

kwal (03-Jui-2009) is conducting analyses on:
ALL speaker tiers

Ratio of morphemes over utterances

CHILDES

CHILDES

Child Language Data Exchange System



Transcription CHAT et
programmes CLAN

*MOT: is the bunny rabbit jumping

Florence Chenu

Dynamique du Langage

Plan :

- Présentation générale
- Transcription
- Réglages
- Codages
- Analyses



CHILDES

CHILDES

Child Language Data Exchange System



Présentation Générale

Introduction

➤ Analyse linguistique et TAL

➤ Logiciels d'analyses linguistiques

- Types de corpus
- Types d'analyses
- Exemples : **Hyperbase**, **BDLEX**, **Lexique**, **SCP**, **Concordance**, **Winsnoori**, **Praat**, **Nooj**, **Shoebox...**



Philosophie CHILDES

- Libre
- À condition de contribuer



Researchers who use TalkBank or CHILDES data and programs are asked to abide by these rules:

- The use of TalkBank data is governed by the [Gnu Public License \(GPL\)](#).
- You should read and follow the TalkBank [Code of Ethics](#).
- You should become a member of the TalkBank system by contacting [Brian MacWhinney](#) (macw@cmu.edu) and sending him your academic affiliation, postal address, and email address. Also, please describe how you plan to use the data and whether you will eventually be able to make your own contribution to the database.
- We maintain separate membership lists and datasets by interest groups. So, when joining, please select your major area of interest from these options:
 - CHILDES-- child language acquisition
 - LIDES-- bilingualism and code-switching
 - Gesture
 - AphasiaBank
 - Linguistic Exploration
 - Text and Discourse, including Conversation Analysis
 - Classroom Discourse
 - Animal Communication
- If at all possible, we would like to encourage you to subscribe to one or more of the TalkBank GoogleGroups mailing lists. You can find information on how to subscribe [here](#).
- All articles using TalkBank or CHILDES corpora should cite the references for those corpora as listed in the documentation. **This is crucial!**
- In addition,
 - articles based on the use of the CHILDES database should cite: MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates.
 - articles based on the use of the TalkBank database should cite: MacWhinney, B. (2007). The TalkBank Project. In J. C. Beal, K. P. Corrigan & H. L. Moisl (Eds.), *Creating and Digitizing Language Corpora: Synchronic Databases, Vol. 1*. Houndmills: Palgrave-Macmillan.
- Please read the relevant sections of the manual for the data you use. These sections describe important aspects of the data that may impact on conclusions based on them.
- If you collect new corpora, please contribute them to TalkBank or CHILDES and urge your colleagues to contribute new corpora. This is key to the growth of the database.
- Faculty members who abide by the above rules are welcome to use the programs and the non-password-protected database freely in their classes and teaching. The programs and data can be freely installed on university computer systems for free use by students, as long as the responsible faculty person is a member of the CHILDES or TalkBank user groups. However, TalkBank or CHILDES files that are password protected cannot be used without explicit permission from the contributor. Also, because they were student projects, the files in the /Conversation/CMU directory can be used for teaching purposes, but not for research purposes.



Télécharger et installer CLAN

Sur <http://childes.psy.cmu.edu/>

CHILDES - Child Language Data Exchange System

CHILDES is the child language component of the [TalkBank](#) system. TalkBank is a system for sharing and studying conversational interactions.

System

- [Ground rules](#)
- [Guidelines for Contributors](#)
- [Overviews and Introductions](#)
- [Membership list](#)
- How to subscribe to [Mailing Lists](#)

Links

- The [TalkBank](#) Database
- [IASCL](#) information
- [Other Child Language sites](#)
- [Research based on CHILDES](#)
- [Related Software](#)

Programs and Database

- Downloadable [Database](#)
- Browsable [Database](#)
- The [CLAN Program](#)
- [WebCLAN](#)
- [Training Videos](#)

Manuals

- [CHAT Transcription](#)
- [CLAN Programs](#)
- [Database Manuals](#)
- [BTS sign transcription system](#)

Windows	Mac	Unix
CLANWin	CLAN	UnixCLAN
V 30-Aug-2011 16:00	V 30-Aug-2011 16:00	V 30-Aug-2011 16:00

Vérifier régulièrement qu'on dispose d'une version récente

Help

- About Clan...
- Commands
- Save Them

About Clan

V 07-Mar-2008 15:00

Close


For help on CLAN please download the manual from <http://childes.psy.cmu.edu>. Send specific questions to macw@cmu.edu. c 1991-2006

Se reporter aux manuels lorsqu'on n'est pas sûr



Composants

- Transcription (CHAT)
- Analyses (CLAN)
- Banque de données
 - 30 langues (danois, français, allemand, grec, hébreux, hongrois, tamil, turc, etc. mais principalement anglais)
 - Corpus transcrits et analysés enfants (développement normal et troubles du langage), adultes aphasiques, bilingues, apprenants L2.
- Manuels
- Compléments (gram)
- Listes de diffusion
 - info-childes@googlegroups.com
 - chibolts@googlegroups.com

CHILDES Child Language Data Exchange System 

CHILDES is the child language component of the [TalkBank](#) system. TalkBank is a system for sharing and studying conversational interactions.

System	Programs and Database
Ground rules	Downloadable Database
Guidelines for Contributors	Browsable Database
Overviews and Introductions	The CLAN Program
Membership list	WebCLAN
How to subscribe to Mailing Lists	Training Videos
Links	Manuals
The TalkBank Database	CHAT Transcription
IASCL information	CLAN Programs
Other Child Language sites	Database Manuals
Research based on CHILDES	BTS sign transcription system
Related Software	
Phonology and Fonts	Teaching with CHILDES
Phon & PhonBank	Topics in language acquisition.
Unicode and IPA for Mac	Teaching Tips and Resources .
Unicode and IPA for Windows	Child Language Bibliographies
Special Procedures	Morphology and Lexicon
Procedures and tools for CA analysis	Part of Speech Analysis by MOR
Working with digitized video	Parental frequency count: view , download , documentation
Working with digitized audio	MRC lexical dictionary
The Computerized Comprehension Task	Syntactic analysis by GRASP



Environnement CHILDES

The screenshot displays the CLAN software interface. The main window, titled "Clan - CLAN Output", contains a text editor with the following content:

```
FCorpus-4.mor.txt
%typ: quotation
%suj: kick
%add: OP
*PDF: hiiiiiii@gb .
%mor:
%lan: ANG-COR
%t
%a
*P
1 är
%m
28 ça
%l
1 éck
%t
1 école
%a
1 écoutait
*P
1 ég
1 élève
1 épée
1 étais
3 était
1 étant
1 étudier
2 été
4 être

-----
997 Total number of different word types used
2855 Total number of words (tokens)
0.349 Type/Token ratio

>
CLAN [E][TEXT] * 1008
```

Overlaid on this window are two smaller windows:

- Commands**: A window with a list of paths: "working" (D:\Travaux de la semaine\tests), "output" (C:\NEWCHILDES\FRA\), "lib" (C:\NEWCHILDES\FRA\), and "mor lib" (C:\NEWCHILDES\FRA\). It also features a "CLAN" icon, a "FILE IN" icon, a "Help" button, and a "Run" button.
- Results**: A window showing the output of a command: "freq fra" and "freq fcorpus-4.mor.txt".

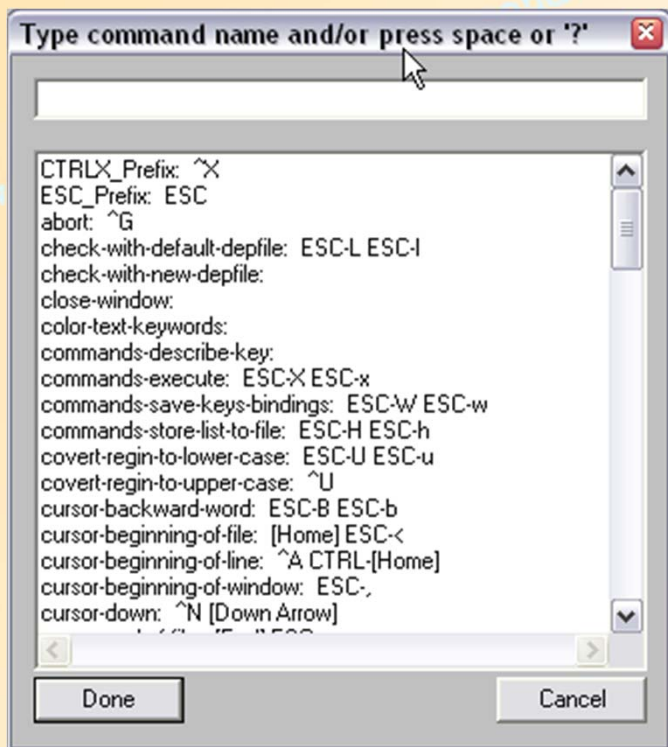
Orange circles and lines highlight the "FCorpus-4.mor.txt" window, the "Commands" window, and the "Results" window, with labels on the right side of the slide.

