

聴覚障害学生支援における音声認識を活用した通訳システムの構築

——利用者の観点に基づいた字幕呈示の検討——

*松崎 丈・*藤島 省太

Developing interpreting system using the automatic speech recognition for deaf and hard of hearing students: The Real-time captioning from the user's point of view

MATSUZAKI Jo and FUJISHIMA Shota

要旨

近年、聴覚障害学生支援において、音声認識を活用した通訳システムが新たな技術として導入されてきている。音声認識は、話者の音声情報をほぼ全文に近い内容に通訳できる技術として注目を集めており、聴覚障害学生の入学が増加している宮城教育大学でも情報保障の質的向上のために導入した。筆者らは、聴覚障害学生支援とは利用者の観点に基づいて支援のあり方を検討しながら実践していくものと考えており、本稿では、講義で音声認識を活用した通訳システムを実験的に導入して行うとともに、利用者である聴覚障害学生にとって「読みにくさ」「理解しにくさ」を与える字幕呈示とはどのようなものであるのか、この字幕提示をどのように改善していくのかを実践的に検討した。

Key words : 聴覚障害学生 (Deaf and hard of hearing students)
音声認識 (Automatic speech recognition)
通訳システム (Interpreting system)
利用者の観点 (User's point of view)
字幕呈示 (Real-time captioning)

I はじめに

現在、高等教育機関に在籍する聴覚障害学生の人数は、聾・難聴あわせて1,326名いる(日本学生支援機構, 2008)。一般に、聴覚障害学生は、補聴器や人工内耳を装用しているが、聴者と同等のレベルで聴覚を活用することは困難である。宮城教育大学でも、2008年度の時点で聴覚障害学生が7名在籍しており、そのうち6名は聴覚活用のみで授業や学校生活における音声情報を獲得することに困難がある(藤島・松崎, 2007)。

そこで、聴覚障害学生が授業の音声情報を獲得することを保障するための支援方法として、支援者が話者の音声情報を文字や手話に変換して聴覚障害学生に伝えるという情報保障を行う。情報保障による支援は、聴覚障害学生が、授業を通して健聴学生と対等に専門的な知識、技能、慣習を学ぶ機会を保障し、それによって地域社会や企業・機関に必要とされる人材に成長していくことを目指している。高等教育機関としても、あらゆる学生が地域社会で雇用されるように育成しなければならないため、学生の個々のニーズに適した

* 特別支援教育講座

サービスを提供するのは当然のことであるといった考えに基づいて、障害学生支援に真摯に取り組むことは重要であろう。

宮城教育大学では、従来、手書きノートテイクによる情報保障を中心に行ってきたが、近年、聴覚障害学生から話者の音声情報をより多く獲得したいという要望が出されるようになった。そもそも手書きノートテイクは、話者の音声情報の20%程度しか文字化できないという技術的な限界があり、情報保障の質的向上が必要である。また、ノートテイクを行う支援者の人数がいつも不足しており、かつ近年は聴覚障害学生の入学が増加していることから(藤島・松崎, 2007)、支援者の量的確保が急務の課題になっている。

そこで、前述の情報保障の質的向上や支援者の量的確保の課題に対し、1つの解決策として、近年聴覚障害者支援の最新技術として注目されている音声認識を活用した通訳システム(以下、音声認識通訳システム)を導入して実施することとした。

音声認識は、人間の音声文字に自動変換する技術である。手書きノートテイクやパソコンノートテイクでは、それぞれ話者の音声情報の20%、40~80%程度文字化されるが、音声認識では要約からほぼ全文までの文字化が可能である(三好, 2008a)。それゆえ、音声認識は、話者の音声情報の命題内容だけでなく、モダリティ(話し手の判断や心的態度)を表す細かな言語表現も文字化することが可能になるため、聴覚障害学生に「臨場感あふれる(中野・牧原・金澤・菊池・黒木・井野・伊福部・福島, 2006)」形で伝えられることも期待される。

ただし一方で、音声認識を活用して音声を100%正確に認識して字幕化できるわけではなく、誤認識が多少生じていることも現状である。誤認識の発生は、音声認識ソフトのデータと、実際の認識すべき音声の間に、何らかのずれがあることに起因している(古井, 1998)。そのずれの原因には、①音声(声質、発声速度、話し方)の個人差が極めて大きいこと、②部屋の反響を含む種々の雑音が入ること、③自発的な話しことばに特有のさまざまな現象があること(例えば、助詞の脱落、言い淀み、言い直し、間投詞、倒置など)が挙げられる(古井, 1998; 田中, 2001)。例えば、話者が「この魚はいくらですか」と話すと、音声認識では「この坂俳句らですか」のように誤った認識結果が出

る。音声情報を獲得できない聴覚障害学生にとっては、この認識結果は誤認識のようだと判断できても、正しい情報を推測することは不可能である(三好, 2008a)。

そこで、音声認識通訳システムについて先駆的に取り組んでいる筑波技術大学(三好・黒木・河野・白澤・石原・小林, 2007)や群馬大学(菊池・金澤・中野・黒木・井野・伊福部・堀, 2004)では、上記のような音声認識の問題点に対して次のような2つの方策をとっている。1つは、復唱者が事前トレーニング(エンロール)と復唱のトレーニングの両方を行うことで認識の精度を向上させること、もう1つは、1~数名の修正者をおいて復唱者の音声入力で生じた誤認識を修正することである。黒木・井野・中野・加藤・渡邊・堀・伊福部(2003)の実地研究によると、訓練を受けた復唱者に同時復唱させることで、認識精度は90%程度で文字化されたが、これに修正者4名が同時修正を行うことで、話者の音声情報発信から平均11秒後に、97.2%の精度で字幕を表示することができたとの報告がある。

以上から、手書きノートテイクやパソコンノートテイクよりも音声情報を多く文字化できるという「最新技術」を無批判に受け入れ、音声認識ソフトのみを安易に導入するべきではないだろう。そして、音声認識を活用した通訳には、復唱・修正の作業を組み合わせる必要がある。支援者は、音声認識技術の特徴と限界を理解し、この限界を補って精度を高めるために復唱・修正作業の技術を習得していくことが求められる。ただし一方で、日本の教育機関における音声認識通訳システムの導入実績は非常に少なく(三好・黒木・河野・白澤・石原・小林, 2007; 菊池・金澤・中野・黒木・井野・伊福部・堀, 2004)、音声認識ソフトの技術的限界を補うための復唱や修正作業のノウハウは開発されているものの、現行の音声認識通訳システムで産出される字幕がはたして利用者にとって情報保障のツールとして使いやすいものになっているかについては不明な点が多い。

