

令和 5 年 6 月 12 日現在

機関番号：33801

研究種目：基盤研究(C)（一般）

研究期間：2018～2022

課題番号：18K00722

研究課題名（和文）CEFR読解CDSに基づいた問題仕様書の開発と妥当性の検証

研究課題名（英文）Development and validation of task item specifications based on the CEFR Reading CDS

研究代表者

谷 誠司（Tani, Seiji）

常葉大学・外国語学部・教授

研究者番号：80514827

交付決定額（研究期間全体）：（直接経費） 3,500,000円

研究成果の概要（和文）：本研究はCommon European Framework of Reference for Languages（以下CEFR）の読解Can-Do Statements（以下CDS）からの読解問題（例文と質問）作成を支援することを目的とし、以下4点を行った：1）日本語例文をCEFR読解CDSに自動分類できる「Webアプリケーション」の実装、2）CEFR読解CDSに対応したサンプル問題の提供、3）CEFR読解CDSに記載されていない情報の提供、4）CEFR読解CDSからの問題作成の支援のための問題仕様書の作成。研究成果は1）は完了、2）と3）は一部完了、4）は2）・3）が完了後作成予定である。

研究成果の学術的意義や社会的意義

英語に比べ、日本語は基準特性（CEFRの各レベルを規定する言語特徴）に関する研究が進んでいないため、利用できる情報がCDSの記載情報以外しかなく、テストや授業で使用する読解問題（例文と質問文）を作成することは負担が大きい。そのため、日本語例文をCEFR読解CDSに自動分類できる「Webアプリケーション」の実装は、負担軽減に大いに寄与できる。また、一部ではあるが、CEFR読解CDSから実際に日本語読解問題を作成し、分析した結果、CDSには記述されていない情報も提供できるようになり、問題作成の負担軽減につながることを期待できる。

研究成果の概要（英文）：The purpose of this study was to support the creation of reading comprehension tasks (passages and questions) from reading comprehension CDS (Can-Do Statements, hereafter CDS) in the Common European Framework of Reference for Languages (CEFR). Specifically, the following four points were addressed: 1) implementation of a web application that can automatically classify Japanese passages into the CEFR reading comprehension CDS, 2) provision of sample tasks corresponding to the CEFR reading comprehension CDS, 3) provision of additional information not described in the CEFR reading comprehension CDS, and 4) creation of specifications for support for creating tasks from the CEFR reading comprehension CDS.

The results of the research are as follows: 1) is completed, 2) and 3) are partly completed, and 4) will be created after 2) and 3) are completed.

研究分野：日本語教育

キーワード：日本語教育 CEFR 読解 Can Do Statements 例文 自動分類 問題仕様書

1. 研究開始当初の背景

CEFR (Common European Framework of Reference for Languages、以下 CEFR) は 2001 年に欧州評議会が発表して以来、世界の外国語教育に影響を与えている。特に技能別に 6 つのレベルで具体的にどのようなことができるかを記述した CDS (Can-Do Statements、以下 CDS) を利用してコース・教材・テストの開発が数多くされている。

CEFR は教材やテストなどに幅広く利用される一方、CEFR のテスト利用については批判もある。Green (2012) は CEFR に準拠してテスト開発をする場合、基準にする CEFR の CDS の記述は具体性が欠け、下位尺度を加えたりするだけでは問題開発者や評価者のニーズは満たせず、どのようなことがどのくらいできるのか、どのような状況であるのかに関する詳細な記述の追加が必要であると指摘している。例えば、B1 (中級) レベルの CEFR 読解 CDS では、「機器に関する、明瞭に書かれた簡単な使用説明を理解することができる」とあるが、「明瞭」や「簡単」がどのような言語的特徴を持っているのか、「理解する」はどのような理解なのかは CDS からは分からない。英語に比べ、日本語は基準特性 (CEFR の各レベルを規定する言語特徴) に関する研究が進んでいないため、利用できる情報が CDS の記載情報以外しかなく、テストや授業で使用する読解問題 (例文と質問文) を作成することは負担が大きいというのが現状である。

