2T-08

述語の連想情報を用いた日本語ゼロ照応解析

伊藤 清晃　　寺岡 丈博
拓殖大学工学部情報工学科

1 背景と目的

日本語では項の省略，代名詞や指示詞での言い換えが頻繁に見られる．この現象を照応という．また，省略された項をはじめ代名詞や指示詞を照応詞と呼び，照応解析はそれらの指示対象である先行詞を同定する処理である．ゼロ照応とは，照応詞が省略されている場合の照応である．ゼロ照応解析は，このような照応詞が省略されているものに対してどこに省略部分があるか，何が省略されているかを明らかにする．とりわけ，人間にとって当たり前である事柄が省略される傾向にあるため，このような表現が含まれる文は理解が困難である．

松林らによれば，文内ゼロ照応に関する事例の中で世界知識や文脈を読み解く必要がある事例や，その他一般化されていない雑多な手がかりを用いる事例に関しては解析精度の向上の余地があり [1]，このことからも，ゼロ照応解析の精度は十分とはいえない．

そのため，世界知識や文脈を読み解く必要がある事例 [1] を対象に，動詞連想概念辞書から得られる連想情報を用いてゼロ照応解析の精度向上を目的とする．この動詞連想概念辞書は，連想実験のデータから構築された辞書で，実際に人間が言葉に対して連想した内容をまとめたものである [2]．この辞書を用いることで人間が普段，文章を読む際に言葉に対して連想する情報をコンピュータが利用することができるため，ゼロ照応の精度向上が期待できる．そこで本研究では，手始めに単文に対して照応解析を行い，出力された先行詞の候補と人間の解釈との比較をした．

2 提案手法

2.1 手法の概要

本手法では，文を対象に動詞連想概念辞書を用いた日本語ゼロ照応解析を行う．図 1 は提案手法の概要を表している．動詞連想概念辞書 [2] は刺激語約 800 語に対して連想語が約 308,000 語，異なり語が約 58,000 語となっている [3].

まず，CaboCha[4] を用いて，形態素解析・係り受け解析を行う．そして文中の動詞および名詞，動詞に係る項を抽出する．助詞に注目することで動詞連想概念辞書の連想課題と動詞に係る項を対応させる．照応詞を検出し，動詞連想概念辞書から対応する連想課題の連想語を抽出する．抽出した連想語に対して，Word2Vec の単語分散表現からベクトルの平均を求める．求めたベクトルから得られた単語とコサイン類似度が高い単語を抽出する．これにより，動詞連想概念辞書外の単語も先行詞の候補として考慮することができる．本研究で利用する類似度はベクトルから得られるコサイン類似度とする．さらに，文中の単語と類似度が高い単語，文中の動詞と類似度が高い単語を抽出する．抽出した単語の中から共通する単語を抽出する．抽出した単語のそれぞれの類似度を計算して，類似度が高いものを先行詞として出力する．

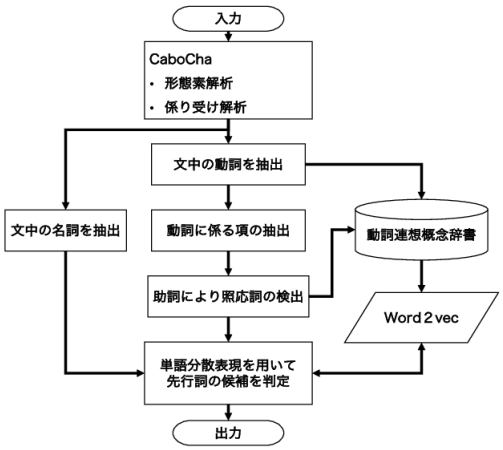

図 1: 提案手法の概要

2.2 照応詞の検出

日本語の文において理解に重要な要素は「誰が」,「何を」,「どうした」に絞られる．そのため，照応解析の対象にする照応詞は「動作主」と「対象」のみとする．照応詞が省略されている場合とみなすものは動詞連想概念辞書の連想課題の中の「動作主」と「対象」のみとする．

照応詞の検出に関しては代名詞と助詞に注目して行う．動詞に係る項の中で代名詞＋助詞となっている項を照応詞の候補とする．そして，助詞の種類によって連想概念辞書の連想課題に対応させ，照応詞として検

Japanase Zero Anaphora Resolution with Associative Information of Predicates
Kiyoaki ITO and Takehiro TERAOKA
Department of Computer Science, Faculty of Engineering, Takushoku University

表 1: 評価実験の結果

	動作主	対象	動作主＋対象
1 位正解率	0.64 (30/47)	0.24 (10/42)	0.45 (40/89)
3 位正解率	0.83 (39/47)	0.48 (20/42)	0.66 (59/89)
5 位正解率	0.91 (43/47)	0.69 (29/42)	0.81 (72/89)

出する.「動作主」と「対象」に対応する項が存在しない場合，照応詞が省略されているものとみなす.