そこで、本稿では、筑波技術大学や群馬大学の取り組みを参考にして宮城教育大学に音声認識通訳システムを導入した上で、その音声認識通訳システムで作成した字幕が利用者である聴覚障害学生に「読みにくさ」「理解しにくさ」を招く障害状況がどのように生

じているのか、どのような字幕呈示が求められるのかを実証的に検討することを目的とする。

II 音声認識通訳システムの概要

まず、宮城教育大学で導入した音声認識通訳システムの概要を説明した上で、その通訳を実施するなかでどのように利用者にとって使いやすい情報保障になっているのかを検討する手続きや方法を述べる。

音声認識通訳システムを実施するために、図1のようなシステムを作った。このシステムの準備や作動確認を行うためには10～30分程度の時間を要する。そのため、別室で復唱・修正作業を行うようにセッティングしておき、講義室の教員の音声や聴覚障害学生が見る字幕はインターネット等で送受信する方法がある。しかし宮城教育大学では、初めての導入でありお互いの状況を確認して対処できるように全員が講義室に集まって行うことにした。図1内で番号がついている機器の詳細を表1に示す。復唱作業は、事前トレーニングを済ませた宮城教育大学教務補佐員（障害学生支援コーディネーター）1名と学生1名が担当して15分程度で交替することとし、修正作業については学生2名が連携して行う。

音声認識通訳システムに用いるソフトウェアについては、市販の音声認識ソフトウェア「ドラゴンスピーチ 2005 Professional」、字幕修正作業用のソフトウェア「IPtalk9t（下記 Web より無償ダウンロード可能：<http://iptalk.hp.infoseek.co.jp/>）」、音声認識ソフトウェアで認識された字幕データを自動的に修正用パソコンや表示用パソコンに送信するソフトウェア「SR-MODULE_ver.5」を使用する。「SR-MODULE_ver.5」は、筑波技術大学の三好・黒木・河野・白澤・石原・小林（2007）が独自に開発、今回の導入のために無償提供されたものである。

音声認識通訳システムの作業過程については次の通

表1 音声認識通訳で用いる機器

No.	機器	メーカー・製品
1	音声遅延装置	B. U. G Videobox
2	ポータブルマルチミキサー	audio-Technica AT-PMX5P
3	ヘッドホンアンプ	audio-Technica AT-HA20
4	マスク型マイク	TALK TECHNOLOGIES SM300
5	サウンドブラスター	クリエイティブメディア(株) SB-DM-PXV
6	ヘッドホン	Zennheiser HDA200

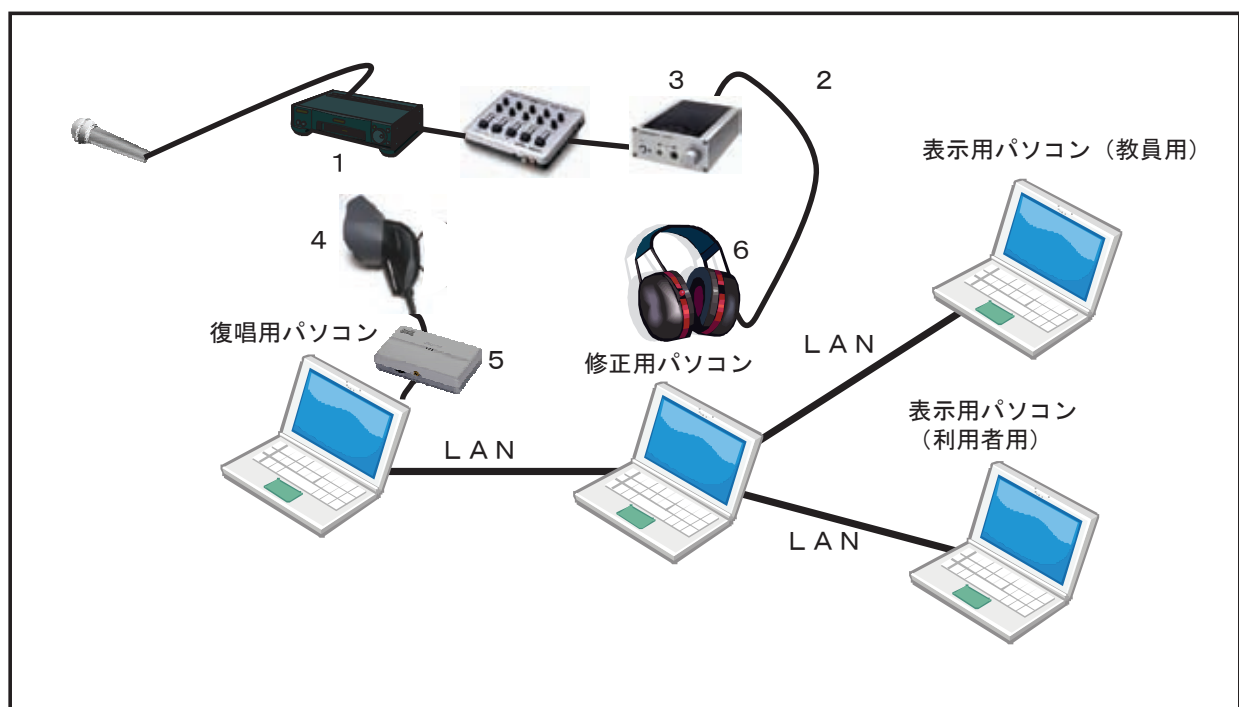


図1 音声認識通訳システムの概要

りである。

まず、教員が音声を発信すると、復唱者が図2のマスク型マイクで復唱して復唱用パソコンに音声入力する。復唱作業では、周囲の雑音が入らないようにマスク型マイクを口周辺に押し付けて覆うようにし、(1)抑揚のない淡々とした話し方、(2)パ行、マ行、イ列、エ列といった不明瞭な音になりがちな発音もあるため口を大きく明確に動かすこと、(3)タ行やパ行の無声音や語尾では大きめに発音すること、といった通常の発話とは異なった復唱(三好, 2008b)を行う。また、助詞の脱落、倒置などが含まれる話しことばを書きことば風に言い直すことも行う必要がある(安藤, 2006)が、現時点でどのように言い直せばよいのか具体的な方策となるものがなく、今回は復唱作業を担う支援者が各自これまでの通訳経験をもとに言い直すことにした。復唱作業のあり方については稿を改めて検討したい。



