

韓国語学習者作文コーパス (KC Corpus) と

韓国語教育への活用¹

林 炫情(山口県立大学)²

李 在鎬 (国際交流基金日本語試験センター)³

黄 晟煥(立命館アジア太平洋大学)⁴

浅尾 仁彦(SUNY Buffalo/京都大学)⁵

要 旨

本稿では、韓国語学習者作文コーパスである「KC Corpus (Korean L2 Learner's Written Composition Corpus)」について、その概要と教育現場での活用の一例を示した。そして、より効果的な韓国語教育のシステムを構築し、有効なものとして活用していくための KC Corpus の可能性について検討を行った。KC Corpus は、韓国語学習者における産出的言語使用（作文）の問題点を客観的データに基づいて分析することが可能であり、分析結果を容易に教育現場へのフィードバックできるという点で、韓国語教育のための有効なツールであると考えられる。現時点での KC Corpus の課題としては、(1) 学習者の主な韓国語学習機関が限定されていること、(2) 上級レベルのデータが不足している。また、(3) 添削や形態素の誤りがまだ一部残っていることがあげられる。今後は、韓国語学習者作文コーパスの規模の拡充を図りながら、複数の添削者による添削を実施していきたいと考えている。またより効果的な第2言語教育、とりわけ韓国語教育への応用をも視野に入れ、現在開発中の韓国語教育スタンダード (can do) との有機関係を模索しながら、韓国語教育への活用可能性を模索していく予定である。

キーワード：韓国語学習者、KC Corpus、誤用分析、韓国語教育への活用

1. はじめに

本稿は、韓国語学習者作文コーパスである「KC Corpus (Korean L2 Learner's Written Composition Corpus)」を紹介する。そして、KC Corpus の誤用分析の集計から得られた知見を韓国語教育現場で活用するための一例を提示する。また、より効果的な韓国語教育のシステムを構築し、有効なものとして活用していくための KC Corpus の可能性について考えてみることを目的とする。

¹ 本稿は、2010年9月に行われた第4回山口県立大学学術研究会（優秀研究報告）の報告資料をもとに、加筆を行ったものである。

² Hyunjung, Lim, Associate Professor, Yamaguchi Prefectural University, Japan; hylim@yamaguchi-pu.ac.jp

³ Jae-ho, Lee, Researcher, The Japan Foundation, Center for Japanese-Language Testing, Japan; jhlee.n@gmail.com

⁴ Jung-nan, Hwang, Senior Lecturers, Ritsumeikan Asia Pacific University, Japan; nancy0518@yahoo.co.jp

⁵ Yoshihiko, Asao, Doctoral Student, Grad.Sch. SUNY Buffalo, USA/ Kyoto University, Japan; asaokitan@ling.bun.kyoto-u.ac.jp

2. 「KC Corpus (Korean L2 Learner's Written Composition Corpus)」

2.1 コーパス (Corpus) と言語教育

一般に、コーパスは言語研究に使用するために大量に収集された書き言葉および話し言葉のテキストを電子化し、コンピュータ処理できるようにした「電子化テキスト資料」を指す。このようなコーパスは多量のデータを踏まえて語の隠れた特徴や言語の傾向性を明らかにすることができるため、語法研究、社会言語学、辞書編纂、文体論などといった言語テキストを扱う広範な研究領域において活用することができる (石川, 2008)。特に、学習者の発話・作文データを集めた学習者誤用コーパスは、学習者が産出した大量の生データを客観的かつ数量的に分析することで信頼性の高い誤用分析を可能にするとともに、主観的経験から得られた知見とはまた異なった知見を提供してくれる。また、その結果は学習者の立場から間違いやすい目標言語の特性を把握し、学習者に目標言語をより正しく理解させ、できるだけ誤用を少なくするための有効な教授方法や教材開発にもつなげることができる点で、コーパスの言語教育への貢献の余地と利用価値は高い (서상규 외, 2002; 石川, 2008) といえよう。

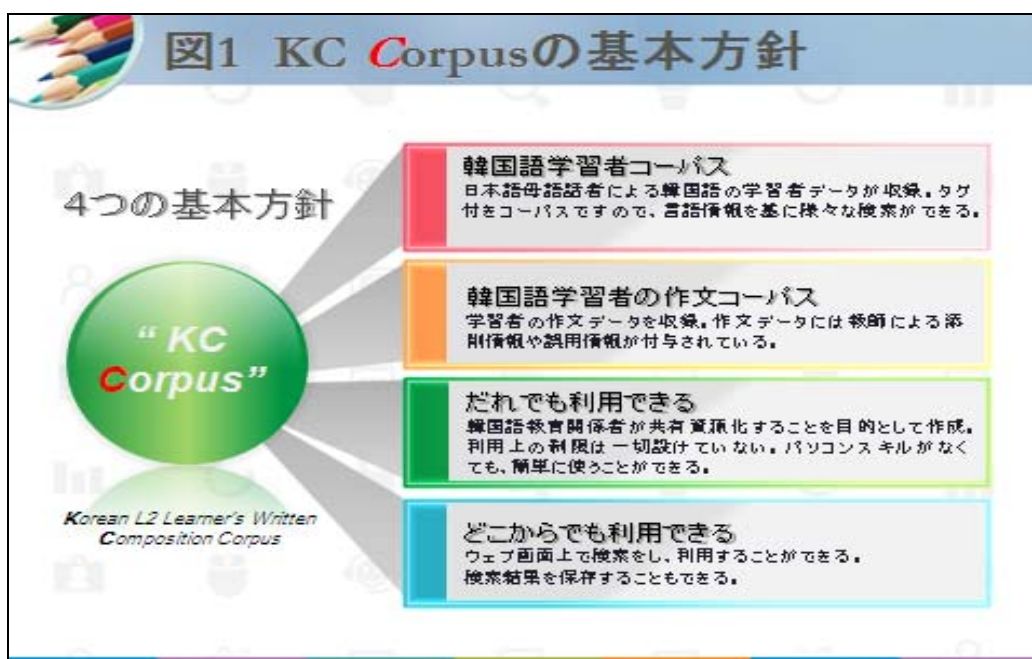
2.2 「KC Corpus」の開発背景と目的

近年、韓国語のみならず、語学学習者のニーズは一層多様化されており、語学教員に対してはより質の高い教育、学習者のニーズや習熟度に柔軟に対応できる教育コンテンツの開発、そして効果的な教授法の支援などが求められている。このような状況のなか、学習者コーパスはその活用が多方面において期待されている。

