

音声認識で文字化されたテキストにおける 誤認識と誤変換 － 漢字における問題を中心に －

増地ひとみ¹

Misrecognition and Misconversion of Text Transcribed
by Automatic Speech Recognition:
Focusing on Problems with Kanji

Hitomi MASUJI

【キーワード】 音声認識、文字化、文字起こし、誤認識、誤変換、Speech Recognition

1. はじめに

近年の音声認識と自動文字化²技術の発展により、人が発した音声や、録音された発話を、パソコン等を用いて容易に文字化できるようになった。「音声入力」「自動文字起こし」等と呼ばれるものである。音声の文字化は、インタビューや会話データの分析、放送における字幕制作、講演・会議等の記録作成など、さまざまな場面で需要がある。この音声の文字化作業は以前は「テープ起こし」とも呼ばれ、人間の手作業によって行われてきた。筆者自身も、この音声文字化の業務に25年携わってきている。

人間の手作業による場合、文字化に必要な時間は元の音声の長さの4～6倍程度とも言われ³、多くの時間と労力が必要である。テレビやラジオの業界では「まさに「地獄」」⁴とさえ形容される業務である。それが最近では、音声認識と自動文字化のシステムを利用することで、作業時間を大幅に短縮することができるようになった。生の音声や、録音した音声データが短時間でテキストに変換される。人間が手でタイピングする段階を挟まないため作業効率が格段に上がったことに加え、パソコンのほかスマートフォンで利用できる手軽さも利点の一つである。アイデアをメモする代わりにスマホに向かって話しかけ、文字化する光景も見られる。そのための「スマホ用文字起こしアプリ

1 h-masuji@otemon.ac.jp 専門は日本語学（現代の文字・表記）

2 人間の手によるタイピングを経ずになされた文字化を、本稿では「自動文字化」と呼ぶ。

3 三島剛ほか（2021）p.118

4 小林祥子（2021）p.566

り」を発売している事業者も多い。

一方で、課題も指摘されている。代表的なものが、誤認識と誤変換である。これらの精度の問題により、現在、音声認識で文字化されたテキストをそのまま直ちに使用することはできない。人手による修正が必要である。もっとも、人間が音声を文字化した場合であっても、聞き間違いや誤変換は起こる。しかし、音声認識によって文字化されたテキストにおける誤認識と誤変換には、人間による聞き間違いや誤変換とは性質を異にするものも多く含まれる。誤変換の原因といえば同音異義語の存在が挙げられるが、そればかりではない。筆者は文字化の作業に従事する中で、最近の音声認識と文字化における誤認識と誤変換を目の当たりにしている。

先行研究においては、誤認識と誤変換が問題視されてはいるものの、その実態を具体的に詳細に報告したものは見当たらない。本稿では、音声認識を利用して文字化されたテキストにはどのような誤認識や誤変換が見られるのか、漢字における問題を中心に報告する。

2. 音声を文字化する作業と音声認識技術、本稿で扱うテキストに関して

人間による発話の音声文字化して記録する作業は、手書きで文字に起こす時代を含めれば古くからおこなわれてきている。本稿で扱うのは、この10年ほどで急速に発展したデジタル技術による音声認識と、認識結果が自動文字化されたテキストである。筆者が1人のユーザーの立場から分類すると、デジタル技術による音声認識には次の2通りがある。

- ① 人間の発話をタイムリーに認識するもの
- ② 録音された音声データを認識するもの

筆者が文字起こしの作業に携わる中で、20年ほど前であったらうか、パソコンの音声認識機能による文字化が可能であることを知り、試してみた。カセットテープに録音された音声を聞きながら、同時にそれを読み上げてパソコン（に搭載されたワープロソフト Word をオフライン環境にて使用）に認識させるのである。この音声認識は上記①に当たる。認識された音声は、直ちに文字になって Word 上に出現する。当時は、録音された音声を聞いて1文字ずつ手で入力し、文字化していた頃である。キーボードに触れることなく文字化されたテキストが生成されていく様子を目の当たりにした時の驚きと感動は、忘れることができない⁵。

その後しばらくは、この①の方法で文字化をおこなっていた。全てを手作業で文字にしていた頃に比べると格段に作業効率が上がり、何よりも身体的な負担が大幅に軽減された。しかしながら、オフライン環境でのパソコンによる文字化作業には音声の不認識や誤認識、誤変換が極めて多く、誤認識されやすい語は他の語に言い換えながら読み上げ、後でまとめて置換するなどの工夫が必要であった。この間に、音源はカセットテープから MD/CD、デジタルデータ（mp3、wav など）へと変化した。その後、Google ドキュメントの「音声入力」機能を利用して、録音データをパソコンで再生しながら同時に認識させて文字化をおこなった時期もある。この方法はオンライン環境で

5 当時はこの音声認識を利用するにあたり、自分の声を何度か登録して学習させる必要があったが、現在はその作業は不要となっている。

の作業であるためか、オフラインの頃と比べると誤認識や誤変換がかなり減少した。

さらに、近年になって②の方法が実用化され、一般向けに提供されるようになった。デジタルデータの音源をアップロードすることで、録音されている音声は文字に変換されるのである。このようなシステムや「文字起こしソフト」は複数の事業者が提供しており、原則有料であるが、例えば「5分までは無料」とする事業者や「無料トライアル」を提供する事業者も存在する。筆者は、Microsoft 365 の Word に搭載されている「ディクテーション」機能を用いて、デジタルデータからの文字化をおこなうようになった。現在も継続して利用中である。1か月あたり音声データ 300 分まで利用でき⁶、60 分から 90 分程度の音声は、3～5 分程度で文字化できる。本稿で扱うのは、この Word の機能によって自動文字化されたテキストである。

