九州大学学術情報リポジトリ
Kyushu University Institutional Repository

韓国語学習者のためのWeb辞書の開発

曺, 美庚
九州大学大学院言語文化研究院言語環境学部門 : 准教授

李, 相穆
東京大学

https://hdl.handle.net/2324/10056

出版情報：e-Learning教育研究. 2, pp.63-73, 2008-03. e-Learning 教育学会
バージョン：
権利関係：

韓国語学習者のための Web 辞書の開発

曺 美庚　李 相穆

『e-Learning 教育研究』第 2 巻（2007）抜刷

論文

韓国語学習者のためのWeb辞書の開発

曺 美庚　　李 相穆
(九州大学)　(東京大学)

1.　はじめに

近年、公共交通機関の案内表示や飲食店の看板、TV コマーシャルなど、街のいたる所でハングル文字を見かけるようになった。韓国語に対する親密感の増加に伴い、韓国語の学習を希望する人の数も増えつつある。ここ数年間の韓国語学習者の増加傾向については、国際文化フォーラム（2005）に詳しく取り上げられている。国際文化フォーラム（2005）によれば、2003年現在、韓国語教育を実施している高校は、全国の高校5,435校中の219校（4.03%）で、履修者の数では6,476名となっている。これらの数値は、1999年の同131校（2.4%）、3,972名と比較すると、倍増に近い伸びを示している。さらに、韓国語科目を開講している四年制大学も、1995年度の143校から2003年度には335校に増加している。筆者の一人が勤めている九州大学においても、第2外国語として韓国語の学習を希望する学生数は、2000年度の815名から2006年度には1,897名へと増加している。このような数値からも学習者数の増加傾向を類推することができる（曺，2007）。

韓国語は日本語と文法構造や語順が一致し、言語運用の思考形態が類似しているため、学習しやすい外国語のはずである。しかしながら、多岐にわたる用言の活用変化を習得することは必ずしも容易なことではない。韓国語の用言の活用変化については、不規則活用の数の多さはいうまでもなく、いわゆる規則活用においても語幹の種類や語尾の種類によって実現される活用形は複雑多岐にわたる。文章の中で表現されている活用変化形から基本形を見出すことが簡単ではないため、初級学習者は辞書が引けるようになると「一人前」といえる。もっとも、既存のWeb辞書や電子辞書、紙媒体の辞書などにおいては、基本形からしか辞書を引くことができない。学習者は、多様な活用形から基本形を見出すためのルールを覚える必要があり、これがかなりの学習負担となっている。このような現状を背景に、活用形から引く学習辞書の開発が強く求められているといえよう。

本稿では、あらゆる活用形をクリックするだけで活用形から基本形が見出せる韓国語

Web 辞書の開発について、主として見出し語の検索の仕組みとプロセスを中心に検討を行う。

２．　韓国語学習者用 Web 辞書と見出し語

２．１　見出し語の選定

韓国語の場合、初・中級学習者が学習の際に接する語彙は約 3,000 語程度と推測される。韓国の국립국어원（以下、国立国語院）が定める韓国語の学習語彙は、初級（A レベル）のものが 982 語彙、中級（B レベル）のものが 2,111 語彙、上級（C レベル）のものが 2,837 語彙となっている（韓国国立国語院ホームページ参照）。また、韓国語学習辞典の代表として最も利用度の高い『외국인을 위한 한국어 학습 사전（外国人のための韓国語学習辞典）』（2005）（以下、学習辞典（2005）と略記）においては、初・中級学習者用の重要基礎語彙は約 2,975 語彙とされている。さらに、「ハングル」能力検定試験を主管するハングル能力検定協会（2006）では、学習レベルの目安として、初級語彙を約 1,400 語彙（5 級:約 450 語彙、4 級:約 950 語彙）、中級前半（3 級）の語彙を約 1,500 語彙とし、およそ 2,900 語彙を初・中級学習語彙として提示している。上記の三者において学習基礎語彙の数に若干の違いはあるものの、初・中級学習者が学習の際に接する語彙は、およそ 3,000 語彙程度といえよう。このうち、活用形を持つ動詞・形容詞などの用言の数については、学習辞典（2005）では 1,120 語彙、国立国語院による学習語彙では 863 語彙（A レベル 230 語彙、B レベル 633 語彙）となっており、およそ 1,000 語彙といえそうである。これらの用言の基本形は「語幹＋다」の形を成している。