一方、韓国語における学習者コーパスは、延世コーパス⁶のほか、いくつかの学習者コーパスがあるが、一般公開されているものは極めて少なく、検索ツールにおいてもその使用方法が容易でないため、学習者コーパスを活用した研究は十分とはいえない (고석주 외, 2001; 서상규 외, 2002)。また、学習者の誤用分析においては学習者の学習環境は重要な変数要因になりうる。しかし、既存の韓国語学習者コーパスは韓国国内で韓国語を学習する学習者のデータを収集したものがほとんどであることから、海外で、特に日本で韓国語を学習する学習者の特性を反映していないのが現状である。

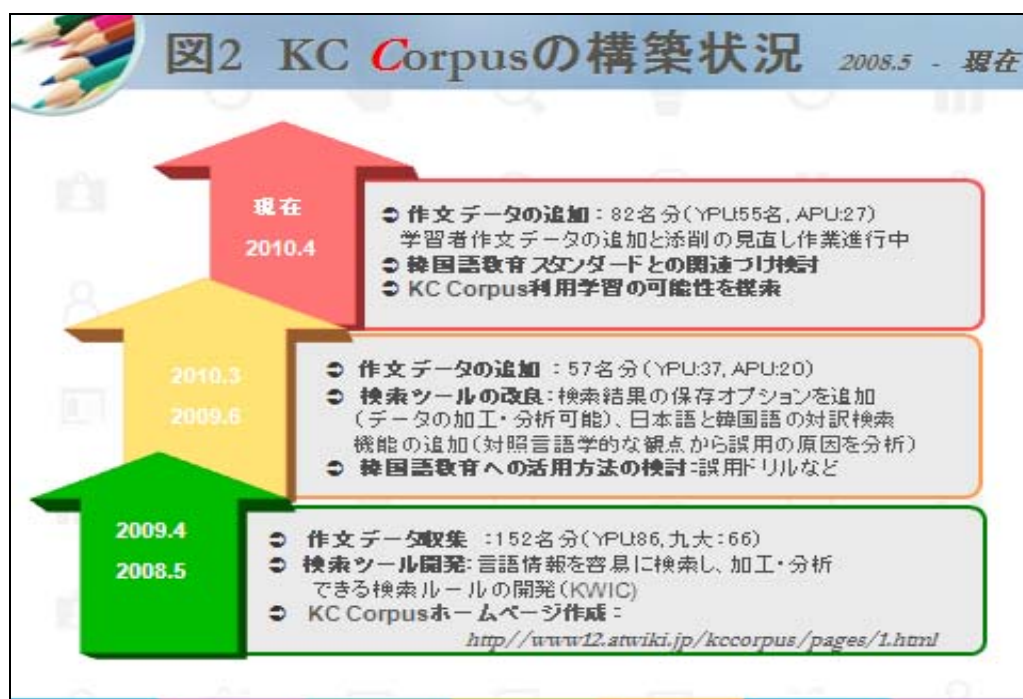
そこで、以上の問題点を踏まえ、KC Corpus では、日本の大学で韓国語を学習する日本語母語話者を対象に、作文データを収集し、電子データ化を行う。また、教師への共有化を前提にした使いやすい学習者コーパスの構築を目的とした。具体的には、複数の言語情報 (形態素情報、誤用タグ、添削情報) を付与することで、誤用分析に適したコーパスの開発、また、研究・教育者の様々な利用目的に応じた情報抽出が可能で、Web ブラウザ上での検索環境を提供するといったユーザフレンドリーな検索環境構築を目指した。さらに、KC Corpus のデータにおいては「集めたけど公開できない」という事態を避けるため、データ収集時に著作権処理も同時に行うことで、著作権問題においてもクリアしている。図1はKC Corpus の基本方針を示したものである。

⁶ 延世コーパスは延世大学の語学堂の韓国語学習者を対象にデータを収集したもので、規模は26万語節 (2002年現在)。



2.3 KC Corpus の構築状況⁷

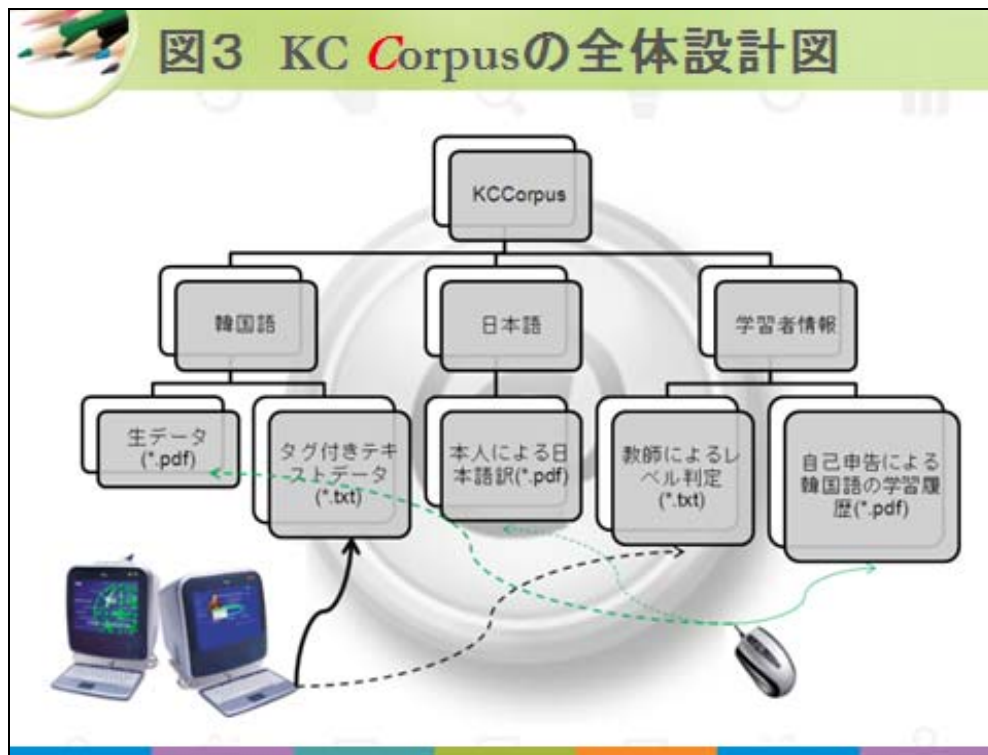
KC Corpus は 2008 年度から開発をスタートし、2009 年度に第 2 回目のデータ追加と検索ツールの改良を経て、2010 年度に第 3 回目のデータ追加のための作文データの収集を終え、添削の修正を行っている段階である。KC Corpus 構築の現在までの進行状況を年度別に示したのが図 2 である。



