

外国語会話訓練システムの運用において 収集した発話データの解析

佐藤 勇貴^{†1} 松本章代^{†1} 佐伯 啓^{†1}

我々は、数年前からスマートフォンを用いた外国語会話訓練システムを開発している。本研究では学習効果を検証するための実験を行い、学習者から発話データの収集と発話内容に関する解析を行ったので、これらについて報告する。この実験では、ドイツ語の授業の全履修者を対象として本システムを4週間運用した。

Analysis of Utterance Data Collected Operation of a Foreign Language Conversation Training System

YUKI SATO,^{†1} AKIYO MATSUMOTO^{†1} and KEI SAEKI^{†1}

We have developed a foreign language conversation training system using smartphone for several years now. In this study, we conduct an evaluation experiment for verifying the learning effect, report on results of experiment. It is collection of utterance data from learners and analysis to utterance detail. The subjects are all the students in a German class, and the term is four weeks.

1. はじめに

経済産業省の調査によると、外国語会話教室に通う生徒数は2016年11月の時点でおおよそ40万人いる¹⁾。実際に外国語会話教室に通っているのは金銭面・時間面で恵まれている人だと考えられるため、実際に外国語会話を身につけたいと思っている人は40万人より多いと考えられる。

外国語会話に必要なリスニング能力を鍛えるためには毎日の訓練が必要となるが、CDなどを利用した訓練にしても、積極的に学ぶ姿勢や意欲がなければ、継続することは難しい。

そこで、我々は強制力があり、少々学習意欲に欠ける学習者でも継続して外国語会話の訓練を行うことができるシステムの開発を、数年前から継続して進めている。

本システムは、教員が指定した時間に学習者へ電話をかけ、あらかじめ用意した音声を再生するものである。実際の会話を想定して配信には発話する時間が設けられているため、会話における間の取り方、レスポンスのトレーニングが可能である。その際、学習者の発話は録音される。教員が電話をかけて半強制的に学習される仕組みになっているため、学習が継続させやすくなっている。

2. これまでの経緯と本研究の位置付け

我々は、数年前より外国語の会話訓練を目的として「教員が指定した日時に学習者へ電話をかけ、あらかじめ用意した音声を再生するシステム」を開発している。

先行研究²⁾では、Twilio APIを利用しており、電話やショートメッセージサービス(SMS)の送受信をプログラムにすることが可能になっている。また、通話には一般電話回線を利用している。昨年度、東北学院大学教養学部言語文化学科1年生のドイツ語履修者全員に対して4週間の運用実験を行っている。内容は学生に毎日1回電話をかけ音声データを再生し、電話を取り逃した場合、再配信設定をウェブページから学生自身に行ってもらおうというもので、本システムを利用することによる学習の継続性を検証したが、学習効果についてはアンケートでの調査に留まっている。また、システムのさらなる改善が課題として挙げられている。

本研究は先行研究²⁾のシステムに対し、運用に適した仕様にするため、各種機能の改善を施す。そして、言語文化学科2年生のドイツ語履修者全員を対象に1か月程度の運用実験を行い、レスポンスタイムの短縮とレベル判定という新たな基準から学習効果の調査を行う。

3. システム概要

学習者はスマートフォン、教員はPC上のウェブブラウザを用いて本システムを利用することを想定している。なお、本システムのCGI部分はRubyを用いて実装しているが、Twilioの制御はsinatraというフレームワークを使用している。

教員用メニューは以下の構成となっている。

(1) 配信設定機能

^{†1} 東北学院大学教養学部
Faculty of Liberal Arts, Tohoku Gakuin University

- (2) 学生のグループ登録機能
- (3) 学習者の学習履歴確認機能

(1) は教員がウェブブラウザから配信の日時、配信する音声データを指定する機能である。配信設定はタスクスケジューラに登録され、指定した時間になったら学習者に電話をかけ音声を再生する。

(2) は配信を行う際のグループを登録する機能である。グループを登録することによりグループ単位での配信設定を行うことができる。グループで設定した場合の処理については3.1節で述べる。

