

## 韓国語文章リーダビリティ判定システム「kReadability」

浅尾 仁彦(情報通信研究機構)

林 炫情(山口県立大学)・李 在鎬(早稲田大学)

須賀井 義教(近畿大学)・斉藤 信浩(九州大学)

### 1. はじめに

本稿では、韓国語の文章難易度を自動判定する韓国語文章リーダビリティ判定システム「kReadability」の概要と構築状況について述べる。

このシステムは、近年、社会的ニーズが高まっている内容言語統合型学習(Content and Language Integrated Learning; CLIL)を柱とする韓国語教育の実践と拡充を目指して、韓国語学習者が自分の語学レベルに応じて適切な学習用教材をアダプティブに利用するためのアプリケーションとして実装することを目的としている。CLILでは真正性の高い新聞、雑誌、ウェブサイトなどの生素材が推奨されるが、このような生素材は学習者にとっての難易度が明らかでないため、どのような文章を、どのレベルの学習者に与えるべきかという問題が生じる。この問題に対処するために、韓国語に特化したリーダビリティ測定尺度を考案し、ウェブ上で誰でも利用できる韓国語文章リーダビリティ判定システム「kReadability」を開発した。今後、この「kReadability」が韓国語の読解教育支援に貢献を果たすことが期待される。

### 2. 研究の背景

近年、内容言語統合型学習(CLIL)の重要性が盛んに指摘されており、その利点として、学術的な内容を扱いながら、同時に外国語スキルを伸ばすことができること、文化理解を深めつつ批判的思考力とリサーチスキルを養えること、学習意欲の高まりによって、学習の達成度をより強く味わえることなどが挙げられている。このような内容言語統合型学習のためにはどのような教材を用いるかも重要であるが、既存の韓国語関連の市販教科書は、固定的な知識を定着させ、現実にはあり得ないような特化した状況での言語能力や付随的内容をいかに学習させるかという視点で作られたものが多い。こうした教材の問題点を克服するために、非母語話者向けの言語学習用に作られた教材ではなく、母語話者対象に書かれたテキストを活用し、

現実のコミュニケーションの場で役立つ韓国語の能力の実態をリアルに学習者に伝えることが強く求められている。一方で、このような新聞、雑誌、ウェブサイトなどのオーセンティックな生教材（以下、生教材）は、学習者にとっての難易度が明らかでないため、どのような文章を、どのレベルの学習者に与えるべきかという問題が生じる。この問題に対して、現状では教師の個人的経験や直感を頼りに学習者に提供することが多く、チームティーチングのような一貫性を必要とするカリキュラムにおいては教育的効果や妥当性が検証できない場合が多々ある。そうした統制しにくい指標によって教育上の教材が選別されることは、学習内容や評価に差が生じてしまうため好ましくない。

そこで、新聞記事、案内文やパンフレット、最近話題の本、ウェブ上のブログなど生教材を使用するに当たり、リーダビリティ測定尺度を活用し、文章難易度を自動判定するウェブシステムを開発することは、学習者のレベルに見合った良質の教材を提供するだけではなく、教員の教材準備の手間や生教材の選別における不安要素の軽減にも役立つ。つまり、kReadability の実用化は韓国語の生教材選定において教師・学習者双方の支援に直結するものであり、強いては内容言語統合型学習の社会的ニーズへの対応としても期待できる。

## 2.1. 内容言語統合型学習 (Content and Language Integrated Learning; CLIL)

グローバル化が加速するなか、現在の言語教育には、流暢で円滑なコミュニケーションを目指す「言語訓練」や「コミュニケーション・スキル」の育成から一歩進んで、言語「教育」の「内容」が求められている（林 2018a,b,2019a,b; 李ほか 2019）。近年、ヨーロッパや北米では、教科内容と外国語学習を統合し、質の高い外国語教育の実現を目指す「内容言語統合型学習」(Content and Language Integrated Learning; CLIL) が急速に広まっている。CLIL は外国語で表された内容を外国語教育の手法を用いて学習し、外国語を「学習手段として使うことによって実践力を伸ばす」ことを特徴としている（池田 2011）。単なる教養としての外国語学習ではなく、使えることばの学習に重点をおいた言語教育は、グローバル社会に適応した知識、態度や技能を効果的に提供するための学習形態に対応したアプローチであるといえる。

