

3B-4

概念モデルに基づく

概念記述の一次作成方法と評価

岸本行生、三輪和弘、塚田康博、小寺次夫、大橋利香、照屋哲

(株)日本電子化辞書研究所

富士ソフトウェア(株)

まえがき

電子化辞書は言葉に依存した単語辞書とその言葉の間にある事実を知識ベースの形で表現した概念辞書から構成される。その概念辞書の中で、概念記述は各概念について常識として成り立つ事実を記述する。事実は概念見出し間に成り立つ関係を与えることで言葉間の概念として記述される。

我々は概念記述の部分実験として概念モデルに基づいて概念記述データの一次作成方法を検討した。概念モデルにおいて概念をBase ConceptとCase Conceptの2つに分けて考えている[1]。このCase Conceptを概念記述データとして記述する。以下に概念記述データの作成と評価について報告する。

1. 概念記述データの一次作成

我々はCase Conceptを記述することから文章中における言葉の共起関係を重視し、以下の手順で概念記述データの作成を行った。

1.1 作成方法

日常用いられる基本的な語を選び、その語の語義文に手を加えたものを記述対象の文章として用意した。その文章中において共起する言葉間の関係を、その言葉の意味を考え、関係子の定義にしたがって概念記述データとして記述した。関係子の定義の例を以下に示す。[2]

例

part of(部分) : あるものの部分
agent(動作主) : 有意志行為を引き起こす主体
sequence(連続) : ある出来事に引き続いて行われる出来事

また概念記述データは次の形式で表した。

[概念]:(概念1,関係子,概念2);

...

1.2 一次作成データの考察

一次作成した概念記述データの記述の揺れは以下の2点に大別できる。

- ・作業によって言葉の概念が異なり、共起関係を文章から抽出した時に個人間に揺れが生じた。

例

(a) [平らになる];

(平ら, goal, なる).

(b) [事物が相寄って一緒になる];

(事物, object, 一緒になる);

(相寄る, sequence, 一緒になる).

(a) を記述した作業者は「平らになる」が「平ら」と「なる」との2つの概念に分かれると考えたが、(b) を記述した作業者は「一緒になる」で1つの概念として考えている。

- ・一人の作業者でも視点の置き方で関係子に揺れが生じた。

例

(c) [身に痛みを感じる];

(c) において「身」と「感じる」との間の関係を「身」を感じる主体ととるか、感じた場所とするか、その両方ととるかによって次の3つの関係子の候補があがった。

(1) (身, object, 感じる);

(2) (身, location, 感じる);

(3) (身, place, 感じる);

揺れが生じたのは、関係子と助詞などの表層的情報の対応関係が不明確であったからである。言葉間の共起関係をとらえるための指針を作る必要があることが分かった。

2. 概念記述データの二次作成

2.1 揺れの減少方法

文章中の共起関係を表している助詞などの表層的情報に注目し、節1.1で作成した概念記述データ及び助詞の文法的役割などをもとに関係子とそれらの対応関係を検討した [3]。

A method and evaluation of conceptual description based on Conceptual model

Yukio KISHIMOTO, Kazuhiro MIWA, Yasuhiro TUKADA,
Tuguo KODERA, Rika OOHASHI, Satoshi TERUYA

Japan Electronic Dictionary Research Institute, Ltd. Fuji Software Inc.

検討した対応関係の例を図1に示す。その対応関係を用いて文章中の言葉間の関係を記述した。

関係子	助詞などの表層的情報	記述例
agent	が【する、なる】	私がボールを <u>投げる</u> 。 (私, agent, 投げる);
	(の)	人の <u>行く</u> 場所 (人, agent, 行く);
object	が	空が <u>青い</u> 。 (空, object, 青い);
	を	私が <u>ボール</u> を <u>投げる</u> 。 (ボール, object, 投げる);
	は	私は <u>眠い</u> 。 (私, object, 眠い);

図1 関係子と助詞などの表層的情報の対応関係例

2. 2 二次作成方法の考察

二次作成で関係子と助詞などの表層的情報の対応関係を用いて関係付けを行うことによって、一次作成と比較して次の点で効果があった。

- ・共起関係の抽出における揺れがなくなった。

例

上記の(a)(b)について言えば助詞などに注目して共起関係を抽出した結果(b)も(a)と同じように分けるように統一された。

- (b) (事物, object, なる);
(一緒, goal, なる);
(寄る, sequence, なる);
(事物, object, 寄る);
(相, manner, 寄る)。

- ・関係子の付与における揺れが減少した。

例

上記の(c)については「身」が助詞「に」によって関係付けられていることから感じる対象としての場所と考え、二次作成において次のように記述した。

- (c) [身に痛みを感じる]:
(痛み, object, 感じる);
(身, location, 感じる)。

以下の原因と考えられる揺れは個々に修正を加え、他のものに適合させた。

- ・関係子が助詞などの表層的情報と1対1対応になっていない。特に「に」と「の」についてはそれらの助詞と対応する関係子が多いため、どの関係子を付与したら良いか分かりにくい。

例

(d) [人が心に思うことを言葉に出す]:
「心に思う」という表現や「言葉に出す」等の表現の関係子が決め難い。

- 「心に思う」
(心, location, 思う);
(心, place, 思う);
「言葉に出す」
(言葉, goal, 出す);
(言葉, imprint, 出す);

- ・上記の対応関係において表層的情報は格助詞を中心にしており、複合助詞や慣用的な表現については関係付けを行っていない。その例としては以下のものがある。

例

- [物に力を加えても丈夫である]:
(物, object, 丈夫);
(加える, condition, 丈夫);
(力, object, 加える);
(物, goal, 加える)。

まとめ

概念モデルに基づき、関係子と助詞などの表層的情報の対応関係を用いた概念記述データの作成及びその評価について報告した。上記の評価を踏まえ今後の概念記述データ作成にあたって以下の点について順次行っていきたい。

- ・概念間の関係付けを行うのに必要な関係子を再検討する。
- ・関係子と表層的情報の対応関係を見直す。
- ・今回作業をほとんどを人間の手によってを行い、かなり時間を要したので、大量のデータ作成に向けてのより効果的な計算機による支援方法を検討する。

参考文献

- [1]三輪, 概念モデルに基づく概念記述の表現方法の一考察, 情報処理学会第37回全国大会, 1988
[2]Fillmore C., The Case for Case, Reinhart and Winston(田中春美 船越道雄訳『格文法の原理』, 三省堂, 1975)
[3]草薙裕, 南, et al., 文法形式が担う意味, 朝倉日本語新講座4『文法と意味II』p1~38, 1985