2. 研究の目的

本研究の目的は CEFR 読解 CDS からの読解問題 (例文と質問) 作成の支援である。具体的には以下の 4 点である。

1) 日本語例文を CEFR 読解 CDS に自動分類できる「Web アプリケーション」の実装

収集済みの CEFR 読解 CDS に対応した日本語例文データを機械学習により他の CDS と区別する特徴量 (専門性や文長など) を抽出する。抽出された特徴量を使って、日本語例文を CEFR 読解 CDS に自動分類できる「Web アプリケーション」を実装する。またこの「Web アプリケーション」では CDS との適合度の高い例文を提示する機能も備える。

2) CEFR 読解 CDS に対応したサンプル問題の提供

CEFR 読解 CDS との適合度が高い日本語例文からサンプル問題 (例文 + 質問) を作り、予備テストを実施し、サンプル問題の項目分析をする。項目分析を通して質の高いサンプル問題を提供する。

3) CEFR 読解 CDS に記載されていない情報の提供

サンプル問題の分析を通して、CEFR 読解 CDS では記述されていない情報を提供する。

4) CEFR 読解 CDS からの問題作成の支援のための問題仕様書の作成

1) ~ 3) をまとめて、問題仕様書を作成し、CEFR 読解 CDS からの問題作成を支援する。

3. 研究の方法

本研究の方法は 4 つの研究の目的に合わせて次のように計画した。

1) 日本語例文を CEFR 読解 CDS に自動分類できる「Web アプリケーション」の実装

CEFR 読解 CDS に対応する例文収集は、内容が高度で母語話者でも到達が難しい C1 と C2 レベルと、読解力より語彙力の記述をしている CDS を除いた 27 個の CDS に対して、既に行っており、約 1000 例文が収集してある。収集済みの例文を使用して機械学習により、日本語例文を CDS に自動分類できる「Web アプリケーション」を実装する。またこの「Web アプリケーション」には CDS に対応する例文を、その適合度の高い順に出力する機能も備える。

2) CEFR 読解 CDS に対応したサンプル問題の提供

CEFR 読解 CDS は、複数の下位尺度 (「必要な情報を探し出す」、「手紙やメールを読む」、「情報や要点を読み取る」、「説明を読む」、「読むこと全般」) から構成されている。下位尺度別に A1 ~ B2 に属する CDS を対象にサンプル問題を作成する。サンプル問題には「Web アプリケーション」が提示した CDS との適合度の高い例文を使い、韓国人日本語学習者 (韓国の大学で日本語を専攻している大学生) を対象に予備テストを実施する。サンプル問題の項目分析 (項目安易値や項目弁別力指数、CEFR のレベルと項目安易値の相関関係など) を通して質の高いサンプル問題も提供する。

3) CEFR 読解 CDS に記載されていない情報の提供

サンプル問題の解答時の読解行動情報を提供するために、予備テストの受験協力者を対象にアンケート調査を実施する。

4) CEFR 読解 CDS からの問題作成の支援のための問題仕様書の作成

1) ~3) をまとめて、問題仕様書を作成し、CEFR 読解 CDS からの問題作成を支援する。

4. 研究成果

4 つの研究の目的に応じて、次のような研究成果が得られた。

1) 日本語例文を CEFR 読解 CDS に自動分類できる「Web アプリケーション」の実装

収集済みのレベル群 (A1~B2) の例文を使用して機械学習 (Support Vector Machine) により、日本語例文を CDS に自動分類できる「Web アプリケーション」を実装した。またこの「Web アプリケーション」には CDS に対応する例文を、その適合度の高い順に出力する機能も装備した。

