

CLAN 临床研究应用概要

林枫¹、陈珍珍¹

1. 南京医科大学康复医学院

2018 年 01 月 08 日更新

目录

简介：为什么要使用 CLAN	3
1、什么是 CLAN 和 CHAT	3
2、下载和安装 CLAN	3
3、安装配套软件	5
4、确认 CLAN 和电脑的设置	5
5、获取和使用失语库数据（初学者可略过）	6
6、下载与查看 CHAT 和 CLAN 手册	7
7、文件夹管理	8
8、创建新的译文文件	8
9、译文转写过程	11
9.1 步控窗口设置	11
9.1 译文示例：	14
9.2 F5 转写模式	15
9.3 译文与媒体进行关联	15
9.4 手工编辑打点符（Mode>Expand bullets）	16
10、文件校验	16
10.1 逐条显示差错：Esc-L 或 Mode>Check Opened Files	17
10.2 一次显示所有差错：check 指令	17
10.3 MOR 校验：mor +xb filename.cha	18
10.4 KIDEVAL 和 EVAL 分析	20
11、CLAN 统计分析概览	21
11.1 → dir	23
11.2 → mlu +t*CHI sample.cha 或 mlu -t%mor sample.cha	23
11.3 → 通配符*（以 freq 指令为例）	25
11.4 → 定向覆盖符> 和 定向添加符>>（以 freq 指令为例）	25
11.5 → 重复（repetition）和回述（retracing）参数：+r6	25
11.6 → 关键词指令：KWAL	25

本文内容主要参考 CLAN 的英文版 SLP 用户指南（<https://talkbank.org/manuals/Clin-CLAN.pdf>），结合本课题组使用该系统的经验编译。欢迎探讨，敬请指教！

Email: 林枫 peterduuslin@gmail.com; peterduus@njmu.edu.cn; 陈珍珍 doc_zzchen0502@163.com

11.7 合并字符检索指令: COMBO.....	26
11.8 块检索指令和镶嵌符: gem 和 gemfreq	27
常见问题	30
下载的媒体文件无法与.chat 文件关联怎么办?	30
如何鉴别话语单位?	30
话语重叠怎么办? → [>] [<] +<.....	30
如何处理重复语? → <> [/] [//] [x n]	31
语言中断问题? → +//. +//? +... ++.....	31
遇到不需要转录的部分怎么办? → www %exp.....	32
难以理解的部分怎么办? → xxx	32
会话者引用其他人的话怎么办? → +"/	32
错误语怎么标注? → [::] [:] @u.....	32
复合词怎么标注? → _	32
如何标注体态语? → &=bodyPart:behavior %fac %gpx	32
&的其他用法→填补词、语音片段、&=laughs	33
@的用法→ @n.....	33
如何标注被省略的词、缩略语和连读: 0 ().....	33
如何编辑停顿? → (.) (..) (...) (min)	34
声音效果或无声→ [=!].....	34
胡言乱语或杂乱语的标注→	34
百分号的作用→ %com:TAB	34

简介：为什么要使用 CLAN

在临床相关科研中，与患者的交流是最常见的，也是最重要的资料之一。它们可以是最日常的对话，也可以是最专业的神经心理和行为学评估过程。这些材料包含了患者的疾病线索，也包含了研究者所关注问题的答案。但是，任何从事于言语语言病理相关临床与科研工作的人员，都会碰到一个棘手的问题：如何处理采集来的语料？

事实上，多数研究者仅仅在研究生阶段接触过语料分析，而即使在今后的研究生涯中使用相关语料，也只是作为质性分析材料。如何有效地管理自己采集到的患者语料，并且进行细致的深加工，最终形成可保存研究资料、发掘研究问题、支持研究假设的数据库呢？CLAN 可以为此提供良好的平台。

通过使用 CLAN，研究者将可以做到：

1. 快速处理语料，完善自己专有的数据库。
2. 提供便捷且准确的统计结果，提高科研质量。
3. 对语言材料进行从语义、语法到语言的多个平面全方位的管理。
4. 对于言语语言康复治疗，帮助记录临床治疗过程，评估和分析疗效。

本概要专门介绍 CLAN 进行文件管理和转写（transcribe）的流程。所谓转写，就是把语音材料转录成文本材料，并且对文本资料进行诸如词语切分、词性标注、体态记录等操作，最终使纯粹的语音资料转变成可供统计分析的专业资料。转写所生成的材料称为译文，在狭义上是指带有.chat 后缀的文件。

在学习和使用 CLAN 时，有一些看似繁琐的细节，可能会使用户最初感到棘手。但它是任何一项专业记录活动的必要阶段。当用户熟悉了 CLAN 系统的逻辑和操作规则后，这些细节所导致的成果，反而会形成掌控语料的愉悦。

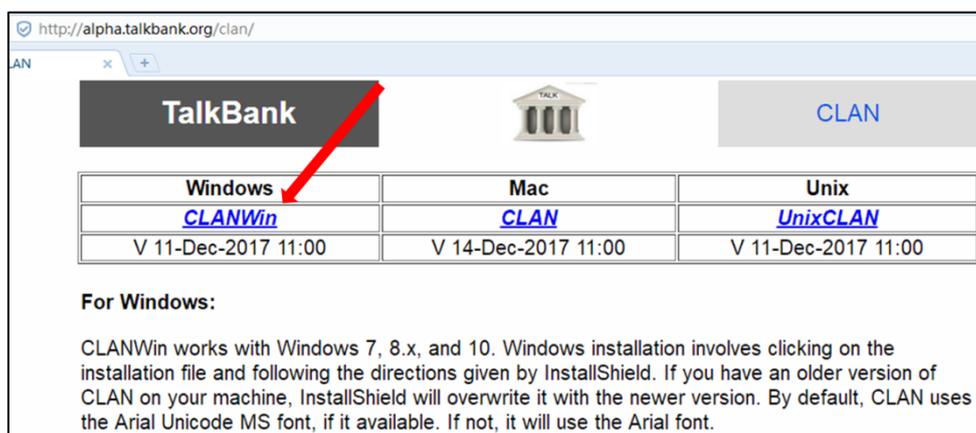
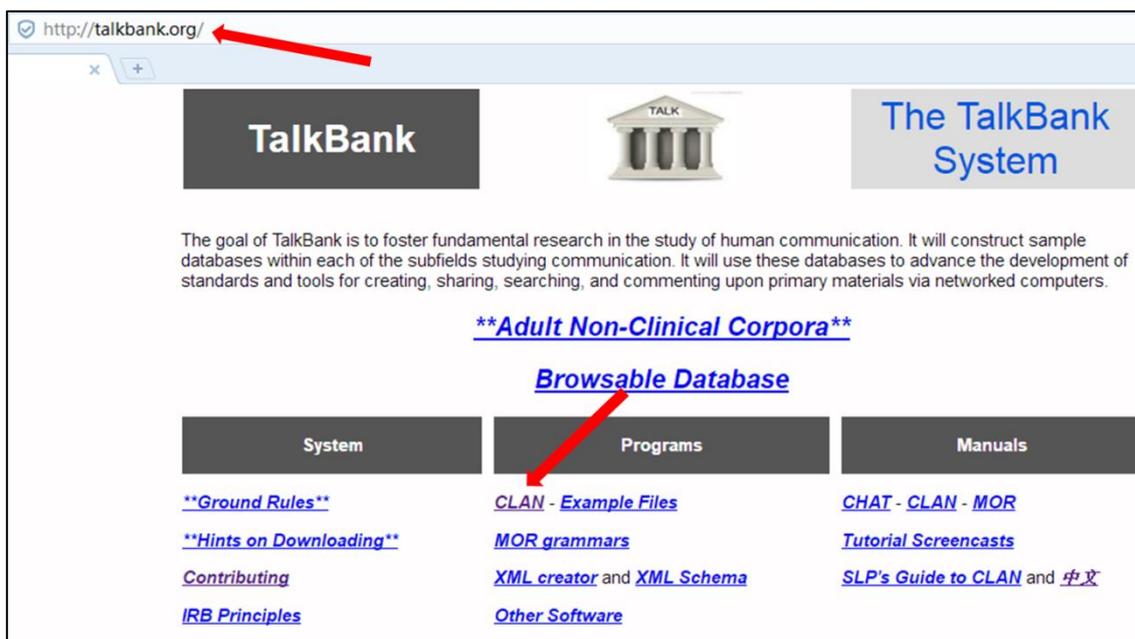
1、什么是 CLAN 和 CHAT

CLAN 和 CHAT 是 CHILDES（child language data exchange system，儿童语料交互系统）的一个子系统。它最初用于记录和管理儿童的会话材料，但是最终发展成为对世界上任何一种语言的记录和管理系统。CHAT 是一个用于对语音材料进行标准化转写的程序。该程序通常对应于.chat 格式的文件。CLAN 是一个对经过转写过的材料进行分析的程序，尤其是分析.chat 格式的语料。由于 CHAT 的材料最终使在 CLAN 中使用，因而在一般讨论时通常统一使用 CLAN 来表示这一整套分析系统。

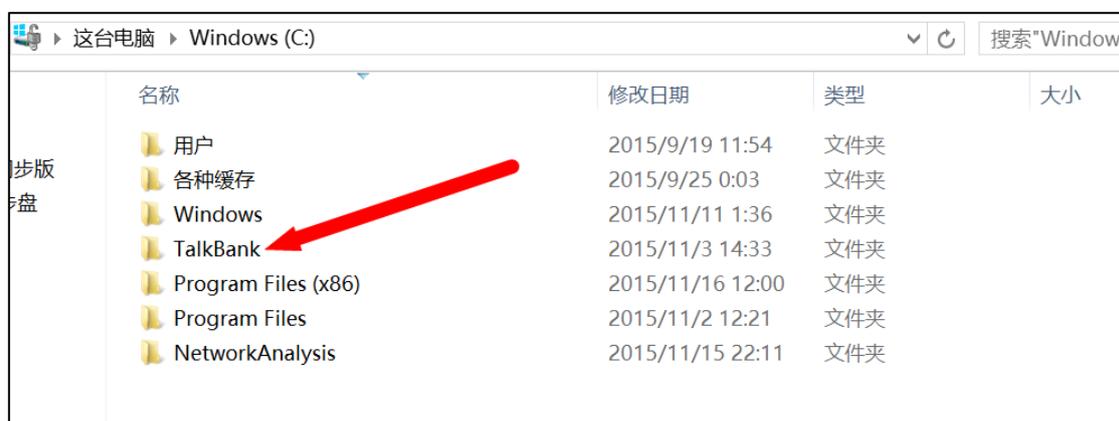
2、下载和安装 CLAN

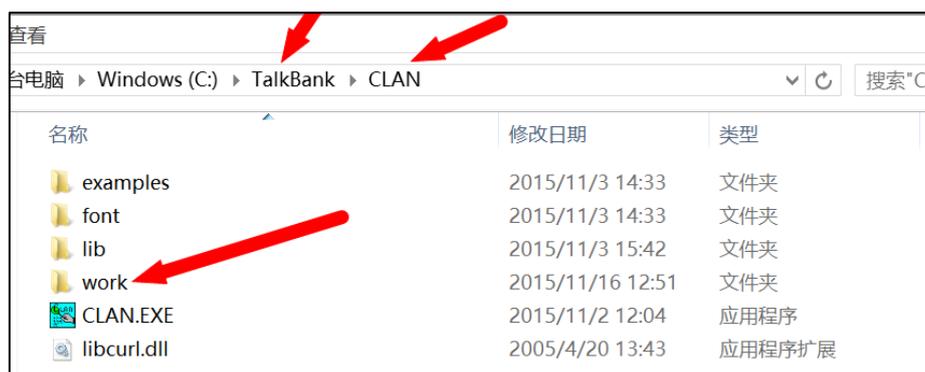
请注意，由于本文件的写作者仅有基于 Windows 的 CLAN 操作经验，因而以下概要仅针对视窗系统，包括 win7 和 win10。

CLAN 的下载地址为：<http://talkbank.org/>。以下是下载界面。

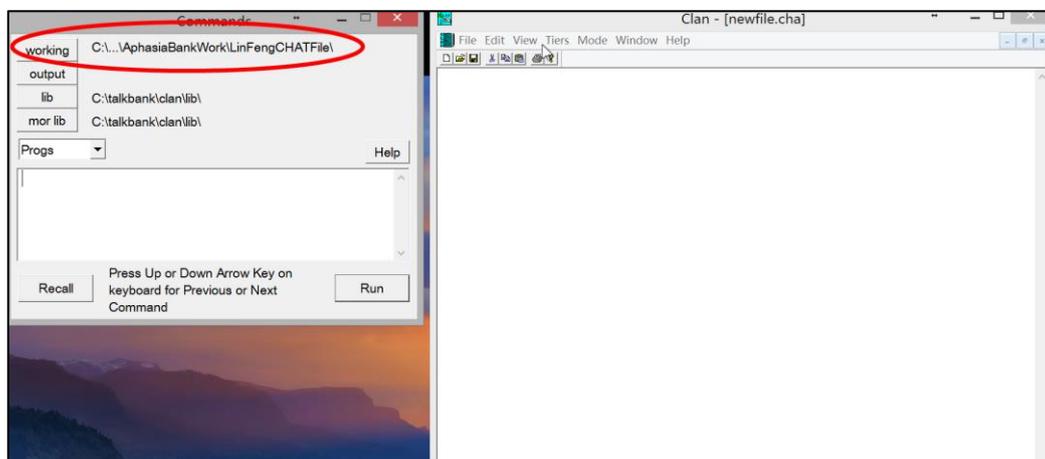


下载 clanwin.exe 安装文件之后，安装过程与常用软件并无区别。需要注意的是，由于 CLAN 需要查找和管理相关的音频或视频，而这些文件的位置对于 CLAN 的查询非常重要，所以 CLAN 最好不要安装在临时盘符（例如 U 盘或其他可拆卸硬盘）中。英文版指南建议安装在 C 盘中，但是考虑到大多数用户的 C 盘都空间很小，因而笔者个人建议可以安装在 D 盘。请注意，最好安装在 D:\TalkBank 文件夹下。但是，在实际安装之前，需要读者注意文件的组织结构。通常文件路径如下：





