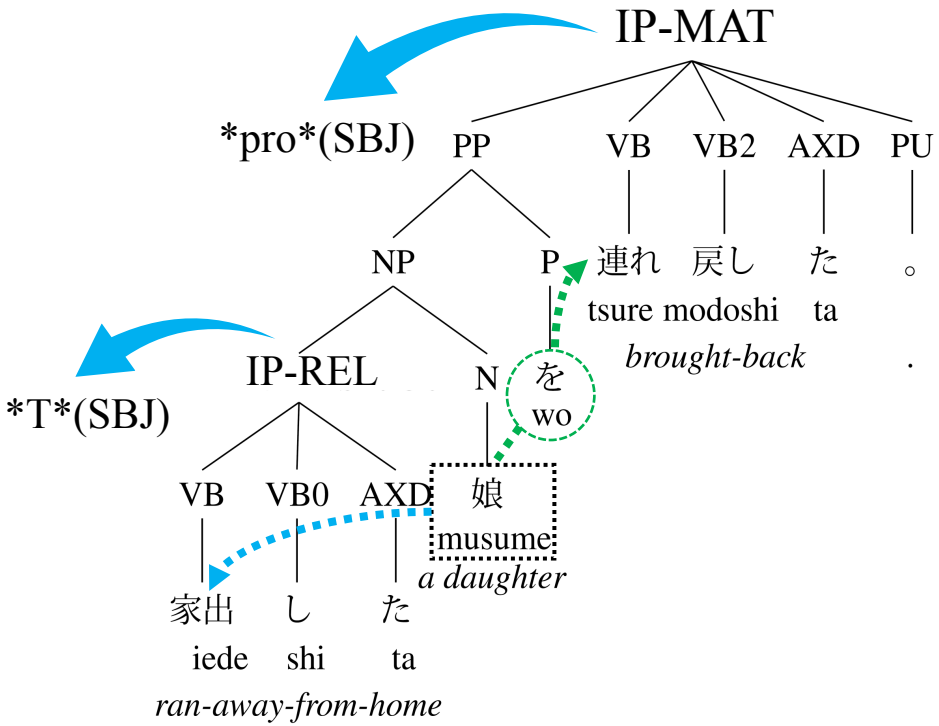
Empty categories(EC) are phonetically null but syntactically exists such as *dropped pronoun*(***pro***) and *trace* (***T***) of NP.

Previous work has built discriminative EC detection model as *classification problem* to each nodes using structural info.

Max-Entropy model for EC detection

$$P(e_1^n|T) = \prod_{i=1}^{n} P(e_i|e_1^{i-1}, T)$$

$$= \prod_{i=1}^{n} \frac{\exp(\boldsymbol{\theta} \cdot \boldsymbol{\phi}(e_i, e_1^{i-1}, T))}{Z(e_1^{i-1}, T)}$$



Problems on the integration of PBSMT

GENERAL IDEA:

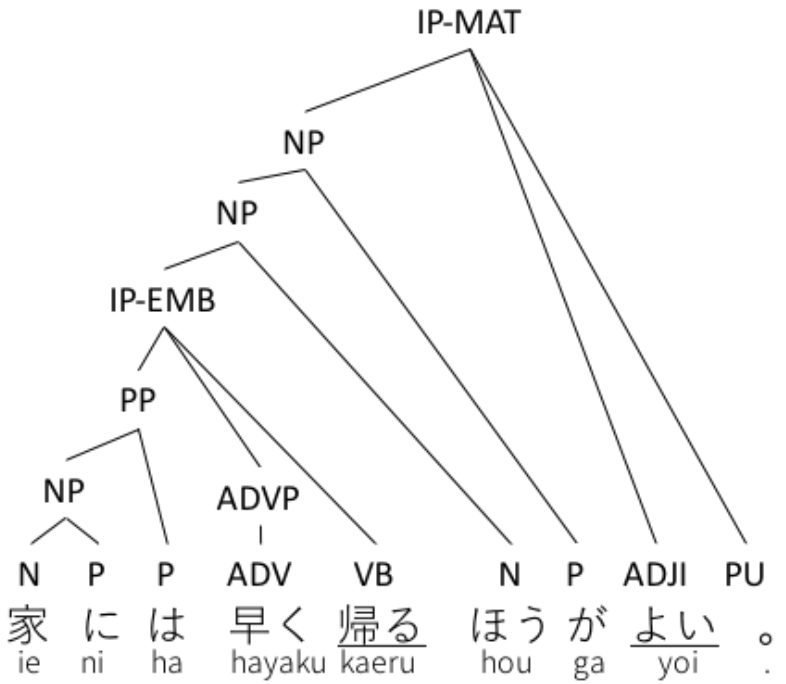
Simply insert detected empty categories as words

家 に は 早く 帰る ほう が よい 。
ie ni ha hayaku kaeru hou ga yoi .

