自 然 言 語 処 理 103－18
（1994．9．16）

埋め込み法を用いた訳文生成

畠中 伸敏
キヤノン株式会社 事務管理部

【概要】 原文の構文構造は、連用修飾語あるいは連体修飾語などの互いに繋がりのある語群から構成されている。この語群に対して下位範疇化フレームという概念を導入し、語群を一つの単位として扱うことができる様にすることにより、埋め込み法の適用を容易にした。更に、FRL の知識情報を群の形で構成すると言う性格を用いて、埋め込み法で必要となる情報を一括して整理できる様にし、繋がりのある語群の情報を訳文生成時まで保持できる様にした。以上の方法の導入により、構文構造の簡素化や訳者の意図した訳の感性を訳文中に実現する一方法を提供した。

Embending Method for Generating Phrase Structure and Translation

NOBUTOSI HATANAKA

Business Information Div.
CANON Inc.

Abstruct

The Case Structure of Japanese language is contained with the modifier which are in lexical cohesion relation with each other a lexical chain. The concept of Subcategorizaton Frame is introduced in order to deal easily with Embending Method appling for these modifier. Accordingly, the lexical cohesion such as the modifier is enable as one coordination. Furthermore, since it is found out that a FRL knowledge base is composed in a good formed group, FRL is introduced too so that the required knowledge by Embending Method is able to make into a bundle. The related information and inheritance between each other lexical cohesion is kept until the result of translation will be genarated. The evaluation show that the two methods increase the abstruction level of Phrase Structure and improve the intuition level of the generated translation result by my proposed methodology.

１．はじめに

ここでの結論は、複写機NP6030ｼﾘｰｽﾞで用いられているｻｰﾋﾞｽ･ﾏﾝ 向けﾏﾆｱﾙを分析して得られたものである。従って、結論の成り立つﾄﾞﾒｲﾝを複写機ｻｰﾋﾞｽ･ﾏﾆｱﾙ の文書空間と言うことで限定するが、適用においては、同様のｱﾌﾟﾛｰﾁ で取り組むことができ、それぞれの事例を重ね合わせて行くと、いずれは全空間を覆い隠すことになると考える。まず、最初に現状の機械翻訳処理部の持つ問題点について述べ、その解決策として下位範疇化ﾌﾚｰﾑ及びFRL の導入を試み、実験結果の例を示した。

２．埋め込み方式の解決しようとする問題点

従来の機械翻訳方式には、翻訳プロセスとして、形態素解析および形態素生成の共通した各プロセスを経て翻訳処理がなされる。また、『トランスファ方式』および『中間言語方式』には、加えるに構文構造解析、意味構造解析、語彙・構造変換、意味構造生成、構文構造生成の処理過程が存在する。（参考文献【３】）従って、従来の機械翻訳方式には、一旦、文章をバラバラに語の最小単位に分割し、目標言語の構文構造に従って訳語を並び換えると言う共通した処理過程が存在する。この方法であると、言い回し表現や慣用的表現、熟語表現などにおいて、解析則や生成則に巧く乗らなかったり、発火しても原文に従って忠実に正確に訳すため、逆に、実際の表現からかけ離れたぎこちない表現になる場合が多く、また、係り結びの先が複数存在した時は多義性の解消などにおいて精度よく結果が得られる保証は充分とは言えない。

これは、形態素解析を経て、一旦、繋がりのある語群を語の最小単位に分割するため、繋がりのある語であることの情報を訳文生成時まで、保持することができないことに起因している。例えば、日本語文から英語文の変換において、連用修飾語、連体修飾語は副詞句と形容句に変換され、互いに繋がりのある語群として翻訳される。しかしながら、機械翻訳処理において、語群からなる複数の修飾語が存在すると、語群としてまとまりのある情報を使用しないために、複数の語群の間で組み合わせるべき語が違うグループの語群に入れ子になるなど、意味不明の翻訳生成結果になることがしばしば出現する。この原因は、形態素解析の段階での品詞情報が精度よく確定しなかったり、語の係り結びの先が複数存在した時の多義性の解消が巧くいかないことに起因している。これは、また、目標文らしくない語順や修飾位置として翻訳文が生成されることもある。

