

Project Next NLP 構文解析タスク

河原 大輔

金山 博

颯々野 学

京都大学 日本アイ・ビー・エム株式会社 ヤフー株式会社

dk@i.kyoto-u.ac.jp, hkana@jp.ibm.com, msassano@yahoo-corp.jp

1 はじめに

構文解析は、英語、日本語などの言語で2000年頃には90%の精度が達成され、さまざまなアプリケーションにおいて利用されてきた。それ以降も、精度向上のための手法が提案されているものの、95%の精度にはまだ及ばないという現状である。従って、構文解析誤りの数%に何があるのか、より精度を上げるには何をすればよいのかを明らかにすることは重要である。また、これらの構文解析誤りによって後続のアプリケーションにどのような影響があるのか、それは誤りのタイプによって異なるのかを明らかにすることは、構文解析のアプリケーションにとって非常に意味のあることである。

本稿では、Project Next NLP 構文解析タスクにおいて行った日本語係り受け解析の誤り分類を概観する。次に、現状において大きな誤り原因の一つとなっている係り受けアノテーションの問題への対処方法について議論する。また、構文解析誤りの後続アプリケーションへの影響について考察する。

2 日本語係り受け解析の誤り分類

日本語構文・述語項構造解析システム KNP¹ による係り受け解析結果の誤りを分類した。KNP は、文法規則を用いて文節ごとに係り先候補を作り出し、述語項構造を単位として格フレームに基づく確率を計算する。その上で、もっとも確率の高い係り受け構造および述語項構造を出力するシステムである [15]。

係り受け解析誤りの分類結果は以下のとおりである。なお、ある文節の係り先の誤り原因として、複数の分類に属することがありえる。

1. 並列構造の誤り

並列構造は、日本語文中に頻出する表現であり、係り受けに直結しているため、並列構造の認識お

よびスコープ同定を誤ると、係り受け解析も大きく誤ることになる [5]。次の例では、「性格と、」と「真空調理を」が並列と解析され誤っているが、正解は「かみ合っの」に係り、並列構造ではない。

- (1) まさに、仕出しという仕事の**性格と**、真空調理を×使ってできる合理化、能率化がうまくかみ合っの○成功である。

なお、本稿中の例文の表記として、太字部を係り元、○下線部を正解係り先、×下線部を自動解析結果の誤り係り先とする。

2. アノテーションの問題

係り受けの正解アノテーションは、文節ごとに一つの係り先を決める必要がある。すなわち、意味的には複数の文節と関係があり、それぞれに係ることができても、正解としてはアノテーション基準に従って一つに決めている。そのため、自動解析の係り先が意味的には誤りではなくても、正解係り先とずれることがある。また、実際問題としてアノテーションの誤りも存在する。次の例では、「期間中は」が○下線部と×下線部のどちらとも意味的に関係があるが、正解と自動解析で係り先が異なっている。

- (2) この**期間中は**交通規制などで皆さんにご不便を おかけしますが、○ ご協力をお願いします。×

3. 節間の係り受け誤り

文中に節が3つ以上ある場合に、節間の係り受けの曖昧性が生じる。次の例では、「かさんで、」の係り先候補として○下線部と×下線部があるが、「かさんで、」の読点を加味して、より遠い方の「いない」に係ると解析し、誤っている。

- (3) ある女性は「交通費も**かさんで**、お金がないから○薬は いない×」と訴えた。

¹<http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP>

このような係り受けを改善するためには、節間の係り受け選好の知識が必要となる [3]。また、節間の係り受けは、前述のアノテーションの問題に深く関係している。

4. 文法規則のカバレッジ不足

KNP では、文法規則をベースにして、係り先候補を作り出しているため、例外的な係り受けに対して文法規則が不足し、正解係り先に係ると解析できない場合がある。たとえば、次の例の「飛ばして」の係り先は「5時間ほどの」であるが、このような「動詞連用形→名詞」の係り受けは文法規則として許されていない。