図2 マスク型マイク

次に、修正者が、復唱用パソコンから送られた字幕データの誤認識について、教員の音声を4秒程度遅延させて送られた音声情報を聞きながら修正用パソコンで修正する。この音声遅延は、(株)B.U.G.の音声遅延装置 Videobox を用いる。修正方法については、菊池・金澤・中野・黒木・井野・伊福部・堀(2004)の取り組みを参考に、次のように修正することとした。(1)語句が抜けているところは挿入する、(2)複数の語句が1つの語句のように誤認識されたものを修正する、(3)助詞の脱落については、文脈から明らかに誤解を生じさせ、かつ正しい内容の推測が不可能な場合に修正する、(4)漢字の誤変換(例えば、誤:不戦を→正:付箋を)や単語の誤認識(例えば、誤:こと→正:音)の

ように正しい音情報を推測できる誤認識については、修正作業の負担がかなり軽減されている状況であれば修正する。

このような作業過程で話者の音声情報は、例えば以下のような内容で字幕化されていく。こうして修正された字幕結果は、聴覚障害学生や教員の目前に設置された表示用パソコンで呈示される。

【話者の音声】

実際に測定すると皆さんの場合も右の耳と左の耳が全く同じという人は少ないと思います。って、そのゼロの方が、えー、本当は、えー、音があるんだけど、皆さんがきっと初めて聞こえたところ感じる点になるということです。

【復唱作業によって認識された字幕結果】

実際に測定をすると、みなさんの場合も、新富右と左が全く同じという人は少ないと思います。ゼロというのは、本当は、夫があるのだけれども、皆さんが聞くと、初めてことが聞こえたというところです。

【修正作業によって修正された字幕結果】

実際に測定をすると、みなさんの場合も、右と左が全く同じという人は少ないと思います。ゼロというのは、本当は、音があるのだけれども、皆さんが聞くと、初めてことが聞こえたというところです。

III 利用者の観点に立った字幕呈示の検討

音声認識通訳システムの実施は、第二執筆(以下、「F」)が担当する講義「聴覚・言語障害の心理・生理・病理」(2008年4～7月,全15回)で行う。本講義では、特別支援教育講座2年次学生が履修対象で、聴覚や言語に障害のある幼児、児童又は生徒の心理等についてパワーポイントで作成したスライドを投影しながら説明している。この講義を受講する聴覚障害学生は、自力で音声情報を獲得することは困難であり、講義で手書きノートテイクを利用している。

音声認識通訳システムで利用者の観点に基づいた字幕呈示になっているのかを検証するために、聴覚障害者であり10年程の情報保障経験を有する第一執筆(以下、「M」)が、受講している聴覚障害学生と字幕呈示を見るとともに、聴覚障害学生、復唱者、修正者、教員の行動を観察したり講義中修正作業を行うノートパソコンの画面をデジタルビデオカメラで収録した。

また、適宜、講義後で聴覚障害学生、復唱者、修正者に音声認識通訳システムについて感じたことをインタビューした。特に、聴覚障害学生へのインタビューについては、講義中の聴覚障害学生の行動や字幕呈示について気になった事柄を中心に質問することとし、例えば、「正しい意味が推測できる漢字の誤変換は修正しないままにしましたが、その誤変換について一瞬戸惑うようなことはありましたか？ それは具体的にどんな誤変換のところでしたか？」「1、2語で字幕が出て改行が多いように思いましたが、それで読みにくかったり一時的に意味が把握できないようなことは生じていましたか？」のような内容で質問した。その上で、聴覚障害学生とは字幕呈示について利用者の観点からどのような障害状況に直面したか、どのような解決策が考えられるのかを検討した。

聴覚障害学生との検討で考えられた解決策は、復唱・修正を担当する支援者に提案し、支援技術の程度に応じながらどのように対処していくのかを協議して実施することとした。

実際に講義で音声認識通訳システムを行った回数は9回分であった。支援者にとって当初、復唱・修正作業は不慣れで作業の滞りが時々みられたが、回数を重ねるにつれて大きな支障なく実施できるようになり、支援者同士がそれぞれの作業の不十分さを補うように連携する様子もみられた。復唱・修正後の認識精度については、英数字や特殊な専門用語が使われる字幕を除いて話者の説明内容や意図が把握できるレベルの字幕結果が得られていた。この認識精度の割合や誤認識の種類・発生率等について詳しく検討する必要はあるが、本稿の目的とは異なるためここでは行わない。一方で、講義中の行動観察や聴覚障害学生との協議等の結果から、受講している聴覚障害学生にとって「使いにくい」障害状況が生じており、かつこれらは当初筆者らが想定していなかったものであることが明らかになった。1つは、表示用パソコンに呈示する字幕作成のあり方であり、もう1つは、聴覚障害学生に対する字幕呈示のあり方であった。

まず、前者の字幕作成のあり方について、聴覚障害学生にとって「使いにくい」障害状況を引き起こした主な字幕例を紹介し、この字幕例に対してどのような対処を検討したのかを報告する。

IV 表示用パソコンに表示される字幕作成に関する検討

1. 句読点の付加に関する問題

本稿の音声認識通訳システムで用いた音声認識ソフトウェア「Dragon Naturally Speaking 2005 Professional」では、句読点の自動挿入機能がついている。従来の音声認識ソフトでは「てん」「まる」と発音して句読点を挿入するが、このソフトウェアでは、句読点を置きたい箇所にポーズを入れたり動詞の過去表現（例えば、「見えた」）の音声を入力したりすると自動的に挿入される仕組みになっている。ただし、意図的にポーズを入れても句読点が挿入されない場合もあり、句読点を挿入するための発話調整が難しいことが課題である。

そのために音声認識通訳システムでは、次の字幕例1と2のような字幕結果が呈示されていた（行番号は筆者による）。なお、聴覚障害学生が見る表示用パソコンでは、以下の各行の字幕結果が1行ごとに約1～5秒の間隔で順次表示される。

字幕例1 句読点の過度な付加（下線部参照）

- 1 その上の切り離しと、
- 2 下の切れ端が、
- 3 分かれているのか、
- 4 くっついているのか、
- 5 を調べるのが、
- 6 ベンベツといいます。

