

## 音声認識技術による字幕運用の課題

—— 音声言語を文字化することの問題 ——

牧原 功<sup>1)</sup>・金澤 貴之<sup>2)</sup>・福島 智<sup>3)</sup>・井野 秀一<sup>3)</sup>  
伊福部 達<sup>3)</sup>・黒木 速人<sup>4)</sup>・中野 泰志<sup>5)</sup>・中野 聡子<sup>3)</sup>

<sup>1)</sup>群馬大学留学生センター

<sup>2)</sup>群馬大学教育学部

<sup>3)</sup>東京大学先端科学技術研究センター

<sup>4)</sup>筑波技術大学

<sup>5)</sup>慶應義塾大学

### 要 旨

音声同時字幕システムの運用を進める中で、いかに誤変換を少なくしても字幕が理解しにくいこと、つまり音声言語を文字化するという自体に何らかの問題が含まれることが明らかになってきた。本稿は、理解の困難さが生じる原因を考察するものである。字幕ログを分析し、理解しにくい呈示文を生む要因を、「呼応関係の消失」「主語・述語等の省略」「複文生成時の統語的な制約違反」「主題の省略」に分類した。また、それら理解しにくい文もフィラーやポーズなどのパラ言語的な音声情報を視覚的に付与することで、理解を助けることができることを示した。

【キーワード】 音声言語 文字言語 音声同時字幕システム 音声文字化

### 1. はじめに

近年、高等教育に学ぶ聴覚障害学生の次世代の情報保障手段として、音声認識技術を活用した字幕呈示技術に対する期待が高まっている。話者の音声を正確に字幕化することができれば、講義を最も臨場感あふれる形で伝えられると考えられるからである。しかしながら、現在の技術レベルでは、特定の話者の音声しか認識できないことに加え、話者の音声を直接認識させるだけでは誤認識が多く発生してしまうという問題も残されている。上記の問題の解決のために、本研究の研究分担者でもある伊福部・井野らは、「音声同時字幕システム」を開発した。これは、コンピューターによる音声認識に、復唱作業や修正作業といった人間の手による作業を介在させることで、誤認識の発生の問題を最小限に押さえようとするものであり、(株)ビー・ユー・ジー（以下(株)BUG）により、すでに国際会議等で実

運用がなされ、字幕の精度は97%に達している。

我々研究グループは、平成16年度から、「音声同時字幕システム」を大学での聴覚障害学生支援の手段に最適化するための研究に着手した。平成16～17年度には、聴覚障害学生と同じ教室内に中間支援者を配置し、字幕の修正作業を担当する方法について検討し（研究代表者：福島智、科研費基盤（C）(2)、課題番号16530617）、大学で利用可能なシステムを構築することが可能となった。

しかし、その運用を進めるに従って、音声同時字幕システムにより呈示される字幕には、ある種の読みにくさ・理解しにくさが伴うものであることが明らかとなった。これは、聴覚障害者だけでなく、健聴者も感じる理解しにくさであることが確認された。これまで、字幕が読みにくいという意見が出ても、それは障害者の言語能力が低いからであると考えられる傾向が少なからず見られたが、実際にはそのような原因に帰着させることができない現象である。

話し言葉を書き言葉として呈示すると何故わかりにくいものになってしまうのかについては、幾つかの可能性を考えることができるが、その一つは、音声言語の理解にはパラ言語要素が重要な役割を果たし、言いよどみやポーズなどのフィラーが談話標識として機能しているが、字幕化の際にはそれらの情報が省略されることにより、理解し難さが生じているということである。

本研究においては、まず、呈示された字幕が何故理解しにくいのかを統語的に考察し、音声言語においては示されているが文字化されると捨象されてしまう情報を字幕に反映させることで字幕の理解しやすさを高められる可能性を検討した。

## 2. 音声字幕システムの概要と運用

伊福部・井野らは、「音声同時字幕システム」を(株)ビー・ユー・ジー（以下(株)BUGとする）と共同開発した。「音声同時字幕システム」は以下のようなシステムとなっている。

話者→（電話回線・LAN回線による音声の送信）→同時復唱者→（文字化・文字データ送信）→同時修正者（話者と同じ場所で文字データと音声データを比較・修正）→利用者（PCやスクリーンに提示）

このシステムの特徴は、同時復唱者の存在であるが、同時復唱者を入れる理由は、

- 1) 不特定話者環境を特定話者環境に変換する
- 2) 適切な発音と一定の発声強度を保つことにより、音声認識率を高める

の2つである。

話者の音声を聞き取ると同時に復唱を脱落なく、かつ音声認識装置にとって正しく認識しやすい発音・発声強度で行うことはたやすいことではない。井野ら（2003）は、発声・発話トレーニングを受けた民放アナウンサーと一般の大学生を対象に、復唱課題を行ったところ、復唱精度は、アナウンサー96%、大学生86%、音声認識精度に至っては、アナウンサー90%、大学生65%と、大きな開きがあったと報告している。