Fenêtre corpus

Fenêtre commande

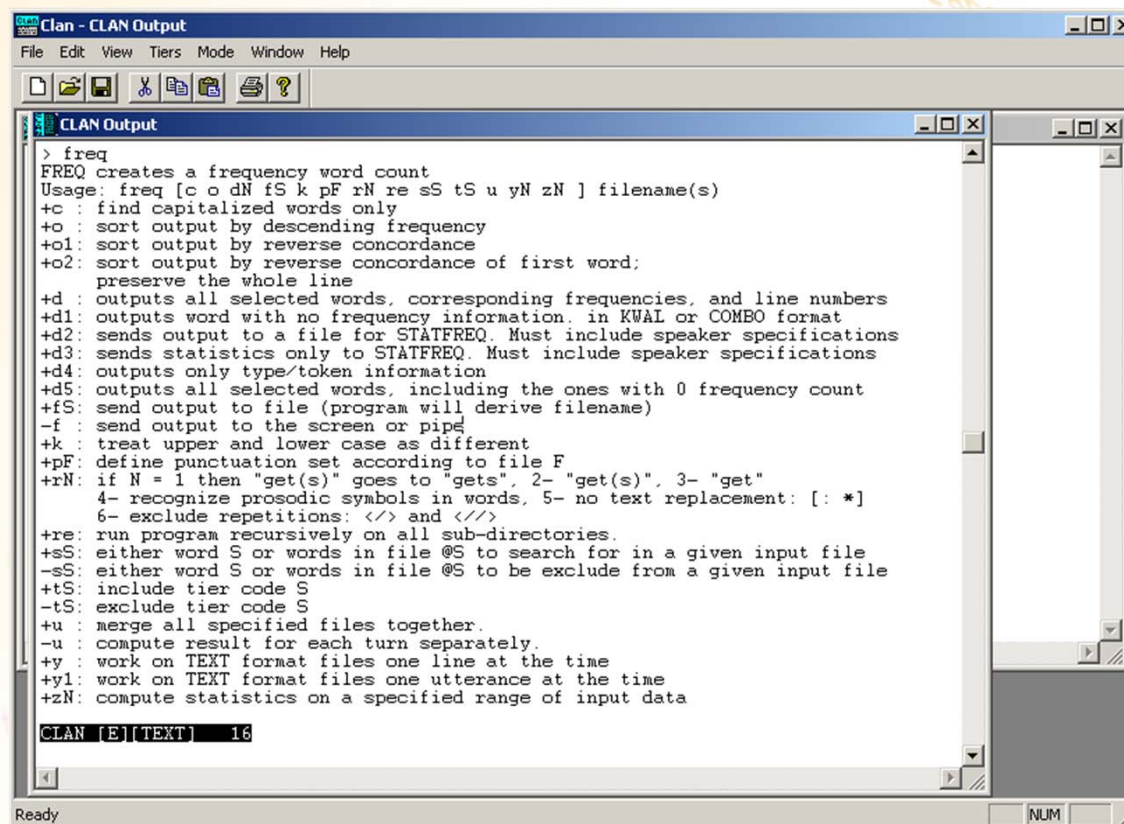
Fenêtre Résultats

Modules d'Aide

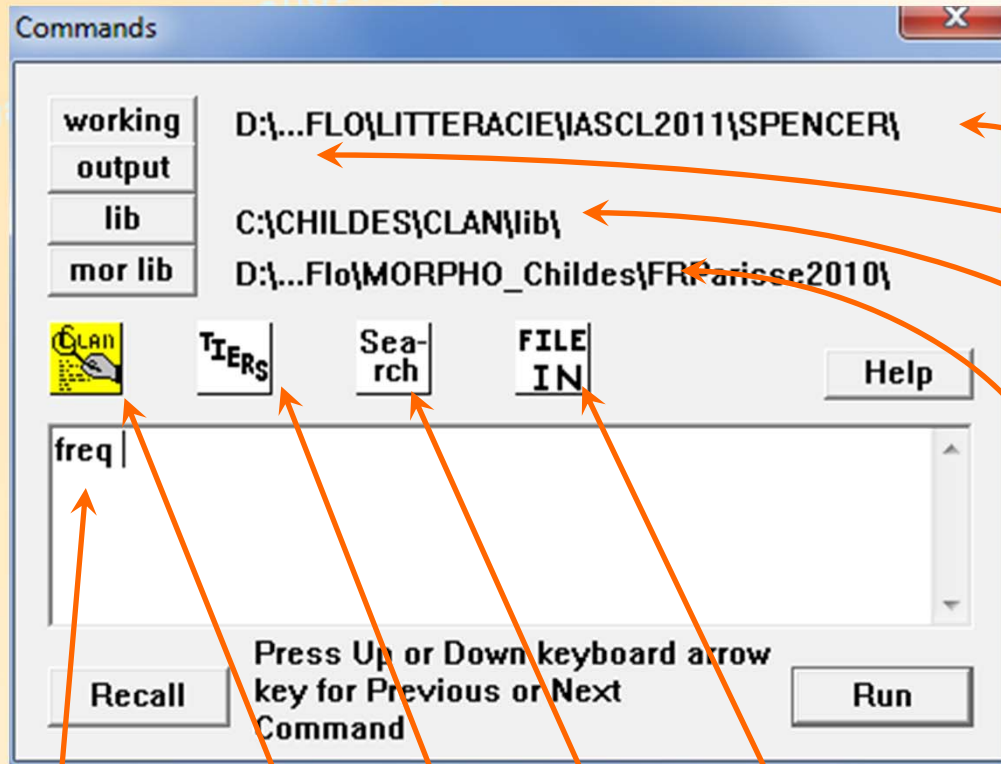


Raccourcis clavier

Aide sur les commandes



Module de Commandes



Commencer par indiquer dans quels dossiers Clan doit chercher/placer les fichiers pour travailler :

1. Dossier contenant les fichiers à analyser (dossier de travail)
2. Dossier qui contiendra les fichiers résultats d'analyse (si on veut qu'ils soient déposés ailleurs que dans le dossier de travail)
3. Dossier contenant les fichiers de configuration de Clan
4. Dossier contenant les fichiers nécessaires au fonctionnement de la grammaire

Sélectionneur de commandes
 Sélectionneur de fichiers
 Chaines à rechercher
 Zone de commandes
 Sélectionneur de lignes dépendantes



CHILDES

kwai (02-001-2009) is conducting analyses on:
ALL speaker diaries

Ratio of morphemes over utterances

CHILDES

Child Language Data Exchange System



Réglages

*MOT: is the bunny rabbit jumping ?

