

はじめに

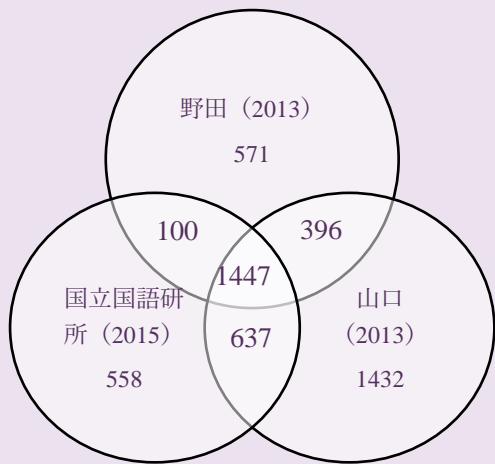
- 複合動詞は、UniDicの形態素解析を用いて網羅的に検索することが難しい語彙項目の一つ。
- 既存の複合動詞データベース3件(野田2013, 山口2013, 国立国語研究所2015)のデータを統合し、複合動詞のリストを作成し、UniDicの解析に基づき検索を試みる。
- 既存の複合動詞データベースには統語的複合動詞が含まれず、主として語彙的複合動詞のみ収録されている(山口2013, 国立国語研究所2015)。
- データベースに含まれない統語的複合動詞や、これらのデータベースにない「新しい」複合動詞を発見するための仕組みを考案する。

複合動詞検索システムの構築

- 目標は谷崎潤一郎(1886-1965)の作品コーパスからの複合動詞の網羅的検索。今回は、青空文庫から取得した9作品について、本文をUniDicで解析して検索に用いる。
- 複合動詞データベースを統合し、登録されている複合動詞をUniDicで解析したうえで、そのパターンを調査。解析パターンを利用してコーパスから用例を取得するプログラムを作成する。
- 統語的な複合動詞については、「動詞連用形+動詞」というパターンで検索をおこない、複合動詞データベースの項目を差し引いた項目を「可能性のある」複合動詞の候補とする。
- UniDicによる複合動詞の「誤解析」のパターンを整理して検索に利用することにより、通常の検索で取りこぼした複合動詞や未知の複合動詞の用例の検出を試みる。

谷崎潤一郎作品からの複合動詞の抽出実験

抽出結果 (1) 統合データベースを用いた用例検索



データベースの収録語彙の比較結果

- 解析された「複合動詞統合データベース」の複合動詞を用いて谷崎コーパスを検索した結果、1語彙素の複合動詞として2,852

例(異なり847種)、2語彙素からなる複合動詞として278例(85種)、3語彙素の解析結果(実際には解析に誤りがあるもの)から25例の用例が検索された。4語彙素から検索できた用例は1例のみであった。2語彙素で解析された前項動詞の解析結果を拡張し、データベースの解析(名詞+動詞)とは異なる品詞(動詞+動詞)で検索することで追加の用例が取得できた(以下は「落ち着き払う」の例)。



	書字形	語彙素	品詞	活用形	語種
統合DBの解析	落ち着き	落ち着き	名詞-普通名詞-一般		和
	払う	払う	動詞-一般	終止形-一般	和
コーパス解析	落ち着き	落ち着く	動詞-一般	連用形-一般	和
	払っ	払う	動詞-一般	連用形-促音便	和

抽出結果 (2) 未知の用例の検索

「動詞連用形+動詞」パターンによる検索では、異なり語数で164種類(延べ1,722語)の後項動詞が見つかった。そのうち統語的複合動詞候補として影山(1993)および姫野(2018)が挙げている30語種類中、22語が含まれていた(述べ786語)。

- 谷崎コーパスから見つかる統語的複合動詞候補：
～出す (234), ～かける (106), ～始める (78), ～切る (74), ～付ける (55), ～合う (41), ～過ぎる (35), ～得る (34), ～続ける (33), ～かかる (30), ～かねる (17), ～尽くす (9), ～慣れる (9), ～直す (8), ～通す (7), ～終わる (6), ～終わる (5), ～抜く (5), ～損なう (5), ～そびれる (5), ～果てる (5), ～誤る (1), ～忘れる (1)。
 - 統語的複合動詞を形成する以下の後項動詞は谷崎コーパスには見つからなかった：～まくる, ～損じる, ～遅れる, ～残す, ～あぐねる, ～飽きる, ～損ねる
- 語彙的複合動詞と考えられる後項動詞として、後項動詞「置く」の例(「(依頼)致し置く」)のような統合データベースに収録されていない動詞が複数見つかった。

