

テレビ生放送番組に付与された字幕に関する調査

福 島 孝 博

Survey on Closed Captions of a Live TV program

Takahiro FUKUSHIMA

1. はじめに

テレビ番組への字幕付与は、視聴覚障害者への情報保障の一環として総務省の指針の基に、NHK および民放各社が取り組みを進め、生放送でない番組については、そのほとんどにおいて字幕が付くまでになってきている。しかし、欧米の状況と比較すると、特にアメリカ、イギリスにおいては、生放送の番組にも字幕が付与されていることが当然のこととなっており、日本での字幕付与番組が増えたとはいえ、テレビ番組全体からみるとまだ字幕が付く番組の割合が多いとはいえないのが現状である。日本では、アメリカやイギリスと違って、生放送番組が多く、一方再放送番組が少ないことも、字幕付与番組が少なくなる一因となっている。NHK の総合では、64% が生放送であるのに対して、アメリカの3大ネットワークである CBS、ABC、NBC では36、32、34% であり、イギリスの BBC 1 では34% である。また、再放送番組についても、アメリカの再放送率の平均は日本の2.1倍、BBC 1の再放送率は、日本の3倍である^[1]。

これらの状況を踏まえて、2007年に総務省は、視聴覚障害者向け放送普及行政の指針を発表し、字幕付与の対象となる番組を大幅に拡大し、一部を除く生放送番組を含めた全ての番組を対象とし、平成29年度（2017年度）までに実施することを目標とした^[2]。これにより、放送事業者にとっては、生放送番組への字幕の付与が重点課題となっている。

本論文では、生放送番組の一つとして2010年に開催されたFIFAワールドカップサッカー南アフリカ大会での試合を対象として、生放送番組においてリアルタイムに付与された字幕に関して、音声の書き起こしデータとの比較を行いながら、その文字数、表示速度、固有表現の頻度を中心とする基本的な調査を行った。以下では、調査の対象、方法の詳細と調査結果を報告し、結果に対する考察を行う。

2. 調査対象

調査の対象としたテレビ番組は、2010年6月25日放送のFIFAワールドカップサッカー南アフリカ大会日本対デンマークの試合である。この番組は、日本テレビ系列において字幕付きの生番組として放送された。その試合の開始直前から試合開始後10分までの約10分間を調査の対象とした。

番組の音声は、アナウンサーは1名、解説者2名、ピッチレポーター1名で構成されている。また、字幕は、一回に2行まで、一行当たり16文字まで表示されている。字幕は、句読点および話者表示がされている。

3. 調査方法

まず、調査対象の音声を書き起こしテキストにした。また、その音声に対応して付与された字幕を書き写しテキストにした。(今後は、前者を「音声データ」、後者を「字幕データ」と呼ぶこととする。)

書き起こしにおいては、フィラー、間投詞や2者が同時に話しされる場合などでもできるだけ文字化した、また、選手名、監督名などの固有名については字幕に使われているものを音声データでも採用した。音声データには句読点ないものを、字幕データは字幕そのもの、つまり、句読点および話者表示がされたものをデータとしている。

音声データおよび字幕データの一例を示すと、以下のようである。

〈音声データ〉

さあ前半45分が始まりました

右にエンドを取った日本青いユニフォーム赤と白がデンマーク代表です

〈字幕データ〉

前半45分が始まりました。右に
エンドをとった日本、青いユ

ニホーム。赤と白がデンマーク代表
です。

この例では、音声データに対して、2枚の字幕が表示されている。字幕の表示は2行分が一度になされ、切り替わる時も2行分が一度になされる、その為ここでは表示される字幕1回分

を1枚と数えている。アメリカのテレビ番組の字幕（Closed Captions）は、表示が3行であるが、最上行がせり上がり一行ずつ縦にスクロールしながら入れ替わる方式であり、日本での字幕の表示方式とは異なっている¹⁾。

字幕データの調査では、文字数だけでなく句読点や話者表示についても、それらの頻度を集計した。句読点については、音声データにはないものであるが、字幕を読みやすくするためのものであり、話者の表示については、複数の話し手が居る場合に、誰の発言かを明確にするために行われるものである。

