

Predicate-Argument Structure-based Preordering for Japanese-English Translation of Scientific Papers

Kenichi Ohwada, Ryosuke Miyazaki and Mamoru Komachi
 Graduate School of System Design, Tokyo Metropolitan University

Preordering Method between Japanese and English

Japanese : 1. Head-final 2. SOV 3. Postposition English : 1. Head-initial 2. SVO 3. Preposition

↓ Preordering

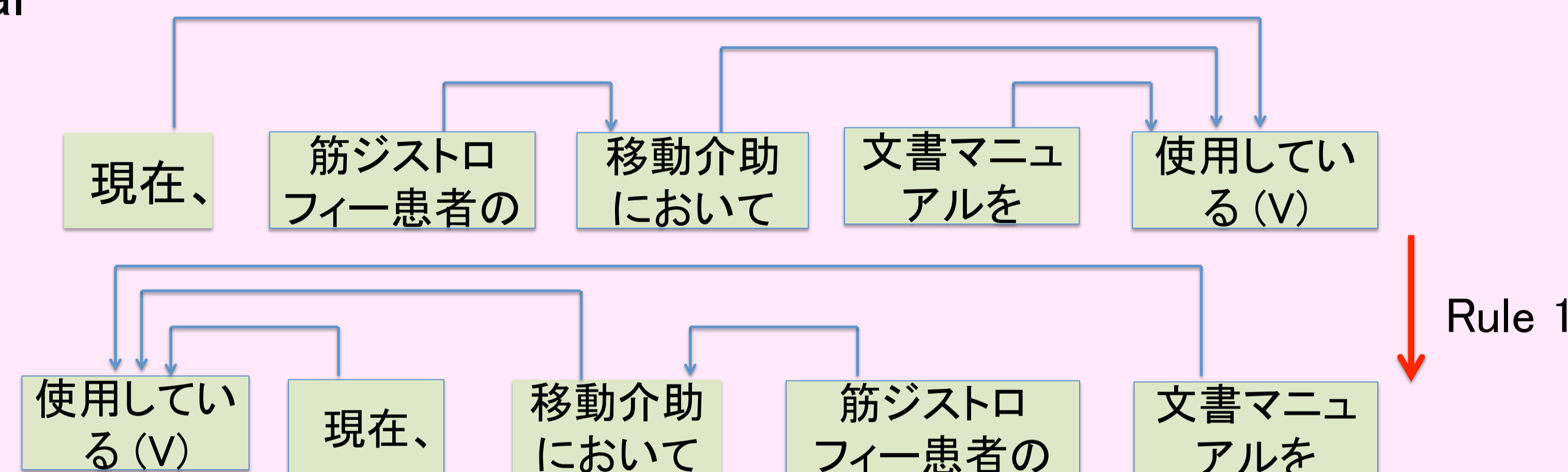
Improvement of long distance word alignment and performance of machine translation

Previous work

OHoshino et al., (2013)

• Rules of sentence-level

Rule 1. Transform a dependency tree from Head-final to Head-initial



Rule 2. Transfer of a predicate of a sentence to make SVO

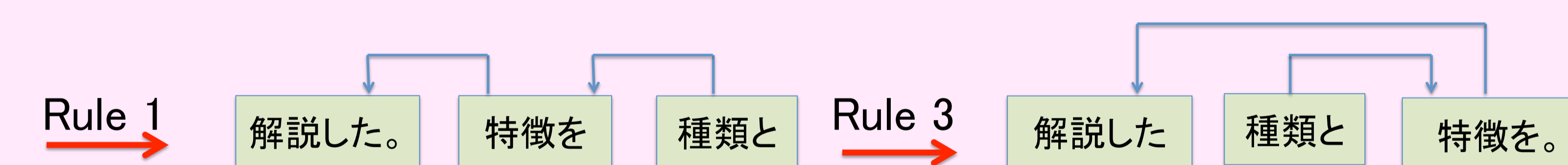
If there's a subject in sentence → just after it

Else if there's an object → just before it

Else → just before the predicate's rightmost dependent

(We used a predicate-argument structure analyzer and judged case of が (ga) as subject and case of を (wo), に (ni) as object.)

Rule 3. Correct coordinate expressions and punctuations



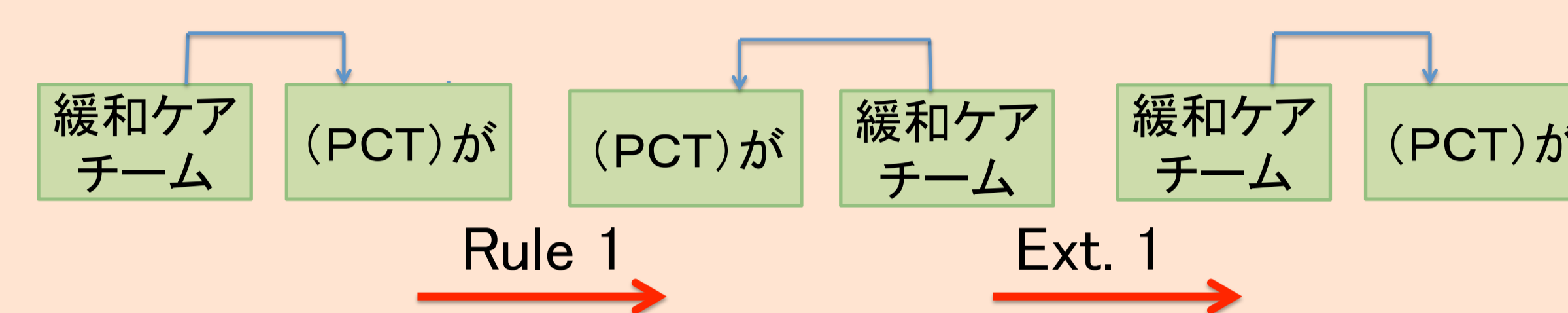
• Rule of phrase (Bunsetsu)-level

Rule 4. Reverse content words and function words of each phrase.

