



## 韓国語学習者作文コーパス (KC Corpus) について

- 李 在鎬 (国際交流基金)
- 林 炫情 (山口県立大学)
- 浅尾仁彦 (京都大学[院])
- 曹 美庚 (九州大学大学院)

## お品書き

1. はじめに：背景と現状など
2. KC Corpusの概要
3. 全体の設計と開発手順
4. 利用例
5. 統計情報
6. 今後の課題



## はじめに：背景 1/4

- 日本国内の韓国語学習に対するニーズの多様化：質の高い教育へのニーズ  
→習熟度に合わせてコンテンツの開発
- 習得の実態が把握できる資料が必要：学習者のoutputをデータベース化する  
→学習者コーパスの開発

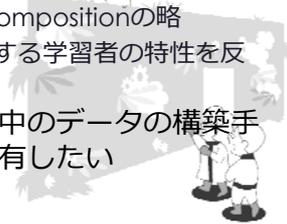
## はじめに：現状 2/4

- 韓国語の学習者コーパスをめぐる現状
  - 연세 학습자 말뭉치 (26万語節/作文)  
→ 詳細は서상규(他)2002及び김민애2006参照  
corpusの有用性については共有できている現状
- ユニバーサルサービスとして共有
  - 要件(ユニバーサルサービスの4条件)
    1. どこでも - 地理的
    2. だれでも - 社会的
    3. 負担可能な料金の - 経済的
    4. 均一なサービス - 技術的



## はじめに：目的 3/4

- 共有化を前提にした、使いやすい学習者コーパスを作成→KC Corpus
  - ウェブ上で利用できる日本語母語話者の作文コーパス。KC CorpusとはKorean L2 Learners' written Compositionの略
  - 海外で韓国語を学習する学習者の特性を反映したデータ
- 本発表では現在公開中のデータの構築手順を示し、成果を共有したい



## はじめに：狙い 4/4

- 本研究が想定する利用例→基礎資料として
  1. 授業で活用：カリキュラムの開発および教授法の開発のために利用
  2. コンテンツ開発で活用：学習者用の辞書作成や教科書の開発のために利用
  3. テスト開発：言語能力測定のために利用



韓国語教育の活性化に貢献

## KCCorpusの概要 1/5

### • 四つの方針

1. KCCorpusは韓国語の学習者コーパスです。
2. KCCorpusは学習者の作文データベースです。
3. KCCorpusは誰でも利用できます。
4. KCCorpusはどこからでも利用できます。



## KCCorpusの概要 2/5

### • 四つの方針

1. KCCorpusは韓国語の学習者コーパスです。
  - 日本語母語話者による韓国語の学習者データが収録。
  - タグ付きコーパスですので、言語情報を元に様々な検索ができる。
2. KCCorpusは学習者の作文データベースです。
3. KCCorpusは誰でも利用できます。
4. KCCorpusはどこからでも利用できます。

## KCCorpusの概要 3/5

### • 四つの方針

1. KCCorpusは韓国語の学習者コーパスです。
2. KCCorpusは学習者の作文データベースです。
  - 学習者の作文データが収録されています。
  - 作文データには教師による添削情報や誤用情報が付与されている。
3. KCCorpusは誰でも利用できます。
4. KCCorpusはどこからでも利用できます。

## KCCorpusの概要 4/5

### • 四つの方針

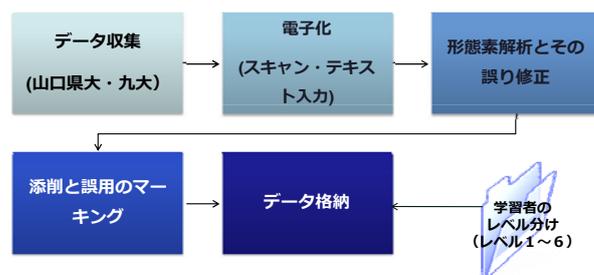
1. KCCorpusは韓国語の学習者コーパスです。
2. KCCorpusは学習者の作文データベースです。
3. KCCorpusは誰でも利用できます。
  - 韓国語教育関係者が共有資源化することを目的として作成。利用上の制限は一切設けていない。
  - パソコンスキルがなくても、簡単に使うことができる。
4. KCCorpusはどこからでも利用できます。

## KCCorpusの概要 5/5

### • 四つの方針

1. KCCorpusは韓国語の学習者コーパスです。
2. KCCorpusは学習者の作文データベースです。
3. KCCorpusは誰でも利用できます。
4. KCCorpusはどこからでも利用できます。
  - ウェブ画面上で検索をし、利用することができる。
  - 検索結果を保存することもできる。