日本でも、英語教育や日本語教育の分野ではすでに CLIL 理論の検討と実践的導入が始まっており、それをサポートするためのツールも作成されている（英語では、渡辺ほか編 2011; 和泉ほか編 2012; [http://someya-net.com/wlc/index\\_J.html](http://someya-net.com/wlc/index_J.html)、日本語では、佐藤ほか編 2015; 奥野 2018; <http://language.tiu.ac.jp/>;

<https://jreadability.net> など)。日本語では、読解学習支援システムとして「リーディング チュウ太」(<https://chuta.cegloc.tsukuba.ac.jp/>) (図 1) のように、文章を入力すれば、検定級の判定や品詞解析などを行ってくれるツールが公開されており、このシステムを利用した応用実践例の報告もなされている (川村 2009; 栗野ほか 2012)。



図 1 日本語の読解学習支援システム「リーディング チュウ太」

ところが、韓国語教育では英語教育や日本語教育に比べてリーダビリティ測定尺度を活用した読解教育支援システムの開発とその実用性の研究が立ち後れており、CLIL の応用実践例の報告もまだ少ない (林 2018a)。

## 2.2. リーダビリティ

リーダビリティ公式は、母語話者や移民児童の英語識字率を高めるために、入力された英文テキストに対し一つの実数値を返し、その英文テキストの読みやすさや読みにくさを客観的に示すために開発された (野本 2016)。リーダビリティは、リーディング教材の選択や、文章作成の指標として用いられるほか (柴崎 2014)、入試問題の作成と品質管理への応用可能性などが指摘されている (染谷 2010)。

最近では、第二言語として英語や日本語を学ぶ学習者のためのリーダビリティ公式も考案されている (李 2011; 李・柴崎 2012)。日本語リーダビリティ公式に基づいて文章の難易度を測定するツールとしては jReadability というシステムが公開されており (<https://jreadability.net>; Hasebe and Lee 2015)、その応用実践例の報告も数多くなされている (安井ほか 2015; 鎌田 2016; 森川 2017; 横山

2019)<sup>1</sup>。しかしながら、韓国語教育ではリーダビリティ測定尺度を活用した読解教育支援システムの開発とその実用性の研究が立ち後れており、韓国での研究においてもまだ十分な実用化の成果が得られたものは見当たらない。このような国内外の研究動向を踏まえ、本研究では韓国語読解教育のための文章難易度自動判定システムを開発し、それを教育現場で共有できるウェブシステムとして実用化することを目指す。

### 3. kReadability の構築

kReadability は 2019 年度から開発をスタートしており、現在は韓国語リーダビリティ公式の試案作成を終え、プロトタイプシステムの検証段階にある。本節では、kReadability の開発手順、そして、韓国語リーダビリティ公式の試案採用までの作業手順について紹介する。

#### 3.1. kReadability の基本方針

kReadability 開発にあたっては、具体的に次の 4 つの基本方針を立てた。第一に、韓国語に特化したリーダビリティ測定ツールを実用化すること、第二に、教師・学習者双方の支援を考慮したシステムであること、第三に、誰でも利用できることを前提とした使いやすいシステムであること、第四に、どこからでも利用できるようウェブブラウザ上で利用できるユーザフレンドリーな検索環境であることの 4 つである。

#### 3.2. データ収集と開発手順

韓国語リーダビリティシステム開発は、3 セクション、11 工程で行った。図 2 は、kReadability の開発手順を示したものである。第 1 セクション (①から③) は「基準テキストデータの作成」、第 2 セクション (④から⑦) は「文章の読みやすさを測定するためのリーダビリティ公式開発段階」、第 3 セクション (⑧から⑪) は「ウェブシステムへ組み込み」の過程となる。第 1 セクションの「基準テキストデータ作成」は日韓で市販されている韓国語テキストと韓国延世大学言語情報院で構築した「한국어 교재 말뭉치 (韓国語教材コーパス)」に収録されているテキスト (初級 11 冊、中級 10 冊、上級 5 冊) を用いて、初級、中級、上級の 3 レベル別に作成した。

---

<sup>1</sup> 本研究で開発する kReadability は、jReadability を参考にしつつ、独自にシステム開発を行うものであり、日本語の対訳を表示するなど、日本語話者向けの機能を充実させることを意図している。