- (4) フリーウェイを **飛ばして** 5時間ほどの距離だ。×

学習ベースの解析器においても、このような例外的な係り受けがタグ付きコーパスにほとんど存在しなければ、学習することは難しいと考えられる。

5. 品詞誤り

判定詞と格助詞の曖昧性、複合辞と動詞の曖昧性 [14]、動詞連用形と名詞の曖昧性 [12]、副詞と動詞の曖昧性など、品詞の曖昧性が構文解析に悪影響を与える。つまり、品詞の曖昧性によって生じた品詞誤りが係り受け解析の誤りにつながる。次の例における「電灯で」の「で」の品詞は判定詞であるが、格助詞と誤って判定されており、そうすると項を受けることができないため、「今は」の係り先となることができない。

- (5) **今は** 電灯で。保久良神社の海拔185メートルの社頭に ある。×

現状の多くの解析システムでは、形態素解析と係り受け解析が独立して行われているが、これらを統合して、品詞の曖昧性についても構文解析で解消することが考えられる。

6. 語彙的選好知識のカバレッジ不足

語彙的な選好知識がないと、正しい係り先を解析することが難しい場合がある。次の例では「旅行で」の係り先として、「参詣する」が他の係り先候補よりも語彙的に優勢であると考えられる。

- (6) **旅行などで**、日頃馴染みのない寺院や神社に 参詣する。際にも、幅広く寺社札を受けて来た。× ということなのでしょう。

文中の【…】で示されることばが強く関係をもつ部分を選択肢から選んでください。

例1) 【クロールで】 [1 泳いでいる] 男の子を [2 見た。]
→ 【クロールで】は [1 泳いでいる] と強く関係をもっていますので、1 を選択します。

例2) 【双眼鏡で】 [1 泳いでいる] 男の子を [2 見た。]
→ 【双眼鏡で】は [2 見た。] と強く関係をもっていますので、2 を選択します。

図 1: クラウドソーシングにおける係り先タグ付けのインストラクション

この問題は、格フレームなどの語彙的選好知識を大規模コーパスから学習することによって、ある程度解決できる [10]。

7. その他

上記以外に、文脈依存、原文がおかしいなどの原因による誤りがある。

Web 文書の係り受け解析結果の誤りにおいては、「並列構造の誤り」「アノテーションの問題」「節間の係り受け誤り」の3つで約50%を占めていた。

3 アノテーション問題への対処方法の検討

従来の係り受けアノテーションは、専門家がそれぞれの文節に対して一つの係り先をタグ付けしている。この手法だと、例(2)のように、意味的にはどちらにもとれるような係り受けに対して、係り先をどちらかに決める必要があり問題となる。

本タスクでは、クラウドソーシングによる係り受けのアノテーションを試行した。一つのタスクは、一つの文節の係り先を判定するタスクとした。ワーカーへのインストラクションを図1に示す。係り先の候補としては、KNP が出力する係り先候補とタグ付きコーパスの正解係り先とした。タグ付きコーパスとしては、京都大学 Web 文書リードコーパス²を用いた。このコーパスは京大コーパスと同じ基準で係り受けがタグ付けされている。

Yahoo!クラウドソーシング³を用いて、一つの文節の係り先につき10人のワーカーに判定を依頼した。

²<http://nlp.ist.i.kyoto-u.ac.jp/index.php?KWDL>

³<http://crowdsourcing.yahoo.co.jp/>

その判定結果に基づき、Whitehill らの手法 [11] を利用して、係り先の確率を推定した。最大確率の係り先が、コーパスに付与されている正解係り先と一致していない事例の多くは、次に示すように、アノテーション基準が定められている場合であった。