Problems on the integration of PBSMT

GENERAL IDEA:

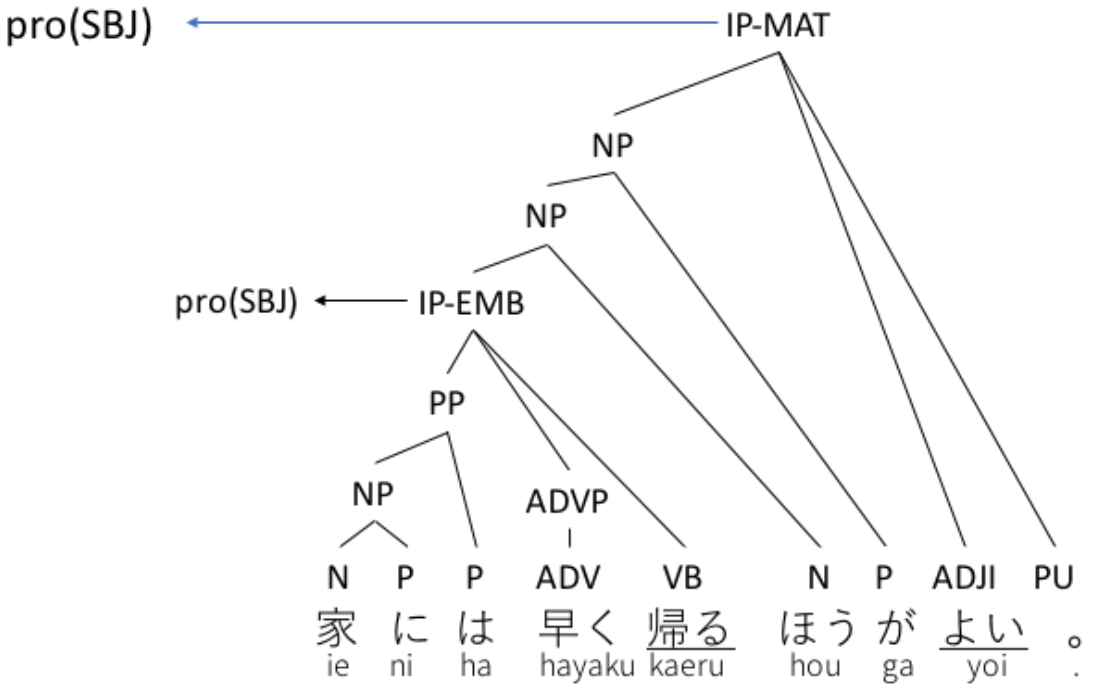
Simply insert detected empty categories as words



Problems on the integration of PBSMT

GENERAL IDEA:

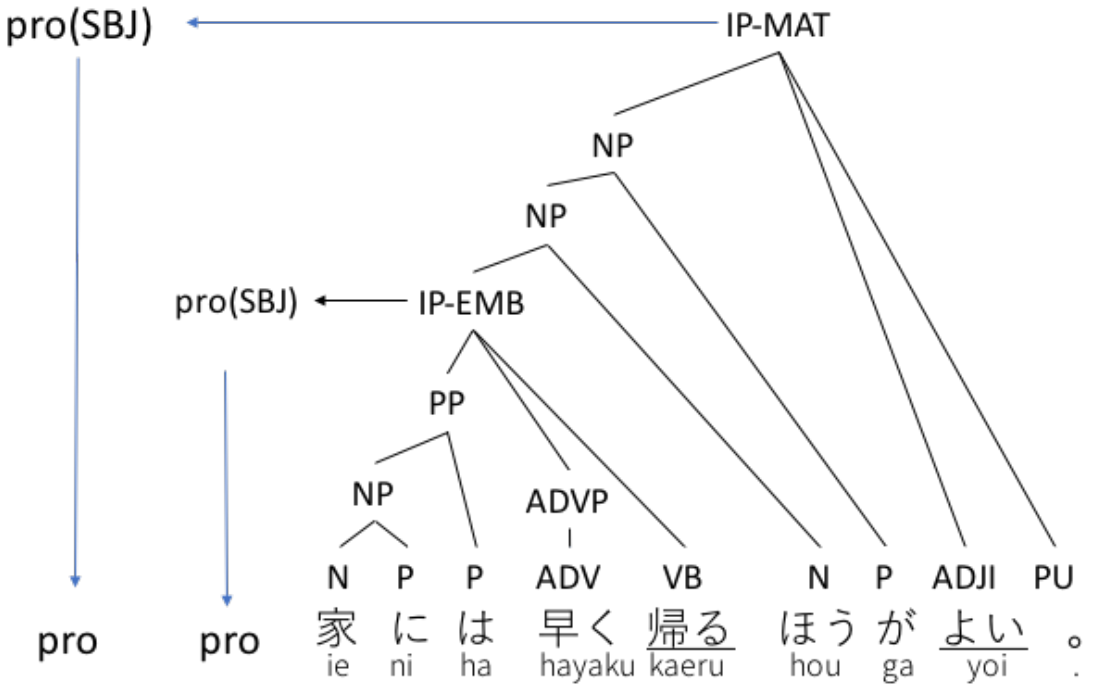
Simply insert detected empty categories as words