今回の Web 辞書の開発に当たっては、学習辞典（2005）の重要語彙 2,975 語彙を基準にしている。このうち活用のある用言は 1,120 語彙で、見出し語全体の 1/3 を占める。この 1,120 語彙を活用による語幹のタイプ別に分類したのが図表 1 である（曺，2007）。図表 1 において特徴的なことは、母音語幹（841 語彙）が全体の 75.1%を占める点である。そのため、一音節内に活用形が混在する場合が非常に多く、初級学習者は辞書を引く際に語幹の判別に戸惑うことになる。さらに、母音語幹と子音語幹のいずれにおいても、規則活用の語幹を持つ語彙数は半分程度に止まっており、不規則活用の語幹を持つ語彙数の多さをうかがい知ることができる。一音節の中に活用形が混在する母音語幹の数が多いこと、ならびに不規則活用の語幹を持つ語彙数が多いことなどが韓国語学習の困難さを高め、さらには学習の支援を目的とする機械的なシステム構築の難しさをも高めている。

語幹	母音語幹　841 語　(75. 1%)					子音語幹						合計
	498 語 (44. 5%)				343 語 (30. 6%)	279 語 (24. 9%)						1120 (100%)
	規則	ㅇ変	ㄹ変	러変	～하語幹	規則	ㄹ変	ㅂ変	ㄷ変	ㅎ変	ㅅ変	
語彙数	445	21	27	5	343	147	58	49	6	12	7	

図表１　重要基礎語彙（初・中級）の用言語幹タイプ

２．２　用言の活用

韓国語は、日本語と同様、文の主要成分とその文法的意味を表す補助成分によって語節が構成される。韓国語は語節を分かち書きすることによって文の成分間をはっきりと区切っている。例えば、以下の例のように分かち書きがなされる。韓国語の分かち書きを日本語の文に「/」を入れた形で表してみよう。

韓）어제 서점에 가서 책을 샀습니다.

日）昨日/本屋に/行って/本を/買いました。

英語や日本語の語節又は文節は、主要成分に付随成分を付け加えることによって成り立っている。例えば、以下の例にみるように、英語の場合は単語「help」に過去形の語尾 ed をつけて、日本語の場合は「助ける」の語幹である「助け」に過去形「た」をつけることによって、それぞれ動詞の過去形を表すことができる。学習者は語節の後ろから一文字ずつ切り取った状態を想定しながら、前の方から一音節ずつ辞書を引いていく。そのうち、前の部分が辞書に収録されている単語（例えば、help）、あるいは原型が類推できる単語「助け（助ける）」にたどりつく。その時点で、前の単語の要素に語尾や助詞が付いたとみなし、その意味を調べることができる。

英）He helped her, but　⇒　help　＋　ed

日）彼は彼女を助けたけど　⇒　助け（語幹）＋ た(過去)　＋ けど（逆接）

韓）그는 그녀를 도왔지만 ⇒ 돕(語幹)　＋　았（過去）　＋ 지만（…けど: 逆接）

しかし、韓国語の場合、検索対象の語節「도왔지만」を日本語や英語と同様な方法で区切っても、動詞語幹に該当する「돕」を割り出すことはできず、辞書を引くことは不可能になる。このような韓国語の特徴を考慮し、本稿では、語節を音節単位で分析するのではなく、字母単位で分析するシステムを構築することにした。

韓国語の場合、語幹に語尾が接続する際に起こりうる活用変化のパターンは非常に複雑多岐にわたる。活用パターンは、語幹と語尾の特徴によって分類することができる。語幹の特徴による分類では、語尾の接続に際して、①語幹が変化しないもの、②語幹が変化するもの、③語幹末尾が脱落するもの、といった３つのタイプが考えられる。さらに、語幹

が変化する場合には、その語幹が母音語幹か子音語幹か、あるいはㄹ語幹かによって活用変化が異なる。また、母音語幹の場合は、陽母音語幹か陰母音語幹かによっても活用変化には違いが生じる[1)]。

韓国語の用言タイプを網羅的に分類している小学館・金星出版社（1993）の分類によれば、韓国語における用言タイプは 67 種類となっている[2)]。そこでは、用言を規則変化用言、不規則変化用言に分けるとともに、その音韻変化をも考慮に入れている。すなわち、語幹は用言の基本形から最後音節の‘-다’をとった形になるが、語尾との結合において、下記の例のように語幹部分に変化が伴う変則が存在するのである。