なお、レベル分けは教材の指定レベルに従って判断した。

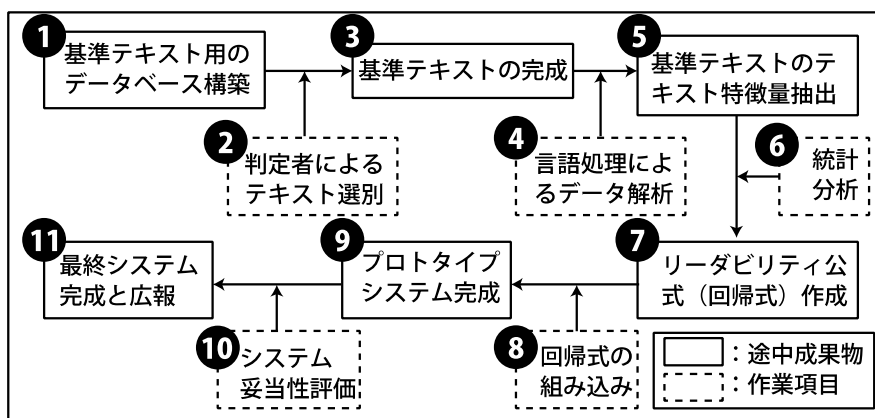


図2 「kReadability」の開発手順

### 3.3. リーダビリティ公式

この節では第2セクションの「リーダビリティ公式開発段階」について説明する。まず、第1セクションで収集したテキストを形態素解析ツール MeCab (Kudo et al. 2004) および韓国語形態素解析用辞書 HanDic (須賀井 2013) を利用して形態素単位に分割した<sup>2</sup>。次に、形態素解析結果を用いて各テキストを 1,000 形態素ずつに分割し、それぞれを文章 1 件とみなして統計分析の対象とした。文章の内訳は初級が 104 件、中級が 255 件、上級が 320 件である。次に、統計分析で用いる各文章の特徴量を求めた。用いた特徴量は以下の通りである。

連体と終止の差  $(\text{終止形の数} - \text{連体形の数}) / (\text{終止形の数} + \text{連体形の数})$

連体の頻度  $(\text{連体形の数} / \text{総形態素数}) * 10$

名詞率  $\text{名詞の数} / \text{総形態素数}$

動詞率  $\text{動詞の数} / \text{総形態素数}$

助詞率  $\text{助詞の数} / \text{総形態素数}$

派生動詞率  $\text{動詞的接尾辞の数} / \text{総形態素数}$

派生形容詞率  $\text{形容詞的接尾辞の数} / \text{総形態素数}$

<sup>2</sup> MeCab は工藤拓氏によって開発された高速な形態素解析エンジンであり、ウェブサイト <https://taku910.github.io/mecab/> から入手可能である。また、HanDic は MeCab 上で利用可能な韓国語辞書であり、以下の URL で公開されている。  
<https://porocise.sakura.ne.jp/wiki/korean/mecab>

派生述語率  $(\text{動詞的接尾辞の数} + \text{形容詞的接尾辞の数}) / \text{総形態素数}$

平均文長  $\text{総形態素数} / \text{文数}$

語の長さ  $\text{字数} / \text{総形態素数}$

特徴量としては文章のレベルを予測するうえで有用性が高いと思われるものを選択している。例えば、「連体と終止の差」および「連体の頻度」は、レベルが上がるにつれて文が複雑になり、修飾語や関係節が増加することで、総形態素数に占める連体形の割合が上昇するであろうという予想に基づいている。実際、「連体と終止の差」を例にとって、初級・中級・上級のそれぞれのテキストごとに分布を図示すると図 3 のようになり、レベルが上がるにつれて終止形の割合が低下し、連体形の割合が上昇することが読み取れる。