如上图所示，CLAN 的文件安装路径及文件组织路径并不复杂。但需要注意的是，当安装完成后，只要开始使用 CLAN，就需要关注你是在哪一个文件夹下工作。这个指定的工作文件夹，是你保存和管理自己的音频视频文件的重要位置，随意变动可能会造成 CLAN 无法匹配文件的后果。当然，读者可以自己指定相应的工作文件夹，例如笔者本人指定的文件夹如下（请参看[文件夹管理](#)一节）。



关于升级问题：可以直接到下载界面查找最新版安装。

3、安装配套软件

CLAN 在编译过程中需要使用 QuickTime 浏览器来打开视频文件。因此必须要安装和使用 QuickTime：用必应搜索 quicktime 即可下载。

CLAN 最常识别的多媒体格式是 .mp3、.wav 和 .mov。虽然它几乎可以识别所有的多媒体格式，但是笔者个人建议将所有视频文件统一转制成 .mov 格式。这可以使用“格式工厂”软件完成。请用户自行下载使用。

4、确认 CLAN 和电脑的设置

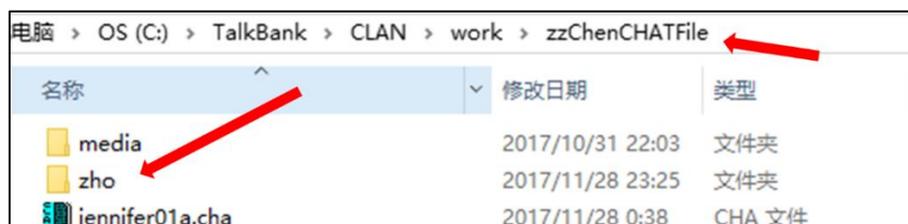
在安装成功后，要在对 CLAN 的配置先进行一番检查，通常以下内容在初装时是默认的，但如果你一段时间不使用 CLAN 了，再次使用时需要确认之。

在 CLAN 主菜单上点击 *Edit* 菜单栏，在下拉菜单中选择 *Select F5 option*。在弹出的 *F5 Option* 选项卡上，勾选 *Bullet on every TIER* 选项，并把 *Segment length* 设为 0，然后单击 *OK* 关闭选项卡。

再次点击 *Edit* 菜单栏，在下拉菜单中选择 *CLAN Options*。把 *Check point every* 右侧框盒中的数字设为 15。这个选项是告诉 CLAN 每编译多少行语料就保存一次。把 *Limit of lines in CLAN window* 右侧框盒中的数字设为 500（初装默认是 1000，如果显示行数过多，可以再此修改）。在 *Tier for disambiguation* 右侧框盒中键入 %MOR:（请注意，这里大小写有区别）。请确保勾选以下两个选项：*Open Commands window at startup* 和 *Check for CLAN updates on childe's web site*。

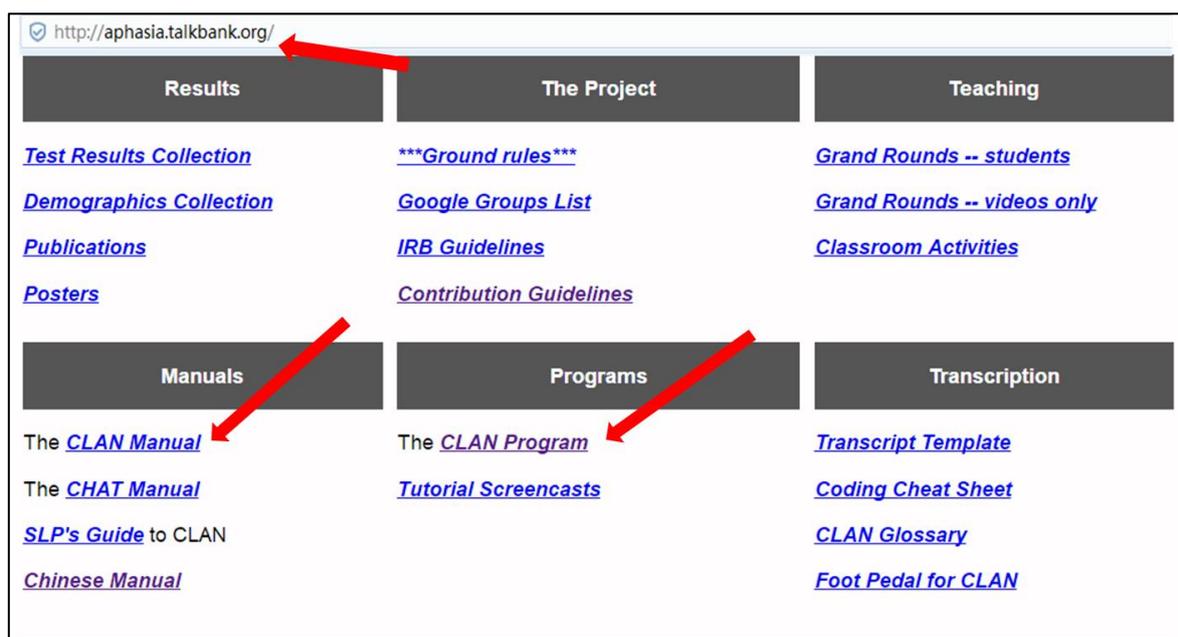
CLAN 还需要知道 *depfile* 的保存位置，因为这个文件指定用什么标准来校对译文。你需要先知道 CLAN 在电脑中的安装路径，才能设置 CLAN 识别 *depfile* 的路径。首先找到 CLAN 程序，双击打开它。确保 *Commands* 窗口也处于打开状态。如果找不到 *Commands* 窗口，请在主菜单上单击 *Windows* 菜单栏，在其中选择 *Commands*，或者也可以用快捷方式 *ctrl-D*。点击 *lib* 按钮来设置 *lib* 文件夹。点击后会弹出一个对话框，询问文库（*library*）的位置在哪里。一般初装时是默认 *C:\talkbank\clan\lib*。这里的路径没有大小写区别。

需要在电脑上安装 MOR 语法才能做专门针对特定语种的统计分析。该语法文件下载地址为：<http://talkbank.org/morgrams/>。在这个网址上你可以看到有许多语种的 MOR 语法。找到你需要的语种。例如 *Chinese(zho)* 是中文普通话，*English(eng)* 是英语。点击链接后会下载一个 .zip 压缩文件。将压缩文件夹解压即可，再把 *zho* 或 *eng* 这样的文件夹拖放到你为创建的 *CHATFiles* 文件夹中（请参见 *文件夹管理* 一节）。例如笔者本人的文件夹如下：



5、获取和使用失语库数据（初学者可略过）

失语库主页 <http://aphasia.talkbank.org/>，也有 CLAN 下载链接。



请注意：初学者可以略过这节。如果你仅想使用数据库作为教学或研究，而不想编辑和管理自己的数据文件，请了解该内容。但需要注意的是，所有的文件使用都需要遵循患者隐私原

则和版权。

在 *Database* 标题下，你可以看到 *Trasncript Database* (译文数据库) 和 *Media Database* (媒体数据库) 这两个链接。

如果想下载译文数据，请点击译文数据库的链接。然后会弹出一个对话框要求你输入用户名和密码。在用户名一栏键入 *broca*。至于密码，请通过电子邮件 *macw@cmu.edu* 与布莱恩·麦克威尼联系。(在输入用户名和密码的时候，你可以选择记住密码，以后就不用反复输入了。) 登陆到数据库界面后，你可以看到一张表单，其中的文件夹按照语言类型排列。在失语库主页的 *Database* 栏目下可以找到数据库指南 (*Database Guide*)。

在 *English* 目录下，英语失语库项目现有的数据都保存在 *Aphaisa* 和 *Control* 这两个文件夹中。至于研究者在失语库项目之前提交的数据，以及其他失语症研究项目的数据，都保存在另外的文件夹中。在这些目录中，你想了解哪个语料库，就请点击它。然后你会发现有一个与你所选的语料库名称相同的文件夹出现在下载路径中。如果同时出现了一个压缩文档 (.zip)，你可以删除它。但是，如果只出现一个压缩文档，那就请双击它解压缩，从而释放出与所选语料库名称相同的文件夹。如果这时候你还能看到一个下载对话框，那么你可以关掉它了。你所下载的文件夹 (译文文档) 应该含有译文数据文件 (扩展名为 .cha)，在这个文件夹下面还有一个名为 *media* 的子文件夹。在这个子文件夹中，有多媒体文件列表，这些文件可以打开存放于服务器上的多媒体文件。

失语库的译文都同时与音频和视频媒体文件关联。如果你想要连续回放原始媒体文件，请确保 CLAN 能够定位这些文件。要用 CLAN 查看译文以及相关的媒体文件，可以有三种方法：

1) 如果你只是想听听译文，而不想改动它。那么你可以使用 *WebData* 工具，在打开译文的同时，通过网络播放译文相关的媒体文件。从流媒体服务器上调用的媒体文件在在线播放时，会受到网络连接速度的影响，*Quicktime* 播放器会对媒体格式进行调整，以适应网络速度。当译文文件与媒体文件连接上时 (这时你可以在每一行结尾看到编号)，你可以在菜单栏上点选 *Mode* (播放模式) 下拉列表，选择 *Continuous Playback* (连续回放，或用快捷键 Esc-8)。

2) 如果你想要对译文进行编辑，但又不想同时下载与该译文关联的媒体文件，你可以只下载译文文件 (下载方法见前)，然后在 CLAN 中打开它，并且用 Esc-8 快捷键来回放媒体文件。CLAN 将会从流媒体服务器获取媒体文件进行回放，而且会再次根据你的网速来调整回放格式。

3) 如果你想对译文进行更高级的编辑，例如使用 F5 或步控窗口进行转写、编码或链接 (这些会在“转写”小节介绍)，或者你只是想避免使用网络 (例如你的拨号速度很慢)，那就要既下载译文文件，又下载原始媒体文件。

建议把 *media* 文件夹的原始名称改成 *refmedia*。这个文件夹中有两个文件，一个是译文数据文件，另一个是参考视频，两者都从服务器下载。然后，你可以再次创建一个 *media* 文件夹，用来放置下载的原始媒体文件，并且把这个媒体文件夹放到下载下来的译文数据文件夹中。这样译文文件和原始媒体文件就可以保存在同一个位置。

请注意：当你从失语库下载了译文数据文件之后，请查看一下文件夹中的内容。如果数据文件有 .cha 扩展名，那就是 CHAT 数据。如果标有 CLAN 的图标表现为其中的 CLAN 字样在图标左侧垂直排列，看起来像是一本淡蓝色的书，那么你已经把所有东西都设置好了。但是，如果默认打开 .cha 文件的是其他电脑程序，你需要重新指定 CLAN 程序为默认程序。

6、下载与查看 CHAT 和 CLAN 手册

在 CLAN 下载界面中，还有一个详细的 166 页的参考手册。请不要紧张，相信很少有人把整个手册都看完再使用 CLAN，请把它当做帮助文档。

7、文件夹管理

你可能会要预先在硬盘上为自己的 CHAT 文件专门建立一个文件夹。你可以命名为 *CHATFiles*，或者如果有不止一个人会用这台电脑，你可以用自己的名字或缩写，例如 *DFCHATFiles* 或 *LoriCHATFiles*。

注意!!! 所用名称中间不要有空格，而且绝不要在文件名中使用空格符。

注意：最好不要安装在中文路径下。笔者本人使用百度云是因为百度云可以设置英文路径名 *baiduyun*。

在 *CHATFile* 文件夹下面，可以为媒体文件建立一个子文件夹，视频和音频文件都可以放到这个子文件夹里面。媒体文件夹命名为 *media*。

请注意，media 文件夹名字要区分大小写，而且必须保持不变。

你可以 *CHATFile* 文件夹保留在桌面上，也可以把它移到硬盘的其他位置。我们建议建一个称为 *AphasiaBank* 的路径，然后在其中建立称为 *CHATFile* 的子路径，并且进一步建立更深一级的称为 *media* 的子路径。

在准备开始进行转写前，你要先把音频/视频媒体文件放到已经建好的 *media* 子文件夹里面。如果要从失语库下载原始的媒体数据，请进入失语库的主页 (<http://aphasia.talkbank.org/>)，在“Database”标题栏下方点击 *protocol Database*，里面有各种语言的数据库，比如中文普通话失语库，点击 *Mandarin Aphasia*，在作者信息简介下方有 *Download transcripts*（转录文档下载）和 *Media folder*（视频文档）的下载链接，需输入用户名和密码后（请参见第 5 节）可下载。*Non_protocol* 数据库不需要密码，可直接下载。请把下载的媒体文件拖放到你建立的 *CHAT* 文件夹下的 *media* 子文件夹中。

请确保：

- 媒体文件必须要放在子文件夹中，这个子文件夹的父文件夹中要存放该媒体文件所对应的 *CHAT* 译文文件。
- 确保媒体文件的文件名不含空格。

8、创建新的译文文件

不建议使用 Office Word 或其他文字处理软件，强烈建议在 *CLAN* 程序中进行转写。接下来将会介绍两种转写技术：步控窗口模式和 *F5* 模式。

首先，请打开 *CLAN* 程序。到菜单栏的 *File* 下拉菜单中选择 *New*，或用 *Ctrl-N*，建立一个空白译文文件。

在第一次做转写时，必须要输入最初几行译文。然后你就可以把这些起始行做为模板保存下来，下次用的时候只要修改一下参与人的信息或一些有必要增加或修改的细节，从而编制一个个专门的译文文件。

务必先要注意两点：

- 1) 每一行必须以回车键结束，无空格。
- 2) 冒号后面，紧接“Tab”键，不允许使用空格键。

文件头如下所示：

```
@Begin
@Languages: zho
@Participants: PAR peterduus01b Participant, INV WeihuaXIANG Investigator
@ID: zho|fengLIN|PAR|48;|male|TMA|Asian|Participant|15|
```

