

言い換え表現の述語項構造への正規化と テキスト含意関係認識での利用

小谷 通隆

柴田 知秀

黒橋 禎夫

京都大学大学院 情報学研究科

{odani,shibata,kuro}@nlp.kuee.kyoto-u.ac.jp

1 はじめに

自然言語は自由度が高いため、同じ内容を表現するのに様々な表現を使うことが出来る。そのような様々な表現のずれをいかにして柔軟に吸収するかは自然言語処理における重要な課題である。本論文ではまず、言い換え表現を網羅的に整理し、構文・格・省略解析に基づいて述語項構造に正規化することでこれらを同一視する手法を提案する。次に、正規化された述語項構造を単位としてマッチングをとることにより、テキスト間の含意関係を認識する。評価セットを用いて実験を行なったところ、本手法の有効性を示すことができた。

2 言い換え表現の述語項構造への正規化

これまで様々な観点から言い換え事例の分類・整理が行われ、言い換えの生成、認識の手法が提案されている [5, 7]。我々は言い換え表現を述語項構造の観点から網羅的に整理し、述語項構造へ正規化することでこれらを同一視する手法を提案する。言い換えを以下の 4 つに分類し、形態素解析器 JUMAN、構文解析器 KNP による構文・格解析 [2]、照応・省略解析 [3] の結果に基づいて以下のように言い換え表現の正規化を行う。

2.1 述語または項単体における言い換え

表 1 に述語または項単体における言い換え表現を整理し、それぞれの現象の正規化について例を示す。JUMAN の辞書では各単語に対してその代表表記が与えられている。例えば「子ども」や「こども」に対しては「子供」が代表表記として与えられており、この情報をもとに 1-b の正規化を行う。さらに、JUMAN の辞書には「話せる」が「話す」の可能動詞であることや「おっしゃる」が「言う」の尊敬動詞であることなどが記述されており、3-b, 3-d の正規化を可能とする。また、KNP では機能語列のパターンに対してそ

表 1: 述語または項単体における言い換え

分類	表現	正規化
1-a	子供が話す	話す ガ 子供
1-b	子どもが話す (異表記)	話す ガ 子供
2-a	彼に本を返した	返した ヲ 本 ニ 彼
2-b	彼に本を返却した (同義語)	返した ヲ 本 ニ 彼
3-a	英語を話す	話す ヲ 英語
3-b	英語を話せる (機能表現)	話す (可能) ヲ 英語
3-c	英語を話すことができる (機能表現)	話す (可能) ヲ 英語
3-d	先生がおっしゃった (機能表現)	言った (尊敬) ガ 先生
4-a	大きい	大きい
4-b	大きくくはない (二重否定)	大きい
4-c	小さくない (反義語の否定)	大きい
4-d	お茶だけある (限定と否定)	ある ガ お茶 (限定)
4-e	お茶しかない (限定と否定)	ある ガ お茶 (限定)
5-a	必要以上に食べた	食べた 修飾 必要以上に
5-b	食べ過ぎた (複合動詞)	食べた 修飾 必要以上に

の役割が定義されており「することができる」が可能表現を表すことなどを認識でき、3-c などの正規化を可能とする。さらに、我々は国語辞典とウェブテキストを利用して語や句の同義・上位下位・反義関係を自動的に獲得しており、2-b などの正規化を可能とする [4]。

2.2 述語項構造における言い換え

表 2 に述語項構造における言い換え表現を整理し、それぞれの現象の正規化について例を示す。Kawahara らは格フレームをウェブから自動構築しており、これを用いて格解析を行っている [2]。さらにこの格フレームには、例えば受身表現「慕われる」のガ格が「慕う」