例）듣다（聞く/聴く）　：　듣（語幹）　＋　어　⇒　들어（聴いて）

これらを含む、すべての用言タイプを辞書データベースから検索するためには、存在しうるすべての用言タイプごとに語尾の結合規則を識別できる固有のコードを与える必要がある。図表２は、検索システムのための細かい分類例を示したものである。

変則の種類	グループ記号**	分　　類	用　例
正規	V1a01	語幹がㅏでおわるもの	만나다　会う
	V1a02	語幹がㅓでおわるもの	서다　立つ
	V1a03	語幹がㅐ,ㅑ,ㅔ,ㅕでおわるもの	보내다　送る
	V1a04	語幹がㅞ, ㅟ, ㅢでおわるもの	뛰다　走る
	V1a05	語幹がㅗでおわるもの	보다　見る
	V1a06	語幹がㅚでおわるもの	되다　なる
	V1a07	語幹がㅜでおわるもの	주다　与える
	V1a08	語幹がㅣでおわるもの	기다　這う
	V1a09	거라変則	가다　行く
	V1a10	너라変則	오다　来る
	V1a11	陽母音の子音語幹のもの	받다　受ける
으変則	V2a01	으の前の音節が陽母音を含むもの	따르다　従う
	V2a02	으の前の音節が陰母音を含むもの	쓰다　書く
르変則	V2b01	르の前の音節が陽母音を含むもの	오르다　上る
	V2b02	르の前の音節が陰母音を含むもの	부르다　呼ぶ
러変則	V2c01	러変則動詞はこれのみ	이르다　至る
우変則	V2d01	우変則動詞はこれのみ	푸다　汲む

図表２　検索のための用言タイプの記述（一例）*

* 小学館・金星出版社（1993）の分類をもとに、グループ記号を追加して作成した。
** V：品詞記号、数字 1～：用言タイプ、a～：変則のタイプ、数字 01～：変則の下位分類を各々表している。以下、同じ。

２．３　活用語尾の結合規則

次に語尾の特徴を見てみよう。語尾は語幹と結合して意味を付加する機能を持つ。国立国語院（2005b）によれば、韓国語の語尾の数は、異形態を除いて 800～900 種類となっている。そのうち、初・中級学習者が比較的頻繁に接する基本語尾はおおよそ 1 割程度と考えられる（曺，2007）[3)]。しかし、語尾と語尾が膠着して結合する場合も多いため、その数は相当な数にのぼる。そうした結合のなかには、「語幹＋語尾」の音節間の接続だけで生成される用言の正規変化と、語幹と語尾の結合により語幹の音韻変化がともなう変則が存在する。また正規変化の場合であっても、語幹の終わりの音韻が子音か母音かによって語尾の接続形態が変化することもある。さらに、動詞か形容詞かによってもその接続パターンは異なる。これらの語尾の特徴から、①語幹のパターンに影響されず一定の形で語幹に接続するタイプ、②母音語幹・子音語幹・ㄹ語幹の区別によってその形を変えて接続するタイプ、③陽母音語幹・陰母音語幹・하다語幹の区別によってその形を変えて接続するタイプ等が考えられる。さらに、母音語幹は縮約を伴う場合もある。次の例を見てみよう。

例）지으셨지만　お作りになりましたけれども

⇒　짓　＋　(으)시　＋　었　＋　지만

짓다の語幹　尊敬の先語末語尾　時制の先語末語尾　語尾（逆接）

⇒　지　＋　으셨　＋　지만

짓다の音韻変化　尊敬の先語末語尾＋時制の先語末語尾　語尾（逆接）

上記の例では、「짓다」の語幹「짓」に尊敬を表す「시」が接続しているが、「짓」は子音で終わる語幹であるため「(으)」を伴うとともに、「짓」は「지」に音韻変化を起こしている。また、過去時制を表す先語末語尾「었」と逆接語尾「지만」が接続しているが、「시」と「었」は縮約され、「셨」として実現されている。