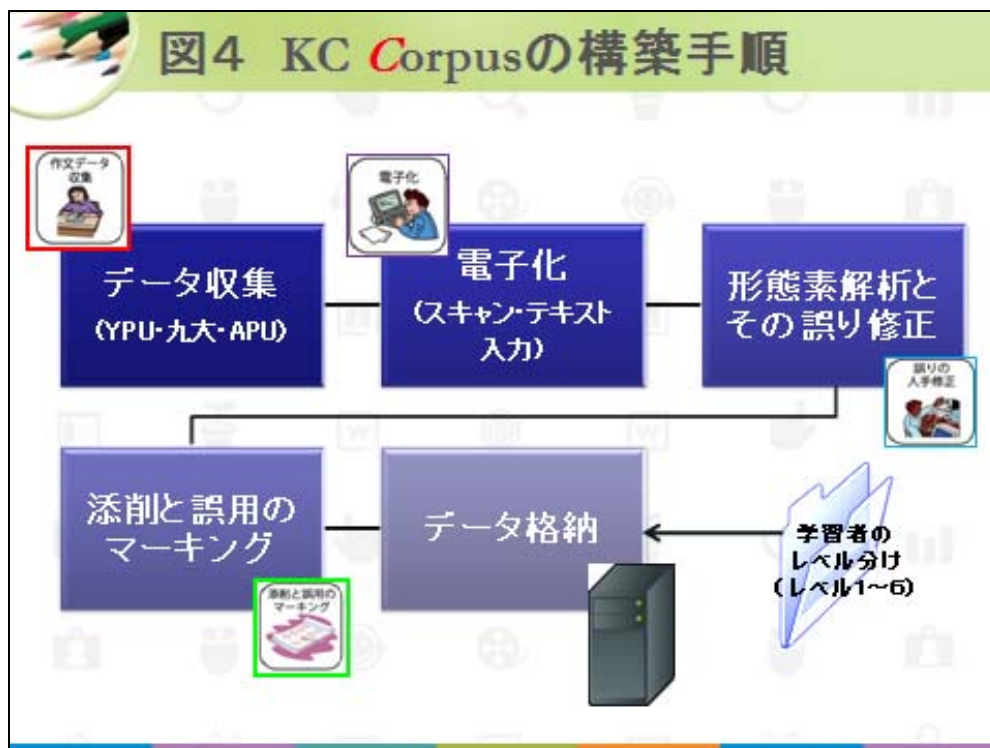
⁷ 本研究は、【Korean Studies Grant 2008 (課題番号：AKS-2008-R15)：2008 年 5 月-2009 年 4 月 (研究代表者：林炫情)「韓国語教育支援共有資源化のためのコーパス構築：日本語母語話者の作文データベース化を中心に」】と【平成 21 年度・山口県立大学学内研究助成金 (基盤研究 (A))：2009 年 6 月-2010 年 3 月 (研究代表者：林炫情)「韓国語学習者作文コーパス (KC Corpus) と韓国語教育への活用」】による補助を受けた。

2.4 KC Corpusの全体設計図と構築手順

KC Corpusは(1)「学習者による韓国語作文」、(2)「その母語訳(日本語)」、(3)「学習者情報」の3つのデータベースからなる。また、「学習者による韓国語作文」には「生データ(*pdf)」と「タグ付きテキストデータ(*txt)」, 「学習者情報」には「教師によるレベル判定(*txt)」と「自己申告による韓国語の学習履歴(*pdf)」がそれぞれ含まれている(図3)。



また、構築手順は図4に示した通りである。(1)データ収集では、韓国語の作文、その日本語訳、学習者履歴、公開に関する確認書(使用承諾書)を紙媒体で作成してもらった。(2)電子化では収集したすべてのドキュメントを電子ファイル化した。そして、(3)지능형 형태소 분석기(知能型形態素分析機)を使用して形態素解析を行い、必要な言語情報を付与した。解析の誤りについては人手で修正を行った。基本的なタグセットは「21 세기 세종계획말뭉치(21世紀セジョンコーパス)」のタグセットと同じである。(4)添削と誤用のマーキングでは、すべて人手で語節単位での添削を行いながら、文法的ミス、表記的ミス、文体的ミスの3種類に分け、それぞれの誤用タグを付与した。複数の誤用については、区切り文字を入れるとともに複数の誤用タグをつけ、いずれの検索条件でもヒットするようにした。(5)学習者のレベル判定はデータの格納時に行った。学習者のレベル判定の基準は、学習者の韓国語関連の検定試験資格、作文作成時の学習レベルと学習歴、留学の有無と期間などを総合して6段階レベルで判定を行った。構築手順の詳細は、林・李・曹・浅尾(2008)、임현정(2009)、李・林・浅尾・曹(2009)李・林・曹・浅尾(2010)を参照。



2.5 KC Corpus の使用法

現在、KC Corpus は、Web 上 (<http://www12.atwiki.jp/kccorpus/>) にて全文データを無償でダウンロードできるようになっているが、アクセスからデータ保存までの使用方法は下記の通りである。