Problems on the integration of PBSMT

GENERAL IDEA:

Simply insert detected empty categories as words

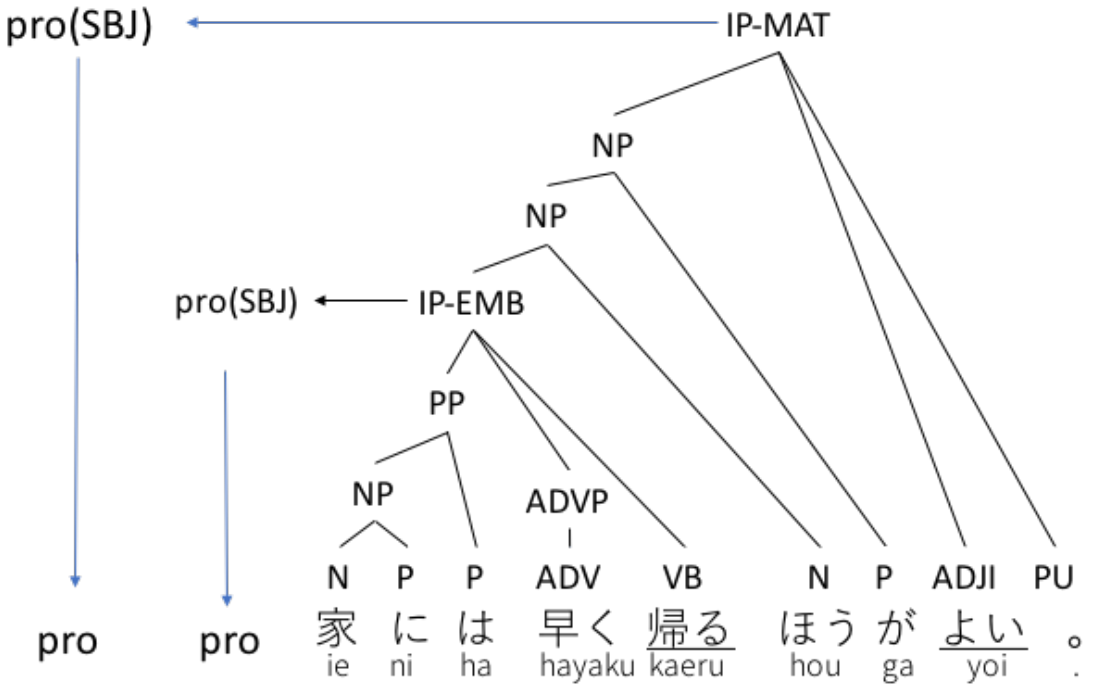


Problems on the integration of PBSMT

GENERAL IDEA:

Simply insert detected empty categories as words

Little improvement on machine translation(BLEU: 33.1 → 33.6)



