

論文 / 著書情報
Article / Book Information

論題(和文)	日本語学習者のためのWeb-Concordancerの開発
Title(English)	
著者(和文)	澤谷 孝志, 仁科喜久子, 赤堀侃司
Authors(English)	TAKASHI SAWATANI, KIKUKO NISHINA, KANJI AKAHORI
出典(和文)	日本教育工学会代 1 7 回大会講演論文集, Vol. , No. , pp. 469-470
Citation(English)	PROCEEDINGS OF THE 2001 CONFERENCE OF JET, Vol. , No. , pp. 469-470
発行日 / Pub. date	2001, 11
権利情報 / Copyright	本著作物の著作権は日本教育工学会に帰属します。 Copyright (c) 2001 Japan Society for Educational Technology.

日本語学習者のための Web-Concordancer の開発

Development of Web-Concordancer for Japanese language learners

澤谷孝志 仁科喜久子 赤堀侃司

Takashi SAWAYA, Kikuko NISHINA, Kanji AKAHORI

東京工業大学大学院 社会理工学研究科 人間行動システム専攻

Dept. of Human System Science, Graduate School of Decision Sciences and Technology,
Tokyo Institute of Technology

<あらまし> 英語を中心とする第二言語学習において、大量の言語データとそれを扱うツールによる探索的学習、Data-Driven Learning が提唱されている。入力されたキーワードを含む文をコーパスの中から大量に探し出し、前後の文脈とともに提示する Concordancer はその代表である。しかし、前後の文脈の提示が不十分であることや、学習者の学習レベルに合った文が提示できないなどの理由であまり普及はしていない。また、日本語教育においては Concordancer 自体がほとんどない。著者らは日本語教育において使用できる Web-Concordancer を開発した。本システムは使用者が指定した語彙レベルに合った文を提示することができ、また1文だけでなく前後数文を提示することができる。教師が適当と判断する文のみを集められるように、チェックした文のみを再提示する機能も持っている。

<キーワード> 日本語教育 CALL コーパス 検索 文脈 学習者レベル

1. 背景

言語学習者に大量の言語データとそれを扱うツールを与え、探索的学習を行うという Data-Driven Learning が Jones[1]などによって提唱されている。Data-Driven Learning の利点は帰納的、探索的学習を自立的に行える点である。

Data-Driven Learning を現実化したものとして、入力されたキーワードを含む文をコーパスの中から大量に探し出し、前後の文脈とともに提示する Concordancer が学習者支援のツールとして使われるようになった。Concordancer はもともと語学学習以外の用途で使われていた既存のものもあるが、第二言語学習用に開発されたものもあり、そのような Concordancer やそれを含む CALL システムは Corpus-based CALL と呼ばれることもある。近年では Web 上に置き、誰もが使えるようにした Concordancer もいくつかあり、Chen[2] に詳しく紹介されている。

しかし、その可能性が期待されるほどには実際の第二言語教育の現場には普及していない。Flowerdew[3]は問題点として(1)学習者のレベルを超えた文を提示していることが多い、(2)一行のみを表示する Single-line Concordancer では文脈情報が少なすぎる、(3)

コーパスのサイズによって、提示される文が多すぎたり少なすぎたりする、という3点を指摘している。

2. 開発

これまで開発されてきた Concordancer は英語を中心とする欧米言語のものがほとんどであり、日本語教育用のものはなかった。著者らは日本語教育において使用できる Web-Concordancer を開発した。特徴は(1)学習者が指定した語彙レベルの文のみを提示することができる、(2)前後数文を同時に提示することができる、(3)使用者がチェックした文のみを再提示できる、の3点である。

2-1. システムの概要

入力画面に調べたい語と語彙レベルを入力し検索すると、語は形態素解析ツール「茶筌」[4]に送られ、基本形と品詞が特定される。それを元にデータベースの中のその語を含む文が検索される。語彙レベルチェックを通った文のリストが画面に表示される。(図1)

2-2. 語彙レベルの指定

語彙レベルは日本語能力試験の語彙レベル表を規準として用いた。本システムを用いて学習することができるのは中級以上の学習者であろうと考え、レベルの指定は2級(中級)、1級(上級)、制限なし(超上級)の3段階とした。