2.5.1 KC Corpus (<http://www12.atwiki.jp/kccorpus/>) のホームページにアクセスし、メニューの使用法をクリックする。

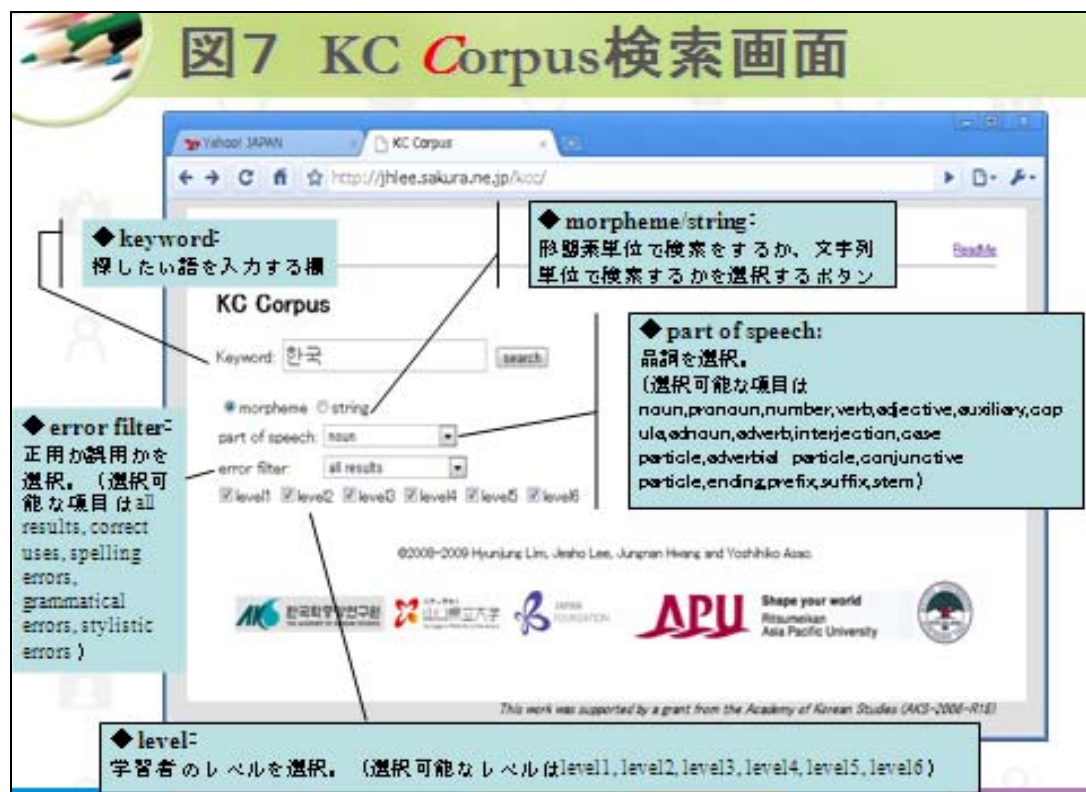


2.5.2 <http://jhlee.sakura.ne.jp/kcc> をクリックする。

接続ダイアログが表示されたら ID 「kor」とパスワード「kwickwic」を入力する。パスワードを設定したのは、入りを一本化するためのもので利用を制限するためのものではない。



2.5.3 検索オプション (morpheme/ string) を指定して「検索」ボタンをクリックする。



■形態素単位での検索

簡単な検索の場合、すべての単語は形態素単位で入れ、動詞、形容詞の場合は「-다 da」を除いた語幹のみ入力する。また、助詞の場合は、前接名詞のバッチムによる異形は別々に検索する。絞り込みで検索したい場合は、「検索語+品詞」で検索を行うことができる。例えば「가 ga」と入力するだけでは、動詞の「가다 gada(行く)」と助詞の「가 ga (が)」が共にヒットする。動詞のみヒットさせたい場合、あるいは助詞のみヒットさせたい場合には、part of speech で品詞を指定する。また、「検索語+添削情報」で検索を行うことも可能である。例えば「가다 gada(行く)」の正用例だけをヒットさせたい場合は「가 ga」+「correct uses」で検索する。誤用例を見たい場合は「errors」を選択する。誤用例の中でも文法的な誤用例のみを見たい場合は「grammatical errors」、綴りの誤用例を見たい場合は「spelling errors」、文体的なエラーを見たい場合は「stylistic errors」を選択する。「検索語+レベル」で検索を行う場合は、初級レベル(level1)から上級レベル(level6)の選択が可能である。