そこで、「音声同時字幕システム」では、アナウンサーなど発声・発話トレーニングを受けた者に同時復唱させることによって、復唱精度は95%以上、音声認識精度は90%程度で文字化し、これに修正クライアントを使用して4名で同時修正を行うことで、話者の発話から平均11秒後に、97.2%の精度で字幕呈示を行うことを可能にした（黒木ら、2003）。

群馬大学では、平成16年度から、この「音声同時字幕システム」を大学での聴覚障害学生支援の手段に最適化するための研究に着手した。ネットワークを㈱BUG（札幌）につなぎ、復唱作業は㈱BUG側で行いつつ、字幕の修正作業を教室内で行う形で、安定した運用ができるようにした。修正作業は、特別なスキルを必要としない為、学生でも行えること、また講義を履修したことのある者の方が、専門用語などの修正に対応しやすいからである。なお、現在は、さらなるコスト削減のため、学内において復唱者養成にも取り組んでいる。

修正者は、4名から1名に減らしたため、字幕呈示までのタイムラグが大きくなっているが、発話→同時復唱→同時修正→字幕呈示という「音声同時字幕システム」のプロセスは同じであり、誤認識の発生を最小限に抑えた字幕呈示が可能となっている。

### 3. 問題の所在

音声同時字幕システムは、

発話者の発話→（発話の同時復唱）→〈音声の文字化〉→（文字化されたデータの同時修正）→字幕呈示

というプロセスで行われる。文字化は機械によって行われるが、機械の音声の誤認識を極力少なくするため、発話者の発話をそのまま入力するのではなく、発話を聞いて同時復唱する復唱者を経て入力される。また字幕として呈示する前に、数名の修正者が、音声を参照しながら字幕のエラーを修正する。

このうち、文法的なエラーなどを修正するのは同時修正者による修正作業の段階であるが、ここで与えられる時間は短く、活用のミス、格助詞の誤用等の明らかなエラーしか修正することができない。そのため、「私が言いたいのは、彼は間違っていると思う」というような文に見られる主語と述語の呼応のズレや、多くの節が続いていく長大な文などは、修正作業を行うことができず、そのまま呈示しているのが現状である。

上記のような修正者が容易に修正できないエラーは、発話者の発話に元々含まれるものでもあり、字幕化の過程で取り除くことができないものである。

このような状況にあって、我々は、字幕として呈示されると理解しにくい文が、何故音声言語としては、いくつものエラーを含んでいるにもかかわらずスムーズに理解されるのかということ进行を明らかにしたいと考えた。また、この点を明らかにすることによって、音声の認識に近い形で字幕を理解させるためには、どのような工夫が必要となるのかを明らかにできるのではないと思われる。

## 4. なぜ字幕は読みにくくなるのか一言語的な特徴からの考察一

### 4-1. 方法

字幕運用実験を見学した研究分担者から、音声を聞きながら字幕を読む場合には特に困難を感じなかったが、耳をふさぐなどし、音声を聞かないようにして字幕のみを読んでいると、話の内容が非常に理解しにくかったという意見が寄せられた。そこで、字幕運用のログを資料として、内容の理解が困難であると思われる箇所を抽出する作業を進めた。

その結果いくつかの難読箇所が確認されたが、それらの文について群馬大学の学生に意見を求めたところ、わかりにくいという意見とそれほどわかりにくいとは思わないという意見とに分かれた。

そこで、読み手の理解が実際に正確になされているのか否かを調べることにし、難読箇所を資料として読ませ、内容を理解したと被験者が判断した段階で内容についての簡単な設問に回答することを求めた。

このような確認作業を経て難読箇所と認定した文について、なぜ読み手の理解が困難になっているのかを統語的な特徴から分析した。あわせて、他の字幕ログの中で理解が難しいと思われるものを抽出し、そのパターン化を行った。

### 4-2. 難読文

字幕ログの分析を行い、難読箇所を含む文として抽出したものの例を以下に示す。なお、これらの文は13名の群馬大学学生（健聴者）に読ませ、内容が理解できているかを確認した。内容理解の設問は細かな部分を問うものは避け、文章の大意が理解できていれば回答が可能と考えられるものを使用した。

調査の結果、回答者への対面聞き取り調査では、さほどわかりにくくないと回答している文であっても、正確な意味の理解ができていないことが多いことがわかった。これらの文のわかりにくさの分析については難読文を生じさせる要因について検討した後に改めて行うこととし、ここでは難読箇所を含む文とはどのようなものであるのか、読み手の理解はどうであったのかを示す。

#### 文章1

ヘルパーさんはボランティアじゃなくて国がお金を払っているわけだから、本人的には依存しないで他人に依存しないでやってもらっていると、いやヘルパーに頼まれていること自体依存しているんだと言われたらですね、例えば家にダスキンの人に来てもらって掃除してもらっているのは依存していないのかという話ですよ、あるいはさっき言ったように、国立大学で授業を受けている、あるいは国立大学の教員であること自体国で養われているのかな、じゃあ依存してないのか、それはそんなことはないと思うんです。

質問：ヘルパーさんを頼んでいるということは「依存している」ということですか？

回答：（選択式）

(正答) いいえ、依存しているということではありません。	6名
はい、依存しているということです。	6名
無回答	1名

## 文章2

先週の木曜日にテレビ局から電話がかかってきました。(中略)

