

日独辞書ツールの開発とその評価 Development and Evaluation of a Japanese-Germany Dictionary Tool

東京国際大学/ウィーン大学
川村 よし子
kawamura@tiu.ac.jp

1. はじめに

筆者らの開発した読解学習支援システム「リーディング・チュウ太」(川村・北村 2001)には学習支援ツールとして辞書ツールが組み込まれている。これは入力されたテキストの形態素解析および辞書引き作業を自動的にを行い、個々の単語の読みや意味を表示するツールである。辞書ツールとしてはすでに日英辞書ツールと日日辞書ツールを開発してきたが、今回新たに意味情報をドイツ語で説明する日独辞書ツールの開発をすすめている(Kitamura & Kawamura 2001)。この日独辞書ツール 版の完成を受けて今回これを試用した評価実験を行った。本発表ではその実験結果を報告するとともに、改善方法について提案する。

2. 日独辞書ツールの概要

日独辞書ツールはインターネット上で公開されている読解学習支援システム「リーディング・チュウ太」(<http://language.tiu.ac.jp>)の学習支援ツールのひとつとして開発したものである。学習支援ツールはすべて「チュウ太の工具箱」にまとめられ、その入力画面に読みたい日本語の文章を入力し、ボタンによって支援ツールを選択する仕組みになっている。

図1がその入力画面である。この「辞書ツール」のテキストボックスに文章を入力し、「日 -> 独」のボタンを選択すると日独辞書ツールが辞書引き作業を自動的に行う。文章はテキストボックス内に直接打ち込むことも、インターネット上の情報やメール等の電子化された文章をコピー&ペーストで入力することも可能である。日独辞書ツールは、この入力された文章をまず形態素解析システム『茶釜』によって形態素(基本的には単語レベルの文要素)に区切る(松本ほか1999)。次にその分析の結果を日独辞書(Apel 2001)と照合し、個々の単語の読みおよび意味の辞書情報を表示する。

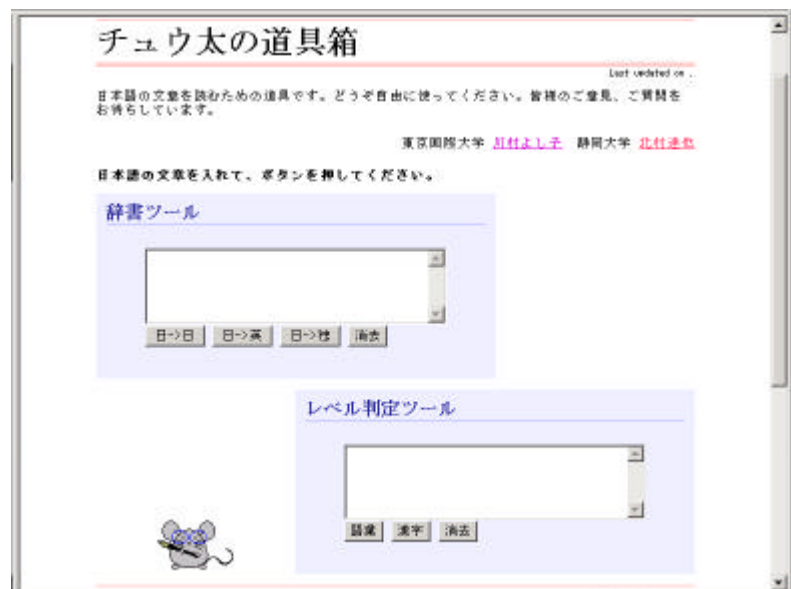


図1 日独辞書ツールの入力画面

図2が日独辞書ツールの出力画面である。画面左が本文で文中の単語は右の辞書情報とリンクされている。本文中の単語をクリックすると右の辞書が自動的にスクロールし当該語が一番上に表示される仕組みになっている。文中にでてきた同じ単語を何度調べることが可能である。さらに、一連の読解作業が終わった時点で本文下の「list」ボタンを押すと、学習履歴が単語リストの形で表示される。このリストも辞書情報とリンクされているため学習

