

## リアルタイム要約としての要約筆記における要約の手法

福島 孝博

追手門学院大学 文学部英語文化学科

江原 暉将

通信・放送機構 渋谷上原リサーチセンター/NHK 放送技術研究所

### 1. はじめに

1999年11月に通信・放送機構渋谷上原リサーチセンター主催の国際ワークショップ「聴覚障害者のためのテレビ字幕制作に関する国際ワークショップ」が開催された([1])。この国際ワークショップでは、耳の不自由な方々への情報保障として、要約筆記、手話通訳が行われ、発表は、日本語、英語で行われたが、英語は日本語に、日本語は英語に同時通訳された。また、講演時には、内容を示したスライドが使われたが、日本語のスライドには、対応する英語のスライドが同時に示された。この結果、参加者にとっては、日本語音声、日本語筆記、日本語手話、英語音声、英語筆記の5種類のチャンネルによる情報の保障がなされ、これだけのチャンネルで情報保障を行ったおそらく最初の国際会議ではなかったかと思われる。

要約筆記は、聞こえてくる音声をその場で、即時に(リアルタイムに)パソコン入力または手書きで文字化を行い、スクリーンなどに拡大して見せ、文字を読んでもらうことにより聴覚障害者への情報保障を行う活動である([2],[3])。

この国際ワークショップでの要約筆記の結果をリアルタイムでの要約とみなし、発表の書き起しと比較することにより、リアルタイム要約としての要約筆記における要約の手法を分析し、その調査結果を報告する。

### 2 調査対象

1999年11月に通信・放送機構渋谷上原リサーチセンター主催で開催された「聴覚障害者のためのテレビ字幕制作に関する国際ワークショップ」における講演者の発表は、その場で、リアルタイムに要約筆記者によって要約筆記された。ここでは、その講演うちの10件の発表の内容を書き起したものと、会場で要約筆記者によって入力された要約筆記の内容を調査と分析の対象とした。発表内容の種類と使用された言語は次のよう

ある。<sup>1</sup>

発表番号	種類	使用言語
1	挨拶	日本語
2	挨拶	日本語
3	基調講演	英語
4	講演	日本語
5	講演	日本語
6	講演	日本語
7	講演	日本語
8	講演	英語
9	講演	日本語
10	講演	日本語

表1 発表の種類と使用言語

### 3 分析結果

#### 3.1 文字数、速度からみた分析

各挨拶、講演をその発表時間(但し、質疑応答は一切含まれていない)、要約筆記された結果に基づいて、以下の事項について計算を行った。

- ・ 「要約筆記速度」: 要約筆記結果における一分間当たりの要約筆記された文字数
- ・ 「発話速度」: 書き起こしの一分間当たりの文字数
- ・ 「要約率」(文字数): 要約筆記文字数を書き起こし文字数で割ったもの
- ・ 「要約率」(文数): 要約筆記文数を書き起こし文数で割ったもの

「要約率」は、その値が高いほど要約がされないことを示すことになる。また、同時に、1文当たりの文字数も計算した。最後に、各項目の平均値を算出した。結果は、表2のとおりである。

表2から、要約筆記速度は、発話速度ほどばら

<sup>1</sup> 実際には、もう3件の発表があったが、著作権の関係上調査の対象とはしていない。

つきがないことがわかる。最も大きな値を最も小さな値で割ったものを計算すると、要約筆記速度では、1.28 であるの対して、発話速度では、1.57 となる。

また、文数での要約率は、多くの場合、1.0 以下であるが、1.0 以上になる場合もあり、要約筆記をすることにより、文数が必ずしも減るとは限らないことがわかる。

発表 番号	要約筆記結果		書き起こし		要約率	
	1文 当り 文字 数	要約 筆記 速度	1文 当り 文字 数	発話 速度	文字 数	文数
1	42.18	178.46	75.00	346.15	0.52	0.92
2	47.85	164.41	67.32	338.06	0.49	0.68
3	27.04	179.60	31.92	282.31	0.64	0.75
4	27.45	160.97	29.28	345.86	0.47	0.50
5	39.68	164.54	53.61	219.66	0.75	1.01
6	33.87	150.46	52.50	251.70	0.60	0.93
7	36.05	166.14	53.05	262.52	0.63	0.93
8	26.13	191.66	36.91	325.29	0.59	0.83
9	57.39	192.94	89.76	313.19	0.62	0.96
10	41.51	162.60	79.12	247.96	0.66	1.25
平均	37.91	171.18	56.84	293.27	0.59	0.88

