

科学研究費助成事業 研究成果報告書

平成 30 年 6 月 15 日現在

機関番号：34418

研究種目：基盤研究(B) (一般)

研究期間：2015～2017

課題番号：15H02940

研究課題名(和文) 学習者話し言葉コーパスに基づく学習支援システムに関する研究

研究課題名(英文) Research of computer-assisted language learning system using spoken learner corpus

研究代表者

小谷 克則 (Kotani, Katsunori)

関西外国語大学・英語キャリア学部・教授

研究者番号：30440994

交付決定額(研究期間全体)：(直接経費) 10,200,000円

研究成果の概要(和文)：英語学習支援システムの様々な効用の中でもインターネットで配信される英語話者向けの素材を用いての実践的な学習による学習効果の高さは注目に値する。しかし、インターネットなどを通じて配信される素材は教材と異なり学習者への適正さが不明であり、学習効果の低下を招く恐れがある。また、これらの素材には学習者の問題点を検出するための設問などが準備されていない等の問題がある。そこで、実用的な素材を用いての英語学習支援システムによる学習支援効果を高めるべく、本研究は(1)素材の難易度を推定する技術、(2)学習者の言語運用を評価する技術の開発を行った。

研究成果の概要(英文)：Computer-assisted language learning has brought effective language learning by using the internet as learning material resource. However, the use of materials on the internet could lead to the decrease of learning motivation due to the inappropriate, uncontrolled difficulty of materials, unlike materials on textbooks. In addition, materials on the internet have no questionnaires to identify learners' problems. Thus, this study aimed to improve the effects of computer-assisted language learning. The purpose of this study was to develop (a) evaluation techniques for the difficulty of materials, and (b) evaluation techniques for learners' proficiency.

研究分野：教育工学、言語学

キーワード：学習者コーパス 英語学習支援システム 話し言葉

1. 研究開始当初の背景

国際化が進み、国内の大学生には専攻分野に関係なく「読む、書く、聴く、話す」の四技能の英語能力が必要とされていた。しかし、多くの大学新入生の英語能力が中学生程度との報告 [1] があるように、大学生の英語学習のより一層の効果促進・効率化が必須であった。このような課題への解決策として、学習者が実際にインターネットを通じ英語を運用する機会を自学自習の環境として提供できるなどの理由から、英語学習支援システムの導入が注目されていた [2]。

英語学習支援システムの教材としてインターネットで配信されるニュースなどの素材を利用する場合、問題集などの教材を利用する場合と異なり、学習者には素材の難易度は不明であり、また、どのような語彙や文法を適切に使えなかったかは判断できなかった。そこで、素材の難易度を自動的に推定する技術(以下、推定技術)と未習得要素を自動的に検出する技術(以下、検出技術)が必要であった。推定・検出技術の開発手法として、学習者の英語の運用を分析するためのデータベース(学習者コーパス)を用いた、多変量解析などの統計的分析による推定・検出手法 [3,4] があった。学習者コーパスに基づく開発手法の利点は、学習者コーパスとしてデータ収集することで、どのような教材、学習者にも対応できる点であった。一方で、この開発手法の課題は、読解コーパスや聴解コーパスの作成が容易でなかったため、対象が作文と発話に限定されてきたことであった。

こうした研究動向を踏まえ、学習者四技能コーパスに基づく推定・検出技術の研究を位置づけ、本研究開始まで、書き言葉(読解・作文)処理の研究を行い、書き言葉(読解・作文)を対象に200名規模の学習者コーパスに基づく推定・検出技術の研究 [5,6,7] に取り組みを終えていた。そして、推定・検出技術において統計的分析(重回帰分析やサポートベクターマシン(機械学習)など)に有用な要因として、母語からの直訳的表現かどうかを示す直訳性や読解・作文速度などを特定していた。

