

書き起こし・タグ付けマニュアル（日本語）

- ・日本語音声については、「漢字仮名交じり」による書き起こし、「カタカナ」による書き起こしの2種類の書き起こしをおこなう。「漢字仮名交じり」による書き起こしは、検索時の便を考え、原則として一通りの表記によっておこなうものとする。「カタカナ」での書き起こしでは、実際の音声を、カタカナで表記可能な範囲で最大限正確に表記するものとする。
- ・いい間違いや、文字では表現しきれない音声などは、「タグ」を用いて表す。「タグ」は、「漢字仮名交じり」、「カタカナ」両方の書き起こしに対して付与するが、タグの種類によっては片方にしか付与されないこともある。詳細は p.102 以下参照。
- ・漢字仮名交じり表記、カタカナ表記いずれにおいても句読点は使用しない。

（漢字仮名交じり）

- ・漢字仮名混じり表記での文字起こしの際、「朗読」の課題については、元原稿どおりの表記で電子化をおこなう（ただし、句読点・記号類は使用しない）。
- ・「スピーチ」と「ロールプレイ」においては、漢字で書けるものはできるだけ漢字で表すことを原則とする（小磯ほか 2001 p.51）。つまり、漢字・仮名いずれで表記してもいい語については、漢字での表記を優先する。ただし、極端な当て字の類は使用しない。

以下、「スピーチ」と「ロールプレイ」の書き起こしにおける表記の原則を記す¹。

<漢字・仮名の使い分け>

* 熟字訓・当て字は、常用漢字表の「付表」に記載されたもののみを使用する。

* 「こと」「もの」「ところ」等、実質名詞としても形式名詞としても使用される語については、一律仮名表記とする。

* 「お」「ご」と読み分ける接頭辞「御」は、一律仮名表記とする。

* 「言う」は、「発言する」という動作性を明らかにした場合のみ漢字表記とし、「～という人」「ああいったもの」のように、動作性が失われてしまっている場合には仮名表記とする。仮名表記にするのは、具体的には以下のような場合である。

・ 指示副詞（ああ/こう/そう/どう）+いう/いった+体言

◇ 例：そういう人、ああいった家

・ 引用助詞（と/って）+いう/いった+体言

◇ 山下っていう人、学校というところ

* 「行く」「来る」「置く」「貰う」「参る」等、本動詞としても補助動詞としても用い

¹ ここに挙げたものは表記の「原則」であり、この原則によってはカバーしきれない個々の語の表記細則については別途作業用マニュアルを作成している。

られる動詞は、原則として、本動詞の場合は漢字表記（例：「本を置く」「学校に行く」）、動詞テ形のあとに用いられる補助動詞の場合は仮名表記とする（例：「食べておく」「読んでいく」）。

例外：ただし次の動詞に関しては、本動詞として用いられていても仮名表記とする。

ある（×有る、在る）、いる（×居る）、できる（×出来る）、なる（×成る、為る）、（～に）よる（×因る）、くれる（×呉れる）、やる（×遣る）、におう（×匂う、臭う）、かかわる（×関わる、係わる、拘わる）

* 指示詞は仮名表記とする（「此れ」「其れ」とはしない）。

* ロールプレイで「呼びかけ」に使われる人名（例：たかはしさん、あいこちゃん、など）は、すべて架空の人名であり、漢字で表記する必然性がないためすべて仮名表記とする。

<漢字の使い分け>

* 「聞く」と「聴く」、「会う」と「逢う」と「遭う」のように、微妙な意味の違いがあるものの客観的な書き分け規則が立てにくいものについては、より包括的な意味を持つほうの表記（上記の例でいえば「聞く」「会う」）を採用する。

* 「会う」と「合う」、「計る」と「図る」のように、意味的に明確な違いが存在する場合は慣習どおりに書き分ける。

* 同じ語を異なる漢字によって表記でき、かつ意味的にほとんど違いがないと考えられる語については、より画数の少ない方を優先する（例：「輿論」ではなく「世論」、「稀薄」ではなく「希薄」、「充分」でなく「十分」を使用）。

<送り仮名の使い分け>

* 「行う」と「行なう」のように、動詞で送り仮名に複数の可能性が存在する場合は、多く送るほうを採用する。同様に名詞の場合も、「取り組み」と「取組」などのように、送り仮名のつける・つけないに揺れがあれば、原則として、送り仮名を付ける方を採用する。

<数字・アルファベット>

* 原則として数字は漢字表記、アルファベットはカタカナに直して入力する。しかし、漢字仮名以外の文字を使って表記するのが一般的な単語であり、漢字仮名に置き換えてしまえば極めて読みにくくなってしまふ語については、後述の(A)タグを用いて2種類の表記を併記する。

(カタカナ)