ちなみに、なぜ私が、私に声がかかってきたか。日本で手話で研究をしている人で一番権威がある。それは嘘です。一番有名だ。これも嘘です。手話の研究をしている人の中でらしいです。いろんな人に電話を掛けまくった時に金澤先生だったら割とひげが濃いからいけるのではないかそんな話だったみたいです。というのはですね何で私なのですかというふうに食い下がって聞いてみました。

手話の言語学をやってるわけではないです。語源の話ですから言語学者に聞くべきですね。僕は言語学をやってるわけではなくて、あえて言えば手話の社会学ですね。もっと専門はろう教育の中で手話はどう取り入れられているかということを経済的に研究することですけれども。

質問：だれが、何を言語学者に聞くべきだと言っていましたか。

回答：(自由記述)

(正答) テレビ局の人が語源を言語学者に聞くべきだ	6名
テレビ局の人が…	1名
金澤先生が語源を言語学者に聞くべきだ	4名
金澤先生がTVの出演依頼を聞くべきだ	1名
金澤先生が…を言語学者に聞くべきだ	1名

## 文章3

そしてもうひとつは常者があたり前のように食事をし服を選びと言う風になっている、そしてそれは自分が食べたいものを食べ自分が着たい服を着る、そういうこちらのほうがですからQOLという話により近いと思いますが、そういうあたり前の生活が営むというその場所の問題と生活の中身と2つの意味でそのどちらもノーマルでなければいけない、だからわれわれにとってあたり前であるのと同じように彼らにとってもそれはあたり前でなければいけないそう考えると細かい部分でいろいろやりにくいことが出てきます。やりにくいってというのは誰にとってかっていうと、障害者を管理しておくほうが楽だという発想からすれば、やりにくいと思います。

質問：ノーマルでなければならぬものは何ですか。

回答：(自由記述)

(正答) 生活の場所と中身	4名
生活の場所	2名
生活の中身	1名
あたり前な生活	1名

生活すること	1名
着たいものを着て食べたいものを食べる	2名
QOL	2名

これらの例は、研究グループでは難読文であると判断したが、予備調査ではさほどわかりにくくないという声も多かった。読み手はさほどわかりにくくないと感じている文であっても、中心的な意味の把握にも困難が生じているものが多いことがわかる。

### 4-3. 難読文を生じさせる要因

前節で考察したように、音声言語を文字化することによって表示される字幕には、音声言語に特有の性質によって生じる問題点が含まれている。

以下では、他の字幕ログを参照しつつ、その問題点を整理したい。

#### 4-3-1. 呼応関係の消失

文意が理解しにくくなる原因の一つに、「呼応関係の消失」を挙げることができる。呼応関係が消失している例は字幕ログ中に多数見られた。以下ではその中の代表的なものを取り上げる。

##### 文章4

例えば、ホーキング青山は、テレビでよく最近テレビでよく見かけませんが、大丈夫でしょうか。彼がいうのには、障害を持った人という特別扱いをしないようにしないようにしている方がよっぽど差別です。

↓

「彼がいうのには」の係り先が省略されており、意味的な大まかな把握は可能であっても、正確に読み込もうとすると困難な点が含まれる。

↓

例えば、ホーキング青山は、テレビで、最近テレビであまり見かけませんが、大丈夫でしょうか。彼がいうのには、障害を持った人という特別扱いをしないようにしないようにしている方がよっぽど差別だということです。

##### 文章5

ちょっと違う例ではあるけれどある種わかりやすい例だと思うのは、女の子が、こういう男の子をいいと挙げる男性像の例として、ちょっと重い荷物を持っている時に、重いでしょうかわざわざ言わずに、さっと、荷物を二三段積み重ねて、歩いているときに上の荷物を持ってくれたりとか、あるいは、何も言わず自分の方が重い荷物を持っていてくれる。高いところにある荷物が届かない時に、さっと出してくれる。その時におれが出してやったじゃん、とか言わないですよ。言われたら単に不愉快になりますよね。

↓

「男性像の例として」の係り先が消失している。



ちょっと違う例ではあるけれどある種わかりやすい例だと思うのは、女の子が、こういう男の子をいいと挙げる男性像の例として、ちょっと重い荷物を持っている時に、重いでしょうとかわざわざ言わずに、さっと、荷物を二三段積み重ねて、歩いているときに上の荷物を持ってくれたりとか、あるいは、何も言わず自分の方が重い荷物を持っていてくれる。高いところにある荷物が届かない時に、さっと出してくれるというのがありますよね。その時におれが出してやったじゃん、とか言わないですよ。言われたら単に不愉快になりますよね

#### 文章 6

それとも、あともうひとつは、逆にこれはむしろ反対に地域の手話通訳者の方が、より高い専門性が必要だといえる面かもしれないと思うのは、いろんな聾者に対応しなければいけないわけです。地域の初通訳の方は。ですが、大学での情報保障に関して言えば、少なくとも今の群大では群大にいる二人の学生の必要とする手話あるいは、日本語力そういったもののニーズにこたえる、言い換えるとたとえば、ろう者の中には、日本語があまり得意ではない人がいっぱいいるわけです。そうしたときに、わかるように通訳しなければいけない。そういった専門性も出てくるわけです。