字幕例1では、聴覚障害学生は、一行ずつ呈示される字幕結果を見るだけでは文全体の内容が把握できない。そこで、文全体が表示されてしまってから、不要と思われる読点を削除し、「その上の切り離しと、下の切れ端が」あるいは「…を調べるのが」のいずれがその文の主題となるのか判断する必要になる。

字幕例2 句点と読点の置換（下線部参照）

- 1 左の耳に、
- 2 到達する音の時間差
- 3 を
- 4 みなさんの頭はちゃんと計算しているのです。
- 5 すごいことですね。

- 6 生まれてすぐの赤ちゃんはそれは難しいです。
- 7 何か音がするというのはわかりますが。
- 8 どこで、
- 9 音が鳴っているのか、
- 10 それができるのは、
- 11 そのあとの段階
- 12 です。

字幕例2では、「何か音がするというのはわかりますが」の直後に句点があるため、その直前の句・節との関係があるだろうと仮定して読みなおそうとする。しかしながら、その直前の「… 赤ちゃんはそれは難しいです。」の「それ」は、「到達する音の時間差を計算する」ということを指しており、先の仮定ではおかしいことに気づく。そこで、7行目の「何か音がするというのはわかりますが」の直後にある「どこで、音が鳴っているのか、それが分かるのは、そのあとの段階です」の部分と関連があると仮定して読んでみると、ここでようやく句点ではなく読点が正確であるらしいことに気づく。

このような字幕例は、手書きノートテイクやパソコンノートテイクでは生じ得ないものである。聴覚障害学生にこの字幕例の問題点について聞いたところ、「多少の読みづらさはあるが、慣れれば大丈夫かなと思う」とのことであった。しかしながら、講義「聴覚・言語障害の心理・生理・病理」の内容は、回数を重ねていくにつれて高度で専門的な内容に発展してきたため、字幕の読解過程で誤解や混乱を招くような句読点の表示はできる限り回避しておく必要があるのではないかということを、聴覚障害学生と相互確認した。

日本語の話しことばのポーズ全てが、句読点、すなわち文やフレーズの区切りに対応するとは限らないため(田中, 2001)、復唱者は、話者の音声情報から文やフレーズの区切りに相応するパラ言語情報を抽出し、音声認識ソフトが句読点に変換できるようポーズを意図的に作らなければならない。しかし前述したように、現実には句読点に変換させるためのポーズを作る時間を作るために発話を調整するのは非常に難しい。

そこで、修正作業の段階で、作業方法を一部変更することにした。本来、修正者は、修正用パソコン画面内にある修正パレットの各段の窓に順次表示される字

幕を、4秒後に遅延される音声情報を聞きながら確認・修正して表示用パソコンに順次送ることにしている。変更点の1つは、修正パレットに1段目だけでなく2段目の窓にも字幕が表示されてから句読点の付加の有無が適切か否かを判断して表示用パソコンに送ること、もう1つは、修正者が1、2段目の窓に表示される両方の字幕をあわせて確認しやすいように音声遅延装置の遅延時間を4秒から9秒程度に設定変更することである。特に、読点の付加の適切さについては、例えば、文部省教科書局調査課国語調査室(1946)が次のような使い方を示している。読点は、「文の中止にうつ」、「終止の形をとっていても、その文意が続く場合にはうつ」、「副詞的語句の前後にうつ」など13原則が出されており、これを句読点の付加が適切かを判断する基準の1つとして参考にした。このような修正作業は、修正者のソフトウェア操作技術のレベルも考慮して、講義の14回目以降からこのような修正方法に変えて行った。

その結果、話者の音声発信から表示用パソコンに文字情報が表示されるまでのタイムラグは9秒ほど長くなったが、表示用パソコンに表示される字幕の句読点は適度に付加され、句読点の置換も皆無になってきた。この音声認識通訳システムの字幕結果をみた聴覚障害学生から、「以前よりも読みやすくなった。自然な日本語になっている。」との評価が得られた。タイムラグが長くなった問題点については、聴覚障害学生からとりあえず今は読みやすさを重視したいという指摘があり、今後の検討課題とした。現時点でタイムラグを少しでも短縮する対策としては、復唱者が認識の精度を上げるよう復唱したり修正者の修正技術を向上させるとともに、修正用ソフトウェアのインターフェースもより簡潔で効率的に修正できるようなものに改良する方法が考えられるだろう。

2. 改行および空行の挿入に関する問題

日本語における改行とは、通常、段落の終わりを視覚的に表示する区切り符号として使うものである。修正用ソフトウェア IPTalk9t は、話者の音声情報を文字化した字幕が画面で上方向に順次表示される。音声認識通訳システムにおける修正作業では、段落の終わりを示す改行方法に加えて、段落内であっても適度に改行が入ることで、読みやすさを重視した改行方略を

とっている（日本聴覚障害学生高等教育支援ネットワーク、2008）。

今回の音声認識通訳システムでは、意味的にまとまりのある発話単位で復唱して修正用パソコンに送り、内容や前述の句読点を確認・修正してから表示用パソコンに送った。しかしながら次のように内容的に短い単位の字幕に改行が入ったような形で表示用パソコンに表示されていた（行番号及び括弧は筆者による）。

字幕例3 検討前の字幕表示

- 1 では一つの例、
- 2 ボールだと思ってもらえば、
- 3 いいとおもう。
- 4 ボールには空気が入っていますよね？
- 5 もし空気のない場合は、
- 6 ぺちゃんこ。
- 7 その空気も、
- 8 （空行）
- 9 （空行）
- 10 （空行）
- 11 外の空気
- 12 よりも、
- 13 圧力が少なかったら
- 14 凹みますよね。
- 15 （空行）
- 16 だから
- 17 ボールがボールらしく存在するためには、
- 18 外の空気の
- 19 圧力と、
- 20 中の空気の圧力が、
- 21 （空行）
- 22 バランスを保っている必要があります。
- 23 そうしなければ、
- 24 丸いボールはできないわけですね。

この字幕例3では、2つの特徴が聴覚障害学生に読みにくさを招いていると思われる。

1つは、表示用パソコンに、1行あたりの字幕の情報量が少なく、そのまま1行ごとに順次表示されていることである。これは、表示パソコンの画面全体に表示されるとき字幕の情報量も少なくなり、結果として直前の節や文と関連付けながら読むことが難しくな