[>]

Outils et Préférences

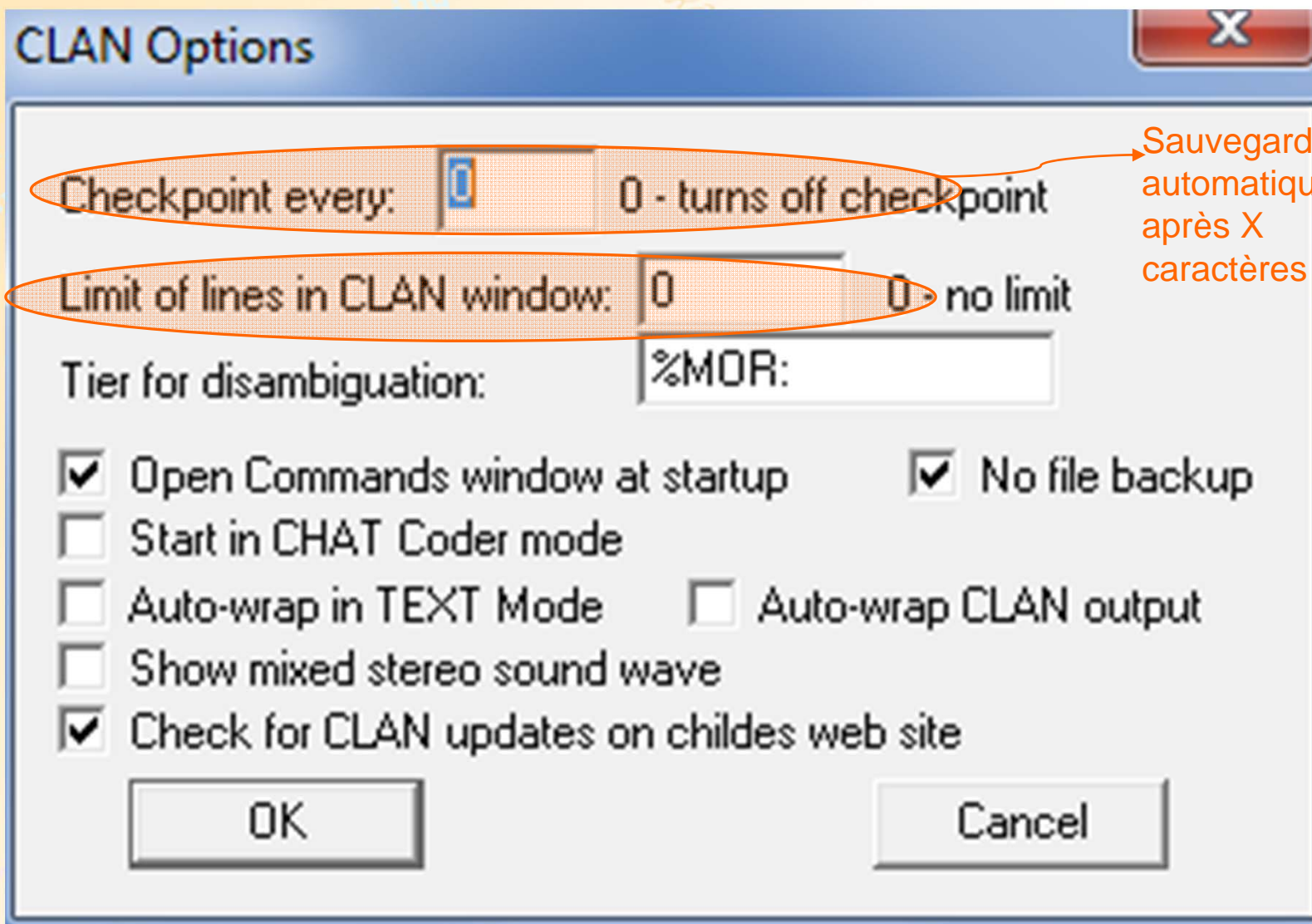
CLAN-
test

➤ Outils ergonomiques (modes d'affichage)

- Numéros de lignes (LINES – **Edit/define line numbers** ; **Mode/show line numbers**)
- Transcription simplifiée (FLO)
- Sélection de lignes (principale/dépendantes) à afficher (**Mode/Hide tiers**)

➤ Préférences





CHILDES

kwal (02-001-2009) is conducting analyses on:
ALL speaker diaries

Ratio of morphemes over utterances

CHILDES

Child Language Data Exchange System



Transcription

*MOT: is the bunny rabbit jumping ?

[>]

Transcription : Chat

Types de transcription

- Corpus oraux :
 - Phonétique : UNIBET, PHONASCII, SAMPA, IPA (cf Font IPA)
 - Orthographique

- Corpus écrits (ex. Projet Spencer, ReFlex)
 - Consignes de transcription pour l'analyse des erreurs

- Corpus Langue des signes (BTS)
 - Ex.
 - *MOT: COWBOY (mount)-cl'VP-cl'TL-gol'VP_TOP-pst'STR .
 - %gls: the cowboy got up on the horse's back



➤ Ex. Corpus Spencer (Viguié-Simon 2001)

➤ Oral

- *SBJ: donc je suis là.
- *SBJ: pour vous présenter les actions à mener envers vos camarades.
- *SBJ: pour euh une euh meilleure cohabitation euh dans la classe.
- *SBJ: donc euh tout d'+abord euh il faut être euh régulier.

➤ Écrit

- *SBJ: nous allons aborder les problèmes.
- *SBJ: que nous pouvons tous rencontrer.
- *SBJ: qui sont inhérents à la vie en société.
- *SBJ: qu' il s' agisse de violence orale ou verbale de rejet.
- *SBJ: mais vous pouvez certainement compléter la liste.



Table 17: Errors in Writing

Function	Example	Coding
omitted apostrophe	dont	don0't
blank for apostrophe	don t	don_0't
apostrophe for blank	will'not	will*' /not
extra apostrophe	wan't	wan*'t
extra punctuation	extra comma	*,
omitted punctuation	missing comma	0,

Extrait de Table 17 Manuel Childes p. 119



analyses on:
Table 18: SAMPA Vowels

IPA	SAMPA	ASCII	Openness	Front-Back	Round
a	a	97	open	front	-
ɑ	A	65	near-open	back	-
æ	{	123	near-open	central	-
ɒ	Q	81	open	back	+
ɔ	O	79	open-mid	back	+
e	e	101	close-mid	front	-
ɛ	E	69	open-mid	front	-
ə	@	64	mid	central	-
ɜ	3	51	mid	central	-

kwai
Extrait de Table 18 Manuel Childes p. 121



Exemple de transcription

@Begin
@Languages: fr
@Participants: CHI [redacted] Target Child, MOT Mother, VAL Valérie Investigator
@ID: fr|lyon|CHI|1;1.11|female||Target_Child||
@Date of birth of CHI: 28 FEB 2001
@Age of CHI: 1;1.11
@Sex of CHI: female
@File name: MAS02VR
@Recording date: 09 APR 2002
@Recorder: VR
@Transcriber: VR
@Checker: GH
@Corpus: Lyon

***MOT:** mettre le pantalon et les chaussures.
***MAS:** tiens.
%pho: /tja/
%act: tend son doudou
***MOT:** merci.
***MAS:** yy.
%pho: /pe/
***MOT:** ouais ça c'est joli.
***MOT:** ## tu sais on pourra jouer après # pa(r)ce+que c'est pas très pratique.
***MAS:** yyy.
%pho: /ka ka:/
***MOT:** c'est lui que tu veux?
%act: donne le doudou à MAS

@End



CHAT

En-têtes (headers):

- Commencent par le symbole @
- Informations constantes, valables pour tout le fichier et utiles pour l'utilisation de la transcription (@Begin, @End, @Languages, @Participants, @ID, ...)

Lignes principales:

- Commencent par le symbole * suivi du code locuteur
- Chaque ligne principale ne contient qu'un seul énoncé
- L'énoncé doit être terminé par un signe de ponctuation (. ! ?)
 - ✦ Exemple:
 - *MOT: oui là tu peux aller.
 - *MOT: # mais les livres # tu arrêtes.

Ligne(s) dépendante(s):

- Commencent par le symbole %
- Informations associées (commentaires, descriptions, codages) suivent immédiatement la ligne principale
- Beaucoup sont prédéfinies par CHILDES: %act, %sit, %com
- Système de codage systématique pour la phonologie (%pho), la morphologie (%mor) et la syntaxe (%syn)



En-têtes (Headers)

➤ Obligatoires :

- @Begin, @End
- @Languages: ⇒ Code_langue₁, Code_langue₂, etc.
 - français=fra; anglais=eng; japonais=jpn, etc.
- @Participants: ⇒ SpeakerID₁ Nom₁ Role₁, SpeakerID₂ Nom₂ Role₂, etc.
 - Target_Child, Child, Mother, Father, Brother, Sister, Sibling, Grandmother, Grandfather, Aunt, Uncle, Cousin, Family_Friend, Playmate, Visitor, Student, Teacher, Investigator, Examiner, Observer, Camera_Operator, Doctor, Nurse, Patient, Client, Subject, Unidentified, Adult, Teenager, Non_Human, OffScript, and Narrator.
 - Child, Adult, Teenager
- @ID: ⇒ language|corpus|code|age|sex|group|SES|role|education|
- @Media: ⇒ filename, type
 - Extension pas obligatoire
 - sound ou movie



En-têtes (Headers)

➤ Autres en-têtes constantes :

- **@Birth of XXX:** date de naissance de XXX
- **@Coder:** nom du transcripteur/codeur
- **@Language of XXX:** les langues de chaque locuteur
- **@Font:** police par défaut
- **@Warning:** pour marquer les défauts du fichier
- Etc.



