

概念パターンと概念関係子の検出

荻野孝野

日本電子化辞書研究所; 神戸大学

ogino@edr.co.jp

1. はじめに

EDR 辞書には、概念関係の利用を想定した概念体系辞書と概念記述辞書概念辞書がある。EDR 電子化辞書の開発当初(1986年度~)、EDR 概念辞書は機械翻訳などの利用を想定し、異なる言語でも中核となる概念部分は同じはずであるという方針に基づいて、構文解析の精度向上、意味選択などの利用を想定し、概念関係を記述し、大量のコーパスデータを作成した。その後、機械翻訳などの開発のための概念辞書という見方は薄らいできたが、概念関係を記述したコーパスは、現在、情報検索の分野などにおいて着目されつつあり、概念を導入した言語分析はまさにこれからのテーマでもある。

ここでは、表層格を手がかりに概念関係を導くという立場で、述語の概念パターンあるいは表層格から概念関係への変換を目的として検討したデータについて述べる。

2. 概念記述

ここで、概念記述とは、文の意味的表現として、構文的な係り受け関係にある文内構成要素間の関係に、関係子という概念関係ラベルを用いて表現するものである。別の言葉で、「概念関係パターンによる表現」「中間言語による表現」などとも表現されている。

例1 人が歩く⇒

人 ←agent ← 歩く

例1は、概念「人」が概念「歩く」に対し、agent という関係子でかかることを示す表現である。

例えば、「立つ」という言葉には、さまざまな意味がある。例2は、格助詞の組み合わせとしては全く同じ関係になりながら、関係子を導入して記述することによって、概念の違いが明確になる例である。

例2 子供[人間/agent]が地面[場所/place]に立つ

{物理的に立ち状態になること}

父[人間/agent]が選挙[選挙/goal]に立つ

{立候補する}

一方で、例3は異なる概念どうしであるが、概念関係で記述すると全く同じ形になる例である。

このレベルの記述では、概念関係表現は概念の違いを反映していないが、概念記述パターンとして同じものが概念的に近いグループとなることがわかる。

例3

幸子[人間/agent]が太郎[人間/agent]と会う。

幸子[人間/agent]が太郎[人間/agent]と結婚する。

3. 概念記述のパターン

「EDR 概念記述の抽象化による動詞パターンについて」(文献1)では、概念記述データに置き換えた動詞をパターン化して、おおまかな動詞の意味的分類を試みている。この段階の検討は、表層格と関係子の組み合わせにおけるパターンであって、体言の意味グループの情報は含んでいない。

ここで用いたデータは、EDR コーパスに出現する14000概念(異なり)のうち、それぞれに概念体系のパスをたどって、概念体系のパスが同じになるものから1概念ずつ選択した1400の動詞概念について作業を行なったものである。作業量の問題もあって、同じパスをたどるものは類似の概念記述になるだろうという想定で作業対象となる概念を絞り込んで行なったものである。

概念記述の核となる関係子として、41個の関係子を用いて記述した。以下にそれらの関係子の部分を示す。

agent; object; a-object; cause; implement;
material; source; goal; place; scene, basis,
manner, time, time-from, time-to, quantity,
purpose, content, role,....

これらの概念記述をパターンとして類似のものを調整し階層化すると、表1に示すように、概念記述パターンの階層に対応して、動詞の概念体系の階層に近いものが導かれることがわかる。

表1 概念記述パターンと概念分類

階層番号	関係子と格助詞の組み合わせによる概念関係パターン				格パターンに対する分類ラベル	EDR 3、4階層ラベル	格パターンに相当する単語
1	1	agent(が)			自身行為	身体的活動	運動、泳ぐ、乗り降り、かけっこ、動く、騒ぐ、突っ走る、出直す、
						移動	家出、再来、往復、泳ぐ、乗り降り、突っ走る
						自身行為	働く、休む、遊ぶ、入浴、休業、働き続ける、自習、
						感情活動	きょろきょろ、沈黙、騒ぐ
						自身行為	武装、蜂起、焼身自殺、家出
2	1	agent(が)	content(と)		「情報」にかかわる行為	ものを対象とする行為	計算、研究、判別
2	2	agent(が)	content(と)	range(について)		情報の移動	語る、発表、話す
2	3	agent(が)	content(と)	recipient(に)		対人行為/情報の移動	解説、偽る