それとも、あともうひとつは、逆にこれはむしろ反対に地域の手話通訳者の方が、より高い専門性が必要だといえる面かもしれないと思うのは、いろんな聾者に対応しなければいけないということ、つまり様々な日本語のニーズに応えなければいけないということです。地域の初通訳の方は。ですが、大学での情報保障に関して言えば、少なくとも今の群大では群大にいる二人の学生の必要とする手話あるいは、日本語力そういったもののニーズにこたえるだけでよいのです。言い換えるとたとえば、ろう者の中には、日本語があまり得意ではない人がいっぱいいるわけです。そうしたときに、わかるように通訳しなければいけない。そういった専門性も出てくるわけです。

上記の例から、呼応関係が消失することが、文字言語の理解においては大きな影響を与えることがわかる。

#### 4-3-2. 省略

字幕の読みにくさを生じさせる原因として、文中の必要な成分の省略が挙げられる。話し言葉においては頻繁に用いられているものであるが、書き言葉として理解しようとした場合には、その妨げになることが多いものである。

## 文章7（述語の省略）

結局のところは、そんなことを前回も話したと思うんです。書く側がこれはいらぬとかいう判断をするべきではなく、伝えるのが情報保障ということです。話し言葉と比べて、ここは追いつかないところがあるとしても、基本姿勢としてありのまま伝えるということです。別の形で、こんなことを書いた方がいます。この情報はいらぬと、だれかが判断するのはなく、情報保障と聴覚障害学生に他の学生と同じ情報というのは、本当に必要だけれども、これを達成できているかといえばそうではないので、という悔しさは、健全だと思うんです。

省略されている成分を補足すると以下のようなになる。

別の形で、こんなことを書いた方がいます。この情報はいらぬとだれかが判断するのではなく、情報保障と聴覚障害学生に他の学生と同じ情報を**与える**というのは、本当に必要だけれども、これを達成できているかといえばそうではないので、ということに悔しさを感じることは、健全だと思うんです。

## 文章8（主語の省略）

手話の言語学をやっているわけではないです。語源の話ですから言語学者に聞くべきですね。僕は言語学をやっているわけではなくて、あえて言えば手話の社会学ですね。もっと専門はろう教育の中で手話はどう取り入れられているかということを経済的に研究することですけれども。

最初の文「手話の言語学をやっているわけではないです」の主語は「私」である。そのため、それ以後の文で主語が明示されない場合、通常は主語は「私」であると解釈される。そのため、「言語学者に聞くべき」主体も「私」と解釈され、そこで意味的な齟齬が生じる。

書き言葉においては、主語が変わった場合は主語を明示するが、話し言葉ではこれまでのコンテキストで主語が想像可能であると話し手が考えた場合、或いは話し手の意識の中では主語の入れ替えがスムーズに行われると考えられている場合など、主語が明示されない場合が多い。この文は上記の点を考慮して以下のようにすると理解しやすいものとなる。

**私は**手話の言語学をやっているわけではないです。これは語源の話ですから、**テレビ局の人は**言語学者に聞くべきですね。僕は言語学をやっているわけではなくて、あえて言えば手話の社会学ですね。もっと専門はろう教育の中で手話はどう取り入れられているかということを経済的に研究することですけれども。

主語の省略の例では、文意が全くとれないということではないが、正確な理解が必要とされる場合には、誤解を生じさせることも十分にありうると考える。



### 4-3-3. 句読点の箇所

句読点の位置が実際の発話と異なっていたり、句点と読点が区別されずに変換されることにより、理解しにくくなる例がある。

#### 文章9

だからこそろう教育では、聴覚を進められてきたということがいえると思います。ろう教育の世界では、ただし僕は、ろう教育に意見を持っていますが、それはそれでわかりやすい意見だったと思います。

本来、句点を打つ場所に読点を打ち、読点を打つ位置に句点を付けることによって、「ろう教育の世界では」と他の文との関係が把握できなくなっている。この点を修正すると次のようになる。

だからこそろう教育では、聴覚を進められてきたということがいえると思います、ろう教育の世界では。ただし僕は、ろう教育に意見を持っていますが、それはそれでわかりやすい意見だったと思います。

### 4-3-4. 主題の省略による内容の混乱

以下は、文章中の主題が省略されることによって、理解しにくくなっている例である。

#### 文章10

(ノーマライゼーションという言葉もあります。つまりいい、だれもがノーマルな当たり前の生活をする権利がある。二つの意味があるわけです。大きな施設に行って、収容されていた歴史があって、収容という言葉を使つてはいけなさとされますが、施設は収容ではなく入所といわれていますが、開いている大きな施設に収容されて批判されて、そうではなくてだれもが自分が住んでいる地域で生活する権利がある。それら一つともう一つは、いろんな生活水準ですね。同じように当たり前のような生活をする権利がある。この二つは結局とどのつまりは同じことを云っています。)