表 2 発表の速度と要約率

発表速度が、各調査項目とどの程度相関性があるのかをみるために、相関係数を計算した。

要約率 (文字数)	-0.883229
要約率 (文数)	-0.655347
要約筆記での1文文字数	0.043766
要約筆記速度	0.420846

表 3 発話速度との相関関係

表3からわかるように、発表速度と文字数での要約率は高い相関性がみられ、文数での要約率ともある程度の相関性がみられる。しかし、発話速度と要約筆記での1文の文数は、相関性がないことがわかり、要約筆記速度との相関性もあるとは言えない。

相関性の高い発話速度と要約率 (文字数) との関係散布図にすると図1のようになる。

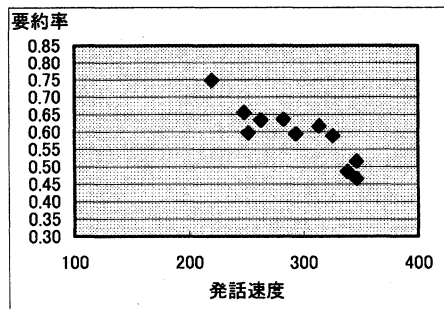


図 1 発話速度と要約率 (文字数) の関係

表3、図1からいえることは、発話速度が大きくなる、つまり、話し手の話す速度が速くなると、要約率が低下する、つまり、より多くの要約がなされることになる。これは、発話速度が上がっても、リアルタイムでのパソコン要約筆記の速度には限度があることに原因していると思われる。また、発話速度が遅い場合、特に250文字以下の場合、要約筆記がかなり追いついており、要約率が高い値であることがわかる。

### 3.2 表現からみた要約筆記の要約手法

次に、要約筆記結果を書き起こしされたテキストと比較することにより、要約が行われている場合、どのような要約が行われているのかを、分析してみる。この分析に使用したのは、発表番号1の挨拶である。

まず、書き言葉には見られない、話し言葉特有の言いよどみ、文途中での言い換えが少しみられた。例をあげると以下のとおりである。

言いよどみ 括弧部分

(あの、えー)

(えー)

文途中での言い換え 《 》で囲まれた部分

「…《この》(ま) これで…」

「…いろいろな《広い》実際に…」

発表中で要約が行われた箇所は、全部で27箇所あった。それらの要約箇所を要約の手法別に大まかな分類と細かく分類できる場合は、更に詳細として分類した。分類の結果をまとめると、以下のようになる。

分類	詳細
単語、語句レベルでの要約	語句の単純化
	丁寧表現の簡略化
	指示的な表現・時間表現の削除
	修飾語句の削除
節レベルでの要約	
意味を考えての言い換え	
その他	複数箇所の削除・言い換え
	定型的な言い換え

表 4 要約筆記での要約手法の分類

次に、実際の要約箇所いくつかをこの分類に従って列記する（より詳細な分析は[4]を参照）。但し、言いよどみは削除して示している。また、2つの行が対となっているが、上が書き起こし（発話内容に対応するもの）であり、下の行が要約筆記結果であり、行中の空白部分が削除された部分となっている。

