

令和 4 年 6 月 7 日現在

機関番号：62618

研究種目：挑戦的研究(萌芽)

研究期間：2018～2021

課題番号：18K18519

研究課題名(和文)コーパスからの比喩表現収集とその分析

研究課題名(英文)Figurative Expression Extraction from Corpora and Analysis

研究代表者

浅原 正幸 (ASAHARA, MASAYUKI)

大学共同利用機関法人人間文化研究機構国立国語研究所・コーパス開発センター・教授

研究者番号：80379528

交付決定額(研究期間全体)：(直接経費) 4,800,000円

研究成果の概要(和文)：『現代日本語書き言葉均衡コーパス』を元にした指標比喩(直喩)データベース・語義に基づく比喩データベースを構築した。前者の指標比喩データベースについては、中村明著『比喩表現の理論と分類』に基づく指標比喩データベースとともに整理をしたうえで、2022年夏に公開する。語義に基づくデータベースについては、国際会議 RaAM 15 (15th Researching and Applying Metaphor Conference: 2022年9月 online)に採択された。今後、論文誌に投稿するとともに、語義に基づく比喩データベースの公開準備を進める。

研究成果の学術的意義や社会的意義

日本語において、電子化された大規模な比喩データベースは限定的であった。今回、文学作品を主とした中村明著『比喩表現の理論と分類』の電子化データに加えて、新たに『現代日本語書き言葉均衡コーパス』の書籍・雑誌・新聞データを対象に指標比喩・結合比喩の情報を大規模に付与した。これにより定量的に比喩表現の使用実態が評価できるようになった。さらに同データには一般の方が比喩と捉えるかどうかについての知覚実験結果を付与した。定量的に比喩表現の受容過程を評価できるようになった。

研究成果の概要(英文)：We constructed simile and metaphor expression databases based on the Balanced Corpus of Contemporary Written Japanese. We will release the simile expression database based on the Balanced Corpus of Contemporary Written Japanese and 'A stylistic study of the figurative' by Akira Nakamura in 2022.

The oral presentation of the metaphor expression database will be made at an international conference RaAM15 (15th Researching and Applying Metaphor Conference: September 2022, online). We will also write an article on the metaphor expression database, and release it.

研究分野：コーパス言語学

キーワード：比喩表現 直喩 隠喩

1. 研究開始当初の背景

日本語において比喩表現用例は、『比喩表現の理論と分類』(中村 1977)『比喩表現辞典』(中村 1995) など冊子体のものがほとんどで、用例採取対象も文学作品に限定されていた。

一方、他言語においては英語の **VU Amsterdam Metaphor Corpus (Steen 2010)** が **British National Corpus** の **academic texts, conversation, fiction, and news texts** を対象にするなど複数のレジスタを対象にしたデータが整備されていた。

2. 研究の目的

本研究の目的は、『現代日本語書き言葉均衡コーパス』を元にした比喩データベースの構築である。文学作品のみならず、他のジャンルの書籍・新聞・雑誌などを対象とし、現代日本語の比喩表現の実態を定量的に評価する。対象は指標比喩(直喩)・結合比喩(隠喩)とする。

3. 研究の方法

基礎的な資料として、国立国語研究所が著作権を所有する『比喩表現の理論と分類』(中村 1977)の指標比喩・結合比喩のデータベース化を行った。具体的には、同書籍を電子化したうえで、**Microsoft Excel** の帳票形式に変換した。同データに対して、新たに比喩の分類を付与したほか、被喩辞・喩辞などの情報も付与した。さらに、一般の方が比喩表現と認識できるか否かについての評価を実施した。

『現代日本語書き言葉均衡コーパス』(以降 **BCCWJ**) を元にした比喩データベースは

- 指標比喩データベース(中村 (1977)による A型把握(第1類)):
概念の移行や転換のほかに、「表現主体の比喩意識の反映と見られる何らかの言語形式をそなえて(pp.167-168)」いるグループ
- **MIP VU (Steen 2010)** に基づく比喩データベース:
の2種類の比喩表現を対象に **BCCWJ** のコアデータに基づき構築した。

指標比喩データベースの構築は、中村 (1977) の指標表現のリストを **BCCWJ** に適合させるほか、指標表現と同じ分類語彙表番号の語彙も指標表現候補として展開し、比喩表現の抽出を行った。

