

韓国語学習者コーパスの構築 -韓国語学習者作文コーパスにおける検索ツールの開発-

林 炫情 (Hyunjung Lim) 山口県立大学
李 在鎬 (Jae-ho Lee) 情報通信研究機構
曹 美庚 (Mikyung Cho) 九州大学
浅尾 仁彦 (Yoshihiko Asao) 京都大学

1

研究背景

- 近年、日本における韓国語学習者のニーズは一層多様化されており、韓国語教員に対してはより質の高い教育、学習者のニーズや学習熟度に合わせた教育コンテンツの開発、効果的な教授法の支援などが求められている。

2

韓国語学習者コーパスの教育上の活用

- カリキュラムの開発および教授法の開発に関する基礎データとしての利用
- 学習者辞書の作成および韓国語学習者のための教科書の開発
- 学習者の言語能力テスト開発の基礎データとしての活用

3

韓国語学習者コーパスの問題点

- 韓国語における学習者コーパスは、いくつか構築されているものの、一般公開されているものは極めて少ない。
- 誤用分析のための検索ツールの使用方法が容易でない。
- 現存する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ため、日本で、韓国語を学習する学習者の誤用要因を把握するには限界がある。

4

本研究の目的

- 日本の大学での教室学習を主とする学習者の作文データを、電子データ化・コーパス化。
- 言語情報(形態素解析情報・誤用タグ・添削情報)を付与したコーパスの提供。
- 言語情報を容易に検索し、加工・分析できる検索ツール(「Korean E-KWIC (仮)」「Korean E-KWIC Web (仮)」)の開発。

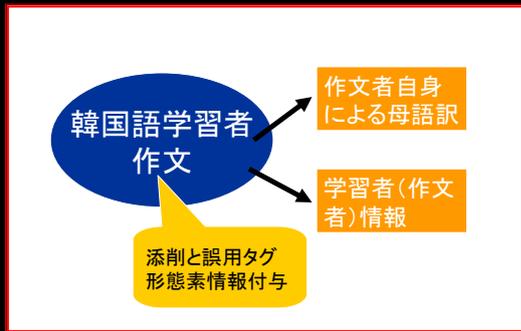
成果物は無償で公開し、韓国語学習者データに関する資源の共有化をはかる。

「韓国語学習者作文コーパス」の構築状況

- データベースの全体像
- データベースの規模
- コーパス構築の手順
- 検索ツール開発概要
- 配布方法について

6

データベースの全体像



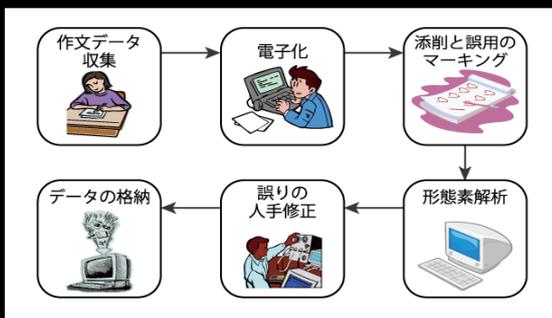
7

データベースの規模

- 現在までに、山口県立大学と九州大学の韓国語学習者96名分のデータを収集済み。
- 最終的には約200名分のデータの収録を目標。
- 将来的には他機関への協力を要請し、規模を拡大させていく予定。

8

コーパス構築の手順



9



作文データの収集

データ	内容	
韓国語作文	①日本の行事や流行について ②日本と韓国の文化の比較 ③私が健康のためにやっていること ④整形手術についてのあなたの考えは？	作文作成は授業のなかで実施。制限時間は1時間。
作文者自身による母語訳	日本語訳	
作文者情報	①生年月日②性別③韓国語学習歴④主な学習場所⑤韓国語検定試験資格の有無⑥韓国で1か月以上生活した経験の有無⑦家で韓国語を話す人の有無⑧辞書使用の有無	