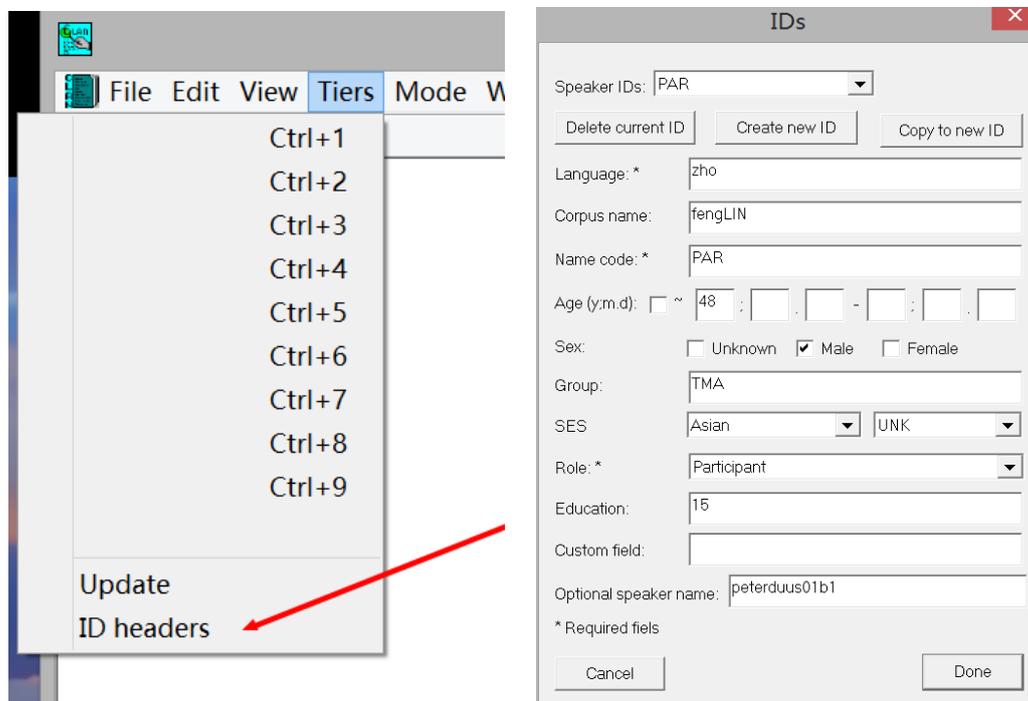
@ID: zho|fengLIN|INV|25;|||Asian|Investigator|18||

@Media: peterduus01b,video

%%以上为强制数据头

@Begin 表示文件开始。接下来几行都是必要的数据头，称为**强制数据头**。

- *Language* 代码来自国际 ISO 标准，它必须以如上形式输入：3 个小写字母。zho 表示中文普通话。
- *Participants* 项，必须以 3 个大写字母构成的 PAR 字母串开始，它表示受试者的 ID。INV 表示检查者的 ID。示例中的 peterduus01b 表示的意思是：由 peterduus 这个检查者提供的第 01 号受试者的第二次检查。失语库项目既包括语料库/检查者的名字，也包括受试者的化名。2 个数字序号后面加个 a 表示首次检查，此后 b 和 c 等字母顺次表示同一患者的第几次复查。程序设置了几种固定的角色。也可以在 lib 文件夹中的 *depfile.cut* 文件的 @Participants 这一行中找到。
- ID Header 是身份资料。这些资料可以由 CLAN 程序自动载入，或者你也可以从其他译文文件中拷贝过来，只需要把与当前受试者相关的信息稍加修改即可。载入资料的方式如下：请将光标放到 *Participants* 这一行的末尾，按一下 ESC 键，放开 ESC 键后，再按一下 L 键（请不要同时按下这两个键，也不要再按 L 键时不放开 ESC 键，只需要按 ESC，放开，再按 L）。或者也可以在 CLAN 界面中使用 *Mode>Check opened file* 菜单。这两种操作方式都会调用一个名为 *Check* 的 CLAN 过程。它会把 @ID 行加入到译文中，还有可能会在屏幕上弹出警示框。你现在可以不需要管提示消息是什么，只要点击“确定”关闭框盒即可。然后请在单词 *Participant* 前面的两条竖线之间键入受试者的身份信息（例如，holland01a）。在这个身份资料行中，还可以输入其他信息，例如失语症类型、年龄、性别等等。但是这种输入方法比较繁琐，笔者推荐使用 *Tiers>ID headers* 菜单编写 PAR（被试）和 INV（主试）的资料。



最后一项强制数据头是 *Media*，其中包括媒体文件名，其后带有一个逗号（半角），一个空格，最后是媒体类型（视频或音频）。注意不需要带媒体文件后缀名。注意：**@Media** 文件头里面的媒体文件名一定要与 media 文件夹里放置的媒体文件的名称一致。也就是说：**视频或音频一定要放在 CHAT 文件夹下面的 media 子文件夹里面，并且视频或音频文件名一定要与 .cha 文件的 @Media 文件头指定的名称相同。**

保存文件名：你现在得把这个才起头的译文保存成它的原名，它以一种特定的方式来表明受试者是谁。点开 *File*，选择 *Save As*，然后给文件命名。失语库使用的文件名原则是：数据提供人的名字，后面加上两位数字（01 第一个受试者，02 代表第二个受试者，以此类推），然后再加上一个字母（a 指受试者第一次测评，b 指受试者第二次测评，以此类推），最后给文件加上 *.cha* 后缀名。因此，如果 peterduus 是数据提供人，现在要记录第 05 个受试者，而且是这个受试者的第一次测评，那么文件名就记录为“peterduus05a.cha”。如果一个给定的测试被分成了 2 个部分，你可以记录为“holland05a1”和“holland05a2”。所有 CHAT 文件都必须是 *.cha* 后缀名。需要再次提醒的是，这个文件必须放在你的 CHATFiles 文件夹当中。

保存模板：编辑完文件头之后，你可以把文件保存为模板，以便用于今后编译。还可以给这个文件起个名字，可以是产生手头这些数据的受试者的名字。点击 *COMMAND-S* 功能键，或到 CLAN 菜单栏的 *File* 下拉菜单中，选择 *Save As*，然后把这个文件命名为 *template.cha*（或者任何你想用的名字，只要后缀名保持 *.cha* 不变），保存到你的 CHATFiles 文件夹中。在你点击窗口下方的 *Save File* 按钮前，请确保文件指向正确的文件夹（例如，DFChatFiles）。如果你准备编写一个新的 CHAT 文件，就可以到主菜单的 *File* 菜单栏中，选择 *Open*，然后打开上述模板文件。然后就只要修改名字、化名或其他任何需修改的细节，把修改过的文件转存为新的文件名，而译文中的已有强制数据头都还保留在原来的位置。

其他非强制数据头：

现在你已经完成了强制数据头（Obligatory Headers）的编写，可以继续往下编辑了。译文文件还有其他一些数据头需要编辑，包括固定数据头（Constant Header）和可变数据头（Changeable Headers）。这些内容都在 CHAT 手册中有详细介绍，其中有一些在本手册后面也会讨论到。但是，我们强烈建议你在自己的各个译文文件中统一采用以下数据头：

1) 检测时间

@Date: 05-JUN-2008

时间格式是：两位数字表示日期连字符、连字符、月份最初三个字母大写、连字符、四位数字表示年份。在这里，月份必须用英文表示。

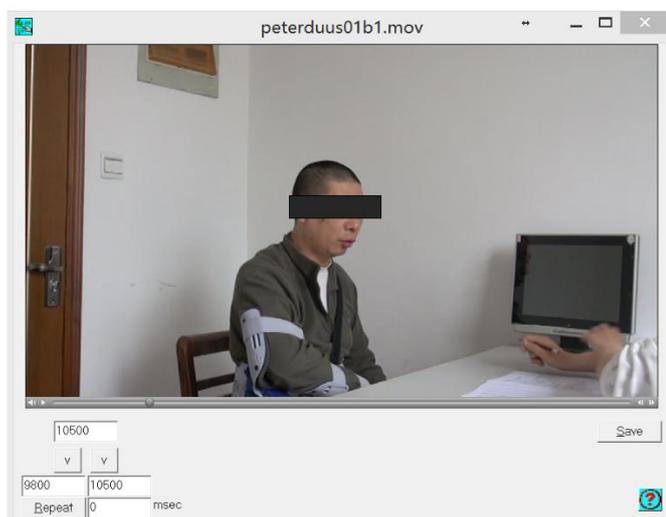
2) 用 @G 标识符来提示译文记录中的不同任务块。例如：@G: Cinderella 表示“讲述灰姑娘的故事”这个任务，“@G: Repetition”表示“复述”这个任务。

注意：

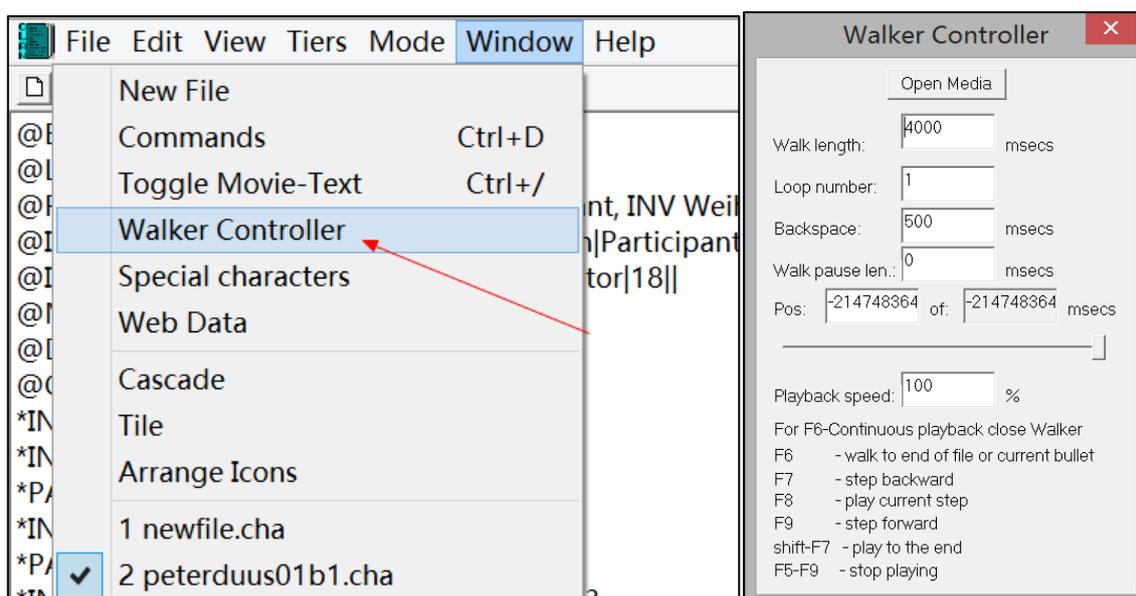
- ① 这些数据头的冒号后面都是加上一个制表符，而不是空格！！！！。
- ② 转录文本使用半角标点符号，每行末尾的标点符号则是任意的，但末尾标点符号之前要有个空格。
- ③ @End 是译文最后的强制数据代码，这个代码后面不允许出现任何标点符号。它是译文结束时的唯一标记。
- ④ 笔者建议：@G 所标注的任务块的名称中不要出现空格，可以用下划线连起来形成复合词表示任务名。例如 @G: cat_rescue。否则有可能会在分块检索的时候出现难以匹配任务名的问题。但同时也建议不要用驼峰命名法（如 catRescue），因为今后处理时难以切分单词。另外，请参看 [块检索指令和镶嵌符](#)。

9、译文转写过程

9.1 步控窗口设置



如果你在新译文文件中已经有了起始行，你就可以开始做转写工作了。有许多方法可以来完成转写过程。步控窗口（Walker Controller）有一个视频/音频回放系统，你可以随意停止、开始、快进和快退。另外，你可以设置回放指定长度的内容，并且自动重复循环播放该内容。窗口如图所示。左下角有三排格子。第一排的一个格子显示的是要从第几毫秒开始播放，在其中键入时间后点击右边的保存，再回车，即可从该时间点播放。中间两个格子是需要重复的时间起止点，在填好止点的格子保存后回车，即可在两个时间点之间回放。第三排有一个repeat，表示在设置时间框内重复播放。比如：第二排设置的两个时间点为：10000-20000毫秒，repeat内容为空或为0，即在10000-20000毫秒内播放；repeat设置为1000或+1000，即在10000-21000毫秒内重复播放；repeat设置为-1000.即在9000-20000毫秒内播放。注：设置了任何之间点，一定要保存后按repeat即可播放。



如上图所示，在CLAN程序的菜单栏上找到Windows下拉菜单，选取其中的Walker Controller。然后就会在屏幕上出现一个名为“Walker Controller”的对话面板。点击该面板顶端的“Open media”（打开媒体文件）按钮。它会弹出一个对话框，其中显示的是你电脑上的路径。找到并选定你需要打开的音频/视频文件。这个文件应该已经事先放到了你桌面的CHATFiles文件夹下的media子文件夹中，或者是在你硬盘的AphasiaBank\CHATFiles\media子文件夹中。在开始使用步控窗口前，需要确认以下设置。

- *Walk length*（步长）设置的是你希望一节播放多长时间，这样你就可以控制视频/音频分节播放的长短。默认从4000msecs（4000毫秒）开始，然后再调整长度。可以随时改变这项设置。
- *Loop number*（循环次数）设置的是你希望每一节循环播放的次数。如果你设置步长为4000毫秒，循环次数为2，那么视频/音频就会自动播放一段长度为4000毫秒的内容，并且自动重复播放两次，随后进入到下一节4000毫秒长度的内容，再次重复两次，以此类推。如果你把循环次数设为1，每节内容就只会播放一次。我们建议把它设置为1，但是你可以尝试一下，看看自动循环多少次比较符合你的习惯。
- *Backspace*（备份空间）设置的是你希望程序在播放下一节新的内容之前，对多长时间（以毫秒计）的内容进行备份。我们建议设置为0。
- *Walk pause length*（休止长度）设置的是各次循环之间以及各节之间的休止时间长度。这个选项仅适用于音频文件，对视频文件无效。
- *Playback speed*（回放速度）的功能就是控制媒体文件的播放速度。慢速播放有时候对处理某些人的媒体材料是很有用的。