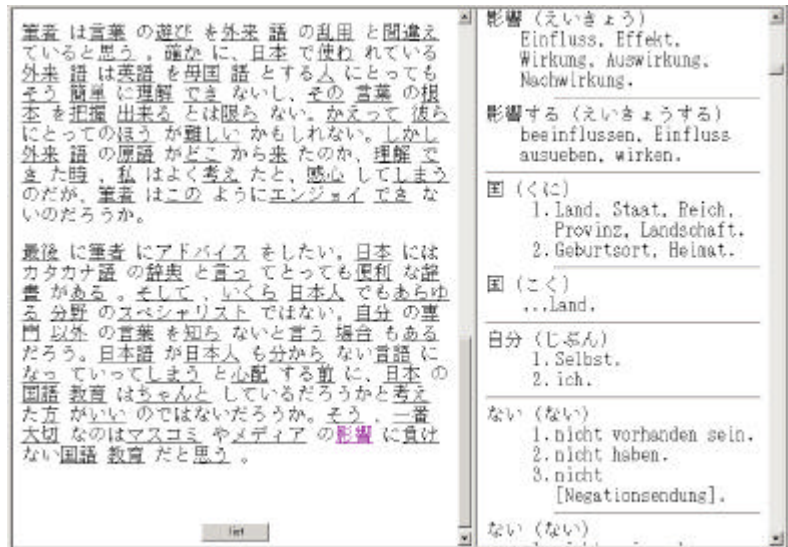


図2 日独辞書ツールの出力画面

単語の意味や読みの確認が可能である。これらの機能はすべて従来の辞書ツールに準拠したもののだが、この学習履歴機能を利用することで学習効果が上がることが検証されている。(北村・川村ほか1999)

3. 日独辞書ツール 版の評価

これまで開発してきた日英辞書ツール(川村・北村・保原 2001)および日独辞書ツール(川村・北村 2000)では、問題点の洗い出しおよび改良作業は基本的には開発者が単独で行ってきた。今回、日独辞書ツールの開発にあたり、版の辞書ツールがどの程度正確な辞書情報を提供できているのか調べるための分析評価を行った。評価方法は次の通りである。

3.1 分析対象

分析した文章は2001年度筆者が担当したウィーン大学の日本語上級クラス「応用日本語」の参加者9名のうちドイツ語が母語あるいはそれに近い状態にある学習者6名の作成した意見文である。対象となった学生の日本語学習歴はいずれも3年以上で上級レベルである。

3.2 分析方法

分析方法は次の通りである。

- i. 意見文を日独辞書ツールで自動処理して辞書情報を表示させる
- ii. 意見文に含まれる語の辞書情報が適切かどうか意見文の作成者本人がチェックする
- iii. 問題のある項目に関して、何が問題でどう改善すべきかを調査者と本人とで詳しく分析する

3.3 調査結果

今回調査対象となった文章は、全体で158文、総文字数7268文字である。また延べ語数で3872語、異なり語数で890語である。

調査の結果、意見文作成者本人によって辞書情報に何らかの修正が必要だとされた語が59語あった。その中には訳語の修正・追加・削除によって修正が可能な語が29語含まれていたが、それ以外の30語は辞書の作成方法あるいは辞書ツールの表示方法に関する修正が必要なものであった。

4. 辞書情報はどうかあるべきか

この調査結果をもとに対訳辞書ツールはどうかあるべきかという視点から現在の辞書ツールの抱えている問題点を分析した。主な問題点とその解決方法は以下の通りである。

4.1 表記

【問題点】辞書項目には含まれているはずの語句の辞書情報が表示されないことがある。

通常漢字仮名交じりで表記される語句がひらがなで表記されている場合、辞書ツールでは表示できないことがある。また送りがなの表記法が複数ある場合にも問題が生じる。今回の調査では「たえた」「理解しあう」等の語がこれにあたり、辞書には「讚える」「理解し合う」という形でのみ登録されていた。