・カタカナは、聞こえたままをできるだけそのまま表記するものとする。カナでは表記で

きない場合の処理は【使用タグ】の項を参照。

- 助詞「を」「へ」「は」は、発音どおり<オ><エ><ワ>と表記する。
- 現代仮名遣いで「ぢ」「づ」を用いることになっている語でも、カタカナ表記では一律<ジ><ズ>を使用する（従って「縮む（ちぢむ）」「続く（つづく）」も、カタカナ表記では<チジム><ツズク>とする）。
- 長音は、原則として長音符号<ー>を用いて表記する。従って「オウサマ」「オネエサン」ではなく、<オーサマ><オネーサン>とする。
- 通常長音として発音される音が、現代仮名遣いの影響などを受けて母音連続として発音される場合（例：「先生」は、「センセイ」というのが通常の発音であるが、文字通り「センセイ」と発音されることもある）は、発音どおり母音連続として書く（この場合は<センセイ>と表記）。
- 通常長音として発音される音が短く発音された場合は、発音どおり長音符号は用いずに書く（例：「先生」が「センセイ」ではなく「センセ」と発音されている場合は、そのとおり<センセ>と書く）。
- 算用数字、ローマ字は用いず、すべてカタカナで書く（例：サンジューゴネン、シーディーロム）。

（漢字仮名交じり、カタカナ共通注意事項）

- 一行に、同じ部分の音声に対する「漢字仮名交じり」と「カタカナ」を、「&（半角アンパサンド）」で隔てて書く。
- ひとつの書き起こし単位が長くなり、エディタ上で1行に収まりきらなくなった場合は適宜改行してよいが、その際単語・文節の途中では改行しない。
- 書き起こしに句読点は使用しない。ひとつの書き起こし単位の中にわずかな無音区間がある場合も、そこで読点「、」をうったりはしない。

【使用タグ】

- 音声特徴などをあらわすタグとしては、以下のものを使用する。これらの多くは小磯ほか(2001) p.55に出ているものと共通であるが、本データベースのために独自に定義し直したタグや、新たに設定したタグもある。タグはすべて半角英数字のみを使用する。またタグの中に文字列を書き込む場合、文字列の前には半角スペースを入れる。

1. 文字範囲指定タイプ

このタイプのタグは、「書き起こされた文字範囲に対応する音声について、何らかの現象が起こっている」ということを示すためのものである。

1.1 (W) タグ

概要

言い誤りや発音の怠け、転訛等があった場合、その部分をこのタグで囲み、同時に「言い誤り・発音の怠け等が起こらなかったときに想定される語形」を併せ示す。このタグは、「音声上、標準的と考えられる発音から一時的に逸脱している部分」に対して使用するものであり、文法的な間違い（例：「学校をいきます」）や、事実認定の誤り（例：「日本には大統領がいる」）にはこのタグをつけない。また、本来「発音の怠け」から生じた発音形態であっても、一個人に限らず広く観察される「口語的表現」（例えば「なっちゃった」（本来は「なってしまった」）²）については、現代語としては「言い誤り、発音の怠け」とは考えず、タグをつけない。

タグを付与する表記形

このタグは「カタカナ表記」に対してだけ使用し「漢字仮名交じり表記」では使用しない。「漢字仮名交じり表記」の方では、「本来はこう発音されるべきであったらう」と想定される形だけをタグなしで表記する。

書式： (W ***;###)

***は、発音どおりの表記、###は、「発音の怠け等」が起こらなかったときに想定される語形を示す。

例：

できるだろうと思います & デキル(W ラロー;ダロー)トオモイマス
一人暮らし & (W シトリグラシ;ヒトリグラシ)

タグを付与する範囲

このタグは、「短単位」³を単位として使用する。言い誤り、転訛、発音の怠け等が、隣接する「短単位」で連続して起きている場合は、「短単位」ごとにタグを打ち直す。

例：「郵便貯金」を「ユビンジョキン」と読んでいる場合

○(W ユビン;ユービン)(W ジョキン;チヨキン)

×(W ユビンジョキン;ユービンチヨキン)

² くだけた場面においてあらわれる口語的表現について、「小説やシナリオ等の中のせりふとして、発音どおりの表記で書かれうるもの」については、「一個人の発話においてだけでなく、社会習慣上広範囲にあらわれうる口語的表現」と考え、タグをつけない。「小説やシナリオ等の中のせりふとしても、発音どおりに書かれることは通常ない」表現については、「言い誤り・発音の怠け」と考え、(W)タグをつける。

³ 「短単位」については p.103 以下で詳細に述べる。この段階で「短単位」とは、一般に「単語」として理解されている単位とほぼ同じ、と考えておいて差し支えない。

ただし、二つの短単位にまたがって音の入れ替えが起こっているような場合（例えば「食+文化」を「ショブクンカ」と発音しているような場合）は、複数の短単位をひとつのタグで囲むこともある。