## 構築手順 1/5



## 構築手順 2/5

### • データ収集

1. 韓国語の作文：教室内でテーマを指定して書かせる。
2. 日本語訳：作文意図を把握するために本人に書いてもらう。
3. 自己申告による学習履歴
4. 公開に関する確認書：公開を前提にした資料収集のため、最初から権利についての確認書を提出してもらう

## 構築手順 3/5

### • 形態素解析

1. 지능형 형태소분석기 (<http://www.sejong.or.kr/>) を使用。
2. 21세기 세종계획 말뭉치의タグセットと互換性あり。

### • 添削・誤用タグ

1. 語節単位に付与（形態素単位の付与から変更）
2. 添削者間で揺れが少ないレベルで決定
  - 文法ミス、表記ミス、文体ミス

## 構築手順 4/5

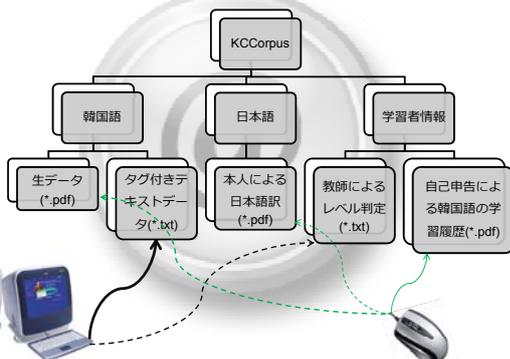
1. 연세말뭉치(고석주 他, 2004)を参考にタグセットを検討
2. 判定の客観性を保つため、三つの大きなカテゴリーに分類
  - 表記ミス：表記レベルの誤用。
  - 文法ミス：活用や文法要素の誤用
  - 文体ミス：表現や語用論的側面の誤用
3. 複数化(文法ミスでかつ文体的ミス)を認める