次に、従来の機械翻訳の通常の処理では、語の単位で並び変えて訳文を生成するため機械翻訳システムがもつ通常の変換則と構文生成則に縛られ、逆に、実態に合った訳の導出に困難が生じる。例えば、原文での品詞は名詞であるが、訳文では節を生成するとか、動詞が副詞に、合成名詞の一方が形容詞に変換されるなどの原文と訳文とにおいて構文構造変換パターンが一般化できない場合は通常の機械翻訳システムがもつ一般的な変換則や構文生成則に縛られるため、変換則や生成則が発火しても、逆に、実態に合わない訳を生成することになる。『今日は良い天気です。』を機械翻訳により翻訳すると訳文結果は『Today is good weather 』となる。『スィッチON』は『switch ON』となる。通常使用される表現は『It's fine today 』であり、『switch is turned ON 』と訳させたい場合など通常の機械翻訳処理では難しい。つまり、特定の言い回し表現は、機械翻訳の通常の変換則や構文生成則に乗りにくい。構文構造の変換のパターンを対応づけにくい場合や、原文と訳文間での品詞の違いは、正しい訳の生成ができないばかりか、誤訳になる場合がある。

解決すべき課題

①繋がりのある語群の情報を訳文生成時まで保持すること。

②機械翻訳システムが持つ変換則や構文生成則などが持つ一般則としてのルールの束縛を解き、原文から目標文へ、適切な表現を用いた対応のある訳文生成を図る。

3．適用の範囲とモデル化

ここで埋め込み法の適用の対象としたのは、①連用修飾語、②連体修飾語、③全文について試みた。次の例は①連用修飾語の例であるが、対象とする部分には、変化する表現と固定して使用される表現からなり、固定して繰り返し使用される表現は固定表現部分として登録し、変化する表現は埋め込み語として処理を施す。

①連用修飾語の例
原文　：上下ローラが過熱した状態で測定する。
目標文：The measurements should be taken after the upper and lower rollers have heated.

