

CLAN

NICER

CLAN で NICER を処理する

CHAT に基づく NICER のフォーマット

■ 共通点

- ・ファイルは @Begin で始まり @End で終わる。
- ・「ヘッダー情報」は、ファイルの先頭部分で、行頭に @ 項目名で示す
- ・「本文部分」は、行頭にアスタリスク+話者記号で、一発話一行で記述
- ・各発話の関連情報は、発話の下（次の発話との間）に、行頭に % 項目名で示す
- ・行の先頭の項目名と内容との間は、コロンとタブで区切る

■ CHAT フォーマット

minCHAT という最低限の枠

■ NICER のフォーマット

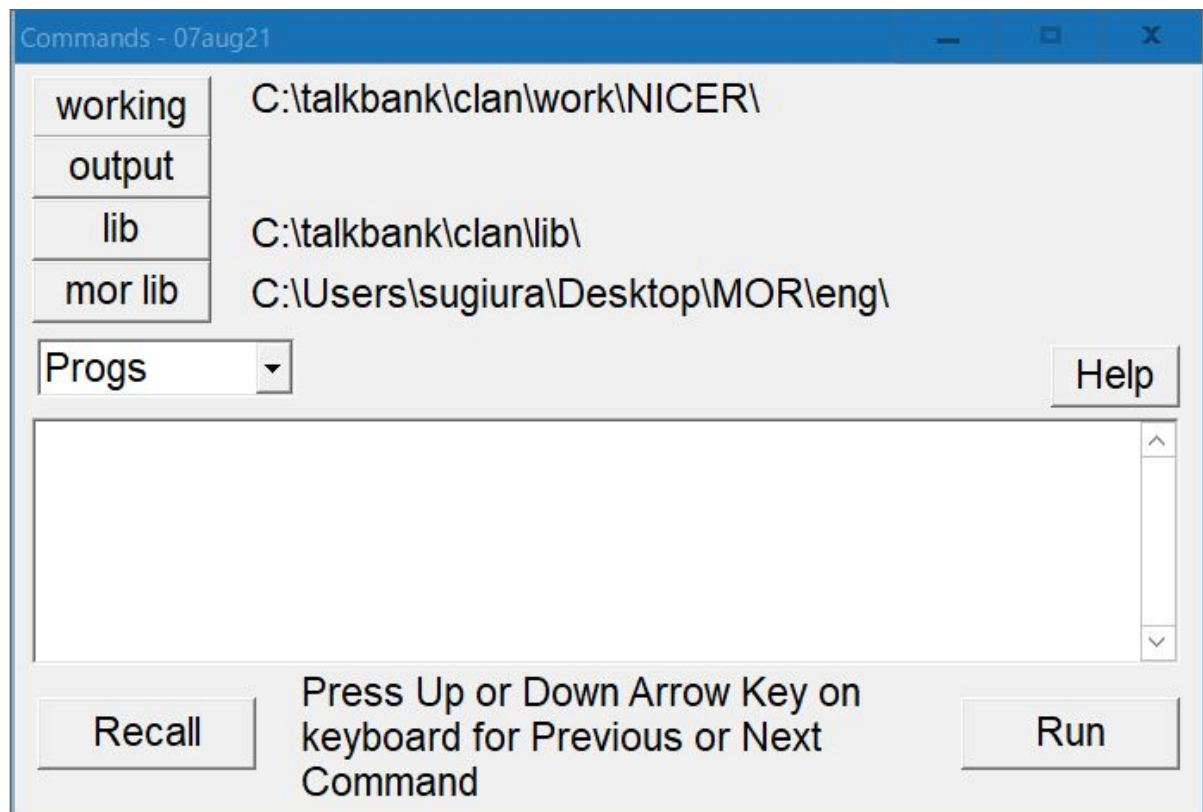
- ・ヘッダー情報の注意点
 - ・ @Qualification で英語力を示す試験のスコア情報
 - ・ @Abroad で海外滞在経験
 - ・ @Topic でエッセイのトピック
 - ・ @Criterion で、そのエッセイを Criterion にかけた時の評価スコア（6 点満点）
- ・本文部分
 - ・ *JPN501 は話者記号
 - ・ %NTV は英語母語の添削者による訂正（書き換え）文
 - ・ %COM は添削時のコメント
- ・この 3 行がセットとなって、発話の数だけ含まれる
- ・ 実例 : JPN501.txt

