

論文 / 著書情報
Article / Book Information

論題(和文)	日本語学習のための Web Concordancer の開発と評価
Title(English)	Development and Evaluation of Web-Concordancer for Japanese Language Learners
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)	第3回「日本語教育とコンピュータ」国際会議, Vol. , No. , pp. 75-78
Citation(English)	Castel/J 2002 Proceedings-The Third International Conference On Computer Assisted Systems For Teaching & Learning/Japanese:Computer Technology and Japanese Language Education, Vol. , No. , pp. 75-78
発行日 / Pub. date	2002, 7

日本語学習者のための Web-Concordancer の開発と評価

Development and Evaluation of Web-Concordancer for Japanese Language Learners

澤谷孝志 仁科喜久子 赤堀侃司 (東京工業大学)

Takashi Sawaya, Kikuko Nishina, Kanji Akahori (Tokyo Institute of Technology)

Abstract : We developed and evaluated Web-Concordancer which can extract only usage examples including the same keyword from annotated corpus and show these to the learner.

The Web-Concordancer allows the user to restrict the vocabulary level of the usage example to be shown. It also allows the selection of usage examples from the given list for the preparation of teaching materials.

The evaluation experiment of composing sentences with Web-Concordancer and some dictionaries was conducted. The results of this experiment showed that Web-Concordancer is able to support learners in the situations whereby the learner would like to know the detailed definition, meanings or usage of his/her known word.

キーワード : Data Driven Learning corpus 生の例文 文脈 語彙レベル

1. はじめに

1-1. 背景

言語学習者は言語データを必要としているが、教室で与えられる言語データは少なく、ほとんどは教科書の中の例文や教師が作った作例である。一方、大量の文字データから特定の部分を検索するのは、人間よりもコンピュータのほうがはるかに早く行える作業である。

このことを踏まえ、本研究は、日本語学習者の学習支援を行うために、大量の文字言語データから調べたい語を含む文だけを検索するツール、Concordancer を開発し、評価することを目的とした。

1-2. Data-Driven Learning

Concordancer を言語学習に使う理論的背景のひとつが Data-Driven Learning(DDL)である。DDL は「学習者に大量の言語データとそれを扱うツールを与え、学習者は言語の仕組みを自ら説明していく」[1]学習方法である。DDL における最も重要なツールが Concordancer である。DDL では「学習者も言語の研究者」[2]であると考え、「その学習は言語データにアクセスすることによって推し進められる必要がある」[2]としている。

DDL では「まずデータがあり、そこから学習者がどんな言語規則やパターンを発見するかは教師も事前にはわからない」としているので、授業のあり方も教師の役割も従来のものから大きく変わることが求められる。それに対して、学習する言語規則やパターンを事前に決めた形で Concordancer を使おうとするシステム開発や研究も行われている。この場合、授業や教師の

役割はそれほどの変化は求められない。本システムは DDL を想定したものであるが、他の理論による学習を妨げるものではない。

英語教育目的の Web-Concordancer はすでいくつかある [3] が、日本語教育目的の Concordancer はほとんどなく、開発や教育への応用の研究もほとんどない。

2. Concordancer の開発

2-1. 開発の方針

Flowerdew(1996)[4]は Concordancerの問題点として(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)一行のみを表示している場合、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。

これをふまえ、著者らは日本語 Concordancer の開発の方針を次のように定めた。

- ①できるだけ学習者の学習レベルに合った文を表示する
- ②1行だけや単文だけではなく、前後の文脈も同時に表示する
- ③提示する文の数を指定できるなど、使いやすいインターフェースのシステムにする
- ④教師の使い勝手をよくする

2-2. Corpus

Corpus は学習の目的別に様々なものを用意するのが望ましいが、まず初めに用意すべき、広い範囲で使える一般性の高い Corpus として、筆者らは新聞記事を選んだ。新聞記事は専門的な校閲を経ているという点で、誤字、脱字や非文法的な文は非常に少ないと考えられ、また、広範な読者層を想定した媒体

Corpus の種類	文の数	のべ語数
信濃毎日新聞社説	66,776	1,242,824
信濃毎日新聞コラム	22,754	352,226
毎日中学生新聞	2,602	53,039
淑徳日本語例文集	1,523	24,947

表1. 各 Corpus のデータ量

なので、語彙や表現も一般性のあるものが多いからである。信濃毎日新聞と毎日中学生新聞から記事の使用と Web 上での公開の許可が得られたので、それぞれの記事を使用した。各 Corpus の現在のデータ量を表1に示す。

2-3. タグ付け

通常、Corpus は各形態素に分解され、様々な情報がタグ付けしてある。これによって、副助詞の「しか」を検索しようとして「しかし」を含む文を検索してしまうようなことが避けられる。

しかし、特定の文法体系に従って文法的な情報を付けてしまうことで、DDL が目指す「学習者が何を発見するかわからない」というダイナミックな学習の可能性は減じてしまう。筆者らは相互の長短所を勘案した結果、品詞情報だけをタグ付けすることにした。

2-4. レベルチェック

日本語能力試験出題基準[5]の語彙表を基に、検索対象の例文を、2級以下の語彙しか含まない文、1級以下の語彙しか含まない文、制限なし、の3段階で指定できるようにした。3、4級まで分けなかったのは、Concordancer による学習は初級では難しいと考えたからである。