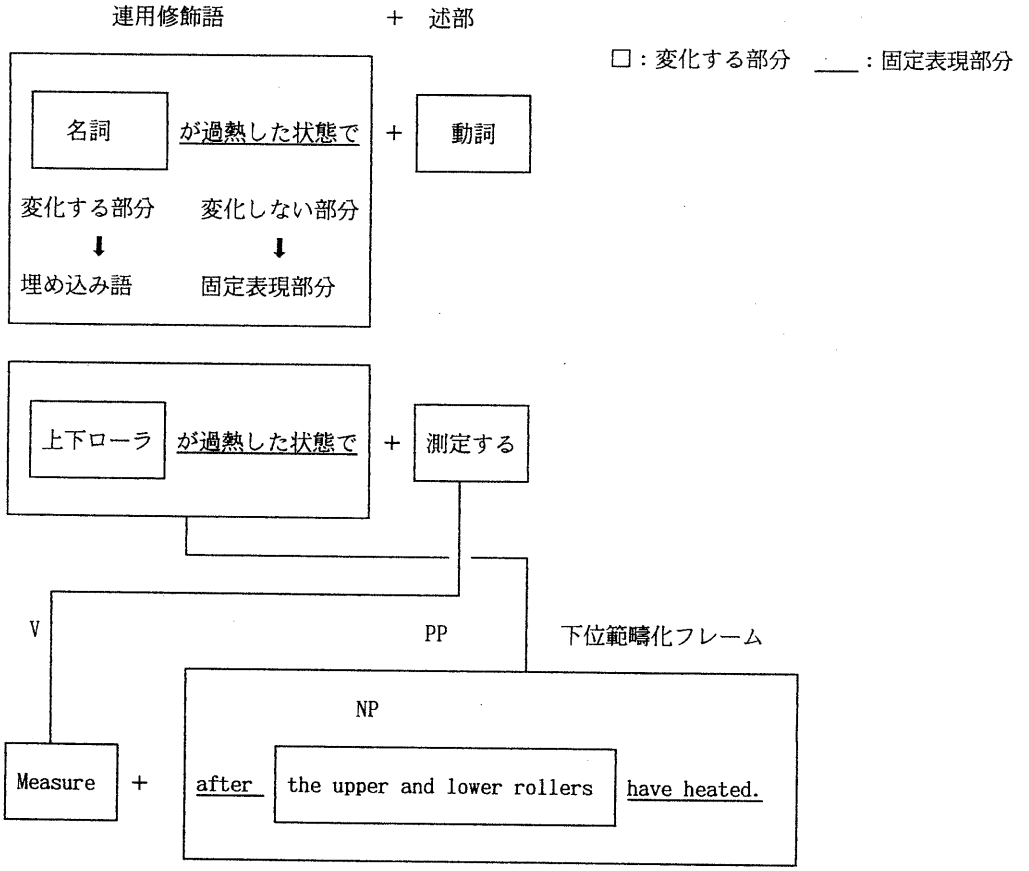

図１．連用修飾語の埋め込みパターンの例

この場合の語彙記述は、measure,V,__ PP となるが、勿論、動詞の部分も変化する可能性がある。copy,V,__ PP　test, V,__ PP　などである。ここで、語彙記述の最後の部分は生成された動詞句の下位範疇化フレームになり（参考文献【２】）、measure はPPを下位範疇化[PP,VP] する。ところで、after を主要部とする語彙記述は、after,PP, __S　となり、動詞が前置詞句の『外に』あることにな

る。この『外に』あるものを埋め込み語としない理由は固定表現部分と組み合わせて用いる語の振れの要素を取り除き、語群としての繋がりのある情報を保持するためである。更に、動詞を埋め込み語の対象とすると、様相と命題の扱いが必要となる。

after はthe upper and lower rollers を下位範疇化 [NP,PP]する。この関係は、主要部の前置詞after に直接支配された名詞句であることを示している。この直接支配された名詞句を埋め込み語の対象とする。この埋め込み語の『上下ローラ』は他の文章では、『ドラム表面』、『一次耐電体』などに変化する。

②連体修飾語での例
原文　：選択倍率に応じた位置を通過するスキャナー。
目標文：The scaner is moved past the position corresponding the ratio selected.

