

## 科学研究費助成事業 研究成果報告書

平成 30 年 6 月 19 日現在

機関番号：12603

研究種目：基盤研究(A) (一般)

研究期間：2014～2017

課題番号：26240051

研究課題名(和文)大規模会話コーパスに基づくラーニングマイニングの深化とテーラーメイド日本語教育

研究課題名(英文) Tailormade Japanese language education and advancing learning mining based on a large conversation corpus

研究代表者

芝野 耕司 (SHIBANO, Kohji)

東京外国語大学・アジア・アフリカ言語文化研究所・教授

研究者番号：50216024

交付決定額(研究期間全体)：(直接経費) 31,700,000円

研究成果の概要(和文)：独自開発のeラーニングシステムJPLANGを従来のPC中心からスマホ・タブレット対応とするとともに、独自ログ解析に加えて、Google Analyticsに加えLearning Analyticsへの対応を行った。システム環境を合理化するため、Docker、Git対応を行った。これらの成果については、AACE及びEuroCall、AAALで発表した。

字幕コーパスに関しては、順調に運用を続け、10億語規模に到達した。研究成果については、AAAL、AiLAなどの国際学会を中心に査読付き論文発表を行った

研究成果の概要(英文)：In order to modernize the original proprietary Japanese e-Learning system called JPLANG, we first change to the the smartphones/tablets environment. In order to support Learning Analytics we added Google Analytics support in addition to custom log analysis. In order to rationalize the system environment, Docker and Git support were added. These results were published at AACE and EuroCall, and AAAL conferences.

Regarding the closed caption corpus, it continued to operate smoothly, reaching the scale of 1 billion words. Regarding the research results, we presented peer-reviewed papers mainly on international academic societies such as AAAL and AiLA.

研究分野：情報工学

キーワード：eラーニング 日本語教育 コーパス言語学

## 1. 研究開始当初の背景

ICT 技術の進化によって、情報爆発が起きている。時代は、メガからギガを超え、既にペタ、エクサバイトの時代に移行しようとしている。産業界では、Cloud Computing 環境を構築するとともに、この新たな Computing 環境を利用し、新たに得られた情報を活用したデータマイニングやテキストマイニングが広く普及し、Amazon や楽天、Google など、こうした状況、すなわち、**ビッグデータ**を積極的に利用し、ビジネスの革新が起きている。しかし、教育研究分野では、こうした状況は具現してはいず、また、対応するラーニングマイニングの試みもほとんどないのが現状である。

大規模な言語情報を活用することによって、ビジネス分野では、検索精度や Apple Siri に代表される音声認識精度の大幅な向上が見られた。一方、言語教育学分野では、書き言葉に関しては上記データをはじめ十分な蓄積があるが、言語教育で重要な話し言葉に関しては、基盤となるコーパス構築すら充分でないのが現状である。

言語教育では、ヨーロッパ発の国際標準「ヨーロッパ共通言語教育参照枠(Common European Framework of Reference for Language Learning, Teaching, and Assessment, CEFR, Council of Europe 2001)」をもとに、聞き、話し、読み、書くという4技能を独話と会話の二つに分け、“話す”ことを強化したコミュニカティブアプローチに重点を移した。しかし、特殊機器としてのLLが姿を消し、PCがとって変わったが、企業の撤退によってLLの代替は実現されてはいるものの、新技術を用いて、Spoken Interactionを含む5技能への対応はなされていず、ICTの進化から言語教育が取り残されているのが現状である。

## 2. 研究の目的

この研究で言う大規模言語運用データとは、(1)日本語e-LearningシステムJPLANGの2000万件以上の大量の学習ログ、(2)318冊の教科書本体及び指導書データ7000万語からなる教科書コーパス、及び(3)地デジ字幕データから構築する1億語以上の大規模話し言葉コーパスの三つを指す。この研究では、これらの独自データをもとに、学習ログのラーニングマイニング研究の成果を活かし、二つの大規模コーパスのデータを活用することによって、言語教育の国際標準であるヨーロッパ共通言語教育参照枠(CEFR)で要求される聞く、読む、独話、会話、書くの5技能教育をe-Learning環境で実現するテラーメイド日本語教育に関連する教授法及び新世代の教員養成方法の開発を行う。また、Web会話ラボ教材の拡張、LMS拡張、海外普及及び海外サポートを強化する。

## 3. 研究の方法

研究組織は、代表者の芝野が全体を統括し、システム・評価班及び日本語教育班の二つで構成する。各研究班の主な役割は、次の通り。

システム・評価班では、ラーニングマイニングの深化のため、数千万件のログ解析を可能にし、学習効果の評価及び統計解析を行うとともに、最近急速に統計手法等の環境が整ってきたRを前提にJPLANGシステムへの統合を行う。また、大規模教科書・話し言葉コーパスを構築し利用するためのMap/Reduceシステム開発を担当する。

