

若者語処理方法の未知語獲得システムへの適用

原田 俊信 亀田 弘之 劔持 直子

東京工科大学 工学部 情報工学科

1 はじめに

人間社会で重要な役割を担う計算機と、それを利用する人間との、話しことばを用いた柔軟でかつ高度なコミュニケーションの実現が期待されている[1, 2]。しかしながら、このような知的な HCI (Human-Computer Interface) を実現するためには、システムが必要に応じて若者語などの未知語を処理(理解・生成)できなければならない。

このような観点から、筆者らは、従来の自然言語処理技術では十分な処理が実現されていない若者語などの未知語について、その処理に関する研究に従事している[3, 4]。

本稿では、そのうち、若者語処理方法を未知語獲得システム(未知語獲得が可能な日本語統語解析システム)に適用し動作確認を行った結果について報告する。

2 若者語とは

2.1 若者語の定義

筆者らは、米川明彦氏による定義[5]に基づいて、工学的な観点から、若者語を、「既知語に、若者語変換規則を適用することで生成される語」と定義した。

2.2 若者語変換規則

若者語変換規則とは、既知語から若者語を生成するための「変形型」「連想型」「その他」の3種類からなる変換規則である。

A) 変形型

変形型は、既知語の表記から若者語の表記を生成する、6種類の変換規則の総称である。変形型によって生成された若者語は、表記のみが変化するだけで、意味はその元となった既知語と変わらない。変形型の下位分類は以下の通りである。

【省略型】 既知語の表記の要素を一箇所以上省略して若者語を生成する変換規則で、単省

略型と複合省略型の下位分類がある。

【倒置型】 既知語の表記の前部と後部を倒置して若者語を生成する変換規則である。

【転訛型】 既知語の表記の一部を、転訛規則を用いて若者語を生成する変換規則である。

【動詞の派生型】 既知語の表記(またはその一部)を語幹として、若者語(ラ行五段動詞)を生成する変換規則である。

【名詞の派生型】 既知語の表記(またはその一部)に接尾語を付けることで、若者語(名詞)を生成する変換規則である。

【形容詞の派生型】 既知語の表記(またはその一部)に接尾語を付けることで、若者語(形容詞)を生成する変換規則である。

B) 連想型

連想型は、ある既知語から、それとは異なる既知語を連想し、どちらか一方の表記に他の一方の意味を関係付けることで若者語を生成する、2種類の変換規則の総称である。連想型によって生成された若者語は、元の語の意味を失い、新たな意味を持つ。なお、連想型はその変換規則に応じて「意味連想型」「表記連想型」の下位分類がある。

C) その他

変形型、連想型以外の変換規則である。

3 若者語の処理方法

本研究では、若者語のうち、省略型によって生成された若者語(名詞)についての処理方法を未知語獲得システムに適用した。以下では、省略型によって生成された若者語を既知語に復元するための処理方法について説明する。

省略型による若者語の復元処理は、単省略型処理、複合省略型処理 A、そして B の順に、図 1 のような流れで行う。なお、各処理の詳細は以下の通りである。

【単省略型処理】 若者語の表記をそのままの形で含む既知語をシステム搭載の辞書から検索する。例えば、若者語“コンビニ”の場合、辞書を検索し、既知語“コンビニエンスストア”が検出される。

【複合省略型処理 A】 若者語の表記を二分割し、それらが含まれる既知語一語を検索する。例えば、若者語“フリマ”の場合、“フリ”

Application of younger's word processing method to the unknown word acquisition system

Toshinobu HARADA, Hiroyuki KAMEDA, Naoko KENMOCHI

Department of Information Technology, School of Engineering, Tokyo University of Technology

と“マ”に分割し，辞書を検索し，それらが含まれる既知語“フリーマーケット”が検出される。

【複合省略型処理 B】若者語の表記を二分割し，それらが含まれる既知語二語をそれぞれ検索する．例えば，“就活”という若者語の場合，“就”と“活”に分割し，辞書を検索し，それぞれが含まれる既知語“就職”“活動”が検出される。

