

5G-3

概念体系を用いた
格関係子の機械的変換について三浦靖子 玉越靖司 重永信一
松下電器産業(株)東京研究所

1. はじめに

自然言語処理技術の実用化と共に、電子化辞書の必要性が高まっている。この要求に対して、汎用・大規模の電子化辞書の開発が進められている[1]。この辞書は汎用であるため、特定の用途の辞書には変換が必要になる。更に、辞書の大規模性を考慮すると、自動的に変換することが望まれる。

本稿では、その変換の中で最も困難である「格関係子の変換」に関して述べる。

2. 格関係子の対応づけ

既に仕様が与えられている汎用辞書[1]から、我々が仕様を想定した日本語生成用辞書への格関係子の変換を扱う。

汎用辞書においても、日本語生成用辞書においても、格関係子是用言的概念と体言的概念の2概念の関係を表すものである。両辞書での違いはその格関係子の種類の数にある。汎用辞書の格関係子の数は13であるのに対し、日本語生成用辞書のそれは34である。(図1)

汎用辞書(13)		日本語生成用辞書(34)
agent	⇒	agent implement possessor :
object	⇒	object product experiencer :
:		:

図1 二つの辞書の関係子の種類

従って、汎用辞書においてある格関係子で結ばれている概念の組が、日本語生成用辞書では複数の格関係子の候補のうちどれで結ばれているかを特定する作業が必要になる。例1は、汎用辞書では同一の格関係子「object」によって結ばれている概念の組でも、日本語生成用辞書では「object」で結ばれるものと「product」で結ばれるものがあることを示している。

例1 格関係子の変換

例文: 1. (太郎が) 林檎を食べる。
2. (太郎が) ケーキを作る。

汎用辞書		日本語生成用辞書
(object, 食べる, 林檎)	⇒	(object, 食べる, 林檎)
(object, 作る, ケーキ)	⇒	(product, 作る, ケーキ)

両辞書間の格関係子の対応づけを分析した結果、表1のようになった。

表1 格関係子の対応づけ

汎用辞書と日本語生成用辞書の格関係子の対応関係	日本語生成用辞書の格関係子の種類の数
(1) 1対1に対応する	14
(2) 1対多に対応する	16
(3) 汎用辞書に該当する関係子なし	4

(1)の場合、汎用辞書の格関係子をそのまま対応する日本語生成用辞書の格関係子に置き換えるだけでよい。また、(3)の場合、日本語生成用辞書の格関係子が表す関係を持つ概念の組が汎用辞書に存在しない。従って、本稿では、(2)の場合の変換について述べる。

3. 変換について

この変換に必要なサンプルを我々の手で作成し、次に示す変換規則と概念体系を用いて、その格関係子を日本語生成用辞書の格関係子に付け替える実験を行った。

汎用辞書には、式(1)のように、用言的概念と体言的概念の格関係を表す格関係子が記述されている。

(汎用辞書の格関係子A,
用言的概念,
体言的概念) . . . (1)

我々は、ここで記述されている2つの概念の組み合わせを分析し、分類することによって汎用辞書での格関係を日本語生成用辞書での格関係に対応させることが可能であることに着目した。すなわち、(2)のような規則によって変換が可能であると考えた。

(汎用辞書の格関係子 A,
用言的概念,
体言的概念,
日本語生成用辞書の格関係子 B) . . . (2)

しかし、全ての概念の組み合わせについて変換規則を記述すると、規則の作成のために非常に大きなコストがかかり現実的でない。

概念体系は、概念間の関係を記述しやすくすることを目的として分類されている[2]。そこで、概念体系の用言的概念のノードと体言的概念のノードの組み合わせを用いて(3)のような規則を作成した。

(汎用辞書の格関係子 A,
用言的概念のノード,
体言的概念のノード,
日本語生成用辞書の格関係子 B) . . . (3)

この規則により、それぞれのノードの下位概念となるものが汎用辞書の格関係子 A で結ばれていたら、それを日本語生成用辞書の格関係子 B に書き換えることが可能となる。例2がこの変換の実例である。

例2 変換規則を用いた格関係子の変換

変換規則: (object, 創造する, 具体物, product)	
汎用辞書の概念記述 (object, 作る, ケーキ)	⇒ 日本語生成用辞書 (product, 作る, ケーキ)

しかし、手作業で規則を作成するには、まだその負担が大きい。そこで、実際にはまず(4)のような規則を16個作った。

(汎用辞書の格関係子 A,
用言的概念のノードの集合,
体言的概念のノードの集合,
日本語生成用辞書の格関係子 B) . . . (4)

そして、(4)の各々の規則における用言的概念と体言的概念のノードを組み合わせで展開し、最終的に(3)の規則を1252個を作成し、変換を実行した。

4. 結果

概念体系のサンプルと、汎用辞書のサンプル437組

を用意した。概念体系のサンプルは、分類項目[1]を参考にして作成し、汎用辞書のサンプルは、例文から抽出して作成した。変換規則を用いて機械的な変換の実験を行った結果は表2の通りである。

表2 機械的変換の結果

2辞書の格関係子の対応	正しく変換された	正しく変換されない	計
(1) 1 対 1	101	---	101 (23%)
(2) 1 対 多	208	128	336 (77%)
計	309 (71%)	128 (29%)	437

用意した概念の組437のうち、1対1に対応するものも含めて71%が正しく変換された。正しく変換されなかったものを分析すると大きく分けて2つの問題点が明らかになる。

一つは、我々が作成した概念体系や汎用辞書のサンプルといったデータの問題である。正しく変換されなかった128組のうち、概念体系に起因するものが102組(80%)あった。また、汎用辞書のサンプルデータに起因するものが62組(48%)あった。こうしたことから、高品質な辞書データの作成に関する検討[2, 3]が重要であるといえる。

もう一つは、変換規則の問題である。正しく変換されなかった128組のうち、変換規則に起因するものが3組(2%)あった。これらは全て比喩的な表現の例文から抽出したサンプルであった。今後の検討を要する。

また、表1の(3)に当たる日本語生成用辞書の格関係子(「direction」「point of change」など)については、比較的少数だと予測されるので、手作業による記述が可能である。

以上の結果から、概念体系を用いた格関係子の機械的変換は、辞書の変換において有効な手段であるといえる。

5. おわりに

辞書の変換で最も困難である格関係子の機械的変換を試みた。その結果、辞書開発の有効な手段として、辞書の機械的変換が可能であるとの見通しを得た。更に、これにより、汎用辞書の有効性も明らかになった。

6. 謝辞

本研究を進めるに際してご討論頂いた、(株)日本電子化辞書研究所の清野氏、縄田氏に感謝致します。

参考文献

- [1] 概念辞書(第1版), EDR TR-007, 1988.
- [2] 井原他: 格関係に基づく概念体系, 情報処理学会第38回全国大会, 1989.
- [3] 岸本他: 概念モデルに基づく概念記述の一作成方法と評価, 情報処理学会第37回全国大会, 1988.