

[IMADE 会話転記基準]

Ver.0 : 070218 ○○

Ver.0.1 : 070301 ○○

3. 転記基準

3.1. 基本フォーマット

- ・日本語話し言葉コーパス（以下 CSJ）の転記フォーマットとする.
- ・開始/終了時刻は今回の各作業ファイルの冒頭を 0 分 0 秒とし, 10 進法で表記する.
- ・基本情報部 : 転記基本 ID 開始時刻-終了時刻 話者 ID:
※「話者 ID」は 1, 2, 3 ではなく, それぞれ A, B, C にしてください.
- ・発話部 : 基本形 & 発音形
- ・発話部では, 文節ごとに転記行を改める.

例)

0223 00578.104-00580.441 A:

しかし	&	シカシ
コンシステントに	&	コンシステントニ
差が	&	サガ
出てるという	&	デテルトイウ
ことであります	&	コトデアリマス

3.2. 転記基本単位の認定

- ・下記の「非言語発声」の場合を除き, 200ms 以上の休止で挟まれた区間を 1 つの「転記基本単位」(間休止単位)とする.
- ・200ms 以上の休止が語中にある場合には単位を分けるのではなく, その箇所に P タグ (3.4.1) を挿入する.

3.3. 非言語発声についてのタグ

- ・笑いや咳などの「非言語発声」は独立した転記基本単位とし、200 ミリ秒以上の無音をはさまずに音声区間と連続していても、単位境界を設ける。
- ・「非言語発声」としては以下ののみをタグで表記する。当該話者による息や鼻すすり、当該話者の音声でないもの（他の話者の発話の回りこみ、雑音など）は「非言語発声」には含まない。

●笑い： {笑}

●咳・咳払い： {咳}

- ・非言語発声は言語的発声とは独立の単位になるので、言語的発声と同様に、「基本情報部」の情報を明記する。

例)

0223 00578. 104-00580. 441 A:

{笑}

- ・笑いなど非言語発声を伴いながらの発話などは通常の音声区間として扱い、3.5 のタグを付与する。
- ・発話の前後に非言語発声のみの区間が存在する場合は、上記にしたがって、その部分のみ独立した転記基本単位とする。

〈笑いのみ〉 〈笑い+発話〉〈発話のみ〉

|<----->|<----->| 転記基本単位

〈発話のみ〉〈発話+笑い〉 〈笑いのみ〉

|<----->|<----->| 転記基本単位

- ・「笑いの最中に咳き込む」場合などには、{笑}を優先し、{咳}はマークしなくてよい。

3.4. 言語的発話についてのタグ

3.4.1. 基本形に付与されるタグ

・発話部中の「基本形」に部分においては、以下の現象をタグ（括弧や記号は半角）でマークする。

●フィラーF (F XXX)

- ・CSJ 基準に準拠する。
- ・ただし、「応答表現」であることが明らかなものについては、F タグは不要であるが、同じような音調の応答表現については、可能な限り表記を統一させる。
- ・似たような音で応答表現かフィラーかが区別しにくい箇所については、個々の例ごとにより妥当そうな方にしておく。

例)

応答表現	フィラー
ああ	(F あー)
ええ	(F えー)
うん	(F うーん)

●言いさしD (D XXX)

- ・CSJ 基準に準拠する。
- ・非言語発声（3.3）によって語が途切れた場合にも、語断片をD タグでマークする。
- ・言いさしによって生じた語未満の断片に対処するためのものなので、D タグを付与する単位は語未満である。

例)

- × (D ケンキュウシ)
- 研究(D シ)

●固有名詞N (N XXX)

・会話参加者や話題となっている人物（いわゆる「有名人」は除く）の氏名やあだ名など、その情報から会話参加者が特定される可能性がある固有名詞等はNタグで囲む。

例) (N 河原)先生

- ・表記は「ジャーゴンリスト」を参照し、分からなければ、当て字やカタカナでよい。
- ・Nに該当する可能性がある要素については、(? XXX)よりもNタグを優先する。

●聞き取り困難? (? XXX)

- ・XXXを発話していると思われるが、聞き取りに自信のない箇所をマークする。
- ・XXXはカタカナとする。
- ・基本形に(? XXX)を用いる場合には、発音形の該当部分もXXXと表記する。
- ・断片的に認定できる語がある場合にはなるべくその語だけでも書き起こすことを原則とするが、大部分が言語的に理解できないような場合には、まとめて(?)だけでよい。

●聞き取り不能(?) (?)