３．　韓国語学習者用 Web 辞書における検索の仕組み

３．１　仮想語幹と仮想語尾の概念

学習者による用言の基本形の抽出は、学習の積み重ねによる経験や文脈の把握、助詞や語尾の習得など総合的な学習と判断により可能となる。しかしながら、人間の脳の中の働きとは異なる機械のシステムによって基本形を抽出するためには、基本形抽出のための新たなロジックが必要である。そこで、曺（2007）が提案した仮想語幹と仮想語尾の概念を用い、活用形から基本形を抽出するシステムを構築することにした。仮想語幹とは、学習者がクリックした語彙を字母単位にまで分解し、組み合わせ直したときに変化を伴わない部分を指す。これに予め固有コードを与え、仮想語幹データベースに登録しておくのである。同様に、仮想語尾も変化を伴わないものと考え、それぞれに予め固有コードを与え仮

想語尾データベースに登録しておく。このシステムの下では、学習者がクリックした語彙は瞬時に字母に分解され、仮想語幹と仮想語尾とに分けられる。そして、それぞれのデータベースから一致する仮想語幹と仮想語尾の検索を行うのである。

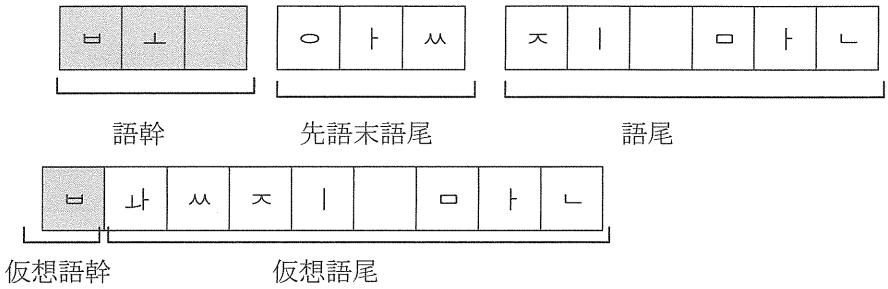
システムの中では、学習者がクリックした語彙に対して常に三段階にわたる検索が行われる。まずは、クリックしたそのままの形が仮想語幹データベースに存在するか否かを確認し、辞書データを出力する。次の段階として、その語彙が「体言＋助詞」のタイプであるか否かを判断する。すなわち、用言語節ではなく体言語節であると想定した分析を行い、検索結果を出力する。その後、第三段階として、用言語節としての分析に移る。すなわち、当該語彙が「用言の活用形を含むもの」と想定し分析を行うのである。分解された字母を仮想語幹テーブルならびに仮想語尾テーブルと照合し、基本形と接続語尾の見出し語を抽出する。その後、基本形と接続語尾の見出し語の意味を辞書記述データ INDEX から参照し、辞書として表示するのである。

このような三段階にわたる検索が必要なのは、韓国語の特徴として、「体言＋助詞」の組み合わせと「用言語幹＋活用語尾」の組み合わせが同じ形である場合が存在するからである。

３．２　仮想語幹と仮想語尾を用いた用言検索の仕組み

仮想語幹と仮想語尾の概念を導入することで、用言の活用形は「変化を伴わない語幹（仮想語幹）＋変化を伴わない語尾（仮想語尾）」の結合として捉えることができる。図表３の例を見てみよう。「봤지만（見たけれども）」は用言「보다（見る）」の語幹「보」に過去時制語尾「았」＋逆接語尾「지만」が接続したものである。「보다（見る）」の実際の語幹は「보」であり、活用形で実現される場合は「보지만」「보니」「보고」のように語幹「보」が変化せず表示されるが、「봐요」「봐도」のように「봐」（音韻変化：ㅗ→ㅘ）になる場合もある。このとき、「봐」は辞書の見出し語に登録されていない場合がほとんどで、用言と語尾の意味が分かっていても音韻変化の規則を適用できない限り、実際の語幹「보」を見つけるのは難しい。

そこで筆者らは、「보다」という用言において、どのような活用形にも現れる最初の子音「ㅂ」を語幹と仮定し、「仮想語幹」と名づけた。語尾については、時制・尊敬などの先語末語尾を含んだもの全部を「仮想語尾」と呼ぶことにした。図表３は、「仮想語幹」と「仮想語尾」の考え方を図示したものである。