日本語教育班では、ラーニングマイニングの結果を日本語教育に接続するためのテラーメイド日本語教育の研究を行うとともに、大規模教科書・話し言葉コーパスとCEFRとに基づく教材開発の研究を行う。

また、海外への普及活動を継続するとともに、海外へのサポート体制を拡充する。

これまでの研究開発では、合計数十人の

大学院生の協力を得て開発を行ってきたが、今後も大学院言語応用専攻日本語教育学専修コースを中心に大学院前期及び後期課程の学生の参加を積極的に促すとともに、中国、タイ、イタリアなどの海外の協力大学との共同も強化する。【各研究班の役割】この研究の中心課題であるラーニングマイニングの深化、大規模教科書・話し言葉コーパスの構築及びテラーメイド日本語教育の研究には、情報学、教育工学、統計学、応用言語学、日本語教育学、教育学の各分野の研究を統合する必要がある。このため、情報学、教育工学の理系学問分野で「システム・評価班」を組織し、日本語教育班と二つの研究班を組織した。

なお、研究代表者の芝野は、情報学分野が主ではあるが、教育工学、応用言語学、日本言語分野においても、研究発表を行っており、全体を統括することができる。

#### 4. 研究成果

独自開発の e ラーニングシステム JPLANG を従来の PC 中心からスマホ・タブレット対応とするとともに、独自ログ解析に加えて、Google Analytics 加え Learning Analytics への対応を行った。システム環境を合理化するため、Docker、GIT 対応を行った。これらの成果については、AAACE 及び EuroCall、AAAL で発表した。字幕コーパスに関しては、順調に運用を続け、10 億語規模に到達した。研究成果については、AAAL、AiLA などの国際学会を中心に査読付き論文発表を行った

#### 5. 主な発表論文等

〔雑誌論文〕(計 14 件)

1. Hajime Mochizuki and Kohji Shibano, Searching Discourse Segments for Formulaic Sequences in a Closed Caption TV Corpus for Language Learning, 2017, World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2017, 19,27, 有, -
2. Hajime Mochizuki, Augmented

Reality Applications for Multilingual Learning with Intuitive Understanding, 2017, Proceedings of World Conference on Educational Media and Technology (EDMEDIA) 2017, 1205, 1213, 有, -

3. 藤森弘子, 初級教材『大学の日本語 初級 ともだち』の開発 その理念と実践 . 2018, 韓国言語教育学会第 31 回国際学術大会予稿集, 131, 135, 有, -
4. 藤森弘子, 教壇実習における実習生の学び - e ポートフォリオによる実習記録と自己評価結果から -, 2018, 東京外国語大学留学生日本語教育センター論集, 83, 9, 有, -
5. 鈴木美加, 『JLPTUF アカデミック日本語 Can-do リスト』の開発: Can-do リスト開発概要と項目間の難度の分析, 2018, 東京外国語大学留学生日本語教育センター論, 219, 234, 有, -
6. Hajime Mochizuki and Kohji Shibano, Analyzing Attractiveness of Specific Location Names of Tourist Destination from a Closed Caption TV Corpus, Hawaii University Conferences, Arts, Humanities, Social Sciences & Education (AHSE), 有, 2016, 2016, 1-12, -
7. Hajime Mochizuki and Kohji Shibano, Detecting Topics Popular in the Recent Past from a Closed Caption TV Corpus as a Categorized Chronicle, the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KMIS), 有, 2015, 2015, 342-349, -
8. 芝野耕司, データベース・データ工学研究者・技術者にとっての SQL 規格の意

- 味, 日本データベース学会  
Newsletter, 無, 9, 2016, -
9. Hajime Mochizuki and Kohji Shibano, Extracting Formulaic Sequences Containing Useful Expressions for Language Learning from Closed Caption TV Corpus, E-Learn 2016, 有, 2016, 2016, 29-37, -
  10. Hajime Mochizuki and Kohji Shibano, Re-Mining Topics Popular in the Recent Past from a Large-Scale Closed Caption TV Corpus, Journal of Future Computer and Communication, 有, Vol. 4, No. 2, 2015, 98-103
  11. Hajime Mochizuki and Kohji Shibano, Re-Mining Topics Popular in the Recent Past from a Large-Scale Closed Caption TV Corpus, International Association of Computer Science and Information Technology Press (IACSIT), 有, 4, 2015, 98-103, -
  12. Hajime Mochizuki and Kohji Shibano, Development of a Closed Caption TV Corpus Retrieval System to Seek Video Scenes Containing Useful Expressions for Language Learning, Proceedings of World Conference on Educational Media and Technology (EDMEDIA), 有, 2015, 2015, 1760-1768, -
  13. Hajime Mochizuki and Kohji Shibano, Building Very Large Corpus Containing Useful Rich Materials for Language Learning from Closed Caption TV, World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, 有, No.1, 2014, 1381-1389
  14. Hajime Mochizuki, The Effectiveness of Slight Modification to Supplement Programs in the Development of AR Applications, World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, 有, No.1, 2014, 1390-1396
- (学会発表)(計13件)
1. Kohji Shibano, Analyzing formulaic sequences in spoken Japanese from a large Japanese TV closed caption corpus, The 18th World Congress of Applied Linguistics (AILA 2017), 2017年
  2. XIAO Tingting, Kohji Shibano, Developing Intimacy by Style-shifting in Japanese: A TV Subtitle Corpus-based Study, The 2017 conference of the American Association for Applied Linguistics (AAAL 2017), 2017年
  3. Hajime Mochizuki, Building a Very Large Spoken Language Corpus from Closed Caption TV and Extracting Practical Formulaic Sequences for Language Learning, The 10th International Conference on Advanced Computer Theory and Engineering (招待講演), 2017年
  4. Hajime Mochizuki and Kohji Shibano, Discourse Segment Clustering with Word Embedding based on Formulaic Sequences for Language Education, 2017 International Conference on Education and Multimedia Technology (ICEMT 2017), 2017年
  5. Hajime Mochizuki and Kohji