在你进行转写时，可以随时对以上设置进行更改。只要点击想要修改的项目，在框栏中填入所需内容即可。你可能会发现，不同受试者在语言交流上有不同的特点，如果要适应这些特点，可能要在转写时采用不同的设置。

使用步控窗口进行转写时，你可以使用面板上列出来的快捷键（见上图设置框下部区域所示）：F6播放到文件尾部或本行打点处；F7回退到上一节；F8播放本节；F9快进到下一节。这些快捷键实际上就类似于录音机的控制键。如果发现功能键不能正常工作，请按照手册对电脑进行配置。

笔者建议：由于快捷键太过靠近，容易误按操作，只建议用以下快捷键：

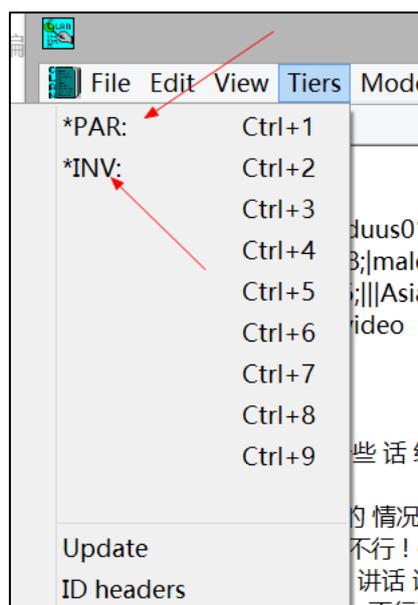
- F4进行按行回放；
- F5进行重新打点（打点操作请参考下文）；
- F6逐行播放到结尾（播放时会逐行深色显示）；
- ESC-8也是逐行播放到结尾，按法是先按ESC，然后松开，然后再按8，不能同时按。

电脑技巧：在一些复合键盘（如SurfacePro的键盘），F系列按键需要与Fn功能键联合使用，如何才能直接使用F系列键呢？可以用Fn+Caps键来让F系列按键直接就表示F1~F12。

以下是一行例子：

*PAR: 猫 正在 爬 树 .

每一段话语都需要另起一行单独记录，以星号开始，然后是说话人的身份ID（在强制数据头的Participants这一行已经设置了用3个大写字母来表示身份ID），接下来就是一个冒号加上制表符。如何切分话语单位，请参看文末。



为了能快捷设置每行的数据头，你可以如上图所示，在菜单栏中打开*Tiers*下拉菜单（此处*Tiers*表示层级，即按照某种指标把数据进行分类，此处是按照说话人身份分类），然后点击*Update*。它会自动把你的第一发言人身份（已经在*Participants*行中设置好，通常是PAR）绑定到或Ctrl+1快捷键，第二发言人身份绑定到Ctrl+2功能键（通常是INV）。这样你就不用每次都输入星号、三个大写字母、冒号和制表符，而只需要按Ctrl+1或Ctrl+2快捷键，就可以把每一行的起始代码填好，从而直接进入这一行的转写工作。不过，除非你没有用模板而从草稿开始创建新的译文文件，或者你在打开译文文件之后又改动过强制数据头的*Participants*这一行的内容，否则你就不需要*Update*，因为发言人的分层排列在你前期准备的模板中早已设置好了。

注意事项：

- 译文中只允许专有名词和“**I**”可以用大写字母。因此，请务必小心，不要在每行话语起始时出现大写字母。当然，在中文录入时没有这些问题。
- 每行译文都会以标点符号结尾，通常是英文句号、问号或感叹号。结尾符号前需要保留一个空格。
- 其他符号也可以用来标注话语中的停顿，或捕获不完整话语，用法请参见文末问题解答部分。
- 逗号可以在需要的时候用于表示短语之间的衔接，但CLAN程序不会用到它们，也没有在逗号与语音韵律的关系上做出严格规定。但是要注意，逗号后面需要有一个空格，然后再接写下一个词。

请阅读文末常见问题解答部分，然后浏览“链接由步控窗口转写的文件”和“文件校验”这两节。

9.1 译文示例:

以下是一则译文（笔者自制，请用户谨慎参考）。在窗口中译文会自动换行。但这种自动换行只是意味着显示上的差异，在处理分析时，仍视为同一节会话（utterance）。

```
@Begin
@Languages: zho
@Participants: PAR peterduus01b Participant, INV WeihuaXIANG Investigator
@ID: zho|fengLIN|PAR|48;|male|TMA|Asian|Participant|15||
@ID: zho|fengLIN|INV|25;||Asian|Investigator|18||
@Media: peterduus01b1, video
@Date: 17-MAY-2014
@G: greeting
*INV: 好, 现在 我 要 . •0_7440•
*INV: 一会儿 要 请 你 讲 一 些 话 给 我 听 . •7440_9800•
*PAR: <好> [>]! •9800_10500•
*INV: <看看 你 讲话> [<] 的 情况 . •10500_11100•
*PAR: &=left_hand:wave 不行! •10800_11440•
*INV: 好, 你 现在 觉 得 自 己 讲 话 讲 得 怎 么 样 ? •11440_14568•
*PAR: &ne &=head:shake 不行 了 ! •14568_16447•
*INV: 那 你 讲 话 是 很 困 难 啦 ? •16500_19700•
*PAR: &哎 &ne &=left_hand:sagittal_plan_left_hand_helix 试 [] 试试 [] 试试 看
&=left_palm_poen! •19577_22550•
*INV: 试 试 看 啊 ? •22550_23800•
*PAR: <&=head:nod 对> [<]. •23500_24000•
*INV: <好的> [>]. •23800_24500•
@G: cat
*INV: 我 们, 我 待 会 儿 要 给 你 看 一 些 <图 片> [>]. •24500_27200•
*PAR: <对> [<]! •26700_27600•
*INV: 然 后 你 花 一 点 点 时 间 来 看 这 些 图 片, 然 后 它 们 讲 了 一 个 故 事 . •27500_33856•
*PAR: 故 事 . •33450_34366•
*INV: 你 一 边 看 一 边 讲 给 我 听 好 不 好 ? •34366_36736•
*PAR: 这 个 +... •36736_37627•
*INV: 故 事 要 有 开 头 经 过 和 结 尾, 你 可 以 一 边 看 图 一 边 跟 我 说 . •37500_43000•
*PAR: &=left_hand_on_picture 他 那 个 &ze
&=left_hand_thum_index_finger_measuring 那 个 &yuxi 拔 出 来
&=left_hand_palm_upturning &=left_hand:sagittal_plan_left_hand_helix
&e 这 &ke &guozi &ke 是 &kuohe 是 &wuwu &ka 他
&=left_hand_index_finger_pointing_to_picture 有 一 个 &kao个 这 个 +... •42940_60000•
*INV: 我 们 停 一 下, 哦, 让 我 们 停 一 下 子 . •60000_63715•
@End
```

9.2 F5 转写模式

当你已经设置好强制数据头，并且准备开始转写时，首先需要做的是打开媒体文件，听取其中内容，然后输入**打点符**（bullets），这种标记的作用是结束一段话语，并把这段话语与一行译文关联起来。关于如何对话料进行分节，请参考文末问题解答一节。

按下F5键，或点击CLAN菜单栏上的Mode下拉菜单并选择*Transcribe Sound or Movie*。在按下F5或进行如上菜单操作之后，就会弹出一个对话框，要求你定位将要转写的媒体文件。选中目标媒体文件，然后点击*Open File*（打开文件）按钮。媒体文件随之在屏幕上打开，并且开始播放。你可以先听一会儿以适应媒体播放的状况。如果准备好了，就按动F5键停止播放，然后在准备要输入打点符前，再次按动F5键。

如果要输入标注符，请在说话人结束一段话语时按动**空格键**。程序会自动添加一个星号、冒号、制表符以及打点符•。如果你不主动停止，播放和按空格键打点的过程可以继续下去。停止播放的方法有两个，一个是再次按下F5键，另一个是点击屏幕上的CLAN转写界面。

在结束打点后，请保存文件。如果文件非常长，请在编辑过程中随时注意保存。可以设置让CLAN程序定时自动保存：到菜单栏上点击*Edit*下拉菜单，然后选择*CLAN Options*。在*Checkpoint every*的框盒中填入数字，然后CLAN就会按照指定长度来保存文件（如果填入15，那么大约每点击空白键15次就会保存一次）。你可以用这个选项来控制保存的量。

当你不再输入打点符时，你可以回到译文第一行，然后开始输入文本。你可以把光标放在第一行的任意位置。在开始输入之前，建议你到CLAN菜单的Tier下拉菜单中，点击*Update*。这样的话，CLAN就会把Ctrl+1和Ctrl+2这两个快捷键与强制数据头的*Participants*行所设置的第1和第2说话人身份联系起来（通常分别是PAR和INV）。

如果你把光标放到打点符标记的第一行上，然后按下F4，或到CLAN菜单栏的Mode下拉菜单并选择*Play bullet media*（播放打过点的媒体内容），CLAN就会播放与该行相应的那段媒体内容。你也可以按下F5，这样媒体内容就会忽略每行的打点符，不在各行停顿，而是按照自然速度连续播放下去。你可以再按一下F5，或者点击鼠标，从而停止播放。之后你就可以把光标放到你想重播的另一行上，再次按下F5。你可以反复执行这些操作，并且经常保存，直到媒体文件结束。

9.3 译文与媒体进行关联

一旦你已经在步控窗口中完成了转写，你需要把它“关联”到媒体文件上，就像你给多媒体文件配字幕一样，你需要指定每一段会话出现在多媒体文件中的哪个时间段。这种关联是在每段会话末尾“打点”来完成的。

如果你刚刚完成步控模式转写，请先关闭步控窗口（*Walker Controller*面板），并且关闭媒体文件，但是不要关闭已经保存好的CHAT译文文件。如果你早就完成了转写，只是想要再回来完成关联操作，那就打开CLAN程序，在File下拉菜单中点击*Open*，然后选择想要操作的CHAT文件打开它。

当CHAT文件打开时，请把光标放在第一行译文上。你现在要做的事情是重新打开并播放媒体文件。当然，这种打开和播放应该是以另外一种模式来实现。你需要听取媒体的内容，然后在每段话语结束时按动空格键。这种操作会在译文的每段话语上添加提示信息，表明该段话语出现在媒体文件中的什么地方，这也就是文件关联操作。

在准备好开始关联操作时，请按下F5键。媒体文件就开始播放。请集中精神，每当你到达一段话语结尾时就按动空格键。在按完空格键后，就会在译文的相应一行的末尾看到一个打点

符 (bullet)。这个打点符含有文件关联信息。光标会自动移行到下一行，而媒体文件会继续播放下去，所以在进行这些操作时要聚精会神才能确保每行末尾都及时按下空格键。

如果想要停止操作，只需要在编辑界面点击鼠标或再次按下F5键。

如果没有及时按下空格键，或者觉得自己按错了，你可以重新关联这一行。只需要把光标放到最后一个正确的打点符上，然后按下F5键，就可以再次开始关联操作。这时候媒体会从你光标所在行开始播放，当你按下空格键时，新打点符会覆盖旧打点符。

一旦把译文文件和媒体关联起来，你就有两种方式来同步播放译文和媒体：

- 1) F4键是单步播放光标所在行；
- 2) ESC-8会连续播放。注意先按EXC然后松开，然后再按8。

这些功能键都列在菜单栏的Mode下拉菜单中。任何一行译文都可以随时接受重新关联的操作。在完成整个译文的关联之后，请再次保存它。

9.4 手工编辑打点符 (Mode>Expand bullets)

打点其实是一种很需要技巧的操作，有时候很难做到在译文特定位置打下准确的时间点。CLAN提供了精确微调打点符的手工编辑途径。

首先你需要扩展译文文件中的打点符。打开译文文件，在菜单栏上点开Mode下拉菜单，选择Expand bullets，或者按下ESC后放开，再按一下A键。此时，每一行译文都会显示数字（它表示的是各行译文与媒体文件中的什么时间相对应）。

例如，如果你需要把一段话语切分成两小段句子，那么只需要看一下第二行结尾的数字，然后把这些数字剪切并粘帖到你想要创建的新句子中即可。如果你想把两段话语合并成一段，以如下话语为例：

```
*PAR: 好 .•250830_252448•
*PAR: 没问题 .•252448_253770
```

你现在可以把后面那句“没问题”剪切并粘帖到“好”后面，然后把最原来第二段结尾的数字剪切并粘帖覆盖原第一段结尾的数字。结果如下：

```
*PAR: 好, 没问题 .•250830_253770•
```