2. 研究の目的

本研究の目的は、200名規模の学習者話し言葉(聴解・発話)コーパスを作成し、聴解と発話を対象に、素材の難易度を文単位で推定する技術(図1)、学習者の未習得の語彙や文法を検出する技術(図2)の開発に取り組

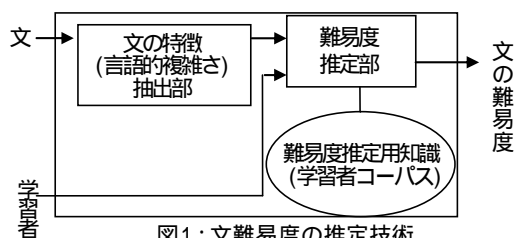


図1: 文難易度の推定技術

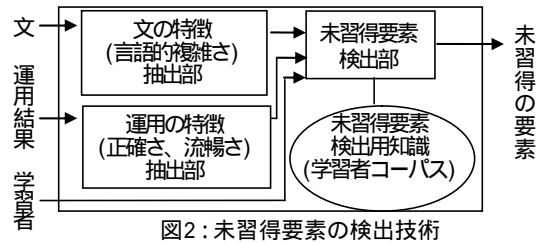


図2: 未習得要素の検出技術

むことにあった。

文難易度の推定技術では、文の特徴として自然言語処理技術により抽出された言語的複雑さと学習者により入力された学習者の特徴から、難易度推定用知識に基づき難易度推定部が文の難易度を自動的に推定する。未習得要素の検出技術では、文の特徴と学習者の特徴に加え、運用の特徴として聴解と発話の正確さ、流暢さから、未習得要素検出用知識に基づき未習得要素検出部が未習得の語彙や文法を自動的に検出する。推定用知識、検出用知識は学習者話し言葉コーパスから作成する。

3. 研究の方法

本研究は、より実践的な英語学習の実現に向けて英語学習支援用の教材の難易度を文単位で推定する技術と学習者の未習得の語彙や文法を検出する技術の研究開発を行った。まず、推定・検出技術の開発に必要な学習者話し言葉コーパス(試用版)を作成した。次に、既存の推定・検出技術を用いて、評価法の基本版を開発し、その有用性と学習者話し言葉コーパス(試用版)を検証した。そして、音声など新たな要因にも基づき推定・検出技術の発展版の開発に向け、大規模学習者コーパスを開発した。

4. 研究成果

まず、学習者コーパスの仕様などを検討し、聴覚資料の検討を行い、国際音声学会などにより作成された英語の発音を網羅したテキストを採用することとした。このテキストに対して英語母語話者による読み上げ音声を録音した音声資料を作成した。この音声・テキスト資料を通じて50名の英語学習者(大学生)を対象に聴き取り、読み上げデータを収集した。このデータに対して聞き取り・読み上げ評価用のアノテーション情報として編集距離情報を用いた。その他、学習者の習熟度や主観的な判断による難易度といったアノテーション情報も追加し、話し言葉学習者コーパスの試用版(以下、試用版コーパス)を開発した。学習者の誤用に対して決定木による推定法の妥当性を確認した。

次に、この試用版コーパスを通じて、学習者データ収集の効率化を課題として位置付け、データ収集用のアプリケーション(Android OS版)を開発した。また、試用版コーパスを用いて聴き取り・発音資料の難易度推定法を開発し、一定度の信頼性と妥当性

を確認した。

そして、話し言葉学習者コーパスのアノテーション情報として編集距離の有用性を確認できたため、大規模な話し言葉学習者コーパスの開発に取り組んだ。英語学習者 200 名（大学生）がインターネットで配信されているニュース記事を読み上げた際の発音データを基に英語母語話者による転写データを作成した。この転写データと実際のニュース記事データから編集距離を算出するためである。