MIP VU に基づく比喩データベースは、**VU Amsterdam Metaphor Corpus** が採用した **MIP VU (Steen 2010)** を用い、分類語彙表番号の語義に基づく検討を行い、比喩表現の抽出を行った。

いずれのデータも比喩の分類を付与したほか、被喩辞・喩辞などの情報も付与した。さらに、一般の方が比喩表現と認識できるか否かについての評価を実施した。

4 . 研究成果

主な本研究の成果として、以下の論文・学会発表がある。

加藤祥, 菊地礼, 浅原正幸 (2020). 「『現代日本語書き言葉均衡コーパス』に基づく指標比喩データベース」『自然言語処理』 **27(4):853-887**.

加藤祥(2020). 「日本語比喩情報付与コーパスの作成と新聞における比喩実態調査の試み」『認知言語学の羽ばたき—実証性の高い言語研究を目指して—』 **144-159**. 東京：開拓社.

加藤祥, 菊地礼, 浅原正幸 (2022). 「『現代日本語書き言葉均衡コーパス』に対する MIP に基づく比喩表現情報の付与」『言語処理学会第 28 回年次大会発表論文集』: **1427-1432**.

結合比喩データベースについては、国際会議 **RaAM 15 (15th Researching and Applying Metaphor Conference: 2022 年 9 月 online)** に採択された。今後、論文誌に投稿するとともに、結合比喩データベースの公開準備を進める。

さらに、以下のようなデータベースを作成した。**2022** 年度以降、順次公開予定である。

『比喩表現の理論と分類』のデータベース版（指標比喩・結合比喩） (**2022** 年度中に公開)

『現代日本語書き言葉均衡コーパス』指標比喩データベース (**2022** 年度中に公開)

MIP VU に基づく『現代日本語書き言葉均衡コーパス』比喩データベース(仮称) (**2023** 年以降に公開)

本データベースに基づき、書籍・雑誌と比べて新聞記事に比喩表現が多用されていることが確認された。また新聞記事で多用されている比喩表現は、一般の方が比喩性を認めることの少ない用例が多かった。今後、同一性否定の観点や、英語の比喩データベースとの対照分析を進めたい。

【参考文献】

中村明(1977). 『比喩表現の理論と分類』国立国語研究所報告 **57**, 秀英出版.

中村明(1995). 『比喩表現辞典』角川書店.

Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., & Krennmayr, T. (2010). VU Amsterdam Metaphor Corpus. Design, Oxford Text Archives.

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 4件/うち国際共著 0件/うちオープンアクセス 4件）

1. 著者名 加藤 祥, 浅原 正幸	4. 巻 33
2. 論文標題 IPAL 用言例文への印象評定情報付与と代表義・典型用例の抽出	5. 発行年 2021年
3. 雑誌名 計量国語学	6. 最初と最後の頁 178-193
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Kato Sachi, Kikuchi Rei, Asahara Masayuki	4. 巻 27
2. 論文標題 Database of Figurative Expressions with Indicators from the 'Balanced Corpus of Contemporary Written Japanese'	5. 発行年 2020年
3. 雑誌名 Journal of Natural Language Processing	6. 最初と最後の頁 853 ~ 887
掲載論文のDOI (デジタルオブジェクト識別子) 10.5715/jnlp.27.853	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 加藤祥	4. 巻 なし
2. 論文標題 日本語比喩情報付与コーパスの作成と新聞における比喩実態調査の試み	5. 発行年 2020年
3. 雑誌名 認知言語学の羽ばたき 実証性の高い言語研究を目指して	6. 最初と最後の頁 144 ~ 159
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 加藤 祥, 浅原 正幸, 山崎 誠	4. 巻 15,2
2. 論文標題 分類語彙表番号を付与した『現代日本語書き言葉均衡コーパス』の書籍・新聞・雑誌データ	5. 発行年 2019年
3. 雑誌名 日本語の研究	6. 最初と最後の頁 134-144
掲載論文のDOI (デジタルオブジェクト識別子) 10.20666/nihongonokenkyu.15.2_134	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計33件（うち招待講演 0件 / うち国際学会 10件）

1. 発表者名 加藤 祥, 菊地 礼, 浅原 正幸
2. 発表標題 『現代日本語書き言葉均衡コーパス』に対するMIPに基づく比喩表現情報の付与
3. 学会等名 言語処理学会第28回年次大会
4. 発表年 2022年