- 定型表現の係り先

時間表現における個々の時間要素の係り先など、定型表現の係り先。次の例における「2010年」の正解係り先は「7月」である。

(7) 2010年 7月_{0.00} 17日に_{0.00}
注文住宅ナビカウンター立川が
オープンいたしました。_{1.00}

なお、下線部の数値は、Whitehill らの手法によって推定した係り先の確率値である。

- 並列句・節の係り先

並列句や並列節の係り先。アノテーション基準としては、隣の並列要素に係るとするため、次の例の「仕事・」の正解係り先は「デート・」となる。

(8) 仕事・デート・_{0.00} 飲み会などで_{0.00}
遅くなくても_{1.00} 深夜バスならまだ
運行しています。_{0.00}

クラウドソーシングによる係り受けアノテーションにおいて、これらの事例に対応することは今後の課題である。

2 節の例 (2) のように、意味的に複数の係り先が考えられる場合については、次のような例があった。

(9) 今回のフォーラムでは、こうした課題について日米両国の比較分析を 交えながら_{0.42} 議論を行います。_{0.58}

(10) 大病院で受診を希望するときは、まず主治医に 話し_{0.77} 紹介状を書いてもらうようにしましょう。_{0.23}

コーパスに付与されている正解係り先は、それぞれ「行います。」「話し、」であり、確率最大の係り先と一致している。また、正解係り先以外の「交えながら、」「書いてもらうようにしましょう。」は意味的には係り先になりうると考えられ、これらにも低い確率がきちんとついていることがわかる。今後、このような確率的アノテーションを利用した係り受け解析の評価方法について検討していきたいと考えている。

4 アプリケーションへの影響

2 節～3 節で見た通り、さまざまな種類の構文解析の誤り（コーパスとの不一致）があり、それらを一元的に扱った構文解析の性能評価が問題となることがある。ここでは、構文解析の良さや問題点を、構文構造を使う側の観点から検討してみる。

同様の試みとして、英語の構文解析器を、タンパク質間相互作用の関係抽出のタスクの観点から比較した Miayo らによる研究 [7] があり、異なる手法・出力形式の構文解析の性能やそのチューニングの効果を応用面から量的・客観的に比較することに成功している。

本稿では、日本語の構文解析における誤りがアプリケーションに与える影響を、質的な観点で考えてみたい。ケーススタディとして、構文規則と統計を併用した構文解析器 [2] を用いた評価表現抽出 [13] を取り上げる。

まず、構文解析の誤りに起因して正しい評価表現や評価の対象が得られなかった場合として、以下のようなものがあった。

(11) 最初のパンは悪くなかったけど○ メインの肉の塩味がきつかった。×

(12) 浴室が汚れていて愛想が悪かったことも あつて○ もう行きたくない。×

(11) の場合、「パンは」の係り先が誤っているため、「悪い」という評価の対象物が「パン」であることと同定に失敗している。また、(12) では、「行きたくない」の原因を示唆する「…こともあつて」という表現の部分が取得できないことにより、「浴室が汚れている」「愛想が悪い」という2つの不評の表現が得られない⁴ ことになり、構文解析誤りのなかでも特に影響が大きい。

一方で、構文解析の性能評価の中で誤りとされる現象のなかには、2 節で挙げられたものの一部を含め、評価表現の抽出に直接は影響しないものも多い。すなわち、構文解析の正解コーパスを用いた評価結果と、応用から見た構文解析器の性能の間に、ギャップが生じうる。経験的には、日本語の場合は (12) のように節間の関係の解析に失敗した場合、英語は節レベルの等位接続の多さから、並列構造の解析に失敗した場合に大きな影響がある。

今回は、構文解析器の後段のコンポーネントが特定の構文解析器の出力に依存した規則を用いているた

⁴適合率を重視し、構文木の頂点から順に構文パターンを適用していく方法を採用しているため。

め、この方法により構文解析器の複数の実装を直接比較することはできないが、同じ実装の構文解析器を再学習させた場合の優劣などを比較する場合には有効である。このように、本質的な問題点を見出す誤り分析の手法を模索していきたい。

5 関連研究

構文解析の誤り分析について、4節で挙げたもの [7] のほかに、英語に関してはいくつかの先行研究がある。Kummerfeldらは、複数の英語句構造解析器に対する誤り分析を行っている [4]。この研究では、誤りの構文木を正解の構文木に変換する部分木の變形操作によって、誤りを分類している。Haraらは、HPSG 解析器に対して、複合した誤り原因を切り分ける手法を提案している [1]。RimellらおよびNivreらは、gapがある難しい文の解析結果の評価、分析し、現状の構文解析器そのままでは解析が難しいことを報告している [9, 8]。McDonaldらは、2種類の係り受け解析器に対して、品詞、係り受け距離などごとに精度を比較している [6]。