Council of Europe が 2018 年に "CEFR Companion Volume with New Descriptors" を発表し、最初歩レベルである Pre-A1 (レベル) の CDS が新たに 7 個追加されることで言語能力レベルを 7 段階とするなどの改定を行った。この新たに追加された Pre-A1 用の 149 の例文を準備し、今までの「専門性」, 「文長」, 「文書タイプ」に加え、例文内の「漢字出現比率」を加えて 4 パラメータとした上で、機械学習により自動分類を行うプログラムを実装した。実験を行って精度を検証した結果、全体の精度は 69.66%であったが、これはマルチレベル数の多寡にも大きく依存していることが考えられた。

Pre-A1 レベルは従前のレベル群 (A1~B2) との難易度に大きな差があると考え、例文を全ての CDS に一度に分類する手法に対し、先に Pre-A1 レベルと従前のレベル群のいずれかを判定 (2 値分類) した後、CDS に分類するという階層的 CDS 分類を試みた。その結果、総体的には精度は改善したと考えられるが、1 つの例文に対して予測する CDS の数が平均 2.76 個に対して約 2 倍の数の推測を行ったため、必然的に正の適合率に良好な結果が得られなかった。また、従前のレベル群のみに対応していた「Web アプリケーション」を Pre-A1 レベルに対応できるようにした。

分類のための特徴量として「文書タイプ」, 「専門性」, 「文長」, 「漢字率」を採用していたが、その中の「文書タイプ」と「専門性」の同定のために現行の版では fastText を用いており、これに対し新たに BERT アルゴリズムを適用することで予測精度の向上を試みた。また、「Web アプリケーション」のシステムに BERT を組み込むことに取り組んだ。

「Web アプリケーション」の精度改善を目的として、34 種の CDS を対象に実験を行った。第 1 段階として推定対象の文章の難易度から CEFR レベルを推定し、その後 CEFR レベルに属する CDS のみを対象とした CDS 推定を行った。

2) CEFR 読解 CDS に対応したサンプル問題の提供

CEFR の読解 CDS は「必要な情報を探し出す」, 「手紙やメールを読む」, 「情報や要点を読み取る」, 「説明を読む」, 「読むこと全般」といった複数の下位尺度から構成されるが、その中の「必要な情報を探し出す」と「手紙やメールを読む」の A1~B2 レベルの CDS に限定して、「Web アプリケーション」が提示した CDS との適合度の高い例文を使って、サンプル問題を作成した。韓国の大学で日本語を専攻している大学生 53 名を対象にサンプル問題を使って予備テストを実施した。また、サンプル問題の項目分析 (項目安易値や項目弁別力指数、CEFR のレベルと項目安易値の相関関係など) も行った。項目分析の結果、全体としては CEFR のレベルに対応していたが、項目安易値が CEFR のレベルに対応していない項目や項目弁別力指数が低い項目があった。また、作問者から CDS と例文のレベルの不一致、CDS で記述された読み方と例文の不一致等の指摘もあった。また同じサンプル問題を日本人大学生 21 名に受検してもらい、項目分析を行った。分析結果から日本人大学生であっても想定よりも項目安易値が低い項目があり、設問の仕方や選択肢に改善すべき点があることが明らかになった。

「情報や要点を読み取る」, 「説明を読む」, 「読むこと全般」の下位尺度に属する A1~B2 レベルの CDS については、サンプル問題が既に作成され、予備テスト実施と項目分析を計画しているところである。

3) CEFR 読解 CDS に記載されていない情報の提供

サンプル問題の解答時の読解行動情報を提供するために、前述の予備テスト受検協力者 (韓国の大学で日本語を専攻している大学生 53 名) を対象にアンケート調査を実施した。アンケート調査はサンプル問題を解答するときを使用したと思う行動を選択肢から選んでもらうパート (Weir, Hawkey, Green & Devi, 2009; Wu, 2011; Box, 2013 を参考にした) とテスト全体へ感想を求めるパートで構成されている。解答行動に関しては全受検協力者 53 名を得点上位群 (得点上位約 1/3 にあたる 17 名) と得点下位群 (得点下位約 1/3 にあたる 17 名) に分けて分析した。得点上位群はレベルや文章タイプに合わせて解答行動が多様であるに対し、得点下位群は単調であることが明らかになった。自由記述については文字サイズの小ささや可読性の悪さに関する指摘があった。また日本人大学生受検協力者 21 名にも同じアンケート調査を実施した。韓国日本語学習者 (得点上位群) は一部の問題に対して日本人大学生の解答行動と近いことが明らかになった。