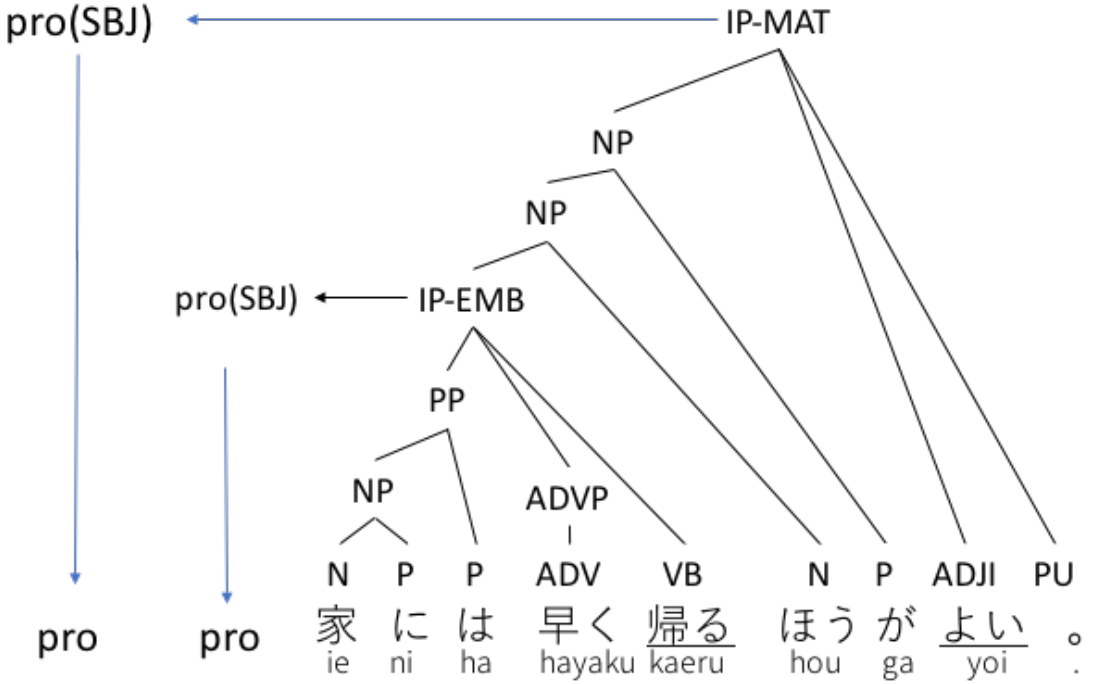
Problems on the integration of PBSMT

GENERAL IDEA:

Simply insert detected empty categories as words

Little improvement on machine translation(BLEU: 33.1 → 33.6)

Why?



Problems on the integration of PBSMT

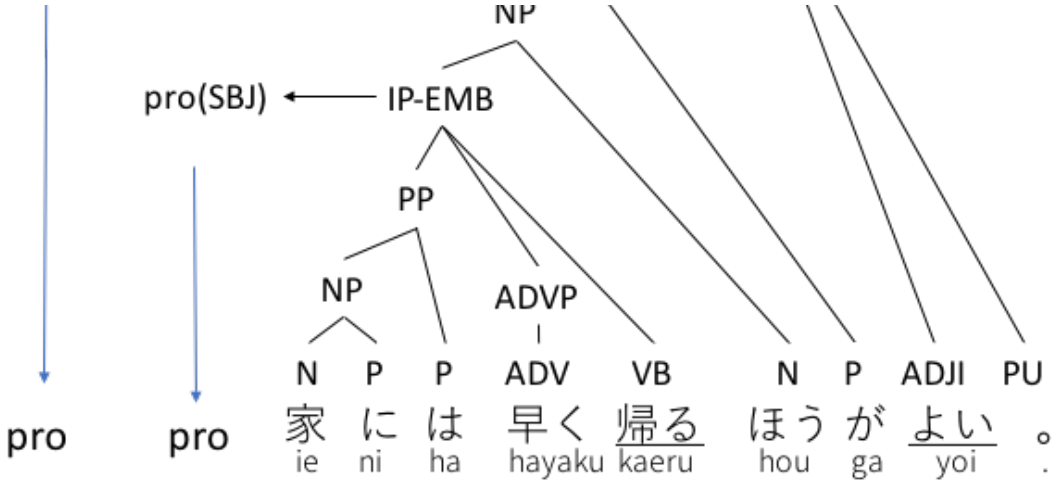
GENERAL IDEA:

Simply insert detected empty categories as words

Little improvement on machine translation(BLEU: 33.1 → 33.6)

Why?

1. Noisy empty category detection
2. Word order problem w.r.t empty categories



Problem1: Noisy EC detection

Suffering from imprecise prediction due to insufficient accuracy of parser

- **Solution: Eliminating** empty categories detected whose confidence is under the threshold(from dev).

Logistic regression model for EC detection

$$P(e_1^n|T) = \prod_{i=1}^{n} P(e_i|e_1^{i-1}, T)$$

$$= \prod_{i=1}^{n} \frac{\exp(\boldsymbol{\theta} \cdot \boldsymbol{\phi}(e_i, e_1^{i-1}, T))}{Z(e_1^{i-1}, T)}$$

types	INPUT	P	R	F
pro	GOLDEN	74.3	75.6	74.9
T	GOLDEN	89.0	95.0	91.9
pro	SYSTEM	60.9	66.2	63.4
T	SYSTEM	50.0	42.2	45.8

Problem2: Word order problem w.r.t ECs

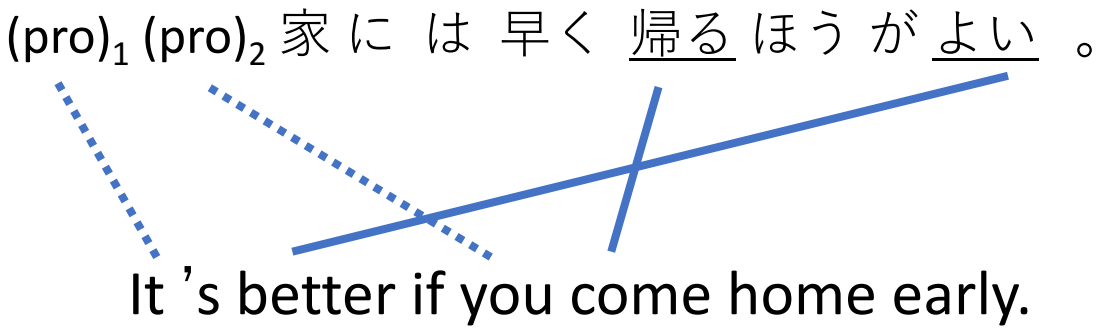
Word order w/ EC in source sentence is different from the word order in target side

$(pro)_1$ $(pro)_2$ 家 に は 早く 帰る ほう が よい 。

It ’s better if you come home early.

