

科学研究費助成事業 研究成果報告書

平成 29 年 9 月 14 日現在

機関番号：12603

研究種目：若手研究(B)

研究期間：2015～2016

課題番号：15K16792

研究課題名(和文) 学習者コーパスに基づく習熟度別ポルトガル語学習語彙目標の策定

研究課題名(英文) Portuguese Vocabulary List based on Learner Corpora for Teaching Portuguese as a Foreign Language in Japan

研究代表者

鳥越 慎太郎(Torigoe, Shintaro)

東京外国語大学・大学院総合国際学研究院・研究員

研究者番号：20743511

交付決定額(研究期間全体)：(直接経費) 800,000円

研究成果の概要(和文)：本研究ではポルトガル語の学習目標語彙リストを、欧州言語共通参照枠の習熟度で区分された学習者コーパスに基づいて作成することを目的とする。学習者コーパスから得られた習熟度別特徴語彙リストを一次データとし、母語話者コーパスや教科書コーパスの語彙頻度情報で調整することで、最終的に初級約1500語、中級約1500語の語彙リストが得られた。

従来、シラバスや教材における使用語彙の選定は、共通基準が不在であるために各教師や執筆者の直観や経験則に委ねられてきたが、目標習熟度の学習者が実際に用いることができる語彙からリストを作成することで、今後、目指すべき語彙が適切に収録されるようになることが期待される。

研究成果の概要(英文)：This project aimed at establishing a Portuguese vocabulary list for L1 Japanese learners based on information from CEFR-based learner corpora. The researcher adopted learner corpora as a primary data source, in addition to a native speaker corpus and textbook corpus used for comparison purpose, and finally gained a vocabulary list of 3000 words (1000 words for A level, 500 words for A+ level, and 1500 words for B level). Because of the absence of common guidelines for Portuguese education in Japan, the vocabulary used for syllabi and textbooks has been decided according to each teacher or writer's intuition and experience. This final vocabulary list based on actual learner usage may help teachers and textbook writers to consistently adopt appropriate words for the targeted level.

研究分野：ポルトガル語教育

キーワード：ポルトガル語 学習目標語彙リスト 学習者コーパス 欧州言語共通参照枠 教科書コーパス

1. 研究開始当初の背景

ポルトガル語は日本国内で約 15 万人の母語話者コミュニティを有する言語で、特に 1 万人規模のブラジル人母語話者を有する静岡県、愛知県、三重県と群馬県では、多くの大学でポルトガル語講座が開講されるなど、学習外国語としてのプレゼンスを高めている (Torigoe 2017)。このような事情に加え、2014 年にブラジルで開催されたサッカーワールドカップ、2016 年にリオデジャネイロで開催されたオリンピックを控え、2010 年頃から様々な種類のポルトガル語学習教材が多数出版されてきた (鳥越 & 山田 2015)。しかし、教材の内容はほとんど初学者向けであるが、最終的な学習到達目標は不揃いであり、タスクや場面内容も様々である。特に収録語彙については、最初の数ページから直観的に初級語彙とは考えにくいような語彙が使用されているケースが散見された。

日本国内では、ポルトガル語を始めとする英語以外の外国語教育の詳細な学習項目の規定は、事実上存在しない状況である。英語教育では学習指導要領が文法や語彙などの習熟目標を詳細に規定しているが、その他の外国語については「英語に準じる」とされる。ポルトガル語の文法指導については伝統文法に沿ってある程度の共通性があるが、語彙については伝統的な規範も存在しない。そのため、ポルトガル語講座の指導語彙やポルトガル語教材の収録語彙は、教師や執筆者の直観や経験に一任される。これは必ずしも現実の言語使用とは一致しない (McEnery, Xiao, & Tono 2008)。

このような実情から、本研究ではポルトガル語教育の語彙面における共通指標を、欧州

言語共通参照枠 (以下 CEFR) に基づいて作成、提案していくことを目指した。

2. 研究の目的

本研究の目的は客観的な言語使用のデータ (コーパス) に基づき、初級、中級、上級の各習熟度の学習者が学習、習得すべき語彙を明らかにすることである。そして、完成した学習目標語彙リストを提案し、日本国内のポルトガル語教材の語彙面での質の向上と、現在ほとんど作成されていない中級以上の教材の開発に寄与していくことを目指した。

また、本研究は CEFR の抽象的な習熟定義を各学習言語の実情に合わせて具体化していく試みのひとつでもある。これは英語教育においても CEFR 研究の課題となっている。

3. 研究の方法

本研究では、英語教育研究における CEFR 習熟度別学習目標語彙リスト、English Vocabulary Profile (EVP) (Capel 2010, 2012) や CEFR-J Wordlist (投野 2013) を参考とし、これらのポルトガル語版を目指した。ただし、母語話者データを一次データとする EVP と CEFR-J Wordlist に対し、本研究では学習者データを一次データとすることで、CEFR の理念である学習者指向主義 (投野 2013) により近づくことを目指した。

