

2E-3

基本動詞の格記述継承による
大規模格辞書の構築

稲垣博人, 斎藤珠喜, 小橋史彦
NTTヒューマンインタフェース研究所

1. はじめに

自然言語の解析、理解、生成などにおいて、格辞書は必須辞書の1つである。特に、格辞書を用いた格解析は、格間の共起関係や格の深層の意味を明確にするため、種々の自然言語アプリケーションへの適用が期待される。我々も、係り受け解析技術[1]の精度向上及び、自動索引抽出、抄録作成などへの格解析の適用を考えている。

しかし、従来の格辞書[2]は、個々の動詞について詳細に記述しようとしているため、実用レベルの格辞書を構築するには至っていない。

そこで、我々は実用レベルの語彙数を持つ大規模な格辞書の構築法として、基本動詞の格記述継承を有効利用した格辞書構築法を提案する。

2. 格辞書の基本的枠組み

一般に動詞が持つ意味は、動詞とその動詞に共起する格と格要素(格を形成する名詞句)から形成される。そのため、同じ意味を持つ動詞は、類似の格記述(動詞が持つ必須格、任意格、表層格、深層格等の動詞の格に関する全ての情報)を持つことが多い[3]。このような動詞の意味と格記述の類似性に着目し、基本的な動詞の格記述を同(類)義の意味を持つ動詞に継承させる継承型格辞書を基本的枠組みとして捉えた。つまり、動詞を基本動詞の同義語または下位の類義語に分類し、上位の基本動詞の格記述を継承させることを特徴とする格辞書である。そのため、下位に分類された動詞については、独自に格記述する必要性がない。

基本動詞として、『学ぶ』を例にとると、その格記述は表1に示すようになる。

表1 『学ぶ』の格記述

【学ぶ】(基本動詞)		
① N1が	(N2に/から)	N3を
動作主	相手	対象
N1, N2 = 人間、機関, N3 = 抽象物		
② N1が	(N2で)	N3を
動作主	手段	対象
N1 = 人間, N2 = 生産物、機関, N3 = 抽象物		
(但し、Nは名詞及び名詞句)		

表2 『学ぶ』の同(類)義語の格記述

【学習する】
《格記述》 格(【学ぶ】①②)
《同義語》 【学ぶ】

【独習する】
《格記述》 格(【学ぶ】②)+数量("一人")
《類義語》 【学ぶ】

【勉強する】
《格記述》 格(【学ぶ】②+N1 [+機関])
《類義語》 【学ぶ】

文中での言い替えが可能な同義語の場合、基本動詞と同じ格記述を持たせる。例えば、同義語『学習する』の辞書項目は、表2に示すように『学ぶ』へのポインターのみとなる。

類義語の場合、格記述の変化は表3に示した2つに分類されるが、基本動詞の格記述との差分のみ新たに記述するだけでよい。格記述の変化例を表2に示す。『独習する』(数量格 "一人で"の追加)は任意格相当部が変化した場合の例、『勉強する』(動作主N1の意味素性追加 [+機関])は必須格が変化した場合の例。