请注意：

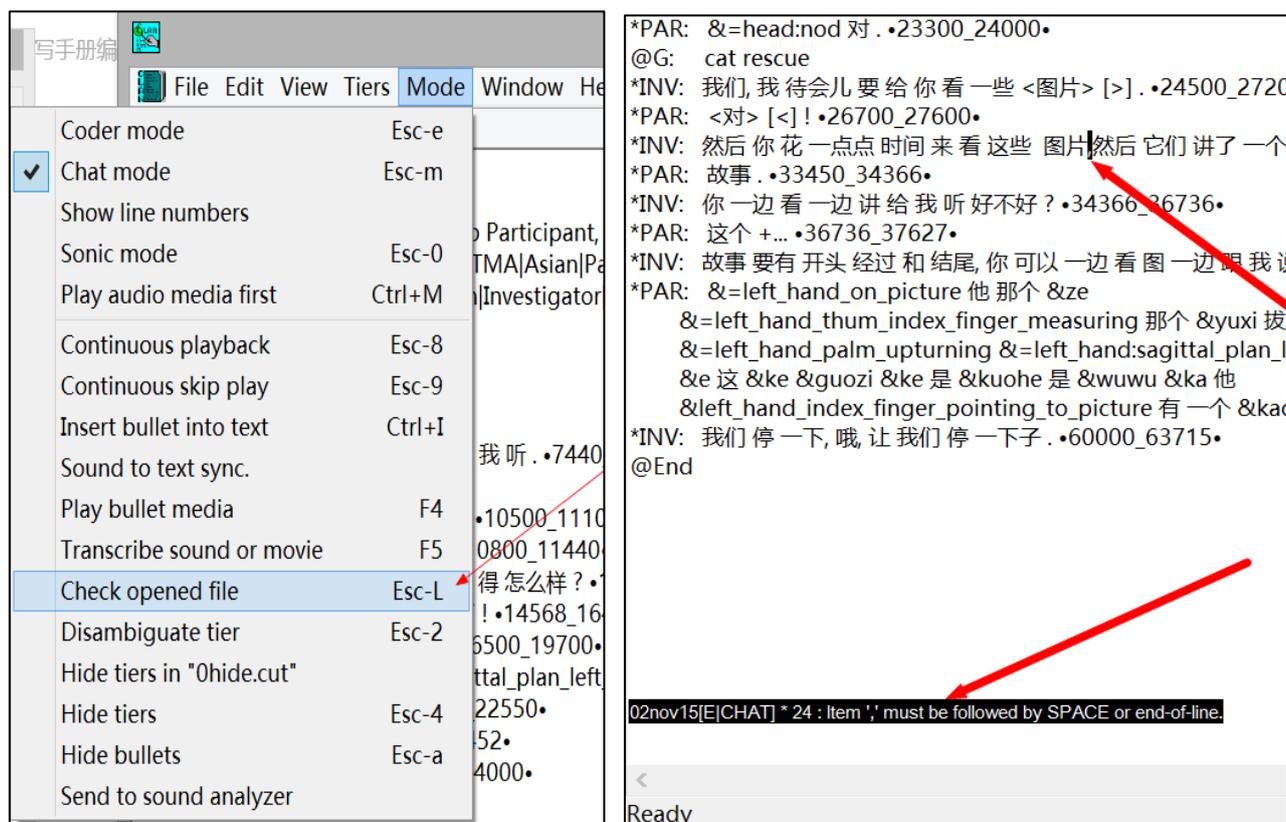
- 当光标停留在某一行时，可以按下F4键播放该行内容。
- 打点符包括了每行的两个数字之间的下划线和两端的点号。
- 剪切和粘帖全部或部分打点符都是可操作的，但是如果你需要控制某些重叠的话语时（即两个人同时在讲话），就会发现直接手工编辑数字的操作就变得很有用。

笔者建议：转写时先将译文与媒体关联，然后进行F5撰写模式，结束后利用手工编辑打点符对译文分句进行精确调控和核对。

10、文件校验

无论你多用心去编辑和关联文件，CLAN的文件校验功能总是会发现一些差错。因此，对文件进行校验是不可避免的一项操作。文件校验最终目的是为了报表界面做出提示：*Success! No errors found.*（校验成功，未发现差错）。

10.1 逐条显示差错：Esc-L 或 Mode>Check Opend Files



ESC-L方式：打开已经保存好的CHAT译文文件，点击ESC键，放开它，然后再点击L键。
菜单方式：点击菜单栏上的Mode菜单，选择Check Opened File。

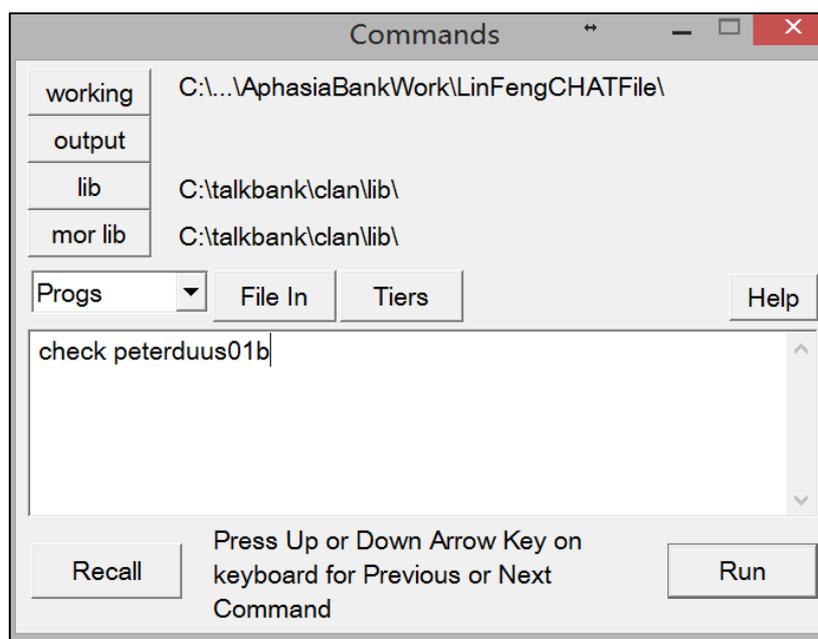
这两种方式都会在你的译文页面底端显示一条错误提示信息。光标则会自动转移到出现错误信息所对应的位置。

你可以修正错误，然后再次使用ESC-L功能键（或再次使用菜单操作调用校验功能），继续对译文进行校验。当然，你也可以一边转写一边调用校验功能，不过这样会使你无法集中精力进行转写。请注意，进行修改后，务必要及时保存。你可以不断进行上述校验操作，直到出现校验成功的提示信息。

10.2 一次显示所有差错：check 指令

另外还有一种方式，可以一次性把译文中的所有差错都显示出来。你需要打开指令界面（Command）：用Ctrl-D，或者在Windows下拉菜单中点击Commands。

在指令界面里面，请确保工作路径已经设置成你现在正操作的文件所在的路径。要进行这种设置，请先点击Working按钮，然后找到你的目标文件夹（例如CHATFiles或任何你自己命名的文件夹），然后点击Select Folder。然后请把光标放到指令界面的框中，先输入check，然后按一下空格键，然后再输入你想要校验的文件的名字，或者在输入check并按空格键以后，在指令界面的中央找到FILE IN按钮。点击这个按钮就会弹出另外一个窗口，请在左侧选择你想要的CHAT文件，然后把它加入到右侧的框中。请确保你在正确的位置查找文件。高亮显示你的文件，并点击Add或双击文件然后点击Done。



点击 *Run*（运行）按钮，校验程序会创建一张表单，其中列出了所有差错。它们会输出并平铺在CLAN文件界面中。在表单中从上往下寻找含有***和File字样的提示行。在这种提示行上，用鼠标快速点击三次，如果被校验的CHAT文件没有打开，那么它会被打开，而其中含有差错的那一行就会高亮显示。这样你就可以看到对差错的具体描述，并能够找到相应位置进行修改。

在完成修改后，请再次保存文件，然后再次运行校验程序，直到获得校验成功的提示信息。再次运行校验程序的方法有两种，一种是重新执行上述校验操作，另一种是把光标放到指令界面指令框中，点击键盘上的↑键，然后就会看到最近一次执行的指令“check @”出现在框中，而在框上方会出现正方形的按钮，这时请点击*Run*按钮重新运行校验程序，在CLAN输出屏幕上就会出现新的报表。

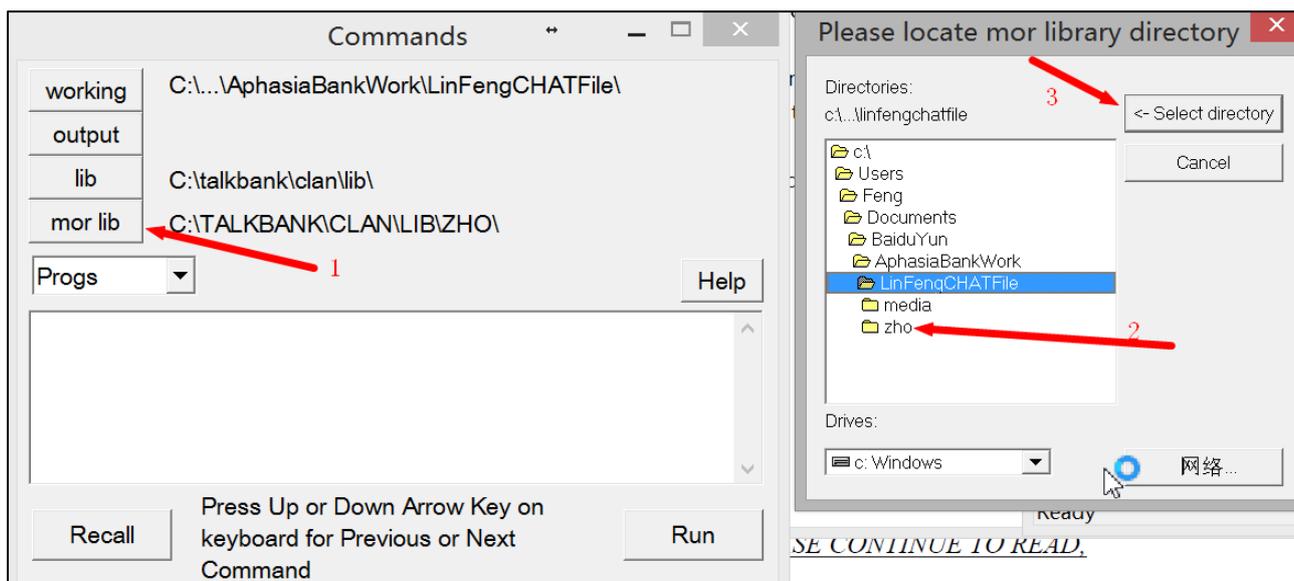
请注意：

- ✧ 请确保媒体文件的文件名中没有空格符。否则校验程序会把媒体文件名中的空格视为每一行话语的差错。媒体文件名实际上嵌入在打点符中，你可以用两种方式来看它：点击ESC再点击A，或者打开*Mode*下拉菜单点击*Expand bullets*。
- ✧ 有时候在一行话语结束时的句号前需要输入一个空格。如果在某行被提示有错的译文中实在找不到任何差错，请尝试句号前输入一空格。
- ✧ 因为CLAN没有中文提示，请仔细阅读英文提示，确保看懂了内容。

10.3 MOR 校验: mor +xb filename.cha

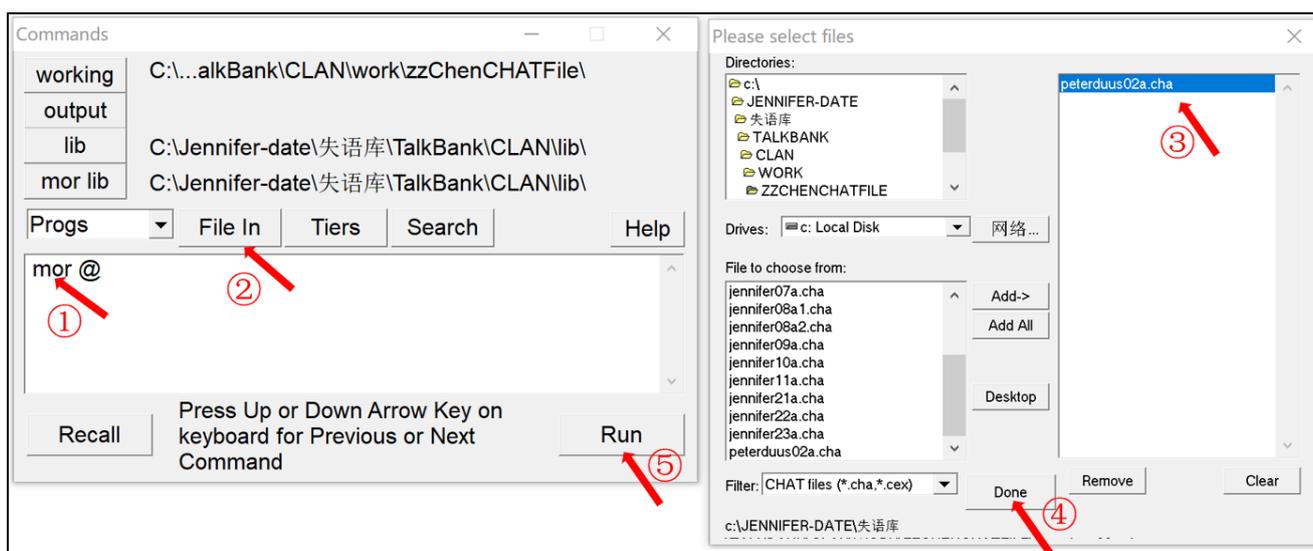
MOR是morphology的缩写，该程序提供了它可以检查是否能识别译文中的所有词。MOR校验会用于分析词类（morphological categories）和句型（syntactic categories）。

语法文件下载地址为：<http://talkbank.org/morgrams/>。在这个网址上你可以看到有许多语种的MOR语法。找到你需要的语种。例如Chinese(zho)是中文普通话，English(eng)是英语。点击链接后会下载一个.zip压缩文件。将压缩文件解压出来，再把zho或eng这样的文件夹拖放到你为创建的CHATFiles文件夹中。然后再指令界面进行如下设置：

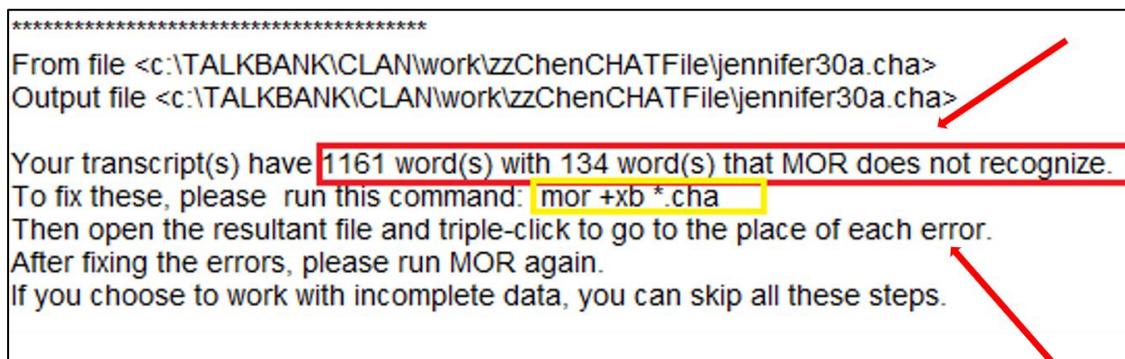


在执行框内输入“mor”，就会出现如②中的三个按钮，以供设置mor所需的参数。三个按钮代表的意义请参照第11条。点击File in，选中目标文件后点击Done，mor就会出现一个@符号。用户也可以在空白栏中的mor后面加一个空白键，手工输入.cha目标文件的名字。

