

基本文関係に分解した含意関係認識日本語評価データの構築

金子 貴美[†] 宮尾 祐介[‡] 戸次 大介[†]

[†]お茶の水女子大学大学院 人間文化創成科学研究科 理学専攻 情報科学コース

[‡]国立情報学研究所

[†]{kaneko.kimi, bekki}@is.ocha.ac.jp

[‡]yusuke@nii.ac.jp

1 はじめに

自然言語処理の分野では、1990年代以降、統計的言語処理の研究が主流となり、基礎的な解析技術が飛躍的に発展してきた。しかし、これらの技術は意味の観点から見ると表層的な処理に留まっており、長年の懸案である「意味処理」の実現には至っていない現状がある。したがって、その実現に向け、一歩踏み込む必要性が強まってきている。そのような「意味処理」の1つとして、どの表現とどの表現が同じ情報を伝えているかの判断がある。その判断を行う技術が含意関係認識である。

含意関係認識とは、与えられた2文の間に含意関係が成り立つかどうかを自動認識する技術であり、文 t1 と t2 の間の含意関係とは、人間が「t1 が正しいとき、t2 も (ほぼ) 正しい」と判断する関係のことである¹。以下、これを「t1 は t2 より狭い意味を持つ」、または「t2 は t1 より広い意味を持つ」ということもある。以下に例を示す。

例 1) 含意関係認識の例

- 含意関係: 成立
- t1: 川端康成は、「雪国」などの作品でノーベル文学賞を受賞した。
- t2: 川端康成は「雪国」の作者である。

この判断を計算機が行うのは難しく、まずは文を個別の言語的・意味的現象に書き下した評価データを構築しなければならない。しかし、そのような日本語評価データとその構築手法が存在しないという問題がある。

本研究では、上記のような問題を解決すべく、より基本的な含意関係の連鎖として書き下した日本語評価データの構築手法を提案する。本手法により、個々の関係に焦点を当てて解析することが可能となるため、含意関係認識手法の進展が期待される。また、計算機に文のブレイクダウンを自動的に行わせる方法を考える材料となる可能性がある。

本論文では、構築手法の妥当性を考察する。第2節では、関連研究と研究背景について述べ、第3節では、日本語評

価データの書き下し方と含意関係ラベルの分類方法について述べる。第4節では、実際に本手法に基づいて評価データを構築し、過去の手法により構築されたデータと比較して評価した。

2 研究背景

含意関係認識の既存研究の問題点を指摘したのとして、Sammons ら [1] の研究がある。彼らは、含意関係を正しく捉えるために、より基本的な含意関係の連鎖として文を書き下す方法を確立させる必要があると指摘すると共に、それによって、自然言語処理の他の既存手法をどう組み合わせることが出来るか、また、解決できない問題は何か、が把握可能となると述べている。また、彼らの提案手法により、Pascal RTE Challenges の RTE-5 データセットから抽出した 210 組の文に対し、22 種類の関係ラベルを用いてアノテーションを行い、手法の妥当性を評価する際に、原文データよりも構築したデータのほうが解析精度が高くなると報告している。

また、含意関係認識のために文を書き下した評価データの構築手法を提案した関連研究として、Bentivogli ら [2] の研究がある。彼らは Sammons らと同様に、RTE-5 データセットから抽出した、90 組の文のペアを提案手法により詳細な含意関係に書き下し、36 種類の関係ラベルで各文にアノテーションを行っている。

これらの手法では、英文のみから成る RTE-5 データセットを用いた場合の妥当性しか評価されておらず、これらの評価結果から日本語データに適用した際の妥当性を判断するのは難しい。日本語データにおける妥当性を判断するには、日本語の評価データを用いる必要がある。日本語の評価データとして小谷ら [3] のデータセットが存在するが、基本的な含意関係が1組の文のペアに1つしか現れないよう構築されているため、これで実際の文のペアを書き下した際の妥当性を判断するのは難しい。したがって、今回は RITE-2²のバイナリクラスサブタスク (以下、「BC サブタスク」) の一部のデータを用いて評価データを作成しながら、日本語の場合の評価