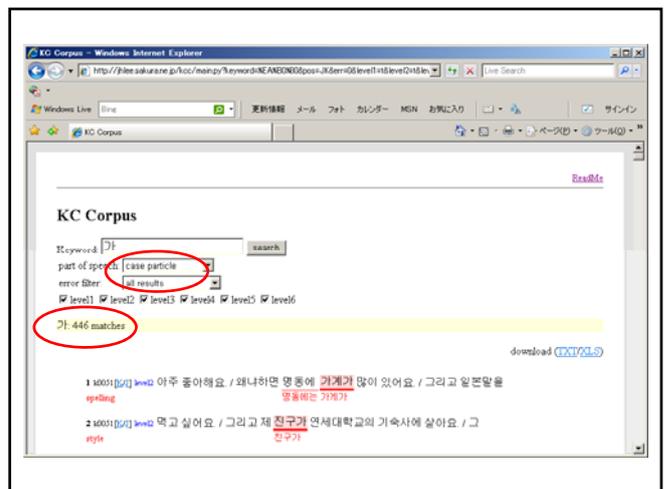
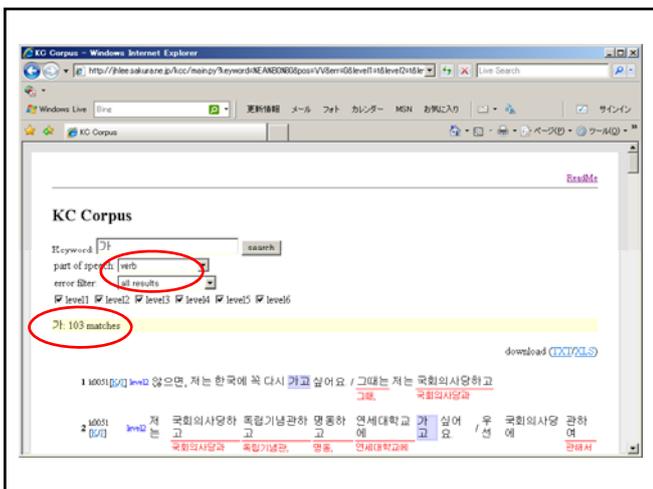
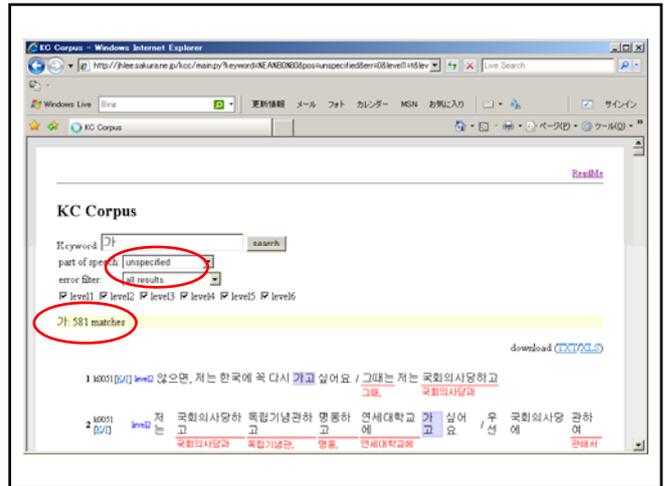
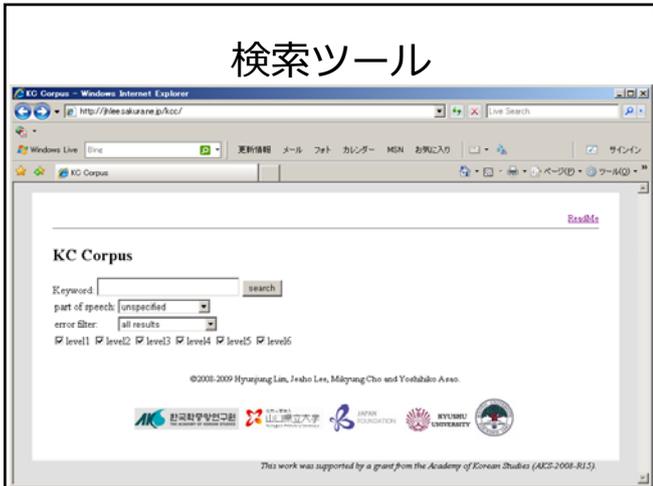
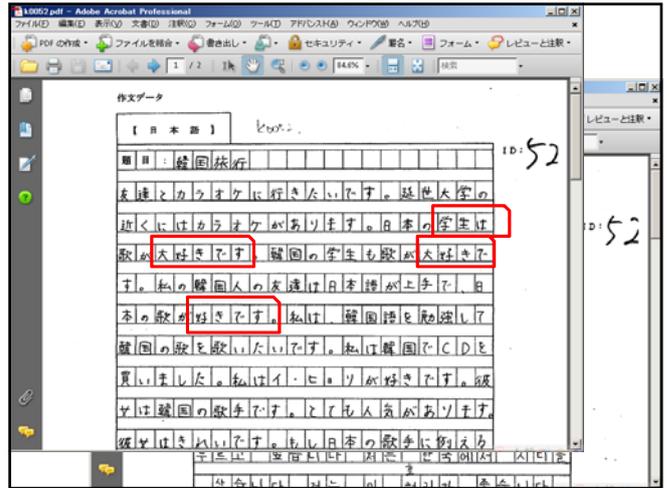
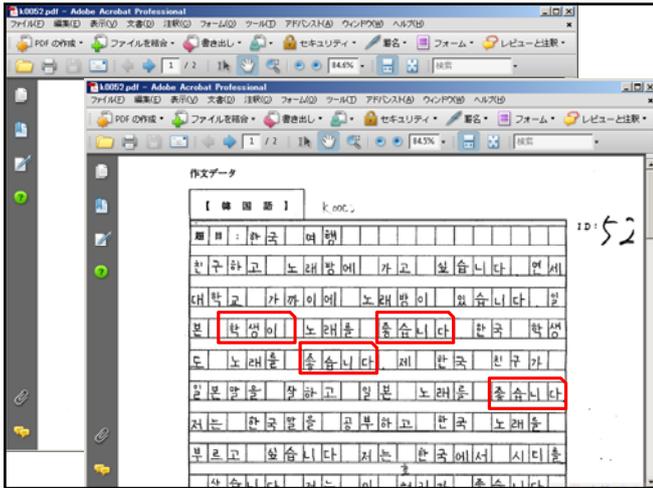
## 構築手順 5/5

### • レベル分け

- 6レベル；1(最下位級)→6(最上位級)
- レベル分けの根拠
  1. 韓国語関連の検定試験の合格有無や合格した級
  2. 作文作成時の学習歴
  3. 留学の有無と機関
 詳細は<http://www12.atwiki.jp/kccorpus/pages/18.html>

