

E-001

2 項関係に基づく概念グラフによる自然言語処理の提案 A proposal for the Natural Language Processor system using conceptual graph structure based on binary relation

今泉 克己[†]
Katsumi Imaizumi

田原 恭二[†]
Kyoji Tahara

1. はじめに

本稿は、語と語の関係を2項関係とし、2項関係を重ねあわせる事で3項以上の関係をあらわす概念グラフを自然言語から得る方法を提案する。

2項関係は、主語と述語、動詞と動詞の取る目的格、動詞と動詞の取るその他の格、その他の辞書上の情報が語と語の関係をあらわす関係子である場合に、項と項の関係 [項 A — 関係子 — 項 B] に直し記憶する事で得られる。

その結果、1つの文・句・任意の参照範囲から複数の関係構造が得られる場合これを許し、それぞれを処理結果とする事ができる。

また、並立する文の間で共有される語がある場合、共有語にも関係子を張り、文と文を連結する事ができる。

さらに、適切な文脈中で同じ主体を現す等、組となる2項関係があれば、文脈中でまとまりを認識し組として記憶する事ができる。

1.1 手順の骨子

(1) 文を形態素解析にかけ、語の区切りの認識を行う。
(2) 依存文法 (係り受け文法) の規則と以下の方法を用いて文節に分ける。

- ・ 近くの文節にかかる。
- ・ 読点など手がかりを使う。
- ・ 共起確率付きの語彙辞書を使い認識する。

(3) 文節から2項関係を作成する。

1.2 入力と出力の例

入力文: [美しい 日本の 私]

出力例1: [私 — attr — 美しい]

[私 — partof — 日本]

出力例2: [私 — partof — 日本]

[日本 — attr — 美しい]

2. 関係子について

以下に関係子の構成例を示す。

2.1 attr

属性や性質を示す関係子。「ビル — attr — 高い」。

2.2 hasa

所有や所属を示す関係子。「自動車 — hasa — エンジン」。

2.3 isa

概念的な上下関係を示す関係子。「人間 — isa — 哺乳類」。

2.4 partof

部分を示す関係子。「エンジン — partof — 自動車」。

2.5 on

接する事や上にある事を示す関係子。「本 — on — 机」。

2.6 in

空間にある事を示す関係子。「葛飾区 — in — 東京」。

2.7 between

間にある事を示す関係子。「9時 — between — 12時」。

2.8 動詞格 (動作主格, 対象格, 道具格)

2.9 相同の関係子 (同一の事象)

between や動詞格 (動作主格, 対象格, 道具格) は3項関係にも見えるが、辞書を併用し文にも関係子を適用する事で2項関係として扱ってよい。

3. 複数の文意を得る処理結果の例

3.1 入力文 A: [報告書に書かれた意見を取り入れた]

(1) 出力例 A1:

[(主体) — 取り入れた — 意見]
[意見 — isa — 報告書に書かれた]
注: 中間表現

(2) 出力例 A2:

[(主体) — 取り入れた — 意見]
[意見 — attr — 書かれた]
[意見 — on — 報告書]

3.2 入力文 B: [報告書に言われた意見を取り入れた]

(1) 出力例 B1:

[主体 A — 言われた — 主体 B]
[主体 A — 言われた — 意見]
[主体 A — 取り入れた — 意見 (を)]
[主体 A — 取り入れた — 報告書 (に)]

3.3 入力文 C: [今日と昨日で何も変わらない]

(1) 典型的な構文木 C1:

```

今日と —
      |
      | 昨日で —
      |         |
      |         | 何も —
      |         |         |
      |         |         | 変わらない
  
```

(2) 本提案事例 C2:

```

[ 今日 — hasa — (状態) ]
      |
      | (相同)      ... 変わらない
      |
      | [ 昨日 — hasa — (状態) ]
  
```

[†]凸版印刷株式会社 Toppan Printing Co., Ltd.