3. 先行研究－音声認識技術、技術の使用等に関して

音声認識技術の変遷については、河原達也（2013・2018）に詳しく述べられている。音声認識の研究が開始されたのは、2023 年の現在から遡れば 60 年以上も前とのことである（河原 2013、p.364）。ここでの「音声認識」は、文字化が目的ではないものも広く含む。河原（2013）は、音声認識の技術は飛躍的な進歩を見せたとしながらも、音声を正確に認識させるうえでの困難さや課題を述べている。それでも、「音声の書き起こし」について、2013 年時点において「ニュース番組や議会・学会講演など、公の場で話される音声に関して、個別のモデル化を行うことで実用的なレベルに達しつつある」と評していた（p.367）。一方で、「現在（筆者注：2013 年）の音声認識はかなり高度になったとはいえ、しょせん外国語話者の域を出ない。一般人の話し言葉にはほとんど対応できないし、騒音下ではとたんに性能が低下する」と当時の状況を報告している（p.367）。そして、河原（2018）は音声認識技術の 2018 年当時における最先端モデルを解説した論考であるが、その中では「IBM と Microsoft が電話会話音声認識のタスクで約 95 % の認識率を達成し、「人間と同等の認識精度を実現した」と発表した」ことに言及している（p.6）。もっとも、IBM と Microsoft 両社の言う「電話会話音声」は英語によるものと推測され、日本語の場合はまた事情が異なるであろう。実際、日本語のインタビューにおける認識精度（単語誤り率）は 19.5 %、記者会見では 23.7 % であったとの報告もある（三島ほか 2021、p.119）。難聴者向けに「みえる電話」サービスを開発した河田隆弘・廣橋道夫（2018）も、認識精度の向上を大きな課題としていた（p.15）。それでも、音声言語処理の分野で継続して日々研究がなされていることを踏まえれば、2024 年となった現在、音声認識における認識率は言語を問わずさらに向上しているはずである。それにもかかわらず、実際に音声認識と文字化のシステムを使用していると、現在も多くの誤認識が見られる。本稿は、その実際を報告するものである。

さらに、たとえ認識率が向上しても、日本語の場合、認識した音声を文字化しようとする場面においては、その先に誤変換の問題も存在する。本稿の焦点は「誤認識と誤変換」にあるが、音声認

6 2024 年 1 月現在。「ディクテーション機能」は、Microsoft 365 のサブスクリプション版で利用できる。

識によって文字化されたテキストには誤変換が非常に多く見られる。特に自然会話における音声を正確に認識し適切に変換することには、人間が行う文字化においてさえ困難が伴うが、音声認識と文字化によるテキストには、人間による誤変換とは性質が異なるものも多い。それが一因となって、音声認識による文字化のシステムを活用したくとも十分に活用できない状況も生じる。例えば、聴覚障害のある子どもへの教育の現場である。大塚とよみ・安田喜一（2022・2023）は、聾学校での教育を充実させるために AI 音声認識文字変換システムを導入し、教員に対してはその効果を、保護者に対してはニーズを、アンケート調査によって尋ねた研究である。教員に対する調査結果の自由記述（大塚・安田 2022、p.76）からは、誤変換が多いことが文字化システムの活用の妨げになっていることがうかがえる。このシステムが役立つことは認めながらも、「誤変換が減少すれば（正確に変換されれば）活用したい」等の条件付きの意見も多かったという（p.77）。これらの調査結果を踏まえ、今後の課題として、「誤変換への対応」が挙げられている（p.78）。保護者の自由記述内容にも誤変換に言及するものが複数見られ、「誤変換があると逆に障害になる」という記述もあったとのことである（大塚・安田 2023、p.88）。

音声認識による文字化の誤変換が減少することで恩恵が及ぶ分野は、他にも多く存在する。例えば、音声認識と文字化が活用されている場として、テレビ番組の制作現場（三島ほか 2021）や公共交通機関の車内アナウンス（永峯蒼士ほか 2021）等もあり、誤変換を減少させることには社会的な意義がある。もちろん議事録作成やメモなど、システムが利用されるあらゆる場面に資するものである。そして、先述のとおり誤変換の前には誤認識の問題があるため、誤認識の減少も同時に目指すべきである。一方で、音声認識の研究が始められて 60 年経ち、研究が継続されているにもかかわらず誤認識や誤変換がなくなる現状に鑑みれば、利用する側が何らかの対策を講じ、工夫をするのが得策だとも言えそうである。

利用する側において何らかの対策や工夫をするのであれば、誤認識や誤変換の現状を知ることが第一歩となるだろう。しかし、実際にどのような誤認識や誤変換が文字化システムにおいて生じているのかが具体的に報告された例はほとんど見当たらない⁷。本稿では、音声認識で文字化したテキストに見られる誤認識と誤変換に関して、特に漢字における問題を中心に報告する。本稿はまた、文字化の現場からの報告を、音声認識と文字化の精度向上に役立てる方法を模索するものでもある。情報処理、音声言語処理の分野では、本稿が提示するようなデータは必要としないのであろうか。情報処理、音声言語処理等の知見をお持ちの方は、ご教授いただければ幸いである。

4. 調査の概要

本稿で扱うテキストは、2 種類の音声を文字化したものである。音声の 1 種類目は、増地ひとみ（2021）で使用したものと同じである。筆者が研究用に録音したもので、長さは 27 分である（以

7 AI 音声認識による面白い誤変換を募るコンテストは存在し、コンテスト結果を公表する Web サイトではいくつかの実例を確認できる。また、誤変換の例を集めたサイトは多数存在するが、それらは本稿の焦点である音声認識による自動文字化ではなく、スマートフォンやパソコンの予測変換で人間がミスをした例を扱っている。

下、「音声①」)。音声の2種類目は、成人男性による講演の発話である。本研究用に提供されたもので、長さは合計240分である(以下、「音声②」)。

音声①は「お便り紹介」と称して多様な話題で構成した内容である。音声②は鎌倉仏教等にまつわる内容で、専門用語や固有名詞を含む。

音声①②とも、先述のWordのトランスクリプト機能を用いて、音声のデータ(mp3ファイル)をアップロードして2023年9月・10月に文字化した。その後、文字化されたテキストを目視確認し、期待する結果となっていない文字列を抽出した。抽出した結果はExcelで一覧化した。音声①は、増地(2021)において表記のゆれの調査に使用したものである。音声の文字化を専門とする複数の技術者が、同一の音声①を文字化した。その結果も参照し、音声①については、人間の手による文字化においても何らかの問題(ゆれの発生、使い分けの困難さ等)が生じる語か否かについても確認した。