CLAN の準備

■ インストール

■ working ディレクトリ の設定

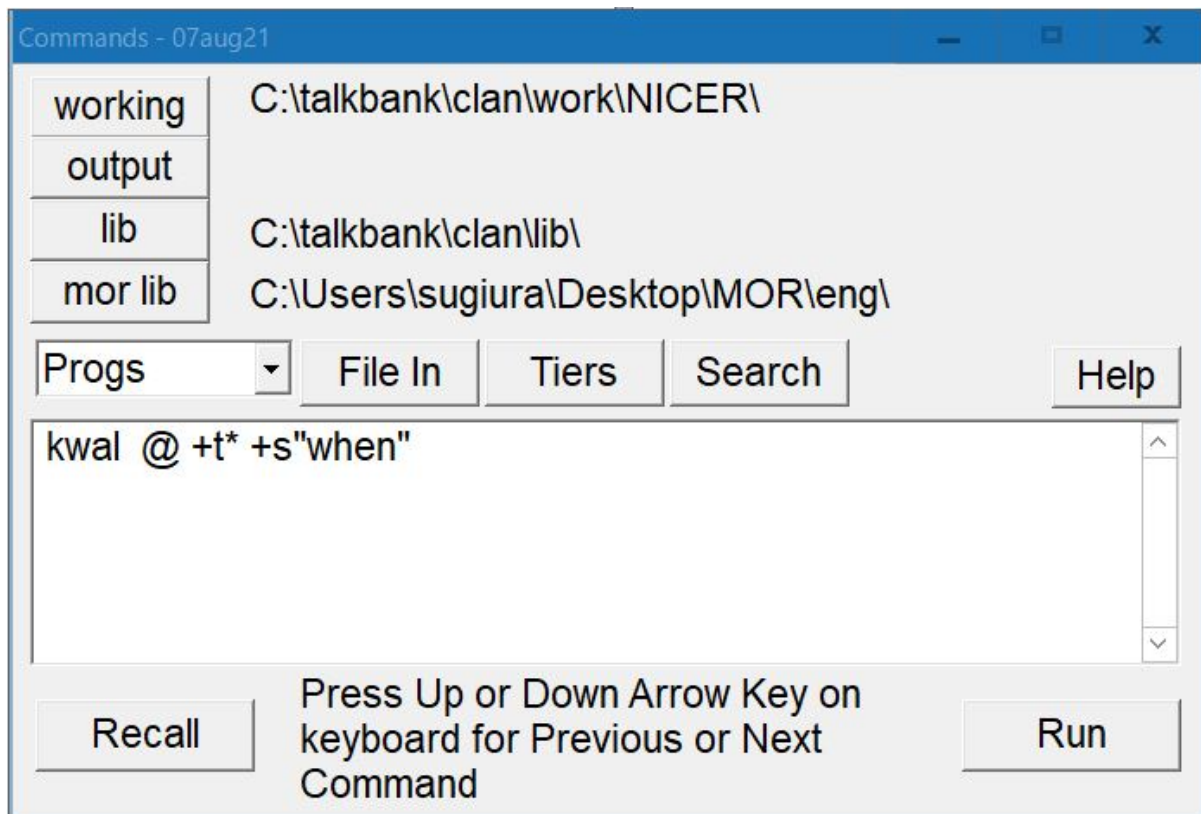
- ・ CLAN 起動後、コマンド・ウィンドウ内の一番上「working」で設定
 - ・ 分析対象とするファイルを入れたフォルダーを指定
 - ・ あらかじめ、分析対象とするファイルを選んでおくのがコツ