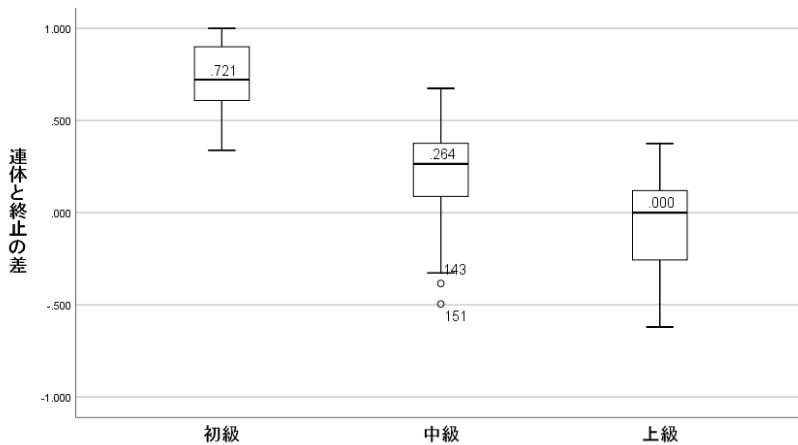


図 3 レベルごとの「連体と終止の差」の分布

次に、判別分析を用いて、各文章のレベルと判別分析の結果が一致するかどうかを調べ、一致しないものはリーダビリティ公式の計算から除外することとした。この結果、リーダビリティ公式作成のために使用するテキストは初級が 98 件、中級が 146 件、上級が 249 件となった。

リーダビリティ公式の開発にあたっては、初級を 1、中級を 2、上級を 3 という整数値に変換し、上で述べた特徴量を説明変数として用いて、文章のレベルを予測する重回帰分析モデルをフィッティングした。モデルが予測する値がリーダビリティ値となる。

モデルは、できるだけ少ない変数かつ高い精度で予測できるものを選択した。モ

デルの精度比較の結果を表 1 に示す。モデル 1 からモデル 7 へと変数を追加するごとに予測精度が向上していることが読み取れるが、モデル 6 以降は変数の増加に対してモデルの予測精度の上昇がわずかであるため、モデル 5 を採用することとした。

表 1 重回帰分析のモデルの比較

モデル ID	詳細	R <sup>2</sup> 値 (予測精度)
1	定数、連体と終止の差	0.628
2	定数、連体と終止の差、連体の頻度	0.692
3	定数、連体と終止の差、連体の頻度、名詞率	0.787
4	定数、連体と終止の差、連体の頻度、名詞率、 派生述語率	0.825
5	定数、連体と終止の差、連体の頻度、名詞率、 派生述語率、平均文長	0.840
6	定数、連体と終止の差、連体の頻度、名詞率、 派生述語率、平均文長、語の長さ	0.843
7	定数、連体と終止の差、連体の頻度、名詞率、 派生述語率、平均文長、語の長さ、助詞率	0.847

なお、リーダビリティ公式に含まれる変数のうち「連体と終止の差」と「連体の頻度」は相関を示す。一般に回帰分析ではこのように説明変数同士に共線性がある状況は避けられるが、リーダビリティ分析においては共線性があることは有害でない(柴崎・原 2010)。実際、「連体と終止の差」と「連体の頻度」のどちらかを削除して分析を行うと、予測精度が低下した。

最終的に得られたリーダビリティ公式の各係数は表 2 の通りである。

表 2 採用したリーダビリティ公式における各係数

変数	係数
定数	7.570
連体と終止の差	-5.178
連体の頻度	-5.970
名詞率	-7.637
派生述語率	1.831
平均文長	-0.027

### 3.4. ウェブ判定システム

この節では第 3 セクションの「ウェブシステムへの組み込み」について説明する。このセクションは、既に述べたように、kReadability を誰でも・どこからでも利用できるようにすることで、生教材選定にあたっての教員負担の軽減や学習者支援に寄与することを目的としている。ウェブ判定システムは現在、プロトタイプシステムが作成された段階であり、今後、機能の拡充や動作の検証を行い、一般公開される予定である<sup>3</sup>。

現段階のプロトタイプシステムでは図 4 のようなインターフェースを用意している。



図 4 開発中の kReadability 画面の例 (入力画面)

ユーザがテキストボックスに韓国語文章を入力 (コピー&ペースト) して「分析する」をクリックすると、図 5 のように分析結果が表示される。まず、リーダビリティ値とそれに基づいたレベル分類が表示され、また、リーダビリティ値計算の根拠となった変数の値が表示される。この例では、リーダビリティ値は 2.70 となっている。また、この数値を四捨五入して元々のレベルと対応付けると分類は「上級」となる。

<sup>3</sup> kReadability のプロトタイプシステムは <http://asaokitan.net/kreadability> からアクセスできる。



