

名詞にかかる形容詞を対象とした Super-Function の拡張

桂 康[†] 松本 和幸[†] 任 福継[†]

[†]徳島大学

1 はじめに

インターネットの普及により、世界中から必要な情報を容易に集めることができるようになったことで、日本人も英語に接する機会が増えてきた。しかし英語を不得意とする日本人が少なくないため、英文を読解するために翻訳が必要となってくる。

英語が苦手な人が日本語を英語に翻訳する時、最もよく利用されているのがエキサイト翻訳 [1] などの機械翻訳システムだろう。しかし、これらの機械翻訳システムには“品質が良くない”、“長文の翻訳精度が悪い”といった問題がある。

そこで、この問題を解決するための手法として Super-Function の理論に基づく翻訳手法を用いる。この手法は自然な訳文を得られやすく、長文にも強いといった特徴がある。

本稿では、Super-Function[2](以下 SF) に基づく機械翻訳を拡張するために、名詞にかかる形容詞を対象とした研究を行った。名詞にかかる形容詞を名詞と合わせることで定数が減り、重複した SF を削除することで SF データベースの圧縮が可能になる。また、変数の応用の幅が広がり、多様な文に対応できる。

2 Super-Function

SF とは、任意の一部分を変数とし残りを定数と定義した場合、変数だけを入れ替えることにより、元の構造を保ったまま別の事象に変換できる関数である。

SF に基づいて機械翻訳を行う場合、SF を原言語と目標言語 (ここでは日本語と英語) の対応を示す関数と定義する。それぞれの言語は変数と定数から成ると定義する。変数を名詞、それ以外を定数とした場合、図 1 のような関係が得られる。図の円は定数、矢印は変数を表す。図 1 の定数と、変数の対応を示したのが表 1 である。

変数には位置と制約条件が記述されている。図および表の ϕ は空文字を示す。文の最初、および変数の間に定数が存在しない場合に用いられる。制約条件はそれぞれ一人称主格、冠詞、指定なしを示す。

SF に基づく機械翻訳ではまず、入力文を形態素解

Expansion of Super-Function for Adjective to affect Noun

[†] Yasushi Katsura

[†] Kazuyuki Matsumoto

[†] Fujii Ren

Tokushima University (†)

析し定数と変数に分類する。そして定数と、変数の条件が一致する SF をデータベースから検索し、目標言語の変数位置を得る。最後に変数を対訳単語辞書で翻訳し、並び替えることで、翻訳結果が得られる。

3 提案手法

本手法では、名詞を修飾している形容詞を元の名詞と合わせて一つの変数として SF データベースに登録する。本手法の流れを図 2 に示す。

まず日英対訳コーパス内の日文、英文をそれぞれ形態素解析し、名詞、形容詞とその他の部分に分ける。名詞と名詞を修飾している語を判定するために、篠山 [3] らの提案した分類規則を用いる。日英それぞれに名詞分類規則、日本語には名詞修飾語分類規則を用いて変数と定数を決定する。日英の名詞分類規則の一部、名詞修飾語分類規則の一部を表 2 - 4 に示す。

判定タグは名詞および形容詞の形態素解析結果と、完全一致する。次に一致するタグの判定位置から隣接する品詞を調べ、部分一致による判定を行う。一致した場合に判定タグに結果を付与する。変数は“EDGE”，定数は“NODE”，修飾語は“SYU”で表されている。

次に変数の中で、修飾語と判定された語を対訳単語辞書を用いて対応付けを行い、名詞と合わせて一つの単語とする。最後に名詞の対応付けを行い、SF データベースに登録する。

図 1 の文に提案手法を用いた場合の変数と定数の関係を図 3、表 5 に示す。“花”を修飾していた“かわいい”に対応関係が見つかり、“かわいい花”という一つの変数として認識されるようになる。

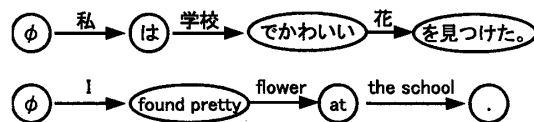


図 1: 例文の関係

表 1: 図 1 の変数と定数の対応関係

J	E	L_j	L_e	condition
ϕ	ϕ	1	1	1p
は	found pretty	2	3	the
でかわいい を見つけた。	at	3	2	Z
	.	-	-	-