る。

もう1つは、行番号7～14、16～22のそれぞれが1つの文を構成しているが、同じ行番号の文の途中で空行が挿入されていることである。行番号8～10の空行の意図は不明である。また、修正パレットの各段の修正字幕が、表示パソコンで約1～5秒の間隔で順次表示されているために、聴覚障害学生にとって、行番号7の字幕が表示されてから行番号11の字幕が表示されるまでの間は、復唱・修正作業が滞っているのか、あるいは行番号7以降の音声情報を文字化できなかったのか判断しにくくなる可能性がある。

そこで、聴覚障害学生、復唱担当の障害学生支援コーディネーターと筆者は協議の上、次のような対処をとることとした。まず、「IPTalk9t」の「表示1」にある「空行のみ改行」にチェックを入れて表示の設定を変更することで、変換候補が確定したらEnterキーを押し、その時点で改行するときはもう1度Enterキーを押すことにした。そして修正者には、①修正した字幕を表示用パソコンに送る際、1、2語の単位ではなく1つの句読点が付加された句節の単位で表示するために修正した各段の字幕同士をつなげること、②空行の挿入は段落単位の発現や終結を表示するために実施すること、の2点に留意して修正・表示してもらうようにした。一方で、誤認識の内容を把握できず修正できなかった場合は、修正できなかった箇所はここであるということ把握できる記号を用いることとする。具体的には、修正用パソコンにあるIPTalk9tの入力設定項目にあるファンクションキーを活用した入力支援機能を使って、F3キーを押すと「—————？」と表示する。

以上の検討によって音声認識通訳システムを行った結果、表示用パソコンには、次のような字幕例4のような字幕結果が表示されるようになった（行番号は筆者による）。

字幕例4 検討後の字幕表示

- 1 私たちを取り巻く世界にはいろいろなものがありますが、
- 2 私たちの世界を取り巻くもの、
- 3 その一つ一つを仮に実物と読んでみると、
- 4 それは信号源。信号の源としての実物というのがある。

- 5 その実物にはいろいろな特徴があります。
- 6 先ほどリンゴの話をしたときに、リンゴという
- 7 実物がある。

こうした字幕結果について、聴覚障害学生から、「このように単語だけで途切れないで1つの文章として示してくれたほうが読みやすい。」との評価を得た。

しかしながら一方で、修正者にとっては、話者から発信される専門的な内容をききながら認識された字幕を確認・修正すると同時に、表示用パソコンに表示される日本語の読みやすさを考慮して句読点や改行の表示を確認・修正することは、非常に負担のかかる修正作業であったとの報告が出された。もう1人の待機中の修正者が修正・表示作業を分担して行ったり、2人で分担して修正できるようなソフトウェアの開発が必要であるかもしれない。

3. 話しことば特有の現象に関する問題

話しことばは、音声による伝達のためその場で消えてしまう(瞬間性)のために、「挿入」「フィラー」「倒置」「語断片」「言いさし」「言い直し」の現象が現れたり、構成要素間の結合が弱く、主述や話線のねじれが生じやすい(堂下・新美・白井・田中・溝口, 1998)。こうした現象は耳で聞くと違和感は覚えませんが、音声認識通訳システムの場合は、そのまま文字に変換して表示するとむしろ理解しにくい文章になる可能性が考えられる。

中野・牧原・金澤・菊池・黒木・井野・伊福部・福島(2006)は、音声認識通訳システムを利用した聴覚障害者から「パソコン要約筆記の方がわかりやすい」といった意見が多く出されていることから、音声言語を単純に文字化するだけでは適切な文字言語として機能できていないことがあるのではないかと考えている。今回の音声認識通訳システムにおいても、話しことば特有の現象をそのまま字幕化することが時々あったために不自然な文章になっている字幕結果がみられた。この字幕結果については聴覚障害学生と十分に協議できなかったため、Mが今後の音声認識通訳システムに向けてどのような対処が考えられるかを検討した。以下、主に呈示されていた2つの字幕例を報告する。1つは「挿入節の呈示」、もう1つは「接続関係の消失」である。

字幕例5 挿入節の呈示(下線部筆者)

そういうみなさんを取り巻く周囲の光、音、においなどそういったものがみなさんの周りには存在しています。それはみなさんがそこにいれば、すべてその刺激を受け取っているかという、実は音というものが空気の振動だと説明しましたよね。いわゆる空気の振動という形で伝わってくる。

字幕例5の下線部のような挿入節や挿入文は、発話中に発話プランが変更されることにより、発話の途中で「但し書き」のような形で急遽挿入されるものである(丸山・高梨・内元, 2006)。

手書きノートテイクやパソコンノートテイクの場合、音声認識通訳システムよりも要約作業の割合が多いため、字幕例5のような音声情報を文字化するとき、例えば「刺激を受け取る時、音は空気の振動という形で伝わってくる」といった内容に要約することで、挿入文の存在が除外されるかもしれない。一方、音声認識通訳システムでは字幕例5のような字幕が呈示されることが生じるため、聴覚障害学生には、下線部の字幕が表示されると日本語文としての不自然さ・不適切さに戸惑い、その直後の字幕が表示されるまでの間は、直前の節との関係が整合できず、一時的に混乱することが予測される。

このような挿入節に対する対処については、次のような検討が必要ではないと思われる。

1つは、復唱作業の段階で、挿入節といった話しことば特有の現象が含まれる音声情報を文語的表現に整理して復唱する方略をとることである。例えば、字幕例5の場合は、「その刺激を受け取っているかについては、以前に音は空気の振動だと説明したように、いわゆる空気の振動という形で伝わってくる」のような内容で復唱する方法である。しかし復唱者にとっては、「挿入節」だけでなく「フィラー」「倒置」や「ねじれ」も含めて実にさまざまな現象を含む話しことばの情報を多角的に処理して復唱する作業は、予想以上に負担が大きくかかるのかもしれない。復唱者がどのように復唱しているのかを把握し、その復唱者のレベルに合わせてどのように復唱作業の改善を図るのかを詳しく検討していく必要があると思われる。