今後の課題として、コーパスデータの拡張が挙げられる。本研究では学習者の発音データの転写を人手作業により行った。予備実験として自動音声認識技術と人手作業の比較を行い、学習者発音の場合、自動認識の精度低下を確認した。一定度の精度を確保し、学習者音声データに対して自動音声認識技術を利用し、データの拡張性を高める必要がある。また、学習者や資料を多角化させることで、より実践的な場面を想定し、コーパスデータを収集し、そのデータに基づき開発された評価法の信頼性と妥当性を検証する必要がある。

<引用文献>

- [1] 小野博他. 2005. 日本の大学生の基礎学力構造とリメディアル教育. NIME 研究報告. メディア教育開発センター.
- [2] Beatty, K. 2013. Teaching and Researching Computer-Assisted Language Learning. Routledge: New York.
- [3] Brooke, J. et al. 2012. Measuring Interlanguage: Native Language Identification with L1-influence Metrics. Proc. of LREC.
- [4] Izumi, E., K. Uchimoto, H. Isahara. 2004. The Overview of the SST Speech Corpus of Japanese Learner English and Evaluation Through the Experiment on Automatic Detection of Learners' Errors. Proc. of LREC.
- [5] Kotani, K., T. Yoshimi, H. Nanjo, and H. Isahara. 2013. Corpus Materials for Constructing Learner Corpus Compiling Speaking, Writing, Listening, and Reading Data. International Journal of Computational Linguistics and Applications, pp.77-92.
- [6] Kotani, K., T. Yoshimi, and H. Isahara. 2013. Application of Reading Data in an Integrated Learner Corpus. Procedia Social and Behavioral Sciences, pp.513-521.
- [7] Kotani, K., T. Yoshimi, M. Uchida, and H. Isahara. 2013. Classification of Sentence Written by Learners of English Based on Linguistic Features and Learner Features. Proc. of the 6th ICERI.

5. 主な発表論文等

(雑誌論文)(計10件)

- Katsunori Kotani, Takehiko Yoshimi, Measuring Readability for Learners of English as a Foreign Language by Linguistic and Learner Features, Computational Linguistics: Communications in Computer and Science Information, 査読有, 593 巻, 2016, 211-222.
DOI: 10.1007/978-981-10-0515-2_15
- Katsunori Kotani, Takehiko Yoshimi, Mayumi Uchida, A Corpus for Analysis of Unlearned Lexical and Syntactic Elements among Learners of English as a Foreign Language, Input a Word, Analyze the World Selected Approaches to Corpus Linguistics, 査読有, 2016, 455-460.
- Katsunori Kotani, Takehiko Yoshimi, Analysis of the Association between Reading Proficiency and Frequency of Writing Errors Using an Error-tagged Learner Corpus for Multiple Linguistic Skills, ICIC Express Letters: An International Journal of Research and Surveys, 査読有, 10 巻 3 号, 2016, 569-574.
- Katsunori Kotani, Takehiko Yoshimi, Effectiveness of Linguistic and Learner Features to Listenability Measurement Using a Decision Tree Classifier, Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications, 査読有, 2016, 6-10.
- Katsunori Kotani, Takehiko Yoshimi, Learner Feature Variation in Measuring the Listenability for Learners of English as a Foreign Language, Emerging Technologies for Education (Lecture Notes in Computer Science), 査読有, 10108 巻, 2017, 339-348.
DOI: 10.1007/978-3-319-52836-6_36
- Katsunori Kotani, Takehiko Yoshimi, Annotation of a Learner Corpus toward Development of an Error-cause Presenting Technique, Advances in Social Science, Education and Humanities Research, 査読有, 68 巻, 2017, 78-81.
DOI: 10.2991/icat2e-17.2016.19
- Katsunori Kotani, Takehiko Yoshimi, Listenability Measurement based on Learners' Dictation Performance, Proceedings of 2017 Conference of the Pacific Association for Computational Linguistics, 査読有, 2017, 357-366.
- Katsunori Kotani, Takehiko Yoshimi, Effectiveness of Linguistic and

Learner Features for Listenability Measurement Using a Decision Tree Classifier, The Journal of Information and Systems in Education, 査読有, 16巻, 2017, 7-11.