Problem2: Word order problem w.r.t ECs

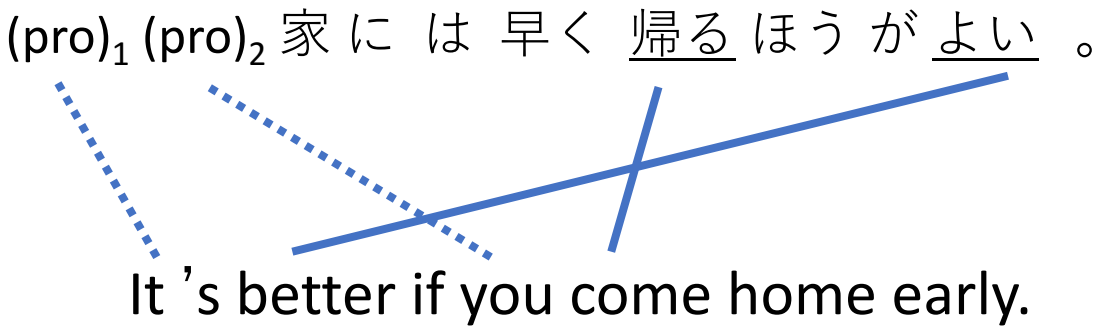
Word order w/ EC in source sentence is different from the word order in target side



Problem2: Word order problem w.r.t ECs

Word order w/ EC in source sentence is different from the word order in target side

Solution: **Preordering model w/ EC**



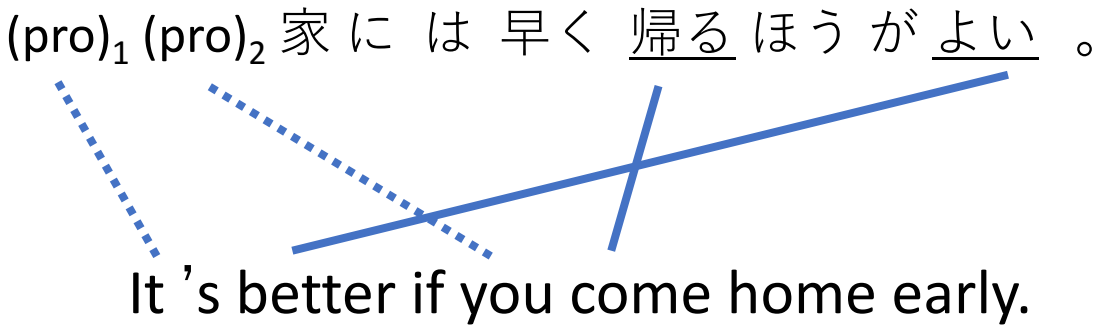
Problem2: Word order problem w.r.t ECs

Word order w/ EC in source sentence is different from the word order in target side