分析結果	
リーダビリティスコア	2.70
レベル	上級
連体と終止の差	-0.21
連体の頻度	0.46
名詞率	0.39
派生述語率	0.39
平均文長	36.00

図 5 開発中の kReadability 画面の例 (分析結果画面)

さらに、形態素解析結果が表示され、形態素ごとの品詞、辞書形、漢字表記が一覧表示される。それに加えて、情報が利用可能な場合は、語彙レベルや日本語訳も表示される (図 6) 4。

81	산발	Noun-普通-動作	산발02	散發		
82	적	Suffix-名詞派生-*	적18	的		
83	연쇄	Noun-普通-*	연쇄01	連鎖		
84	감염	Noun-普通-動作	감염	感染		
85	도	Ending-助詞-とりたて	도	*		
86	발생	Noun-普通-動作	발생	發生	3級	發生
87	해	Suffix-動詞派生-*	하다02	*		
88	안심	Noun-普通-動作	안심04	安心	3級	安心
89	하	Suffix-動詞派生-*	하다02	*		
90	ㄹ	Ending-語尾-連体形	ㄹ	*		
91	수	Noun-依存名詞-*	수02	*	5級	數
92	없	Sonzaishi-自立-*	없다01	*		
93	는	Ending-語尾-連体形	는	*		
94	상황	Noun-普通-*	상황02	狀況	3級	狀況
95	이	Siteisi-非自立-*	이다	*		
96	다	Ending-語尾-終止形	다	*		
97	.	Symbol-ピリオド-*	.	*		

図 6 開発中の kReadability 画面の例 (形態素解析結果画面の一部)

「語彙レベル」は、リーダビリティ値からは得られない情報を補うものである。リーダビリティ値は文長や品詞の分布など、文章の統計的特徴だけから算出されており、文章で使用されている語彙や文法事項のレベルについて、直接、情報が得られる性質のものではない。そのため、ハングル検定のレベル別語彙リストと照合する

4 現段階では、一部の語彙レベルについてデータを準備中であることや、同綴異義語を見分けることが難しいなどの理由により、仮に一部の語についてのみ語彙レベルや日本語訳を表示している。より網羅的に語彙レベルや日本語訳を表示することは今後の課題となる。「韓国語学習辞典」(<http://koreanlexis.webcrow.jp/Vector/index.php>)では、網羅的な語彙レベルや対訳の表示が実現されているので、参考にしたいと考えている。

ことで、文章中で用いられている個々の語彙項目についてレベル情報を提示することにした。現在は実装していないがハングル検定のほか韓国語能力試験 (TOPIK) の語彙リストも利用可能である。今後は語彙レベルの集計結果をグラフで表示するなど、「語彙レベル」欄から得られる情報を要約して視覚的にユーザに提示することで、リーダビリティ値と合わせて総合的なレベルの評価を行うことができるようにすることを検討している。

その他の機能として、使用方法やリーダビリティ値の読み方に関する説明や、各レベルごとのモデルテキストを掲載する予定である。

### 3.5. 今後の課題

現状のプロトタイプシステムについてはいくつか課題が指摘できる。まず、現状では形態素解析結果をほぼそのまま表示しているが、これはユーザにとって必ずしも直感的ではない。例えば、形態素解析結果では、句読点や括弧などの記号類も形態素扱いされ、通常の単語と並列で一覧表示されてしまうが、そのような情報は有用性が低い。

形態素解析の単位が学習項目の単位に比べて細かすぎるという点も指摘できる。例えば、「할 수 없다」という表現は現状の形態素解析結果では「하/動詞+ㄴ/語尾+수/名詞+없다/動詞+다/語尾」と分解される。ハングル検定の語彙レベルリストには「할 수 없다」全体で 4 級の語彙に掲載されているが、形態素解析結果では、より細かい単位に分割されてしまっているため、自動的に対応付けることが簡単にはできない。

また、形態素解析は必ず正しい結果が得られるとは限らない。特に、ニュース記事などを題材とする場合、「문재인」や「코로나」のような固有名詞や時事性の高い語が辞書に登録されていないことが多いため、「코/名詞+로/助詞+나/助詞」といった誤った解析結果が表示され、またリーダビリティ値もそれに影響を受けてしまうという問題がある。これらの例のような重要な語については、形態素解析辞書をアップデートすることで形態素解析の精度を改善していく必要がある。他の問題として、仮に形態素解析結果が正しくとも、同綴異義語がある場合、それを語彙レベルなどと対応付ける段階でも誤りが生じるおそれがあることも指摘できる。