予備テストの作問者から「Web アプリケーション」が選定した適合度の高い例文が必ずしもテスト問題として適切とは限らない可能性が示唆されたため、CDS と例文の一致度や作問可能度

ついて調査を実施した。調査は「必要な情報を探し出す」、「手紙やメールを読む」、「情報や要点を読み取る」、「説明を読む」、「読むこと全般」の下位尺度に該当する A1～B2 レベルの CDS を対象に行った。「Web アプリケーション」から各 CDS との適合度上位 5 例文を選択し、a) 対応する CDS とどのくらい一致しているか(「一致度」)、b) CDS で記述されている「(読解)行動」を問える問題にどのくらいできるか(「作問可能度」)の 2 点から 5 段階評定と評定理由のコメントを韓国での日本語教育経験者 4 名に依頼した。これにより、Web アプリケーションによる CDS との適合度上位 5 例文に対して、韓国での日本語教育経験者 4 名による「一致度」と「作問可能度」の評定点が付与されることになり、例文に対してより複合的な情報を提供できるようになった。

また、日本語教育経験者 4 名から下位尺度別に以下のようなコメントがあり、CDS の記載内容を補完できるようになった。

「必要な情報を探し出す」：同じ CDS (B2 と A2) の例文でも文字数にかなりの差がある。

「手紙やメールを読む」：A1 レベルの CDS にある「短く簡単な」とは基本的な定型表現レベルである。B2 レベルの CDS にある「専門分野」は具体的に高い専門的知識(例えば、精密機械に関する知識)を必要になる水準ではない。

「情報や要点を読み取る」：B1 レベルの例文は背景知識を必要としない例文が望ましい。A1 レベルの例文では商品広告のキャッチコピーが使われる場合、表現が難しく、背景知識が必要で、情報量が少なすぎて作問が難しい。

「説明を読む」：A2 レベルの例文は語彙レベルが A2 にしては高い例文がある。また CDS に「短く簡単」とあるので、例文も文が短いもの多かったが、文が短くなったため表現が難しくなっていたり、作問が難しくなっていたりする例文がある。

「読むこと全般」：A1 レベルの例文の中には語彙レベルが A1 レベルにしては高すぎる例文や、文章が短すぎて作問が難しい例文がある。また A1 レベルの例文では日本社会に関する背景知識を要求しない例文が望ましい。

4) CEFR 読解 CDS からの問題作成の支援のための問題仕様書の作成

1) の Web アプリケーションの実装はできている。2) のサンプル問題の提供は「情報や要点を読み取る」、「説明を読む」、「読むこと全般」の下位尺度に属する A1～B2 レベルの CDS についてはサンプル問題はできているが、予備テスト実施をこれから計画しており、サンプル問題の項目分析や 3) の読解行動の調査はこれからである。2) と 3) が終わり次第、4) に問題仕様書を作成する予定である。