- Shibano, Modification of word2vec by Formulaic Sequences and Extraction of Useful Expressions for Language Learning from Closed Caption TV Corpus, The IAFOR International Conference on Language Learning, 2017年01月08日~2017年01月10日, Hawaii, USA
6. XIAO Tingting, Kohji Shibano, Developing Intimacy by Style-shifting in Japanese: A TV Subtitle Corpus-based Study, The 2017 conference of the American Association for Applied Linguistics (AAAL 2017), 2017年03月18日~2017年03月21日, Portland, USA
7. 芝野耕司, SQL 言語の開発と日本の講演, EIM2016 日本データベース学会功労賞記念講演 (招待講演), 2016年02月29日~2016年03月02日, 福岡
8. Hajime Mochizuki, Development of a Closed Caption TV Corpus Retrieval System for Language Learning, 8th International Conference on Education Technology and Computers (ICETC 2016), 2016年09月28日~2016年09月30日, Singapore, Singapore
9. Hajime Mochizuki and Kohji Shibano, Analyzing Usefulness of Dialogues from Closed Caption TV Corpus as an Example of Can-do Statements for Language Learning, 2018 Hawaii University Conference, Arts, Humanities, Social Sciences & Education (AHSE), 2018年
10. Kohji Shibano, A Quantitative Formulaic Analysis of Large TV Closed Caption Corpus &#8211;
- Pragmatic Use of Utterance End in Japanese Animation Languages, 14th International Pragmatics Conference, 2015年07月26日~2015年07月31日, Antwerp Belgium
11. 芝野耕司, 日本語話し言葉コーパスの構築と会話用例検索システム, 6th CASTEL/J Hawaii, 2015年08月07日~2015年08月08日, Honolulu, USA
12. Hajime Mochizuki and Kohji Shibano Re-Mining Topics Popular in the Recent Past from a Large-Scale Closed Caption TV Corpus, 5th International Conference on Networking and Information Technology (ICNIT 2014), 2014年11月21日~2014年11月23日 Singapore
13. Hajime Mochizuki MyScale: Making Personal Paraphrases and Replacement of Scales, Ninth IEEE International Conference on Semantic Computing (IEEE-ICSC 2015) 2015年02月07日~2015年02月09日, California
- 〔図書〕(計2件)
1. 芝野耕司共訳, 日本規格協会, JIS X 3005-2: 2015 データベース言語 SQL 第2部:基本機能 (SQL/Foundation), 2015,1112
2. 芝野耕司共訳, 日本規格協会, JIS X 3005-14: 2015 データベース言語 SQL 第14部:XML 関連仕様 (SQL/XML), 2015, 334
- 〔その他〕  
ホームページ等  
<https://jplang.tufts.ac.jp>
6. 研究組織  
(1)研究代表者

芝野 耕司 (Kohji SHIBANO) 東京外国語  
大学・アジア・アフリカ言語文化研究所・教  
授

研究者番号：50216024

(2)研究分担者

藤村 知子 (Tomoko Fujimura) 東京外  
国語大学・大学院国際日本学研究院・教授

研究者番号：20229049

大津 友美 (Tomomi Ootsu) 東京外国語  
大学・大学院国際日本学研究院・准教授

研究者番号：20537073

鈴木 美加 (Mika Suzuki) 東京外国語  
大学・大学院国際日本学研究院・准教授

研究者番号：90226556

藤森 弘子 (Hiroko Fujimura) 東京外  
国語大学・大学院国際日本学研究院・教授

研究者番号：50282778

佐野 洋 (Hioshi Sano) 東京外国語大  
学・大学院国際学研究院・教授

研究者番号：30282776

望月 源 (Hajime Motizuki) 東京外国  
語大学・大学院国際学研究院・准教授

研究者番号：30282776

中村 美奈子 (Minako Nakamura) お茶の  
水女子大学・基幹研究院・准教授

研究者番号：20345408