図表３　仮想語幹と仮想語尾

このような考えの下、潜在的に存在しうる全ての仮想語幹と仮想語尾が辞書のデータベースに登録されている。仮想語幹の意味は何か、仮想語尾はどのような語尾の結合なのかが詳細に登録されていれば、活用形の基本形を見出したとき、基本形の意味提示と共に活用形と関連した文法構造が理解できる。「仮想語幹」と「仮想語尾」の概念を導入した検索により、正確な意味抽出が可能となるばかりでなく、高速な検索も期待できるのである。

４．　データベース構築と見出し語検索

４．１　仮想語幹と仮想語尾のデータベース構築

韓国語の文字の処理においては、UNICODE のハングルコードの配列を利用する。このコード自体には、文字を構成する子音(C)と母音(V)の情報が含まれていない。ただし、CV の後に C が来る順番によって文字が羅列されているため、文字を構成している CVC の字母を文字コードで特定することは可能である。つまり、ハングルは 16 進数 AC00 から始まり、子音は 588 ごとに、母音は 28 ごとに反復される。

ハングルコード　= 44032（AC00）　+ 子音 * 588 + 母音 * 28 + 子音

図表４　UNICODE ハングルコード表

UNICODE のハングル配列と単語を記述した体言や用言の固有コードを参照し、下記のような仮想語幹データベースを作ることができる。このデータベースには、韓国語 Web 辞書の見出し語を CVC に分解し仮想語幹にしたものと、体言や用言の種類による固有番

号、本辞書でのインデックス番号が含まれている。このインデックス・フィールドは、用言の基本形を検索するときに用いられることになる。図表５は、仮想語幹テーブルの一例を示したものである。

ID	WORD	IDX*	意味
2815	ㄱ	V1a08,2932	動詞の仮想語幹
2816	ㄱㅏ	NN,1;V1a09,2819; XV,2820; JX・・・・	「端・辺」(名詞)、「行く」(動詞・補助動詞)の語幹、「ガ」(助詞)・・・・
2912	ㄱㅏ ㄱㅔ	NN,3	「店」(名詞)
3117	ㅁㅏ ㄴ ㄴㅏ	V1a01,	「会う」(動詞)の語幹

*V：動詞、NN：名詞、XV：補助動詞、JX：助詞

図表５　仮想語幹テーブルの例

仮想語尾データベースには、活用形の語尾連結規則を考慮し、用言のすべてのグループに対応できる先語末語尾（例えば、尊敬、時制など）と語末語尾の組み合わせのあらゆるパターンを収録した。現時点では、用言の種類（30 種）　×　尊敬の有無（2 種）　×　時制（8 種）×　語尾（69 種）の組み合わせに対してインデックスしたもの 10,000 個程度を登録している。図表６は、仮想語尾テーブルの一例を示している。

ID	KATUYOU	IDX
5286	ㅡ ㄱㅔㅆㄴㅡ ㄴㅑ	V2a01,02031; V2a02,02031
5287	ㅡ ㄱㅔㅆㄴㅡㄴㄱㅏ	V2a01,02032; V2a02,02032
5289	ㅡ ㄱㅔㅆㄱㅜ ㄴㅏ	V2a01,02034; V2a02,02034