Solution: **Preordering model w/ EC**

(Hoshino+2015) train discriminative model from **word alignment**

Swap the nodes so that maximize Kendall distance between sentences



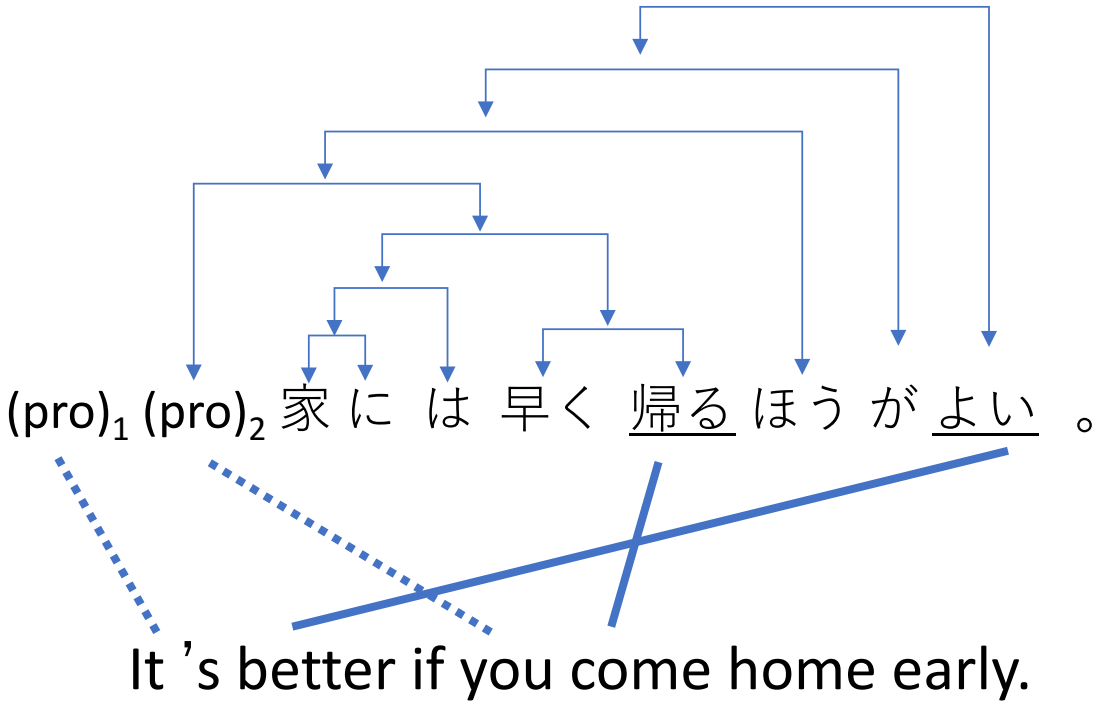
Problem2: Word order problem w.r.t ECs

Word order w/ EC in source sentence is different from the word order in target side

Solution: **Preordering model w/ EC**

(Hoshino+2015) train discriminative model from **word alignment**

Swap the nodes so that maximize Kendall distance between sentences



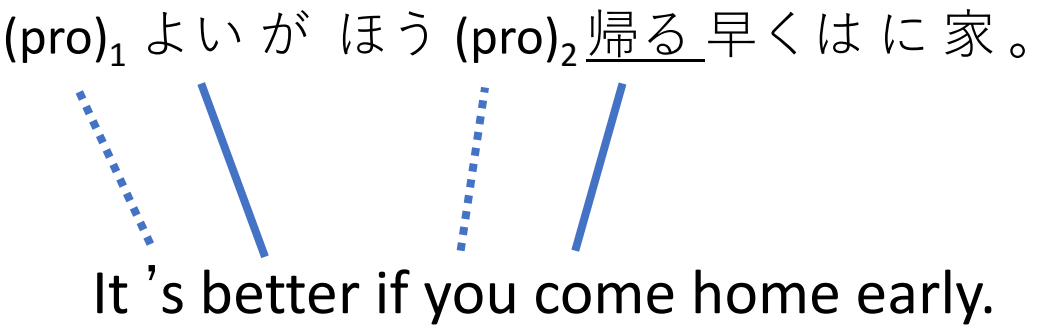
Problem2: Word order problem w.r.t ECs

Word order w/ EC in source sentence is different from the word order in target side

Solution: **Preordering model w/ EC**

(Hoshino+2015) train discriminative model from **word alignment**

Swap the nodes so that maximize Kendall distance between sentences



Problem2: Word order problem w.r.t ECs

Word order w/ EC in source sentence is different from the word order in target side

Solution: **Preordering model w/ EC**

(Hoshino+2015) train discriminative model from **word alignment**

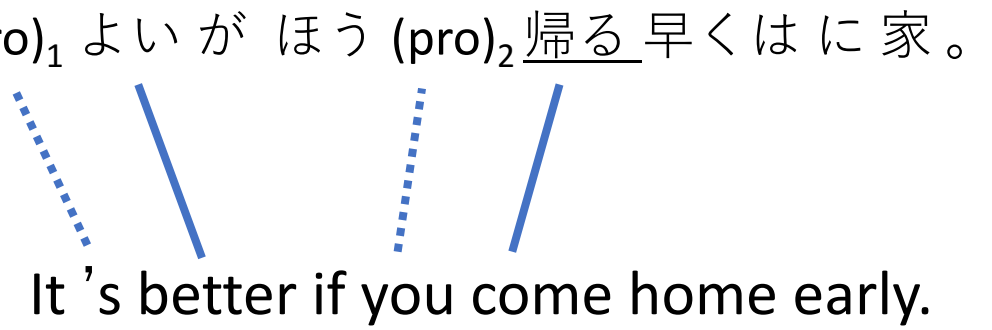
Swap the nodes so that maximize Kendall distance between sentences

Chicken and Egg problem:

Can we build preordering model from automatic word alignment?

REORDERING(C) : Automatic word alignment w/ EC

or REORDERING(H) : Manual word alignment w/o EC



Experiment - settings

DATASETS:

- *IWSLT 2005 JE Translation Task* (19,980 sent.)
  - **Small size** but **many empty categories** (Spoken language corpus)
- *The Kyoto Free Translation Task* (KFTT) (440,000 sent.)
  - **Medium size** but **fewer empty categories** (Written language corpus)

METRICS: BLEU, RIBES

MODELS:

Evaluating each model w/ or w/o EC

- **BASELINE**:
  Plain translation model (Moses)
- **REODERING(C):**
  Built from **automatic word alignments (i.e GIZA++)**
  The word alignment include **EC alignment (EC as *known words*)**
- **REODERING(H)**: (～ 5,319 sent. pairs)
  Built from **manual word alignments**
  The word alignment **doesn't include EC alignment (EC as *unknown words*)**