これらの観点を合わせていくと、聞こえにくいということは、その人のものとして、最初からあるわけですから、権利としてあって当然なわけです。ですからそこに努力が必要なのです。そして、努力しなくてもいいことなんで、聴覚が障害ないからわれわれは授業を普通に聞きます。普通に努力をしなくつまらない話を寝ないようにする努力とか、それはおいといて、普通に聞こえてくる。聴覚に障害があることでそこに努力を強いられるのはなぜか。本人の努力には限界がある。回りが努力することも必要である。という指摘です。

文法的には「権利としてあって当然」の事柄は「聞こえにくいということ」と解釈されるはずである。しかし、読み進めると意味的に整合しないことによって、再度読み返すという行為が必要となると考

えられる。「権利としてあって当然」であるものが何か明示することによって、理解しやすい文となる。

これらの観点を合わせていくと、聞こえにくいということは、その人のものとして、最初からあるわけですから、ノーマルな当たり前の生活をするということは権利としてあって当然なわけです。ですからそこに我々健聴者側の努力が必要なのです。そして、努力しなくてもいいことなんで、聴覚が障害ないからわれわれは授業を普通に聞きます。普通に努力をしなくつまらない話を寝ないようにする努力とか、それはおいといて、普通に聞こえてくる。聴覚に障害があることでそこに努力を強いられるのはなぜか。本人の努力には限界がある。回りが努力することも必要である。という指摘です。

#### 4-3-5. 複文生成時の統語的な制約違反

複文を用いる際に、話し言葉ではポーズやイントネーションによって、どこに区切りがあるかを示すことが多い。しかし、それを文字言語として提示してしまうと、読み手は文法的な手がかりをもとに、文の切れ目を探さねばならない。このような原因によって生じた読みにくさは非常に重大であり、文意が全くとれないということもしばしば起こっている。以下にその例を示す。

##### 文章11

自分が全くわからない言語、わからない中でも頑張って交渉しようとするときにも、そのバックに自分の味方がいる、自分の味方がいると思っているのか、自分の何それに対して自分の他に誰もいないという状況の中で過ごすのか、結果的に社交性がどちらに身に付くのかというのは、そうですね、いろいろな聞こえない人、インテグレーションを経験した人の話を聞くと、逆に、同じ聞こえない人の出会いによって自分の世界が広がったというわけです。

「そのバックに自分の味方がいる、自分の味方がいると思っているのか、自分の何それに対して自分の他に誰もいないという状況の中で過ごすのか、結果的に社交性がどちらに身に付くのかというのは、」を一つの節として認識してしまうことによって、「結果的に」以降の文が、それ以前とは異なる文脈の中で用いられていることが理解できなくなっている。

文を区切ることでこの点のわかりにくさは解消される。

自分が全くわからない言語、わからない中でも頑張って交渉しようとするときにも、そのバックに自分の味方がいる、自分の味方がいると思っている場合と、自分の何それに対して自分の他に誰もいないという状況の中で過ごす場合とで、社交性の身に付き方が変わってくるということがあるように思います。

結果的に社交性がどちらに身に付くのかというのは、そうですね、いろいろな聞こえない人、インテグレーションを経験した人の話を聞くと、逆に、同じ聞こえない人の出会いによって自分の世界が広がったという人が多いようです。

## 文章12

もっと言ってしまうと菊池さんには悪いんですが、菊池さんはあくまで、今回の実験の聾学生役という役目なので、字幕が例えば止まってしまったりして混乱したとしても、菊池さんのことはほおっておくという言い方へんですが、実験上はそうなったということを理解してもらうということですが、受講している学生は、実験が失敗したから受講できなくなったということだと困るわけですから、常に支援者が頑張っているという形を取っています。

統語的なルールに従うと、従属節「～ので」の中に従属節「～が」を含むことはできない。そのため、この例では、「～が」で構成される従属節中に「～ので」が含まれる構造として理解され、本来「ので」で切るべき意味単位がそこで切れず、意味が不明確になってしまっていると考えられる。

もっと言ってしまうと菊池さんには悪いんですが、菊池さんはあくまで、今回の実験を行っている聾学生という役目です。

ですから、字幕が例えば止まってしまったりして混乱したとしても、菊池さんのことはほおっておくという言い方はへんですが、実験上はそうなったということを理解してもらうということになります。しかし、受講している学生は、実験が失敗したから受講できなくなったということだと困るわけですから、常に支援者がスタンバっているという形をとっています。

## 4-3-6. 調査に使用した例文の分析

上で考察した統語的な問題点は、実際には文中で重複して現れることが多く、そのような場合は文の理解を更に大きく妨げるものとなっている。また、このような問題点は、話し言葉が根本的に内包する性質であり、発話者が発話時に発話から除外することは困難なものであると思われる。この点について、調査に用いた難読文例を用いて再度検討しておく。

