

CHAT

The CHAT Transcription Format オンラインマニュアル

基本単位

■ utterance

- ・イントネーションを考慮に入れる
- ・“ c-unit ” or conversational unit
 - ・ "T-unit" と事実上同じ
 - ・主節のみの発話
 - ・主節と従属節の発話
 - ・等位接続（ and, but, or, so ）でつながれた節は、別の発話に区切る
 - ・一発話に主節はひとつのみ
- ・「 main line 」：一発話一行
 - ・発話の終端記号は「 . 」 「 ? 」 「 ! 」 のいずれか
- ・行頭
 - ・大文字でも小文字でもよいが、原則、小文字
 - ・行頭の単語の繰り返しとか言い直しとかあるので
- ・：主語と動詞を含むをつなぐ場合には別発話
 - ・名詞や名詞句ををつなぐ場合は、別の発話にしてはいけない。
 - ・

Pauses：無音のポーズ

- ・長さによって3段階
 - ・ (.)
 - ・ (..)
 - ・ (...)

■

- ・問題点：3段階に分けるのは難しい
- ・修正案：(..) と (...) の二種類。短めと長め。
 - ・いち・に、と数えられるくらいなら (..)
 - ・いち・に・さん、くらい以上なら (...)

推測

- ・直前の語 [?]

*SAR: I want a frog [?]

- ・一連の語の場合 < > で囲む

*SAR: <going away with my mommy> [?] ?

■

- ・問題点：直前の一語か連語かで、<> で囲むかどうか分けるのは煩雑
- ・修正案：語数にかかわらずその範囲を <> で囲む

繰り返し Repetition [/]

■ 「繰り返し」と「言い換え」の違いは微妙

- ・特に、部分的に重複するところがある場合

<I> [/] <I> [/] <I go to> [/] <I> [/] <I go> [/] I go to school

- ・部分になっていたら部分的な繰り返し

■ 取り扱い

- ・語彙リストを作成する際に、繰り返した語を token として数えない。

■ 例

- ・直前の語 [/]

*CHI: I practice dance to from [/] from three thirty to five ten .

- ・一連の語の場合 <> で囲む

*BET: <I wanted> [/] I wanted to invite Margie.

■

- ・問題点：直前の一語か連語かで、<> で囲むかどうか分けるのは煩雑
- ・修正案：語数にかかわらずその範囲を <> で囲む
 - ・CLAN は問題なく動く。

- ・直前の語 [/]

*CHI: I practice dance to <from> [/] from three thirty to five ten .

■ 複数回の繰り返し

- ・いろんな表記法があるが、、、
- ・繰り返しが複数回ある場合は、そのまま表記する。

<milk> [/] <milk> [/] <milk> [/] milk

言い直し

■ 言い直し Retracing [/]

- ・同じことを言い直している

*CHI: <the fish is> [//] the [/] the fish are swimming.

- ・ <the fish is> と言い出したが、その部分を言い直している [//]
- ・ the が繰り返されている [/]
- ・ 発話としては the fish are swimming が最終的に産出されたと見なす。

■ 言い換え Reformulation [//]

- ・ (意味はほぼ同じで) 全く別の表現に言い換えている場合

*BET: <all of my friends had> [///] &-uh we all decided to go home for lunch.

■ 言い直し False Start Without Retracing [/-]

*BET: <I wanted> [/-] uh when is Margie coming?

■

- ・ 問題点: 言い直しか言い換えか、判断に迷う。
- ・ 修正案: いずれも、該当範囲を <> で囲み、[//] を使う。
 - ・ 言い直しか言い換えかの区別はしない (保留)

わからない単語 xxx

*SAR: I want xxx .

フィラー &- フィラー

- ・ フィラーの語そのもの (例: uh) の前に「&-」を付ける

*BET: <I wanted> [//] &-uh I thought I wanted to invite Margie.