Experiment – Results for Spoken Language

- Plain insertion(BASELINE) yields **only slight improvement**
- Preordering with EC detection yields **much improvements**

	BLEU		RIBES	
	w/o EC	w/ EC	w/o EC	w/ EC
BASELINE	33.1	33.6 (+0.5)	74.2	75.7 (+1.5)
REORDERING(C)	33.2 (+0.1)	**34.3** (**+1.2**)	76.3 (+2.1)	**78.8** (**+4.6**)
REORDERING(H)	33.8	34.1	76.8	78.6

Experiment – Results for Written Language

- EC detection has little effect on KFTT
  - Difficulty of EC detection in longer sentence(~ 24 words in src.)
  - Frequently confusing person information(*pro* <–> *it* or *he*)

	BLEU		RIBES	
	w/o EC	w/ EC	w/o EC	w/ EC
BASELINE	18.5	18.6 (+0.1)	62.4	62.5 (+0.1)
REORDERING(C)	19.3 (**+1.4**)	19.8 (+0.3)	64.8 (**+2.8**)	65.2 (+0.3)
REORDERING(H)	19.9	**20.2**	65.2	**65.5**

Conclusion

- Propose the integration of EC detection into PBSMT
  - **Preordering model** alleviate the word order problem w/ EC
    - Plain insertion of EC slightly improve due to ***word order*** problem including ECs
      Word alignments about EC are needed for building the model
  - **Elimination of unreliable ECs** refines EC detection
    - Accuracy of structural parse is insufficient for practical usage
    - Cutting lower confidence of ECs alleviate the problem
- Future works:
  Recovering linguistic information of EC
  i.e. person, animacy or gender

Experiment – Results for Written Language

- KFTT Evaluation (Written language)
- EC detection has little effect on KFTT
  - Difficulty of EC detection in longer sentence(~ 24 words in src.)
  - Frequently confusing person information(*pro* <–> *it* or *he*)

	BLEU		RIBES	
	w/o EC	w/ EC	w/o EC	w/ EC
BASELINE	18.5	18.6 (+0.1)	66.4	65.4 (-1.0)
REORDERING(C)	19.3	19.8(**+0.5**)	65.7	66.0 (**+**0.3)
REORDERING(H)	19.9	**20.2**(+0.3)	66.2	**66.3**(+0.1)

Experiment – Results for Spoken Language

- IWSLT 2005 JE Task Evaluation (Spoken language)
- Plain insertion slightly improve the result.
- Combining preordering with EC detection yields much improvements.

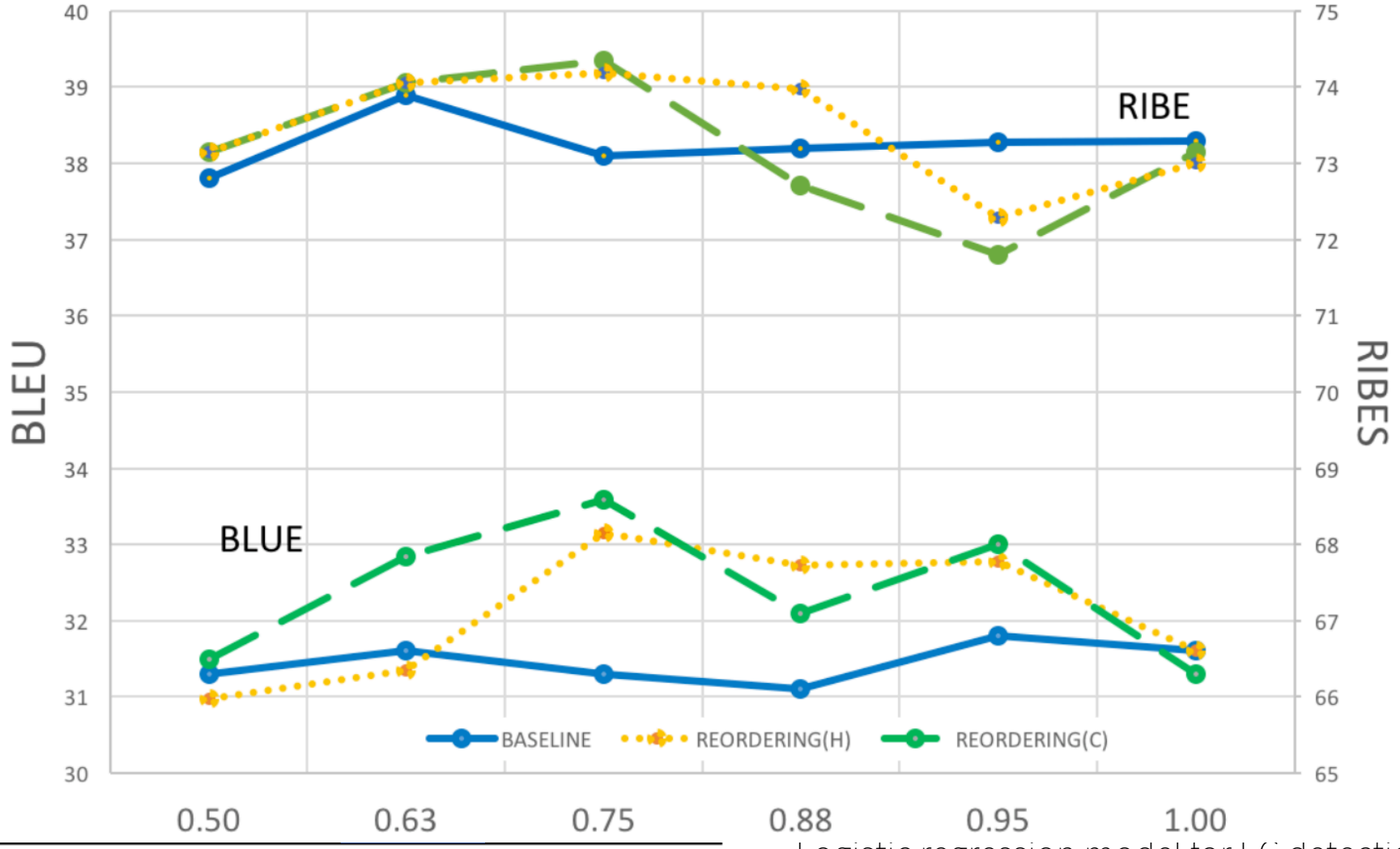
	BLEU		RIBES	
	w/o EC	w/ EC	w/o EC	w/ EC
BASELINE	33.1	33.6 (+0.5)	74.2	75.7 (+1.5)
REORDERING(C)	33.2	**34.3 (+1.1)**	76.3	**78.8 (+2.5)**
REORDERING(H)	33.8	34.1 (+0.3)	76.8	78.6 (+1.8)