En-têtes (Headers)

➤ En-têtes relatives au contexte :

- **@Activities:** activités
- **@Bg** and **@Bg:** marqueur de début de bloc
- **@Eg** and **@Eg:** marqueur de fin de bloc
- **@G:** marqueur de bloc pour les fainéants
- **@Bck:** info d'arrière plan
- **@Blank** pour les lignes blanches dans les textes écrits
- **@Comment:** commentaires
- **@Date:** date de l'interaction
- **@Location:** localisation géographique
- **@New Episode:** début d'un nouvel épisode
- **@New Language:** point pour nouvelle langue (shift)
- **@Room Layout:** configuration de la scène
- **@Situation:** atmosphère générale de l'interaction
- **@Tape Location:** repère sur l'enregistrement
- **@Time Duration:** heure de début et de fin
- **@Time Start:** heure de fin

Transcription

➤ Quel type de données ?

- Oral : orthographique aménagée + API
- Écrit: orthographique aménagée + spécificités pour l'écrit
- Autres (langues des signes)

➤ Lignes principales :

*Code_locuteur: ⇒ *texte* marqueur_de_fin_d'énoncé



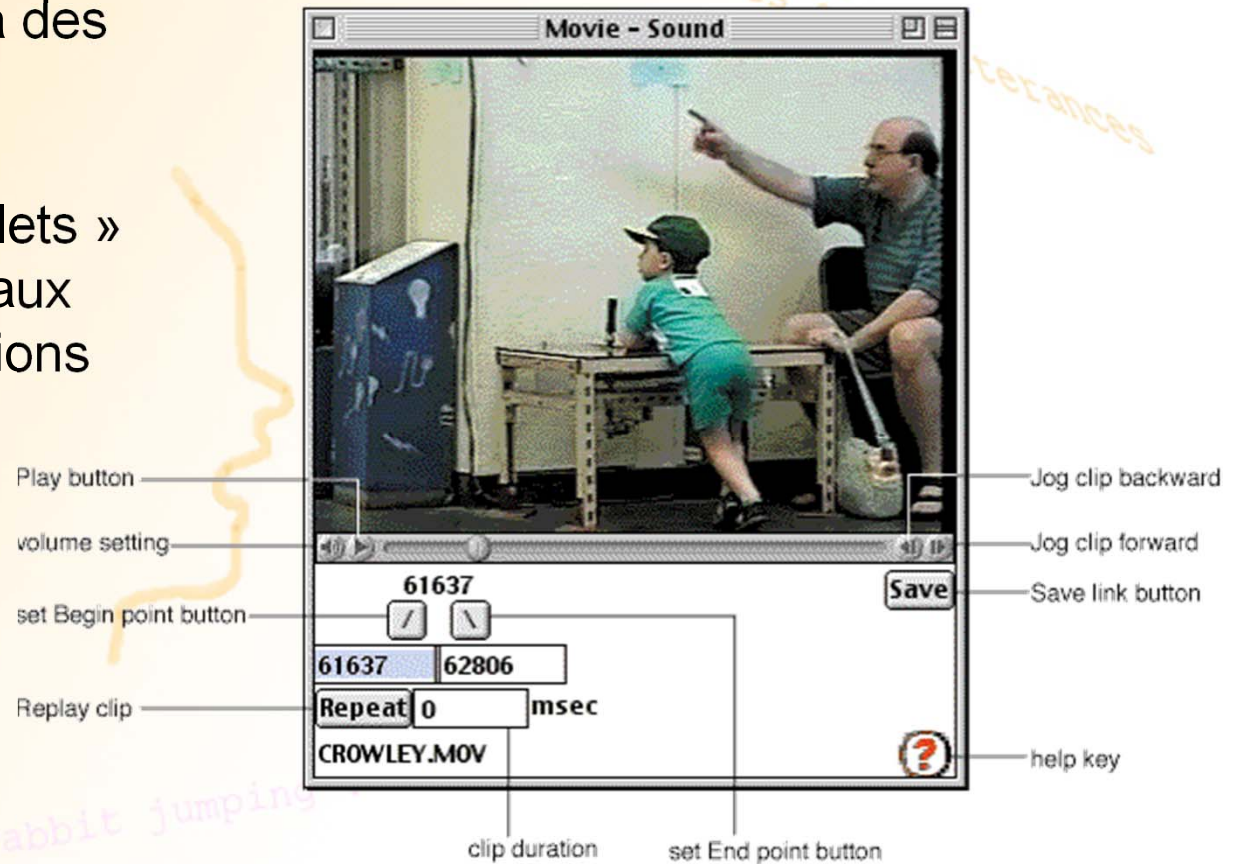
Ligne principale

- Toute séquence de caractères séparée par des espaces est considérée comme un mot
 - Mettre une espace après les apostrophes
 - Rassembler les mots composés par des +
- Les majuscules ne sont pas autorisées à l'exception des noms propres
- Les nombres ne sont pas autorisés
- On peut indiquer le découpage en morphèmes avec des tirets



➤ Liens audio/vidéo

- Relier les transcriptions à des fichiers vidéos numérisés
- Créer des « bullets » correspondant aux différentes sections sélectionnées



Ajout d'un en-tête

- Lorsqu'on souhaite lier un fichier média à une transcription, la première chose à faire est d'insérer une ligne d'en-tête « @Media »
- **@Media: nom_du_fichier, type_de_media**
 - Le premier champ est le nom du fichier média. Il n'est pas nécessaire d'indiquer l'extension du fichier.
 - Chaque transcription ne peut être associée qu'à un seul fichier media
 - Le second champ indique si le media est audio ou video.



Liens vidéo : méthode 1

- Ouvrir le fichier de transcription CLAN.
- Dans le menu **Mode**, sélectionner **Transcribe Sound or Movie (F5)**.
- Localiser la vidéo. Une fenêtre Quicktime va s'ouvrir permettant de visualiser la vidéo. La première portion du texte se met en surbrillance.
- Alors que la vidéo se joue, on peut définir des « bullets » (liens vers des parties de vidéo) en appuyant sur la touche espace.



Liens vidéo : méthode 2 (modifs)

- Ouvrir le fichier de transcription CLAN.
- Dans le menu **Mode**, sélectionner **Transcribe Sound or Movie (F5)**.
- Sélectionner le fichier vidéo
- Cliquer dans la transcription.
- Placer le curseur sur une ligne de transcription pour laquelle on veut une *bullet*.
- Jouer la vidéo jusqu'au début du passage à intégrer dans la *bullet*.
- Cliquer sur le bouton d'initialisation de bullet
- Jouer la vidéo jusqu'à la fin du passage à intégrer dans la *bullet*.
- Cliquer sur le bouton d'initialisation de fin de *bullet*
- Cliquer sur **save**
- Une *bullet* s'installe à la place du curseur



Liens audio

- Ouvrir le fichier de transcription CLAN.
- Dans le menu **Mode**, sélectionner **Transcribe Sound or Movie (F5)**.
- Localiser le son. Le signal s'affiche. Alors que le son se joue, on peut définir des « bullets » (liens vers des parties de son) en appuyant sur la touche espace.



Liens audio : méthode 2 (modifs)

- Ouvrir le fichier de transcription CLAN.
- Dans le menu **Mode**, sélectionner **Transcribe Sound or Movie (F5)/sonic mode**.
- Sélectionner le fichier son
- Cliquer dans la transcription.
- Placer le curseur sur une ligne de transcription pour laquelle on veut une *bullet*.
- Sélectionner la partie de signal adéquate, puis cliquer sur S.

- Modifs par triple clic PUIS modification des bornes par Shift+clic suivi de clic sur S



CHILDES

CHILDES

Child Language Data Exchange System



Exercice Transcription/liens

➤ Codage adaptable suivant le type de corpus et/ou le type d'analyses

➤ *Special form markers* sur la ligne principale

- Ex. données multilingues, néologismes, vocabulaire enfantin, etc.