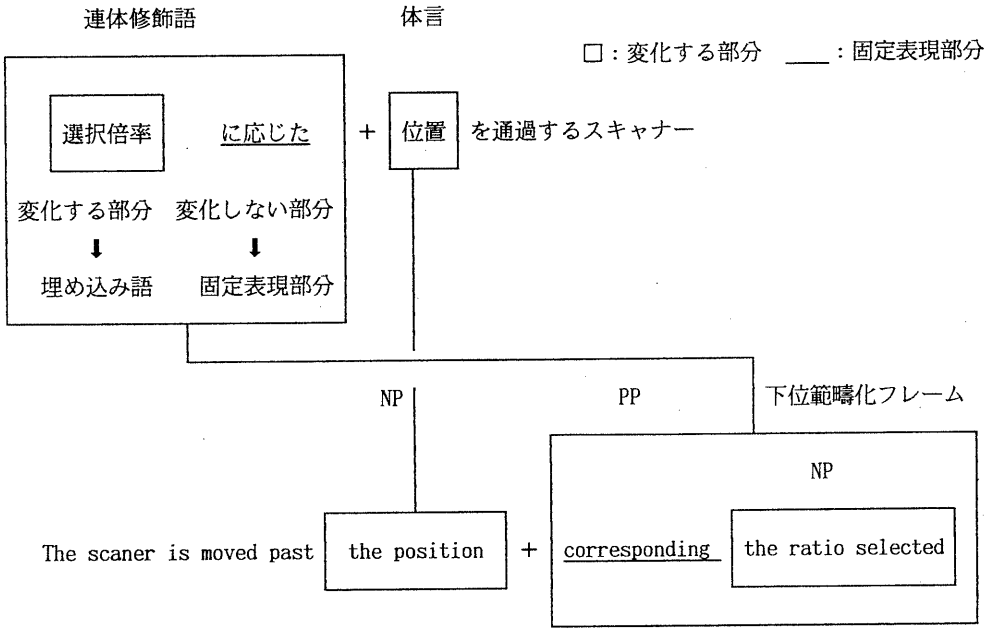

図２．連体修飾語の埋め込みパターンの例

同様に、語彙記述は、position,NP,__PPとなる。他の文章の使用において、『位置』と言う語は、『紙の大きさ』、『電圧制御』などに変化し、それぞれ、size of copy,NP,__PP　control of voltage,NP,__PPとなる。『選択倍率』は『紙の種類』、『濃度』、『紙の質』などに変化する。この例におけるcorresponding を主要部とする語彙記述は、corresponding,PP, __NPとなり、corresponding はratio selectedを直接支配することになり、この部分を埋め込み語としての対象とする。

図３の例において、下位範疇化フレームの概念は導入できないが、文章全体と言うことで『外に』と言う関係がなく、文章全体を独立した語群の集まりとして扱うことができる。文章全体を一つの纏まりのある単位と考え、変化する部分と固定表現部分に分割する。変化する部分の『清水』と言う人名は『田中』、『鈴木』などに置き換えられ、日付けの『1992年 5月 5日』は『1995年 1月 1日』などに違った場面で変更されて使用される。この変化する部分を埋め込み語としての対象とすることができる。

③全文の例
原文　：清水氏の神戸転勤に関する1992年 5月 5日付け貴書を頂いた。
目標文：
We recieved your letter of 5 May 1992 concerinig to Mr.Simizu's transfer to KOBE.

文　▒▒　□：変化する部分　＿＿：固定表現部分

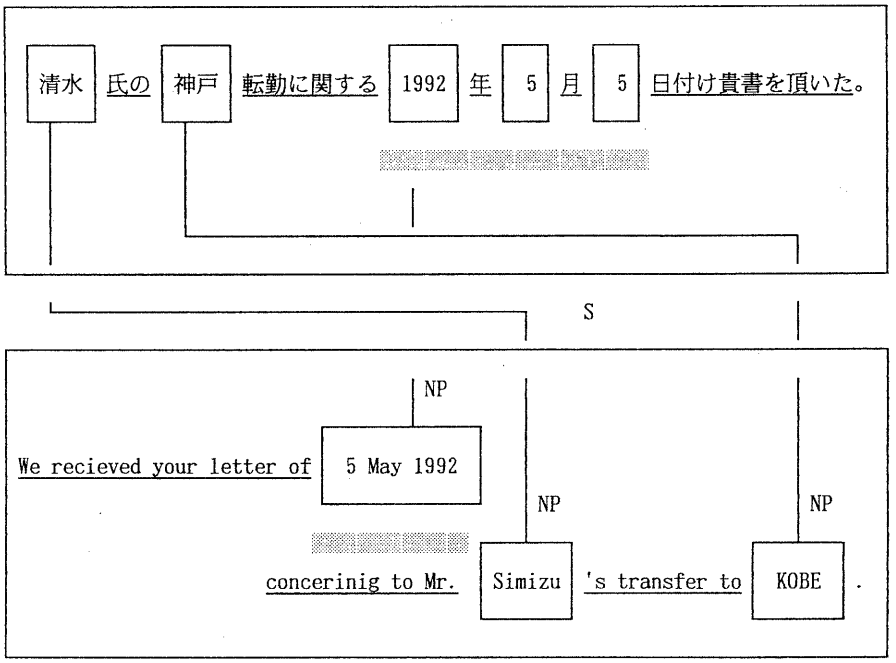

図３．全文での埋め込みパターンの例

ここで、下位範疇化フレームを導入したのは、繋がりのある語群の情報を訳文生成時まで、保持するためで、通常の機械翻訳システムに原文の引き渡しと同時に、これらのフレームの情報を同時に引き渡すことにより、全文もしくは下位範疇化フレームの訳文生成は、独立した別個の埋め込み法の処理プロセスを経て行われる。その分、原文の構文構造が単純になり、通常の機械翻訳処理において、下位範疇化フレーム単位に扱うことができ、修飾位置の決定や係り結びの処理は容易になる。逆に、『外に』あるものは、固定表現部分と組み合わせて用いる対象外とすることにより、語群としての繋がりを高め、構造の抽象化のレベルを高める。