音声①は筆者が読み上げる際、適度な速度と明瞭な発声を心がけたため、音声認識はしやすいはずである。音声②は男性の日本語母語話者による自然な発話である。話者の発声は明瞭であるが、話しことばの特徴である、語の倒置やくり返し、言い換え、フィラー等が見られる。

5. 調査結果

4で抽出した、音声認識と文字化によるテキストにおける問題点(期待する結果とは異なる文字列の出現)は、下に示すように分類できる。Aの「誤認識」は、元の音を適切に認識できていないと見なせるものである。Bの「誤変換」は、元の音は認識できているものを指す。Cの「期待する結果」は、日本語における用字用語の手引きとして広く使用されている『記者ハンドブック』(共同通信社編著、14版、2022年。以下、「ハンドブック」)の基準に沿った表記を指す⁸。

なお、筆者自身は、音声認識と文字化における誤認識と誤変換が解消することにより、生成されるテキストの表記が100%、ハンドブックあるいはその他の「表記の手引き類」に準拠した結果になることを目指しているものではない。以下に挙げるAとBに該当する誤認識と誤変換が、最低限解消されることを望むものである。

A. 誤認識によるもの

- ・ある音を認識しない(不認識)
- ・ある音を他の音として認識する(誤認識)
- ・存在しない音を挿入する(挿入)
- ・複数の原因によるもの

B. 誤変換によるもの

8 2019年に、音声の文字化を請け負う約50の事業者に関して、ホームページ閲覧やメールでの連絡により情報収集をおこなった。結果、表記に関する基準を設けている事業者34のうち25(73.5%)がハンドブックを使用していたため、本研究でもハンドブックを基準とすることにした。

- ・ 同音類義語
 - ・ 同音異義語
 - ・ それ以外（語の切り方が不適切なものなど）
- C. 誤認識でも誤変換でもないが、期待する結果とは異なるもの
- ・ 漢字に関わること
常用漢字の不使用、表外字の使用、表外音訓の漢字使用、常用漢字表にない熟字訓の使用、
当て字の使用、一般的ではない漢字の使用
 - ・ その他（ハンドブックの基準と異なる用法、表記のゆれ、文字種の使用法）

以下に、A・B・Cそれぞれの事例から一部を示す。AとBでは、音声①と音声②の例を分けた。以下、全ての表と本文における事例の提示部分において、「誤」には「誤りとは言えないが、期待する結果とは異なる」ものを含む。「正」は、正しい変換結果、あるいは期待されていた変換結果である。（ ）内は、前後関係の参照のため適宜補った。活用語の場合は、終止形にせず、テキストの文字列をそのまま抜き出す。

5-1. A. 誤認識によるもの

誤認識は、音声①で31件、音声②で222件抽出された。先に挙げた「不認識」「誤認識」「挿入」は、どのような音において起きた現象かという観点でさらに分類できる。巻末の別表1に音声①、別表2に音声②における誤認識の例を挙げる。特に「誤」「正」のいずれかに漢字を含む例に焦点を当てるが、その他、音声認識における誤認識の実状を示す典型的な例と考えられたものも提示する。なお、別表の「現象」と「原因」欄の記述は、キーボードでローマ字入力する際にタイプする文字を基準とした。例えば「シソ」が「思想」と文字化された例は、音声上は長音の添加と見なせるが、「ウ (u)」つまり「母音の挿入」としてある。

別表1・2に挙げたもの以外で、面白い例としては「(誤) お得にデラー／(正) 乙訓寺 (おとくにでら)」「(誤) シリーズをいって／(正) 退いて」「(誤) 開催入らない／(正) 介在は要らない」「(誤) はたち／20歳 (音声は「にじゅっさい)」」などが見られた。また、漢字に関わりのない部分における誤認識の例としては、「(誤) スーパームーン／(正) スーパーマン」「(誤) ノコノコ熱心に／(正) なかなか熱心に」等があった。

5-2. B. 誤変換によるもの

誤変換は、音声①で34件、音声②において159件抽出された。先に分類して挙げた「同音類義語」「同音異義語」「それ以外」は、今述べた順に音声認識による誤変換の重症度が高くなる。すなわち、「同音類義語」の正しい変換には、人間が行う場合でもしばしば困難が伴い、判定が難しい場合もある。そして、同音異義語の誤変換は、人間においては主に不注意によるミスとして発生する可能性がある。一方で、同音類義語でも同音異義語でもない「それ以外」の例については、人間

による文字化においては通常は起こりえないと考えられ、音声認識に特有の現象と想定されるものである。該当するのは、パソコンでの入力時には変換候補として出てこない漢字の組み合わせや、語の切り方が想定とは異なるものである。

以下に、それぞれの例を挙げる。／の前を「誤」、後ろを「正」として、「誤／正」の形でカギ括弧に入れて示す。

◆音声①

〈同音類義語〉「変えて／(スマホに) 替えて」「固い／(靴の中底が) 硬い」

(同音異義語)

「感じ／漢字」「持を／字を(書かない)」「変えます／飼えます」「橋／(ベランダの) 端」「面／
麺」「選択／洗濯」「海路／懷炉」「日々／ひび(とかあかざれ)」「付近／布巾」「聞きます／(やけどに) 効きます」「言って／(パチンコへ) 行って」「紙さん／かみさん(奥さんの意味)」「胃腸／
イチョウ」「依然／以前」「拗ね／すね(体の)」