### 1 単語、語句レベルでの要約

#### ・語句の単純化

「いかにして」  
「いかに」

「有効になるように作ることができるかと。」  
「有効に 作れるか。」

#### ・丁寧表現の簡略化

「ご参加いただきまして、ありがとうございます。」  
「ご参加 ありがとうございます。」

「プロジェクトとして進めてまいりました。」  
「プロジェクトとして進めています。」

「まだ問題を抱えておりますので、」  
「まだ問題がありますので」

#### ・指示的な表現・時間表現の削除

「そういうことをその字幕を」  
「 字幕を」

「そちらのユーザーの側のご意見というようなことを」

「 ユーザーの 意見など、」

「デジタル放送が1年後に始まるということに」  
「デジタル放送が 始まるということで、」

#### ・修飾語句の削除

「幅広くいろいろいただいて、」  
「幅広く いただいて、」

「それから実際に障害をお持ちの方」  
「 障害をおもちの方」

### 2 節レベルでの要約

「今ちょっと紹介がありましたけれども、このプロジェクトですね、」

「 このプロジェクト、 」

「当初立てた目的の範囲でいえばかなりの程度できあがってきたと。」

「 かなりの程度できあがっています。」

「このシンポジウムを機会にして、またいろいろご意見を賜れたい〜んありがたいというふうに思っております。」

「 いろいろご意見をお願いしたいと思います。」

### 3 意味を考えての言い換え

「このプロジェクトをさらに実りあるものにするということの目的で、」

「 いいものにする目的で、」

「いろいろな意見交換をぜひ活発にしていればありがたいと、いうふうに存じます。」

「 活発な意見をお願いします。」

### 4 その他

#### ・複数箇所の削除・言い換え

「いろいろな方にいろいろなご意見をいただいたりしながらこのプロジェクト進んでおるわけですけども、」

「いろいろな助言を いただきながら 進んでいますが、」

「できるだけ合わさった形でこのプロジェクトの内容が役に立つようにと我々期待しているわけですが。」

「できるだけ合わせた形で プロジェクトが役立つように 期待しています。」

#### ・ 定型的な言い換え

「どうも、本日はたくさんのご参加をありがとうございました。」

「 ありがとうございます。」

分類された要約をみると、以下のような特徴があると言える。

- ・ リアルタイムでの入力のため、漢字変換に時間をかけていない場合がみられ、語句が平仮名のまま残されている場合がある。  
例 「障害をおもちの方」、書き起しでは、「障害をお持ちの方」となっている。
- ・ 単純な語句の簡略にとどまらず、いろいろな種類の要約方法がなされていることが、要約表現の分類からわかる。
- ・ 指示的な表現や、修飾語句が削除される
- ・ 挨拶（講演）に見られる丁寧表現が、平易化されている。
- ・ 場合によっては、節の単位で削除され、意味をとって大幅な簡略化が行われることがある。
- ・ 要約の手法は、1文を要約する際に、1つの手法に限らず、いくつかの手法を組み合わせで用いている。
- ・ 定型的と思われる簡略化もみられた。

#### 4 考察

テレビニュース番組における原稿と字幕を分析した結果では、字幕に見られる要約手法として文頭・文末の表現を簡略化、文末の丁寧表現の平易化などがあった([5])。それに対して、パソコン要約筆記の結果では、同様な要約手法も見られたが、全体をとおしてみると、文末よりむしろ文の途中の表現を削除、言い換えしている場合のほうが多くみられた。これは、やはり、リアルタイム性を求められるためであると考えられ、パソコン要約筆記の特徴と思われる。

発話速度と要約筆記の関係を見ると、発話速度が上がると、要約筆記速度には限度があるため、

要約される部分が多くなり、要約率が下がることが分かった。発話速度は、要約率、特に、文字数での要約率とで高い負の相関関係があることが分かった。これも、リアルタイム性を求められるパソコン要約筆記ならではの問題であると考えられる。

また、要約筆記で用いられる要約手法も、多種類あり、場合によっては、一節全てを削除したり、意味をとっての短い表現への言い換えがなされていた。このように、場合によっては、大胆な削除や言い換えがあるが、要約筆記結果全体をみると、要約筆者である大田氏の著書([2])で述べられているように、要約筆者が、話し手の話す内容の要点をつかみつつ、限られた時間の中で工夫をしながら要約筆記をされていることが分かる。

今後は、今回1つの発表だけを対象として表現の調査・分析を行ったが、その他の発表についても調査・分析を進めていきたい。また、同時通訳を介しているが、英語の講演が日本語に要約筆記されているので、その要約筆記結果と原文である英語の書き起しについても、調査と分析を行っていく予定である。

#### 参考文献

- [1] Proceedings of TAO Workshop On TV Closed Captions For The Hearing Impaired People, Shibuya Uehara Research Center Telecommunications Advancement Organization of Japan. 1999 (<http://www.shibuya.tao.go.jp/workshop.htm>)
- [2] パソコン要約筆記入門 聞こえを支えるボランティア太田晴康 人間★社 1998年
- [3] 「要約筆記への招待」 大田晴康「言語」大修館書店 1999年9月号 (Vol.28 No.9) pp73-79
- [4] 「リアルタイム要約である要約筆記に見られる要約の手法」福島 孝博、江原 暉将 追手門学院大学文学部紀要 2000年12月
- [5] 「テレビニュース字幕のための自動要約」 若尾 孝博、江原 暉将、沢村 英治、丸山 一郎、白井 克彦 言語処理学会第4回年次大会併設ワークショップ 1998年3月