- ・発話していることは分かるが、全く聞き取れないものをマークする。
- ・基本形が(?)の場合には、発音形の該当部分も(?)でよい。

●語中の休止P (P XXX)

- ・語中の促音および破裂音の閉鎖区間などによって生じる「語中の休止」では転記基本単位を切らないため(3.2)、代わりに当該の位置に休止時間XXX(200以上)をPタグで記入する。
- ・当該の「語」がすでにFやD, W, ?などのタグで囲まれている場合には、Pタグの付与は省略する。

※CSJにおけるAタグは本作業では使用しない。

- ・一般的な略称や短縮形などが定着しているものについては、実際の発音は発音形を見れば分かるため、基本形の表記においては、可読性を考慮し、略称を使用する。

例)

N I C T & ニクト

N I C T & エヌアイシーティー

3.4.2. 基本形と発音形に付与されるタグ

●非標準的発音 W (W XXX|YYY)

・意図された語は特定できるものの、発音に忠実な表記では意味が通じなくなるような非標準的な発音のうち、語中で生じた非語彙的な長音と促音 (3.6) 以外のものについては、W タグでマークする。・口語表現や方言などはある程度定着し、また規則性もあるため、普通書き起こすことができる限り、「非標準的発音」には含まれない。

・XXX を聞き取った通りの音 (カタカナ)、YYY を意図されていた発話の標準的な表記とする。

・3.6 の(H)や(Q)と同様、W も発音についてのタグだが、表記と発音の対応付けの問題に対処するためのものなので、基本形の方にのみタグを付与し、発音形は発音のまま書く。

例) (W シューシャ|水車)小屋の & シューシャゴヤノ

・語の一部が聞こえない場合も W タグを使用して音の列と単語を併記する。

例) 「だから」の「ら」が聞こえない: (W ダカ|だから) & ダカ

例) 「日刊スポーツ」の「にっ」が聞こえない: (W カン|日刊)スポーツ & カンスポーツ

・F や D, N, ?タグでマークされる箇所については、これらのタグを優先し、重複して W タグで囲む必要はない。

・表記が標準的でなくなる原因がタグ (特に P タグ) の挿入によるものである場合には、タグを削除すれば通常の表記になるため、重複して W タグで囲む必要はない。

例) びっ(P 210)くりした ← (W びっ(P 210)くりした|びっくしりた)とする必要はない。

3.4.3. 発音形に付与されるタグ

●語中で生じた非語彙的な長音 H (H)

●語中で生じた非語彙的な促音 Q (Q)

- ・ CSJ と同じ基準で、基本形の欄ではなく発音形の欄にのみ付与する。
- ・ CSJ とは異なり、〈H〉、〈Q〉ではなく、(H)、(Q)とする（機械処理の都合上〈〉は避ける）。
例)
左下 & ヒダリ(H)シタ
左側に沿って & ヒ(Q)ダリガワ
- ・ 文（節）末の長音も同様にマークする。
例)
それは & ソレワ(H)
確かにね & タシカニネ(H)
大丈夫かな & ダイジョウブカナ(H)
- ・ 関西弁の発話における非語彙的な長音も同様にマークする。
例)
目痛い & メ(H)イタイ
手見せて & テ(H)ミセテ
- ・ 引き伸ばしの長さに応じて H, Q の数を変える必要はない。
- ・ すでに基本形に F や D, N, ? などのタグが付与されている場合には、発音形における H, Q の付与は省略する。
- ・ 語中の休止と H, Q タグ付与が同時に必要な箇所については、基本形の方に P タグ、発音形の方に H, Q タグを使用する。
例) す(P 235)ごい & ス(Q)ゴイ

3.5. 言語的発声と非言語発声などの融合

- ・ 言語的発声が非言語発声などと融合している区間のうち、以下のものについてはタグやコメントでマークする。

●話者が言語的内容を笑いながら発話 L (L XXX L)

- ・ 基本形の転記は通常通りとした上で、該当区間を下記のタグで囲む（他のタグとは異な

り、閉じ括弧にも記号がつくことに注意)。

・語ごとに別のタグで囲む必要はなく、転記基本単位内であれば転記行をまたいでタグを付与してよい。

例)

0223 00578.104-00580.441 A:

しかし & シカシ

コンシステントに & コンシステントニ

(L 差が & サガ

出てるという & デテルトイウ

ことで L)あります & コトデアリマス

・漢字かな混じりの基本形の表記の可読性を優先し、通常の変換を変えなくてもよいよう、範囲は適宜「広めに」マークする。

・言語的内容とは時間的に独立の笑いについては 3.3 の {笑い} 参照。

●当該話者の発話に他の話者の発話や笑いなどが重複している場合 0 #0

・対象としている話者の発話内容を聞き取るのに邪魔になるほど大きな音量の重複がある場合のみ、当該転記行にコメントをつける。

例)

0223 00578.104-00580.441 A:

しかし & シカシ

コンシステントに & コンシステントニ

差が#0 & サガ

出てるという#0 & デテルトイウ

ことであります#0 & コトデアリマス

・重複している音が他者の言語音や笑い、咳などではなく、非発声的な雑音の場合には、タグの付与は不要。

※（特に会話 B や C では）大半の転記行に「#0」が付くという可能性もあるため、B や C の作業が始まってから再度変更することも検討する（おそらくその場合には 0 に関連する作業は何も行わないことになる）。

3.6. タグの入れ子やクロス

- ・ 3.3 のタグは 3.4 や 3.5 のタグと同じ転記基本単位に生じることがないため、入れ子やクロスは生じない。
- ・ 3.4 の各タグが互いに入れ子やクロスになることは原則としてない（問題があれば要報告）。
- ・ 3.4 のタグと 3.5 のタグの入れ子やクロス、3.5 の L タグと 0 タグの入れ子やクロスは許容される。