〈それ以外〉

「マスク思ってた／マスクを持ってた」「オーク／多く」「を食らって／オクラって」

「神戸便／缶と瓶」「うどんお家で／うどんを家で」「金金／キンキン(のビール)」

◆音声②

〈同音類義語〉

「薦める／勧める」「書き表す／書き著す」「聖誕／生誕」「開校／開講」「教典／經典」

(同音異義語)

「隕石／姻戚」「開設／解説」「改装／開創」「階段／戒壇」「革新／確信」「換金／監禁」「関節／
間接」「寒天／観点」「機械／機会」「帰属／貴族」「今日に入る／京に入る」「携帯／形態」「検診／
献身」「謙信／献身」「公開／後悔」「高尚／交渉」「拘留／交流」「西条／最上」「最上京／再上京」
「最朝／最澄」「最長／最澄」「才長／最澄」「惨状／山上」「自習／時宗」「死傷／師匠」「自信／自
身」「実態／実体」「指名／使命」「就活／(老後の) 終活」「省庁／象徴」「浸透／神道」「真皮／神
秘」「心理／真理」「造影／造営」「早期／想起」「総研／(寺を) 創建」「送検／創建」「促進／即身」
「体型／体系」「体制／大成」「大勢／大成」「遅延／地縁」「調停／朝廷」「電報／伝法」「頭／唐」
「党／唐」「当／唐」「東映／投影」「当時／東寺」「同情／道場」「半／班」「秘宝／秘法」「不況／布
教」「豊穰／北条」「迷走／瞑想」「表す／(書を) 著す」「行きたい／生きたい」「行きました／(～
歳まで) 生きました」「映る／(都が) 移る」「折れる／おれる(「居ることができる」の意味)」「描
き／(文章を) 書き」「型／方」「悔い／杭」「組んで／汲んで」「式／士気」「軸通り／字句どおり」
「進められ／(本を) 勧められ」「住んだ／済んだ」「尋ね／訪ね」「東京都も／東京とも」「説かれ／
解かれ」「二次／虹」「素／許(もと)」「妖精／養成(※文脈から「要請」の可能性もあり、人間で
も「養成」か「要請」かの判断は難しい例)」

〈それ以外〉

「第五次派／醍醐寺派」「アンドロメダ性／アンドロメダ星」「オリオン生／オリオン星」「北条良時／北条義時」「さねとも／実朝」「北条雅子／北条政子」「物延市／物部氏」「韓武天皇／桓武天皇」「水滸天皇／推古天皇」「銃段階／10段階」「12で死／十二弟子」「重大弟子／十大弟子」「14／重視」「にに通って／に似通って」「雄平されて／幽閉されて」「京成市／形成し」「常樹市／成就し」「木はしていた／気はしていた」「気候文／紀行文」「汚念仏／お念仏」「お台紙信仰／お大師信仰」「長岡教／長岡京」「核月刊／隔月刊」「中高の祖／中興の祖」「神皮／神秘」「呉国／護国」「年物／念仏」「長岡教師／長岡京市」「開放し／会報誌」「解放し／会報誌」「手ん／天」「はん／班」「オゴマ／お護摩」「ルザイ／流罪」「クラス／暮らす」「コーチ／高知」

先にも述べたとおり、同音類義語は人間が文字化する場合であっても使い分けが困難な場合がある。同音異義語は、不注意による変換ミスが起こりうる。一方で、「それ以外」には、人間が文字化する場合であれば、文脈から考えて、また漢字の組み合わせ上、一般的には現れないと想定される変換結果が多く含まれる。例えば、歴史上の人物の「北条政子」などは日本で義務教育を受けていれば知り得ているであろう名前であり、「雅子」に変換することは通常は起きにくい。また、「韓武天皇」も、「見たことがない」という内省が働くであろうし、そもそも今日のパソコンでは「カンムテンノウ」と入力して変換する場合に「韓武」は変換候補として出現しない。

5-3. C. 誤認識でも誤変換でもないが、期待する結果とは異なるもの

誤認識でも誤変換でもない、つまり誤りではないものの、文字化の結果が期待されるものにならない場合は極めて多い。先述のとおり、それらは「漢字に関わること」と「その他（ハンドブックの基準と異なる用法、表記のゆれ、文字種の使用法）」に大別できる。「漢字に関わること」は、さらに下位分類できる。各々につき、例を挙げる。

／の前が音声認識と文字化による変換結果である。／の後ろは、ハンドブックに準拠しておこなう文字化作業で採用される、期待される表記である。このCについては、音声①と②とで共通する例が複数あるため併せて扱う。

【漢字に関わること】

〈常用漢字の不使用〉

「あかすり／あか擦り」「あげて／（証拠を）挙げて」「あたって／当たって」「あたり／（うちの）辺り」「あとで／後で」「あわてる／慌てる」「いのり／祈り」「おいて／置いて」「おきて／（事が）起きて」「おすすめ／お勧め・お薦め」「おすそ分け／お裾分け」「かける／（発破を）掛ける」「かす／（本を）貸す」「かねて／兼ねて」「気づいて／気付いて」「さびくさい／さび臭い」「すこしずつ／少しずつ」「すみ／住み」「そのへん／その辺」「たずね／訪ね」「ついて／着いて」「つきません／尽きません」「つけて／（気を）付けて」「とって／取って」「とる／（年を）取る」「中ぞこ／

(靴の) 中底」「根ざした／根差した」「のど／喉」「はきます／吐きます」「ふとさ／太さ」「まる聞こえ／丸聞こえ」「みない／見ない」「みる／見る」「やすい／安い」「わからない／分からない」「ごほん／ご本」「たいてい／大抵」「だめ／駄目」「とうきょう／東京」「満がん／満願」