学習者のレベル分け

レベル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、留学の有無と期間等を総合して下記の6段階レベルで判定した。

レベル	目安
レベル1	90分授業の受講経験が30時間未満。約450語の基礎的な語彙や基本文法について学習経験をもつ。辞書を引きながら簡単な文を作ることができる。ハングル検定試験5級以下。
レベル2	90分授業の受講経験が28~30時間程度。450語程度の基礎的な語彙と基本文法に対する理解が定着しており、決まり文句としての挨拶や簡単な質疑応答ができる。ハングル検定試験5級程度。
レベル3	90分授業の受講経験が56~60時間程度。自己紹介・買い物、飲食店での注文など生活に必要な基礎的な言語を駆使でき、身近な話題の内容を理解、表現できる。950語程度の基礎的な語彙と基本文法が理解でき、簡単な文章が作れる。ハングル検定4級・韓国語能力試験初級程度。

学習者のレベル分け

レベル	目安
レベル4	90分授業の受講経験が70~120時間程度。電話やお願程度度の日常生活に必要な言語や、郵便局、銀行などの公共機関での会話ができる。1500~2000語程度の語彙を用いた文章が理解でき使用できる。ハングル検定3級・韓国語能力試験中級Ⅰ(3級)程度。
レベル5	90分授業の受講経験が160~200時間程度。日常生活に支障がなく、様々な公共施設の利用や社会的関係を維持するための言語使用が可能。文章語と口語の基本的な特性が理解でき、使用が可能である。ハングル検定準2級・韓国語能力試験中級Ⅱ(4級)程度。
レベル6	社会的常識の範囲内にある話題が殆ど理解できる。また、新聞の社説などを読んで理解できる。韓国語で論理的な文が書け、話ができる。ハングル検定2級・韓国語能力試験上級(5級・6級)以上。

検索ツール「Korean E-KWIC」

- Koren E-KWICはエクセルファイルになっており、マクロを実行するだけで利用できる。

人文系の研究者も日常的に使用している場合が多く、実行にあたっては障壁が少ない

- 検索結果はエクセルファイルとして出力できる。

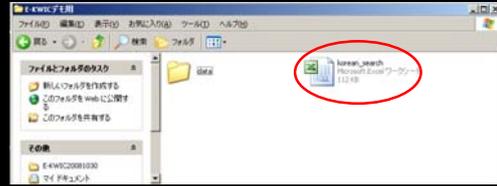
検索結果をさらに加工・分析することが容易

- 指定された検索語に対して、KWIC形式で学習者の用例を表示することができる。
- 誤用の場合は、誤用の種別(文法/文体/文字)に加え、添削結果も表示される。

19

「Korean E-KWIC」利用手順

1. データフォルダ内のkorean_search.xlsをクリック

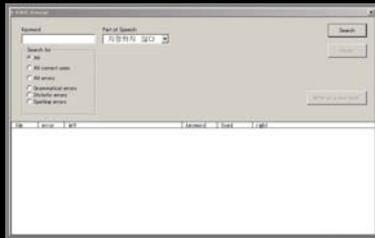


2. マクロを実行するかどうか聞いてくるが、「マクロを有効にする」をクリック

20

「Korean E-KWIC」利用手順

3. 検索用のウィンドウが開く



21

「Korean E-KWIC」利用手順

4. 検索語と品詞を指定し、フォルダ内のすべてのファイル内を検索することができる。また、検索語を指定せず、品詞のみで検索することも可能。

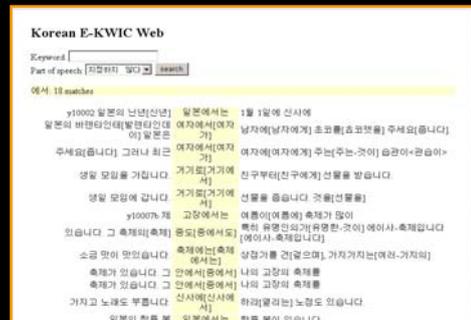


22

「Korean E-KWIC」実演

23

ウェブ版Korean E-KWIC



24

今後の検索ツール開発

- 現段階では、キーワードと品詞のみによる検索に対応しているが、将来的には**学習者レベルの絞り込み機能**や**誤用例に対するオプション**なども追加していく予定。