本研究の手順は以下である。

- (1) CEFR 習熟度別学習者コーパスに基づく、習熟度別特徴語彙リストの作成

コーパスデータにはポルトガル語学習者コーパスの Corpora do PLE (リスボン大学、約 7 万語) と Corpus de PEAPL2 (コインブ

ラ大学、約 12 万語) を使用した。両コーパスは共通の方法によって収集、構築された作文コーパスで、PLE は 471、PEAPL2 は 546 のデータファイルからなる。特筆すべき点として、各データファイルにはインフォーマントの自己評価に基づく CEFR 習熟度が明示されている。これにより、両コーパスを統合した後、習熟度別に再区分した。なお、PLE では CEFR 習熟度が A1-A2、B1-B2、C1-C2 の 3 段階に区分されているため、本研究でもこれに合わせている。

この学習者コーパスから、習熟度別の学習者語彙頻度リスト(約 4200 語) を得て、各語彙の頻度を習熟度別サブコーパス間で統計処理(カイ二乗検定)をし、有意に多く産出されている習熟度に振り分け、習熟度別特徴語彙リストを作成した。

(2) ポルトガル語教科書のコーパス化と、語彙頻度リストの作成

日本国内で流通するポルトガル語教科書をコーパス化し、語彙頻度リストを作成した。使用した教科書は、モデル会話、文法解説、例文と練習問題からなる「総合型教材」6 編で、OCR でスキャンしたのち、TreeTagger で各単語に品詞タグを付与し、コーパスとして整形した。

この教科書コーパスから、教科書語彙頻度リストを作成した。教科書の語彙使用は母語話者の語彙使用、及び学習者が実際に習得する語彙と乖離しているのではないかという設問に基づき、教科書語彙頻度リストと(1)の学習者語彙リスト、母語話者語彙リストを比較した。母語話者語彙リストには LMCPC (リスボン大学、約 10 万語のうち上位 5000

語) を使用した。

産出語彙のリストである(1)の学習者コーパス語彙リストに対し、教科書語彙リストは受容語彙知識からなるが、頻度中心主義では見落とされがちな低頻度の基礎語彙をカバーすることができるため (cf. Capel 2010)、比較データとして最終語彙リストの作成に活用していく。

(3) 最終的な学習目標語彙リストを完成させる。

(2)で作成した教科書語彙頻度リストを用いて、(1)の習熟度別学習者特徴語彙リストを修正した。学習者特徴語彙リストで中級と判定された語彙のうち、4 つ以上の教科書で用いられた語彙を新たに初級語彙 (A レベル) として分類した。また、学習者語彙リストで初級に判定されていなかった語彙のうち、頻度が 5 以上のものを A+レベルとした。中級語彙のうち、初級学習者の使用が 5 未満で、教科書でも使用が見られなかった語彙は中級語彙 (B レベル) とした。また、学習者に用いられていなくても、教科書で用いられている語彙は同じく B レベルとした。なお、上級 (C レベル) は、日本国内で CEFR の C1 レベル以上の学習者を「指導」する局面が想定されにくいため、割愛した (cf. 投野 2013)。

4 . 研究成果

(1) CEFR 習熟度別学習者特徴語彙リストとその課題 (Torigoe 2016)

学習者コーパスから、初級 (A1-A2 レベル)、中級 (B1-B2 レベル)、上級 (C1-C2 レベル) 各約 500 語からなる約 1500 語の特徴語彙リストが得られた。これは先行の EVP と

CEFR-J Wordlist と比較すると約 3 分の 1 程度の量である。

加えて、学習者コーパスの頻度情報にのみ基づいているため、例えば数詞や曜日の表現、色の表現など、直観的には基礎的と判断される語彙が、中級や上級語彙として過大評価されてしまった。一方で、学生生活やリスボンやコインブラの街の生活と関係が深い語彙が高頻度のために初級語彙として評価された。

(2) 教科書の使用語彙と、母語話者及び学習者の使用語彙との乖離 (鳥越 2016)

教科書語彙リストと母語話者語彙リスト (LMCPC) を比較すると、両者に乖離があることが確認された。母語話者データでは教科書と比較して政治や社会に関する語彙、空間や時間に関する抽象的な語彙、副詞や接続詞語彙が多く用いられた。一方で教科書では食事に関係する語彙、施設や店舗の語彙、その他日常生活に関する動詞語彙が母語話者データと比較すると多く用いられていた。

なお、教科書語彙リストでは頻度に加えてレンジ(使用されている教科書の数)も重要視しているため、統計学的な検討は行っていない。

(3) 最終語彙リスト (Torigoe 2017)