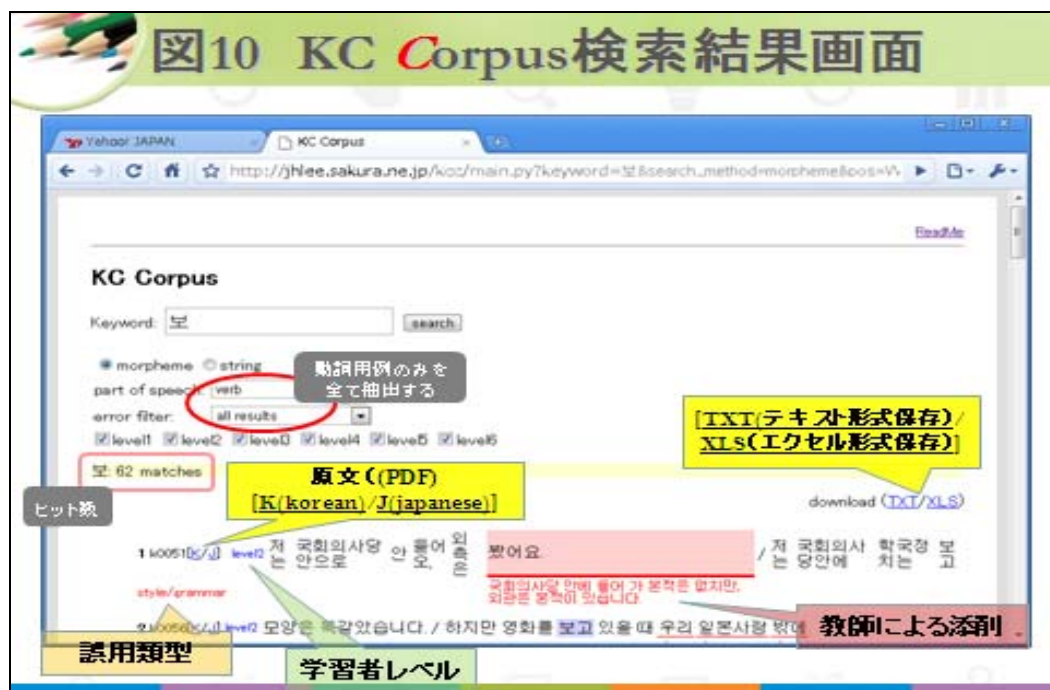
■文字列のみの検索

形態素を超える単位を検索する(複合的な要素)場合は、string ボタンをクリックの上、キーワードを入力する。



2.5.4 検索結果を確認する。

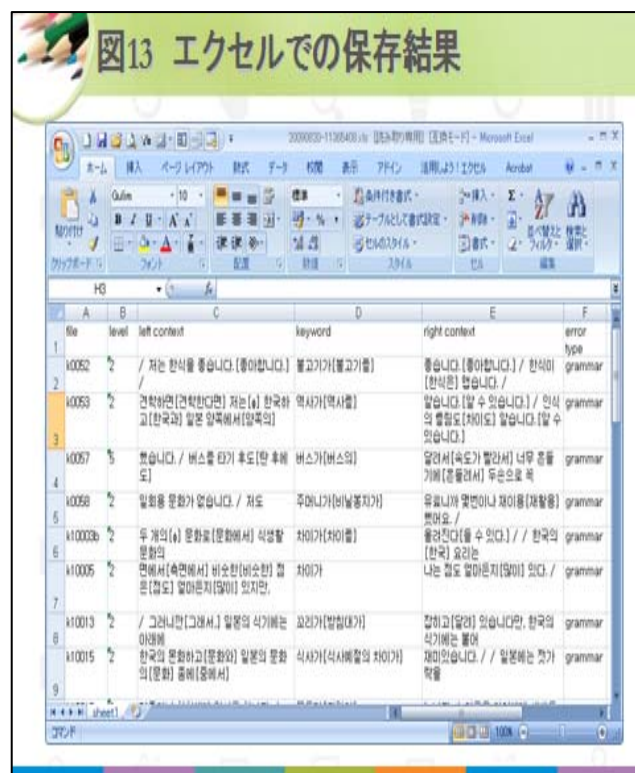
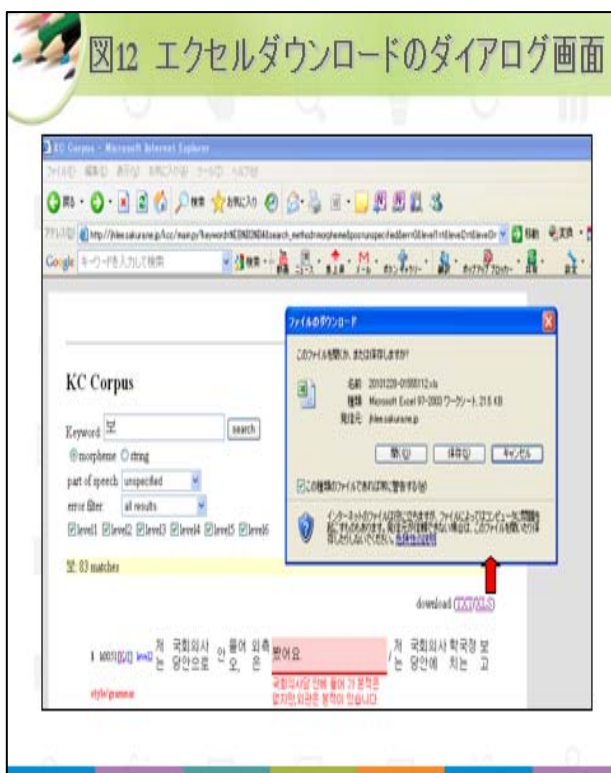
全体のヒット件数が表示される。また、学習者 ID | 元の PDF へのリンク【K/J】 | レベル【level2-6】 | KWIC データ | 添削情報(誤用タグ) の順で結果が表示される。元の PDF へのリンク【K/J】をクリックすると、元の作文データを PDF ファイルで見ることができる。日本語データを見たい場合は「J」を、韓国語のデータを見たい場合は「K」をクリックする。例えば、「보다 boda」の全用例をヒットさせてみると図 10 のような画面が表示される。ただし、検索結果画面に関する説明は表示されない。



2.5.5 検索結果を保存する。

download(TXT/XLS)では、検索結果を保存することができる。テキストファイル形式で保存したい場合はTXT をクリックする。クリックするとウェブ画面上 (図 11) で結果が表示される。コピー&ペーストで結果を保存することができる。また、XLS はエクセルファイル形式で保存する場合にクリックする。クリックするとエクセルのダウンロードダイアログが表示される (図 12)。図 13 はエクセルファイル形式での保存結果である。





3. KC Corpus の統計分析

3.1 全学習者データの集計

2010年10月現在まで集めた学習者作文データの総数は291名分（内訳は、山口県立大学178名分、九州大学66名分、立命館アジア太平洋大学47名分）である。ただし、2009年4月以後に収集したデータについては、現在添削の見直しを行っていることから、本稿では、第1回目に収集した152名分のデータのみを数値化し、統計処理を行った。152人の学習者から産出した総語節数（語節のトークン頻度）は20905語節で、一人あたりが産出した平均は137.5語節であった。外国人学習者が産出したデータの平均語節が、初級の場合は60-70語節を超えるのが難しく、中・上級の学習者の場合も100-120語節以内が多い（고석주 외, 2004）ことを考慮すると、少ない数のなかでも本データの産出量はある一定のレベルはクリアしているといえよう。また、誤用総語節数は4988語節のうち、誤用語節が占める割合は平均23.9であった。

図14で分かるように誤用類型は、文法ミス（形態・活用、文法要素の脱落などといった活用や文法的側面での誤用）が3808語節と最も多く、次に表記ミス（韓国語にはない文字などの誤用）、文体ミス（語順、表現や語用論的側面での誤用）の順であった。総語節のなかで誤用語節が占める割合は、各18.2%、2.9%、2.7%であるが、全体誤用を100でみた場合、それぞれの誤用が占める割合は文法誤用が76.3%、表記誤用が12.3%、文体誤用が11.3%となる。