¹自然言語処理でいう「含意」は、論理学でいう意味論的含意とは異なり、現象に対する一種のカバータムである。形式言語論でいう「含意」「前提」「会話の含意」を含む。

²<http://www.cl.ecei.tohoku.ac.jp/rite2/doku.php>

データ構築手法を策定することとした。RITE-2とは、含意関係認識の研究を促進させることを目的とした評価型ワークショップである。BCサブタスクという、与えられた文のペアの含意関係が成り立つか否かを判別する2値分類問題をはじめとした、4つのサブタスクから構成されている。BCサブタスクのデータの一部を用いた理由は、この評価データを作成することによって、解析精度の向上に寄与することが期待されると共に、作成したデータをBCサブタスクに用いることにより、提案手法の妥当性の検証もできるためである。

3 提案手法

含意関係を基本文関係にブレイクダウンし、含意関係認識のための日本語評価データを作成する方法を考案した。基本文関係の一覧は表2に示す。基本的に書き換えていく度に意味が広がるよう書き下し、t1を徐々にt2に近づけるよう変換することとした。また、第1節で述べたように、2文の関係を自動認識させるために必要な言語現象が書き出せるよう、t1とt2の書き換えは関係の繋がりを考慮して行うこととし、個々の現象にフォーカスし分析できるよう、書き下した各ペアでは、複数の関係が成立しないように書き換えることとした。

たとえば、先の例1のような文のペアが与えられたとする。この2文のペアを以下のように書き下し、関係ラベルを付与した。

例1') 例1を書き下した例

- **[フレーズ:含意・前提]** 川端康成は「雪国」などの作品の作者である。
- **[集合・リスト]** 川端康成は「雪国」の作者である。
- **[一致]** 川端康成は「雪国」の作者である。

一方、含意関係が成り立たない2文が与えられた場合の例を以下に示す。

例2) 含意関係が成り立たない場合の例

- ラベル=N
- t1:パルテノン神殿は、ドーリア式神殿の最高傑作と言える作品である。
- t2:パルテノン神殿は、ヘレニズム文化の影響下で建設された。

ここで、1行目の「ラベル」とは例1の1行目同様、含意関係の成立・不成立を表す。例1のように、t1とt2の間に含意関係があれば「成立」を表す「Y」のラベルが、なければ「不成立」を表す「N」のラベルが付与される。

例2') 例2を書き下した例a

- **[修飾句削除]** パルテノン神殿は、ドーリア式神殿の作品である。
- **[主辞削除]** パルテノン神殿はドーリア式神殿である。
- **[フレーズ:含意・前提]** パルテノン神殿は、ドーリア式神殿として建設された。

- **[フレーズの不一致]** パルテノン神殿は、ヘレニズム文化の影響下で建設された。

あるいは、書き換え易さを重視し、以下のように「含意関係が成り立たない場合の関係」ラベルの途中使用も良いとする。また、t1とt2が全く無関係で書き換えるのが難しい場合など、必要に応じて複数回使っても良いとした。

例2') 例2を書き下した例b

- **[修飾句削除]** パルテノン神殿は、ドーリア式神殿の作品である。
- **[主辞削除]** パルテノン神殿はドーリア式神殿である。
- **[単語の不一致]** パルテノン神殿は、ヘレニズム式神殿である。
- **[フレーズ:含意・前提]** パルテノン神殿は、ヘレニズム文化の影響下で建設された。

3.1 t2の変換

前述までの例で示したように、原則的には、t2より狭い意味をもつt1を変換してt1より広い意味をもつt2に近づけるように書き換えることとした。t2の書き換えも許すと、含意関係が成立しない場合であっても、t1とt2が示す状況(集合)が一致するよう書き換えることができってしまうからである。逆に、t1とt2が矛盾関係にあり、t1とt2の状況の集合に重なりがない場合であれば、t2の意味を広げても結果としてt1の状況と重ならなければ含意判定が変わらないため、その場合に限り、t2を変換可能とした。以下に例を示す。

例3) t2の変換の例

- ラベル=N
- t1:ポイントカードには、クレジット機能付きとクレジット機能なしの2つの種類がある。
- t2:ポイントカードは現金専用でクレジット機能は付いていない。