もう1つは、修正作業の段階で、話しことば特有の

現象が出現していることを把握できる記号を活用することである。日本語の話しことばコーパスの構築方法を研究している丸山・高梨・内元（2006）は、文中における挿入文、挿入節の範囲を丸括弧（ ）で括弧することによって視覚化することを提案している。前項で述べた修正用ソフトウェア IPTalk のファンクションキーによる入力支援機能に、入力文の前後を2種類までの括弧で使って括弧の機能が含まれている。その機能の設定内容として、引用表現部分の範囲を示す鍵括弧「 」と、挿入文・挿入節の範囲を示す丸括弧（ ）の2種類を登録すればよいのではないかと考える。丸山・高梨・内元（2006）は、挿入節・挿入文の他に、倒置やフィラー文の範囲を表示する記号も提案しており、音声認識通訳システムでも、この記号を参考にして修正技術を開発することができると思われる。

字幕例6 接続関係の消失（下線部筆者）

たとえばみなさん、タバコを吸っている人の部屋に入った時に、タバコの匂いを感じますよね。でもそこから辺にずっといると、タバコの臭いを感じてる。たばこのにおいがするということを感じなくなります。このように考えると、人の神経というものは、さっき電気で伝わるといいましたが、豊富な知識と続いていると慣れてしまうということもあります。

この字幕例6の下線部は、補助動詞「いる」の終止の形で終わる中止法であると思われる。音声言語における中止法は、耳で聞くと違和感を持たれないようであるが、単純に文字化すると、下線部の部分とその後の部分との関係を把握しにくくなる可能性がある。字幕例6の場合は、「タバコの臭いを感じてる」ことと「たばこのにおいがするということを感じなくなります」との接続関係が捉えにくいものになっていると思われる。

こうした両者の接続関係が不明瞭な字幕については、例えば、次のような表現（下線部）を挿入する方法が考えられる。「でもそこから辺にずっといると、タバコの臭いを感じてるけど、徐々にたばこのにおいがするということを感じなくなります。」あるいは、字幕例6の下線部を削除して「でもそこから辺にずっといると、たばこのにおいがするということを感じなくなります。」と表示する方法もあるだろう。

以上から、音声認識通訳システムの利点は、話者の音声情報をほぼ全文に近い形で字幕化することにあるが、それゆえ字幕例5・6のように手書きノートテイクやパソコンノートテイクではみられない話しことば特有の現象がそのまま字幕化されることも生じてくることが明らかになった。なお、その他の現象として、話者の音声にあるイントネーション、リズム、ポーズ、トーンといった韻律的特徴の存在も指摘されており、これらは話者が発話の流れを調整したり聞き手に誤解を与えない内容の発話を産出するために重要な役割を果たしている。中野・牧原・金澤・菊池・黒木・井野・伊福部・福島（2006）は、「韻律的要素のすべてを表記することは困難であるが、修正段階で、適切な記号を使用することによって、意味を理解しやすくし、パラ言語情報欠落による理解のしにくさを緩和させることができる可能性がある」と提案している。

このように話しことば特有の現象は聴覚障害学生にとっても獲得したい情報ではあるが、単純に文字に変換するのではなく、復唱あるいは修正の段階で書きことばとしての文章に変換したり適切な記号を使用することで対処していく必要があると思われる。

V 聴覚障害学生に対する字幕呈示のあり方に関する検討

情報保障の「情報」とは、話者の音声情報を通訳した情報でなく、話者の表情や身振り、板書・スライド・配布資料のように文字や図絵で表される視覚的情報、さらには授業に直接関連しない音声情報（例えば、校内放送、携帯着信等の雑音）も含まれると考えている。こうした情報は、聴覚障害学生にとって講義の内容や全体の様子を把握するために必要な資源である。

音声認識通訳システムを実施した講義では、教員は、パワーポイントで作成したスライドをスクリーンに投影して口頭説明する方法をとっているため、聴覚障害学生の机に表示用パソコンをおき、聴覚障害学生の座席の前に復唱者と修正者を配置することで、聴覚障害学生の視野内に、表示用パソコン、教員、黒板、スクリーン、支援者が入るようにセッティングした。しかし講義の回数を重ねていくにつれて、聴覚障害学生が表示用パソコンだけでなく修正用パソコンも覗き見る行動が頻繁に観察されるようになった。講義後の

インタビューによると、「字幕結果が早くほしいので、修正用パソコン画面にある修正パレットも一緒に見たい」とのことであった。そこで、聴覚障害学生と障害学生支援コーディネーターと次のような検討を行った。

第1に、聴覚障害学生が、表示用パソコンで呈示される字幕と修正用パソコンの修正パレットに表示される字幕の両方を読めるように、聴覚障害学生の座席位置を変更した。

第2に、表示用パソコンの使い方については、修正用パソコンの修正パレットでは語句の単位で示される傾向があるため、発話の全体的流れを整理・把握する目的で用いることにした。その理由は、聴覚障害学生から「修正用パソコンの修正パレットでその時々の語句を読むので、表示用パソコンでは本を読んでいるようなイメージで読みたい」とのニーズが出されたためである。そこで、表示用パソコンの画面で一度に表示される文字情報量を増やすとともに、表示用パソコンに表示される白色のフォントと黒色の背景をそれぞれ白黒反転させて修正パソコンの修正パレットのフォント色や背景色と統一させた。

当初、筆者ら支援する側は、音声認識通訳システムに関しては表示用パソコンに表示される文字情報が聴覚障害学生にとって資源となるものと思っていたが、聴覚障害学生から「リアルタイムでの情報獲得」を少しでも保障するために修正用パソコン画面内の修正パレットも必要とすることが明らかになった。そして、検討の結果、支援者にとっての修正用パソコンと表示用パソコンは、聴覚障害学生にとってそれぞれ情報を早く獲得するためのツール、談話全体の流れを適切に把握するためのツールとして活用できることがわかった。

VI 考察

1. 利用者の観点に基づいて支援することについて

利用者である聴覚障害学生にとって音声認識通訳システムは使いやすいものであるのか、授業に参加する上で何らかの障害状況に直面していないかを検討するために、Mが、講義に同席して全体の様子や聴覚障害学生の行動を観察し、講義の後で聴覚障害学生に講義中の聴覚障害学生の行動や字幕呈示について気になった事柄を中心にインタビューを行った。そのイン