(3) は教員が学習者の学習記録を一覧で確認することができる機能である。学習者の学習記録については3.4節で述べる。

学習者用のメニュー画面を図1に示す。

学習者用メニューは以下の構成となっている。

- (1) 学習者情報の登録機能
- (2) 再配信機能
- (3) 自分の学習記録確認機能

(1) は学習者が配信時に使用する電話番号や学生番号、氏名、配信希望時間帯を登録する機能である。登録画面を図2に示す。学習者が必要な項目に記入し、登録ボタンを押すと記入されている電話番号宛にパスワードが記入されたSMSが送信される。学習者はそのパスワードを入力することで登録が完了する。

(2) については3.2節で述べる。

(3) は自分の学習記録について、何時に電話がかかってきたか、電話に応答したかを確認することができる。

本システムの音声配信は、音声データと音声データの間に学習者が発話する時間を設けている。学習者は音声と会話する感覚で発話することになっており、学習者の発話は録音されることになる。また、発信終了時にSMSを送信することができる。再配信設定をするウェブページへのURLを学習者にSMSで送信することで、学習者は電話に回答できなかった場合、スムーズに再配信設定を行うことができる。

3.1~3.4では、今年度追加した機能・改善した機能の一部について説明する。

3.1 グループ配信機能の改良

昨年度のシステムではグループで設定した場合、学習者との通話が終わり次第順次、次の学習者に電話をかけるように配信設定をタスクスケジューラに登録されていた。今年度から



図1 学習者用のメニュー

は1分おきに順次、次の学習者に並行して電話をかけられるように改良を行った。この改良によって、配信人数や音声データの長さへの依存度を抑えた配信システムを構築することができた。

3.2 再配信機能の改良

本システムでは、電話に出ることができなかった学習者がもう一度電話を受けるための再配信機能が実装されている。昨年度のシステムでは一日に一回まで再配信設定を行うことができるように管理されていたが、一日という範囲指定では各配信ごとに区別することができていなかった。そこで、今年度からは各学習者の配信希望時間帯ごとに、配信希望時間帯の開始時間から24時間の範囲で再配信設定ができるように改良した。

3.3 レベル判定機能の追加

本システムでは、学習者ごとに適した難易度の音声データを配信することで学習者の内容理解を促し、「適切な間で話し始められること」、「たくさん話せること」の2つを達成することを目的として、レベル判定機能が実装されている。今年度は学習者の発話レベルを判定する部分に焦点を当てて実装を行った。

発話レベルについては、Twilio通話での録音記録がURLとして発行されるため、その

学生情報登録ページ

必要事項を記入して、送信ボタンを押してください。

学生番号(半角英数7文字)

氏名

電話番号(ハイフン無し)
ex)090-1234-5678 → 09012345678

連絡先(メールアドレス)

配信希望時間帯
 21:00~21:30
 22:00~22:30
 23:00~23:30

図 2 学習者情報の登録画面

URL から wav 形式でダウンロードすることで得られる発話データから判定を行う。また、発話データの分析方法として音声認識を採用した。音声認識は、Google Cloud Speech API というサービスを利用したものであり、発話データのテキスト化を行うことで、テキストに対する単語数と信頼度を抽出することができる。

判定については、はじめに基準となる単語数を事前に教員が指定し、発話データとの比較をすることで、基準単語数に応じた段階に分類を行う。その際、信頼度が一定値を下回った場合、テキスト内容に誤りがある可能性が高いため、基準となる単語数との比較を行わないようにする。次に各段階を数値化して配信日ごとに平均を求めることで各配信日の判定結果を出し、最後に各配信日の判定結果について平均を求めることで、総合としての判定結果を定める。

3.4 学習者の通話記録を保存する機能

学習者に電話をかける際に何時に電話をかけたか、学習者が電話に応答したか、どの音声データが配信されたかを通話記録として保存する機能である。また、学習者が電話に応答し、発話をした場合は発話内容を保存する。