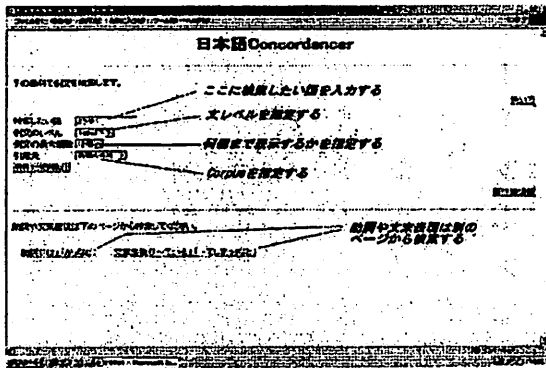


図1：検索画面

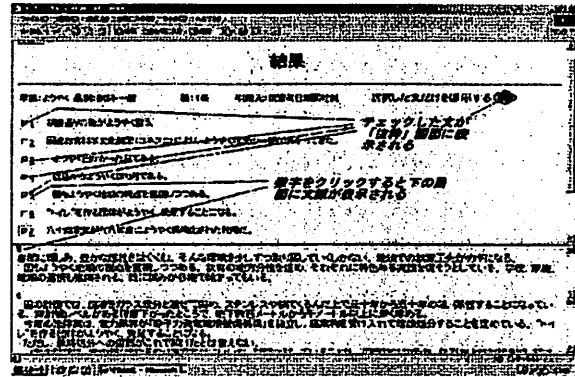


図2：結果表示画面

2-5. 画面

図1は検索画面である。ここに検索したい語を入力し、Corpus とレベルを指定して検索すると、図2のように結果が上の画面に一文ずつ、下の画面には前後3文とともに表示される。

また、教師が教材作りに利用しやすいように、結果表示画面の使いたい文だけを別の画面に表示する機能も加えた。

3. 日本語学習者を被験者とした評価実験

3-1. 目的

本システムはひとつの学習活動だけに限ったものではなく、様々な学習活動への支援が考えられるので、評価も様々な面から行わなければならないが、今回は自由作文への支援効果の評価を辞書と比較する形で行った。

3-2. 方法

被験者 日本語学習者20名

タスク 4つの4コマ漫画の内容を自由記述。その際に2タスクは辞書、残り2タスクはConcordancerを自由に使う。タスクの一つを図に示す。

評価 3人の日本語教師が作文を評価する。Concordancerの支援と辞書の支援の効果の差を測定するために、上の評価を標準化し、各被験者ごとにConcordancerを使ったタスクの成績と辞書を使ったタスクの成績の差を算出、比較する。

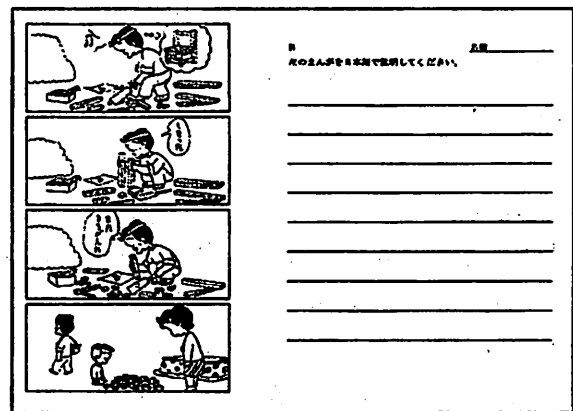


図3：タスクの例

3-3. 結果

図4は各被験者の<Concordancerを使ったタスクの成績>から<辞書を使ったタスクの成

績>を引いた数値を表している。全体の平均は-2.8だが、0よりも有意に低くはなかった。つまり、全体としては Concordancer と辞書の支援効果に有意な差は見られなかった。しかし、Concordancer と辞書の成績に大きな差のある被験者3名にインタビューを行ったところ、その行動に大きな違いがあることがわかった。

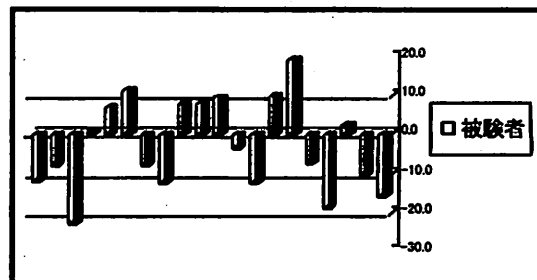


図4：各被験者の<Concordancer - 辞書>

Concordancer を使ったほうが成績がよかった被験者の場合、既知の日本語の単語の更に正確な使い方や意味を知ろうとするという顕著な傾向が見られた。

一方、辞書を使ったほうが成績がよかった被験者の場合、母語から考えて、その日本語訳を知ろうとする顕著な傾向が見られた。この方略は Concordancer ではあまり使えず、成績が悪かったと考えられる。

以上の結果から、作文に対する Concordancer の支援効果は、学習者の方略に影響を受けると考えられる。

4. 結論

以上、Web 上で使える日本語 Concordancer を開発し、評価した。本システムは汎用的なツールであるので、完全な評価は難しく未だ行われていないが、文章作成過程に限っては学習者の方略によっては有効な支援を与えられようと考えられる。

今後の課題としては、

① Corpus の拡充 ② より速い検索の実現 ③ インターフェースの改良 ④ 実際に Concordancer を使った授業の実践が挙げられる。

Reference

- [1] Wauschauer, M. & Healey, D., Computers and language learning: An overview, *Language Teaching*, Vol.31, pp.57-71, 1998
- [2] Johns, T., Should you be persuaded - Two samples of Data-Driven Learning Materials, *English Language Research Journal*, Vol.4, pp.1-16, 1991
- [3] Chen, H. J., Developing a Web Concordancer for English as Foreign Language Learners, *Proceeding of International Conference of Computer for Education 2000*, pp.340-347, 2000
- [4] Flowerdew, J., Concordancing in Language Learning, In Martha C., *Power of CALL*, pp.97-113, 1996
- [5] 日本語能力試験出題基準, 国際交流基金・日本語国際教育協会, 1994