これらの問題は辞書や手法を工夫することで一定程度の改善は見込めるが、いずれにしても自動処理は 100%の正確さが保証できるものではないため、ユーザがその分析結果を常に正しいものと受け取ってしまわないよう、見せ方に工夫が必要になると言える。

#### 4. おわりに

本稿では、韓国語の文章難易度を自動判定する韓国語文章リーダビリティ判定システム「kReadability」の概要と構築状況について述べた。今後の開発方針として、電子辞書や語彙学習支援のためのシステムと連動し、学習者の自律的な学習を支援する「アダプティブラーニング読解学習支援システム」を構築するという目標を掲げている。より具体的には、電子辞書、語彙レベル別の単語帳、語彙能力測定ツールなどとリーダビリティ判定システムとを連携させ、学習者がツール群を自由に行き来できるシステムを構築することで、学習者の学習スキル向上を支援することを目標としている。

また、今回開発したシステムを利用した教育現場での教材開発、読解テスト作成などの実践に繋げていくとともに、その知見をシステム開発へフィードバックし、システムの改良を行っていきたいと考えている。

#### 謝辞

韓国語リーダビリティ開発の基準テキストデータ作成に際し、九州産業大学の長谷川由起子先生、延世大学言語情報院の Kim Hansem 先生、福岡大学の丁仁京先生からテキストデータのご提供をいただきました。感謝申し上げます。なお、本研究は JSPS 科研費 JP19K00794 の助成を受けたものです。

#### 参考論著

Hasebe, Yoichiro and Lee, Jae-Ho (2015) Introducing a Readability Evaluation System for Japanese Language Education. *Proceedings of the 6th International Conference on Computer Assisted Systems for Teaching & Learning Japanese (CASTEL/J)*, pp.19-22.

池田真 (2011) 「CLIL の基本理念」、渡辺他編『CLIL 上智大学の外国語教育の新たな挑戦』、pp.1-13、上智大学出版

和泉伸一・池田真・渡部良典 (編) (2012) 『CLIL 上智大学の外国語教育の新たな挑戦 第2巻 実用と応用』、上智大学出版

鎌田美千子 (2016) 「日本語を第二言語とする子どもたちのためのリライト教材作成に関する方法論的検討—日常会話レベルから教科書レベルへの橋渡し—」、『宇都宮大学留学生教育研究論集 7』、pp.3-10.

川村よし子 (2009) 『チュウ太の虎の巻：日本語教育のためのインターネット活用

術』、くろしお出版

Kudo, T., Yamamoto, K., and Matsumoto, Y. (2004) Applying Conditional Random Fields to Japanese Morphological Analysis. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004)*, pp.230-237.

林炫情 (2018a) 「CLIL (内容言語統合型学習) の教育実践と展望—一日中での韓国語学習を例に」、山口県立大学国際文化学部 (編) 『多文化・多言語を生きるために: 地域でみかくコミュニケーションの力』、pp.101-124、山口県立大学

林炫情 (2018b) 「CLIL を活用した韓国語の授業設計とその課題」、第 79 回朝鮮語教育学会口頭発表、2018.9、京都女子大学 (京都)

林炫情 (2019a) 「内容言語統合型学習 (CLIL) 韓国語授業における教師の役割」、『多文化研究と学際的教育』国際シンポジウム、2019.9、山東大学 (中国)

林炫情 (2019b) ‘다문화 공생사회 실현을 위한 언어교육-내용언어통합학습 (CLIL) 한국어 교육실천과 과제’、大連外国語大学 韓国語学院 招待講演、2019.9、大連外国語大学 (中国)

李在鎬 (2011) 「大規模テストの読解問題作成過程へのコーパス利用の可能性」、『日本語教育』148、pp.84-98.

