

入れ子構造による文章の意味理解手法の提案 Semantic Understanding Method of Sentence Based on Nested Structure

小谷 涼[†] 吉村枝里子[‡] 土屋誠司[‡] 渡部広一[‡]
Ryo Kotani Eriko Yoshimura Seiji Tsuchiya Hirokazu Watabe

1. はじめに

現在は産業用ロボットが主流であるが、今後、医療や介護の現場で活躍する人のパートナーとなる知的なロボットが求められている。ロボットが人のパートナーとして活躍するためには円滑なコミュニケーションをとれる必要がある。人との円滑なコミュニケーションの方法として会話が挙げられる。会話をするには相手の発言を理解しなければならないため、ロボットは人の発言を理解する必要があると考えられる。そこで、人の発言の意味を理解するシステムが必要となる。本研究では、入力文を整理することにより、意味理解を行うことを目的とする。

2. 提案手法

本手法では入力文を 7W1H と述語のフレーム（以下、文意味理解フレーム）に分類し、7W1H のフレームに分類された語句や述語が体言だった場合その語句をさらに主要部と 7W1H と述語のフレーム（以下、語句意味理解フレーム）に分類することで、入力文の整理をする。文意味理解フレームと語句意味理解フレームを総称し、意味理解フレームとする。表 1 に「今日は父の日なので、私は有名な画家が描いた絵を父に贈った」と入力したときの意味理解フレームの一部を示す。

表 1 意味理解フレームの例

Who	Who+	What	When	Where	How	Why	Whom	述語
			今日					父の日だ

↓ 順接 ↓

Who	Who+	What	When	Where	How	Why	Whom	述語
私		有名な画家が描いた絵					父	贈る

↓ 主要部 ↓

主要部	Who	Who+	What	When	Where	How	Why	Whom	述語
絵	有名な画家								描いた

表 1 に示した意味理解フレームの一番上と上から二番目のフレームが文意味理解フレーム、一番下のフレームが語句意味理解フレームである。2 つの文意味理解フレームを繋ぐ「順接」は 2 つの文を繋ぐ関係である。7W1H は一般的に用いられる 6W1H に「誰と」を示す Who+を追加したものである。動作を共にする者も情報として必要であると考られるため、分類に 7W1H を用いた。ここでは「有名な画家が描いた絵」に対する語句意味理解フレームのみを示す。

[†] 同志社大学大学院理工学研究科

Graduate School of Science and Engineering, Doshisha University

[‡] 同志社大学理工学部

Faculty of Science and Engineering, Doshisha University

したが、本手法では、文意味理解フレームに格納された全ての語句に対して語句意味理解フレームが作成される。さらに語句意味理解フレームに格納された全ての語句に対して語句意味理解フレームが作成され、これが繰り返され入れ子構造となる。

3. 提案手法の流れ

本手法の流れを以下に説明する。

3.1 文の入力

本手法では文末が述語である文のみを対象とする。本稿では「今日は父の日なので、私は有名な画家が描いた絵を父に贈った」という文を入力例として、以下の処理の説明に用いる。

3.2 入力文の解析

入力文に対して形態素解析と係り受け解析を行う。本研究では形態素解析器として、茶筌^[1]と Mecab^[2]、係り受け解析器として、南瓜^[3]を用いた。

3.3 入力文を分割

複文や重文は接続助詞や用言の連用形+読点などで区切ることによって分割する。文章は句点で区切ることによって分割する。ここで、「今日は父の日なので、私は有名な画家が描いた絵を父に贈った」という文は接続助詞「ので」で分割され、「今日は父の日だ」と「私は有名な画家が描いた絵を父に贈った」の 2 つの文に分割される。

3.4 分割した文をフレーム単位に分割

対象となる入力文では、南瓜で入力文を解析すると最後の文節が述語の文節となり、係り先のない文節となる。意味理解フレームに格納する候補となる文節と、述語の修飾語はそれぞれ述語の文節に係る。述語の文節以外に係る文節は、その文節を修飾する文節となる。そのため、係られていない文節の係り先を辿り、係り先が述語の文節となる文節で止めると、格納される語句と述語の修飾語を獲得することができる。述語の修飾語と述語は 1 つにまとめる。この処理は入力文の分割で分割できた文の数だけ行う。具体例ではこの処理を 2 回行い、「今日は父の日だ」、「私は有名な画家が描いた絵を父に贈った」に分割される。

3.5 フレーム格納処理

フレーム単位に分割された語句を格納するフレームの決定、文意味理解フレーム同士を繋ぐ関係の決定、語句意味理解フレームの展開を行う。格納するフレームは語句に伴う助詞と、格納する語句の最後の名詞のシソーラス^[4]の上位ノードや述語を動詞項構造シソーラス^[5]を基に作成した動詞のデータベースと参照することで決定する。文意味理解フレーム同士を繋ぐ関係は入力文をより細かい文に分割する基準となった接続詞や接続助詞などが示す意味とする。

接続詞、接続助詞とその語が示す関係を 1 対 1 でそれぞれ接続詞知識ベース (170 語) と接続助詞知識ベース (31 語) に格納し、それを利用する。語句意味理解フレームの展開は意味理解フレームに格納された 7W1H の語句と述語が体言の場合その語句に対して行う。語句の最後の用言の次の語から終わりまでをその語句についての語句意味理解フレームの主要部フレームに格納し、最後の用言までを 1 つの文として、文の解析からフレーム格納処理を再帰的に行う。展開できる語句がなくなれば、処理を終了する。