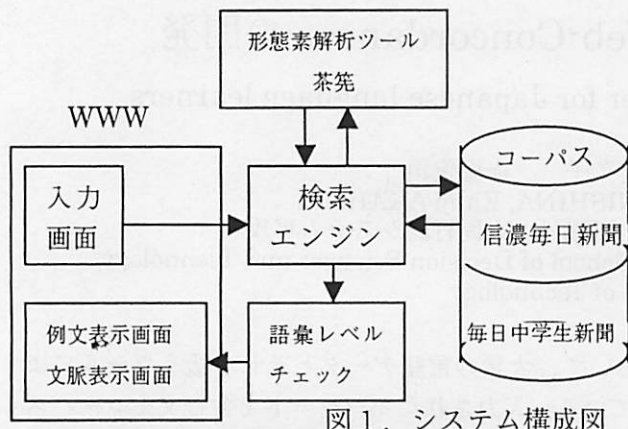


図1. システム構成図

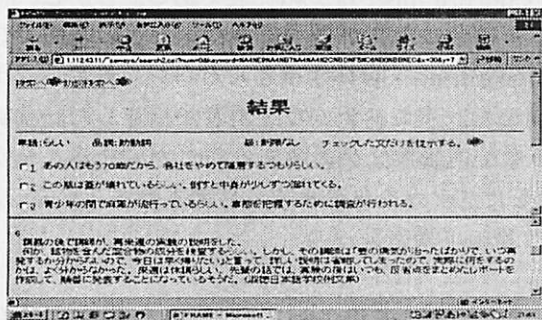


図2. 表示画面

コーパス内の文全てには含まれる語の最高レベルによって上の3段階のレベルがタグ付けされており、使用者が指定したレベル以上の語が含まれている文は提示の対象外となる。但し、固有名詞や「茶筌」が未知語と判断した語についてはレベル判断の対象外とした。

2-3. 文脈の提示

検索結果の表示画面は上下に分かれており、上に検索語が含まれる文のリストが提示され、その中からクリックされたものが、下の画面に前後3文を含めて提示される。(図2)

2-4. チェックした文のみの再提示

Concordancerには学習者が全ての作業を自分で行う使い方だけでなく、教師が与えたい文のみを抜き出して与える使い方もある。このような使い方を容易にするために、チェックした文のみを再提示する機能を付加した。

2-5. コーパス

信濃毎日新聞と毎日中学生新聞の社説、コラム、解説記事からコーパスを作成した。事件記事は体言止めが多く、言語学習に広く使えるものではないので、除いた。また中学生新聞は内容的には大人が読むに耐えるものだが、語彙は

少しやさしいという点で言語学習に適していると判断した。現在(8月末)の時点で、5万文が入れているが、今後もコーパスの量は大きくしていく予定である。

2-6. タグ付け

調べたい語を形態素解析ツール「茶筌」によって基本形と品詞のデータに変換するのに合わせて、コーパスも各語ごとに基本形と品詞がタグ付けされている。これにより、副助詞の「だけ」を入力して「だけど」が検索されるというようなことはない。

3. 今後の予定と課題

本システムの評価実験を年内に行う。被験者は日本語学習者だけでなく、日本語教師による評価も入れる予定である。

今後の課題としては、現在は語彙のみによるレベルの判定を行っているが、文型や表現形式も学習者レベルに関係するので、それによる判定も加える必要があると考える。それ以外にも文の長さなど別の要素も学習者レベルに関するものがあるかもしれない。その特定も課題のひとつである。

また、学習スタイルによって Concordancer の効果が変わってくる可能性がある。Concordancer の効果と学習者の学習スタイルとの関係を明らかにすることも課題のひとつである。

参考文献

[1] Jones, T.(1986), Micro-concord: a language learner's research tool, *System* 14(2) pp151-162
 [2] Howard Hao-Jan Chen (2000). Developing a Web Concordancer for English as Foreign Language Learners. *Proceedings of the Eighth International Conference of Computers in Education (ICCE2000)* pp.340-347.
 [3] Flowerdew, J.(1996), Concordancing in language learning, *The Power of CALL*, Athelstan, pp.97-113
 [4] <http://chasen.aist-nara.ac.jp/>