本年度から教員用の学習履歴確認機能に、音声認識を利用して、発話データの内容をテキ

スト化して確認できる機能を追加した。

4. 運用実験

4.1 学習効果・レベル判定機能の検証実験

学習効果・レベル判定機能の検証実験は、本システムの利用が発話の訓練として有効かどうかをレスポンスタイムに着目して調査することとレベル判定機能の妥当性を検証することを目的としている。レスポンスタイムの分析結果と実験終了後のアンケートを集計し、レベル判定結果の評価を行うことで、学習効果の調査とレベル判定機能の検証を行う。

実験方法は以下の通りである。

- ドイツ語の授業を受講している 17 名を対象に行う。
- 期間は 28 日間 (4 週間) とし、毎日 1 回電話をかける。
- 配信には 2 種類の形式がある。
 - 1 つ目は音声データを 10 個再生し、音声データ間に学習者が発話を行い、全データの再生が終わり次第通話を終了するというものである。発話した内容は録音される。こちらの配信形態は最初の 7 日間のみ行う。
 - 2 つ目は音声データを 11 個再生し、音声データ間に学習者が発話を行い、全データの再生が終わり次第通話を終了するというものである。発話した内容は録音される。なお、11 番目に再生する音声データは 5 種類あり、それぞれ 3 日間ずつ配信した後、ランダムに選択される。また、上記の 7 日間以外はこちらの配信形態で配信を行う。
- 電話を取り逃した場合には、再配信設定をするウェブページから学生自身が日時を決め、再配信の設定をしてもらう。
- レスポンスタイムは録音された音声から話し始めるまでの間以外の無音を取り除き、学習者が録音開始から何秒後に話しているか計測することで取得する。

4.2 レスポンスタイムの分析

1~10, 1~6, 7~10 問目への発話データについて、レスポンスタイムの平均を取ったものと 11 問目も含めたものの実例を図 3, 図 4 にそれぞれ示す。このようなグラフを学習者 17 人に対して作成し、レスポンスタイムの短縮がみられるかを判定する。ここで示したグラフでは、11 問目を含めた場合、値が全体的に上下しているが、11 問目を含まない場合、1~6 問目について、値が安定しているうえに最後の 4 日間は短縮がみられ、7~10 問目についても、最初は値が上下していることがわかるが、後半から低い値で安定してきているこ

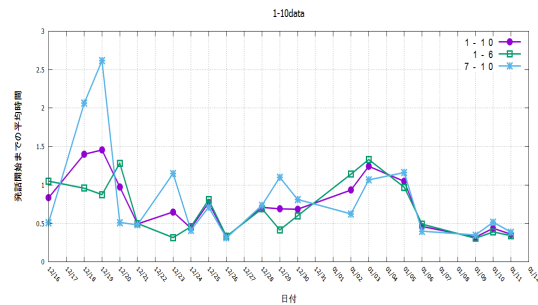


図 3 1～10 問目に対する平均レスポンスタイムの実例

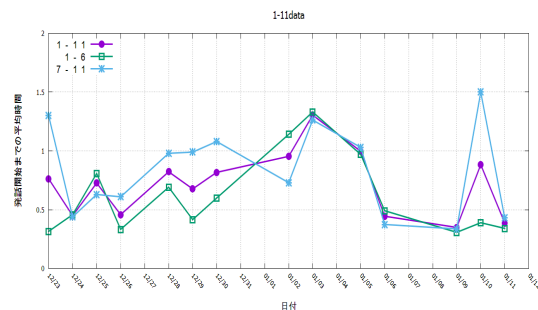


図 4 1～11 問目に対する平均レスポンスタイムの実例

とから、1～10 問目についてレスポンスタイムの短縮が認められるという判定結果となる。

以上の要領で判定を行ったところ、17 名中、レスポンスタイムの短縮がみられた学習者が 12 名、短縮がみられなかった学習者が 3 名、母数が少ないことから判定を行わなかった学習者が 2 名という結果となり、70.6 % の学習者に対して、発話の訓練としての学習効果が認められるという結論に至った。

4.3 レベル判定機能の検証

本実験では、3 つのクラスについて基準となる単語数を設定しておき、発話データとの比較をすることで、学習者を発話レベルの観点から 3 段階に分類を行った。ここでは、3 つのクラスを便宜的に ABC クラスと表現する。各クラスの定義は以下のように設定している。