CLAN の実行

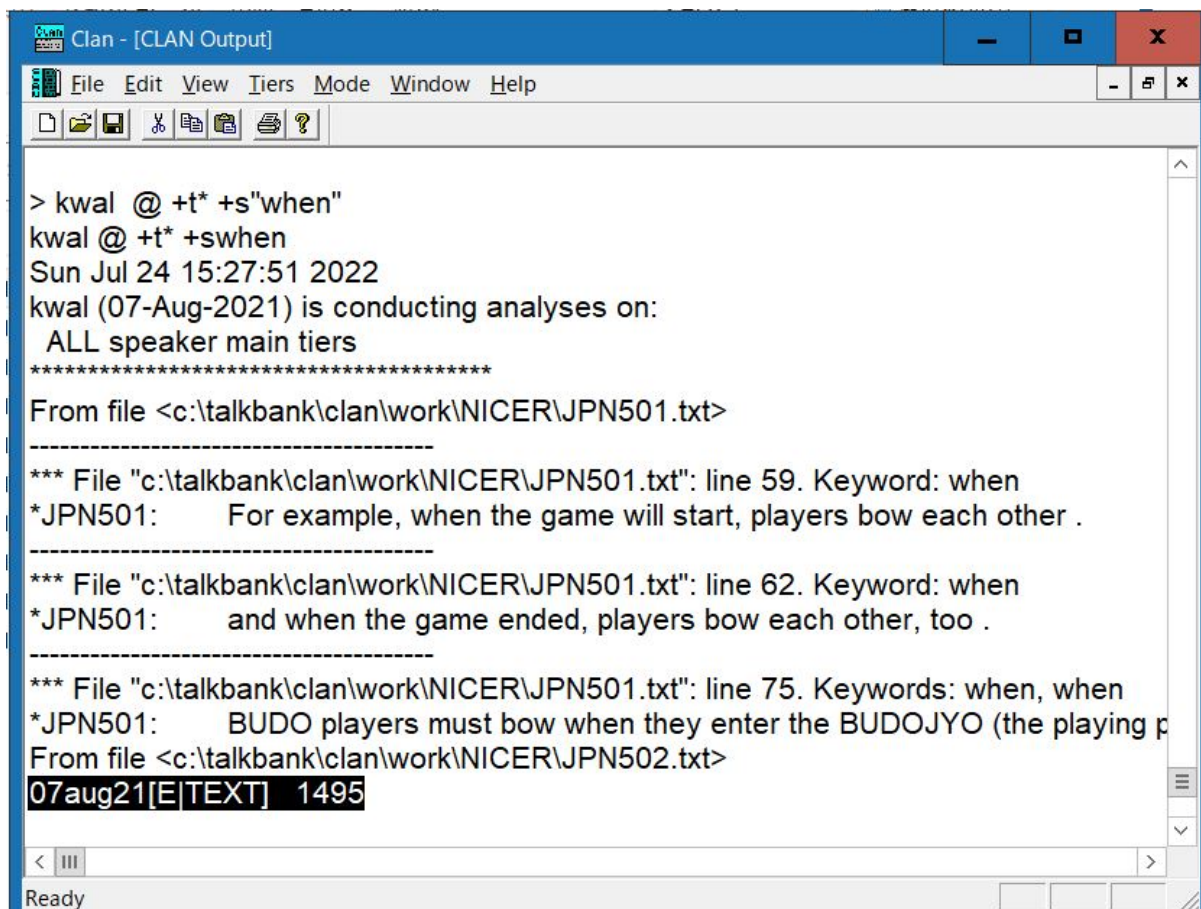
■コマンド・ウィンドウ

- Progs
 - コマンドを選ぶ（例：検索 kwal）
 - コマンドボックス内に選んだコマンドが表示される
 - 選んだコマンドに応じて、ボタンが表示される（例：「File In」「Tiers」「Search」）
- File In
 - 分析対象とするファイルを選ぶ
 - その際に、「Filter」で、「All files (*.*)」にして、拡張子 .txt も読めるように。
 - 「Add All」でフォルダー内ファイル全部
 - 「Done」で設定完了
 - コマンドボックス内に「@」が表示される
- Tiers
 - 対象とするティアを選ぶ。話者行は「*speaker tier」
 - 具体的な話者記号を指定する場合は、右枠内に記入（例：*CHI）
 - コマンドボックス内に「+t*」と表示
- Search
 - 対象とする文字列 word を指定（例：when と入力）
 - コマンドボックス内に「+s"when"」と表示
- 右下の「Run」で実行