Extension of rules that we propose

Ext.1. Parenthesis restoration

As a result of Rule 1, parenthesis expressions is reversed, so we modify the rule to restore them.

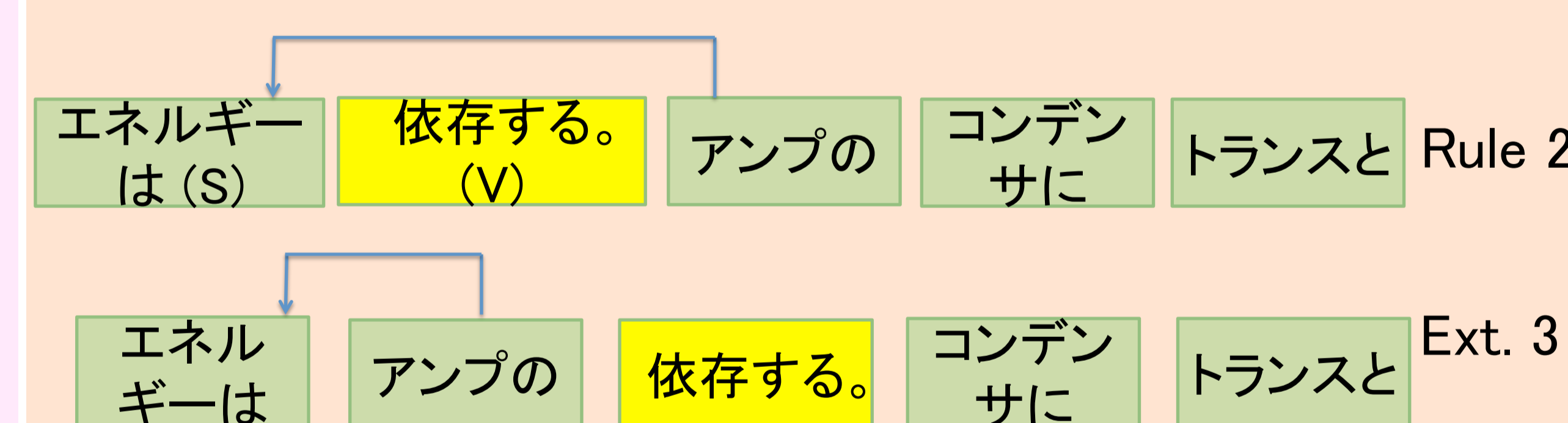


Ext. 2. Passive voice preordering

When there's no subject in a Japanese sentence, we move a predicate to the end of the sentence because a corresponding English sentence in the data of this task is passive voice in many cases.

Ext. 3. Subjective phrase preservation

A predicate is sometimes inserted in between the subject and its modifiers by Rule 2, so we change the rule to the one that the predicate comes after the subjective phrase (In this case, phrase doesn't mean Bunsetsu).



Example of Preordering

ダイナミックミキシング法 | (DM法)による | TiN膜生成技術を |
the dynamic mixing method (DM) by TiN film generation technique
 開発した。 ———— predicate

was developed.

Rule. 1 ↓ The order of “ダイナミックミキシング法” and “(DM法)による” is modified by Ext. 1.

開発した。 | TiN膜生成技術を | ダイナミックミキシング法 | (DM法)による

Rule. 2 ↓ Since there is not a subject, the predicate “開発した” is moved to the end by Ext. 2.

(Rule. 3) is skipped because there are no coordinate expressions and a full stop in an inappropriate position.

TiN膜生成技術を | ダイナミックミキシング法 | (DM法)による | 開発した。

Rule. 4 and minor adjustment ↓ Function word “による” is moved over phrase, because the phrase is parenthesis expression.

をTiN膜生成技術 | によるダイナミックミキシング法 | (DM法) |

た開発し。

TiN film generation technique by the dynamic mixing (DM) method

was developed.

Experimental setting and results

- Training data: 1 million sentences
- Baseline : SRILM 1.7.0, GIZA++ 1.0.7, Moses 2.1.1
- PAS Analyzer : Syncha 0.3 (Mecab 0.996 IPADic 2.7.0)
- Distortion limit : 6 (default setting)

Method	BLEU	RIBES
Phrase-based SMT Baseline (Hoshino et al., 2013)	15.74	0.620162
Proposed Method – Ext. 1	15.73	0.652461
Proposed Method – Ext. 2	15.93	0.654454
Proposed Method – Ext. 3	15.88	0.650964
Proposed Method	16.02	0.654600

Discussion

- Our Proposed Method outperforms our re-implementation of (Hoshino et al., 2013) and the baseline, but the impact on the translation quality is not so large. We think it is because of the property of this data, so we want to try in different domains.
- In terms of RIBES, methods including (Hoshino et al., 2013) outperform the baseline, so we think the effectiveness of preordering is better reflected by RIBES.
- When we subtract three modifications one by one from proposed method, the parenthesis rule has the largest impact. On Japanese side, 16.8% of sentences have parenthesis expressions and about 0.74% of parenthesis expressions cross over phrases in the training data we use.
- The Passive voice modification doesn't have much impact, we think it is because the rate of no-subject sentence is 57% but one of passive voice in English side is not so high.