「I wanted」と言い出して、そこでやめて [//]、「アー」(&-uh) と言った後で、「I thought ...」と言い換える。

- ・ 例

&-ah, &-eh, &-er, &-ew, &-hm, &-mm, &-uh, &-uhm, &-um

■ 日本語と英語と不明な場合と

- ・ 日本語の場合は、フィラーの種類にかかわらず、&-jp としておく
- ・ 英語の場合は、英語の単語と認識できる場合は、単語を書く。
 - ・ 複数の単語の場合は、アンダーバーでつなぐ

&-um
&-well
&-let_me_see
&-you_know

- ・ その他不明な場合は、 &-?

発話の一部に直接引用がある場合は引用符で囲む。

- ・ 引用符の始まりは begin double-quote (“ , Unicode 201C)
- ・ 終わりは end double-quote (” , Unicode 201D)

直接引用が複数の発話の場合、の後に、引用の始まりに をつける。

- ・ 引用自体が、複数の発話になる場合 ↑
 - ・ 発話自体は、複数の発話に分けて書く。
 - ・ 始まり部分行頭に +"

複文の節の区切り [^c]

- ・ 複文中の節を、二つの「発話」として処理したほうが良い場合
- ・ 節の区切りに を入れておく。

分析対象から除きたい部分 [e]

- ・ 調べたい発話部分以外のところを処理から除く。
- ・ とで囲んだ後にを書いておく。

*BET: <I think that maybe> [e] the cat is up the tree.

数字はアルファベットで単語としてつづる

日本語の単語・表現の表記

■ 日本語データの場合

When transcribing Japanese data in Latin script (Romaji) as well as in Japanese script (Kana Kanji), you should add spaces to identify words. The WAKACHI02 system can be downloaded as a part of the complete JPN grammar from <https://talkbank.org/morgrams>. This web page summarizes the rules for word

separation (Wakachigaki). It is crucial to follow these rules in order to get correct results from MOR (automatical morphological analysis) or DSS (Developmental Sentence Score).

■ 英語データの中での日本語の単語

- ・ 発話内容に関係ない「フィラー」の場合（例：「なんだっけ」、「わかんないなあ」）
 - ・ &-jp（日本語のフィラー扱い）
- ・ 発話内容に関係ある場合
 - ・ 日本語部分を [] で囲む
 - ・ 名詞の例：I like [海苔巻き].
 - ・ 形容詞の例：I like [青い] apple.
 - ・ 副詞の例：It's [メチャ] beautiful.
 - ・ 動詞の例：the dog [生んだ] puppy.

- ・ 日本語を英語で言い換えている場合：言い換え表記を追加する
- ・ 言い換え部分を <> で囲み [/] を付与

it speed story <[進む]> [/] going.

本文中にコメントを残す： [% コメント書いておく]

- ・ メモを書きおきたいことがあります。
- ・ スクエアブラケットと % を使う。

誤用分析に関連する表記

- 誤用の直後に を書く。

- 語の欠損：Omitted Word Oword

- ・ 具体的な語がない場合、例えば不定詞の to

*EVE: I want Oto go.

- ・ 冠詞の欠損 Odet
- ・ 動詞の欠損 Ov
- ・ 助動詞の欠損 Oau

“ he Oau running ”

- ・ 主語の欠損 Osubj

- 語の一部の欠損：() で補う

*RAL: I been sit(ting) all day .

minCHAT

- ここまでしないと、CHECK に通らない

- ・ @Participants: に三つの要素が必要
 - ・ 3 文字記号
 - ・ 「名前」
 - ・ 役割
- ・ @ID: に 1 0 のフィールドが必要（縦棒 1 0 本）
 - ・ 言語
 - ・ コーパス名
 - ・ 3 文字記号
 - ・ 役割
- ・ CLAN/lib/depfile.cut に定義が書いてある
 - ・ 必要であればそこを修正

W Data 転記規則