5. 主な発表論文等

〔雑誌論文〕 計7件（うち査読付論文 1件/うち国際共著 5件/うちオープンアクセス 3件）

1. 著者名 谷 誠司, 宮崎 佳典, 安 志英, 元 裕璟	4. 巻 4
2. 論文標題 CEFR読解指標 (CDS) に基づく作題例の妥当性に関わる特性抽出の試み 2	5. 発行年 2023年
3. 雑誌名 常葉大学大学院国際言語文化研究科研究紀要	6. 最初と最後の頁 1-11
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する
1. 著者名 谷 誠司, 宮崎 佳典, 安 志英, 元 裕璟	4. 巻 3
2. 論文標題 CCEFR読解指標 (CDS) に基づく作題例の妥当性に関わる特性抽出の試み	5. 発行年 2022年
3. 雑誌名 常葉大学大学院国際言語文化研究科紀要	6. 最初と最後の頁 33-43
掲載論文のDOI (デジタルオブジェクト識別子) 10.18894/00002386	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する
1. 著者名 谷 誠司, 宮崎 佳典, 安 志英, 元 裕璟	4. 巻 2
2. 論文標題 CEFR読解指標に基づいた日本語能力テストの分析 - 日本人大学生の受験データに基づいて -	5. 発行年 2021年
3. 雑誌名 常葉大学大学院国際言語文化研究科紀要	6. 最初と最後の頁 45-58
掲載論文のDOI (デジタルオブジェクト識別子) 10.18894/00002073	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する
1. 著者名 宮崎 佳典, Vuong Hong Duc, 谷 誠司, 安 志英, 元 裕璟	4. 巻 125
2. 論文標題 CEFR Companion Volume1に対応した日本語例文自動分類手法	5. 発行年 2020年
3. 雑誌名 日本學報	6. 最初と最後の頁 153-175
掲載論文のDOI (デジタルオブジェクト識別子) 10.15532/kaja.2020.11.125.153.	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 谷誠司・宮崎佳典・安志英・元裕璟	4. 巻 1
2. 論文標題 CEFR読解指標に基づく日本語能力テスト開発の試み	5. 発行年 2020年
3. 雑誌名 常葉大学国際言語文化研究科研究紀要	6. 最初と最後の頁 43-54
掲載論文のDOI (デジタルオブジェクト識別子) 10.18894/00002051	査読の有無 無
オープンアクセス オープンアクセスとしている(また、その予定である)	国際共著 該当する

1. 著者名 谷 誠司, 宮崎 佳典	4. 巻 35
2. 論文標題 韓国人日本語学習者による自己評価と韓国人日本語教師による難易度評価の調査結果比較: CEFR読解Can-Do Statementsを対象に	5. 発行年 2019年
3. 雑誌名 常葉大学外国語学部紀要	6. 最初と最後の頁 45-57
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Huynh Nguyen Tra My, Y. Miyazaki, S. Tani	4. 巻 33(2)
2. 論文標題 Inferring CEFR Reading Comprehension Index Based on Japanese Document Classification Method Including Pre-A1 Level	5. 発行年 2018年
3. 雑誌名 教育システム情報学会 研究報告	6. 最初と最後の頁 63-69
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計15件(うち招待講演 0件/うち国際学会 8件)

1. 発表者名 里見 陸, 宮崎 佳典, 谷 誠司, 安 志英, 元 裕璟
2. 発表標題 CEFR準拠CDS自動分類システムの精度改善を目的とした推定改良手法の提案
3. 学会等名 2022年度JSiSE学生研究発表会(東海地区)
4. 発表年 2022年~2023年

1. 発表者名 里見 陸, 宮崎 佳典, 谷 誠司
2. 発表標題 CEFRレベルを用いた2段階分割による日本語文章CDS自動分類システムの精度改善の試み
3. 学会等名 情報学ワークショップ2022(WINF2022)
4. 発表年 2022年

1. 発表者名 里見 陸, 宮崎 佳典, 谷 誠司
2. 発表標題 CEFRを用いた日本語文章CDS自動分類システムの精度改善の試み
3. 学会等名 韓国日本學會第104回國際學術大會(國際学会)
4. 発表年 2022年

1. 発表者名 CAO HOAI GIANG、宮崎 佳典、谷 誠司、安 志英、元 裕璟
2. 発表標題 CEFRに対応した日本語例文自動分類システムのBERT適用による精度改善の試み
3. 学会等名 韓国日本學會第102回國際學術大會(國際学会)
4. 発表年 2021年

1. 発表者名 谷 誠司、宮崎 佳典、安 志英、元 裕璟
2. 発表標題 CEFR読解指標(CDS)に基づく作題例の妥当性に関わる特性抽出の試み
3. 学会等名 The 9th International Conference on Computer Assisted Systems for Teaching & Learning Japanese (CASTEL/J)(國際学会)
4. 発表年 2021年