タビューのなかで、聴覚活用が困難で文字情報に依存する場合にどのような障害状況に直面したのかについて詳しく確認し、どのような改善点が考えられるのかを検討した。この検討の中で聴覚障害学生から指摘された字幕呈示の問題点は、筆者や支援者にとって音声認識通訳システムを始めた当初の段階では予期しないものであった。このように利用者から障害状況の存在を指摘されることによって、筆者や支援者は初めて利用者の観点に基づいた字幕呈示とは何かを検討することが実現できたといえる。ただし、本稿で考案した字幕呈示の技術は、あくまでも今回利用した聴覚障害学生自身の体験やニーズに即して検討したものであって、聴覚障害学生全員に一般化して活用できる技術であるとは断定できないことを付言しておきたい。

また、今回の字幕呈示について指摘された障害状況や解決方法について、聴覚障害学生が自ら発信することは少なく、むしろ「これ位でよいと思う」といったあいまいな回答がみられた。しかし筆者とのインタビューを通して徐々に「読みにくさ」「理解しにくさ」について具体的な障害状況を指摘するようになった。聴覚障害学生自身が、利用者として障害状況の存在を指摘するためには、支援を活用しながら、自分自身と呈示される字幕との相互交渉を分析し、その中で生じた障害状況とその要因を探ることがある程度できる必要があると思われる。しかしながら、日本における聴覚障害学生支援で、大学に情報保障支援技術が積極的に導入されてきているが（白澤，2005；日本学生支援機構，2008）、一方で、聴覚障害学生自身が利用者として情報保障現場で生じる障害状況やその要因をどのように分析・指摘し、解決策をどう見出すのかその手立てを講ずるための研修体制は十分に整備されていない（松崎，2006）。

支援者が利用者の観点に基づいて支援するとは、個々の聴覚障害学生の「体験事実を自身の体験事実として共有し、現象するその事実を確定すること（中野，2006）」であると考えられる。個の体験事実とはその人の心そのものであり、支援する側にある我々は、実際にその人の側においてその人の行動をつぶさに観察することでそこにどのような体験が生まれているのかを捉える。その個々の行動を注意深く観察せずアンケートやインタビューのみで障害状況を間接的に把握する方法では、個の体験事実を深く捉えることに限界が生じる

に違いない。利用者と支援者とが個々の体験事実即してどのように対処していくのかを共同で探り形成していく係わり合いこそ、障害学生支援の根底にあらなければならないと考える。聴覚障害学生も、そうした係わり合いの経験を重ねることによって、支援技術や支援体制の問題を指摘・調整できる利用者に成長することにつながると思われる。

本稿では、聴覚障害者である M が、聴覚障害学生の行動観察とその観察を踏まえて話し合うことで、体験事実を深く捉え、その要因を詳しく分析し、ともに事実として確定することによって、障害状況の指摘や改善点の考案を行う段階へ進めることができたと思われる。

以上から、支援に係わる者は、利用者の観点に基づいて支援するにあたって、特定既存の支援技術のみに依拠せず、聴覚障害学生や支援者等の様子を丁寧に観察して個々の体験事実を自身の体験事実として共有するように係わり、共同で支援システムを構築していく姿勢が求められているといえる。これは、音声認識通訳システムだけでなく、手書きノートテイク、パソコンノートテイク、手話通訳などの情報保障全般に共通する事柄であるということもできる。

2. 音声認識通訳システムにおける復唱・修正方法について

本稿では、音声認識ソフトによる認識のみでは聴覚障害学生にとって正しい情報を推測することが不可能な誤認識が多々生じるため、復唱・修正作業も組み合わせることによってほぼ全文に近い内容を文字に変換するとともに、上記の推測不可能な誤認識を減少させるように取り組んできた。

しかし一方で、音声情報を単純に文字化すると、字幕結果に不適切な句読点や改行が表れたり話しことば特有の現象ゆえに不自然な字幕が呈示されるといった問題点も生じていることがわかった。これは、利用者にとって「読みにくさ」「理解しにくさ」を引き起こす字幕の呈示であり、支援者も筆者も予期していなかった問題点でもあった。そして、本稿で検討してきたように、音声認識通訳システムにおける復唱・修正作業で、推測不可能な誤認識を減少させて認識の精度を高める必要はあるが、日本語の話しことば特有の現象・特徴をどのように処理し、どのような字幕に変

換・呈示していく必要があるのかについて検討を深めていくことも重要であることが確認された。このように支援する側には、音声情報をいかに多くかつ正確に字幕化して呈示するのかわけだけでなく、利用者にとって読みやすい、理解しやすい字幕を作成して呈示しているのかという問題意識も持って復唱・修正のあり方を検討する姿勢が求められている。したがって、音声認識通訳システムは、音声を自動的に文字化できる「最新技術」であると無批判的に受け入れ、利用者の観点をぬきにして聴覚障害学生支援に安易かつ一方的に導入することは、利用者にとって必ずしも望ましいことではないといえる。

今後の課題としては、字幕例 5・6 のような音声サンプルを多く収集して日本語に関する音声学や言語学の知識を活用してどのような処理・変換方略が考えられるのかを分析・考案したり、復唱・修正を行う支援者がその処理・変換方略の作業を行うにあたってどのような障害状況に直面するのかその体験事実を詳しく把握して適切な解決策を考案する必要があると思われる。そうした検討作業に、聴覚障害学生も参加して共同で解決方法を検討していくことも重要なことであろう。また、高等教育機関は、論理・推測・分析・統合・評価等といった言語思考力と専門分野における知識世界を習得するところであるため、手書きノートテイクやパソコンノートテイクと比べて通訳する内容が増加する音声認識通訳システムでは、それ相応の言語力・思考力・専門知識が支援者に要求されるのかもしれない。そこで、講義における教員の様々な話し方や内容を適切に処理・変換して字幕呈示するための方略を検討した上で、その方略を習得するための復唱・修正トレーニングを考案して育成する必要があると思われる。

3. 音声認識通訳システムに対する教員の配慮について

音声認識はほぼ全文に近い内容で通訳することが可能であるが、そのためには音声認識ソフトへの音声入力段階で、復唱者が一定の発話速度と高い発音明瞭度で明確に復唱し、講義中にできる限り一定の認識率を維持・向上していく必要がある。

講義を担当した F は、意味的・文法的にまとまりのある意味単位の区切りでポーズを入れたり比較的ゆっくりしたペースで説明したり支援者の様子を観察して

一時的に話すことを止めたりしていた。そのため復唱者は、かなり短い内容に要約する必要に迫られることなく、ほぼ全文に近い内容に要約して復唱することができた。また、事前にパワーポイント資料を提供してもらうことで、支援者があらかじめどのような用語・説明がなされるのかを把握しておいたり修正用パソコンの日本語入力辞書に単語登録したり、修正用パソコンのそばに資料や単語登録リストをおくことで、修正作業時の変換作業を効率化させることができたと思われる。