【解決方法】辞書の見出し語の登録に際してはさまざまな可能性を配慮し、必要に応じて複数の見出し語を登録する必要がある。特に副詞、形容詞、形式名詞等はひらがな表記されることも多いため、辞書へのひらがなでの登録、あるいは辞書引きの際に読み情報を元にした検索が不可欠である。また送りがなの可能性が複数考えられる語についてはそれらをすべて登録する、補助的に用いられる動詞を含んだ語に関しては漢字表記とひらがな表記の両方で登録する等の配慮が必要である。(例えば「寄り掛かる」については「寄掛る」および「寄りかかる」を登録する必要がある。)

4.2 品詞情報

【問題点】

日独辞書ツールで利用している「和独辞書」の品詞情報が完全ではないため日独辞書ツール 版では品詞情報によるフィルターをかけていない。一方サ変動詞への対応として見出し語に「する」を付加したのも同時に表示している。そのため 版の辞書ツールでは例えば次のような問題が発生していた。接尾辞としての「人」に対して名詞としての「人」の意味や読みが表示される。接尾辞としての「さ」に「する」を付加した「さする」が表示される。

【解決方法】

日独辞書の品詞情報を茶釜の品詞分類にあわせて整備する必要がある。辞書ツールでは品詞によるフィルターをかけ、必要な項目以外の辞書情報を表示しないようにする必要がある。

【問題点】

動詞の意味情報に本動詞としての意味と補助動詞として用いられた場合の意味とがまとめて表示されているため、動詞の意味情報が多くなりわかりにくい。また当該の動詞が持つ意味範囲が把握しにくい。

【解決方法】

形態素解析による分析では、基本的には本動詞と補助動詞とを区別して表示することは可能である。学習者の便宜を考えた場合、本動詞の意味と補助動詞の意味とは別項目に分けて辞書に登録し、これを区別して表示することが望ましい。

4.3 意味の選定

【問題点】

意味情報が多すぎるので単語の意味が捉えにくいことがある。

【解決方法】

辞書としての機能から考えれば各々の語が含む意味を出来るだけ詳しく網羅するのが望ましい。ところがその一方で、記述が詳しくなると学習者にとっては辞書情報が多すぎるため、利用しにくくなるという問題が生じる。そこで、a. 意味情報は中心となる語義から順に並べる必要がある。b. 意味情報は単に羅列するのではなく、概念ブロックごとにまとめる必要がある。c. 派生的な意味まで詳しく扱った場合、例文提示の必要性を検討する必要がある。これらの改良を加えることによって辞書はかなりわかりやすくなると思われる。だが、やさしい語ほど派生的意味も多くなることを考えると、利用者のレベルを考慮した辞書ツール、つまり、初級中級学習者向けの簡易版辞書ツールと上級学習者向けの詳しい辞書情報が得られる辞書ツールとを別個に用意する必要があるのかもしれない。

4.4 複合語の扱い

【問題点】

複合語であっても茶釜の辞書に登録されている場合、茶釜は一語として切り出してしまう。そのため対訳辞書にもその項目が見出し語として存在しないと辞書情報を表示できないという問題が生じる。例えば「要するに」「何らかの」等の語は茶釜の辞書にあるので1語として切り出されるが、日独辞書にはこれが見出し語として入っていないので表示できない。また、逆に、茶釜に登録されていない場合は、辞書項目として見出し語を登録しても表示できない。例えば「的」のつく言葉のうち茶釜には「比較的」は登録されているが「積極的」「印象的」などの語は登録されていない。