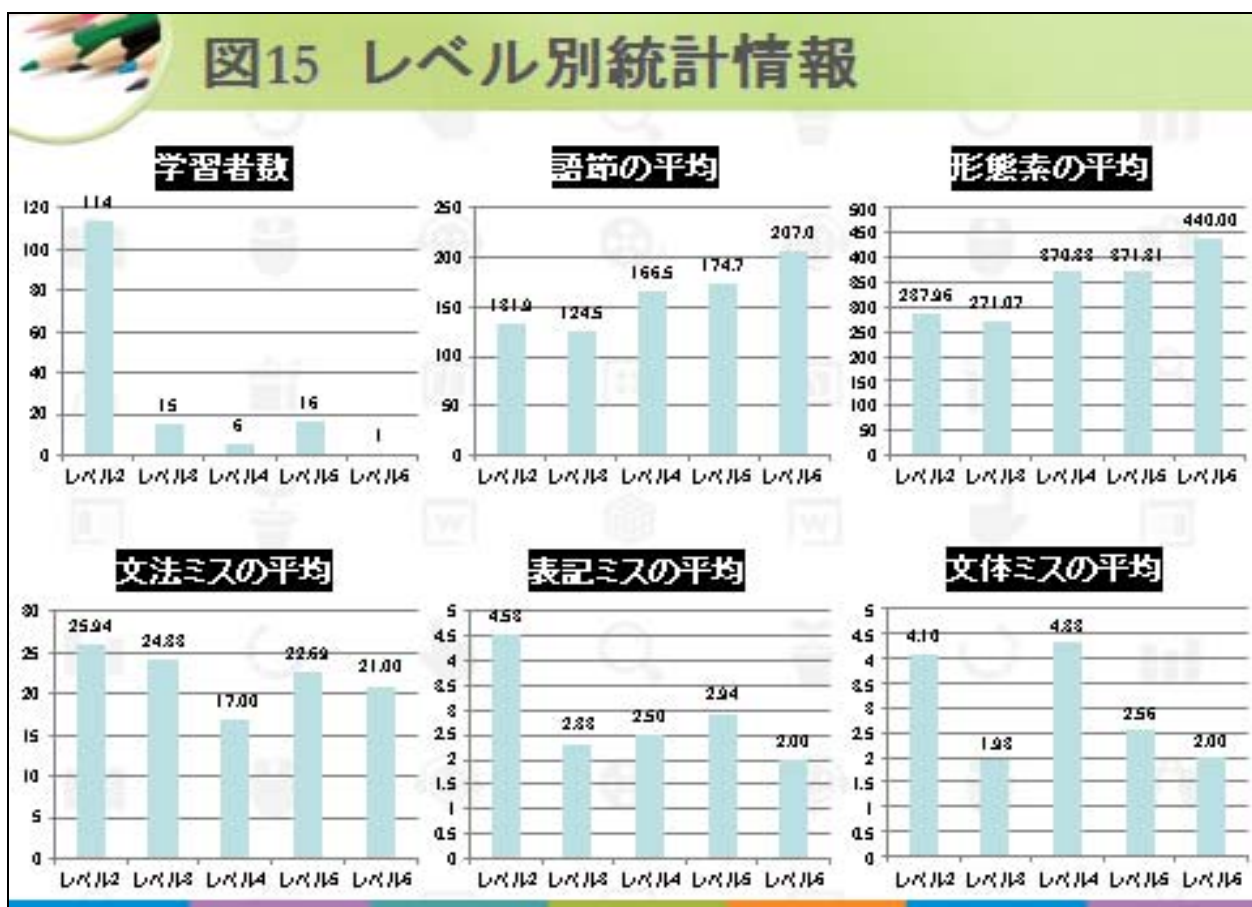


3.2 レベル別・類型別誤用分析

レベル別⁸の誤用分析の結果をグラフで示したのが図 15 である。学習者数はレベル 2 に偏っている。中級から高級にかけてのデータが少ない点が今後のデータ収集の課題として残る。総語節数はレベルが上がるにつれて語節数も増加する傾向が見られた。全レベルにおいて最も多かった誤用類型は文法ミスであった。誤用類型の特徴をレベル別にみると、文法ミスはレベル 2 において顕著に高く、レベル 4 で最も低いことが分かった。つまり、韓国語学習を始めて間もない学習者の場合は、助詞の使い分けに難しさを感じるようであるが、一定の学習期間を経るとその使い分けも徐々に安定してくるようである。レベル 4 より上級のレベル 5 とレベル 6 の学習者の文法ミスが増えたり、再び減ったりすることは、形態や活用の誤用よりは、助詞などの文法要素の脱落などに起因するものと思われる。次に、表記の誤用はレベル 2 において最も多いことから、初級の学習者に対する表記教育をもっと強化する必要があることが示唆された。

最後に、表現や語用論的側面の間違いである文体ミスでは、レベル 3 が最も少なく、レベル 4 が最も多かった。文法ミスで最も少なかったレベル 4 の学習者が文体ミスでは最もミスを犯していることは大変興味深い。この結果から、文法的知識がある程度安定してきたレベル 4 の学習者については、語用論的側面での学習サポートが更に必要であると思われる。

⁸ 本研究では、学習者レベルを全 6 レベルと分けており、レベル 1 からレベル 6 の順でレベルは高くなる。今回の作文データの収集は、「90 分授業の受講経験が 28 から 30 時間程度で、450 語程度の基礎的な語彙と基本文法に対して理解が定着しており、決まり文句としての挨拶やかんたんな質疑応答ができる」ハングル検定能力試験 5 級程度以上のレベル 2 以上の学習者を対象に行った。各レベルの設定基準については林・李・曹・浅尾 (2008)、임현정(2009)を参照。



3.3 誤用分析の一例

誤用分析を韓国語教育現場で活用するための一例として、ここでは、韓国語の多義語副詞である「잘 jal」について取り上げる。韓国語副詞「잘 jal」は『ハングル検定公式ガイド・合格ドウミ』では5級の語彙として掲載されており、その意味は「①立派に、上手に、うまく、②詳しく、十分に、③無事に、④よく、しばしば、⑤正しく、⑥よろしく」などが代表的な日本語訳となっている⁹。つまり、「잘 jal」は初級レベルの語彙ではあるが、その意味が多義に至っていることから、韓国語学習者にとってはその習得が容易でないことが予想される。そこでこの仮説を検証するため、検索ツールから「잘 jal」に対応する日本語を抽出してみた。図16は「잘 jal」に対応する日本語、「よく」に対応する韓国語の用例数をそれぞれグラフで示したものである。

図16から分かるように、韓国語の「잘 jal」には「うまく、よく、元気、うまく、上手」などが対応して使われている。そのなかでも特に「よく」の対応が顕著であった。誤用例に注目してみると、「잘 jal」と「元気」、「잘 jal」と「上手」の対応では特に誤用がみられなかったのに対し、「잘 jal」と「よく」の対応において誤用が集中する傾向がみられた。一方、日本語の対訳データ¹⁰から「よく」をキーワードにして対応する韓国語の表現を抽出してみた結果、日本語の「よく」に対しては「아주aju, 좋다johta, 자주jaju, 잘jal」と対応していることが多かった。そのなかでも誤用が最も多く見られたのは日本語の「よく」を韓国語「잘 jal」に対応させたケースであった。さらに、日本語の「よく」に対応する21の誤用例について、正しい言い方に置き換えた添削後の表現をみると、「자주jaju」を用いるべき文脈に対して「잘 jal」を対応させた誤用例が最も多かった（図17）。