➤ *Scoped Symbols* sur la ligne principale

- Marques temporelles audio et video
- Codes paralinguistiques (toux, rires, cris, accentuation...)
- Explications (éléments déictiques, traductions, formes standard...) et alternatives (doutes de transcription).
- Reprises et chevauchements

➤ *Postcodes* sur la ligne principale

- Inclusion exclusion de lignes principales pour les analyses
- Forcer à considérer un énoncé comme un tour.

➤ Codes lignes dépendantes (type et contenu)

- Codage préétabli pour erreurs, actes de langage, morphologie et syntaxe, phonologie, commentaires.
- Codage modifiable.
- Pas de restriction de création d'autres types de lignes dépendantes



Table 2: Special Form Markers

Letters	Categories	Example	Meaning
@b	babbling	abame@b	-
@c	child-invented form	gumma@c	sticky
@d	dialect form	younz@d	you
@f	family-specific form	bunko@f	broken
@fp	filled pause	huh@fp	-
@fs	filler syllable	n@fs	in? one? and?
@g	general special form	gongga@g	-
@i	interjection, interaction	uhhuh@i	-
@k	multiple letters	ka@k	Japanese "ka"
@l	letter	b@l	letter b
@m	Motherese	baba@m	bottle
@n	neologism	breaked@n	broke
@o	onomatopoeia	woofwoof@o	dog barking
@p	phonol. consistent form	aga@p	-
@pm	protomorpheme	wi@pm	will?
@pr	phrasal repetition	it's a, it's a@pr	for disfluency work
@q	metalinguistic use	no if@q-s or but@q-s	when citing words
@s:x	second-language form	istenem@s:h	Hungarian word
@si	singing	lalala@si	singing
@sl	signed language	apple@sl	apple
@sas	sign & speech	apple@sas	apple and sign
@t	test word	wug@t	small creature
@u	Unibet transcription	binga@u	-
@wp	word play	goobarumba@wp	-
@x	Excluded words	stuff@x	excluded

15.10 Scoped Symbols

·%mov:"*" _0_1073	time alignment marker
[=! text]	paralinguistics, prosodics
[!]	stressing
[!!]	contrastive stressing
["	quotation marks
[: text]	replacement
[= text]	explanation
[=? text]	alternative transcription
[% text]	comment on main line
[\$text]	code on main tier
[?]	best guess
[>]	overlap follows
[<]	overlap precedes
[/]	retracing without correction
word [x N]	word repetition
[//]	retracing with correction
[///]	retracing with reformulation
[/-]	false start without retracing
[/?]	unclear retrace type
[*]	error marking
[+ text]	postcode
[- text]	precode
[+ bck]	excluded utterance
[+ trn]	included utterance

ratio of morphemes over utterances



Postcodes & Precodes

➤ Postcodes [+ text]

- *CHI: not this one. [+ neg] [+ req] [+ inc]

➤ Precodes [- text]

*NOT: is the bunny rabbit jumping ?



Marqueurs de fin d'énoncé

15.9 Special Utterance Terminators

+...	trailing off
+..?	trailing off of a question
+!?	question with exclamation
+/.	interruption
+/?	interruption of a question
+//.	self-interruption
+//?	self-interruption of a question
+"/.	quotation follows on next line
+".	quotation precedes
+"	quoted utterance follows
+^	quick uptake
+<	"lazy" overlap marking
+.	utterance broken for transcription
+,	self-completion
++	other-completion
++.	CA missing final delimiter
+.=	CA begin latch delimiter
[^c]	clause delimiter



CHILDES

CHILDES

Child Language Data Exchange System



Valider la transcription

Vérifications

➤ Commande Check

- Permet d'obtenir la liste des erreurs
- **Mode/Check Opened File (esc I)**

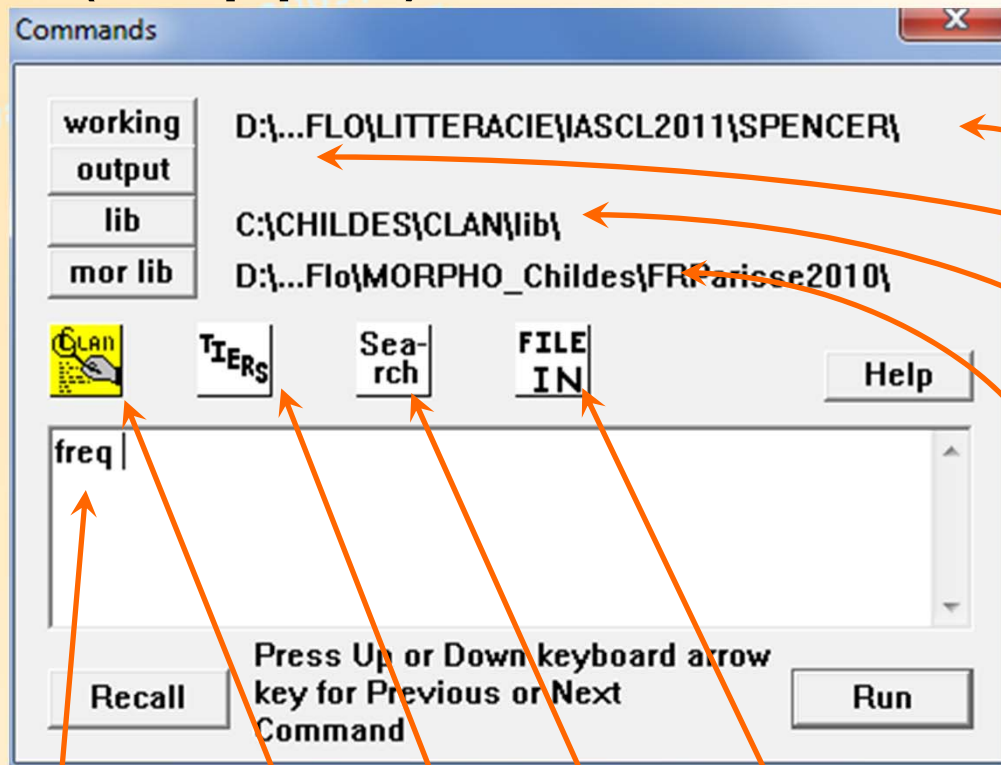
OU

➤ Module de Commandes, check (options possibles)

- Ne pas considérer certaines erreurs
- Ne considérer que certaines lignes, etc.



Module de Commandes (Rappel)



Commencer par indiquer dans quels dossiers Clan doit chercher/placer les fichiers pour travailler :

1. Dossier contenant les fichiers à analyser (dossier de travail)
2. Dossier qui contiendra les fichiers résultats d'analyse (si on veut qu'ils soient déposés ailleurs que dans le dossier de travail)
3. Dossier contenant les fichiers de configuration de Clan
4. Dossier contenant les fichiers nécessaires au fonctionnement de la grammaire

Sélectionneur de commandes
 Sélectionneur de fichiers
 Chaines à rechercher
 Zone de commandes
 Sélectionneur de lignes dépendantes



Options check (module de commandes)

- +c : check bullets consistency
- +eN: report only error(s) number N
- +e : list numbers for every error
- -eN: report all error(s) except number N
- +g1: recognize prosodic delimiter
- -g1: ignore prosodic delimiter (default)
- +g2: check for "CHI Target_Child"
- -g2: do not checking for "CHI Target_Child" (default)
- +g3: check words for upper case letters, numbers and illegal apostrophes
- -g3: do not check words in more details (default)
- +g4: check file(s) for missing @ID tiers
- -g4: do not check file(s) for missing @ID tiers (default)
- +d : attempts to suppress repeated warnings of the same type
- +d1: suppress ALL repeated warnings of the same type
- +d2: display ONLY error messages
- +fS: send output to file (program will derive filename)
- -f : send output to the screen or pipe
- +pF: define punctuation set according to file F
- +re: run program recursively on all sub-directories.
- +tS: include tier code S
- -tS: exclude tier code S



Check

*kwai (03-Jui-2009) is conducting analyses on:
ALL speaker tiers*

Ratio of morphemes over utterances

```
> check @
depfile.cut being used from: "C:\CHILDES\CLANlib\
check @
Wed Nov 28 16:41:39 2007
check (07-Mar-2007) is conducting analyses on:
  ALL speaker tiers
  and those speakers' ALL dependent tiers
  and ALL header tiers
*****
From file <d:\workingflo\formation_childes\FC_Test_fs01feway.cha>
First pass DONE.
Second pass DONE.
Success! No errors found
```

**NOT: is the bunny*



CHILDES

kwai (02-001-2009) is conducting analyses on:
ALL speaker diaries

Ratio of morphemes over utterances

CHILDES Child Language Data Exchange System



Exercice Check

*NOT: is the bunny rabbit jumping ?