1語彙素で解析される例：聞き分ける(5), 見分ける(4), 使い分ける(1), 掻き分ける(1), 引き分ける(1), 呼び分ける(1) *カッコ内は用例数
2語彙素で解析される例
名詞+動詞：咲[き](名詞)+分ける, 蹴り(名詞)+分ける, 弾き(名詞)+分ける, 釣り(名詞)+分ける, 立てる(動詞)+分ける
動詞+動詞：食べる(動詞)+分ける, 張る(動詞)+分ける, 鳴らす(動詞)+分ける, 鳴る(動詞)+分ける, 履く(動詞)+分ける

複合動詞検索システムの評価と今後の課題

- 『刺青』には「軋み合う、誇り合う、疼き出す」のような統語的複合動詞があるが、3つのデータベースのいずれにも入っていない。「動詞連用形+動詞」の検索で得られたのは「誇り合う」「疼き出す」のみ。「軋み合う」「堪え兼ねる」は解析が想定と異なっており、用例として収集できなかった。
- 「生き代わる、死に代わる、結(ゆ)い繞(めぐ)る」のような語彙的複合動詞も「複合動詞統合データベース」を使った検索からは得られなかった。「結い繞る」は統合データベースには収録されていない。「生き代わる」「死に代わる」が検索できない問題には、送りがなの問題があることがわかった(代るvs.代わる)。UniDicは階層的見出しを導入しており、書字形情報をもつことで表記のゆれに柔軟であるという特徴をもつ(伝ほか2007:108)が、複合動詞の検索においては、漢字や送りがなの異表記が検索において依然として課題であることがわかった。
- 複合動詞はUniDicにおいて単純動詞と区別されていないため、その検索には工夫が必要である。UniDicの解析に準じた検索により、データベースにある多くの複合動詞はその用例が検索できるものの、網羅的な検索には至らないことが示された。本研究でおこなった副次的候補による検索で収集できた例文もあったが、解析内容を完全に網羅した検索には至らなかった。
- 今後は本研究で実施に至らなかった、UniDicの辞書をカスタマイズする手法、解析の副次的な候補を充実させる方法に取り組み、より網羅的な例文の収集にむけて取り組んでゆきたい。
- UniDicの品詞体系に含まれない複合動詞をどのように解析情報に加えていくべきかという問題についても今後検討したい。

謝辞

本研究は、国際交流基金の日本研究フェローシップにより共同研究として実現したものである。記して感謝する。

参考文献

- 影山太郎 (1993) 『文法と語形成』 ひつじ書房。
伝 康晴, 小木曾 智信, 小椋 秀樹他 (2007) 「コーパス日本語学のための言語資源：形態素解析用電子化辞書の開発とその応用」 『日本語科学』 22, pp. 101-123.
野田時寛 (2013) 「日本語動詞用法事典について(4)―複合動詞一覧の試み―」 『人文研究』 75, pp. 31-62. 中央大学人文科学研究科。
姫野昌子 (2018) 『新版複合動詞の構造と意味用法』 研究社。
野村雅昭・石井正彦 (1987) 『複合動詞資料集』 国立国語研究所。
山口昌也 (2019) 「『Webデータに基づく複合動詞用例データベース』の構築と評価」 『国立国語研究所論集』 17, pp. 15-34.
由本陽子 (2013) 「動詞+動詞型の複合動詞」 『レキシコンフォーラムNo:6』 (編：影山太郎) pp. 59-78. ひつじ書房。

関連URL

- 国立国語研究所『複合動詞レキシコン』<https://vlexicon.ninjal.ac.jp/>
『Web茶まめ』<https://chamame.ninjal.ac.jp/>
『Webデータに基づく複合動詞用例データベース』<https://csd.ninjal.ac.jp/comp/>