例3') 例3を書き下した例

- **[集合・リスト]** ポイントカードには、クレジット機能付きもある。
- **[フレーズ:同義・類義]** ポイントカードはクレジット機能が付いているものもある。
- **[t2の編集]** ポイントカードは現金専用でクレジット機能は付いていない。
- **[修飾句削除]** ポイントカードはクレジット機能が付いていない。
- **[モダリティの不一致]** ポイントカードはクレジット機能が付いているものもある。

4 評価

BC サブタスクのデータセットの一部に関して、本手法を適用した。173組のデータを1名で書き下し、それらに対して、2名のアノテータがラベル付けを行った。173組のうち112組のデータは本手法の策定に用い、残りの61組について一致率と平均書き換え数を算出した。一致率は、以下の式で算出した。

$$\text{一致率} = \text{ラベル一致数} / \text{全体数}$$

ここで、「全体数」は「ラベル付けを行った文の数から一致ラベルを付けた文の数を引いた数」である。今回扱ったデータにおける一致率は0.83に達しており、本手法は十分に実用性があることが分かる。参考までに、算出方法はDice係数³であるため異なっているが、これは[2]の、書き下した文毎の一致率0.78より高い値となっている。

書き換え数は231回であり、平均3.79回であった。これも[2]の平均2.98回よりも若干高い値となっている。日本語データであることや、複雑なデータが多いことを考慮すると、この値は妥当であると考えられる。

表1に[2]と本研究における関係ラベルの対応を示す。

表1: [2]と本研究における関係ラベルの分布

関係ラベル	文の数	
	[2]	本研究
同義・類義	25	46
上位・下位	5	3
含意・前提	-	1
全体・部分	7	0
品詞の変換	9	0
共参照・照応	49	4
語順入れ替え	-	15
態の変化	7	7
修飾句削除	25	42
主辞削除	6	1
並列・従属節/集合・リスト	6	18
同格	3	1
関係節	1	8
時間の推論	2	1
空間の推論	2	1
数の推論	6	0
暗黙の関係	7	19
単語の不一致	3	2
モダリティの不一致	1	1
対応ラベル無し	42	74
合計	206	244

[2]の分類のうち、‘Donymy’, ‘Statements’は日本語の場合ではあまり問題とならなかったため、本研究の分類では除外した。一方で、「語順入れ替え」を追加した。表1より、[2]は本研究と比較して、「共参照・照応」が多く見られること

³Dice係数とは、単語間の意味距離を示す値で、共起頻度を中心語頻度と共起語頻度の和で割ったものを2倍して求める。アノテーションの一致率計算にも用いられる。

や、本研究では[2]と比べ「語順入れ替え」が頻発していることが分かる。これらのことから、言語の違いが関係ラベルの分布や分類の違いに現れているといえる。

また、書き換え時間は、1組あたり平均12分であった。[2]の書き換え時間が1組あたり平均15分であることから、本手法の作業方針が妥当であると考えられる。

5 問題点

アノテータ2名により付けられたラベルが異なっていた文は41組であった。そのうち、5組が「単語」と「フレーズ」、3組が「関係節」と「書き換え」、8組「同義・類義」「上位・下位」「全体・部分」「含意・前提」、5組が「暗黙の関係」と「同義・類義」といった、人により判断が揺れやすいラベル間の場合であった。このような場合は比較的多く見られたことから、再検討する必要があると考えられる。

t1からt2への書き換え経路には複数のパターンがあると考えられるが、今回の文の書き換えは1人で行ったため、人によって書き換えがどれだけ揺れるかについては考察していない。したがって、複数人の書き換えパターンを比較し、評価する必要がある。

6 まとめ

含意関係を基本含意関係・非含意関係にブレイクダウンした日本語評価データの構築方法を考案すると共に、本手法に従ってBCサブタスクのデータセットの一部を書き下し、日本語評価データを構築した。構築したデータの一致率、平均書き換え数、書き換え時間を既存研究と比較した結果から、本手法の妥当性を示した。また、言語による違いが関係ラベルやアノテーション結果の違いに現れることを指摘した。

ラベル付けの問題の検討、書き換えパターンの比較や、認識精度からの妥当性の考察など、課題は数多く残っている。今後はこれらの課題を検討する必要があると考える。

参考文献

- [1] Sammons, M., Vydiswaran, V.G., Roth, D.(2010). Ask not what textual entailment can do for you... In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. pp. 1199-1208 (2010)
- [2] Luisa, B., Elena, C., Ido, D., Danilo, G., Medea, L.L., Bernardo, M. (2010). Building Textual Entailment Specialized Data Sets: a Methodology for Isolating Linguistic Phenomena Relevant to Inference. In Proceedings of LREC 2010, Malta.
- [3] 小谷 通隆, 柴田 知秀, 中田 貴之, 黒橋 禎夫.(2008). 「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, pp.1140-1143 (2008.3).