〈表外字の使用〉

「呆れる／あきれる」「躰／しつけ」「杏／アンズ」「嘘／うそ」「噂／うわさ」「幼馴染／幼なじみ」「蝦蟇／ガマ」「穢れた／汚れた (けがれた)」「肴／(酒の) さかな」「鮭／サケ」「喋る／しゃべる」「揃える／そろえる」「叩いて／たたいて」「溜まる／たまる」「辻褄／つじつま」「繋がって／つながって」「椿／ツバキ」「雛／ひな」「鶉／ヒヨドリ」「味噌汁／みそ汁」「虚しい／むなしい」「碌に／ろくに」「喧嘩／けんか」「口喧嘩／口げんか」「糞／ふん」「葡萄／ブドウ」「味噌／みそ」「贅沢／ぜいたく」「生姜／ショウガ」「胡椒／コショウ」「鞆／かばん」「癌／がん」「綺麗／きれいい」「誤謬／誤びゅう」「澁刺／はつらつ」「呆然／ほうぜん」「銀杏／ギンナン」「麻痺／まひ」

〈表外音訓の漢字使用〉

「漁った／あさった」「良い／いい」「お寿司／おすし」「可愛い／かわいい」「凄技／すご技」「凄い／すごい」「微笑ましかった／ほほ笑ましかった」「諸刃／もろ刃」「輩／やから」「私達／私たち」「怪我／けが」「人参／ニンジン」

〈常用漢字表にない熟字訓の使用〉

「美味しい／おいしい」「蕎麦／そば」「流行って／はやって」「火傷／やけど」

〈当て字の使用〉「上手く／うまく」

〈一般的ではない漢字の使用〉「酷／こく・コク」

【その他 (ハンドブックの基準と異なる用法、表記のゆれ、文字種の使用法)】

〈ハンドブックの基準と異なる用法〉

「ハンドブックの基準と異なる用法」は、次のように分類できる。

- ・ 外来語の表記：「アイディア／アイデア」
- ・ ひらがなの使用 (カタカナが通例の語)：「みかん／ミカン」
- ・ カタカナの使用 (ひらがなが通例の語)：「ゴミ／ごみ」「ウロウロ／うろうろ」
- ・ カタカナの使用 (常用漢字の不使用)：「カゴ／籠」「シップ／湿布」「ダメ／駄目」
- ・ 漢字使用 (ひらがなが通例の語)：「剣幕／けんまく」「居る／(外に) いる」「寂／(歌の) さび」「子供／子ども」「挨拶／あいさつ」「様々／さまざま」「一か八か／いちかばちか」「仕草／しぐさ」「出来る／できる (可能の意味)」「恐らく／おそらく」「風に／ふうに」
- ・ 漢字使用 (カタカナが通例の語)：「蛇／ヘビ」「餃子／ギョーザ」
- ・ 形式名詞における漢字使用：「時／(打ち合わせの) とき」「事／(楽しい) こと」「方／(隣の) ほう」「物／(餌になるような) もの」
- ・ 複合動詞後部要素における漢字使用：「こすり過ぎた／こすりすぎた」「持って行き／持っていき」

- ・ 補助動詞における漢字使用：「行きたい／(生きて) いきたい」「行く／(教育して) いく」
- ・ 補助動詞における漢字不使用：「悪すぎる／悪過ぎる」
- ・ 同音訓異表記同義語：「寝巻き／寝間着」

〈表記のゆれ〉

人間が文字化する場合であっても、ハンドブックを参照してもなお表記を決定するのが困難な語群が存在する。それらは、音声認識による文字化においてもいわゆる「ゆれ」となって出現する。ゆれにはさまざまな種類があり、増地（2021）では13種類のゆれのタイプ（佐竹式13分類、p.121、表2）を使用して分類した。本稿の調査において見られた例の一部を示す。なお、以下のうち、「匂い・臭い」以外は、人間が文字化した場合であっても表記する者によってゆれが見られた（増地 2021、p.124、表4）。

- ①漢字と平仮名の対立によるゆれ：「(風邪を) 引く・ひく」「何でも・なんでも」「今日・きょう」
- ②漢字と片仮名の対立によるゆれ：「片仮名・カタカナ」「本当・ホント」「首・クビ」
- ③平仮名と片仮名の対立によるゆれ：「ねた・ネタ」「もてる・モてる」
- ④漢字と数字の対立によるゆれ：「一回・1回」「一日・1日」
- ⑥異なる漢字の対立によるゆれ：「匂い・臭い」「付ける・着ける」「鳥肉・鶏肉」「体型・体形」「跡継ぎ・後継ぎ」「錬磨・練磨」「偏頭痛・片頭痛」「荒削り・粗削り」

〈文字種の使用法〉

「ゆれ」とは見なしにくいものをここに分類した。事例は下に示すとおりである。外来語の「ジャーナル」のアルファベット表記や「特別なモノ」はゆれとも言えそうであるが、人間が文字化する場合は、特別な意図がない限り一般的な日本語のテキストでは出現しにくいであろう。

「じゃあなる・journal／ジャーナル」「コジンマリ／こちんまり」「モノ／(特別な) もの」

6. 以上の事実から見える、音声認識による文字化テキストの特徴

以上、5章で提示した実態を踏まえると、まず「音声認識による文字化」には次のような特徴があると言えそうである。①②③は、人間が文字化する場合との大きな違いであると言える。④⑤⑥は人間にも当てはまるが、音声認識による自動文字化のシステムはこれらが人間以上に不得手であるように見受けられる。

- ①コロケーション・文脈（コンテキスト）・アクセント・イントネーション等を考えればありえない、あるいは可能性が低い認識や変換をおこなう（類推ができない）
- ②語の切り方（一語の認定）の不適切さに起因する誤認識をおこなう
- ③手入力の際には変換候補に上がらない変換をおこなう
- ④同じ文書における表記の統一はおこなわない
- ⑤話しことばの認識・変換が不得手である
- ⑥専門用語、固有名詞の認識・変換が不得手である