表3 格記述の変化

①任意格相当部の変化
• 任意格の追加
②必須格の変化
• 格の追加・削除
• 格要素の意味素性の追加・削除

また、同義語、類義語以外の意味類似性の低い関連語(例えば、『感激する』と『(心を)動かす』)も類義語と同様に扱う。

3. 格辞書構築法

格関係を付与する動詞は、現在我々のマスター辞書(語数約10万語)の中の全動詞(約2万語)とし、基本動詞としては、IPAL[2](861語)を使用した。

格辞書構築のための作業としては、まず対象とする動詞の上位語を基本動詞から選択する。さらに、最適な基本動詞の語義（格記述）を決定する。

最後に、付与した基本動詞に対して同義語、類義語、関連語の判別を行い、類義語や関連語については、基本動詞と異なる格記述のみ記入する。

図1に格辞書構築作業の概略図を示す。作業時間の短縮を計るために、機械処理で人手作業の支援を行った。

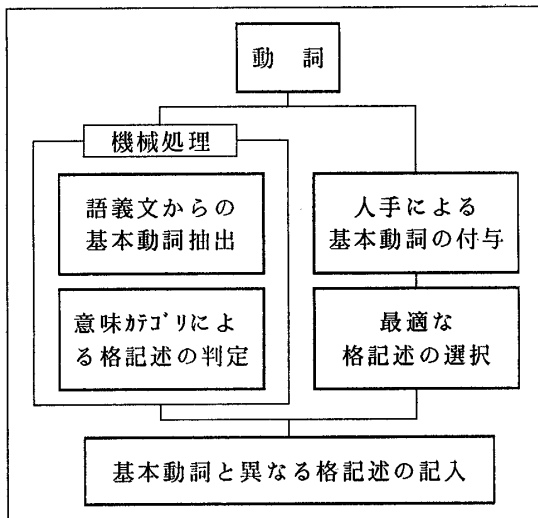


図1 格辞書構築作業の概略

4. 機械処理による格辞書構築支援

機械処理では、国語辞書の語義文に着目した処理を行った。語義文は、人間が文章を書く際役立つように、見出し語の意味を平易な動詞を用いて書き下している。そのため、語義文中の動詞は、見出し語の上位語にあたるものが多く[4]、かつその動詞は同義または、類義の関係を持つことが多い。そこで、動詞の語義文から上位語を自動抽出し、その上位語が基本動詞と一致する場合、その基本動詞を採用した。

今回、処理を単純化するため、

”（和語動詞）+（こと）。”

例 【学習する】習い学ぶこと。

という形式を持つ語義文から上位語を自動的に抽出した。基本動詞が複数の格記述（語義）を持つ場合、意味カテゴリ番号[5]から最適な格記述を決定した。つまり、対象とする動詞に最も近い意味カテゴリ番号を持つ基本動詞の格記述を選択した。意味カテゴリ番号が一致しない場合は、抽出した基本動詞は不適切と判断し、破棄した。

5. 格辞書構築結果と上位語付与結果の検討

3, 4章で述べた格辞書構築法に基づき、格辞書の構築を進め、現在、動詞の上位語付与がほぼ完成している。その結果を表4に示す。

I PALの基本動詞および表記のゆれは5820語あり、動詞全体の約30%を占めている。表記のゆれ

については、機械または人手で付与された格記述を適用することができる。

表4 上位語付与結果（単位 語）

		和語動詞	サ変動詞	合計
I P A L	見出し語	861	0	1885
	使役可能	1024	0	
表記のゆれ		2805	1130	3935
計算機による上位語付与		658	1988	2646
		正解 350	正解 1403	正解 1753
		修正 308	修正 585	修正 893
人手付与		4987	6903	11890
合計		10335	10021	20356

機械処理では、全体の約15%の動詞について、上位語を付与することができた。そのうち、上位語抽出の正解率は66%であった。計算機による主な抽出失敗原因としては、①語義文が不適當（名詞、サ変動詞、形容詞等で表現されており、基本動詞で表現されていない）②語義文、意味カテゴリ番号が付与されていないなどがある。

残りの約50%の動詞については、人手で上位語付与作業を行った。

6. おわりに

本稿では、格解析に必要な大規模な格辞書の構築法について述べた。特に、動詞の意味と格記述の類似性に着目し、基本動詞の格記述を継承させる継承型の格辞書を提案した。現在I PAL辞書を基本動詞として格辞書の構築を進めているが、本構築法は、辞書規模の縮小、構築時間の短縮に有効であると考えられる。

今後は、構築した格辞書の評価を行う予定である。

7. 謝辞

本研究を進めるにあたり、熱心な討論をしていただき、データ収集に多大な協力をして頂いたNTT技術移転(株)の清末三恵子嬢に感謝します。

8. 参考文献

- [1] 稲垣ほか：意味連結パターンを用いた係り受け解析，NL67-5。
- [2] 計算機用日本語基本動詞辞書I PAL (Basic Verb) -辞書編-，情報処理振興事業協会技術センター。
- [3] 仁田ほか：ソフトウェア文書のための日本語処理の研究-7，情報処理振興事業協会。
- [4] 富浦ほか：語義文からの動詞の階層関係の抽出，NL64-3。
- [5] 大野ほか：角川類語新辞典，角川書店，1981。