例：

日本の食文化です & ニホンノ (W ショブクンカ; ショクブンカ) デス

1.2 (?) タグ

概要

聞き取り、語彙同定、漢字表記に自信がない場合に使用する。「おそらく〇〇であろうが自信がない」場合、「〇〇の可能性もあるが、××かもしれない」場合、「まったく分からない」場合、すべてこのタグで対応する。

タグを付与する表記形

このタグは、「漢字仮名交じり表記」、「カタカナ表記」の双方に付与しうるが、場合によっては、片方にしか付与されないこともある。「音声は不分明だが、文脈から漢字表記は明らか」という場合はカタカナ表記にのみタグがつき、「音声は明瞭だが、漢字表記には複数の可能性がある（例：「製糸業」か「製紙業」か分からない）」という場合は漢字表記にのみタグがつく。

書式 1： (? ***)

おそらく***といているのであろうが、自信がない

書式 2： (? ***, ###)

おそらく***といているのであろうが、###の可能性もある。

書式 3： (?)

まったく聞き取れない。

例：

(? 競技, 協議)会の中で & キョーギカイノナカデ

タイの(? 民間)と & タイノミンカント

※音声としては「ミンカン」で明らかであるが、文脈から考えると「民間」なのかどうか不明。

異なっているから & (? コトナツ, クトナツ)テイルカラ

※音声としては「コトナツテ」とも「クトナツテ」とも聞こえるが、文脈を考慮すると明らかに「異なって」。

(? 分裂, 分立)する中で & (? ブンレッツ, ブンリツ)スルナカデ

※音声としては「ブンレッツ」とも「ブンリツ」とも聞こえる。このため漢字表記も一通りに決められない。

タグを付与する範囲

基本的には短単位に対し付与するが、発音が不明瞭なため短単位の切れ目も不分明な場合は、よく聞き取れない部分全体に対し使用してよい。

1.3 (D) タグ

概要

ある単語を言いかけて、完全に発音し終わらないうちに発音をやめてしまったような場合、言いかけた部分をこのタグで囲む。言いかけた部分が単語として意味をなさない場合にもこのタグを使用する。

ある単語を別の単語で言い直したり、あるいは同じ単語を二度以上言い直したりしている場合であっても、単語を完全に発音し終わってから言い直しを起こしているのであればこのタグは用いない。

タグを付与する表記形

このタグは、「漢字仮名交じり表記」、「カタカナ表記」の双方につける。

書式： (D ***)

例：

(D な)何	&	(D ナ)ナニ
(D いそ)急でさ	&	(D イソ)キューデサ
(D す)二週間	&	(D ス)ニシューカン

※タグをつけない例：

私私にとっては	&	ワタシワタシニトッテワ
---------	---	-------------

タグを付与する範囲

このタグは、短単位の断片に対して付与する。複数の短単位がまとめて言い直されている場合、言い直されている部分全体に対しタグを付与するのではなく、あくまで不完全に発音された短単位に対しタグを付与する。

例：

○私の(D ち)私の父に	&	ワタシノ(D チ)ワタシノチチニ
×(D 私のち)私の父に	&	(D ワタシノチ)ワタシノチチニ

1.4 (WR) タグ

概要

朗読の際、原稿に書かれている文字・単語をとり違えて読んでいる場合に使用する。スピーチ・ロールプレイにおいては、このタグは用いない。文字の読み違いではなく、一時的な発音の誤り・怠けが起こったと考えられる場合は (W) タグで対応する。

文字・単語のとり違えと考えられる現象であっても、朗読者がその単語を読み終わらないうちに取り違えに気づき、読み直したような場合は (D) タグで対応する。

タグを付与する表記形

このタグは「漢字仮名交じり表記」、「カタカナ表記」の両方につける。ただし、漢字の振り仮名を読み間違っている場合は、カタカナだけにタグがつく。

書式： (WR ***;###)

***は、発音どおりの表記、###は、実際に原稿に書かれていた表記。

例：

楽しく暮らしましたと (WR き;さ) & タノシククラシマシタト (WR キ;サ)

慌て (WR て;) ました & アワテ (WR テ;) マシタ

※原稿にない、余分な「て」を挿入している。

遠く (WR ;に) & トーク (WR ;ニ)

※原稿にある「に」を読み飛ばしている。

継母が & (WR ハハママ;ママハハ) ガ

※「ママハハ」という振り仮名を読み間違っている。

タグを付与する範囲

このタグは、短単位に対し付与する。複数の短単位にまたがって読み間違いが起こっている場合は、短単位ごとにタグを打ち直す。

ただし、原稿に記載されている単位と、実際に発音された単位とで、切り方がずれる場合がある（例：原稿では「気付い+た」（2 単位）と書かれているところを「気+が+付い+た」（4 単位）と読んでいるような場合）。そのような場合は、より広い範囲をカバーする単位の切れ目を優先する。