【解決方法】

複合語はより多く登録したほうが辞書としての機能はあがる。複合語を多く表示したいのであれば茶釜にそれを登録すればきれいに切り出すことも可能になる。ところが茶釜の登録語が増えれば増えるほど新しい辞書（例えば別の言語への対訳辞書）の作成時にもそれら全ての語を網羅しないかぎり辞書情報を表示できないという問題が生じる。汎用性という視点から考えると茶釜の単語辞書は必要最低限に押さえることが望ましい。これらのことを踏まえたうえで複合語の登録基準を定め、これに準拠した登録を進めていくべきである。

4.5 慣用表現

【問題点】

慣用表現等で派生的に用いられた単語の意味が他の意味に並んで列挙されていると当該の単語の意味を把握しにくくなる。例えば「なる」の意味で「spielen (遊ぶ・演じる)」という訳語が出ている。これは「役目を演じる」という意味で用いられた「なる」の訳語であるが、訳語だけが単独で示されると学習者にはわかりにくい。

【解決方法】

慣用表現等で派生的に発生した特殊な意味を記述する場合、単にその意味を記載するだけでなく慣用表現であることを記述すると共に、その使われ方を表示する必要がある。また現在の辞書ツールは単純に単語レベルで切り出しを行っているが、読解学習支援システムとしては慣用表現自体も何らかの形で示せるのが望ましい。こうした慣用表現に対する対応は早急に考えていく必要がある問題である。

4.6 例文の提示

【問題点】

上記の慣用表現も含め、特に単語の意味が派生的に用いられている場合、意味が列挙されているだけでは分かりにくい。また意味情報が増えれば増えるほど、単語自体の意味範囲の把握も難しくなる。意味ごとに例文があればわかりやすい。

【解決方法】

4.5 とも関連し、何らかの形で例文を提示する機能が必要とされている。これは辞書情報に例文情報を追加することによって、対応自体は可能である。

4.7 単語の分割

【問題点】

言語によって単語の区切りが異なっている。例えばドイツ語では「積極的な」は「aktive」という1語の形容詞として捉えられている。ところが辞書ツールでは「典型・的・な」と3語に分析され、各々の意味をどう表示するかという問題がある。

【解決方法】

これに関しても茶釜の辞書と日独辞書の双方に当該の語句（例えばこの場合「積極的」）を登録することで解決する。今後、複合語の登録基準を考える際に、この問題も含めて検討していきたい。

5. 終わりに

今回の調査によって現在の日独辞書ツールの問題点がすべて網羅できたわけではないが、学習者の視点からみた日独辞書ツールの問題点を明らかにすることができた。今後、この調査結果をもとに日独辞書ツールの整備を進めていく予定である。またこの調査結果は日独辞書ツールばかりでなく他の言語の対訳辞書を作成・整備する場合にも利用可能なはずである。他の言語についても学習者にとって使いやすい対訳辞書が作成されることを期待したい。

参考文献：

Apel, U.(2001) Japanische-deutsche Woerterbuchdatei, <http://133.179.591/WadokuJT/default.htm>
Kitamura & Kawamura (2001) A Japanese Reading Comprehension Support System for German-Speaking Learners, EUROCALL 2001.

川村よし子・北村達也(2001)「インターネットを活用した読解教材バンクの構築」『世界の日本語教育』第6号、pp.241-255.

川村よし子・北村達也(2000)「概念辞書ツールを活用した語彙学習」『日本語教育方法研究会誌』7, No.1, pp.36-37.

川村よし子, 北村達也, 保原麗(2000)「EDR 電子化辞書を活用した日本語教育用辞書ツールの開発」『日本教育工学会誌』 Vol.24, Suppl.2000, pp.7-12.

北村達也・川村よし子・内山潤・寺朱美・奥村学(1999)「学習履歴管理機能を持つ日本語読解支援システムの開発とその評価」『日本教育工学会論文誌』23(3), pp.127-133.

松本裕治・北内啓・山下達雄・平野善隆・松田寛・浅原正幸(1999)「日本語形態素解析システム『茶釜』version 2.0 使用説明書第二版」NAIST-IS-TR99012