

児童向け新聞教材のための言い換え表現対の抽出に関する研究

中村健二[†] 田中成典[‡] 北野光一[†] 田中裕一[†] 大林睦[‡]

関西大学大学院総合情報学研究所[†] 関西大学総合情報学部[‡]

1. はじめに

近年、小中学校では、総合学習の実施に伴い、新聞教材を用いた学習（NIE：Newspaper in Education）が実施[1]されている。教育現場では教師が一般向けの新聞記事を手作業で児童向けの表現に直している。しかし、適切な児童向けの表現を選択することは難しく、作成者にとって負担となる。そこで、文書作成の負担を軽減するため、言い換えの研究が行われている。既存研究では、言い換えのための基礎技術として、対応付けた同一内容の記事から言い換えられている同義語の対を言い換え可能な表現の対として抽出する研究[2]-[5]がある。児童向けの言い換え表現を対象にした場合も、一般向け新聞記事と同一内容の児童向け新聞記事が存在するため、同様にして児童向け表現の抽出が可能であると考えられる。しかし、児童向け新聞記事は、一般向け新聞記事と同一の内容でも本文中で使用される表現が異なるため、言い換え可能な表現の対を抽出することが難しい。そこで、本研究では、まず、シソーラスを用いて、同一内容の一般向け新聞記事と児童向け新聞記事を文書間の類似度を基に対応付ける。そして、対応付けられた文書の間で同一内容の文を対応付け、言い換えられている表現を抽出し、児童向けの言い換え表現対を抽出する。

2. 研究の概要

本研究では、一般向け新聞記事と児童向け新聞記事を対応付ける。そして、同一内容の一般向け新聞記事と児童向け新聞記事から対となる言い換え表現を抽出する。本システムは、一般向け新聞記事と児童向け新聞記事を入力データとし、対応付けられた新聞記事からによって抽出された言い換え表現対を出力データとする。本システムは、1) 記事収集機能、2) 対応付け機能、3) ルール抽出機能の3つの機能により構成される。図1にシステムの概要を示す。

Research for Extracting Paraphrase Acquisition of News Articles for Children.

[†] Kenji Nakamura, Koichi Kitano, Yuichi Tanaka
Graduate School of Informatics, Kansai University, 2-1-1
Ryouzenji-cho, Takatsuki-shi, Osaka 569-1095, Japan

[‡] Shigenori Tanaka, Mutsumi Obayashi
Faculty of Informatics, Kansai University, 2-1-1 Ryouzenji-cho,
Takatsuki-shi, Osaka 569-1095, Japan

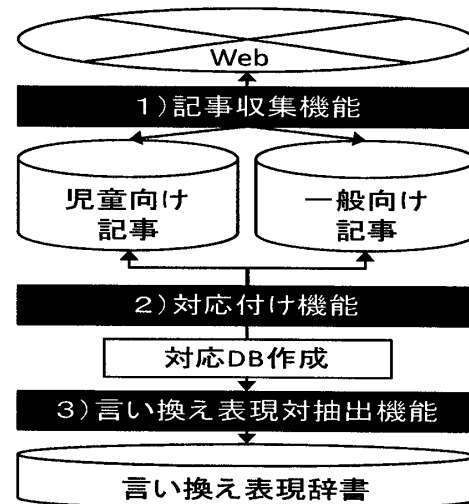


図1 システムの概要

2.1 記事収集機能

本機能では、Webから一般向け記事と児童向け記事を自動で収集する。収集する新聞記事は、毎日新聞がWebで配信している一般向け新聞記事と毎日小学生新聞が配信している児童向け新聞記事の2種類である。

2.2 対応付け機能

本機能では、シソーラス辞書を用いて、記事収集機能で収集した一般向け新聞記事と児童向け新聞記事の中で、同一内容の新聞記事を対応付ける。まず、一般向け記事と子供向け記事の類似度をベクトル空間モデルによって算出し、最も類似度が高い記事を同一内容の記事として対応付ける。本研究で用いるベクトル空間モデルの素性は、名詞、数値と文末表現とする。次に、対応づけた記事の組から固有名詞、一般名詞、サ変名詞と数詞の一致率が最も高い文を同一内容の文として対応付ける。

2.3 言い換え表現対抽出機能

本機能では、対応付け機能で対応づけられた文の組から同義語、広義語や語の位置を手掛かりとして、文中の語に対する言い換えられた表現の対を抽出する。そして、抽出した単語を基に言い換え表現辞書を作成する。