表 2: 述語項構造における言い換え

分類	表現	正規化	分類	表現	正規化
1-a	太郎が向かった	向かった ガ 太郎	8-a	太郎が花子の先生だ	先生だ ガ 太郎 ノ 花子
1-b	太郎は向かった (提題化)	向かった ガ 太郎	8-b	花子が太郎の生徒だ (一方向の関係表現)	先生だ ガ 太郎 ノ 花子
1-c	パソコンが使える (提題化)	使える ヲ パソコン	9-a	太郎が次郎の親友だ (双方向の関係表現)	親友だ ガ 太郎 ガ 次郎
2-a	数学を勉強する	勉強する ヲ 数学	9-b	次郎が太郎の親友だ (双方向の関係表現)	親友だ ガ 太郎 ガ 次郎
2-b	数学の勉強をする (サ変名詞)	勉強する ヲ 数学	10-a	車と自転車が衝突した (相互格)	衝突した ガ 車 ガ 自転車
3-a	取引先が注文する	注文する ガ 取引先	10-b	車が自転車と衝突した (相互格)	衝突した ガ 車 ガ 自転車
3-b	取引先から注文される (受身表現)	注文する ガ 取引先	11-a	彼が背が高い	高い ガ 背 ガ 2 彼
3-c	取引先から注文を受ける (受身表現)	注文する ガ 取引先	11-b	彼の背が高い (二重主語構文)	高い ガ 背 ガ 2 彼
4-a	花子が向かった	向かった ガ 花子	12-a	開示請求が一件あった	あった ガ 開示請求 修飾 一件
4-b	太郎が花子に向かわせた (使役表現)	向かった ガ 花子 使役 太郎	12-b	一件の開示請求があった (数量詞の遊離)	あった ガ 開示請求 修飾 一件
5-a	風が扉を閉めた (自他動詞)	閉めた ガ 風 ヲ 扉	13-a	柳が芽をふく	ふく ガ 柳 ヲ 芽
5-b	風で扉が閉まった (自他動詞)	閉めた ガ 風 ヲ 扉	13-b	柳に芽がふく (湧出動詞)	ふく ガ 柳 ヲ 芽
6-a	弟に父がお金を貸した (授受動詞)	貸した ガ 父 ヲ お金 ニ 弟	14-a	杯に日本酒を満たす	満たす ヲ 日本酒 ニ 杯
6-b	弟が父にお金を借りた (授受動詞)	貸した ガ 父 ヲ お金 ニ 弟	14-b	杯を日本酒で満たす (壁塗り・場所格交替)	満たす ヲ 日本酒 ニ 杯
6-c	弟が父からお金を借りた (授受動詞)	貸した ガ 父 ヲ お金 ニ 弟	15-a	有意義に週末を過ごす	過ごす ヲ 週末 修飾 有意義に
7-a	兄の方が弟より背が高い (比較表現)	高い ガ 背 ガ 2 兄 ヨリ 弟	15-b	有意義な週末を過ごす (修飾要素)	過ごす ヲ 週末 修飾 有意義に
7-b	兄より弟の方が背が低い (比較表現)	高い ガ 背 ガ 2 兄 ヨリ 弟			

のヲ格に対応付けられており、格解析結果と格フレーム対応知識を用いて 2-b のような正規化を行う。これが文の解析結果であり、これによって同じ意味を示す言い換え表現を同一のものとして扱うことが可能となる。格フレーム対応知識において、使役表現のように正規化後の述語項構造に格が存在しないものについては「使役格」など新しい格を定義している。

2.3 格要素が述語項構造を構成する言い換え

表 3 に格要素が述語項構造を構成する言い換え表現を整理し、それぞれの現象について例を示す。KNP は用言以外にもサ変名詞、動作性名詞についても格解析を行っており、これに基づいて正規化を行う。

2.4 構文の変換を伴う言い換え

表 4 に構文の変換を伴う言い換え表現を整理し、それぞれの現象の正規化について例を示す。KNP は強調構文や連体節においても格解析を行っており、それに基づき正規化を行う。

3 テキスト 含意関係認識での利用

近年 Textual Entailment の認識 (Recognizing Textual Entailment, RTE) が注目を集めている [1]。RTE とは、以下に示すような text (以下 t と略記)、hypothesis (以下 h と略記) と t から h が推論されれば YES、そうでなければ NO とする推論関係を人手で付与したデータを入力とし、この推論関係をシステムに正しく判断させることを目的とする。