## 全体設計

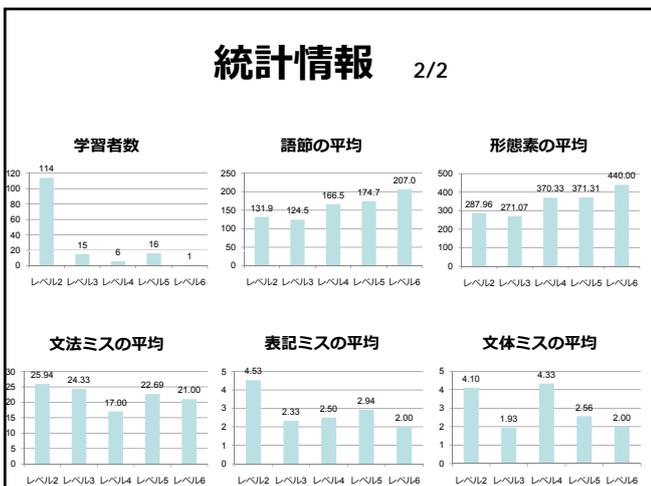




file	level	left context	keyword	right context	error	type
k0052	2	/ 저는 한식을 좋아합니다. [좋아합니다]	좋고기(좋고기들)	좋습니다. [좋아합니다] / 한식이 [한식]입니다.	grammar	
k0053	2	견학하면 [견학한다면] 저는 한국하고 [한국과] 일본 양쪽에서 [양쪽의]	역사(역사들)	알습니다. [알 수 있습니다] / 인식의 [인식]입니다. [알 수 있습니다].	grammar	
k0057	5	맛입니다. / 뷔스를 타기 후도 [한 때]	비스(비스의)	말려서 [속도가 빨라서] 너무 흔들기(흔들려서) 두 손으로 꼭	grammar	
k0058	2	일화를 문화가 있습니다. / 저도	주머니(비닐봉지)	우울(나) 몇번이나 재미(재미들)	grammar	
k1003b	2	두 개의 [4] 문화로 [문화에서] 학생들	차이(차이들)	중요(중요)할 수 있다. // 한국의	grammar	
k1005	2	면에서 [속에서] 비슷한 [비슷한] 점	차이가	나는 점도 알(알지) 있다. /	grammar	
k10013	2	/ 그러나 [그래서] 일본의 식(식)에	고려(고려)가	장하고 [알려] 있습니다. 한국의	grammar	
k10015	2	한국의 문화하고 [문화와] 일본의 문화의 [문화] 중(중)에서	사(사)에(같은) 차이(차이)	재미(재미)입니다. // 일본에는 것이	grammar	

### 統計情報 1/2

区分	頻度
学習者総数	152
語節のトークン頻度	20905
形態素のトークン頻度	45496
文法ミス (語節)	3808
表記ミス (語節)	615
文体ミス (語節)	565



- ### 今後の課題
- 問題点
    1. 上位級のデータが不足→バランス的に良くない
    2. 添削や形態素の誤りが一部残っている
    3. 大学が限定されている
  - 解決に向けての方針
    1. 韓国語学習者作文コーパスの規模の拡充
    2. データの誤用タグの誤りの修正と細分化
    3. 複数の添削者による評価
    4. KC Corpusの検索機能を改良・追加
      - 絞り込み検索機能を追加,日本語PDFと検索結果の関連づけを検討 (対照言語学への応用を模索)

## 参考文献

- [1] 고석주·김미옥·김제열·서상규·정희정·한송화(2004), 한국어 학습자 말뭉치와 오류분석, 한국문화사.
- [2] 서상규·유현경·남윤진(2002), 한국어 학습자 말뭉치와 한국어 교육, 한국어 교육13-1, 국제한국어교육학회. 127-157.
- [3] 김민애(2006), 한국어 학습자 오류의 분석 방법 고찰, 한국어 교육17-2, 국제한국어교육학회. 21-43.
- [4] 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.
- [5] 浅尾仁彦・李在鎬 (2008), 日本語学習者コーパス検索ツールの開発, 言語科学会第10回年次大会(静岡県立大学)大会論文集, p.182
- [6] 李在鎬(2009印刷中), タグ付き学習者コーパスの開発について, 計量国語学, 27-2



### 以下の外部研究資金を得て開発しました。

- Korean Studies Grant 2008(AKS-2008-R15) (2008年度)
  - 研究課題名: 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を中心に
  - 研究代表者: 林 炫情
- 山口県立大学学内研究助成金 (基盤研究 (A)) (2009年度)
  - 研究課題名: 韓国語学習者作文コーパス (KC Corpus) と韓国語教育への活用
  - 研究代表者: 林 炫情