この場合の between は、[(状態) 時制 今日] と [(状態) 時制 昨日] を 1 つにした形で、[(状態) between 今日] と [(状態) between 昨日] の 2 つの 2 項関係として記憶する事ができる。

(3) 本提案事例 C3 :

[(状態) — attr — 変わらない]
[(状態) — between — 今日]
[(状態) — between — 昨日]

3.4 入力文 D : [人物 A と人物 B は隣に住んでいる]

(1) 構文木 D1 :

```

人物 A と —
      |
      |
人物 B は —
      |
      |
      隣に —
      |
      |
      住んでいる

```

(2) 本提案事例 D2 :

[(主体) — isa — 人物 A]
[(主体) — isa — 人物 B]
[(主体) — 住んでいる — (場所)]
[(場所) — attr — 隣]

(3) 本提案事例 D3 :

```

[人物 A — 住む — (場所 A)]
      |
      |
      隣
      |
      |
[人物 B — 住む — (場所 B)]

```

4. これまでの課題

4.1 係り受け文法/依存文法 (Dependency Grammar)

- ・係り受け解析に加え、形態素の単位も考える。[1]
- ・文と文の関係の把握には、係り受け解析は逆に重要。
- ・語の素性だけに頼る (Bag-of-words) のは困難。[2]

4.2 格文法 (Case Grammar)

- ・日本語は句構造だけでは文法的関係が明らかにならない。[3]

4.3 意味素性 (semantic feature)

- ・定義素性、特徴素性を語彙辞書に備える。

4.4 格フレーム (Case Frame)

- ・Fillmore の国際プロジェクト [4] に実践だけでなく理論面も情報が集約されている。

5. この提案の特徴と効果

5.1 データ構造上の特徴

本提案では、矛盾や整理のつかない事態も記述するため、通常非循環と限定する所をラベル付有向グラフを用い、循環を許す。

すなわち、原文そのものと、原文を解析して作成したグラフを対応づけ、更に格辞書スロットと素性情報付語彙辞書スロットにも対応づけする。

これらを実現する際のデータ構造上の利点からリスト構造上にグラフを表現するデータ構造を採用する。なぜなら、グラフ構造を用いる事で、グラフの共通部分を複数の文脈間で共有する事やグラフの構造比較が容易になるからである。

5.2 事例から辞書を補足する

辞書の初期構築時に行う関係子の設定により解析結果の精度が決まる。事例の検索を行いヒットした別解釈によって辞書の情報を補い変更する。辞書を使いながら事例で補足する事ができる。

5.3 C.S.Peirce^[6] の存在グラフ (Existential Graphs)

Sowa が Peirce の解題をしている。[5][7] 1 階述語の論理式のモデルと辞書スロットに付番し参照する事で可能であると考えられる。2 項関係で 3 項以上の関係を代置できるという着想は、ここから得た。

5.4 循環グラフのモデルと公理

集合論的矛盾 (ラッセルのパラドクス) とされる文や自分を指す概念を扱いたい。AFA (Anti-foundation axiom) に注目している。[8]

6. おわりに

この研究の現実的課題として、係りの交差、依存文法と並列関係、「魚を焼くけむり」のような文の処理例の蓄積と分析を挙げる。

また、関係子の着想を溝口^[9]に負うが、is-a, instance-of, subset-of, part-of, member-of などの「関係」が既に同文献中で問いが示した通り基本的な関係であるが故に関係の意味は明確な定義を必要としている。

7. 参考文献

- [1] 岩倉, 高橋, 飯田, 乾: KURA: 統一のかつ宣言的知識記述に基づく言い換えエンジン, 第 63 回情報処理学会全国大会, デ-19, 2001.
- [2] 山崎, 新保, 松本: 系列パターンを素性とした論文概要文の自動分類, 研究報告「知能と複雑系」アブストラクト, No.132-023.
- [3] 山本, 構文解析 (1) 自然言語処理特論 (4), <http://nlp.nagaokaut.ac.jp/~ykaz/edu/NLP/slide/4-Apr28.pdf>
- [4] FrameNet Project
<http://www.icsi.berkeley.edu/~framenet>
<http://www.nak.ics.keio.ac.jp/jfn/ja/index.html>
- [5] John F.Sowa, Conceptual Structures: Information Processing in Mind and Machine, Addison-Wesley, 1984.
- [6] パース, 連続性の哲学, 岩波文庫青 688-1, 岩波書店, 2001.
- [7] Existential Graphs, MS 514 by Charles Sanders Peirce, with commentary by John F.Sowa, <http://www.jfsowa.com/peirce/ms514.htm>
- [8] Hyperset Theory
<http://www.cs.bilkent.edu.tr/~akman/jour-papers/air/node8.html>
- [9] 溝口, オントロジー工学序説—内容指向研究の基盤技術と理論の確立を目指して, 人工知能学会誌, Vol.12, No.4, pp.559-569, 1997.
<http://www.ei.sanken.osaka-u.ac.jp/pub/miz/Ont97.pdf>