3. システムの実証実験と考察

本提案手法の有用性を実証するため、言い換え表現辞書を作成し、辞書へ登録した語彙の有用性を示す実験を行った。

3. 1. 実証実験

本実験では、2008年1月から2008年12月までに毎日新聞が毎日.jpで配信した一般向け新聞記事9191件と毎日小学生新聞がWebで配信した児童向け新聞記事742件を用いて言い換え表現辞書を構築した。構築した言い換え表現辞書の中で名詞の言い換え表現対の一部を表1に示す。また、構築した言い換え表現辞書を用いた結果を図2に示す。

3. 2. 結果と考察

実験結果を表1と図2に示す。表1より、名詞の言い換え表現では、「米」と「アメリカ」や「賃金」と「給料」のように有用な言い換え表現対が抽出できていることが確認できる。しかし、「不良」と「最悪」のように必ずしも適切な言い換え表現対とは言うことができない表現対も存在する。また、図2より、単純に本提案手法で構築した言い換え表現辞書を用いて言い換え可能な表現を言い換えただけでは、前後の文脈を考慮していないため、「パナソニックに一緒した」のような意味が通じない文を生成してしまう。そのため、言い換える表現の前後の文脈を考慮して言い換えを行う必要があると言える。

4. おわりに

本研究では、児童向け新聞記事の作成にかかる負担を低減するため、一般向け新聞記事を児童向け新聞へ言い換えられている同義語の対を抽出した。そして、実証実験の結果から、語彙を変換し、言い換え文を生成することに成功した。今後の課題として、本研究で構築した言い換え表現対を利用した変換では、食用の「米」を「アメリカ」と変換するなどの誤変換が発生していた。そのため、今後は、言い換え表現対の構築に加え、修飾表現や構文の変形を考慮した言い換え表現方法について検討することで、誤変換を無くすと共に、より児童に適した記事を提供する手法の考案を行う予定である。

参考文献

- [1] 榎本聡, 清水康敬: 教育情報のメタデータ化と検索システムの構築, 情報知識学会誌, 情報知識学会, Vol. 16, No. 4, pp. 63-72, 2006. 10.
- [2] 佐藤大, 岩越守孝, 増田英孝, 中川裕志: Webと携帯端末向けの新聞記事の対応コーパスからの言い換え抽出, 自然言語処理研究会研究報告, 情報処理学会, Vol. 2004, No. 1, pp. 193-200, 2004. 1.
- [3] 平尾努, 鈴木潤, 磯崎秀樹, 前田英作: 単一言語コー

表1 名詞の言い換え表現対の一部

	一般向け語彙	児童向け語彙
1	米	アメリカ
2	英国	イギリス
3	賃金	給料
4	応援	支援
5	若者	青年
6	活動	活躍
7	不良	最悪
8	男	男性
9	ジャパン	日本
10	統一	一緒

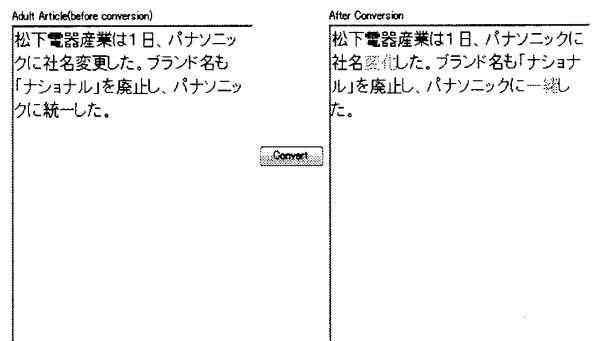


図2: 言い換え結果の例

パスにおける文の自動対応付け手法, 情報処理学会論文誌, 情報処理学会, Vol. 46, No. 10, pp. 2533-2545, 2005. 10.

- [4] Tanaka-Ishii, K., Yamamoto, M., and Nakagawa, H.: Kiwi: A Multilingual Usage Consultation Tool based on Internet Searching, Proceedings of the Interactive Posers/Documentations of the Association for Computational Linguistics 2003, Association for Computational Linguistics, pp. 105-108, 2003. 7.
- [5] Papineni, S., Roukos, S., Ward, T., and Zhu, W.: Blue: A method for Automatic Evaluation of Machine Translation, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, pp. 311-318, 2002. 7.