なお、被験者の回答状況の後に、実験に用いた文を理解しにくいものとしている要因を挙げるため、例文中にはその箇所を示すための番号(①など)や下線を加筆した。

## 文章1

ヘルパーさんはボランティアじゃなくて国がお金を払っているわけだから、本人的には依存しないで他人に依存しないでやってもらっていると①、③いやヘルパーに頼まれていること自体依存しているんだと言われたらですね、例えば家にダスキンの人に来てもらって掃除してもらっているのは依存していないのかという話②ですよね、あるいはさっき言ったように、国立大学で授業を受けている、あるいは国立大学の教員であること自体国で養われているのかな⑤、④じゃあ依存してないのか、それ⑥はそんなことはないと思うんです。

文章1を理解しにくい文としている原因としては以下の点を挙げるができる。

1. 述語の省略
  - ① 「思っているのに」
  - ② 「と同じ」
2. 他の動作主の省略
  - ③ 他の人から
3. 文の省略
  - ④ 「いや、そうじゃない、」
  - ⑤ と言うと
4. 指示詞「それ」の指すものが不明瞭
5. 読み手を混乱させる挿入節「じゃあ依存してないのか」の使用

#### 文章2

先週の木曜日にテレビ局から電話がかかってきました。(中略)

ちなみに、なぜ私が、私に声がかかってきたか。日本で手話で研究をしている人で一番権威がある。それは嘘です。一番有名だ。これも嘘です。手話の研究をしている人の中で①らしいです。いろんな人に電話を掛けまくった時に金澤先生だったら割とひげが濃いからいけるのではないかなそんな話だったみたいです。というのはですね何で私なのですかというふうに食い下がって聞いてみました。

②手話の言語学をやってるわけではないです。③語源の話ですから④言語学者に聞くべきですね。僕は言語学をやってるわけではなくて、⑤あえて言えば手話の社会学ですね。もっと専門はろう教育の中で手話はどう取り入れられているかということを社会的に研究することですけれども。

文章2を理解しにくい文としている原因としては以下の点となろう。

1. 述語の省略
  - ① いろいろな人に声をかけた
2. 主語の省略
  - ② 私は
  - ③ 今回の話は
  - ④ テレビ局の人は
  - ⑤ やっているのは

### 文章 3

そしてもうひとつは常者があたり前のように食事をし服を選びと言う風になっている、そしてそれは自分が食べたいものを食べ自分が着たい服を着る①、そういうこちらのほうがですから QOL という話により近いと思いますが、そういうあたり前の生活が営むというその場所の問題と生活の中身と 2 つの意味でそのどちらもノーマルでなければいけない、だからわれわれにとってあたり前であるのと同じように彼らにとってもそれはあたり前でなければいけないそう考えると細かい部分でいろいろやりにくいことが出てきます。やりにくいっていうのは誰にとってかっていうと、障害者を管理しておくほうが楽だという発想からすれば、やりにくいと思います。

この文を理解しにくいものとしている理由は、一文が非常に長くなっているということである。節を次々と続けていくことによって、文中の呼応関係にもずれが生じ、より一層わかりにくさを生じさせていると考えられる。また句点の省略も多い。

この文をいくつか分割して整理し、①の部分に「もうひとつは」と呼応する述語「ということです」を補うと理解しやすい文にすることができる。

文章 1～3 の例からもわかるとおり、話し言葉に含まれる様々なわかりにくさを生む要因は、重複して生じるものであり、このような要因を避けて発話することは不可能であると考えられる。また、字幕提示において、字幕修正者が瞬時に対応して上記のような文に修正して呈示するということも実現は難しい。

そのため、このような問題箇所を含んだままの文を字幕として提示しつつ、読み手の理解を助ける方策を考えなければならない。以下では、その方策の一つとして、文字化に際して捨象される音声情報を利用する方策を検討する。

## 5. 文字化に際して省略される音声情報

### 5-1. フィラー・言いよどみ・ポーズ

話者データの文字起こし資料を見ると、「えー」を始めとする、「その」「あ」「あの」など、フィラー・言いよどみが散見される。しかし、これらの情報をすべて文字化した場合、字幕が読みにくくなると考え、復唱段階において省略されてきた。

しかし、フィラーは話し手にとっては、思考過程を相手に伝える手段ともなっていると考えられる。たとえば、ある発話の途中で、「えー」「あー」のようなフィラーが用いられた場合、聞き手は「これまで話してきた内容をわかりやすく言い換えようと試みている」「話題が転換する」等の可能性を想定し、以後の発話とそれ以前の発話との関連性に注意を払うということが想像できる。また、ポーズもフィラーの一種として、同様の機能を果たしていると考えられる。

## 5-2. パラ言語情報

話し言葉は、イントネーション、トーン、リズム、スピード、プロミネンスなどの多くのパラ言語情報を含んでいる。これらの情報は、たとえばイントネーションが「～か」という文が疑問であるのか、納得であるのか、反語であるのか等を示したり、数秒のポーズの存在によって、聞き手に話題の転換への準備を進めさせたり、場合によってはアクセントが「はし」という言葉の意味が「橋」なのか「箸」なのかを示すなど、意味の理解に寄与している。

