Description of KYOTO EBMT System in PatentMT at NTCIR-10

Toshiaki Nakazawa,   Sadao Kurohashi

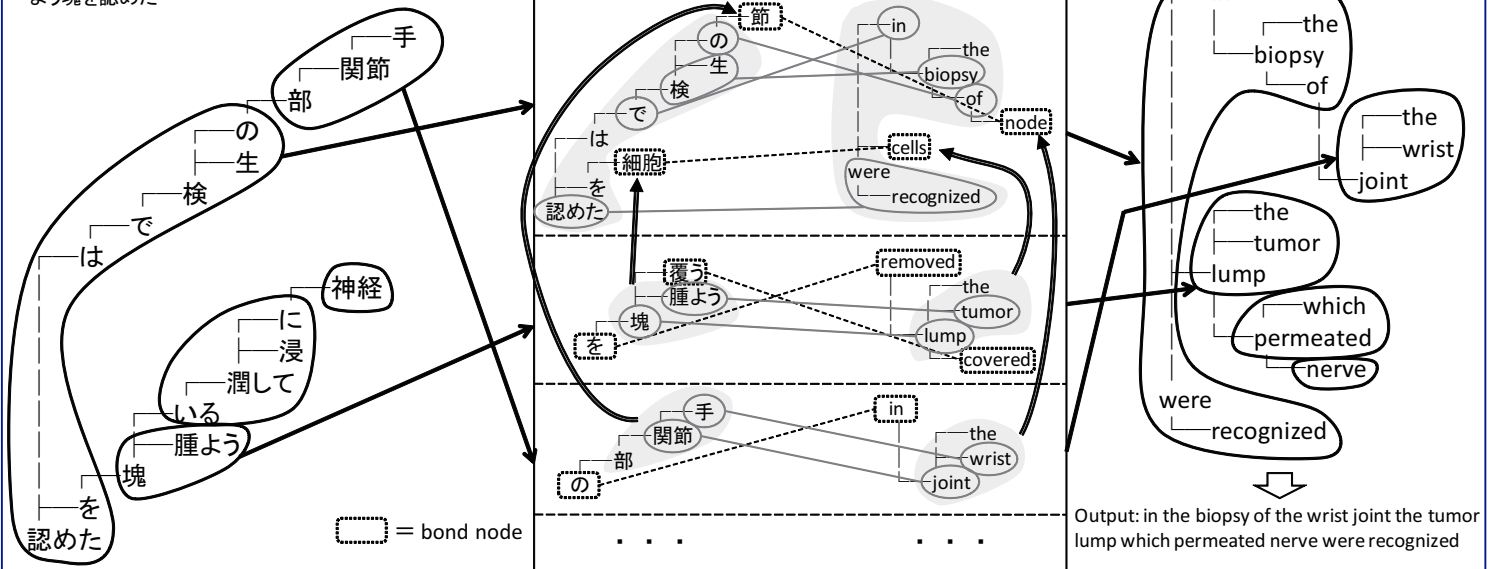nakazawa@pa.jst.jp   kuro@i.kyoto-u.ac.jp

Graduate School of Informatics, Kyoto University

System Description

Input: 手関節部の生検では神経に浸潤している腫よう塊を認めた

Translation Example Database

[ ] = bond node

Output: in the biopsy of the wrist joint the tumor lump which permeated nerve were recognized

## Alignment Model [Nakazawa+, 2012]

· Two types of function words

Type 1: counterparts exist in the other side
  less problematic, can be handled as content words

Type 2: counterparts do not exist in the other side
  problematic, we call them "*unique* function words"

· Proposed alignment model

bilingually generate treelet pairs for content words and Type 1 function words

monolingually derive Type 2, unique function words
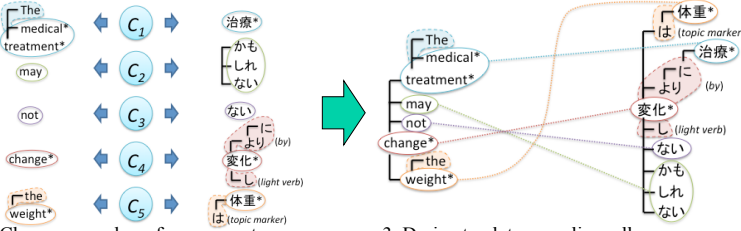
1. Choose a number of components
2. Generate each of <u>core</u> treelet pairs bilingually
3. Derive treelets monolingually
4. Combine treelets so as to create sentences

Special Treatments for Patent Translation

1. Japanese lexicon acquisition

Using the method of [Murawaki+, 2008], we acquired technical terms which include 2 adjectives, 37 verbs and 830 nouns

Lexicon	POS	Form
不溶だ	adjective	ナ形容詞
嵌込む	verb	子音動詞マ行
組付ける	verb	母音動詞
移載	verbal noun	N/A
支持桿	noun	N/A

2. English compound noun extraction

To avoid parsing errors for English compound nouns, which often contain a verb, we automatically extract English compound nouns using the alignment results and Japanese parsing results, because Japanese compound nouns are relatively easy to detect and Japanese parsers correctly analyze the compound nouns.

3. PP-attachment modification

English PPs often have ambiguities on choosing their parent phrase and cause both alignment and translation errors. We modified the parent of English PP again using the alignment results and Japanese parsing results so as to make the dependency relations similar to those of Japanese sentences.

NTCIR-10 PatentMT Results

Intrinsic evaluation result

	Japanese-to-English			English-to-Japanese		
	RIBES	BLEU	Adeq.	RIBES	BLEU	Adeq.
Moses	69.72	28.56	2.81	72.31	32.98	2.69
KYOTO	67.24	24.01	2.74	72.52	26.85	2.50

Chronological evaluation result

	Japanese-to-English		English-to-Japanese	
	RIBES	BLEU	RIBES	BLEU
baseline	69.60	23.41	72.77	25.47
+Ja lexicon	69.48	24.17	73.03	25.80
+En compound noun	69.53	23.89	73.39	25.80
+En PP modification	69.54	23.84	73.39	25.72

Effect of special treatments (tested on intrinsic test set)

	Japanese-to-English		English-to-Japanese	
	RIBES	BLEU	RIBES	BLEU
Moses	70.07	28.47	72.44	32.10
NTCIR-9	65.15	21.49	66.11	24.59
NTCIR-10	69.18	24.65	72.35	26.52

Translation Samples

Input: 図15は、本NA変換光学素子を回動部をもたない照明装置に適用した例を示している。

Output: FIG. 15 shows an example in which the present NA converting optical element is applied to an illumination unit which is not equipped with the rotational section .

Input: 冷却水入口50には入口用配管ニップル54が取り付けられ，冷却水出口52には出口用配管ニップル56が取り付けられる（図3も参照）。

Output: A cooling water inlet 50 is attached with an entrance piping nipple 54 , and a cooling water outlet 52 is fitted with an outlet piping nipple 56 (see also FIG. 3) .

Input: If the circuit shown in FIG. 1 is used in a solid-state imaging apparatus, the analog-to-digital converter will suffer a drop in conversion accuracy.

Output: また、図1に示す回路を固体撮像装置に用いた場合，アナログ／ディジタル変換器が変換精度の低下を受ける。

Input: In this respect, the total volume is more preferably equal to or less than 390 mm3, and particularly preferably equal to or less than 380 mm3.

Output: この観点から総容積は390mm3以下がより好ましく380mm3以下が特に好ましい。