4. 体言の意味と格表示から関係子へ

第3章は、関係子と格助詞の組み合わせから、動詞の分類を得ることができるかどうかの検討であった。ここでは、それらを一つの格項目に分解して、{体言の意味、格表示(表層格)、動詞}のセットから関係子を自動的に付与できるような変換表が求められないかを検討している。表1でわかるように、下位に行くに従って、格関係が付加され限定されていくが、有意志行為の範囲では「{人間}(が) ⇒ agent」になる部分は共通である。ということで関係子の自動付与を2項関係に着目して整理していくものである。

同種の先行研究に大石・松本氏らの「格パターン分析を利用した概念関係パターン獲得」(文献

2)があるが、これは、格パターンの組み合わせを用いて「関係子の自動獲得」を提案したものであるが、本稿の提案は2項関係に分解して検討するものである。

本データは以下のようにして作成した。

- ① EDRコーパスから用言を中心にして、そこにかかる体言、格関係(関係子付与済みのデータ)の共起関係を取りだす。
- ② 体言についてEDR概念体系のパスを用い、体言部分の意味的な抽象化を行う。
- ③ 上記レコードから{体言(表記)、体言意味グループ名、表層格、関係子}のセットを抽出し、用言ごとに「体言+表層格」から関係子を得るためのデータとして整理する。

(1) 元データとなる共起関係データの作成

表2 共起関係データ

STP: 概念体系パスノードで体言の意味特徴のために必要と思われる最下位ノード

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
概念体系全パス(概念ID表示)	概念体系STPまでのパス(概念ID表示)	概念体系全パス(概念説明で表示)	概念体系STPまでのパス	STPの表示	STP末尾語	表層格(実質格)	関係子	体言部分の単語	表層格(表現格)	体言の品詞	体言の概念ID	コーパスID	中核文	原文

(2) 体言部分の意味グループ作成

体言部分の概念にEDR概念IDが付与されているのでこれを用い、概念パスの情報を付与する。

表3 {概念パスつき体言、表層格、関係子} セット

STP 頻度: {STOPNODEに含まれる末尾単語、表層格、関係子} が一致した数

上位5段階までの体系パス	上位5段階までの体系パス	STP頻度	STOPNODE	STPの末尾単語	表層格	関係子	文中の体言部分	中核文	動詞
概念/ものごと/もの/具体的あるいは抽象的存在物	3aa966/3d017c/444d86/444e3f	1	作品	作品	が	'a-object	作品	作品が耳目を集める	集める
概念/ものごと/もの/具体物/静物	3aa966/3d017c/444d86/30f6ae/4444c4	1	情報媒体	情報媒体	が	'a-object	光ディスク	光ディスクが注目を集める	集める
概念/人間または人間と似た振る舞いをする主体/人間/行う行為の内容で捉えた人間	3aa966/3aa911/30f6b0/44474a	1	~する人	人	が	'a-object	科学者	科学者が尊敬を集める	集める
概念/ものごと/もの/具体物/静物	3aa966/3d017c/444d86/30f6ae/4444c4	1	機械	機械	が	'agent	テレビ	テレビが声を集める	集める
概念/人間または人間と似た振る舞いをする主体/自立活動体/組織	3aa966/3aa911/3aa912/30f746	2	国家の政治を司る機関	機関	が	'agent	省,通産省	通産省が24社を省に集める	集める