次に、埋め込み語を名詞に（あるいはNP）に限定したのは、その他の品詞の場合には、文における様相と命題を解決する必要や、訳文生成時の修飾位置の決定が、原文と訳文間においてスムーズに対応が取れないためである。また、入力文書から、固定表現部分の抽出が容易にでき、組み合わされて使用される埋め込み語の辞書引きを容易にできる。下位範疇化フレームは固定した表現と埋め込み語の組み合わせを、一つの概念に結束させることになる。

４．下位範疇フレームの知識表現と訳文生成の手順

人工知能の研究過程で、Minsky(1975)が知識構造をフレームなどの名前で呼んだ。このフレームを使うと知識を情報の便利な群に構成することができ、フレームは概念記述をしたスロットの集まりから構成される。ここでは、FRL(Frame Representation Language)を用いて、下位範疇フレームの知識構造の整理を行う。（参考文献【１】）

表１．FRL を用いた下位範疇フレームの事例

フレーム	スロット	ファセット	データ	コメント
KANETU	フレームの見出し	$VALUE	//が過熱した状態で	
	フレームの訳語	$VALUE	after//have heated	
	埋め込み語品詞条件	$VALUE	名詞	
	埋め込み語	$VALUE		
	意味属性の条件	$DEFAULT	状態	
		$REQUIRED	[MEMBER:	
			VALUE 状態	
			無指定]	
	埋め込みの語順	$VALUE	逆順	
	原文埋め込みコード	$VALUE		
		$DEFAULT	用五末	
		$REQUIRED	[MEMBER:	
			VALUE 用五末	
			用五後	
			用五頭]	
	訳文埋め込みコード	$VALUE	#500&&	
	埋め込み法対象部分の原文での品詞	$VALUE	連用修飾語	
	下位範疇フレームの訳文での品詞属性	$VALUE	副詞句	
	生成時の修飾位置	$VALUE		
		$DEFAULT	文末	(TYPE:
		$REQUIRED	[MEMBER:	typical)
			VALUE 文末	
			主語後	
			文頭]	
	埋め込みパターン	$VALUE	3	
	A-KIND-OF	$VALUE	EMBENDING	

表１の//は埋め込み語の入る位置を示す。下位範疇フレームおよび全文に渡って同様の知識構造を総称してフレームと呼ぶことにする。このフレームを各言い回し表現ごとに登録し、埋め込み語は意味属

性、品詞などの情報を加え単語辞書として構築する。原文の中に一致するフレームの見出しが存在すると入力文章中の該当する埋め込み語位置の単語の辞書引きを行い、埋め込み語の品詞条件、埋め込み語の意味属性の条件が合致するかどうかフレームの照合を行う。一致するとフレームが起動し、埋め込みパターンに従って、訳文生成を行う。表１の生成時の修飾位置が文末であることから、原文での連用修飾語は訳文では副詞句となり、文末より動詞を修飾することとなる。埋め込みパターンの３は、表２から埋め込み語は一つで、訳文では固定表現部分に挟まれて、埋め込み語が訳されることを示す。

表２．埋め込みパターン

ﾊﾟﾀｰﾝNo n(i)	原文パターン	訳文パターン
1	埋め込み語＋固定表現部分	固定表現部分＋埋め込み語
2	第１埋め込み語＋固定表現部分 （＋第２埋め込み語）	固定表現部分＋埋め込み語
3	埋め込み語＋固定表現部分	第１固定表現部分＋埋め込み語 ＋第２固定表現部分
4	第１埋め込み語＋第１固定表現部分 第２埋め込み語＋第２固定表現部分	第１固定表現部分＋第１埋め込み語 ＋第２固定表現部分＋第２埋め込み語