执行界面如下：



结果界面如下图：



结果表示：在jennifer30a.cha这个文档中，共有1161个单词，其中134个单词MOR语言无法

识别。需要执行`mor +xb +文件名.cha`指令确认不能识别的单词，然后连续点击三次即可定位到目标文件中不能识别的位置。

如果在你的文件中含有MOR程序无法识别的词，可能有以下原因：

- 1) 该字词在此前译文中从来没有出现过；
- 2) 拼写错误；
- 3) 该词在对话中的用法不同。

注：MOR校验前，请先备份。因为MOR校验后，CHAT文件会不可逆地加入%mor和%gra行。

mor +xb filename.cha

请确保工作路径（由指令界面上部的`working`按钮设置）已经设置成你的目标CHAT文件所在文件夹。如果工作路径不对，请点击`working`按钮，并选择正确的路径。然后，回到指令界面，点击右下角的`Run`按钮。这样就会产生一个新的文件，这个文件存放在你正在校验的这个.cha文件所处的同一个文件夹中。新文件名与.cha文件相同，但后缀为.ulx.cex。如果你打开这个文件，就会发现这是一张表单，其中列出了未能被MOR校验识别的所有单词。这样你就可以纠正拼写错误，或修正其他遗漏的差错。

在进行深入的统计分析之前，如果你还没有转写好一批文件，就暂时不需要关心mor是如何分析文件的，但需要注意：

- ✧ 掌握上述指令，并且知道打开.ulx.cex文件查看未能识别的单词。

名称	修改日期	类型	大小
media	2015/11/17 16:11	文件夹	
zho	2015/7/23 20:33	文件夹	
baorunjianTemplate.cha	2015/11/17 16:13	CHA 文件	2 KB
peterduus01b1.cha	2015/11/17 22:48	CHA 文件	2 KB
peterduus01b1.mor.cex	2015/11/17 22:47	CEX 文件	5 KB
peterduus01b1.mor.pst.cex	2015/11/17 22:47	CEX 文件	4 KB
peterduus01b1.ulx.cex	2015/11/17 22:41	CEX 文件	1 KB

- ✧ `mor`指令的语法中，如果使用+1（例如 `mor +1 peterduus01b1.cha`）则表示分析结果会添加到`filename.cha`这个被分析文件中。这样做的好处是可以避免在工作文件夹中生成许多如图所示的分析结果文件。如果你有大量分析工作要做，整个文件夹会被填满各种杂七杂八的结果文件。但是，这样做的坏处是有可能你的原始文件会遭到覆盖和破坏。所以笔者建议还是不用+1为好。而且笔者强烈建议，当你生成了一个不需要再加工的.cha文件时，无论你是不是使用覆盖指令，请务必保存一个副本（例如 `peterduus01_bac.cha`），这样在今后分析时一旦文本受到损坏，还可以从副本恢复。
- ✧ 无论是哪个语种，`mor`程序首先调用的是`ar.cut`文件。普通话该文件位于`zho`文件夹下。请确保该文件位于正确的位置。
- ✧ `.mor.cex`文件是在执行`mor`校验后自动标注词性等信息形成的经校验文件。

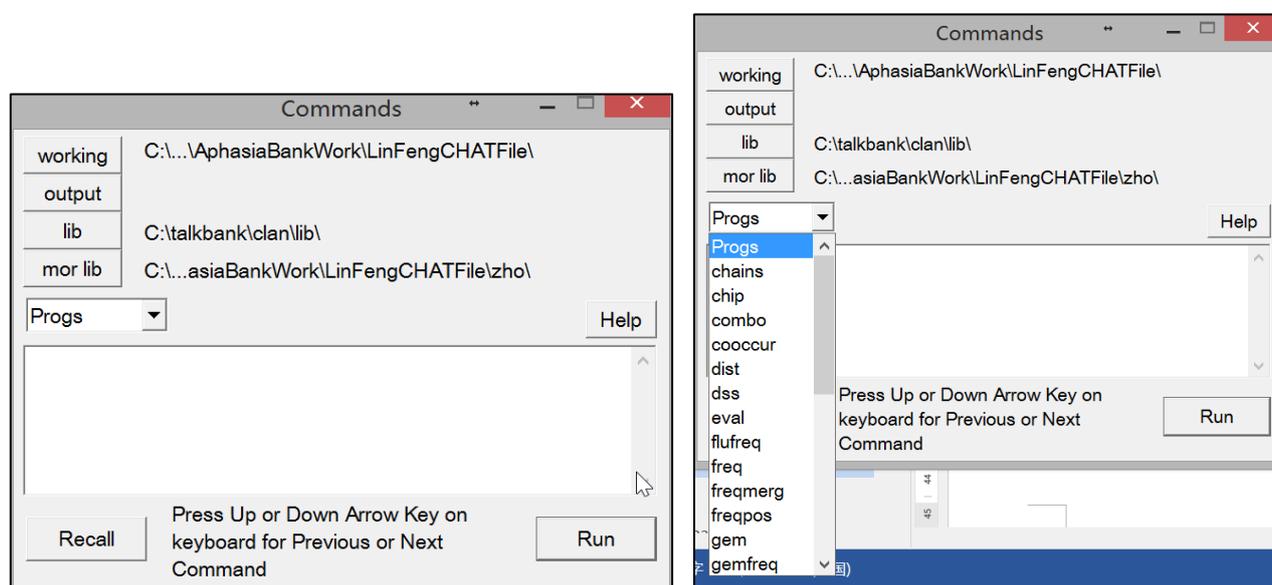
10.4 KIDEVAL 和 EVAL 分析

CLAN从2012年下半年开始提供一个非常有用的分析指令KIDEVAL。该指令用于失语症的版本是EVAL。执行该指令的前提是你已经安装了相应的MOR包。执行该指令后可以生成一

个EXCELL文件，其中包括了许多有用的分析结果。例如：

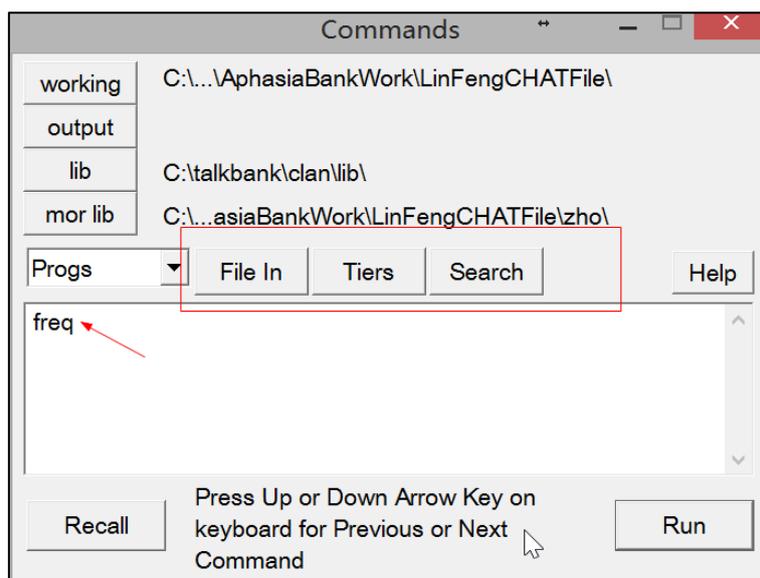
- 语素型话语均长（mean length of utterance in morphemes, MLUm）、
- 单词型平均话语均长（mean length of utterance in words, MLUw）、
- 形次比（type-token ratio, TTR）、
- 从句密度（clause density）、
- 语素量（number of major morphemes observed）。
- 还有包括：样本的持续时间（Duration）、会话单位数量（number of Utterances, Utts）、每句话的平均数量、单词水平错误的数量、句子水平错误的数量、重复的次数、自我纠正的数量、词性(名词、动词、代词、介词、副词、形容词、连词和限定词所占的百分比)、动词时态(第三人称：现在、过去、进行、完成)、和复数等。

11、CLAN 统计分析概览



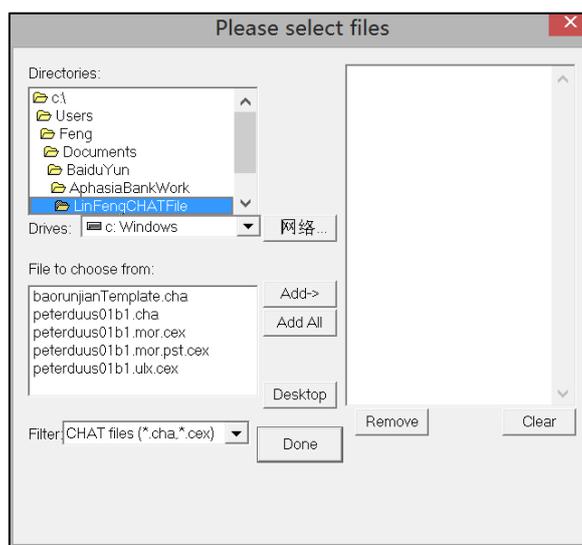
在CLAN的指令界面结构如上图。

- **Recall按钮**：点击它之后，会弹出一个窗口，其中显示了至少20条最近使用过的指令。如果在空白的框栏中放置光标，然后按动键盘上的↑或↓，则可以逐条显示上一条或下一条指令。CLAN的指令采用的是类似于程序编码的形式。
- **Help按钮**：点击后会显示帮助文件。
- **Progs下拉列表框**：里面给出了一些未设置参数的常见指令，在点击某条指令后，还需要用户在指令后面自己手工添加参数。
- 另外还有一些隐藏的按钮，必须要等某些指令输入到框栏中时才会显现。如下图。在输入freq之后，出现三个按钮，以供设置freq所需参数。

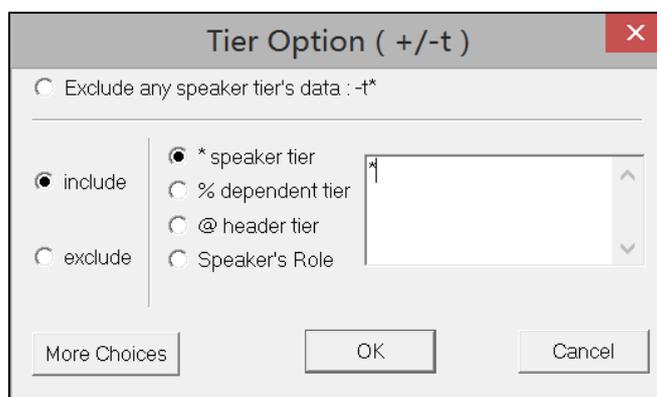


CLAN提供了多种图形用户界面GUI，使用户不仅可以直接编辑代码行进行统计，还可以通过窗口点击操作来选取一些不常见的指令选项。

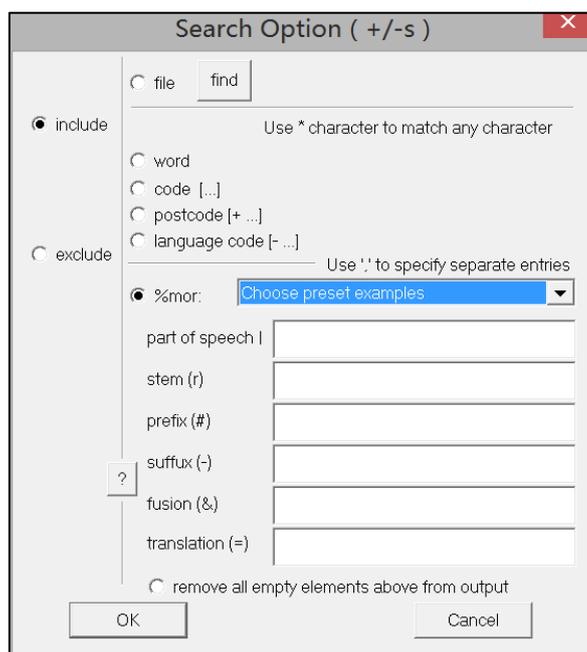
以下分别是执行freq指令时，按动File In、Tiers和Search三个按钮会弹出的窗口：



File In用于指定要统计的.cha目标文件。值得注意的是，如果File In里面选取目标文件，在点击Done按钮后，在前图的空白框栏中不会出现目标文件名，而是用一个@符代替。当然，用户也可以在前图空白栏中的freq后面手工输入.cha目标文件的名称。



Tiers是指对数据进行分层、分区或分类处理。在Tier Option界面中有一个预先填入*号的空白栏。用户可以在*后面填入在freq指令执行时需要筛选出来的数据层面。填入的代码必须是编码过程中在*号后面使用的缩写。例如，表示儿童的CHI，表示受试者的PAR等。



在Search Option中的选项，用于为freq指令进一步圈定特定的搜索范围，例如单词，甚至精确到特定的词性。

也可以用代码行来完成上述选项，以下介绍一些简单指令。其中所用例子是sample.cha，数据分层采用儿童数据，代码为*CHI。

请记住以下指令参数：

- ① +t 是查询分层tiers
- ② +s 是检查字符串
- ③ +f 是保存报表文件
- ④ ^ 是A^B，即B紧接着A

11.1 → dir

该指令显示的是工作文件夹里面的所有文件。执行结果会在报表界面显示如下内容：

```
> dir
media\                zho\
baorunjianTemplate.cha  peterduus01b1.cha
peterduus01b1.mor.cex  peterduus01b1.mor.pst.cex
peterduus01b1.ulx.cex

5 files, 2 directories
```