表 3: 格要素が述語項構造を構成する言い換え

分類	表現	正規化
1-a	言葉が温かいと感じた	感じた 補文 温かい ガ 言葉
1-b	言葉を温かく感じた (補文)	感じた 補文 温かい ガ 言葉
1-c	言葉に温かみを感じた (補文)	感じた 補文 温かい ガ 言葉
1-d	言葉の温かみを感じた (補文)	感じた 補文 温かい ガ 言葉
2-a	友人が結婚したことを喜ぶ	喜ぶ ヲ 結婚した ガ 友人
2-b	友人の結婚を喜ぶ (サ変名詞・動作性名詞)	喜ぶ ヲ 結婚した ガ 友人

t: 彼に借りた本を無くしてしまった

h: 彼が本を貸してくれた

推論関係: YES

このタスクに対し、述語項構造を単位としてマッチングをとることで t と h の含意関係を認識するシステムを構築する。まず、2 節に示した正規化によって表現のずれを吸収し、h に存在する全ての述語項構造についてそれらを含意する述語項構造が t に存在するときに、h が t に含意されていると判断する。

述語項構造の含意関係は以下のように定義する。まず、正規化した述語項構造が一致するものは含意関係にあると判断する。そして格要素や述語の上位下位関係が導く述語項構造の含意関係を表 5 のように定義する [6]。この含意関係は、(1) 述語項構造が性質記述文であるかどうか、(2) 肯定文であるか否定文であるかの 4 つのタイプに分類される。例えば、肯定文の性質記述文では t の格要素が h の格要素の上位表現であるとき t は h を含意し、h の述語が h の述語の上位表現であるとき t は h を含意する。前者では例えば「マナガツオは海の生物だ → カツオは海の生物だ」という含意関係が成り立ち、後者では例えば「カツオは海の生物だ → カツオは生物だ」という含意関係が成り立つ。ここで述語・格要素の上位下位関係は以下のように定義する。

格要素の上位下位関係

- 名詞の上位下位関係:
マナガツオ → カツオ, インコ → 鳥
- 修飾要素の欠如:
チョコレート工場 → 工場, 熟した柿 → 柿

述語の上位下位関係

- 用言の上位下位関係:
昼寝する → 眠る, 大学生だ → 学生だ

表 4: 構文の変換を伴う言い換え

分類	表現	正規化
1-a	太郎が次郎に勝った	勝った ガ 太郎 ニ 次郎
1-b	次郎に勝ったのは太郎だ (強調構文)	勝った ガ 太郎 ニ 次郎
2-a	先生が本を書いた	書いた ガ 先生 ヲ 本
2-b	先生が書いた本 (連体節)	書いた ガ 先生 ヲ 本
2-c	先生の書いた本 (連体節)	書いた ガ 先生 ヲ 本
2-d	体が小さい馬 (連体節)	小さい ガ 2 馬 ガ 体
2-e	体の小さい馬 (連体節)	小さい ガ 2 馬 ガ 体
2-f	小さい体の馬 (連体節)	小さい ガ 2 馬 ガ 体
3-a	英語を徹底的に教育する	教育する ヲ 英語 修飾 徹底的に
3-b	英語の教育を徹底する (副詞句)	教育する ヲ 英語 修飾 徹底的に
4-a	国民の義務は納税だ	納税だ ガ 国民の義務
4-b	国民が納税の義務がある (内容名詞)	納税だ ガ 国民の義務

- 格要素の欠如:

海の生物だ → 生物だ

昨日生まれた → 生まれた

また、マッチングにおいて例外処理を設ける。t の補文節内の述語項構造と h の述語項構造のマッチングにおいては、t の補文節の主節の内容を考慮する必要がある。「勘違いする」「間違う」など特定の述語に係る補文節内の述語項構造は、実際には起こっていないという否定の属性を与えてマッチングを行う。

t: 私が転んだと勘違いした

h: 私は転んだ

推論関係: NO

また、仮定法過去を用いた表現における条件節、およびその主節についても同様の処理を行う。

t: 彼が選ばれたら彼女は後悔しただろう

h: 彼が選ばれた

推論関係: NO

t: 彼が選ばれたら彼女は後悔しただろう

h: 彼女は後悔した

推論関係: NO

表 5: 述語項構造の含意関係

	性質記述文				性質記述文以外			
	t の格要素が上位		h の述語が上位		h の格要素が上位		h の述語が上位	
肯定文	t: 生物だ ガ カツオ ノ 海	h: 生物だ ガ マナカツオ ノ 海	t: 生物だ ガ カツオ ノ 海	h: 生物だ ガ カツオ	t: 飼っている ヲ インコ	h: 飼っている ヲ 鳥	t: 昼寝した ハ 寝た	h: 寝た
否定文	t: 生物でない ガ 家具	h: 生物でない ガ 机	t: 生物でない ガ 机	h: 生物でない ガ 机 ノ 海	t: 飼っていない ヲ 鳥	h: 飼っていない ヲ インコ	t: 寝なかった ハ 昼寝しなかった	h: 昼寝しなかった