表 2: 基本文関係一覧表

基本含意関係		
関係ラベル	例	用法説明
一致		文同士が完全に一致する場合。作成した最後の文がt2と一致するときに使う。
単語の書き換え		
同義・類義	紀元前5世紀→前5世紀, 各ポリス→諸ポリス, アケメネス朝(ペルシア帝国)→アケメネス朝ペルシア, 法廷で→法廷において	「単語」を(ほとんど)同じ意味の別の「単語」に書き換える。逆向きの書き換えが不可能な場合は除く。文脈によらず成り立つ同義表現を補う場合を含む。
上位・下位	最高裁判所裁判官→裁判官, 電子機器→コンピュータ	「単語」をその上位または下位(一方の語が指すものの集合がもう一方の語が指すものの集合を含む、is-a関係)の「単語」に書き換える。
含意・前提	妨害した→止めた, 離婚した→結婚した	動詞や形容詞(イベントを表す名詞も含む)で、前者が成り立っている状況では必ず後者も成り立っている場合。
全体・部分	千代田区→東京都, 顔→口	物とその部品、のように「単語」で指されているものに物理的に含まれるものを指す、別の「単語」に書き換える。「上位・下位」に当てはまるものは除く。
フレーズの置き換え		
同義・類義	広がっていった→形成されるようになった, 西ヨーロッパを代表する国々→西欧諸国	「単語: 同義・類義」同様、書き換え元または先が「フレーズ」の場合。
上位・下位	イエニチェリ→オスマン帝国の常備軍	「単語: 上位・下位」同様、書き換え元または先が「フレーズ」の場合。
含意・前提	ノーベル文学賞を受賞した→作者である, 施政権を行使する→統治権を得た	「単語: 含意・前提」同様、書き換え元または先が「フレーズ」の場合。
全体・部分	パリ→西ヨーロッパを代表する国	「単語: 全体・部分」同様、書き換え元または先が「フレーズ」の場合。
品詞の変換	雑誌の差し止め→雑誌を差し止める	ある品詞(例: 名詞)での表現を、同じ単語の別の品詞(例: 動詞)に書き換える。
共参照・照応	これ→アケメネス朝(ペルシア帝国), (eは)勝利した→ギリシアの諸ポリスは勝利した	代名詞や省略された項(主語や目的語など)を、文中の別の句で埋める。または、文中で同じものを指す単語を別の単語で置き換える。
文構造の書き換え		
語順入れ替え	XがYをZした→YをXがZした, AはBである→BはAである	動詞、名詞、形容詞等に係る複数の句や節の順番を入れ替える。複数の句(「Xが」「Yを」)が同じ句(「Zした」)に係る場合のみ該当。表層的に語順が入れ替わっていても、句・節以外の部分が入れ替わっている場合は、別のラベルが該当する。
態の変化	XによってYが建設された→XがYを建設した, XがYを壊した→Yが壊れた	受身、使役、能格化などの構文変形を行う(または能動態に戻す)。
修飾句削除	一気に征服し→征服し	A→Bという係り受け関係があるときに、Aを削除する。
主辞削除	ドーリア式神殿の作品→ドーリア式神殿, 日本代表が勝利した→日本が勝利した	A→Bという係り受け関係があるときに、Bを削除する。Bを削除しても意味が変わらない場合のみ使う。意味が変わる(逆向きの含意関係不成立)時は除く。
並列・従属節	Aし、Bした→Bした, AであるがB→Aである	連用修飾や接続詞で2つ(以上)の節がつながっているとき、いずれかの節を残して別の節を削除する。助詞化した表現が修飾する場合を除く。
集合・リスト	「雪国」などの作品→「雪国」, バルカン半島, エジプト→バルカン半島,	「と」「,」「.」「など」などによりいくつかの名詞が列挙されているとき、そのうち1つを抽出する。