11.2 → mlu +t*CHI sample.cha 或 mlu -t%mor sample.cha

该代码分为三个部分。

- ① 指令名：mlu
- ② 参数：+t*CHI
- ③ 数据源：sample.cha

代码的三个部分之中，只有指令名必须第一个出现，参数和数据来源的顺序可以是任意的。笔者建议根据按照中文思维顺序编写指令代码，通常是指令名+数据源+参数。因此上述指令代码可以写为：`mlu sample.cha +t*CHI`。它的意思是：需要计算话语均长（mean length of utterance, MLU），数据源是sample.cha，数据层面是儿童数据。

默认的话语均长单位是每句话语所含语素，即语素型话语均长（mean length of utterance in morpheme, MLUm），可以简单地把它看作每句话所含字数。但是，如果.cha文件中没有%mor行，也就是说没有添加过语法标注，那么上述指令是无法给出结果的。笔者建议可用mor.cex或mor.pst.cex文件作为数据源。指令`mlu peterduus01b1.mor.cex +t*PAR`结果如下图：

```

> mlu peterduus01b1.mor.cex +t*PAR
mlu peterduus01b1.mor.cex +t*PAR
Thu Nov 19 12:50:29 2015
mlu (02-Nov-2015) is conducting analyses on:
ONLY dependent tiers matching: %MOR;
*****
From file <peterduus01b1.mor.cex>
MLU for Speaker: *PAR:
MLU (xxx, yyy and www are EXCLUDED from the utterance and morpheme counts):
  Number of: utterances = 9, morphemes = 21
  Ratio of morphemes over utterances = 2.333
  Standard deviation = 3.432
    
```

所用指令代码
指令执行时间
说明指令执行的条件
数据源
数据分层
凡是被标注为xxx yyy 和www的都被排除
有9条话语（句子或编码行），21个语素（字）
语素/话语的比值
标准差

如果数据源文件中确实没有%mor标记，但又确实想要分析该数据源，则需要把参数代码改成`-t%mor`。例如：`mlu peterduus01b1.cha -t%mor`。此时统计的不是语素型话语均长，而是单词型话语均长（mean length of utterance in words, MLUw）。

```

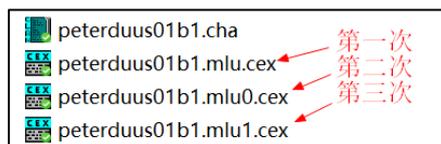
> mlu peterduus01b1.cha -t%mor
mlu peterduus01b1.cha -t%mor
Thu Nov 19 13:02:23 2015
mlu (02-Nov-2015) is conducting analyses on:
ALL speaker tiers
*****
From file <peterduus01b1.cha>
MLU for Speaker: *PAR:
MLU (xxx, yyy and www are EXCLUDED from the utterance and morpheme counts):
  Number of: utterances = 9, words = 21
  Ratio of words over utterances = 2.333
  Standard deviation = 3.432

MLU for Speaker: *INV:
MLU (xxx, yyy and www are EXCLUDED from the utterance and morpheme counts):
  Number of: utterances = 11, words = 92
  Ratio of words over utterances = 8.364
  Standard deviation = 3.699
    
```

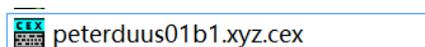
指令代码
如果没有指定分层参数+t*PAR或+t*INV 则会在所有层面都执行mlu指令

上述所有指令代码所输出的结果都显示在CLAN的临时报表文件中，如果想要把结果保存到工作文件夹中，则需要添加+f参数。例如：`mlu peterduus01b1.cha -t%mor +f`会在工作文件夹中生成一个报表文件：`peterduus01b1.mlu.cex`。请注意，这个报表文件的缀名.mlu.cex明确提示是什么指令产生它。

如果反复执行该指令，则会产生如下图所示报表文件序列：



也可以在+f后面添加一段字符串来指定结果文件的缀名。例如`mlu peterduus01b1.cha -t%mor +fxyz`，如下图所示：



11.3 → 通配符*（以 freq 指令为例）

星号用于模糊表示一个或多个真正的字符。*.cha表示所有的cha文件，CH*表示以CH开头的言说者，+s"*ing"表示所有以ing结尾的字符串。

以freq指令为例，freq peterduus01b1.cha +s"*个" 需要用词频指令计算peterduus01b.cha文件中所有以“个”结尾词的出现次数。结果如下图，其末尾给出了类符（different types of items, i.e. word form）和形符（token, an individual word）的比值，即类符/形符比（**type/token ratio, TTR**，又称形次比）。

```

> freq peterduus01b1.cha +s"*个"
freq peterduus01b1.cha +s"*个
Thu Nov 19 13:23:33 2015
freq (02-Nov-2015) is conducting analyses on:
ALL speaker tiers
-----
From file <peterduus01b1.cha>
Speaker: *INV:
1 一个

-----
1 Total number of different item types used
1 Total number of items (tokens)
1.000 Type/Token ratio

Speaker: *PAR:
1 一个
2 这个
2 那个

-----
3 Total number of different item types used
5 Total number of items (tokens)
0.600 Type/Token ratio

```

指令代码
执行时间
不分层，涵盖所有言说者
主试者产生的词
被试者产生的词

11.4 → 定向覆盖符> 和 定向添加符>>（以 freq 指令为例）

可以用redirection symble（定向符）来指定报表内容要输出到什么文件。该符号有三种用法。以下用freq指令为例，输出分析结果，将文件命名为freqAnalysis.txt。

- ① freq peterduus01b1.cha > freqAnalysis.txt 结果覆盖文件原有内容
 - ② freq peterduus01b1.cha >> freqAnalysis.txt 结果添加到文件末尾
 - ③ freq peterduus01b1.cha >& freqAnalysis.txt 结果和系统信息覆盖到文件中
- 笔者建议使用>>添加指令，以免误删原有文件内容。

11.5 → 重复（repetition）和回述（retracing）参数：+r6

freq等指令默认把复述和回述数据纳入统计范围。如果希望不计算复述和回述数据，需要添加+r6参数。例如：

*PAR: &哎 &ne 试 [/] 试试 [/] 试试看

如果计算freq不用+r6，则“试”、“试试”和“试试看”计为三个词。如果+r6，则不计算前两个，而只计入最后一个词“试试看”。

11.6 → 关键词指令：KWAL

关键词（key word）指令KWAL用于检索用户指定关键词的出现位置及其上下文。例如：
kwal +s试试看 +w2 -w2 peterduus01b1.cha 检索出现“试试看”一词的句子（话语行），并且

显示该位置上下各两行的语句。+w和-w就是上下行的意思，如果w后面不接数字，则只显示关键词所在行。结果如下图：

```

> kwal +s试试看 +w2 -w2 peterduus01b1.cha
kwal +s试试看 +w2 -w2 peterduus01b1.cha
Thu Nov 19 15:01:29 2015
kwal (02-Nov-2015) is conducting analyses on:
  ALL speaker tiers
*****
From file <peterduus01b1.cha>
-----
*** File "peterduus01b1.cha": line 17. Keyword: 试试看
*PAR: &ne &=head:shake 不行了!•
*INV: 那你讲话是很困难啦?•
*PAR: &哎 &ne &=left_hand:sagittal_plan_left_hand_helix 试 [/] 试试 [/] 试试看
      &=left_palm_poen!•
*INV: 试试看啊好的!•
*PAR: &=head:nod 对!•
-----
*** File "peterduus01b1.cha": line 19. Keyword: 试试看
*INV: 那你讲话是很困难啦?•
*PAR: &哎 &ne &=left_hand:sagittal_plan_left_hand_helix 试 [/] 试试 [/] 试试看
      &=left_palm_poen!•
*INV: 试试看啊好的!•
*PAR: &=head:nod 对!•
@G: cat rescue

```

在报表界面三连击***所在行，就会切换到原始文件中相应行上，并且深色显示该行。

kwal指令可以使用+r6来排除重复或回述内容。

在+s后面的关键词可以不用英文半角双引号或单引号来作为字符串标记，但如果使用引号作为标记，程序也可以识别。笔者建议最好还是使用引号来表示字符串。

在关键词中也可以使用通配符。例如"*看"可以检索到“看看”、“试试看”和“看”等以“看”字结尾的词。

11.7 合并字符检索指令：COMBO

COMBO指令的可以用Boolean型字符串进行检索。CLAN的布尔型字符串定义如下：

- ◆ A^B B紧随A
- ◆ A+B A或B
- ◆ A! B A非B
- ◆ * 通配符，任意单个或多个字符
- ◆ _ 通配符，任意单个字符
- ◆ \ 转义符。

转义符可以使用户检索诸如*和_这样的字符本身。例如，“看图_说话”这个复合词在检索的时候就需要检索下划线本身。字符串的编制思路可以参看正则表达式相关书籍。

以下举例： combo +s"试^试试" +w -w peterduus01b1.cha

```

> combo +s"试^试试" +w -w peterduus01b1.cha
试^试试
combo +s试^试试 +w -w peterduus01b1.cha
Thu Nov 19 15:41:16 2015
combo (02-Nov-2015) is conducting analyses on:
ALL speaker tiers
*****
From file <peterduus01b1.cha>
-----
*** File "peterduus01b1.cha": line 17.
*PAR: &哎 &ne &=left_hand:sagittal_plan_left_hand_helix (1)试 [/] (1)试试 [/] 试试看
      &=left_palm_poen ! •

Strings matched 1 times
    
```

指令代码 ← 试^试试
 所查字符串 ← 试

“试”紧接着是“试试”

combo +s" _^试" +w -w peterduus01b1.cha

```

> combo +s" _^试" +w -w peterduus01b1.cha
{_^试
combo +s _^试 +w -w peterduus01b1.cha
Thu Nov 19 15:46:47 2015
combo (02-Nov-2015) is conducting analyses on:
ALL speaker tiers
*****
From file <peterduus01b1.cha>
-----
*** File "peterduus01b1.cha": line 17.
*PAR: &哎 &ne (1)&=left_hand:sagittal_plan_left_hand_helix (1)试 [/] 试试 [/] 试试看
      &=left_palm_poen ! •

Strings matched 1 times
    
```

11.8 块检索指令和镶嵌符: gem 和 gemfreq

失语库数据的采集包含了多个任务块，例如汉语普通话失语库的任务块包括以下部分：

- 寒暄 (greeting)，引发谈话过程。
- I、看图说话 (picture_description)
 - 1、打破窗户 (window)
 - 2、拒绝雨伞 (umbrella)
 - 3、拯救小猫 (cat)
 - 4、洪水 (flood)
- II、命题故事 (story_narrative)
 - 1、龟兔赛跑 (tortoise_hare)
 - 2、狼来了 (cry_wolf)
- III、过程讲述 (procedural_discourse)
 - 火腿肠蛋炒饭 (egg_ham_fried_rice)
- IV、自由说话 (speech)
 - 1、重要事件 (important_event)
 - 2、讲述卒中过程 (stroke_story)
 - 3、讲述恢复过程 (stroke_coping)

注意：各个任务的英文名称需要确保完全与上表中括号内所示一致。

GEM和GEMFREQ指令的作用就是从一个较大的含有多个任务块的.cha文件中，提取特定的任务块，以供进一步分析。如果配合上述+f参数，可以生成并保存新的.cha文件，相当于从大文件中获取某些特定任务块的剪辑。

每一块会话过程在.cha文件里都以gem符（译称**镶嵌符**）来表示。它有两种形式：

① 简式镶嵌符@G:（又称懒人镶嵌符）

这种镶嵌符只需要在块的起始部位加入就行，检索时会默认以下一个简式镶嵌符作为这个块的结束以及下一块的开始。例如

- @G: greeting 表示“寒暄”任务块。
- @G: window 表示“打破窗户”任务块。
- @G: cat 表示“拯救小猫”任务块。

笔者建议这里采用下划线连接的复合词作为任务块名字。虽然可以用汉字来做任务块名称，但是为了今后与其他语种的标注结果进行比较，尽量还是使用英文名称。

如果用了简式镶嵌符，那么在使用gem指令时，需要添加参数+n。使用EVAL指令分析时，需将末尾+u改成+n。该参数会让gem指令从一个@G开始，到下一个@。笔者不建议使用简氏镶嵌符@G，因为在执行gem指令分析时容易发生错误。

② 复式镶嵌符@Bg:和@Eg:

分别放置在块的起止位置。复式镶嵌符的好处在于它可以分出嵌套的大块和小块。使用时要注意@Bg和@Eg所在行都要填入相同的快名称。

笔者提醒：复式镶嵌符的不能是bg或BG，一定要一大写一小写。否则会出现“Tier is not declared in the depfile file”。也就是说在C:\TalkBank\CLAN\lib文件夹下的depfile.cut中找不到相应的编码定义。如果在编码时发现类似提示，可以到depfile.cut中查一下标准的写法。

仍以前文所述peterduus01b1.cha文件为例，如果进行嵌套分块，就可以变成如下代码。其中cat是嵌套在picture-description中。

```

@Begin
@Languages: zho
@Participants: PAR peterduus01b Participant, INV WeihuaXIANG Investigator
@ID: zho|fengLIN|PAR|48;|male|TMA|Asian|Participant|15||
@ID: zho|fengLIN|INV|25;||Asian|Investigator|18||
@Media: peterduus01b1,video
@Date: 17-MAY-2014
@Bg: greeting
*INV: 好,现在我要. •0_7440•
*INV: 一会儿请你讲一些话给我听. •7440_9800•
*PAR: <好> [>]! •9800_10500•
*INV: <看看你讲话> [<] 的情况. •10500_11100•
*PAR: &=left_hand:wave 不行! •10800_11440•
*INV: 好,你现在觉得自己讲话讲得怎么样? •11440_14568•
*PAR: &ne &=head:shake 不行了! •14568_16447•
*INV: 那你讲话是很困难啦? •16500_19700•
*PAR: &ne &=left_hand:sagittal_plan_left_hand_helix 试试 [] 试试 [] 试试看
      &=left_palm_poen! •19577_22550•
*INV: 试试看啊? •22550_23800•
*PAR: <&=head:nod 对> [<] . •23500_24000•
*INV: <好的> [>] . •23800_24500•
@Eg: greeting
@Bg: picture_description
@Bg: cat
*INV: 我们,我待会儿要给你看一些<图片> [>] .
*PAR: <对> [<]! .
*INV: 然后你花一点点时间来看这些图片,然后它们讲了一个故事. .
*PAR: 故事. .
*INV: 你一边看一边讲给我听好不好? .
*PAR: 这个+... .
*INV: 故事要有开头经过和结尾,你可以一边看图一边跟我说. .
*PAR: &=left_hand_on_picture 他那个&ze
      &=left_hand_thum_index_finger_measuring 那个&yuxi 拔出来
      &=left_hand_palm_upturning &=left_hand_sagittal_plan_left_hand_helix
      &e 这&ke &guozi &ke 是&kuohe 是&wuwu &ka 他
      &=left_hand_index_finger_pointing_to_picture 有一个&kao 这个+... .
*INV: 我们停一下,哦,让我们停一下子. .
@Eg: cat
@Eg: picture_description
@End
    
```