4 実験と考察

RTE で扱われる事例は「彼から本を借りた 彼が本を持っていた」といった推論的な事例から「彼から借りた本を無くした 彼から本を借りた」といったテキスト含意的な事例まで幅広く、様々な要因に導かれる。我々は推論要因を整理し、1つの要因によって推論を判断できるような事例を集めた推論コーパスを作成している[6]。このうち言い換え表現の正規化によって認識が可能となる推論要因「節」「補文」「主語の変換」「強調構文」の事例を用いて実験を行った。実験結果を表6に示す¹。

システムが正しく推論関係を判断できたものは表の左上と右下の部分となる。システムの誤りについて表の右上の事例数が左下の事例数より多いが、これは事例の大半が正規化を行わなければシステムがNOであると判断するものであるからである。システムが誤った原因は、大半が文の解析失敗によるものであり言い換え表現の述語項構造への正規化による悪影響はなかった。以下は正しく推論関係を判断できた例である。

節	t: 太郎に貸した本が無くなった h: 太郎に本を貸した 推論関係: YES
補文	t: 彼は私が転んだと勘違いした h: 私は転んだ 推論関係: NO
主語の変換	t: 英外相がブラジル外相に謝った h: ブラジル外相が英外相に謝られた 推論関係: YES
強調構文	t: 彼が必死に叩いたのは窓だった h: 彼は窓を必死に叩いた 推論関係: YES

5 おわりに

言い換え表現を述語項構造へ正規化することで表現のずれを吸収する手法を提案し、テキストの含意関係

¹表中では YES を Y, NO を N と表記する。

表 6: 実験結果

「節」				「補文」			
		システム				システム	
		Y	N			Y	N
正解	Y	32	11	正解	Y	15	7
	N	7	29		N	4	23
「主語の変換」				「強調構文」			
		システム				システム	
		Y	N			Y	N
正解	Y	18	31	正解	Y	11	8
	N	3	29		N	0	10

認識においてその有効性を示した。今後の課題としては述語項構造間の関係を辞書やウェブから自動獲得する。また、モダリティの違いが生み出す推論関係について整理する。また、本研究で得られた成果を検索などのアプリケーションで利用する。

参考文献

- [1] Danilo Giampiccolo, Bernardo Magnini, Ido Dagan, and Bill Dolan. The Third PASCAL Recognizing Textual Entailment Challenge. In *Proceedings of ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pp. 1–9, 2007.
- [2] Daisuke Kawahara and Sadao Kurohashi. A fully-lexicalized probabilistic model for japanese syntactic and case structure analysis. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 176–183, 2006.
- [3] Ryohei Sasano, Daisuke Kawahara, and Sadao Kurohashi. A fully-lexicalized probabilistic model for japanese zero anaphora resolution. In *Proceedings of the 22nd International Conference on Computational Linguistics*, pp. 769–776, 2008.
- [4] Tomohide Shibata, Michitaka Odani, Jun Harashima, Takashi Onishi, and Sadao Kurohashi. SYNGRAPH: A flexible matching method based on synonymous expression extraction from an ordinary dictionary and web corpus. In *Proc. of IJCNLP2008*, 2008.
- [5] 乾健太郎, 藤田篤. 言い換え技術に関する研究動向. 自然言語処理, Vol.11, No.5, pp. 151–198, 2004.
- [6] 小谷通隆, 柴田知秀, 中田貴之, 黒橋禎夫. 日本語 textual entailment のデータ構築と自動獲得した類義表現に基づく推論関係の認識. 言語処理学会 第 14 回年次大会 発表論文集, pp. 1140–1143, 2008.
- [7] 黒橋禎夫, 酒井康行. 日本語表現の柔軟な照合. 言語処理学会 第 7 回年次大会 発表論文集, pp. 343–346, 2001.