6 おわりに

本稿では、Project Next NLP 構文解析タスクにおいて行った日本語係り受け解析の誤り分類を概観し、現状において大きな誤り原因の一つとなっている係り受けアノテーションの問題への対処方法について議論した。また、構文解析誤りの後続アプリケーションへの影響について分析し、解析誤りの種類によってアプリケーションの精度への影響が異なることについて考察した。

参考文献

- [1] Tadayoshi Hara, Yusuke Miyao, and Jun'ichi Tsujii. Descriptive and empirical approaches to capturing underlying dependencies among parsing errors. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 1162–1171, 2009.
- [2] Hiroshi Kanayama, Kentaro Torisawa, Yutaka Mitsuishi, and Jun'ichi Tsujii. A hybrid Japanese parser with hand-crafted grammar and statistics. In *Proceedings of the 18th International Conference on Computational Linguistics*, pp. 411–417, 2000.
- [3] Daisuke Kawahara and Sadao Kurohashi. Capturing consistency between intra-clause and inter-clause relations in knowledge-rich dependency and case structure analysis. In *Proceedings of the 11th International Conference on Parsing Technologies (IWPT'09)*, pp. 108–116, 2009.
- [4] Jonathan K. Kummerfeld, David Hall, James R. Curran, and Dan Klein. Parser showdown at the wall street corral: An empirical investigation of error types in parser output. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 1048–1059, 2012.
- [5] Sadao Kurohashi and Makoto Nagao. A syntactic analysis method of long Japanese sentences based on the detection of conjunctive structures. *Computational Linguistics*, Vol. 20, No. 4, pp. 507–534, 1994.
- [6] Ryan McDonald and Joakim Nivre. Characterizing the errors of data-driven dependency parsing models. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 122–131, 2007.
- [7] Yusuke Miyao, Rune Sætre, Kenji Sagae, Takuya Matsuzaki, and Jun'ichi Tsujii. Task-oriented evaluation of syntactic parsers and their representations. In *Proceedings of ACL-08: HLT*, pp. 46–54, 2008.
- [8] Joakim Nivre, Laura Rimell, Ryan McDonald, and Carlos Gómez Rodríguez. Evaluation of dependency parsers on unbounded dependencies. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pp. 833–841, 2010.
- [9] Laura Rimell, Stephen Clark, and Mark Steedman. Unbounded dependency recovery for parser evaluation. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 813–821, 2009.
- [10] Ryohei Sasano, Daisuke Kawahara, and Sadao Kurohashi. The effect of corpus size on case frame acquisition for discourse analysis. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 521–529, 2009.
- [11] Jacob Whitehill, Paul Ruvolo, Ting fan Wu, Jacob Bergsma, and Javier Movellan. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In *Advances in Neural Information Processing Systems 22*, pp. 2035–2043, 2009.
- [12] 麻岡正洋, 颯々野学. 係り受け解析のための動詞連用形用法の曖昧性解消. 言語処理学会第 11 回年次大会, pp. 1209–1212, 2005.
- [13] 金山博, 那須川哲哉, 渡辺日出雄. 木構造変換を利用した評判分析手法. 人工知能学会論文誌, Vol. 26, No. 1, pp. 273–283, 2011.
- [14] 注連隆夫, 土屋雅稔, 松吉俊, 宇津呂武仁, 佐藤理史. 日本語機能表現の自動検出と統計的係り受け解析への応用. 自然言語処理, Vol. 14, No. 5, pp. 167–197, 2007.
- [15] 笹野遼平, 河原大輔, 黒橋禎夫, 奥村学. 構文・述語項構造解析システム KNP の解析の流れと特徴. 言語処理学会第 19 回年次大会, pp. 110–113, 2013.