例：

(WR 気が付い;気付い) た & (WR キガツイ;キズイ) タ

※W タグと WR タグの使い分け

W タグは、「音声上の言い誤り・発音の怠け等」について、WR タグは「文字の読み違い」について用いる。そのように読むことが「誤り」とはいえない事例であっても、読み上げ原稿の表記と異なる読み方がなされていれば WR タグを使用する。「音声上の言い誤り・発音の怠け」が起こっている可能性があると思われる場合は、W タグを優先的に使用する。

例：「手術」を「シジツ」と読んでいる→「発音の怠け」とも、「ゆ」を読み飛ばしたとも考えられるが、こういう場合は「発音の怠け」という解釈を優先し、
W タグ
「通りかかった」を「となりかかった」と読んでいる→なぜ「な」が挿入されたのか、音声的な理由からは説明がつかないので、文字の読み間違いと考え
WR タグ
「私」^{わたし}とルビが振られているのに「アタシ」と読んでいる→文字の読み間違いと考え、WR タグ

W タグと WR タグとが同時に使用される場合もある。

例：「突然」を「とうじえん」と読んでいる場合→「突然」を「当然」と読み違え、なおかつ発音上の誤りが生じているものと考えられるので、

(WR 当然;突然) & (W トウジェン;(WR トウゼン;トツゼン))

とする。

※D タグと WR タグとの使い分け

原稿にない音声切片が挿入されているとき、その音声切片がどの短単位の断片であるか、文脈などを手がかりに明瞭に分かるときには D タグを用いる。どの短単位の断片であるかが分からないときには WR タグを用いる。

例：

泣くのは(D やん)おやめ & ナクノワ(D ヤン)オヤメ

※「やん」の部分は「やめ(る)」の断片であると判断できるのでDタグ。

絶対(WR お;)に忘れないでくれ & ゼツタイ(WR オ;)ニワスレナイデオクレ

※「お」がどの短単位の断片であるか判断できないのでWRタグ。

1.5 (F) タグ

概要

フィラー等をこのタグで囲む。狭義のフィラーだけでなく、応答詞・感情表出系感動詞等にも使用する。どういう音声切片をこのタグで囲むかは、p.106以下の「F タグ付与基準」を参照。

タグを付与する表記形

このタグは、「漢字仮名交じり表記」、「カタカナ表記」の双方につける。

書式： (F ***)

例：

昨日 (F そうですね) 三時ごろかな & キノー (F ソーデスネ) サンジゴロカナ
昨日 (F あの) 経済学の本を & キノー (F アノ) ケーザイガクノホンオ

タグを付与する範囲

フィラー・応答詞・感情表出系感動詞と認定された単位に対し付与する。

1.6 (O) タグ

概要

日本語の発話内に日本語の方言、古語や、日本語以外の言語の単語(地名・人名等も含む)が混じっている場合、その部分をこのタグで囲む。ただし、外来語として日本の中に定着しているような単語についてはこのタグを使用する必要はない。

タグを付与する表記形

このタグは、「漢字仮名交じり表記」、「カタカナ表記」の双方につける。

書式： (O ***)

例：

(O ラームカムヘーン) 大学で & (O ラームカムヘーン) ダイガクデ
気分は (O ケセラセラ) っていう & キブンワ (O ケセラセラ) ッテイウ

1.7 (A) タグ

概要

漢字仮名以外の文字を、漢字仮名と併記するために使用する。英数字は全角文字で表記する。

タグを付与する表記形

このタグは、「漢字仮名交じり表記」にのみつける。

書式： (A ***;###)

***は漢字仮名、###は漢字仮名以外の文字種

例：

(A 千九百七十三;1 9 7 3)年 & センキューハクナナジューサンネン
(A シーディーロム;CD-ROM) & シーディーロム

1.8 (K) タグ

概要

ひとつの漢字で表示される語の途中にフィラーが挿入される、といったような理由で、漢字表記ができなくなってしまったときに使用する。

タグを付与する表記形

このタグは、「漢字仮名交じり表記」にのみつける。

例：

(K む<FV>すめ;娘)が いれば & ム<FV>スメガイレバ

1.9 (LG) タグ

概要

笑いながらの発話に対し使用する。笑いながら話している部分を括弧内に入れる（小磯ほか 2001 では(笑)を用いているが、当データベースでは日本語以外の言語も使用するのでタグに漢字は使用しない）。

話し終わってから笑いが生じている場合は、後述 2. 「音声・事象表現タイプ」の<LG>タグを使用する。

タグを付与する表記形

このタグは、「カタカナ表記」にのみつける。