⑦人間が文字化してゆれる語は、やはりゆれる

そして、漢字に焦点を当てるなら、以下のような傾向が認められる。

- ⑧同音異義語の誤変換が多い
- ⑨和語における常用漢字の不使用が多い
- ⑩表外字を積極的に使用する
- ⑪類音語の誤認識が多い。そこから誤変換も発生する

最後の類音語の誤認識に関して、別表1・別表2で示した例からは、誤認識されやすい類音としてまず以下が指摘できる。

- ・ 調音位置と調音様式が同じ子音 (t と d、s と z など)
- ・ 調音様式が同じ子音 (k と g、n と m)
- ・ 後舌で発音される母音 (u と o)⁹

これらの中には、古くラジオ放送において同音語・類音語の問題が重要視されていた時代から、誤認識されやすい類音と認知されていたものが含まれる。昭和9年に設置された放送用語並発音改善調査委員会（以下「委員会」）による調査や用語の検討といった活動の中では、類音語を蒐集して整理した表の具体案が示され、そこには例えば「ナ」「ダ」「タ」「ラ」の類⁹が挙げられていたとのことである（石野博史 1991、p.42）。これらの音節は、相互に誤認識されやすい類音として注意喚起される対象であったということになる。ここから、人間に認識される類音の範囲は、調音位置や調音様式が同じであること以上にさらに幅広いのであろうことが推測できる。そして、本稿の結果を見る限り、音声認識による類音語の誤認識は、人間の認識よりも精密な、狭い範囲で起きているように思われる。

本稿の別表1・2には、子音 k と h の誤認識が散見される。後続する母音は、a、i、o である。上述した委員会の案では、「ヒ」と「キ」は「イ」「シ」「チ」と共に類音とされている。また、「コ」と「ホ」については「ゴ」と「ホ」が「オ」と共に類音とされている。しかし、母音が a の場合の組み合わせである「カ」と「ハ」には言及がない（石野 1991、pp.42-43）。ka と ha の誤認識は、音声認識に特有のものなのであろうか。

別表1・2の「原因」の欄には「母音の挿入」が一定数見られるが、これは先述のとおり、音声上は長音の添加（母音の長音化）である。「母音の不認識」についても同様に、長音の不認識の場合が含まれる。また、撥音・促音の挿入や不認識も散見される。音声認識においては、特殊拍の認識に課題があると言えそうである。そして、別表1・2において誤認識はほかにもあり、多様である。さらに実例を集めれば傾向が見えるかもしれない。あるいは、音声認識による誤認識には決まったパターンが見いだせない可能性もあるだろう。

9 木村琢也・小林篤志（2010）を参照した。子音は図-2（p.180）、母音は図-4（p.181）より。

7. おわりに

以上、音声認識を利用して文字化されたテキストにはどのような誤認識や誤変換が見られるのか、特に漢字における例を取り上げて報告した。音声認識による文字化の欠点が強調される形となってしまうが、自動文字化のシステムには利点もある。例えば、人間の知識不足を補ってくれるなどである。これは筆者の実体験であるが、「シュゼンジ」の音声に対して「修禅寺」と自動文字化されたことがあった。筆者は「修善寺」とすべきではないかと考え、調べたところ、そのテキストの文脈においては「修禅寺」が正しかったのである。修善寺は地名、修禅寺は寺の名称に用いるという知識が筆者には不足しており、それを文字化システムが補ってくれる結果となった。そして、誤認識や誤変換が多いとはいっても、全てを手作業で文字化するのに比べれば多大な時間と労力の軽減につながる利点があるのは、冒頭に述べたとおりである。

一方で、6章の①～③に挙げた点では人間に及ばず、これらが音声認識による文字化のシステムが広く活用されることを妨げている。例えば、NHKは2022年10月から衆参両院の予算委員会のテレビ中継に字幕をつけているが、この国会中継の字幕の付与は「AIでは難しく、プロの字幕速記者のスキルが必要になる」という。AIには難しいとされる原因として、発話者のクセやイントネーションにより聞き取りが難しいこと、また、法令の正式名称、つまり専門用語や、同音異義語の存在が指摘されている（本間英士 2022）。これらは以前から問題視されている事項であるため、専門用語を事前登録するなどの解決策も当然ながら提案されてはいる（大塚・安田 2022、p.78）。システムを提供する事業者も、文字起こしに特化したソフトでは、専門用語や固有名詞を登録して精度を高める仕組みを提供するなどしているという¹⁰。業界あげて、改善策を模索している状況と言ってよいであろう。

今後、このシステムがいっそう汎用性の高いものとなっていくには、アクセント等を認識させて音声の面から誤認識を減少させることとともに、文脈を考慮した文字変換をさせる仕組みが必要なのであろう。本稿で提示した同音異義語の実例を見るかぎり、文脈を考慮すれば誤変換が防げるものが大多数を占める。MicrosoftによるWordのトランスクリプト機能の説明には、「AI」「ディープラーニング」の類の語は見られないのであるが、素人ながらこうした技術はAIの深層学習によって進化していくものと認識している。実際、音声認識技術は「ディープラーニングによって精度が飛躍的に向上した」と述べる記事もある（翁羽翔・渥美友里 2023、pp.43-44）。同音異義語や類音語の情報を与えてAIに学習させることは、すでになされていると推測する。さらに、単語だけではなくコロケーションの情報や文脈（コンテキスト）まで包括的な情報として与えたり、「同じ文書内では同じ表記を使う」などのコマンドも与えたりすることは可能ではないのだろうか。また、人間が表記を選択するときの「選択要因」¹¹までも学習させることはできないものであろうか、

10 スマート書記ホームページ「【画像で解説】Wordで文字起こしする2つの方法！おすすめの文字起こしソフトもご紹介」<https://www.smartshoki.com/blog/mojiokosi/howto-word-transcription/>（2024年1月11日閲覧）

11 増地（2019）にて、非外来語がカタカナ表記で出現する要因、つまり表記が選択される要因に関して総括的に述

などと夢想する次第である。その実現を待ちながら、当面は人間の側でなしうることをおこなっていくしかない。誤認識されやすい音を含む語においては発音に気をつけたり、誤変換されやすい語を一覧化しておいたりするなどであろうか。