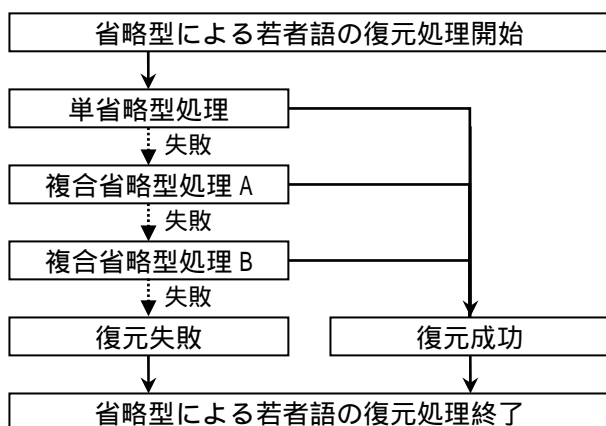


図 1. 省略型若者語処理の流れ

4 若者語処理方法のシステムへの適用

4.1 未知語獲得システムの概要

上記の若者語処理方法を，筆者らが作成中の未知語獲得システムに組み込んだ。

このシステムは，漢字かな交じり文を入力として受け付け，まず，未知語存在の検出と未知語候補文字列の抽出を行う。ついで，必要に応じて，語内構造の分析を行った後，未知語の品詞推定・意味推定を行う。若者語の処理は，語内構造分析の一部として，新たに組み込んだ。

本システムは，未知片仮名異表記語や未知複合語の処理等を行う機能がすでに実現されており，これに付加する形で若者語処理機能を実現した。その結果，処理効率のチューニングが現状ではまだ不十分ではあるが，動作内容の妥当性は確認されたので，その動作例を以下に示す。

4.2 システムの動作例

いくつかの例文を入力し，本システムを動作させた。以下に，若者語処理方法を適用する前のシステムと適用した後のシステムについて，それぞれの動作結果を示す。なお，＜適用前＞の下線部が未知語として処理されたものであり，＜適用後＞の波線部が若者語処理によって復元されたものである。

【例文 1】フリマで壺を買った

＜適用前＞

文(述部(未知語(フリマ)),格助詞(で),名詞(壺),格助詞(を),五段動詞連用形(買う),過去助動詞た終止形(た))。

＜適用後＞

文(述部(名詞(フリーマーケット),格助詞(で),名詞(壺),格助詞(を),五段動詞連用形(買う),過去助動詞た終止形(た))。

【例文 2】就活で忙しい

＜適用前＞

文(述部(未知語(就活),格助詞(で),形容詞終止形(忙しい))。

＜適用後＞

文(述部(名詞(就職),名詞(活動),格助詞(で),形容詞終止形(忙しい))。

【例文 3】私はコンビニの店長です

＜適用前＞

文(呈題部(名詞(私),係助詞(は)),論述部(未知語(コンビニ),格助詞(の),名詞(店長),断定助動詞です終止形(です))。

＜適用後＞

文(呈題部(名詞(私),係助詞(は)),論述部(名詞(コンビニエンスストア),格助詞(の),名詞(店長),断定助動詞です終止形(です))。

5 おわりに

以上，若者語処理方法を未知語獲得システムに適用した動作結果について報告した。

参考文献

- [1] 情報処理学会編：“情報処理ハンドブック，”オーム社（1997）。
- [2] 竹澤 寿幸：“いまこそ話しことば処理技術の研究を，”情報処理学会誌，Vol.42，No.2，pp.173-177（2001）。
- [3] Hiroyuki KAMEDA, Chiaki KUBOMURA：“Unknown Linguistic Knowledge Acquisition System，”Proc. of the Conference PACLING2001，pp.146-152（2001）。
- [4] 原田 俊信，亀田 弘之：“若者語の処理方法とその評価，”信学技報，TL2002-33，pp.1-6（2002）。
- [5] 米川 明彦：“若者語を科学する，”明治書院（1998）。