1. 発表者名 Cao Hoai Giang、宮崎 佳典、谷 誠司
2. 発表標題 CEFR読解指標に基づく日本語例文自動分類Webアプリケーションの精度向上に向けた取り組み
3. 学会等名 教育システム情報学会(JSiSE)第46回全国大会、インタラクティブセッション
4. 発表年 2021年

1. 発表者名 宮崎 佳典、Vuong Hong Duc、安藤 聖野、谷 誠司、安 志英、元 裕璟
2. 発表標題 CEFR Companion Volumeに対応した日本語例文分類手法
3. 学会等名 韓国日本學會第100回國際學術大會（國際学会）
4. 発表年 2020年

1. 発表者名 Vuong Hong Duc, Huynh Nguyen Tra My, Y. Miyazaki, S. Tani
2. 発表標題 A Pilot Study to Infer CEFR Can-Do Statements Based on a Japanese Document Classification Method Including the Pre-A1 Level
3. 学会等名 9th International Congress on Advanced Applied Informatics (IIAI-AAI) (國際学会)
4. 発表年 2020年

1. 発表者名 Cao Hoai Giang、宮崎 佳典、谷 誠司
2. 発表標題 CEFR読解指標に基づく日本語例文の自動分類Webアプリケーション開発について
3. 学会等名 情報学シンポジウム2020
4. 発表年 2020年

1. 発表者名 宮崎 佳典, 谷 誠司, 安 志英, 元 裕璟
2. 発表標題 韓国語学習者のためのCEFR読解指標に基づく例文自動分類 - A0を含めた7段階のCEFRレベルを対象に
3. 学会等名 The 8th International Conference on Computer Assisted Systems for Teaching & Learning Japanese (CASTEL/J) (国際学会)
4. 発表年 2019年

1. 発表者名 安藤 聖野, 宮崎 佳典, 谷 誠司, 安 志英
2. 発表標題 CEFR Companion Volume1に対応した日本語例文自動分類の試み
3. 学会等名 情報学ワークショップ2019(WINF2019)
4. 発表年 2019年

1. 発表者名 安藤 聖野, 宮崎 佳典, 谷 誠司, 安 志英
2. 発表標題 CEFR Companion Volume1に対応した日本語例文自動分類のアルゴリズム開発
3. 学会等名 情報学シンポジウム2019
4. 発表年 2019年

1. 発表者名 Y. Miyazaki, R. Hirakawa, S. Tani, J. An,
2. 発表標題 Document Classification Method Based on CEFR Reading Comprehension Index for Korean Learners of Japanese
3. 学会等名 2018 KACL & ISLI Joint Conference (国際学会)
4. 発表年 2018年

1. 発表者名 Huynh Nguyen Tra My, Y. Miyazaki, S. Tani
2. 発表標題 Inferring CEFR Reading Comprehension Index Based on Japanese Document Classification Method Including Pre-A1 Level
3. 学会等名 第17回情報科学技術フォーラム(FIT)
4. 発表年 2018年

1. 発表者名 谷 誠司, 宮崎 佳典, 安 志英, 元 裕璟
2. 発表標題 CEFR読解指標に基づく日本語能力テスト開発の試み
3. 学会等名 韓国日本學會第98回國際學術大會(國際学会)
4. 発表年 2019年

〔図書〕 計1件

1. 著者名 宮崎 佳典, 平川 遼汰, 高田 宏輝, 谷 誠司	4. 発行年 2019年
2. 出版社 ひつじ書房	5. 総ページ数 14
3. 書名 第4章「CEFR 読解指標に基づく日本語例文分類手法」(『ICT×日本語教育 - 情報通信技術を利用した日本語教育の理論と実践』(當作靖彦 監修・李在鎬 編))	

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担者	宮崎 佳典 (MIYAZAKI Yoshinori) (00308701)	静岡大学・情報学部・教授 (13801)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------