手順を以下の通りに示す。

Ⅰ）原文入力　：上下ローラが過熱した状態で測定する。

Ⅱ）該当するフレーム見出しの検索：//が過熱した状態で

Ⅲ）埋め込み位置の単語の辞書引き：『上下ローラ』の品詞は『名詞』で、意味属性は『状態』

Ⅳ）フレームの照合　：正確に一致

Ⅴ）フレームの起動

a)テーブルの作成　：埋め込み位置情報、フレームの訳の生成結果

訳語にそれぞれ変換する

上下ローラ＋//が過熱した状態で　➡ upper and lower rollers + after//have heated

埋め込みﾊﾟﾀｰﾝ3 を用いてフレーム単位の訳の生成を行う

必要な形態素生成を加える

➡after the upper and lower rollers have heated

b)原文での埋め込みコードの埋め込み：用五末測定する。

Ⅵ）翻訳処理部で翻訳　：Measure #500&&.

Ⅵ）訳文埋め込みコードを

フレームの訳の生成結果に変化：

Measure after the upper and lower rollers have heated.

５．埋め込み法の適用による効果

埋め込み法を下位範疇化フレームに適用することにより、繋がりのある語群を一つの単位で扱うことができ、FRL を用いて埋め込み法で必要となる情報を一括して整理できる。このことにより、繋がりのある語群の情報を訳文生成時まで保持することができた。更に、原文中を通常の翻訳処理部と埋め込み処理部に分割して翻訳生成することにより、通常の翻訳処理部へ引き渡す構文構造が、抽象化された単純な構造に変換される。係り結びの決定などにおいて、候補とすべき語の組み合わせを極端に減少させることができる。

効果

①構文構造の簡素化により、係り結びの決定などにおいて、候補とすべき語の組み合わせを極端に減少させることができる。

②繋がりのある語群を一つの単位で扱うことができ、一括して、情報を訳文生成時まで保持できる。

③登録された適切な表現を、訳文生成時に、通常の機械翻訳処理部と独立に合成でき、訳者の感性を実現できる一方法を提供する。

表３．実験結果

原文	通常の機械翻訳処理	埋め込み法
上下ローラが過熱した状態で測定する。	Measure in the condition which upper and lower rollers overheated.	Measure after the upper and lower rollers have heated.
選択倍率に応じた位置を通過するスキャナー。	The scaner which leaves the position which complied with the ratio selected.	The scaner which leaves the position corresponding the ratio selected.
清水氏の神戸転勤に関する1992年 5月 5日付け貴書を頂いた。	I got May,1992 date -- 5 -- your writing about Koube transfer of spring water.	We recieved your letter of 5 May 1992 concerning to Mr. Simizu's transfer to KOBE.

-- 5 -- ：翻訳処理部が部分的に失敗したことを示す。

実験結果から分かるごとく、訳者の意図した訳の感性は、埋め込み法が数段、優れている。例③は通常の翻訳処理部では構文解析に失敗し、部分的未翻訳となっていることを示している。このことからも、機械翻訳処理部に与える負荷も軽減し、構文構造を簡素化することに大いに役立てることができる。

６．参考文献

【１】Harry T."Natural Language Processing.":Petrocelli Books Inc,1981 森 健一他訳：自然言語処理入門、産業図書株式会社、1986

【２】Peter S."Lectures on Contemporary Syn-tactic Theories.":Leland Stanford Junior University,1985 郡司隆男他訳：現代の文法理論、産業図書株式会社、1989

【３】野村浩郷編：言語処理と機械翻訳、講談社ｻｲｴﾝﾃｨﾌｨｸ 、1991

【４】仁井正治：TOPTRAN における事例学習、情報処理学会自然言語処理研究報告 102-8 (1994)