李在鎬・林炫情・須賀井義教・浅尾仁彦・斉藤信浩 (2019) 「リーダビリティを活用した読解授業実践 (日本語教育から韓国語教育へ)」、朝鮮語教育学会ワークショップ企画、2019.6、京都女子大学 (京都)

李在鎬・柴崎秀子 (2012) 「文章の難易度と語彙の関連性に関する考察」、『コーパスとテキストマイニング』、pp.181-192、共立出版

森川結花 (2017) 「「つながり」の実現を目指した読解授業プラン: 吉本ばなな作「ムーンライト・シャドウ」を用いた読解授業の実践報告と授業案再考」、『言語と文化』21、pp.197-211.

野本忠司 (2016) 「リーダビリティ研究の 100 年」、『情報処理学会 SIG Technical Reports 2016-DC-101』、pp.1-7、情報処理学会

奥野由紀子 (編著) (2018) 『日本語教師のための CLIL 入門』、凡人社

佐藤慎司・高見智子・神吉宇一・熊谷由理 (編) (2015) 『未来をすることばの教育をめぐって: 内容重視の批判的言語教育』、ココ出版

柴崎秀子 (2014) 「リーダビリティ研究と「やさしい日本語」」、『日本語教育』158、

pp.49-64.

- 柴崎秀子・原信一郎 (2010) 「12 学年を難易尺度とする日本語リーダビリティ判定式」、『計量国語学』 27-6、pp.215-232.
- 染谷泰正 (2010) 「オンライン版「英文語彙難易度解析プログラム」(Word Level Checker) の概要とその応用可能性について」、青山学院大学文学部『紀要』 51、pp.97-120.
- 須賀井義教 (2013) 「MeCab (めかぶ) を用いた現代韓国語の形態素解析」、『朝鮮語研究』 5、pp.283-312、ひつじ書房
- 栗野真紀子・川本かず子・松田緑 (2012) 『日本語教師のための多読授業入門』、NPO 法人日本語多読研究会 (監修)、アスク
- 渡部良典・池田真・和泉伸一 (編) (2011) 『CLIL 上智大学の外国語教育の新たな挑戦 第 1 巻 原理と方法』、上智大学出版
- 安井朱美・井手友里子・土居美有紀 (2015) 「中級日本語読解教材の書き換えにおける一考察—リーダビリティスコアと日本語教師の評価を基に—」、『国際教育センター紀要』 15、pp.23-37.
- 横山りえこ (2019) 「多読が作文に与える影響についての一考察: 多読素材と看図作文に焦点をあてて」、『2019 年度日本語教育春季大会予稿集』、pp.406-411.

【要旨】

kReadability: A Korean text readability assessment system

한국어 문장 리더빌리티 판정 시스템 "kReadability"

최근 언어 학습과 내용 학습을 통합한 교육 모델인 내용 언어 통합 학습 (Content and Language Integrated Learning; CLIL)에 대한 사회적 기대가 높아지고 있다. 이 CLIL 은 진정성 있는 신문, 잡지, 웹 사이트 등의 소재가 권장되지만, 이러한 원시 자료는 난이도가 분명하지 않기 때문에 어떤 문장을 어떤 수준의 학습자에게 주어야 할 것인가라는 문제가 생긴다. 이러한 과제를 해결하기 위해 본 연구에서는 한국어 문장 난이도를 자동 판정할 수 있는 한국어 문장 리더빌리티 판정 시스템 "kReadability"를 개발하고 있다. "kReadability" 개발의 첫 단계로 난이도별 교과서 코퍼스를 구축했다. 그리고 교과서 코퍼스를 계량 분석하여 "관형사형과 종결형의 차이, 관형사형 빈도, 명사 비율, 파생 용언 비율, 평균 문장길이"를 변수로 하는 리더빌리티 측정 척도를 고안했다. 이 리더빌리티 측정 척도의 정확도는  $R^2 = .84$  이며, 한국어의 다양한 텍스트의 가독성을 정밀하게 예측하는 것으로 나타났다. CLIL 에서 사용하는 소재의 난이도를 "kReadability"로 객관적으로 판정할 수 있어서 한국어 독해 교육 지원에 공헌할 수 있다. 현재 리더빌리티 측정 도구를 누구나 어디서나 이용할 수 있도록 하기 위해 웹 브라우저에서 "kReadability"가 이용 가능한 환경을 만들고 있으며, 앞으로 교육 현장에서 교재 개발, 독해 테스트 작성 등의 실천에 연결해 나갈 계획이다.

キーワード：リーダビリティ, 教科書コーパス, 重回帰分析, 内容言語統合型学習