また、両方のデータの文字数に加えて、データの内容について、固有表現（国名、チーム名、選手、監督名）の頻度、サッカー用語の頻度についても調査をした。一方、調査対象の10分間を約25から30秒の21区間に分けて、各区間での音声および字幕の文字数を調べ、音声の発話速度、字幕の表示速度についても調査した。

4. 調査結果

4.1 基本統計値

全体の文字数は表1のとおりである。字幕を音声データからの要約と考えると、元データに対する要約率は表2のようになる。字幕そのものの、つまり、句読点、話者表示を含むデータで約56%の要約率となった。サッカーのように目まぐるしく状況が変わる内容の音声をリアルタイムで字幕化すると、音声情報の半分程度しか字幕化できていないことが分かる。

表1 音声データ、字幕データの文字数

音声（書き起こし）	文字数
句読点なし	4006
字幕	
話者表示、句読点あり	2225
話者表示、句読点ともなし	1910

表2 字幕データの要約率

	要約率
話者表示、句読点あり	0.555
話者表示、句読点ともなし	0.477

また、字幕は1回の表示を1枚として数えると、145枚が表示されて、1枚当たりの平均の文字数は15.34となった。

字幕データにおける句読点および話者表示の頻度、文字数は表3のとおりである。

1) 実際には採用されていないものの、日本のデジタル放送の字幕に関する規格においても字幕をスクロール方式にすることは可能である^[3]。

表 3 字幕データの句読点、話者表示の頻度と文字数

	頻度	文字数
句読点		
句点（。）	105	105
読点（、）	25	25
話者表示		
実況>>	23	92
北澤>>	17	68
城>>	7	21
田辺>>	1	4
合計	178	315

話者表示には、話者の名前（実況のアナウンサーの場合は「実況」）と>>のマークが使用されている。それら全部で 315 文字となり、字幕データの約 14% を占めている。

4.2 固有表現およびサッカー用語での比較

データ内容がどの程度保持されているかを調べるために、固有表現（選手名、監督名、国名、チーム名）に注目して調査を行った。選手、監督名では、日本チーム代表の場合とデンマーク代表の場合では、その頻度に差がみられた。

日本代表の選手、監督は音声データに 28 回登場し、字幕データでは 19 回、68% の割合で保持されている（図 1）。一方、デンマーク代表の場合は、音声データに選手、監督が 65 回登場し、字幕データでは 22 回となり、保持率は 34% である（図 2）。デンマーク代表の場合、音声では 11 名の選手全員が登場し頻度が多いにもかかわらず、字幕にはあまり保持されていない結果となった。

