

The logo for TCCM (Taiwan Corpus of Child Mandarin) features the letters 'TCCM' in a bold, white, sans-serif font. The letters are set against a dark green rectangular background that has a subtle gradient and a few small white stars scattered around it.

Taiwan Corpus of Child Mandarin

● 台灣 ● 兒童 ● 語言 ● 語料庫

A stylized, cursive letter 'N' in a golden-yellow color, positioned on a light blue, semi-transparent rectangular card. The card is part of a sequence of three cards that appear to be flying or falling from the top right towards the center of the slide.A stylized, cursive letter 'T' in a golden-yellow color, positioned on a light blue, semi-transparent rectangular card. This card is the middle one in a sequence of three cards flying from the top right.A stylized, cursive letter 'U' in a golden-yellow color, positioned on a light blue, semi-transparent rectangular card. This is the final card in a sequence of three cards flying from the top right.

檔案格式介紹

.chat script 主要檔案格式

3 個主要成分

- ▶ 檔頭 **file headers** @
- ▶ 主層 **the main tier** *
- ▶ 附屬層 **the dependent tiers** %

```
@Begin
@Languages:      zho
@Participants:   CHI 小綜 Child, MOT 雯雯 Mother
@ID:             zho|LabName|CHI|2;5.|male|||Target_Child||
@ID:             zho|Labname|MOT|32;0.|female|||Mother|16|
*MOT:            我們要按照順序數預備備。
%mor:           Npro:1p|我們Vt|要Vt|按照Nn|順序Vt|數Vt|預備備。
*CHI:           一二三四五六七八。
%mor:           IDM|一二三四五六七八。
@End
```

檔頭 Headers

☞ 檔頭格式類別：

- Bare headers
 - @Begin
- Headers with entry
 - @Birth of CHI: 21-MAR-2005

☞ 檔頭功能類別：

- 必要檔頭 Obligatory headers
- 固定檔頭 Constant headers
- 可變檔頭 Changeable headers

必要檔頭 Obligatory Headers (1)

- 此類檔頭為必要資訊，若缺少任何其一，CLAN程式的指令將無法正確執行
- 此類檔頭需遵照固定排序方式

@Begin

@Languages: zho

@Participants: CHI 小綜 Child, MOT 雯雯 Mother

@ID: zho|LabName|CHI|2;5.|male|||Target_Child||

(其他檔頭、主層、附屬層)

@End

必要檔頭 Obligatory Headers (2)

☞ @Begin

- 標記檔案開端

☞ @Languages: zho

- 標記該語料檔案中主要使用的語言; zho for Mandarin

☞ @Participants: XXX Name Role, XXX Name Role, ...

- @Participants: CHI 小綜 Child, MOT 雯雯 Mother

name (omittable)

role

必要檔頭 Obligatory Headers (3)

- ☛ @ID: zho|LabName|CHI|2;5.|male|||Target_Child||
 - 標記對話參與者相關資訊
 - Language|corpus|code|age|sex|group|SES|role|education
 - 可供clan程式根據某一資訊選取語料檔
- ☛ @End
 - 標記檔案結尾

固定檔頭 Constant Headers

- ☞ 非必要性檔頭
- ☞ 位於檔案起始處，所紀錄的資訊在同一語料檔案中維持固定，如參與者年齡、性別等等
 - @Age of CHI: 6;4
 - @Sex of CHI: female

可變檔頭 Changeable Headers (1)

- ☞ 可出現在檔案起始處或正文中
- ☞ 紀錄可變動的資訊
 - @Situation: reading the book "好忙的蜘蛛"
 - General setting of the interaction.
 - Applies to all the material that follows it until a new @Situation header appears.
 - @Room Layout: two chairs and a table
 - Important for experimental setting



主層 The Main Tier (1)

標記互動中話語內容

格式為 *說話者代號: 內容

- *OBY: 吃荷包蛋.
- *CHI: 我最歡吃蛋.

主層 The Main Tier (2)

- ☞ 每一個主層僅標記一個語句 (utterance)
- ☞ 無法辨識內容之標記方式:

■ **XXX**

- 完全無法辨識之語句，計算平均語句長度(MLU)時不會納入
- *CHI: XXX.

■ **XX**

- 無法辨識之詞彙
- *CHI: 我想要 XX.

主層 The Main Tier (3)

■ WWW

- 轉寫者無法或不打算轉寫的內容，通常會伴隨附屬層的註記
- *MOT: www.
%exp: talking on the phone

■ 0 (zero)

- 僅有動作沒有語句
- *CHI: 0 [= cries] .

■ &

- 不完整的音節片段
- *CHI: &t, &k.



附屬層 The Dependent Tiers (1)

緊接在每一主層之下，提供研究者可能感興趣的資訊，如詞類標記、句法結構等等。

■ 格式：**%資訊類別:** **內容**

● *OBY: 吃荷包蛋.

● %mor: Vt|吃 Nn|荷包蛋.

附屬層 The Dependent Tiers (2)

☛ %mor

- Morphosyntax Tier

- 標記詞類

- *CHI: 有什麼小火車?

- %mor: Vt|有 WH|什麼 ADJ|小 Nn|火車?