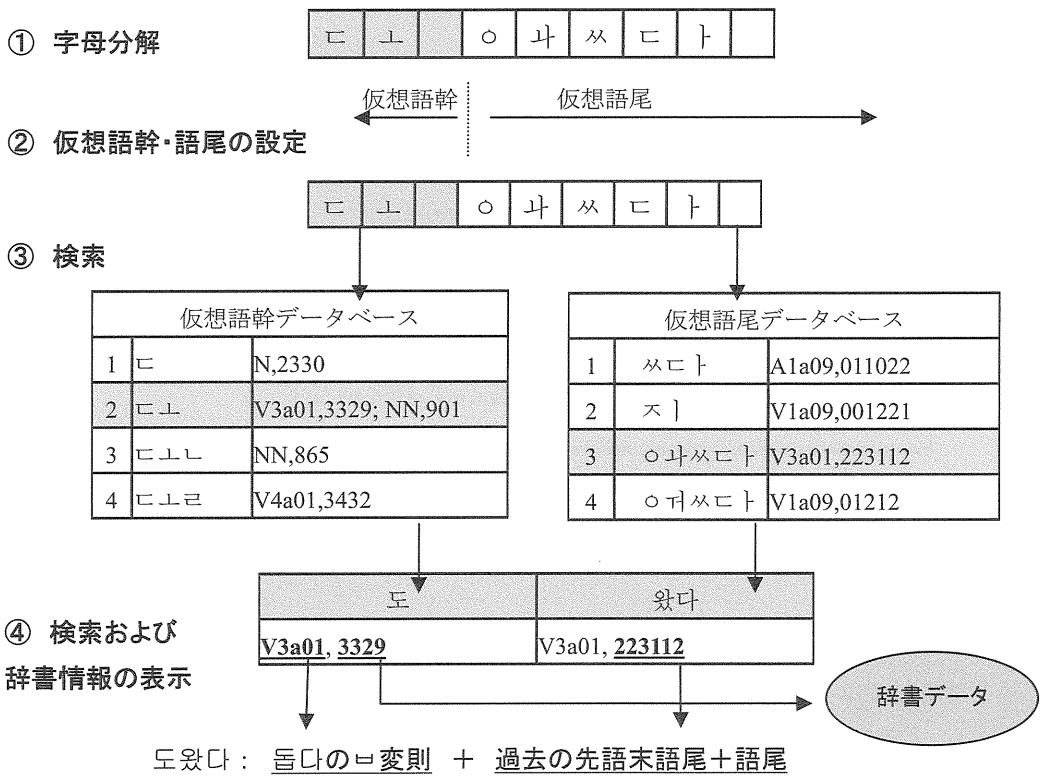
図表６　仮想語尾テーブルの例

４．２　用言の活用形の検索

ここでは、学習者がクリックした語節単位の検索対象語に対して、それがどの活用形なのかを検索し、基本形の意味と活用形の構成要素を見出す処理について概観する(図表７)。まず、検索対象語は「仮想語幹＋仮想語尾」として検索されるため、語節を音節ごとに切り離し、音節を子音＋母音＋子音の字母単位に分解する。一音節は初声・中声・終声がCVC の構造を持つため、例えば、「ㅌ」のように母音で終わる音節については、終声の代わりに空白をマークした。つまり、「ㅌ」の場合、「ㄷ(初声：子音)＋ㅗ(中声：母音)＋ゼロの終声(空白)」を配置し、分解した字母が現れる初声・中声・終声をマークした(図表７

の①)。

字母単位で分解した検索対象語は、後ろからインデックスを増加させながら検索語を仮想語幹と仮想語尾とに分離する(図表７の②)。検索対象語を任意の仮想語幹・仮想語尾に分けた後、それぞれ仮想語幹と仮想語尾のデータベースから一致するものがあるか否かを調べる(図表７の③)。検索の段階では、マッチングする文字列およびその語幹のタイプと語尾のタイプが一致するかを調べる。仮想語幹と仮想語尾に一致する文字列があり用言の分類が一致すると、それをありうる活用形として認識し結果を出力する。語幹データベースにはその語幹をもつ用言が記載されている辞書データのインデックスがあり、それを参照して用言の意味を出力する(図表７の④)。



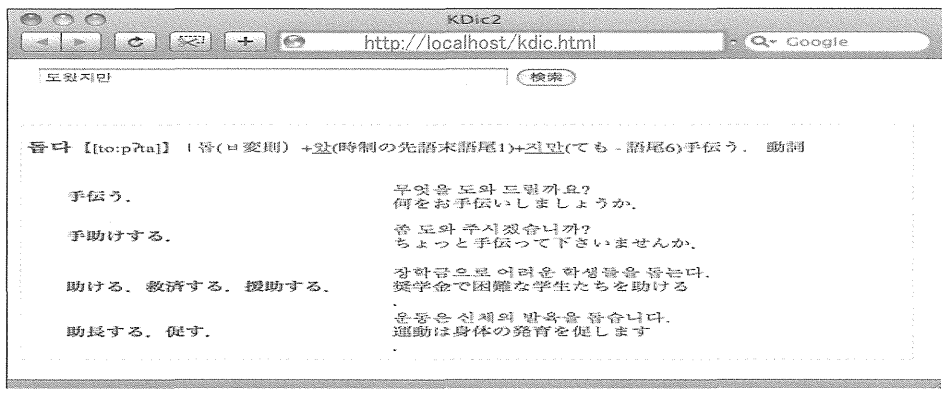
図表７　仮想語幹と仮想語尾データベースからの基本形検索

一方、仮想語尾データベースからは、その活用形がどのような語幹と語尾が結合して音韻変化を起こしたのかを特定するための情報を得ることができる。また、語尾生成システムにより作成されたインデックス番号を記入しておくことによって、検索語の文法事項表示が可能となる (図表７の④) 。