習熟度別学習者特徴語彙リストをベースに教科書語彙頻度リストで修正し、初級 A レベル約 900 語、A+レベル約 500 語、中級 B レベル約 1500 語の最終語彙リストを得た。中級レベル以降の語彙量は EVP や CEFR-J Wordlist と比較して依然不足しているが、初級語彙については遜色ない量となっている。

	最終語彙 リスト	学習者 特徴語彙 リスト	EVP	CEFR-J Wordlist
A1	A: 911	531	601	1000
A2	A+: 534		925	1000
B1		417	1429	2000
B2	1487		1711	2000
C1-C2	n/a	472	2300	n/a

また、最終語彙リストでは、学習者コーパスでの頻度の低さから中上級語彙に「過大評価」されていた多くの語彙群(食事、施設、職業、スポーツ、体、心情、動物、人間関係、その他日常生活の語彙など)が初中級語彙として修正された。

一方で、コーパスデータ量が限られている中で計量的情報と、最終的には主観的な区分方法に依拠しているため、特に初級 A+レベルと中級 B レベルの差異が曖昧である点が課題として残った。また、学習者コーパスがヨーロッパポルトガル語変種のものに限られるため、ヨーロッパ変種とブラジル変種の厳密な区分ができなかった点も課題である。

最終語彙リストは下記のウェブサイトにて学習者特徴語彙リスト、教科書頻度リストとともに公開している。今後は検索機能や整列機能を充実させていく。

<引用論文>

Capel, Annette. (2010). A1-B2 Vocabulary: Insights and Issues Arising from the English Profile Wordlists Project. English Vocabulary Journal, Volume 1.

Capel, Annette. (2012). Completing the English Vocabulary Profile: C1 and C2 Vocabulary. English Vocabulary Journal, Volume 3, 1-14.

McEnery, Tony; Xiao, Richard; & Tono, Yukio.

Corpus-Based Language Studies. (2006).
Abingdon: Routledge.

投野由紀夫 (2013). 『英語到達度指標 CEFR-J
ガイドブック』. 大修館書店.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者に
は下線)

〔雑誌論文〕(計 2 件)

Torigoe, Shintaro. Portuguese Vocabulary
Profile: Uma Lista de Vocabulário a
Aprendentes do PL2/PLE, Baseada nos
Corpora de Aprendentes e de Livros de
Ensino. Revista da Associação
portuguesa de Linguística (2007). 査読
あり. 2017. 掲載ページ未定.

Torigoe, Shintaro. Seeking the
Portuguese Vocabulary Profile. In A. M.
Ortiz & C. Pérez-Hernández (eds).
*CILC2016. 8th International Conference
on Corpus Linguistics (EPiC Series in
Language and Linguistics)*. Asociación
Española de Lingüística de Corpus (ス
ペインコーパス言語学会). 査読あり.
Volume 1. 2016. 396-410.
<https://easychair.org/publications/paper/270269>

〔学会発表〕(計 4 件)

Torigoe, Shintaro. Portuguese Vocabulary
Profile: uma Lista de Vocabulário a
Aprendentes do PL2/PLE Baseada nos
Corpora do PLE e Corpus de PEAPL2.
XXXII Encontro Nacional da Associação

Portuguesa de Linguística (ポルトガル
言語学会). At Universidade de Aveiro
(アヴェイロ大学、ポルトガル).
October, 2016.

鳥越慎太郎. 「ポルトガル語の教材コ
ーパスの構築と教材語彙リストの作
成」. 日本ポルトガルブラジル学会.
大阪大学. 2016年10月.

Torigoe, Shintaro. Seeking the
Portuguese Vocabulary Profile. VIII
International Conference on Corpus
Linguistics (Asociación Española de
Lingüística de Corpus: スペインコー
パス言語学会). At Universidad de Málaga
(マラガ大学、スペイン). March, 2016.

鳥越慎太郎、山田将之. 「CEFR-J に基
づくポルトガル語教材の目標習熟度
の考察」. 外国語教育学会. 東京外国
語大学. 2015年11月.

鳥越慎太郎. 「ポルトガル語の語彙学
習目標—学習者コーパスに基づいて」.
日本ポルトガルブラジル学会. 東京外
国語大学. 2015年10月.

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

名称:

発明者:

権利者:

種類:

番号:

出願年月日：

()

国内外の別：

取得状況(計 0 件)

名称：

発明者：

権利者：

種類：

番号：

取得年月日：

国内外の別：

[その他]

ホームページ等

<http://www.tufs.ac.jp/ts2/society/pvp/>

6. 研究組織

(1) 研究代表者

鳥越慎太郎 (TORIGOE, Shintaro)

東京外国語大学 総合国際学研究院 研究
員

研究者番号：20743511

(2) 研究分担者

()

研究者番号：

(3) 連携研究者

()

研究者番号：

(4) 研究協力者