具体例では「今日は父の日だ」と「私は有名な画家が描いた絵を父に贈った」がそれぞれ別の文意味理解フレームに格納される。例えば、「私は有名な画家が描いた絵を父に贈った」についてであれば、「私」は伴う助詞が「は」であるため Who フレームに、「有名な画家が描いた絵」は伴う助詞が「を」であることと、動詞のデータベースと述語の「贈る」を参照することで What フレームに格納するというように 51 のルールに基づいた処理により分類される。そして、この入力文は「ので」で区切られたため、「ので」が意味する「順接」で 2 つの文意味理解フレームを繋ぐ。次に語句ごとに語句意味理解フレームを展開するが、「有名な画家が描いた絵」以外の語句は用言を含んでいないため、それぞれが主要部フレームに格納され処理が終了する。「有名な画家が描いた絵」についての語句意味理解フレームでは、「描いた」が用言であるため、その次の語から終わりまで「絵」となり、主要部フレームに格納され、「有名な画家が描いた」を 1 つの文とし、文の解析からフレーム格納処理まで行う。その結果が表 1 の一番下のフレームとなる。その後、語句意味理解フレームが展開できなくなるまで展開され、この入力文についての処理が終了する。

4. 評価

4.1 文意味理解フレームの評価

アンケートにより収集した 150 文の単文と 100 文の重文や複文をテストセットとして用いた。それぞれ大学生の被験者 5 名による目視で文意味理解フレームに正確に格納されているかを判断し、正確に格納されているものを○、格納されていないものを×とし、多数決を取ることで評価を行った。文意味理解フレームの精度は単文のテストセットでは 85.3%、重文や複文のテストセットでは 54%となった。

4.2 語句意味理解フレームの評価

文意味理解フレームの評価に用いたテストセットから 50 文をランダムに選択し、それらに対して出力された全ての語句意味理解フレームを評価した。出力された語句意味理解フレームは 185 個となった。大学生の被験者 5 名による目視で語句意味理解フレームに正確に格納されているかを判断し、正確に格納されているものを○、格納されていないものを×として、多数決を取ることで評価を行った。語句意味理解フレームの精度は 83.2%となった。

5. 考察

文意味理解フレームの評価では単文であれば、比較的高い精度が得られたが、複文や重文になると精度が低くなるという結果が得られた。係り受け解析を用いることによってフレームに格納する語句がどのような構成となっていて

も一つのまとまりとして取得することができた。失敗例として「木工家具や竹工芸などに多くのデザインを残す」に対する文意味理解フレームを表 2 に示す。

表 2 文意味理解フレームの失敗例

Wh o	Wh o+	What	Wh en	Whe re	Ho w	Wh y	Wh om	述語
		多くの デザイン						残す

失敗例は「木工家具や竹工芸など」が格納されていない。これは助詞「など」に対応できていないためである。フレーム格納処理で格納するフレームを決定する際に語句に伴っている助詞を用いるが、本手法では全ての助詞に対応できているわけではないため、対応していない助詞が入力文に含まれていると評価が×となった。そのためフレーム格納処理において日本語の全ての助詞に対応させる必要があるが、全ての助詞とシソーラスの上位ノードの組み合わせを記述するのは現実的ではないため、機械学習によるフレーム格納処理の提案が必要になると考えられる。また、複雑な文になると 7W1H のどのフレームも適切でない語句が現れることが多かったため、分類するフレームを増やす必要があると考えられる。

語句意味理解フレームの入力となる語句はほとんどの場合は名詞句となり、主要部を修飾する文は複雑な文であることは少なく、簡単な文であることが多かったため、語句意味理解フレームの精度は単文のテストセットに対しての文意味理解フレームの精度と同程度になったと考えられる。しかし、主要部を修飾する文は 1 つではないこともある。例えば、「人に影響を与えられる魅力的な仕事」という名詞句は「仕事」が主要部になるが、これを修飾しているのは「人に影響を与えられる」と「魅力的な」の 2 つとなる。このような場合、1 つの語句意味理解フレームでは対応することができないため、複数の語句意味理解フレームを用いた格納方法の提案が必要となる。

6. おわりに

本稿では入れ子構造による文章の意味理解手法の提案を行った。評価を行った結果、文意味理解フレームの精度は単文のテストセットに対して 85.3%、重文や複文のテストセットに対しては 54%となった。語句意味理解フレームの精度は 83.2%となった。入力文を整理し、意味理解を行えることが確認できた。

謝辞

本研究の一部は、JSPS 科研費 16K00311 の助成を受けたものです。

参考文献

- [1] ChaSen -- 形態素解析器, 奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座(松本研究室), <http://chasen-legacy.sourceforge.jp/>, 2016/6/20.
- [2] Mecab, MeCab: Yet Another Part-of-Speech and Morphological Analyzer, <http://taku910.github.io/mecab/>, 2016/6/20.
- [3] CaboCha, CaboCha/ 南瓜 :Yet Another Japanese Dependency Structure Analyzer, <http://taku910.github.io/cabocha/>, 2016/1/19.
- [4] NIT コミュニケーション科学研究所, “日本語語彙系”, 岩波書店, 1997.
- [5] 竹内孔一, 動詞項構造シソーラスの構築, 3H2-OS3-5, OS-03 意味と理解のコンピューティング, 人工知能学会全国大会 2011.