(3) {体言(表記)、体言意味グループ名、表層格、関係子} のセットのソート

元データは表3に示す通りだが、表層格から関係子への変換をより単純化するために作成したものが表4である。

表4 {STOPNODE末尾単語、体言、表層格、関係子} セット

STP 頻度: {STOPNODEの末尾単語、表層格、関係子} が一致した数

NO	STP頻度	STPの末尾単語	表層格	関係子	文中の体言部分	中核文	動詞
1	1	作品	が	'a-object	作品	作品が耳目を集める	集める
2	1	情報媒体	が	'a-object	光ディスク	光ディスクが注目を集める	集める
3	1	人	が	'a-object	科学者	科学者が尊敬を集める	集める
4	1	機械	が	'agent	テレビ	テレビが声を集める	集める
5	2	機関	が	'agent	省,通産省	通産省が24社を省に集める	集める
6	3	呼称	が	'agent	私たち,彼女,われわれ	われわれがカラを集める	集める
7	2	国家	が	'agent	チリ,米国	米国が情報を集める	集める
8	1	集団	が	'agent	団体	団体がコート、衣服、靴、おもちゃ、本を集める	集める
9	17	人間	が	'agent	ちびっ子あんこたち,ちびっ子あんこたち	ちびっ子あんこたちが人気を集める	集める
10	17	人間	が	'agent	委員,候補,美女,ちびっ子あんこたち	ちびっ子あんこたちが人気を集める	集める

(4) {体言意味グループ名、表層格、関係子、動詞} のセット

表3, 4のデータを元にして、体言部分を手作業調整して作成したものが表5である。

これによって、動詞側からみて、どのような体言の意味グループと格助詞の組み合わせでどういう関係子になるかの傾向がデータから抽出されてくる。

表5. {意味グループ(手作業記入)、表層格、関係子} の関係子順データ

意味グループ	表層格	関係子	動詞
行為	で	#/implement	売る
指示範囲	が	'agent	売る
組織	が	'agent	売る
人間	が	'agent	売る
システム	は	'agent	売る
指示範囲	に	agent/goal	売る
場所	で	a-object/place	売る
行為	から	'cause	売る
様態	で	'cause	売る
評価	で	'condition	売る
様態	他	'condition	売る
組織	に	'goal	売る
人間	に	'goal	売る
場所	に	'goal	売る
人間	他	'goal	売る
具体物	で	'implement	売る
方向	で	'implement	売る
方法	で	'implement	売る
価格	で	manner	売る
具体物	を	'object	売る
建造物	を	'object	売る
指示物	を	'object	売る
設備	を	'object	売る
土地	を	'object	売る
表象物	を	'object	売る
不動産	を	'object	売る
金銭等価物	を	'object	売る
電気	を	'object	売る
イベント	を	'object	売る
書いた物; 書類	を	'object	売る

証券	を	'object	売る
情報媒体	を	'object	売る
権利	を	'object	売る
抽象物	を	'object	売る
方法	を	'object	売る
誇り	を	'object	売る
情報	を	'object	売る
価格	を	'object	売る
金銭	を	'object	売る
指示範囲	を	'object	売る
数量	を	'object	売る
組織	を	'object	売る

5. 最後に

4章は基本的に元データを自動的に抽出し、データ整理を行なったものであるが、関係子変換表を作成するために、作業過程において手作業の検討も行なった。

これによって、元データの語義の分け方の問題や語義の割り当て誤りのために、それに関係して関係子や体言の意味グループに異質なものが存在してしまっていることがわかる。概念関係の検討は、動詞を概念で分類し、体言の意味や表層格の組み合わせによって概念関係を自動的に獲得する手段であるとともに、辞書の語義の切り分けの検討にもつながる。

概念関係を持ち込むことによって、

- 1) 異なる概念をグループ化することができる
- 2) 体言のまとめかたによっては、同じ概念と思われていたものの詳細化ができることになる。これはまさに相反する方向であるようにみえて、言語分析において必要な二つの方向であると思われる。

参考文献

- 1) 荻野孝野、小林正博「EDR概念記述の抽象化による動詞パターンについて」情報処理学会自然言語処理研究会 138-6 (2000.7)
- 2) 大石亨、松本裕治「格パターンを利用した概念関係パターン獲得手法について」情報処理学会自然言語処理研究会 104-18 (1994.11)
- 3) 荻野孝野、橋田浩一、飯島満、関口知子「関係子の体系の間のアライメント」言語処理学会第7回年次大会 (2001.3)
- 4) 国立国語研究所: 日本語における概念関係パターンと表層格の対応関係, 国立国語研究所研究報告 113, 国立国語研究所, 1997年3月, 三省堂