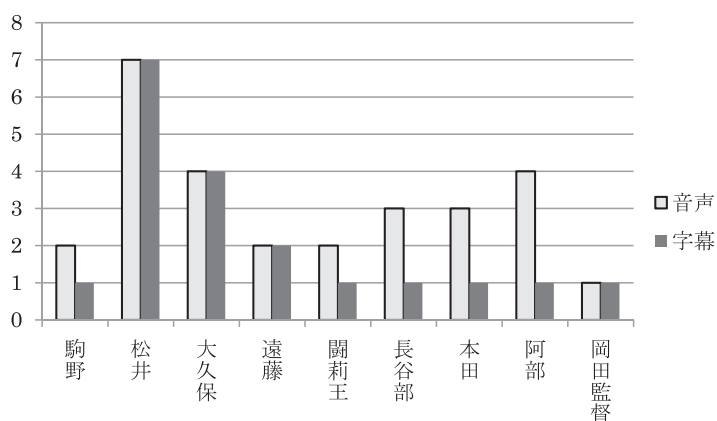


図 1 日本代表の選手、監督名の頻度

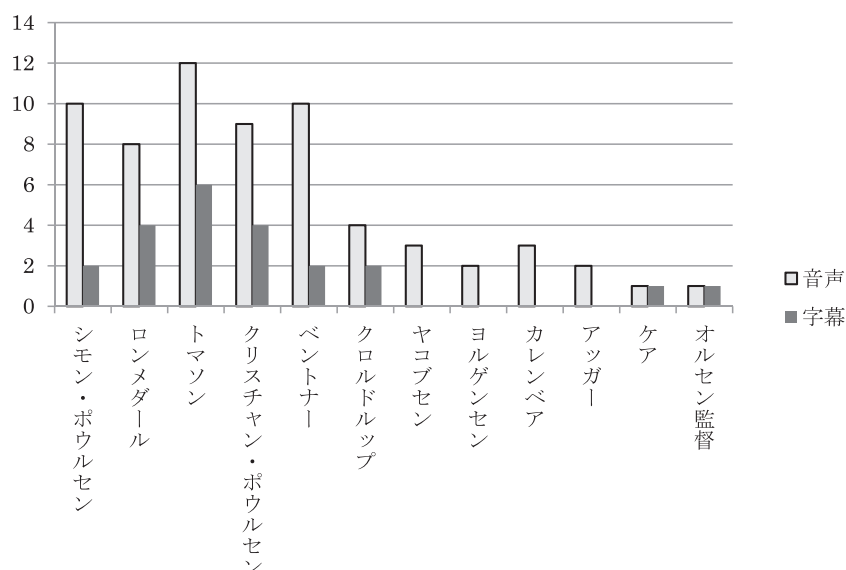


図2 デンマーク代表の選手、監督名の頻度

次に、国名、チーム名の頻度を調べると、以下の表4となり、保持率は64%と高くなっている。

表4 国名、チーム名の頻度

国名・チーム名	音声	字幕
日本	14	11
日本代表	10	4
デンマーク	17	10
デンマーク代表	6	4
フランス	2	1
オランダ	2	2
カメルーン	1	1
日韓	1	1
合計	53	34

表5 サッカー用語の頻度

サッカー関連用語	音声	字幕
エンド	1	1
オーバーヘッド	1	1
カウンター	1	1
サイドアタック	1	1
タッチライン	1	0
バックライン	1	0
ピッチ	1	1
ミッドフィールダー	1	0
ワンタッチ	1	0
セットプレー	2	2
フリー	2	1
フリーキック	2	2
プレス	2	2
ホイッスル	2	1
シュート	5	1
パス	5	5
トップ(下)	7	3
サイド(逆、右、左)	14	5
ボール	23	6
	73	33

また、フリーキック、シュートなどのサッカー関連用語の頻度も集計した。用語については、音声データに出現する語句から任意で選んだ。

サッカー関連用語全体では、保持率は、45.2% となり高いものとはなっていない。

尚、表5の「トップ（下）」とは、「トップ」または「トップ下」を表し、同様に、「サイド（逆、右、左）」は、「サイド」「逆サイド」「右サイド」「左サイド」をまとめたものとなっている。

4.3 30秒間隔に区切った比較

次に、約10分のデータを、約30秒ごとに区切り、音声データと字幕データの比較を行った。

表6 21区間ごとの音声データと字幕データの比較

区間	音声秒数	音声文字数	発話速度	字幕秒数	字幕文字数	字幕表示速度
1	31.3	186	357	32.5	171	316
2	29	175	362	26.7	147	330
3	29.8	269	542	30.1	163	325
4	23.6	140	356	19.9	129	389
5	21.9	138	378	13.4	106	475
6	29.2	225	462	8.6	46	321
7	27.1	232	514	26.8	59	132
8	31.1	226	436	30.1	80	159
9	30.5	272	535	20.8	30	87
10	30.6	218	427	32	123	231
11	29.7	190	384	32.6	145	267
12	27.2	119	263	4.7	24	306
13	29.2	177	364	37.2	93	150
14	27.2	138	304	32.7	128	235
15	29.7	186	376	28.3	107	227
16	31.2	171	329	23.9	94	236
17	30	186	372	40.5	130	193
18	29.3	172	352	48.4	143	177
19	30.5	250	492	25.3	178	422
20	27.9	217	467	7.8	41	315
21	23.5	119	304	12.1	88	436
合計	599.5	4006	401	534.4	2225	250