[>]

CHILDES

kwai (02-001-2009) is conducting analyses on:
ALL speaker diaries

Ratio of morphemes over utterances

CHILDES

Child Language Data Exchange System



Codages

*MOT: is the bunny rabbit jumping ?

[>]

Types de codages

➤ Standards (décrits dans le manuel Chat & Clan)

- Phonétique
- Découpage en morphèmes
- Lemmatisation
- Syntaxe
- Actes de langage
- Erreurs

➤ Personnalisés

- Sémantique, syntaxe, gestes, regards, etc.



*NOT: is the bunny rabbit jumping ?



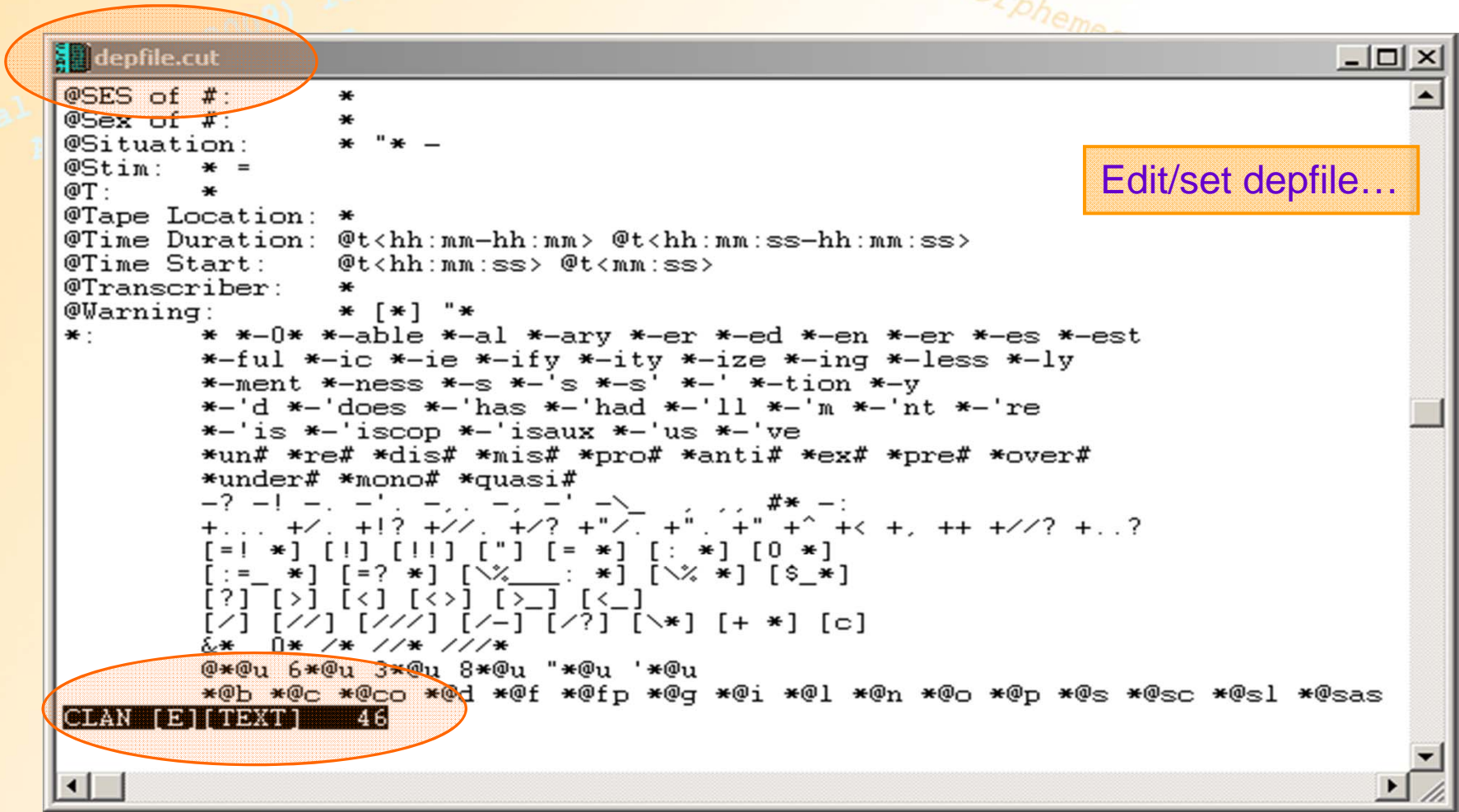
Lignes dépendantes

15.11 Dependent Tiers

%act:	actions
%add:	addressee
%alt:	alternative transcription
%cod:	general purpose coding
%coN:	additional general coding categories, co1, co2
%coh:	cohesion tier
%com:	comments by investigator
%def:	codes from SALT
%eng:	English translation
%err:	error coding
%exp:	explanation
%fac:	facial actions
%flo:	flowing version
%gls:	target language gloss for unclear utterance
%gpx:	gestural and proxemic activity
%int:	intonation
%lan:	language
%mod:	model or target phonology
%mor:	morphemic semantics
%mov:	movie tier
%par:	paralinguistics
%pho:	phonetic transcription
%sit:	situation
%snd:	sonic CHAT sound tier
%spa:	speech act coding
%syn:	syntactic structure notation
%ssy	simple syntactic categories
%tim:	time stamp coding



Ajout/modification de codages



```
depfile.cut
@SES of #: *
@Sex of #: *
@Situation: * "*" -
@Stim: * =
@T: *
@Tape Location: *
@Time Duration: @t<hh:mm-hh:mm> @t<hh:mm:ss-hh:mm:ss>
@Time Start: @t<hh:mm:ss> @t<mm:ss>
@Transcriber: *
@Warning: * [*] "*"
*:
* *-0* *-able *-al *-ary *-er *-ed *-en *-er *-es *-est
*-ful *-ic *-ie *-ify *-ity *-ize *-ing *-less *-ly
*-ment *-ness *-s *-s *-s' *- ' *-tion *-y
*-'d *-'does *-'has *-'had *-'ll *-'m *-'nt *-'re
*-'is *-'iscop *-'isaux *-'us *-'ve
*un# *re# *dis# *mis# *pro# *anti# *ex# *pre# *over#
*under# *mono# *quasi#
-? -! -. -' -.. -\ -\ #* -:
+... +/ . +! ? +//. +/? +"/. +". +^ +< +. ++ +//? +..?
[=! *] [!] [!!] ["] [= *] [: *] [0 *]
[:=_ *] [= ? *] [\%__ : *] [\% *] [$_*]
[?] [>] [<] [<>] [>_] [<_]
[/] [//] [///] [/-] [/?] [/*] [+ *] [c]
&* 0* /* //*/
@*@u 6*@u 3*@u 8*@u "*@u '*@u
*@b *@c *@co *@d *@f *@fp *@g *@i *@l *@n *@o *@p *@s *@sc *@sl *@sas
CLAN [E][TEXT] 46
```



Procédure pour réaliser un codage personnalisé

- Réfléchir à ce qu'on souhaite compter
- Établir le codage
 - Éviter les caractères spéciaux
 - Séparateurs/mots
 - Redondances
 - Ordre ?
- Coder une toute petite partie du corpus à analyser (1 fichier)
- Tester les requêtes
- Automatiser le codage
- Le tester
- Modifier le *depfile.cut* (pour que les codages passent au check)



Utiliser le *Coder Mode*

➤ Définir un fichier « instructions » en mode TEXT

➤ Se mettre en mode « Coder »

➤ Indiquer à Clan quel fichier d'instructions utiliser

➤ Et c'est parti !

```
+b50 +d +l1 +s1
%nsi:
$
?|
su:
pro:per:ani:$
pro:per:ina:$
pro:imp:imp:$
pro:imp:gen:$
pro:imp:col:$
lex:lex:lex:$
aut:
```