本稿では、扱えなかったことも多い。例えば、音声認識による文字化で正確に変換される語と、正確に変換されない語の特徴を比較するなどである。また、現在複数の事業者が音声認識による文字化のシステムを提供しているが、システムによって文字化の精度は異なることが想定され、その実態も把握したいところである。音声認識と文字化に関わる者が目指す共通のゴールは、認識と変換の精度を高めることであり、その先には読みやすくわかりやすい文字情報の提供が目指されている。そのために、本稿で提示したような文字化の現場におけるデータがどのように寄与できるのかを探っていきたい。

付記

本稿は、日本漢字学会第6回研究大会（2023年12月17日）における口頭発表に基づく。

参考文献

- 石野博史（1991）「放送における類音語」『日本語学』10、明治書院、pp.37-46
- 翁羽翔・渥美友里（2023）「イチから理解 生成 AI 10 の疑問」『日経コンピュータ』10/26号、日経 BP、pp.38-48
- 大塚とよみ・安田喜一（2022）「聴覚障害のある子どもの教育の充実にに向けた AI 音声認識文字変換システムの導入効果に関する研究－聾学校教員への意識調査から」『障害者教育・福祉学研究』18、pp.73-78
- 大塚とよみ・安田喜一（2023）「聴覚障害のある子どもの教育の充実にに向けた AI 音声認識文字変換システムの導入効果に関する研究Ⅱ」『障害者教育・福祉学研究』19、pp.83-91
- 河田隆弘・廣橋道夫（2018）「音声テキスト化による新たな通話コミュニケーション支援－難聴者向け「みえる電話」サービスの開発導入」『情報通信学会誌』36(1)、pp.11-16
- 河原達也（2013）「音声認識技術の現状と将来展望」『電気学会誌』133(6)、pp.364-367
- 河原達也（2018）「音声認識技術の変遷と最先端－深層学習による End-to-End モデル」『日本音響学会誌』74(7)、pp.381-386
- 木村琢也・小林篤志（2010）「IPA（国際音声記号）の基礎－言語学・音声学を学んでいない人のために」『日本音響学会誌』66(4)、pp.178-183
- 小林祥子（2021）「「もじこ」～文字起こし地獄を救う WEB アプリ テレビ局ならではの音声認識 AI による文字起こしエディタの開発」『映像情報メディア学会誌』75(4)、pp.566-571
- 永峯蒼士・前川知輝・松本直弥・峯山賢伸（2022）「車内アナウンスをすべての人に」『2022 年度

ㄨ べた。（pp.155-156）

情報処理学会関西支部 支部大会 講演論文集』

本間英士 (2022) 「メディアインサイド 国会審議の「花形」に字幕 速記者が挑んだ高い壁」
『THE SANKEI NEWS』 2022.12.23 <https://www.sankei.com/article/20221223-224BJUIZ5FLE7H4TZIB4SZGPAI/> (2024. 1. 11 閲覧)

増地ひとみ (2019) 「「非外来語のカタカナ表記」研究の現状と今後の展望」『愛知淑徳大学論集
文学部篇』 44、pp.143-159

増地ひとみ (2021) 「「表記の手引き」類に準拠したテキストにおける表記のゆれ－実態調査と発生
理由の検討」『早稲田大学日本語学会設立 60 周年記念論文集 第 1 冊 言葉のしくみ』早稲田大
学日本語学会編、ひつじ書房、pp.117-134

三島剛・萩原愛子・伊藤均・小森智康・佐藤庄衛 (2021) 「番組制作を支援する音声認識を用いた
字起こしシステムの開発」『映像情報メディア学会誌』 75(1)、pp.118-124

別表 1 音声①における誤認識

誤	正	現象	原因
トリガー	鶏がら	ra ⇒ 長音	音節の誤認識
いれば	入れれば	re ⇒ 無	音節の不認識
の	つわもの	語頭から 3 音節 ⇒ 無	音節の不認識
あっちゃった	洗っちゃった	ra ⇒ 無	音節の不認識
で例	デレーツ	促音 ⇒ 無	促音の不認識
法話った	頬張った	b ⇒ w	子音の誤認識
は画像	若造	k ⇒ g	子音の誤認識
好きです	次です	ts ⇒ s、g ⇒ k	子音の誤認識
真っ赤	輪っか	w ⇒ m	子音の誤認識
口	愚痴	g ⇒ k	子音の誤認識
膝	きざ	k ⇒ h	子音の誤認識
ため	種	n ⇒ m	子音の誤認識
詰め	ずれ	z ⇒ ts、r ⇒ m	子音の誤認識
許す	居留守	i ⇒ yu	母音の誤認識、子音の挿入
絵画	声 ^が	oe ⇒ ai	母音の誤認識
打ちます	(ギンナン ^が) 落ちます	o ⇒ u	母音の誤認識
リストから	リスとか	無 ⇒ ra	音節の挿入
持の	物／モノ／もの	tsu ^が 挿入されている？	音節の挿入
落ちて	置いて	i ⇒ chi	子音の挿入
思想	シソ (植物)	無 ⇒ u	母音の挿入
みんな	見も (知らぬ)	mimo ⇒ miNna	母音と子音 (音節) の誤認識、撥音の挿入
逆	お客	o ⇒ 無、k ⇒ g	母音の不認識、子音の誤認識
出ます	言ってます	itte ⇒ de	母音の不認識、子音の誤認識、促音の不認識
日。Jr	秘訣		
mitsuwa artur	3 つはあるっちゃ		