このように音声認識通訳システムの質的向上には、講義担当教員の配慮も重要な要素となる。講義担当教員は、「最新技術」である音声認識通訳システムの技術的限界とそれを補うための通訳システムを適切に理解するとともに、聴覚障害学生が健聴学生と対等に講義の情報を共有することを保障するという教育的観点から、復唱者が内容を把握し一定の明瞭度で復唱できるような速さで話したり、復唱者や修正者が専門知識をあらかじめ少しでも把握するために資料を事前に提供することが望まれる。また、教員用の表示用パソコンの字幕や支援者の様子を時々見て支援作業が滞っているかどうかを確認するといった配慮も必要であるかもしれない。

Ⅶ おわりに

教育機関で音声認識通訳システムの導入実績が非常に少ない現在、本稿で利用者の観点に基づいて音声認識通訳システムで生じる「読みにくさ」「理解しにくさ」の解消を目指して検討したが、その試みはまだ緒についたばかりである。

今後も音声認識通訳システムの質的向上に向けて、既存の支援技術のみに依拠したり安易かつ一方的に支援技術を導入するのではなく、利用者を中心に支援者と教員とがそれぞれ直面している障害状況は何かを共有した上で、新たな支援方法を共同で創り出していく必要がある。この実践は、聴覚障害学生にとってさまざまな人々と共生する道を探求するために人と人との係わり合いに価値を見出す機会と実践がまた閉ざされることのないために重要なことであろう。このように聴覚障害学生、支援者、教員との協同的・創出的な営みを通して聴覚障害学生支援を構築する必要があると

いうことを心に深く銘じておきたいと思う。

【謝 辞】

宮城教育大学における音声認識通訳システムの導入にあたり、三好茂樹先生、金澤貴之先生をはじめ、筑波技術大学や群馬大学の先生方にご指導ご助言頂きました。また、宮城教育大学教務補佐員（障害学生支援コーディネーター）の前原明日香さんには、音声認識通訳システムのセッティングや実施、支援学生への指導など様々な面で大変ご協力頂きました。この場を借りて心より感謝申し上げます。

なお、本稿は、文部科学省の平成19年度「新たな社会的ニーズに対応した学生支援プログラム」に選定された宮城教育大学の「障害学生も共に学べる総合的學生支援－障害学生との共生により人間性豊かな社会人を育成するための入学から就職までの総合的學生支援システム構築－（事業推進責任者 藤島省太）」の補助金を受けて行ったものである。

【文 献】

- 安藤彰男（2006）リアルタイム音声認識. 財団法人電子情報通信学会.
- 堂下修司・新美康永・白井克彦・田中穂積・溝口理一郎（1998）音声による人間と機械の対話. オーム社.
- 藤島省太・松崎丈（2007）教員養成大学の特色を生かした障害学生支援体制の構築－宮城教育大学の障害学生支援プロジェクト, 大学と学生, 46, 31－38.
- 菊池真里・金澤貴之・中野聡子・黒木速人・井野秀一・伊福部達・堀耕太郎（2004）聴覚障害学生の情報保障手段としての音声認識システムの活用－聴覚障害学生のニーズに即したシステムの構築－. 日本特別ニーズ教育学会第10回記念研究大会発表要旨集, 41－42.
- 黒木速人・井野秀一・中野聡子・加藤士雄・渡邊括行・堀耕太郎・伊福部達（2003）聴覚障害者の国際会議参加支援のための遠隔型音声字幕化システム－札幌－横浜間におけるシステム運用とその評価－. ヒューマンインターフェースシンポジウム2003論文集, 729－732.
- 古井貞熙（1998）音声情報処理. 森北出版.
- 松崎丈（2006）コーディネーターの目指す姿. 第2回日本聴覚障害学生高等教育支援シンポジウム「聴覚障害学

- 生支援におけるコーディネーターの役割」パネル
ディスカッション資料, 40-41.
- 丸山岳彦・高梨克也・内元清貴 (2006) 第5章 節単位情報.
国立国語研究所報告124 日本語話し言葉コーパス
の構築法, 独立行政法人国立国語研究所, 255-
321.
- 三好茂樹・黒木速人・河野純大・白澤麻弓・石原保志・小林
正幸 (2007) 音声認識技術を利用した字幕作成担当
者のための支援技術とそのシステム開発. 筑波技術
大学テクノレポート 14, 145-152.
- 三好茂樹 (2008a) PEPNet-Japan Tipsheet 音声認識技術を
用いた情報保障. 日本聴覚障害学生高等教育支援
ネットワーク.
- 三好茂樹 (2008b) 音声認識同時字幕のための復唱者養成セ
ミナー (2008.02.28), 宮城教育大学学生支援 GP 配
布資料.
- 文部省教科書局調査課国語調査室 (1946) くぎり符号の使ひ
方. 文部省.
- 中野尚彦 (2006) 障害児心理学ものがたり I - 小さな秩序系
の記録. 明石書店.
- 中野聡子・牧原功・金澤貴之・菊池真里・黒木速人・井野秀
一・伊福部達・福島智 (2006) 音声認識技術を利用
した字幕呈示システムの現状と課題 - 音声言語と文
字言語の性質の違いに焦点をあてて. 群馬大学教育
実践研究, 23, 251-259.
- 日本学生支援機構 (2008) 平成19年度 (2007年度) 大学・短
期大学・高等専門学校における障害学生の修学支援
に関する実態調査結果報告書.
- 日本聴覚障害学生高等教育支援ネットワーク (2008) パソコ
ンノートテイク導入支援ガイド やってみよう! パソ
コンノートテイク. 日本聴覚障害学生高等教育支援
ネットワーク.
- 白澤麻弓 (2005) 日本における聴覚障害学生高等教育支援の
実態～全国調査の結果～. 第8回聴覚障害者と高等
教育フォーラム資料集, 32-33.
- 田中穂積 (2001) 言語理解と行動制御に関する研究 (課題番
号12NP9201) 平成12年度科学研究費補助金 (創成的
基礎研究) 研究成果報告書.

(平成20年9月29日受理)