⁹ 李・井佐原(2006)では、世宗コーパス（韓国語の均衡コーパス）を使って韓国語の「잘 jal」のデータを収集し、クラスター分析で語義分類を行っているが、分析の結果、様態解釈、程度解釈、頻度解釈、態度解釈の4つの意味パターンが見られたことを報告している。

¹⁰ KC Corpus では、韓国語作文と学習者自身による母語訳が対訳形式で検索できる。対訳形式の検索機能を設けたのは、韓国語と日本語の双方向での検索が可能になることで、母語からの誤用の影響を探ることができる。また、強いては対照言語学的な知見のための有効な資源の提供が可能になると考えたからである。



図16 「잘(jal)」誤用分析



図17 「よく」の誤用例に対応する添削語



学習者ID	行	韓国語	日本語(品詞)	韓国語	添削判定 (添削理由)	正入/修正 の表裏	よくに 対応する 語
K0051	32	2	私もよく新刊へ買い物を 行きました。	나도 잘 신문에 쇼핑했어 요.	1. 과수	잘	
K0056	160	2	それは家でよく見る人々 のけんかや悪人たちの 仲違いの様子です。	그것은 거처에서 잘 보는 사람들 싸움이나 예민들 사 이가 좋은 모양입니다.	1. 과수	잘	
K10019	387	2	悪人でない異性同士が 胸に手を置いて抱いた いたが、異性同士で手をつな いだりという光景がよく 見られる。	전연이 아닌 이성 사이에서 도 어깨 손을 주고, 포옹을 하거나 남성과의 손을 잡아 있거나 하는 광경을 잘 볼 수 있다.	1. 과수	잘	
K10001	489	2	福岡ではよく韓国人を見 ます。	후쿠오카에서 한국 사람들 잘 봐요.	1. 과수	잘	
K10001	647	2	一方、日本人はよく韓人 的な民族だと書かれる。	한편, 일본인은 잘 소국적 인 민족이라고 한어요.	1. 과수	잘	
K30002	681	2	私は結婚前は不規則 でお酒をよく飲み、ま た文化祭のサークルに 入っていることもあり最 近は運動好きとんどしな いため、人とは比較でき ない健康であると思っ て居ないのです が、そのような私でも健 康について気を付けて いることがいくつかあり ます。	나는 취침시간은 불규칙하 고 술도 잘 마셔, 또 문화제 의 세월에 들어가 있기도 해 최근에는 운동도 거의 하지 않거 때문에, 사람과 비교했을 때에 건강하다하 고 자신을 가지고 말할 수 는 없습니다만, 그런 나 도 건강에 대해 조심하고 있는 것이 몇몇가지 있습니 다.	1. 과수	잘	
K0007	1378	2	よく寝ています。	잘 수면을 합니다	1. 과수	잘	
K10020	2767	2	韓国のドラマでは最近 でもよく見ますか?	한국의 드라마에서는 최근 도 잘 보십니까?	1. 과수	잘	
K40025	2671	5	「両眼が赤くなってた 件に、傷をつけるなんて 」という言葉をよく聞 きます。	「두눈이 빨개져 우선 눈에 상 처를 주다니」라는 말을 잘 들었습니다.	1. 과수	잘	

上記の分析結果を踏まえ、学習者には日本語の「よく」は大きく分けて「それをする/した能力・行為・作用・結果が十分であること」とことと、「頻繁に～する」という頻度を表す(市川(編), 2010)。後者の意味で「よく」を用いる場合、韓国語は「잘 jal」ではなく、「자주 jaju」を用いることを強調して指導する必要がある。また、日本語で

は副詞の位置は比較的自由であるが、韓国語では修飾する動詞や形容詞の直前に置くのが自然である（油谷, 2007）こともあわせて説明しておくことが必要であると思われる。

4. より効果的な韓国語教育のシステム構築と KC Corpus の活用

教育の最終目標は、学習したことを的確に使用でき、自ら次の学習目標が設定できる、自律した学習能力をもった自立した学習者の育成である。これは言語教育においても同様のことがいえよう。つまり、言語教育における学習者の自立とは、学習者が学習した内容（言語能力）を的確に正しく運用でき（言語使用）、自ら次の学習設計が可能になる段階のことを意味する。しかし、第2言語習得の分野では、学習者の心理的 affect¹¹が言語習得に少なからず影響を及ぼすことが報告されている（田所, 2001）。つまり、学習者自身の使用可能な言語能力がいつもどこでも発揮できるとは限らず、学習者の不安の状況によって言語活動（使用）と言語能力は異なるのである。そのため、教師は学習者の言語習得を促進させるために、学習者がどのようなところで習得の難しさを感じているのか、また、どのようなスキルが必要かを常に把握しながら、学習者の不安を少しでも取り除かせて的確に学び続けさせられるように指導していく必要がある。

一方、教育現場での学習目標やそれを評価するための学習者側の言語能力を「can do」とすれば、学習者の作文データを集めた KC Corpus は、学習者の産出的言語活動の一つである「書く」に対する「want to do」の産物である。従って、KC Corpus から得られた知見（例えば、学習者個々の習得状況や誤用のパターン）を、カリキュラムや教材開発に繰り返し反映していくことは、最終的には学習者自らが自分の学習を「計画・モニター・評価」するといった、学習者の自立を促す第2言語教育の支援につながると考える。



¹¹ Affective Filter Hypothesis(情意フィルター仮説)

学習者が持つ言語学習に対する感情的な壁が、第2言語および外国語習得に影響を及ぼすとした仮説。学習者がたとえ充分な量の理解可能なインプット (comprehensible input) を受けていても、その学習者側に不安や学習意欲の欠如、自信喪失などの心理的問題があると、それが障壁となり、インプットが学習者の中に入りにくくなる。その結果、言語習得の進展に支障が生じるとされる【第2言語習得 (SLA) 用語集: <http://www.modem.tsukuba.ac.jp/~ushiro/Publishing/SLAglossary.htm>】