別表2 音声②における誤認識

誤	正	現象	原因
決めて	清めて	yo ⇒ 無	音節の不認識
～にり	～により	yo ⇒ 無	音節の不認識
見なされいます	見なされています	te ⇒ 無	音節の不認識
示されいます	示されています	te ⇒ 無	音節の不認識
かもれません	かもしれません	shi ⇒ 無	音節の不認識
あげまよう	あげましょう	shi ⇒ 無	音節の不認識
受け継がました	受け継がれました	re ⇒ 無	音節の不認識
黒麴	袋小路	fu ⇒ 無	音節の不認識（語頭）
ウイン	封印	fu ⇒ 無	音節の不認識（語頭）
理解深めたり	理解を深めたり	wo（助詞）⇒ 無	音節の不認識（助詞）
犬	今	ma ⇒ nu	音節の誤認識
寛容疑あれ	関与を疑われ	wo（助詞）⇒ u、w ⇒ 無	音節（助詞）の誤認識、子音の不認識
喧嘩には	厳格な	g ⇒ k、kuna ⇒ niwa	音節の誤認識、子音の誤認識
高くする	託する	無 ⇒ ka	音節の挿入
その国	その奥に	o ⇒ 無	母音の不認識
着ていく	生きていく	i ⇒ 無	母音の不認識（語頭）
武器	息吹	i ⇒ 無	母音の不認識（語頭）
持って	思っ	o ⇒ 無	母音の不認識（語頭）
越し	興し	o ⇒ 無	母音の不認識（語頭）
体起こして	体を起こして	wo（助詞）⇒ 無	母音の不認識（助詞）
理解	礼拝（らいはい）	ai ⇒ i、h ⇒ k	母音の不認識、子音の誤認識
彼に	仮に	i ⇒ e	母音の誤認識
山頂	尊重	o ⇒ a	母音の誤認識
焼き物	良きもの	o ⇒ a	母音の誤認識
データ	デート	o ⇒ a	母音の誤認識
回収	解消	o ⇒ u	母音の誤認識
受ける	置ける（スペース）	o ⇒ u	母音の誤認識
逃げむどり	逃げ戻り	o ⇒ u	母音の誤認識
記事	傷	u ⇒ i	母音の誤認識
参照	参集	u ⇒ o	母音の誤認識
時代人	時代順	ji ⇒ ju	母音の誤認識（直音と拗音）
首相	師匠	shi ⇒ shu	母音の誤認識（直音と拗音）
口出し穴	口達者な	d ⇒ t、促音・拗音 ⇒ 無	母音の誤認識、促音・拗音の不認識
議員	原因	e ⇒ i、撥音 ⇒ 無	母音の誤認識、撥音の不認識
私	今年	k ⇒ w、o ⇒ a	母音と子音の誤認識
提言	体験	k ⇒ g、a ⇒ e	母音と子音の誤認識
誰たち	成り立ち	n ⇒ d、i ⇒ e	母音と子音の誤認識
異常に	非常に	h ⇒ 無	子音の不認識
日当たり	火渡り	w ⇒ 無	子音の不認識
非常、市場	地上	ch ⇒ h、s	子音の誤認識
健康	原稿	g ⇒ k	子音の誤認識
絆	傷が	g ⇒ n	子音の誤認識

別表 2 (続き) 音声②における誤認識

誤	正	現象	原因
だからわれて	計らわれて	$h \Rightarrow d$	子音の誤認識
高速	法則	$h \Rightarrow k$	子音の誤認識
新調	心境	$k \Rightarrow c$	子音の誤認識
行事	境地	$k \Rightarrow g, ch \Rightarrow j$	子音の誤認識
投票	東京	$k \Rightarrow h$	子音の誤認識
ところ	心	$k \Rightarrow t$	子音の誤認識
抜け	向け	$m \Rightarrow n$	子音の誤認識
出す／出して	(働きを) なす／なして	$n \Rightarrow d$	子音の誤認識
市代将軍	二代将軍	$n \Rightarrow s$	子音の誤認識
全権力	先見力	$s \Rightarrow z$	子音の誤認識
家で	言えて	$t \Rightarrow d$	子音の誤認識
だから	たから	$t \Rightarrow d$	子音の誤認識
ペン	天	$t \Rightarrow p$	子音の誤認識
まで揃い	が出そろい	$g \Rightarrow m$	子音の誤認識、語の切り方
敵を待って	て清まって	$y \Rightarrow w$	子音の誤認識、語の切り方
望遠、公園	御縁	$g \Rightarrow b, k \text{ 無} \Rightarrow u$	子音の誤認識、母音の挿入
秋田の	間 (あいだ) の	$d \Rightarrow t, i \Rightarrow ki$	子音の誤認識、子音の挿入
話	悲しい	$k \Rightarrow h, i \Rightarrow \text{無}$	子音の誤認識、長音の不認識
借りた	書いた	$i \Rightarrow ri$	子音の挿入
向けられて	受けられて	$u \Rightarrow mu$	子音の挿入
いたり	いったり	促音 $\Rightarrow \text{無}$	促音の不認識
走って	発して	促音の位置	促音の誤認識
実習	時宗	無 $\Rightarrow$ 促音	促音の挿入
とっていく	とおっていく (通っていく)	$o \Rightarrow \text{無}$	長音の不認識
とるべき	通るべき	$o \Rightarrow \text{無}$	長音の不認識
映画	絵が	無 $\Rightarrow i$	長音の挿入
かわいいらしい	かわいらしい	無 $\Rightarrow i$	長音の挿入
終わん	手腕	無 $\Rightarrow u$	長音の挿入
主張	主著	無 $\Rightarrow u$	長音の挿入
取引金	声利き	無 $\Rightarrow u$ 、無 $\Rightarrow$ 撥音	長音と撥音の挿入
開放心	会報誌	無 $\Rightarrow$ 撥音	撥音の挿入 (語尾)
基準	傷	無 $\Rightarrow$ 撥音、 $zu \Rightarrow zyu$	撥音の挿入 (語尾)、母音の誤認識 (直音と拗音)
空間抜いていた	空海の言っていた		
リズムが合う	つじつまが合う		
僕えば新車	ほほえましいじゃ		
百合子	より、こう、		
の cosco だから	残すほうだから		
アイアンできて	歩んできて		
メディアったら	に出会ったら		
日記ファン	に規範		
戻られる	も、おられる		
加速度	下足番		
ビデオ	微力		