4．3　辞書のインターフェース

次に、韓国語学習者用 Web 辞書のインターフェースを示すと図表８のとおりである。ある検索語（ここでは、「도왔지만」）に対して、字母分解によって仮想語幹・仮想語尾データベースにおいて検索を行い、用言の基本形を抽出し、その意味を辞書データベースから検索し表示している。さらに、接続されている語尾の種類と語尾の意味へのリンクが表示され（왔（時制の先語末語尾）＋ 지만（でも-語尾））、語尾の意味を把握することが可能となっている。

このように、語節を解析する形の検索システムにより、学習者は未知の語節をクリックすることで、その語節について、基本形の意味と接続されている語尾の文法事項を学習することができるのである。



図表８　韓国語 Web 辞書のインターフェース

5．　むすび

外国語としての韓国語学習において、複雑に変化する用言の活用形を正しく理解することは容易ではない。クリックするだけで活用形から基本形を見出してくれる辞書の実現により、学習者の学習負担は大幅に軽減されるはずである。基本形を見出せない初・中級学習者は、そのような辞書の助けを借りることで、用言の活用に圧倒されることなく学習を続けられるであろう。本 Web 辞書を活用することで、とりわけ日本人学習者は、韓国語を日本語並みに運用できる力を身につけられると期待している。

文中に出ている語節がどのような要素の組み合わせでできているのかを把握せずに辞書を調べても、その文の意味を十分に理解することは難しい。その点、学習者のクリックする語節単位の検索語に対して、その語節の主要成分の意味のみならず文法的構成をも提

示できる本 Web 辞書は、韓国語学習ならびに自動翻訳の精度アップに大いに貢献できるであろう。辞書のあり方を大きく変える可能性をもつ全く新しいタイプの韓国語 Web 辞書の開発に寄せられた期待は大きく、その実用性を検証していくことは今後の研究課題である。今後さらにシステムの精緻化を図っていきたいと考えている。

＊本稿は、平成 19-21 年度科学研究費補助金研究基盤研究 B（課題番号：19320082）「韓国語語彙データベース構築と外国語サイバー･ユニバーシティ用韓国語 Web 辞書開発」による研究成果の一部である。

注

1) 形容詞、指定詞、存在詞などの品詞によって活用変化が影響される場合もある。
2) 曺（2007）では、学習者のための大まかな分類として 21 種類の分類を提示している。しかし、21 種類のパターンが全ての用言タイプを網羅しているわけではない。学習者が同一グループとして判断できそうな細かい部分については、それを同一グループとして処理しているからである。しかしながら、システム設計においては機械的な判断になるため、類推判断はきかず、全ての用言タイプの提示が求められる。そこで、韓国語の用言タイプを網羅的に提示している小学館・金星出版社（1993）の分類を採用することにした。
3) ハングル能力検定協会の主管する「「ハングル」能力検定試験」においては、初級（４・５級）の語尾として約 80 個、中級前半（３級）の語尾として約 170 個が提示されている。

参考文献

국립국어원(2005a)『외국인을 위한 한국어 문법 1 체계편』, 커뮤니케이션북스, 서울.
국립국어원(2005b)『외국인을 위한 한국어 문법 2 용법편』, 커뮤니케이션북스, 서울.
국립국어원･한국어세계화재단기획(2005)『외국인을 위한 한국어 학습 사전』, 신원프라임, 서울.
国立国語研究所ホームページ：http://www.korean.go.kr/06_new/index.jsp
연세대학교 언어정보 개발연구원편(1998)『연세 한국어 사전』, 두산동아, 서울.
国際文化フォーラム(2005)『日本の学校における韓国朝鮮語教育―大学等と高等学校の現状と課題―』, 財団法人国際文化フォーラム, 東京.
小学館・金星出版社(1993)『朝鮮語辞典』, 小学館, 東京.
曺美庚（2007）「活用形から引く韓国語 Web 辞書の開発に関する一考察」『言語科学』, 九州大学大学院言語文化研究院言語研究会, 福岡.
ハングル能力検定協会（2006）『「ハングル」検定公式ガイド：合格トウミ』ハングル能力検定協会, 東京.