

音声認識で ZOOM に字幕を表示。 保存したログを分析し、 プレゼン指導に活用する

オンライン授業で画面に字幕を表示すれば、聴覚障害者などにとって理解の助けとなる。この授業では教員自身の音声情報取得への不安を解消するために、Zoom の授業で話者の音声を認識してリアルタイムに字幕表示するように設定した。それを活用し、保存された発言内容のテキストを分析して、学生の話し方の癖などを可視化したところ、プレゼンの改善指導に効果を発揮したという。



杉中 拓央
早稲田大学 非常勤講師

コミュニケーションの障害を補うために Zoom の発言にリアルタイム字幕を表示する

この授業は 3 年以上を対象にした専門ゼミで 3 名の学生が履修している。2020 年度より新たに担当することになった杉中講師には聴覚障害があり、オンライン授業で初対面となる学生たちとのコミュニケーションに若干不安があった。「オンラインということもあり、十分聴き取れないために教育できない・互いに打ち解けにくいという状況を懸念して、念のために字幕を表示することにしました」。

利用したのは、Web 上で無料公開されている speech-to-text-webcam-overlay というサイトだ。PC のオーディオ設定をしたらウェブブラウザ上にこのページを表示しておく、Zoom の会話音声は自動認識され、字幕として映像に重ねて表示される。また、OBS Studio というフリーソフトを併用することで、スライドや字幕、カメラの映像など、Zoom の表示画面を任意にレイアウトできる。

気になる認識精度については、使用するマイクによってかなりの違いがあることが分かった。「PC に内蔵されているマイクでは口との間に物理的な距離があるために、精度がかなり悪く実用的ではありませんでした。そこで、学生に超指向性のピンマイクを支給して使ってもらったところ、8 ～ 9 割は正しく認識されるようになりました」。この状態で入力された音声であれば、あくまで杉中講師個人の感覚ではあるが、目の前でマスクをした状態で話されるよりもむしろ聴き取りやすくなるという。

発言のログを分析し、 話し方の改善に役立てる

基本的には自分用として字幕を表示していたが、せっかくログが残るのであればこれを活用したいと考え、保存したテキストを使って KH Coder というフリーソフトで分析してみることにした。これはテキストマイニング、すなわちテキスト型のデータを統計的に分析するためのソフトだ。対象となるテキストの中に頻出する語句、言葉のつながりなどをデータとして目に見える形で表示することで、言語による表現方法を客観的に見直すことができる。

この授業では毎回学生に 20 分程度のプレゼンの練習を行わせている。学生に使用目的を説明してデータ利用の許諾をとったうえで、全体のログからこの部分を抜き出して分析作業を行い、1 ～ 2 週間後に結果をフィードバックした。

「どんな言葉を何回使ったかなどが前後関係の関連やランキングから分かるので、つい多用してしまう口癖なども一目瞭然です。本人も自覚していたことが可視化されて納得するケースもあれば、気づいていなかった癖が見つかることもあったようです」。

分析されたデータをいっしょに眺めていると、学生からは「音声認識されやすい話し方というのは、万人に聴き取りやすい話し方なのかもしれない」という感想が聞かれた。「認識されやすいように意識すると、自然とゆっくりハッキリ話すようになります。それは、就活で面接官に好印象を与えることにもつながりそうだという声もありました」。

また、音声認識ではときに同音異義語が誤変換されることがある。「誤変換されやすい言葉というのは、人間にも間違っただけで聴き取られる可能性があるから避けたほうがいいという意見も出ました。学生たち自身でいろいろな気づきを得られ、従来のプレゼン指導にはない効果があったかなと思います。将来的には言語の専門家の力もお借りして、係り受けなど具体的な改善点を示すプログラムなどが作れるといいですね」。

ひとりに対する支援が同時に 他者にも役立つ仕掛けをつくりたい

このテキスト分析は、Zoom の発言をリアルタイムで字幕化する以外に、全体を録画や録音したデータを後からテキスト化して使うこともあった。その際使っているのは、recording-studio という、やはり無料で利用できるサイトだ。マイクから入力した音声をリアルタイムに文字化だけでなく、PC のオーディオ設定をすることで、音声データを再生すると同時にその音を拾って音声認識させることもできる。「あらかじめ音声データになっているときは、ノイズを除去したり、周波数を変換したりするなど多少編集しておく、さらに認識精度が高まります」。

どちらの方法でも、会話のログのテキストデータが自動生成されるという点ではさまざまな使用法が期待できる。「最近は教授会なども Zoom で行われることもあり、マイク等の環境設定が大切ですが、議事録作成など音声認識を使うと楽になると思います」。

自身の研究テーマである特別支援教育は、すべての教育に通じるともいわれている。「障害の有無では量れない、感覚が人より敏感な人など多様な状況の人がいます。そのすべてを包括できるような教室や教材の作り方、教育スタイルなど、インクルーシブな教育環境をいかに実現するかというところから、今回のアイデアも出てきました」。

今回の試みは、情報支援と他の参加者への学習効果を両立させたことに意義があると考えている。「字幕のみならず、こうした支援には多くの方の協力があり、それなりに手間もかかります。他の参加者にも副次的な効果をもたらすというメリットがあれば、結果として、周囲の支援行為に対する興味関心・理解にもつながるのではないかと期待しています」。