同格	ギリシアの都市国家(ポリス)→ギリシアの都市国家はポリスと呼ばれる	A(B)のような同格表現(AとBが同じものを指す)時、「AはBと呼ばれる」「AはBである」と書き換える。A(Bともいう)やA(Bとも呼ばれる)のように、同格を示す表現が括弧内に明示される場合や、「A-Bともいう」の様な、括弧に準ずる表現の場合も含む。単にAやBを削除する場合は、「修飾句削除」とする。
関係節	小アジア半島に建国したオスマン帝国→オスマン帝国は小アジア半島に建国した, さんまを焼くにおいて→さんまを焼く, 後ウマイヤ朝はウマイヤ朝が再興した王朝→後ウマイヤ朝はウマイヤ朝が再興した	関係節+名詞で表されている名詞句を文に変換する。名詞が関係節の中に入る場合(内の関係)と入らない場合(外の関係)の両方共、「関係節」のラベルを付ける。関係節+名詞以外の部分も書き換える場合は、「関係節」ではなく「フレーズの書き換え」とする。
推論		
時間の推論	1931年→1930年代, 1821年→19世紀, Aは18世紀末に発明された→Aは19世紀には存在していた	年代・時間についての包含関係や、時間配置に関する推論が必要な場合。
空間の推論	ギリシャ西南部やエーゲ海→ギリシャ各地, Aの上にBがある→Bの下にAがある	地理の知識や空間配置に関する推論が必要な場合。空間配置を考慮する必要がなく、国名や地名に関する「上位・下位」や「全体・部分」の書き換えに相当する場合は除く。
数の推論	20万とも50万とも言われるペルシア軍→ペルシア軍の大群	数学的推論や数に関する推論が必要な場合。時間に関する推論に該当する場合は除く。
暗黙の関係(主辞・修飾句挿入)	日本の軍隊→日本が保有していた軍隊, 日本が勝った→日本の代表チームが勝った	「AのB」「AからのB」など暗黙的に関係付けられた句の間の関係を明示化する。もしくは、「A」の(文脈から推測できる)暗黙の主辞や修飾句を補って「AのB」「BのA」「AというB」「BというA」とするような場合も含む。変換元より変換先の方が情報が多い(より詳細)な場合に使う。「上位・下位」「含意・前提」「主辞削除」などの逆向きの関係。補う語が元の文中に含まれる場合は除く。
その他の推論	津波は20mに達した→津波は最大波高が20mに達した	上記の関係以外で、一般常識に基づく推論が必要な場合。
基本非含意関係		
単語の不一致	小アジア半島⇔バルカン半島, ヘレニズム文化⇔ドーリア式	文の中で同じ場所にある「単語」が、お互い「同義・類義」「上位・下位」のような関係に無く、書き換えられない場合。
フレーズの不一致	オスマン帝国のスルタン⇔清の皇帝	「単語の不一致」同様、「フレーズ」が書き換えられない場合。
モダリティの不一致	到達した⇔到達しなかった, しなければならない⇔してもよい	動詞に続くモダリティ(主に助動詞)が異なり、書き換えられない場合。
係り受けの不一致	ドイツが⇔ドイツを, ギリシャによって⇔ギリシャに対し, XがYを侵略した⇔(eが)Yを侵略した	係り受け関係にある2つの文節において、それぞれの内容語は同じだが、その関係が異なる場合。必要な項が存在しない場合も含む。
修飾の不一致	わずかに拡大した⇔大幅に拡大した	修飾句が異なり、書き換えられない場合。必要な修飾句が存在しない場合も含む。
時間の不一致	1921年⇔18世紀	時間表減に矛盾がある場合。
空間の不一致	AとBは重なっている⇔Aの下にBがある	空間に関する表現に矛盾がある場合。単語やフレーズの不一致に該当する場合は除外。
数の不一致	3か国⇔4か国, わずか3か国⇔多くの国々	数値表現に矛盾がある場合。時間表現の場合を除く。