+b50 checkpoint buffer
+d backup du fichier original.
+l réordonnancement des codes selon leur fréquence
+s1 pour que les codes encapsulés dans code d'un niveau soient appliqués à tous les codes de même niveau

Mode	Window	Help
Coder mode		Esc-e
Shot mode		Esc-m
Show line numbers		
Sonic mode		Esc-0
UTF8 encoded		
Continuous playback		Esc-8
Insert bullet into text		Ctrl+I
Sound to text sync.		
Play bullet media		F4
Transcribe sound or movie		F5
Check opened file		Esc-L
Disambiguate tier		Esc-2
Exclude tiers		Esc-4
Expand bullets		Esc-a
Send to sound analyzer		

Récapitulatif des commandes en mode Coder

Command	Function
<i>Control -c</i>	insert highlight code at cursor
<i>Esc-c</i>	finish coding current code
<i>Esc-c</i>	finish coding current tier
<i>Control-t or F1</i>	finish coding current tier and go to the next
<i>Esc-t</i>	restrict coding to a particular speaker
<i>Esc-Esc</i>	go on to the next speaker
<i>Esc-s</i>	show subcodes under cursor
<i>Control-g</i>	cancel illegal command



CHILDES

CHILDES

Child Language Data Exchange System



Exercice codage personnalisé

Consignes

➤ Codage des Arguments

- S, V, O, Sujet disloqué, Sujet Trace, Objet Disloqué, Objet Trace, Autres
- Type d'argument Lexical, pronominal
- Type de verbe (aux, mod, lex)

*NOT: is the bunny rabbit jumping ?



Systeme

Proposition :

- S V O Sd St Od Ot A
- Lex, Pro
- aux, mod, lex



Correction fichier code

```
exerciceCode...
\\+b 50 +d +l +s1
%arg:
$S:
lex
pro
$
$O:|
$A:
$Sd:
$Od:
$St:
$Ot:
$V
aux
mod
lex
070308[E][TEXT] 7
```



CHILDES

kwal (02-001-2009) is conducting analyses on:
ALL speaker files

Ratio of morphemes over utterances

CHILDES Child Language Data Exchange System

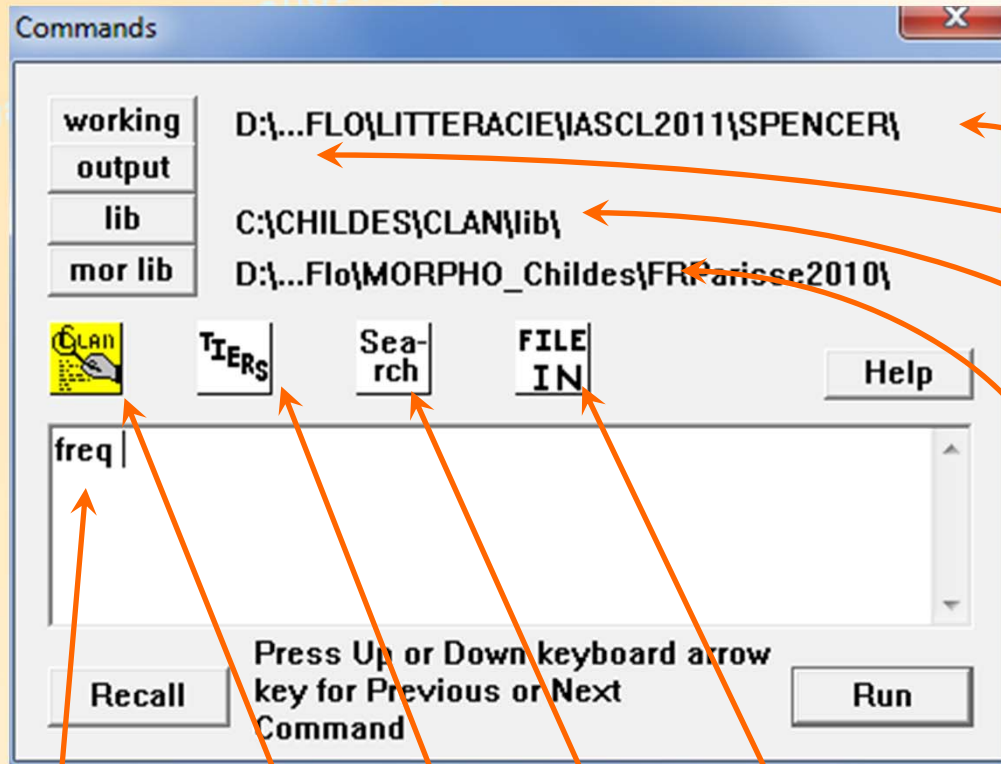


Analyses

*MOT: is the bunny rabbit jumping ?

[>]

Module de Commandes (Rappel)



Commencer par indiquer dans quels dossiers Clan doit chercher/placer les fichiers pour travailler :

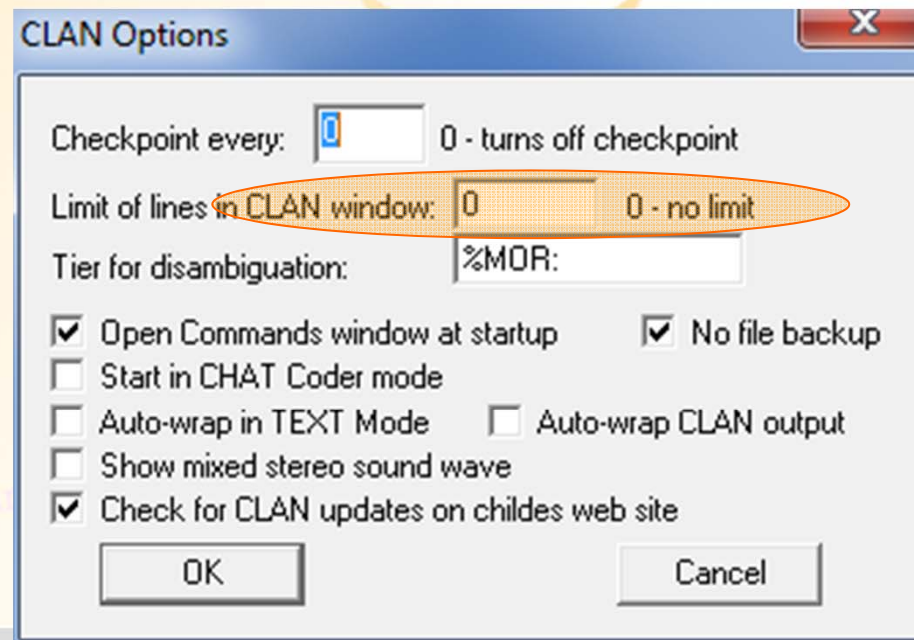
1. Dossier contenant les fichiers à analyser (dossier de travail)
2. Dossier qui contiendra les fichiers résultats d'analyse (si on veut qu'ils soient déposés ailleurs que dans le dossier de travail)
3. Dossier contenant les fichiers de configuration de Clan
4. Dossier contenant les fichiers nécessaires au fonctionnement de la grammaire

Zone de commandes
 Sélectionneur de commandes
 Sélectionneur de lignes dépendantes
 Chaines à rechercher
 Sélectionneur de fichiers



Paramètres

Avant de lancer des analyses, vérifier les paramètres CLAN, notamment Edit/CLAN Options



Commandes préalables

- Validation de la transcription et du codage (CHECK)
- Insertion automatique des lignes @ID (INSERT)
- Examen de *reliability* (RELY)

*NOT: is the bunny rabbit jumping ?



```
> check MAS01VR.cha
depfile.cut being used from: "C:\childes\Clan\lib\depfile.cut"
check MAS01VR.cha
Tue Apr 01 09:23:24 2003
check (06-Mar-2003) is conducting analyses on:
  ALL speaker tiers
  and those speakers' ALL dependent tiers
  and ALL header tiers
*****
*MAS:  [??].
Symbol is not declared in the depfile(s).(11)
*** File "C:\CHILDES\CLAN\Transcriptions
longitudinales\Transcriptions terminées\MAS\MAS01VR.cha":
line 1188.
*MOT:  ## c'+est ça que tu vois?
This item must be followed by text or be removed.(65)
*** File "C:\CHILDES\CLAN\Transcriptions
longitudinales\Transcriptions terminées\MAS\MAS01VR.cha":
line 1189.
Second pass DONE.
Warning: Please repeat CHECK until no error messages are
reported!
```

Check

*kwai (03-Jui-2009) is conducting analyses on:
ALL speaker tiers*

Ratio of morphemes over utterances

```
> check @
depfile.cut being used from: "C:\CHILDES\CLANlib\
check @
Wed Nov 28 16:41:39 2007
check (07-Mar-2007) is conducting analyses on:
  ALL speaker tiers
  and those speakers' ALL dependent tiers
  and ALL header tiers
*****
From file <d:\workingflo\formation_childes\FC_Test_fs01feway.cha>
First pass DONE.
Second pass DONE.
Success! No errors found
```

**NOT: is the bunny*



CHILDES

CHILDES

Child Language Data Exchange System



Exercice CHECK.cha

Éléments pour correction

- Ordre des lignes d'en-tête
- Ligne @ID
- Tabulations
- Apostrophes & nombres
- Majuscules
- Marqueurs de fin d'énoncé
- Pour ne pas contrôler les lignes de codage -t%
- Pour ignorer certains types d'erreur -e *numéro_d'erreur*
- Ne contrôle pas tout !!!