2.3 先行詞の判定

先行詞の判定には，まず，照応詞の連想課題に対応する連想語を抽出する．そして Word2Vec により抽出した連想語の単語分散表現の平均となる単語を決定する．決定された単語と類似度が高い単語，動詞と類似度が高い単語，文中の名詞と類似度が高い単語をそれぞれ求める．3 つの単語の類似語の中で共通する単語を先行詞の候補として抽出する (図 1)．抽出されたものから Word2Vec の類似度が高い上位 5 語を先行詞として出力する.

3 評価実験

3.1 評価方法

優しい日本語コーパス [5] から代名詞を含む文 (例.「あなたはそれを持っていますか。」,「いつそれを買ったの。」) を 50 文を抽出する．そして，抽出した文を対象に照応解析を行い，出力された先行詞についてアンケートを行った．アンケートでは出力された先行詞が文に補完された際に理解できるものを選択してもらった．アンケートには 20 代の学生から 6 名に対して行い，3 名が選択したものを正解データとした．アンケートの結果から，N 位正解率 (Top N accuracy) で評価を行った.

3.2 結果と考察

表 1 を見ると「1 位正解率」では「動作主」の照応・省略箇所が 47 箇所に対して，30 箇所については先行詞として出力した単語が補完された場合に理解できる結果となり,「動作主」の絞り込みの精度は 64 %となった．また,「対象」については 42 箇所に対して 10 箇所が理解でき,「対象」の絞り込み精度は 24 %となった.「5 位正解率」では「動作主」の照応・省略箇所が 47 箇所に対して 43 箇所については先行詞として出力した単語が補完された場合に理解できた．よって,「動作主」の絞り込みの精度は 91 %だった．また,「対象」については 42 箇所に対して 29 箇所が理解でき,「対象」の絞り込み精度は 69 %だった.

このことから,「動作主」の先行詞については先行詞の候補を 5 つ以内に高い精度で絞り込めている．また,「対象」の先行詞については候補を絞り込めていない．なぜなら，単文に対して「対象」の先行詞の候補を絞り込むことは，人間でも難しいからである．これは単文だと先行詞の候補が拡散してしまうからと考えられる.

このことから，提案手法は「動作主」に対する先行詞の絞り込み精度が高く,「対象」に対する先行詞の絞り込み精度は低いと言える.

4 まとめ

本研究では，単文に対して動詞連想概念辞書を用いた照応解析を行い，出力された先行詞の候補を人間の解釈と比較した．提案手法は「動作主」に対して 0.91 と高い精度だった．しかし,「対象」は 0.69 と低い精度だった．結果として提案手法は,「動作主」のみ人間の解釈した先行詞と近いものを出力できた.

今後の課題としては，精度が低かった「対象」の先行詞の絞り込みの精度向上が挙げられる．これは文章を解析対象とし，先行詞の候補に使われる手がかりが増えることによって精度が上がると考えられる. 今後，解析する対象を単文から文章と増えることで，文全体の流れを考慮した先行詞を出力できるようになれば照応解析の更なる精度向上が見込める.

謝辞

本研究の一部は JSPS 科研費 18K12434 の助成を受けたものである.

参考文献

[1] 松林優一郎, 中山周, 乾健太郎. 日本語述語項構造解析タスクにおける項の省略を伴う事例の分析. 自然言語処理, Vol. 22, No. 5, pp. 433–463, 2015.

[2] 寺岡丈博, 東中竜一郎, 岡本潤, 石崎俊. 単語間連想関係を用いた換喩表現の自動検出. 人工知能学会論文誌, Vol. 28, No. 3, pp. 335–346, 2013.

[3] Takehiro Teraoka. Analysis of Associative Information for Second Language Learning of Japanese. In *Proceedings of the 4th Asia Pacific Corpus Linguistics Conference (APCLC)*, pp. 434–439, 2018.

[4] 工藤拓, 松本裕治. チャンキングの段階適用による日本語係り受け解析. Vol. 43, No. 6, pp. 1834–1842, 2002.

[5] Takumi Maruyama and Kazuhide Yamamoto. Simplified corpus with core vocabulary. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.