5. おわりに

以上、KC Corpus の概要と韓国語教育への応用の一例を簡単に示してみた。現時点での KC Corpus の課題としては、(1)学習者の主な韓国語学習機関が限定されていること、(2)上級レベルのデータが不足している。また、(3)添削や形態素の誤りがまだ一部残っていることがあげられる。今後は、韓国語学習者作文コーパスの規模の拡充を図りながら、複数の添削者による添削を実施していきたいと考えている。またより効果的な第2言語教育への更なる応用を視野に入れ、特に現在開発中の韓国語教育スタンダード¹²との有機的関係を模索しながら、KC Corpus のより効果的な韓国語教育への活用可能性を模索していきたい。

引用文献

- 石川保子(編)(2010)『日本語誤用辞典』スリーエーネットワーク
石川慎一郎(2008)『英語コーパスと言語教育』大修館書店
李在鎬・井佐原 均(2006)「統計的手法に基づく韓国語副詞「jal」の一考察」日本言語学会 133 回大会, 304-309.
李在鎬・林炫情・浅尾仁彦・曹美庚(2009)「韓国語学習者作文コーパス (KC Corpus) について」朝鮮語教育研究会第 10 周年大会(東京大学)発表資料
李在鎬・林炫情・曹美庚・浅尾仁彦(2010)「韓国語学習者コーパス (KC Corpus) について」『朝鮮語教育—理論と実践』5, 134-137.
林炫情・李在鎬・曹美庚・浅尾仁彦(2008)「韓国語学習者コーパス構築：韓国語学習者作文コーパスにおける検索ツールの開発」『信学技報(2008-40)』電子情報通信学会, 21-26.
田所真生子(2001)「外国語学習における学習者の情意要因に関する考察」『ことばの科学』14, 303-320.
ハングル能力検定協会(2006)『「ハングル」検定公式ガイド・合格ドウミ(初・中級編)』ハングル能力検定協会
油谷幸利・金恩愛(2007)『間違いやすい韓国語表現 100 初級編』白帝社
고석주・김미옥・김재열・서상규・정희정・한송화(2004)『한국어 학습자 말뭉치와 오류분석』한국문화사
임현정(2009)『한국어 교육자료 공유지원화를 위한 말뭉치 구축: 일본어 모어화자의 작문 데이터베이스화를 중심으로』한국학 중앙연구원 해외한국학지원사업(학술연구) 최종보고서(ASK-2008-R-15)
서상규・유현경・남운진(2002)「한국어 학습자 말뭉치와 한국어교육」『한국어교육』13 (1)
국제한국어교육학회, 127-157.
『第2言語習得(SLA)用語集』: <http://www.modern.tsukuba.ac.jp/~ushiro/Publishing/SLAglossary.htm>
国際交流基金・『JF 日本語教育スタンダード』: <http://jfstandard.jp/top/ja/render.do>

¹² 韓国語教育スタンダードとは、「国際的な言語教育の枠組みである「言語のためのヨーロッパ共通参照枠：学習、教育、評価 (Common European Framework of Reference for Languages; Learning, teaching, assessment, 以下 CEFR)」の研究成果を活用し、本学における韓国語教育のさまざまな教育実践内容を明確に説明できる透明性と、教育実践の比較や連携を可能にする一貫性をもたらす韓国語教育共通参照枠(以下、韓国語教育スタンダード)」を指す。韓国語教育スタンダード開発の目的は、第1に、CEFR の共通参照レベル及び例示的能力文を参照にして、韓国語の能力記述文 (can do 形式の記述) をデータベース化すること。第2に、CEFR に対応した表現の自動抽出ができる韓国語テキストデータベースを作成すること。第3に、1と2をふまえ、「グローバル生活人」として必要な多文化・相互理解のための言語教育スタンダードの観点から本学の韓国語カリキュラムを見直し、改善することである。現在開発中の韓国語教育スタンダード(検討案)では、コミュニケーション言語能力 (communicative language competences) とコミュニケーション言語活動 (communicative language activities) の考え方や構成は CEFR に準じており、現在国際交流基金が開発・運用している『JF 日本語教育スタンダード』<http://jfstandard.jp/top/ja/render.do> を参考にしている。なお、本研究は【平成22年度・山口県立大学学内研究助成金(基盤研究(A))「多文化・相互理解のための韓国語教育スタンダードの開発」(研究代表者：林炫情)】の補助を受けた。

韓国語学習者作文コーパス (KC Corpus) と

韓国語教育への活用

KC Corpus (Korean L2 Learner's Written Composition Corpus) and its application in education

林 炫情(山口県立大学)

Hyunjung, Lim

(Yamaguchi Prefectural University, Japan)

李 在鎬 (国際交流基金日本語試験センター)

Jae-ho, Lee

(The Japan Foundation, Center for Japanese-Language Testing, Japan)

黄 晟媛(立命館アジア太平洋大学)

Jung-nan, Hwang

(Ritsumeikan Asia Pacific University, Japan)

浅尾 仁彦(SUNY Buffalo/京都大学)

Yoshihiko, Asao

(Doctoral Student, Grad.Sch. SUNY Buffalo, USA/ Kyoto University, Japan)

요 지

이 논문에서는 한국어학습자 작문 코퍼스(corpus)인 「KC Corpus (Korean L2 Learner's Written Composition Corpus)」의 개요와 교육 현장에서의 활용의 예를 제시하였다. 그리고 더욱 효과적인 한국어교육의 시스템을 구축하고 효율적으로 활용하기 위한 KC Corpus의 가능성에 대해 검토를 하였다. KC Corpus에서는 한국어학습자가 산출한 언어사용(작문)의 문제점을 객관적 데이터를 바탕으로 분석할 수 있으며, 분석 결과를 쉽게 교육 현장에 피드백할 수 있다는 점에서 한국어교육을 위한 유용한 도구라고 생각한다. 현시점에서의 KC Corpus의 과제로서는 (1) 학습자의 주된 한국어학습 기관이 한정된 점, (2) 상급 수준의 데이터가 부족한 점, 그리고 (3) 첨삭이나 형태소분석의 오류가 아직 일부분 남아 있는 것을 들 수 있다. 앞으로는 한국어학습자 작문 코퍼스(corpus) 규모의 확충을 도모하면서 복수의 첨삭자를 통한 첨삭을 시행해 갈 예정이다. 또한, 더 효과적인 제2언어교육, 특히 한국어교육에의 응용을 시야에 넣어, 현재 개발 중인 한국어교육 기준(can do)과의 유기적 관계를 모색하면서, 한국어교육에의 활용 가능성을 모색해 가고자 한다.

키워드 : 한국어학습자, KC Corpus, 오용분석, 한국어교육에의 활용