DOI: 10.12937/ejsise.16.7

— Katsunori Kotani, Takehiko Yoshimi, Word Error Rate as a Listenability Index for Learners of English as a Foreign Language, Proceedings of the 25th International Conference on Computers in Education, 査読有, 2017, 934-936.

— Katsunori Kotani, Takehiko Yoshimi, A Listenability Index Consisting of Subjective Judgment and Objective Evaluation, Proceedings of the 2017 9th International Conference on Education Technology and Computers, 査読有, 2017, 63-67.

DOI: 10.1145/3175536.3175541

〔学会発表〕(計 1 1 件)

— Katsunori Kotani, Takehiko Yoshimi, Measuring Readability for Learners of English as a Foreign Language by Linguistic and Learner Features, The 14th International Conference of the Pacific Association for Computational Linguistics (国際学会), 2015.

— Katsunori Kotani, Takehiko Yoshimi, Application of a Corpus to Identify Gaps between English Learners and Native Speakers, The 8th Workshop on Building and Using Comparable Corpora (国際学会), 2015.

— Katsunori Kotani, Takehiko Yoshimi, Design of a Learner Corpus for Listening and Speaking Performance, The 29th Pacific Asia Conference on Language, Information and Computation (国際学会), 2015.

— Katsunori Kotani, Takehiko Yoshimi, Design for a Listening Learner Corpus for Computer-assisted Language Learning, The 11th International Symposium on Natural Language Processing (国際学会), 2016.

— 小谷克則, 吉見毅彦, 英語学習者の未習得言語知識習得のための冠詞誤り原因の付与, 第 15 回情報科学技術フォーラム, 2016.

— Katsunori Kotani, Takehiko Yoshimi, Learner Feature Variation in Measuring the Listenability for Learners of English as a Foreign Language, SETE 2016: 1st International Symposium on Emerging Technologies for Education (国際学会), 2016.

— Katsunori Kotani, Takehiko Yoshimi,

Effectiveness of Linguistic and Learner Features to Listenability Measurement Using a Decision Tree Classifier, The 3rd Workshop on Natural Language Processing Techniques for Educational Applications (国際学会), 2016.

— Katsunori Kotani, Takehiko Yoshimi, Annotation of a Learner Corpus toward Development of an Error-cause Presenting Technique, 2017 International Conference on Advanced Technologies Enhancing Education (国際学会), 2017.

— Katsunori Kotani, Takehiko Yoshimi, Listenability Measurement based on Learners' Dictation Performance, The 15th International Conference of the Pacific Association for Computational Linguistics (国際学会), 2017.

— Katsunori Kotani, Takehiko Yoshimi, Word Error Rate as a Listenability Index for Learners of English as a Foreign Language, The 25th International Conference on Computers in Education (国際学会), 2017.

— Katsunori Kotani, Takehiko Yoshimi, A Listenability Index Consisting of Subjective Judgment and Objective Evaluation, The 9th International Conference on Education Technology and Computers (国際学会), 2017.

〔その他〕

ホームページ等

言語・音声メディア研究室

<http://nlp.i.ryukoku.ac.jp/index.html>

科研開発ソフト

<https://www7090ue.sakura.ne.jp:50443/smartvoice/index.html>

6. 研究組織

(1) 研究代表者

小谷 克則 (KOTANI, Katsunori)

関西外国語大学英語キャリア学部・教授

研究者番号: 30440994

(2) 研究分担者

吉見 毅彦 (YOSHIMI, Takehiko)

龍谷大学理工学部・准教授

研究者番号: 50368031

(3) 研究分担者

井佐原 均 (ISAHARA, Hitoshi)

豊橋技術科学大学情報メディア基盤セン

ター・教授/情報メディア基盤センター長

研究者番号: 20358881