Success translation	
Reference	i 'm in a hurry .
Source(EC)	*pro* 急い で いる ん です 。
NO EC	are in a hurry .
ECs	i 'm in a hurry .
Reference	how much to rent it for three days ?
Source	*pro* 三 日間 借り る と いくら に なり ます か 。
Reordered Source	*pro* いくら に ます なり と 借り る 三 日間 か 。
NO ECs	i have a three days and how much will it be ?
ECs	i have a three days and how much will it be ?
Pre-ordered w/o EC	what would you like to hire and three days .
Pre-ordered w/ EC	how much will it cost to three days ?

Failed translation	
Reference	do you have any fruits or plants ?
Source	*pro* 果物 や 植物 を 持っ て い ます か 。
Reordered Source	*pro* い て 持っ ます 果物 や 植物 を か 。
NO ECs	i have a carrying any plants and fruits ?
ECs	i have fruit or plant ?
Pre-ordered w/o EC	do you have some fruit or plants ?
Pre-ordered w/ EC	i have a carrying any plants and fruits ?

Problem2 : inaccurate EC detection

- Reordering empty categories



Logistic regression model for EC detection

Resources and Tools

- Shift-reduce Phrase structure parser : (Hayashi+2015)
- Tokenization: MeCab
- EC detection : (Takeno+2015)
- Lower casing

Experiment – Results for Spoken Language

	BLEU		RIBES	
	w/o EC	w/ EC	w/o EC	w/ EC
BASELINE	33.1	33.6 (+0.5)	74.2	75.7 (+1.5)
REORDERING(C)	33.2	**34.3 (+1.1)**	76.3	**78.8 (+2.5)**
REORDERING(H)	33.8	34.1 (+0.3)	76.8	78.6 (+1.8)

- Plain insertion(BASELINE) improves both scores only slightly
- Combining preordering with EC detection show much improvement

Workflow of Our proposals

- Reordering empty categories almost solve the word order problem.
- As practical problem of Empty Category Detection, We suffer from unstable prediction due to insufficient accuracy of parser.
- We need to eliminate empty categories detected whose confidence is under the *threshold*(from dev).

types	INPUT	P	R	F
pro	GOLDEN	74.3	75.6	74.9
T	GOLDEN	89.0	95.0	91.9
pro	SYSTEM	60.9	66.2	63.4
T	SYSTEM	50.0	42.2	45.8

$$P(e_1^n|T) = \prod_{i=1}^{n} P(e_i|e_1^{i-1}, T)$$

$$= \prod_{i=1}^{n} \frac{\exp(\boldsymbol{\theta} \cdot \boldsymbol{\phi}(e_i, e_1^{i-1}, T))}{Z(e_1^{i-1}, T)}$$

Problem on the integration of PBSMT

INTUNITIVE IDEA:

Simply insert detected empty categories as words