しかし、書き言葉には、パラ言語が含まれない。この点で、本来、イントネーションやポーズが担っていた機能は失われてしまうことになる。

## 6. 音声情報の付与による効果

### 6-1. フィラーの付与

呼応関係が消失している例においては、フィラー・ポーズの付与は効果的である。以下の例を参照されたい。

#### 例1

たとえば、ホーキング青山は、テレビでよく最近テレビでよく見かけませんが、大丈夫でしょうか。彼がいうのには、障害を持った人という特別扱いをしないようにしないようにといている方がよっぽど差別です。

↓

たとえば、ホーキング青山は、テレビでよく最近テレビでよく見かけませんが、大丈夫でしょうか。

彼がいうのには…

障害を持った人という特別扱いをしないようにしないようにといているほうがよっぽど差別です。

この例はさほど意味を取りにくいというタイプの文ではないかもしれないが、字幕を読むということ、つまり、発話と同時に示される文字情報から情報を読み取るという作業を想定した場合、瞬間的な内容理解を妨げる可能性はあるのではないかと考える。

このような文では、「彼がいうのには」とそれ以降の文の間のポーズを「…」で示すことによって、以降が引用であることが比較的読み取りやすくなるように思われる。また、「…」の後で改行を行うことによって、更にわかりやすい文となる。

## 例2

ぼくもいくつか特例子会社を見たことがありますが、例えば、エプソンという会社は解りますね。プリンターとか、いろいろなコンピューター関係のものを作っています。そのエプソンの特例子会社では、名刺作りを例えばしています。他にもいくつかやっていますが、名刺作りというのはどういうことかという、例えば僕たちが自分の名刺を作ることは自分のパソコンがあって名刺作りのキット例えば紙を買えばできますね。あるいは、印刷会社に名刺、朝に頼めばできますね。一人当たり名刺を作る金額も量もたいしたことないと思いますが、何しろ世界のエプソンです。社員証考えたら、その名刺作りだけを一カ所に集めれば、十分に採算がとれるわけです。名刺作りをするために、比較的簡単な作業をして子会社に任せる。そんなことは十分に採算がとれるわけです。エプソンの全社員の名刺を知的障害者を雇用することができるわけです。



ぼくもいくつか特例子会社を見たことがありますが、例えば、エプソンという会社は解りますね。プリンターとか、いろいろなコンピューター関係のものを作っています。  
そのエプソンの特例子会社では、名刺作りを例えばしています。  
他にもいくつかやっていますが、名刺作りというのはどういうことかという…  
例えば僕たちが自分の名刺を作ることは自分のパソコンがあって名刺作りのキット、例えば紙を買えばできますね。…

この文は例1の文よりもわかりにくい文であると考えるが、ポーズを反映させない文だと「名刺作りというのはどういうことかという、例えば僕たちが自分の名刺を作ることは自分のパソコンがあって名刺作りのキット例えば紙を買えばできますね。」が一つの文として認識されてしまう可能性が高いということが、その原因と思われる。

統語的な点から文を修正する場合であれば、既に見たように「どういうことかという、次のようなことです。」として文を切り、次の文を始めることになるが、それが時間的に不可能であると言う場合、次善の策として、「どういうことかという」の後ろにポーズを示す「…」を明記するということが挙げられ、更に改行、或いは一行空けることによって、よりわかりやすい文になると考えられる。

呼応の消失以外でも、従属節を多用した複雑な複文や、従属節の従属度を反映しない包摂関係で解釈しないと意味が取りにくい文においても、フィラーやポーズは文の構造を把握するための手がかりとなると考えられる。以下は従属節の統語制約に違反した文であるが、フィラーやポーズを明示することで理解のしやすさは向上すると思われる。

次の例も同様の処理により理解しやすくなるものである。

## 例3

もっと言ってしまえば菊池さんには悪いんですが、菊池さんはあくまで、今回の実験の聾学生役という役目なので、字幕が例えば止まってしまったりして混乱したとしても、菊池さん

のことはほおっておくという言い方ヘンですが、実験上はそうなったということを理解してもらおうということですが、受講している学生は、実験が失敗したから受講できなくなったということだと困るわけですから、常に支援者が頑張っているという形をとっています。

↓

もっと言ってしまえば菊池さんには悪いんですが…

えー、菊池さんはあくまで、今回の実験の聾学生役という役目なので、字幕が例えば止まったりして混乱したとしても、菊池さんのことはほおっておくという言い方ヘンですが、実験上はそうなったということを理解してもらおうということですが…

えーと、受講している学生は、実験が失敗したから受講できなくなったということだと困るわけですから、常に支援者が頑張っているという形を取っています。

## 6-2. パラ言語情報の付与

話し言葉においてはパラ言語情報は、声の大きさ、早さ、声色、イントネーション、アクセント等によって付与される。

文字言語におけるパラ言語という観点から見解を述べた論文は管見の限りでは見あたらないが、近年のパソコンの普及によって、様々なパラ言語的な表現手段が生まれているように思われる。

