

日本語を測る 言語能力測定システムの日本語教育への応用可能性について

久保 圭[†] 宮部 真衣[‡] 四方 朱子[‡] 李 在鎬^{††} 荒牧 英治[‡]

大阪大学 日本語日本文化教育センター[†]

京都大学 学際融合教育研究推進センター[‡] 筑波大学 人文社会系^{††}

1. はじめに

現在、本邦では16万人を超える外国人留学生在が日本語教育機関に所属し、日本語を学んでいる[1]。それぞれの日本語教育機関には、学習者の習熟度に対して適切な評価を与えることが求められているが、評価の難易度はその対象によって異なる。選択問題など、問いに対する解が明示的に設けられていることの多い読解試験や聴解試験に関しては、その評価が比較的容易であると考えられるが、作文課題や会話テストなど、書く能力や話す能力の評価については、テスターによる判断の揺れがしばしば客観的な評価を困難にする。また、試験実施や評価に時間を要するといった問題点もある[2]。

このような問題をふまえ、本研究では、テキストや音声から言語能力指標（語彙量、語彙の特殊性、語彙の難しきなど）を定量的に自動測定するシステムを紹介し、どのような指標を用いれば、日本語学習者の習熟度を適切に判定することができるのかについて検証することで、システムの日本語教育への応用可能性について探る。

2. 言語能力自動測定システム「言秤」

本研究で用いるシステム「言秤（ことばかり）」は、日本語で綴られた文章や日本語で話された音声から言語能力指標を自動測定する機能を備えている。この言語能力指標の値によって、語彙量、使用語彙のレベル別使用率、文章の具体性、語彙の特殊性などが測定できる。このシステムを用いた研究として、宮部(2014)では、認知症や発達障害といった、症状の進行につれて言語能力に影響を及ぼすことが知られている疾患を対象として、言秤を用いたスクリーニングの可能性について述べている[3]。各言語能力指標の詳細については次節で述べる。

Quantifying Japanese Language: Applying Language Ability Measurement System to Japanese-Language Teaching
Kay KUBO[†] Mai MIYABE[‡] Shuko SHIKATA[‡]
Jaeho LEE^{††} Eiji ARAMAKI[†]

[†]Center for Japanese Language and Culture, Osaka University
[‡]Center for the Promotion of Interdisciplinary Education and Research, Kyoto University

^{††}Faculty of Humanities and Social Sciences, University of Tsukuba

3. 言語能力指標

本研究で用いる指標は、以下における(i)～(vi)である。これらの指標は、荒牧(2014)で加齢による言語能力の変化を調査するために用いられ[4]、上述の宮部(2014)においても医学的観点から応用されている。

本研究では、これらの指標のなかから日本語教育への応用に適するものが何であるかを探る。各指標の説明について、以下では最小限の範囲にとどめるが、さらなる詳細については荒牧(2014)を参照されたい。

(i) 機能表現難度 (Difficulty of Functional Expression; FNC)

機能表現の難度を表す指標。機能表現とは、日本語における機能語と複合辞を総称したものである。難易度の定義は「日本語機能表現辞書つづじ」で設定されているものを1から5の5段階に変換した。文ごとに算出し、その平均をFNCスコアとする。この値が大きいほど、文章内における機能表現の難易度が高いことを示す。

(ii) 頻度・使用者数比 (Frequency per User Popularity; FPU)

語の特殊性を表す指標。語の特殊性の定義は、ある語の出現頻度をそのユーザ数で割った値とする。語ごとに頻度・使用者数比を算出し、全単語の値を平均した値をFPUスコアとする。この値が大きいほど、ユーザ数が少ない特殊な語であることを示す。

(iii) 日本語学習語彙レベル (Japanese Educational Lexicon Level; JEL)

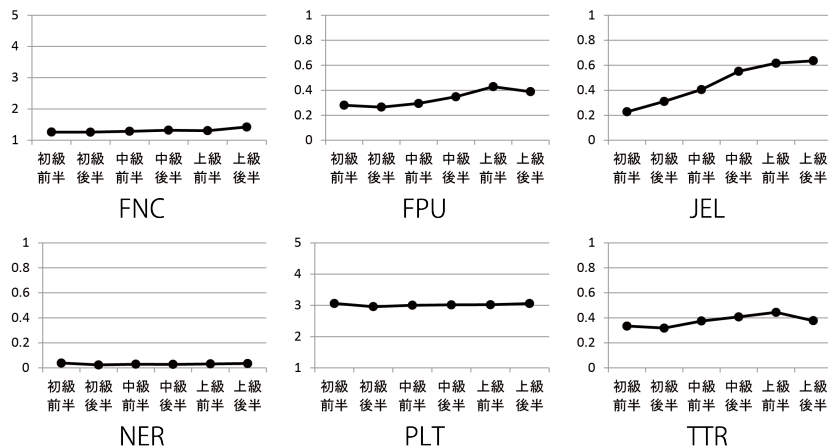
語彙の難易度を表す指標。日本語学習辞書における語彙レベルを難易度として用いた。レベルは1(初級前半)から6(上級後半)の6段階に分けられる。