- A：教員側が期待する単語数に達している。
- B：回答を構成する上で最低限の単語数を有している。

- C：単語数が明らかに不足している。

なお、本実験においては、単語数設定の際に定義をより具体的なものにして設定を行った。その定義は以下のとおりである。

- A：文章として成立している。
- B：名詞のみだが回答としては成立している。
- C：回答になっていない。

信頼度については、言語文化学科の佐伯ゼミ 3 年生 7 名を対象に 2 週間の予備実験を行い、収集した発話データの音声認識結果を参考に、基準となる信頼度を設定した。その際、基準信頼度を下回った発話データが全体の 13.0 % だったことから、2 年生を対象とした場合、13.0 % 以上の発話データが基準信頼度を下回ることが予想された。そこで、基準信頼度を下回った発話データは B クラスと判定し、数値化を行う際に、B クラスのデータ数が増える分 A クラスと C クラスの数値を絶対値の大きい値に設定することで、できるだけ正確な判定が行えるように調整を行った。

以上の内容を踏まえて、学習者ごとにレベル判定を行ったところ、17 名中、A クラスが 7 名、B クラスが 8 名、C クラスが 2 名となった。

次に上記のレベル判定結果について、学習者に対する教員側の評価との関係性を調査した。調査方法は、ドイツ語担当教員にレベル判定結果に対する評価を依頼するというものである。評価内容は以下のようにになっている。

- 学習者ごとに、判定結果と普段の授業の様子に基づく評価との関係性を (1) ～ (4) で提示する。
 - (1) 一致している。
 - (2) どちらかといえば一致している。
 - (3) どちらかといえば一致していない。
 - (4) 一致していない。
- 判定結果全体の評価を記述してもらう。

以上の評価内容をドイツ語担当教員に依頼したところ、1 つ目の評価結果は (1) が 7 名、(2) が 8 名、(3) が 0 名、(4) が 2 名となり、88.2 % の学習者に対して、判定結果と教員による評価との一致が認められた。

また、2 つ目の評価結果については、「概ね普段の授業における評価と一致していると言えるが、授業でのリスボンや筆記試験の成績が悪い受講生の中にも、このシステムでの評価は B クラスに入っている学生が 2 名ほどいたのが興味深い。逆に、筆記試験の成績はトッ

表 1 運用実験における学習効果についての実感
単位：人

項目	あり	少し	なし	有効件数
聞く	6	10	0	16
話す	3	12	1	16

プラスでもこのシステムでの応答力はもう一息という例も見られた。」という評価をいただいた。

以上の結果から、本研究における学習者の発話レベルの判定に焦点を当てたレベル判定機能は、妥当性が十分に確保された機能であると実証することができた。

4.4 アンケート調査

表 1 では運用実験終了後に行われた「28 日間のシステムによる学習を通して、システム利用前と比べて『聞き取れるようになった』、あるいは『質問にスムーズに回答できるようになった』という実感は得られたか」という問いに対してのアンケートの結果を示している。

表 1 から、聞き取れるようになった、質問にスムーズに回答できるようになったと感じる学生がそれぞれ多く、概ね学生は自分の会話能力が向上したと感じていることがわかった。

5. ま と め

一般電話回線を用いた外国語会話訓練システムに機能を追加し、機能の改良を行った。運用実験では学習者の 70 % にレスポンスタイムの短縮が認められたという結果から、本システムを利用することによる学習効果の向上が確認できた。

また、レベル判定機能について学習者の発話レベルの判定結果が、教員側の評価と 90 % 近く一致したことから、妥当性が十分に確保されていることを実証できた。

今後は、さらなるシステムの改善、ドイツ語の品詞解析を行えるツールを利用し、発話内容を文法レベルで判定する機能の実装とその検証実験を検討している。

参 考 文 献

- 1) 経済産業省白書特定サービス産業動態統計調査, <http://www.meti.go.jp/statistics/tyo/tokusabido/result-2.html> (2017.01)
- 2) 松本章代, 蜂谷僚哉, 佐伯啓: 音声通話・SMS を利用した外国語会話訓練システムの構築と運用, 情報処理学会 コンピュータと教育研究会 133 回研究発表会 (2016.02)