1. 発表者名 加藤 祥, 浅原 正幸
2. 発表標題 『現代日本語書き言葉均衡コーパス』に対する印象評定情報付与
3. 学会等名 言語処理学会第28回年次大会
4. 発表年 2022年

1. 発表者名 加藤祥, 浅原正幸
2. 発表標題 多義語語義調査を目指した IPAL 形容詞例文への印象評定情報付与
3. 学会等名 言語処理学会第27回年次大会 (NLP2021)
4. 発表年 2021年

1. 発表者名 浅原正幸, 加藤祥
2. 発表標題 『日本語歴史コーパス』に対する文脈化単語埋め込みに基づく意味空間
3. 学会等名 人文科学とコンピュータシンポジウム「じんもんこん2020」
4. 発表年 2020年

1. 発表者名 加藤祥, 浅原正幸
2. 発表標題 受容主体の比喩性把握における指標と要素結合の影響
3. 学会等名 2020年度日本認知科学会第37回大会
4. 発表年 2020年

1. 発表者名 浅原正幸, 加藤祥
2. 発表標題 『日本語歴史コーパス』に対する文脈化単語埋め込み情報付与
3. 学会等名 日本語学会2020年度春季大会
4. 発表年 2020年

1. 発表者名 Rei Kikuchi, Sachi Kato and Masayuki Asahara
2. 発表標題 Collecting figurative expressions using indicators and semantic tagged Japanese corpus
3. 学会等名 the Fifteenth International Cognitive Linguistics Conference (ICLC-15) (国際学会)
4. 発表年 2019年

1. 発表者名 Sachi Kato and Masayuki Asahara
2. 発表標題 Exploring Metaphorical Expressions in Japanese newspaper-article corpora
3. 学会等名 the Fifteenth International Cognitive Linguistics Conference (ICLC-15) (国際学会)
4. 発表年 2019年

1. 発表者名 加藤 祥, 西内 沙恵, 浅原 正幸
2. 発表標題 多義語用例の類似度による語義の分類; 「遠い」と「近い」を例に
3. 学会等名 日本認知言語学会第20回全国大会
4. 発表年 2019年

1. 発表者名 加藤 祥, 浅原 正幸, 山崎 誠
2. 発表標題 『現代日本語書き言葉均衡コーパス』新聞・書籍・雑誌データの助動詞に対する用法情報付与
3. 学会等名 日本語学会2019年度春季大会
4. 発表年 2019年

1. 発表者名 加藤 祥
2. 発表標題 クラウドソーシングによる語義調査
3. 学会等名 日本言語学会第158回大会
4. 発表年 2019年

1. 発表者名 浅原 正幸, 加藤 祥
2. 発表標題 BERTed-BCCWJ: 多層文脈化単語埋め込み情報を付与した『現代日本語書き言葉均衡コーパス』データ
3. 学会等名 言語処理学会第26回年次大会
4. 発表年 2020年

1. 発表者名 浅原 正幸, 西内 沙恵, 加藤 祥
2. 発表標題 NWJC-BERT: 多義語に対するヒトと文脈化単語埋め込みの類似性判断の対照分析
3. 学会等名 言語処理学会第26回年次大会
4. 発表年 2020年

1. 発表者名 加藤祥, 田邊絢, 浅原正幸, 古宮嘉那子, 新納浩幸
2. 発表標題 多義語の語義分布と語義間の派生関係調査の試み--相の類を中心に
3. 学会等名 言語処理学会第25回年次大会
4. 発表年 2019年

1. 発表者名 菊地礼, 加藤祥, 浅原正幸
2. 発表標題 比喩指標とその類義語句を用いた直喩用例収集の試み
3. 学会等名 日本認知言語学会第19回大会
4. 発表年 2018年

1. 発表者名 Sachi Kato, Masayuki Asahara, Makoto Yamazaki
2. 発表標題 Annotation of `Word List by Semantic Principles' Labels for the Balanced Corpus of Contemporary Written Japanese
3. 学会等名 Proceedings of PACLIC 32 (国際学会)
4. 発表年 2018年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担 者	加藤 祥 (Kato Sachi) (40623004)	目白大学・外国語学部・専任講師 (32414)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------