(iv) 具体性 (Named Entity Ratio; NER)

文章の具体性を表す指標。具体性の定義は、固有名形態素数を全名詞形態素数で割った値とする。文ごとに算出し、平均した値をNERスコアとする。この値が大きいほど、文章の内容が具体的であることを示す。

(v) ポライトネス (Politeness of Functional Expression; PLT)

機能表現の丁寧さ(ポライトネス)の程度を表す指標。ポライトネスは「日本語機能表現辞書つづじ」の分類を用いた。この分類では常体、敬体、口語体、文語体に分



図：日本語教科書コーパスの測定結果

けられており、これを口語体=1, 常体=1, 敬体=3, 文語体=5に変換した。文ごとに算出し、平均した値をPLTスコアとする。この値が大きいほど、機能表現が丁寧であることを示す。

(vi) タイプ・トークン割合 (Type Token Ratio; TTR)

語彙量の多寡を表す指標。この指標は、Type (異なり語数) と Token (延べ語数) の比率 (Type/Token) であり、文章全体で集計した値を TTR スコアとする。この値が大きいほど、語彙量が多いことを示す。

4. 測定結果と考察

本研究では、日本語教科書コーパスから各指標を測定し、その結果を用いて指標の妥当性を検証する。

日本語教科書コーパスとは、日本国内で流通している初級から上級までの日本語教材から、販売実績に基づいて選定しテキスト化したものである。また、テキスト化した後、5段階(初級前半から上級前半)のレベルに分類した。また、上級後半レベルとして、現代日本語書き言葉均衡コーパス(BCCWJ)内の国会会議録のデータを収載した。なお、コーパスの規模は、592,674語(初級前半: 72,691語, 初級後半: 68,748語, 中級前半: 87,433語, 中級後半: 174,968語, 上級前半: 69,270語, 上級後半: 119,564語)である。各指標の測定結果を上図に示す。

日本語教科書コーパスの測定結果について、昇級にもなう順当な値の変化がみられた指標は FPU, JEL, TTRであった。FPU や JEL の上昇は、学習内容に沿ってより特殊/難解な語の使用が増えることを示し、TTR の上昇は、使用する語彙が多様になっていることを示す。これらの指標は、日本語学習者が既習の語彙を実用しているか、また、同じ言葉を多用することなく、ある物事について説明したり換言したりすることができるようになっていくかといった、習熟のプロセスを捉え得るものと考えられる。

5. 日本語教育への応用可能性

今回はテキストコーパスを用いて指標の測定をおこなったが、本システムはテキストのみではなく、音声認識を用いた音声データからの測定も可能である。これは作文課題だけではなく、会話テストの評価にも用いることができる応用性を示唆する。測定結果によって得られる学習者の習熟度や使用語彙レベルの傾向を考慮することで、教授法やコース運営の改善にも活用することが可能となる。また、試験作成の際にも、問題文に含まれる語彙の難易度が適切であるかなどのチェックにも利用することができると考えられる。

6. おわりに

本研究では、言語能力自動測定システム「言秤」を用いて、外国人留学生など日本語学習者の習熟度を適切に評価することを念頭に、どの指標を測定することがより効果的であるかについて、日本語教科書コーパスを用いてそれぞれの指標の妥当性を検証した。

今後は、新たな指標を模索するとともに、今回の分析では有用と思われる測定結果が得られなかった指標について分析を進め、改良につなげる。

参考文献

- [1] 平成 25 年度外国人留学生在籍状況調査結果, 独立行政法人日本学生支援機構, http://www.jasso.go.jp/statistics/intl_student/data13.html
- [2] 鎌田修: 日本語の会話能力とその測定・評価, 日本語学, Vol. 33, No.12, pp.16-27 (2014)
- [3] 宮部真衣 ほか: 音声認識による認知症・発達障害スクリーニングは可能か?: 言語能力測定システム「言秤」の提案, グループウェアとネットワークサービスワークショップ2014論文集, pp. 1-8 (2014)
- [4] 荒牧英治 ほか: 古いと〈ことば〉: プログ・テキストから測る老化, 信学技報, Vol.114, No.173, pp.131-136 (2014)