表6において、音声秒数が必ずしも30秒となっていないのは、発話の区切りや無音声箇所に合わせての結果である。字幕秒数とは、その区間で発話された音声に対応する字幕が表示され

るのに必要とした時間（秒数換算）である。実際の字幕は、音声から遅れて表示されており、音声区間と同じ時間経過（タイムライン）における字幕を調べたものではない。

また、発話速度は、各区間の音声データの文字数を、1 分間の文字数として換算したものであり、同様に字幕表示速度は、各区間に表示された字幕を、1 分間の字幕文字数として計算したものである。

5. 考 察

サッカーのように、動きが激しく、場面変化の多い番組においてのリアルタイムでの字幕付与は、厳しい環境での字幕制作であると言える。そのため、字幕データの文字数は、音声データの文字数の 56% となり、音声情報の約半分少ししか文字化されていないことが判明した。

自動要約の研究で用いられる、情報保持率（Retention Ratio）は、オリジナルドキュメントの情報内容（information content）に対する要約の情報内容の割合とされる^[4]。字幕に保持される内容を少しでも明らかにするために、今回の調査では、固有表現とサッカー用語に注目して、その保持の割合（保持率）を算出した。その結果、日本の選手名とデンマークの選手名では、保持率が異なるという興味深い結果となった。これは、デンマークの選手名が人によっては長いものがあり、リアルタイムの条件下では字幕化がしにくいことが原因ではないかと思われる。偶然ではあるが、ポウルセンという選手が 2 名おり、シモン・ポウルセン選手とクリスチャン・ポウルセン選手を区別する必要があったことも、デンマークの選手名の頻度が少なくなった一因だと考えられる。

サッカー関連用語については、全体としては、字幕に保持される割合は、高いものではなかった。しかし、詳細にみると、音声データにおいて頻度が 1 回の用語は 9 個あるが、保持率は 56% であり、頻度 2 回は、80%、頻度が 5 回の用語は保持率が 60% と高い保持率となっている。一方、頻度が 7 回以上の高頻度用語 3 個（「トップ（下）」、「サイド（逆、右、左）」、「ボール」）は、それぞれ 43%、36%、26% の低い保持率となっている。これは、頻度が低い語であっても、それらの多くがシュートやコーナーキックのように試合の進行上で重要な語であるため保持されたと考えられる。

次に、30 秒区間に区切つての結果をさらに詳しく分析する。まず、区間ごとの文字数であ

表 7 30 秒区間の文字数での統計値

	音声文字数	字幕文字数
平均値	190.8	106.0
標準偏差	45.5	46.3
相関係数	-0.0032	

表 8 30 秒区間の速度での統計値

	音声発話速度	字幕表示速度
平均値	398.8	272.8
標準偏差	78.5	105.1
相関係数	-0.20999	

るが、音声、字幕の平均値、標準偏差値、2セットのデータの相関関係係数は表7とおりである。

同様に、区間ごとの速度、音声では発話速度、字幕ではその表示速度における、平均値、標準偏差値、2セットのデータの相関関係係数は表8とおりである。

字幕データは、表示速度においてばらつきが大きくなっている、つまり表示速度が速くなったり遅くなったりしていることを示している。また、音声と字幕はその文字数および速度において、相関係数は相関があるといえる数字となっていない。

音声の発話速度は、これは標準的なニュース番組でのアナウンサーの発話速度と同程度であるが、一般的には一分間に300から400文字程度であると言われており、今回の音声は速いほうである^[5]。一方、字幕の表示速度は、分速273文字で、字幕放送の文字数制限である4秒で30文字、一分間で450文字から考えると、特に速いとは言えない^[6]。しかし、映画の字幕の世界では、1秒4文字がルールとされており、この基準（1分240文字）と照らし合わせると、今回の字幕は速いものとなる^[7]。また、NHKのテレビ番組「手話ニュース」の字幕速度の目安である180文字／分と比べても字幕の速度は速い^[5]。