たとえば、フォントの大きさ、形、色の変更、太字、斜体、下線等の使用は、音声言語における声の大きさ・質・声色などによく似た機能を果たす場合がある。音声言語において強調する場合に用いられるやや大きめの声を、たとえば太字で表示するというルールを規定した場合、読み手は、話し手が今この文を強調して話しているということを理解できる。或いは、通常の声は黒で、比較的明るい感じで話されている場合の声は赤で、話しにくそうな暗い感じの声は青でというように表示方法を規定することによって、音声言語に頼ることのできない聴覚障害者も、話し手がどのように情報を提供しているのかというイメージをつかむことができるだろう。

また、一定時間中に繰り返して使用される言葉があった場合、その言葉を現在話しているテーマのキーワードとして判断し、フォントを変えるなどして、これが現在のキーワードであるということを示すことができれば、話題の省略による難読文の理解に役立つのではないかと考える。

更に、文末において、音声のイントネーションの変化が一定の範囲を超えた場合、機械的に上昇イントネーションを示すマークや、下降イントネーションを示すマークを付与することにすれば、「～ですか」という文が、質問として発話されているのか、納得したという意味で発話されているのが容易に把握できるようになるのではないだろうか。

## 7. 今後の課題

本稿において、字幕として呈示される文が何故理解しにくいのかを統語的な面から考察し、そのような問題点を含んだままで理解しやすい文として呈示するためにはどのような方策が考えられるかを



検討した。今後は、音声情報を付与した字幕とそうでない字幕との理解しやすさについての比較実験の実施、更に健聴者で有意な差が見られる提示方法が聾者にも有効であるかの検証等が必要となると思われる。

また、話し言葉がそのまま文字化して呈示されるということは、日本語では技術的な問題からこれまであまり想定されていなかったように思う。音声同時字幕システムによって、話し言葉をそのまま文字化することの問題点に、改めて気付かされることとなった。音声言語と文字言語との認知過程の差異についての実証的な研究も今後の大きな課題である。

## 付記

本研究は第1回博報堂「ことばと文化・教育」研究助成金「音声認識技術を応用した話し言葉の字幕化による聴覚障害学生支援」の研究成果の一部をまとめたものである。

## 参考文献

- 井野秀一・黒木速人・加藤士雄・渡邊括行・堀耕太郎・伊福部達（2003）「聴覚障害者の会議参加支援を目的としたリアルタイム音声字幕化システムの設計」『計測自動制御学会第18回生体生理工学シンポジウム論文集』221-224
- 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
- 菊池真里・金澤貴之・中野聡子・黒木速人・井野秀一・伊福部達・福島智（2005）「音声認識技術を活用した高等教育機関における聴覚障害者の情報保障（1）—中間支援者の修正作業に注目して—」『日本特殊教育学会第43回大会論文集』559
- 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・インタフェース・シンポジウム2003論文集』729-732
- 佐久間まゆみ・鈴木香子（1993）「女子学生の日常談話の接続表現」『国文目白』32 31-48
- 中野聡子・黒木速人・井野秀一・金澤貴之・菊池真里・伊福部達（2004）「高等教育機関における聴覚障害学生のための遠隔型音声字幕化システムの活用」『日本特殊教育学会第42回大会論文集』359
- 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（2005）「音声認識技術を活用した高等教育機関における聴覚障害者の情報保障（2）—音声言語と文字言語の性質の違いを中心とした検討—」『日本特殊教育学会第43回大会論文集』556
- 福島智・中野聡子・金澤貴之・黒木速人・井野秀一・伊福部達（2006）「音声認識技術を活用した字幕呈示システムの開発研究及び運用における諸課題—利用者の観点を中心に—」群馬大学教育学部紀要 人文・社会科学編, 第55巻（印刷中）
- 加藤重弘（2001）「談話標識の機能について」『東京大学言語学論集』20 121-138
- 南不二男（1974）『現代日本語の構造』大修館書店
- 野田尚史（1989）「文構成」『日本語と日本語教育』67-96 明治書院
- 森山卓郎（1989）「文の意味とイントネーション」『日本語と日本語教育』172-197 明治書院
- 中野聡子・牧原功・金澤貴之・中野泰志・新井哲也・黒木速人・井野秀一・伊福部達（2007）「音声認識技術を用いた聴覚障害者向け字幕提示システムの課題—話し言葉の性質が字幕の読みに与える影響—」『電子情報通信学会論文誌』Vol.90-D 808-814

## Problem of Speech Recognition Technology — from the real-time captioning system —

MAKIHARA Tsutomu, KANAZAWA Takayuki,  
FUKUSHIMA Satoshi, INO Syuichi, IFUKUBE Tohru,  
KUROKI Hayato, NAKANO Yasushi, NAKANO Satoko

Through the use of speech recognition technology from the real-time captioning system, we discovered that the hearing impaired feels that presented speeches are difficult to understand.

In this research, we exhibited sentences which had characteristics of spoken language and had test subjects read them.

As a result, we obtained the conclusion that the following features seen in the spoken language cause the difficulty of the understanding of sentences. The features are “Disappearance related to responding”, “Omissions of subject and predicate, etc.”, and “Syntax violation of restriction when complex sentence is generated”, “Omission of subject”.

In literary language, the intonation and pose etc. has been deleted. We showed the possibility of becoming easy to understand the sentences by giving those information as the sight.