- ・カーソルキーの上下で、コマンドの履歴一覧を表示・選択できる

■実行結果



分析の仕方

■分析コード

- ・ CLAN を使って、大量のデータに対して分析コードをつける方法は以下を参照

<https://sugiura-ken.org/wiki/wiki.cgi/exp?page=CLAN#p8>

https://talkbank.org/manuals/CHAT.html#_Toc107417330

- ・ CLAN を使ってコードを付けることもできるが、
- ・ CLAN を使わずに、
- ・ ポイントは以下のようなコードがデータについていればよい
- ・ コーディングのルールに従ってコードをつける
 - ・ \$ で始める
 - ・ 階層関係は : で区切る（スペースなし）
 - ・ 複数のコードを一つのティアに書く場合は、間に半角スペースを置く（単語を並べるのと同じ）
 - ・ コードを書くティアは %COD 以外でも、自由にティア名を付けて作ってよい

%COD: \$N:Y \$V:N

■分析コードの頻度算出

- ・ CLAN の FREQ コマンドで、対象のティアを分析コードティアにして、頻度を出す。
- ・ コードの頻度一覧が出る。

分析の例：文のタイプと文型の分析

■コード

- ・ 三つの分析ティア
 - ・ %SNT 文のタイプ
 - ・ DE 平叙, IN 疑問, IW WH 疑問, IM 命令, EX 感嘆
 - ・ S 単文, cSS 従+主, ScS 主+従, StS 主+ that 節
 - ・ %CNS 文型
 - ・ s0 フラグメント, s1, s2, s3, s4, s5, s6, s7
 - ・ A 肯定, N 否定
 - ・ there 構文, PASS 受身
 - ・ %REC 関係節
 - ・ who, which, what, where, when, that
 - ・ nom 主格, obj 目的格, app 同格の that

■実際のコーディング例（部分）

■ %SNT 文のタイプの頻度分析

```
File Edit View Tiers Mode Window Help
[Icons: New, Open, Save, Cut, Copy, Paste, Print, Help]

> freq @ +t%snt
freq @ +t%snt
Sun Jul 24 22:52:57 2022
freq (28-Jul-2020) is conducting analyses on:
  ONLY dependent tiers matching: %SNT;
*****
From file <c:\TalkBank\CLAN\work\NICER\JPN501-cod3.txt>
Speaker: *JPN501:
12 $DE:S
3 $DE:cSS
1 $IN:S
1 $IW:S
-----
4 Total number of different item types used
17 Total number of items (tokens)
0.235 Type/Token ratio
This TTR number was not calculated on the basis of %mor line forms.
If you want a TTR based on lemmas, run FREQ on the %mor line
with option: +sm;*,o%
```

■ %CNS 文型の頻度分析


```

> freq @ +t%cns
freq @ +t%cns
Sun Jul 24 23:07:55 2022
freq (28-Jul-2020) is conducting analyses on:
  ONLY dependent tiers matching: %CNS;
*****

From file <c:\TalkBank\CLAN\work\NICER\JPN501-cod3.txt>
Speaker: *JPN501:
  2 $s1:A
  3 $s1:A:there
  5 $s2:A
 10 $s3:A
  1 $s3:A:PASS
  1 $s3:N
  1 $s5:A:PASS
-----
  7 Total number of different item types used
 23 Total number of items (tokens)

```

■ %REC 関係節の頻度分析

```

> freq @ +t%rec
freq @ +t%rec
Sun Jul 24 23:13:27 2022
freq (28-Jul-2020) is conducting analyses on:
  ONLY dependent tiers matching: %REC;
*****

From file <c:\TalkBank\CLAN\work\NICER\JPN501-cod3.txt>
Speaker: *JPN501:
  1 $that:app
  1 $that:obj
  1 $where
  1 $who:nom
-----
  4 Total number of different item types used
  4 Total number of items (tokens)

```