25

配布について

- 本プロジェクトは、韓国学中央研究院の研究助成によって企画・進行しており、韓国語学習者のデータの共有資源化を目的としているため、**完全フリーで公開する予定**である。
- データの提供者(韓国語学習者)に対してはデータ収集の段階から、本プロジェクトの趣旨を説明した上、公開に関する同意を書面でもらっており、**著作権に関する問題は完全にクリア**できている。

26

今後の計画

- 2009年4月に公開予定
- 一般公開に向けて作業を進めていくほかに、他の研究者との連携をより深めながら韓国語学習者コーパスを拡充し、実用性を高める。
- 日韓両言語の対照研究といった応用分野での利用に関しても普及活動を進める。

27

References

- 浅尾仁彦・李在鎭 (2008) 「日本語学習者コーパス検索ツールの開発」、言語学会第10回年次大会(静岡県立大学)『大会論文集』, p.182
- 石川慎一郎(2008)『英語コーパスと言語教育』大修館書店
- 李在鎭・浅尾仁彦・濱野寛子・佐野香織・井佐原均(2008)「タグ付き日本語学習者コーパスの開発」、2008年度自然言語処理学会全国大会(東京大学)『大会論文集』, pp.658-661.
- 佐渡島沙織 (2001) 「作文コーパス研究における対訳の有効性: 対訳が添削に及ぼす影響」『日本語教育のためのアジア諸言語の対訳作文データの収集とコーパスの構築』国立国語研究所
- Cho, N.H., Lee, U.K., Kim, J.M., & Lee, S.U. 2005 *Korean learner's corpus of construction and use*. Taehaksa. (韓国語)
- Kim, M.A., 2006 A study on the analysis method of KFL learners' error. *Korean Education*, 17(2). 21-43. (韓国語)
- Seo, S.K., Yoo, H.H., & Nam, Y.J. 2002 Korean learner's corpus and Korean education. *Korean Education*, 13(1). 127-157. (韓国語)
- You, S.H., 2001 The need of a learners' corpus for the learners of Korean as a foreign language and the corpus analysis. *Korean Education*, 12(1).165-180. (韓国語)

28

本研究は、2008年度韓国学中央研究院の支援を受けて行っているものである(課題番号AKS-2008-R15)。

29

以下、未使用
(参照まで)

30

韓国語教育関連のコーパス

Seo, Yoo, & Nam (2002)

「標準コーパス」

- ・標準コーパス (standard corpus) :
韓国語母語話者が産出した言語資料
- ・関連コーパス (reference corpus) :
韓国語教材・テキストなどの言語資料

「誤用コーパス」

- ・学習者コーパス (learner's corpus/ error corpus) : 学習者が産出した言語資料

31

韓国語教育関連の「標準コーパス」

「延世韓国語コーパス」

- ・1999年以降、延世大学の言語情報開発研究院で構築したコーパス。
- ・標準コーパス: 100万語節 (文節) (2002年現在)
- ・関連コーパス: 82万語節 (2002年現在)

32

韓国語教育関連の「誤用コーパス」

(1)「延世学習者コーパス」

- ・対象: 韓国語学習者の作文等
(主に、延世大学の語学堂の学生作文)
- ・規模: 26万語節

(2)「Cho, ect., 学習者コーパス」

- ・対象: 韓国語学習者の作文
- ・規模: 50万語節

33

Sejong コーパス



21世紀Sejong計画(韓国語情報化事業(1998-2007)の一環として構築されたコーパス。規模は、2億語節(うち、現代韓国語は9,316万語節、形態素分析は1,230万語節、北朝鮮資料が1,083万語節)。