表 2: 日本語名詞分類規則

判定タグ	隣接する品詞	判定位置	結果
名詞-サ変接続	動詞	NEXT	NODE
名詞-サ変接続	数	NEXT	CUT
名詞-サ変接続	他全	NEXT	EDGE

表 3: 英語名詞分類規則

判定タグ	隣接する品詞	判定位置	結果
JJ	DT	PREV	SYU
JJ	NN	NEXT	SYU
JJ	ALL	PREV	NODE

表 4: 名詞修飾語分類規則

判定タグ	隣接する品詞	判定位置	結果
形容詞-自立	名詞-一般	NEXT	SYU
形容詞-自立	固有名詞	NEXT	SYU
形容詞-自立	形容動詞	NEXT	SYU

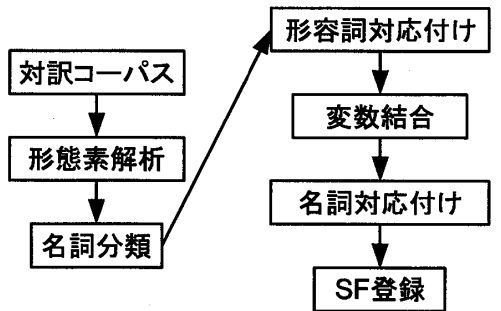


図 2: 提案手法の流れ

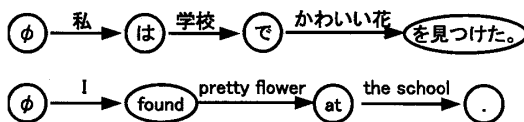


図 3: 図 1 に提案手法を用いた場合の関係

表 5: 図 3 の変数と定数の関係

J	E	L_j	L_e	condition
ϕ	ϕ	1	1	1p
は	found	2	3	the
で	at	3	2	Z
を見つけた。	.	-	-	-

4 評価実験

手法の有効性を示すために従来手法による SF 作成数と提案手法による SF 作成数の比較を行う。日本文の形態素解析には茶釜 [4], 英文の形態素解析には

Brill's Tagger[5] を用いた。日本文は“名詞-形容動詞語幹”, “形容詞-自立”を形容詞として扱う。英文は“JJ(形容詞)”, “JJR(形容詞比較級)”, “JJS(形容詞最上級)”を形容詞として扱う。

実験には日英対訳文 10,250 文, 形容詞を含む文はそれぞれ 3,111 文, 4,036 文のコーパスを用いて行う。

また従来手法との翻訳精度の比較を行うため, 日英対訳コーパス 213,626 文を SF データベースに登録し, 翻訳実験を行う。

5 おわりに

本稿では名詞を修飾する形容詞を対象として Super-Function を拡張する手法を提案した。この手法により, SF データベースの肥大化を防ぎ, 多様な文に対応する SF データベースが構築できるようになる。

今後の課題としては, 名詞から形容詞付き名詞に変換した際の形容詞位置の判別手法について検討する予定である。

参考文献

- [1] <http://www.excite.co.jp/world/>.
- [2] F. Ren, “Super-function based machine translation,” Language Engineering, Proceedings of JSCL and TsingHua University Press, pp.305–312, 1997.
- [3] M. Sasayama, F. Ren, S. Kuroiwa, “Automatic extraction of super-function from corpus and experimental evaluation,” Information-MFCSIT'06, pp.395–398, 2006.
- [4] 松本裕治, 北内啓, 山下達雄, 平野善隆, 浅原 正幸 松田寛, “日本語形態素解析システム『茶釜』 version 2.2.1 使用説明書,” Nara Institute of Science and Technology, 12 2000.
- [5] E. Brill, “Some advances in transformation-based part of speech tagging,” In Proceedings of the 12th National Conference on Artificial Intelligence (AAAI-94), vol.1, pp.722–727, 1994.