以下是几种不同查询的结果：

greeting

```
> gem +s"greeting" +t*PAR peterduus01b1.cha
gem +sgreeting +t*PAR peterduus01b1.cha
Thu Nov 19 18:24:05 2015
gem (02-Nov-2015) is conducting analyses on:
  ONLY speaker main tiers matching: *PAR;
  and ONLY header tiers matching: @BG;; @EG;;
*****
From file <peterduus01b1.cha>
*** File "peterduus01b1.cha": line 9.
@Bg: greeting
*PAR: <好> [>] ! •
*PAR: &=left_hand:wave 不行 ! •
*PAR: &ne &=head:shake 不行了 ! •
*PAR: &哎 &ne &=left_hand:sagittal_plan_left_hand_helix 试 [/] 试试 [/] 试试看
      &=left_palm_poen ! •
*PAR: <&=head:nod 对> [<] . •
@Eg: greeting
@Eg: cat
@Eg: picture_description
```

picture_description

```
> gem +s"picture_description" +t*PAR peterduus01b1.cha
gem +spicture_description +t*PAR peterduus01b1.cha
Thu Nov 19 18:22:19 2015
gem (02-Nov-2015) is conducting analyses on:
  ONLY speaker main tiers matching: *PAR;
  and ONLY header tiers matching: @BG;; @EG;;
*****
From file <peterduus01b1.cha>
@Eg: greeting
*** File "peterduus01b1.cha": line 24.
@Bg: picture_description
@Bg: cat
*PAR: <对> [<] ! •
*PAR: 故事 . •
*PAR: 这个 +... •
*PAR: &=left_hand_on_picture 他 那个 &ze
      &=left_hand_thum_index_finger_measuring 那个 &yuxi 拔出来
      &=left_hand_palm_upturning &=left_hand:sagittal_plan_left_hand_helix
      &e 这 &ke &guozi &ke 是 &kuohe 是 &wuwu &ka 他
      &left_hand_index_finger_pointing_to_picture 有一个 &kao个 这个 +... •
@Eg: cat
@Eg: picture_description
```

常见问题

下载的媒体文件无法与.chat 文件关联怎么办？

如果收到警告“CLAN cannot locate the media file”。那么请按照以下顺序查找原因。

- ① 请确认你的电脑是不是Windows 7, 8.x, and 10. 系统。
- ② 请确认在CHAT译文内的@Media文件头是否指定了正确的媒体文件名。
- ③ 请确认媒体文件名和路径中不含任何空格或非英文文字。
- ④ 请确认.chat文件和媒体文件都保存在同一个文件夹中，或者媒体文件是否处于media文件夹中。
- ⑤ 请确认媒体文件格式，是否是.mp3、.mov等格式。如果不是，请用格式转换器。例如笔者使用的格式工厂。
- ⑥ 重启CLAN。
- ⑦ 重启电脑。
- ⑧ 重装CLAN。
- ⑨ 检查QuickTime浏览器是否有问题。
- ⑩ 检查电脑上的后缀文件名是不是默认隐藏的。如果后缀名被修改或删除，你可能需要手工添加。

如何鉴别话语单位？

会话单位（conversation unit, C-Unit，又称 T-Unit，也称为 utterance），以下三选二即可：

1. 沉默或停顿超过 2 秒。
2. 语调结束。
3. 在语法上形成一个完整的句子，或者词语已经可以形成完整的、适宜的表意，例如：
 - *妈妈： 你要去哪里啊？
 - *宝宝： 家。（虽然是词，但已经是一个话语单位utterance）。

话语重叠怎么办？ → [>] [<] +<

如果重叠部分与研究无关，例如临床对话时医生讲话而同时患者发出连续不断的无意义的“嗯啊”，那么可以不编码。如果重叠部分包含需要研究的内容，可以使用如下方式：

*INV: how did you communicate <with her> [>] ?

*PAR: <I just kept talking> [<] .

也可以插入空白表示重叠发生的位置，CLAN分析时不受影响。如：

*INV: how did you communicate <with her> [>] ?

*PAR: <I just kept talking> [<] .

最简单用法：是使用“+<”作为重叠标记，在话语的开头，来表示之前重叠话语，这表示重叠中的第二句话覆盖前一句重叠，但并没有明确表明是哪些词重叠。

*INV: how did you communicate with her ?

*PAR: +< I just kept talking .

如何处理重复语? → <> [/] [//] [x n]

首先要区分 **非纠正性复述**和**纠正性复述（修正）**。

如果受检者重复句子不做任何改变，使用“<>”重复的句子记录在三角括号里面，并紧跟着“[/]”。如果仅有一个单词被重复一次，那么不使用三角括号标记，直接使用“[/]”。如果当重复的是一些如“嗯，额” (e.g., uh, um)之类，不需使用上面的标记。

*PAR: <it was> [/] it was so bad.

*PAR: and the [/] the window was open.

*PAR: and she &s spill ed <the the the> [/] &uh &uh the water on the floor.

如果是带有纠正的复述（修正），当受检者在测试中修正句子，通常改变语法结构，但句子的含义不变，重新叙述的内容将被包括在三角括号内，其后紧跟“[//]”。用如下方式：

*CHI: <I wanted> [//] uh I thought I wanted to invite Margie

非纠正性复述和纠正性复述可以合用：

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

注意：在大多数统计指令中，重复语(retracing)都是默认被排除的，如果需要计入，则要特地使用+r6指令。

如果只是一个单词变化，就不需要三角括号标注。

*PAR: well <Cinderella was a> [//] &uh Cinderella is a nice girl.

*PAR: and then sometimes we [//] I was scared about the traffic.

如果某单词重复多次，可以在方括号内填入“x 空格 单词重复次数”：

*PAR: it's [x 4] um a dog.

语言中断问题? → +//. +//? +... ++

首先区分是自发言语中断、他人打断。

如果受试者话语是他人打断，使用+//.

*EXP: what did you +//.

*CHI: mommy .

*EXP: +, with your spoon .

如果受检者话语自发中断，又分为跳至下一项表达和停顿两种情况。

如果是跳至下一项表达，则陈述句使用+//. 疑问句使用+..?。

*PAR: well then the [//] &uh you know <the the> [/] the airplane that [//] that his +// .

*PAR: no the airplane that &uh landed +//.

*PAR: no [//] no that's not right.

当受检者停顿或不能说完一句话，使用“+...”表示话语的结束。

*PAR: but um the story goes on from um +...

*PAR: oh gosh.

发生于治疗师诱导患者产生言语时。使用+...标注前一句的中断，++标注下一句的产生及完成整个句子。可以提示性，也可以为非提示性。

*INV: how about ra + ...	*HEL: if Bill had known +...
*PAR: ++ a radio.	*WIN: ++ he would have come.

遇到不需要转录的部分怎么办? → www %exp

非转录部分 (Untranscribed material) 可以记录为 `www`, 表示这部分内容与分析无关。信息转化器不予转换的部分。需要另加一行 `%exp`, 用于注明非转录内容。

```
*PAR: www.
%exp: talking to spouse
*PAR: www.
%exp: looking through pictures
```

难以理解的部分怎么办? → xxx

难以理解的部分 (Unintelligible segments) 可以转写为: `xxx`。

会话者引用其他人的话怎么办? → +"/.

为了标注引用的材料, 使用特殊符号 `+"/.` 标记在引导句的末尾, 然后以 `+` 符号开始于下一行, 包含了一个完整的从句或句子的引用材料。

```
*PAR: and so the prince found the slipper and he said +"/.
```

```
*PAR: +" my gosh I can find the lady who is fitting this slipper.
```

如果引用的材料包括很多行, 那么在每一个引用句子的开头使用 “+” 符号。

如果引用的材料在主从句的前面, 引用前使用 “+”, 主从句结尾标注 “+.”。

```
*PAR: +" my gosh I can find the lady who is fitting this slipper.
```

```
*PAR: the prince said +".
```

错误语怎么标注? → [::] [:] @u

如果错误语是一个真词, 在方括号里标注两个冒号, 和一个空格;

如果错误语是一个不存在的词, 在方括号里标注 1 个冒号, 和一个空格。

这可以保证 CLAN 软件可以识别错误语。

如果错误语是一个不存在的词, 你可以按照国际音标标注, 或者拼写出来。如果使用国际音标标注, 使用 “@u” 标注在错误语后面。

```
*PAR: and the soo [ : two] [*] sisters went to the ball.
```

```
*PAR: no dubs [ : dogs] [*] allowed in the cemetery.
```

```
*PAR: the pints@u [ : prince] [*] want s to know who the slipper fits.
```

复合词怎么标注? → _

用下划线标注复合词。例如: `you_know`, 或者一个人的名字: 如: `Feng_LIN`

如何标注体态语? → &=bodyPart:behavior %fac %gpx

有多种识别体态语的方法。例如, 以 “&=” 作为前缀, 接着 “身体部位”, “冒号”, 再到 “具体动作及意义”。

可以在一行中既记录句子又记录手势语，还可以将两种手势语记录为一行。

- *PAR: &=head:shake .
 *PAR: &=hand:hello .
 *PAR: see you later &=ges:wave .
 *PAR: the woman &=imit:fishing fishing pole water &=casts:pole .

也可以使用“%fac”和“%gpx”分别编码面部表情和身体姿势，来表示一句话中长时间的停顿。

- *PAR: she was fish [/] fish.
 %gpx: raising her arm up and down

注意!!! 在标注体态语的冒号后面，不能有空格，否则会把表示体态语的描述，当作句子里面的词语进行分析。

&的其他用法→填补词、语音片段、&=laughs

填补词，如 额，嗯等，也作为话语的一部分，但记录中要加“&”前缀。虽然在 CHAT 手册中的“Interactive and Communicative Markers”小节中会有一张关于此类填补词的表单，但是手册并不能做到时时更新。最好是在 MOR 词库 (MOR lexicon) 中查找正确的拼写方法。

- *INV: how do you think your language is these days?
 *PAR: well &uh &uh pretty good.

不完整片段（语音学上）：记录中要在字母前加“&”前缀，来描述产生的发音。因此，对于任何不完整词语，请先输入&，然后输入可以表示其发音的字符。

- *PAR: so now I can &sp speak a little bit.
 *PAR: and then &sh &s &w we came home.

当说话中有大笑或叹气，可以记录为：

- *PAR: well &=laughs tell you the truth, I can't say what I said.
 *PAR: &=laughs.

@的用法→@n

话语中的新词，转写中紧跟新词后面加“@n”后缀，作为新词的标记。

- *PAR: o h yes, this is a little sakov@n that's all.

如何标注被省略的词、缩略语和连读：0 ()

省略词 (Omitted Words)：以“0”标记，标注在实际单词的前面或被省略部分的前面。

- *PAR: I 0v kidding.
 *PAR: it was 0art same thing.

缩略语：省略字母的发音，如将“running”说成“runnin”。记录中可记录为 runnin(g)。再如 (be)cause, prob(ab)ly, (a)bout, (re)member, (ex)cept。

连读 (Assimilations)：“gonna” and “kinda”。这些可以被 CLAN 识别，所以无需替换。

笔者建议：根据汉语书写系统缺少形变的特点，省略语和缩略语标注与否取决于研究者更关注语音还是更关注语义。例如“不要”如果标注成语音，则连读的话可以是“表”的谐音，可以记录 biao。但是从汉语语义研究来看，仍旧标为“不要”。

如何编辑停顿? → (.) (..) (...) (min)

“(.)”表示句子未完成的停顿，“(..)”表示较长的停顿，“(...)”表示相当长时间的停顿。以小括号内句号表示停顿的时间长短。

*PAR: 我 不 (.) 知道 .

*PAR: 你 是 (...) 怎么 想 的 ?

如果想记录得更准确, 可以将具体停顿时间记录在括号内。例如 (.5) 表示停顿 0.5 分钟。

*PAR: 我 (0.5) 不 知 道 .

*PAR: 你 是 (1:13.41) 怎么 想 的 ?

声音效果或无声 → [=!]

使用[=!]来表示伴有表意作用的声音效果, 在感叹号后写下对音效的描述。

*MOT: you're just the cutest [=! kissing noises].

如果什么语音都没有, 则用以下表示沉默 (但仍有表意):

*MOT: 0 [=! kissing noises].

或者表示无语音但有动作 (例如玩具等) 引发的声音

*MOT: 0 [=! toys rustling]

长时间沉默无声: *MOT: 0 [=! silence].

注意, 用零来表示没有说什么, 但又有可表达的意思在传递。

注意: 是零, 不是 0。

胡言乱语或杂乱语的标注 →

用 yyy 表示会试着猜测用什么样的词, 用 xxx 表示没有可猜测的地方。

*CHI: yyy [=! dada].

*CHI: xxx [=! vocalizes/laughs/whines, etc].

百分号的作用 → %com: TAB

%com 用于注释, 即制作.cha 文件的研究者可以在后面写下一些说明文字, 以供人读懂代码。单独一行, 后面要有冒号, 然后要有制表符 (注意, 不是空白格!!!)