次に、区間ごとの字幕の要約率と発話速度の変動は図3のとおりである。

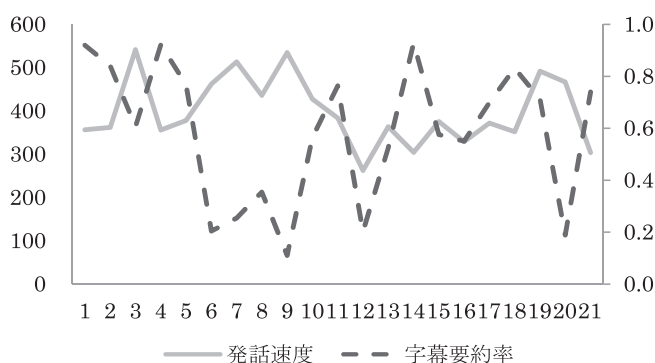


図3 30秒区間での発話速度と字幕の要約率

字幕の要約率は、情報内容の保持の点から発話速度に関係なく、ある一定の率であることが望ましいが、そうはなっておらず、かなりのばらつきがみられる。両者の相関係数をみると0.4693となり、負の相関があると言える。つまり、発話速度が速くなると、要約率は下がる傾向にあることとなる。特に区間6から区間9については、発話速度が400文字／分以上あるのに要約率は0.4以下となっており、音声速度に字幕が追いついていない状態を示している。これは、音声の発話速度が速い場合にどのようにして字幕の要約率を下げることなく字幕制作するのが課題であることを示している。

6. 今後の課題

本研究では、サッカーの試合に付与されたリアルタイム字幕と音声を書き起こしたものを、文字数や固有表現を中心にして比較した。今後は、これらの視点だけでなく、リアルタイムに付与された字幕の読みやすさ、つまり、字幕の言語表現そのものの特徴を明らかにする研究を進める。関連した研究としては、聴覚障害者への情報保障活動である要約筆記における要約の特徴に関するもの^{[8], [9]}、リスピーク方式によるリアルタイム字幕制作における言い換えの効果に関するものなどがある^[6]。

また、今回はサッカーの試合番組だけを調査したが、他の生中継のスポーツ番組、例えば、相撲、野球、ゴルフの中継番組等における字幕についても調査の対象とする予定である。

最後に、テレビ番組への字幕付与は、要約筆記とともに聴覚障害者への情報保障の柱であると言われている。今よりもさらに字幕化を推進するためにも、スポーツ中継番組だけではなく、現状では字幕の付与が困難、または付与されていないが情報保障の上で重要だと考えられる番組、例えば選挙に関する番組等への字幕の付与が可能なかどうかを調査することが次の課題である。

参考文献

- [1] デジタル放送時代の視聴覚障害者向け放送に関する研究会報告書，総務省報道資料，http://www.soumu.go.jp/menu_news/s-news/2007/pdf/070330_19_ts2.pdf, 2007.
- [2] 視聴覚障害者向け放送普及行政の指針の概要，総務省，平成 19 年 10 月 30 日策定，http://www.soumu.go.jp/main_content/000030361.pdf, 2007.
- [3] 八木伸行，吉村俊郎，加井譲次郎，『デジタル放送技術入門』，オーム社，2005.
- [4] 森辰則，情報編集のためのテキスト処理技術，人工知能学会 情報編集研究会第二回研究会チュートリアル資料，2010 年 3 月.
- [5] 上村博一，『字が話す目が聞く』，新樹社，2009.
- [6] 松井淳，本間真一，小早川健，尾上和穂，佐藤庄衛，今井亨，安藤彰男：言い換えを利用したリスピーク方式によるスポーツ中継のリアルタイム字幕制作，電子情報通信学会論文誌 D-II Vol.J 87-D-II No.2 pp.427-435, 2004.
- [7] 太田直子，『字幕屋は銀幕の片隅で日本語が変だと叫ぶ』，光文社，2007.
- [8] 太田晴康，「要約筆記」の手法上の構造と課題，静岡福祉大学紀要第 1 号，pp.21-31, 2005.
- [9] 福島孝博，江原暉将：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，追手門学院大学文学部紀要 36 号，pp.47-57, 2000.