Insertion des lignes @ID

➤ INSERT

- Se base sur la ligne @Participants pour créer les lignes @ID.
- insert *nom_du/des_fichier(s).ext* [options]
- Options :
 - +c*S*: pour spécifier ce qui apparaîtra dans la “case” corpus
 - +f*S*: envoie le résultat dans le fichier *S*
 - -f : envoie le résultat dans la fenêtre *output*
 - +re: explorer les sous-dossiers.
 - +1 : remplacer le fichier original par le nouveau
 - +2 : ne pas créer des versions différentes des fichiers résultats
 - NB: les noms des fichiers peuvent être “*.cha” ou un fichier contenant la liste des fichiers à traiter “@:*nom_du_fichier_liste*”



CHILDES

kwal (02-01-2009) is conducting analyses on:
ALL speaker files

Ratio of morphemes over utterances

CHILDES

Child Language Data Exchange System



Exercice INSERT.cha

*NOT: is the bunny rabbit jumping ?

[>]

Exercice Insertion des lignes @ID

▶ Ex. ExerciceINSERT.cha

▶ Résultat :

```
>insert exerciceinsert.cha
insert exerciceinsert.cha Mon Jun 08 18:53:18 2009
insert (07-Mar-2008) is conducting analyses on:
ALL speaker tiers
*****
From file <ExerciceINSERT.cha>
to file <ExerciceINSERT.ins.cex>
300
Done with file <ExerciceINSERT.ins.cex>
```

```
@Begin
@Languages:          fr
@Participants:       CHI Raphael
Target_Child, MOT Mother, FEY Feyza
Investigator
@ID: fr||CHI||||Target_Child||
@ID: fr||MOT||||Mother||
@ID: fr||FEY||||Investigator||
```



Comparaison de codages *Reliability*

▶ RELY

- ▶ Compare des fichiers codés par des personnes différentes
- ▶ Rely *fichier_maître fichier_2 +t%cod [options]*
- ▶ Options :
 - ▶ +c pour ignorer les modifications de la ligne principale
- ▶ Résultat : un fichier où sont indiquées les lignes différentes



CHILDES

CHILDES

Child Language Data Exchange System



ExerciceRELY1.cha

ExerciceRELY2.cha

Éléments de correction pour l'exercice sur *reliability*

➤ Fichiers : ExerciceRELY1.cha ExerciceRELY2.cha

➤ Commande :

➤ `rely exerciceRely1.cha exerciceRely2.cha +t%coi`

➤ Résultats

```
> rely exerciceRely1.cha exerciceRely2.cha +t%coi
```

```
rely exerciceRely1.cha exerciceRely2.cha +t%coi
```

```
Mon Jun 08 19:03:20 2009rely (07-Mar-2008) is conducting analyses on:
```

```
ALL speaker tiers          and those speakers' ONLY dependent tiers matching: %COI; and  
  ALL header tiers
```

```
*****
```

```
From file <ExerciceRELY1.cha> to file <ExerciceRELY1.rl0.cex>
```

```
Done with file <ExerciceRELY1.rl0.cex>
```

```
From file <ExerciceRELY2.cha> to file <ExerciceRELY2.rly.cex>
```

```
**** Merged data is in a file <ExerciceRELY1.rl0.cex>.
```

```
Done with file <ExerciceRELY2.rly.cex>
```



Éléments de correction pour l'exercice sur *reliability*

Les lignes où des différences apparaissent ressemblent à celle-ci :

*MOT: ah+mais+non les toilettes ça se+sort pas de la maison.

%coi: O:?"ExerciceRELY2.cha" A:?"ExerciceRELY1.cha"



RELY: autre utilisation

→ Rely peut permettre de faire coder le même fichier par deux codeurs différents pour des analyses différentes (ex. %pho et %syn) puis permettra de rassembler les codages dans un seul fichier.



➤ Outils pour l'analyse linguistique

➤ Phonétique, phonologie

➤ Fréquences phonèmes (PHONFREQ, MODREP)



*NOT: is the bunny rabbit jumping ?

Fréquence des sons

➤ PHONFREQ

- Compte la fréquence des sons transcrits en fonction de leur position
- PHONFREQ *nom_du_fichier [options]*
- Options :
 - +b pour spécifier une autre ligne que la ligne %pho
 - +t pour sélectionner des locuteurs
 - +d pour obtenir la liste des mots pris en compte
- Des problèmes restent avec les symboles complexes

Transcription SAMPA rappel:

<http://www.phon.ucl.ac.uk/home/sampa/french.htm>



phonfreq MAS05VR.cha @

Mon Mar 31 16:33:57 2003

phonfreq (06-Mar-2003) is conducting analyses on:

ALL speaker tiers

and those speakers' ONLY dependent tiers matching: %PHO::

From file <c:\childes\clan\transcriptions longitudinales\transcriptions sujets longi\mas\MAS05VR.cha>

396 /	initial =198,	final =198,	other = 0	2 i:	initial = 0,	final = 0,	other = 2
4 2	initial = 0,	final = 0,	other = 4	46 j	initial = 2,	final = 0,	other = 44
20 9	initial = 0,	final = 1,	other = 19	36 k	initial = 4,	final = 1,	other = 31
23 :	initial = 0,	final = 1,	other = 22	11 l	initial = 2,	final = 0,	other = 9
24 @	initial = 0,	final = 4,	other = 20	15 m	initial = 0,	final = 0,	other = 15
20 E	initial = 1,	final = 1,	other = 18	9 n	initial = 0,	final = 0,	other = 9
51 O	initial = 1,	final = 1,	other = 49	12 o	initial = 0,	final = 1,	other = 11
7 S	initial = 0,	final = 0,	other = 7	17 p	initial = 2,	final = 0,	other = 15
1 Z	initial = 0,	final = 0,	other = 1	6 s	initial = 1,	final = 0,	other = 5
138 a	initial = 3,	final = 8,	other = 127	71 t	initial = 1,	final = 0,	other = 70
61 a:	initial = 0,	final = 2,	other = 59	1 tS	initial = 0,	final = 0,	other = 1
20 b	initial = 1,	final = 1,	other = 18	2 u:	initial = 0,	final = 0,	other = 2
20 d	initial = 2,	final = 0,	other = 18	7 v	initial = 1,	final = 0,	other = 6
34 e	initial = 0,	final = 0,	other = 34	9 w	initial = 0,	final = 0,	other = 9
5 e:	initial = 0,	final = 0,	other = 5	1 z	initial = 0,	final = 0,	other = 1
1 f	initial = 0,	final = 0,	other = 1	1 ~	initial = 0,	final = 0,	other = 1
1 g	initial = 0,	final = 0,	other = 1				
3 i	initial = 0,	final = 0,	other = 3				



CHILDES

CHILDES

Child Language Data Exchange System



Exercice PHONFREQsampa.cha
Exercice PHONFREQ_API.cha

Correspondances entre lignes

▶ MODREP

- ▶ Aligne 2 lignes à condition qu'elles aient le même nombre d'éléments

- ▶ Ex.

*MOT: oui@i ça c'est un fauteuil mon+ange.
%pho: wi sa sE e~ fot9j mo~naz

- ▶ `modrep nom_du_fichier +bligne_base +cligne_à_aligner [options]`
 - ▶ +k pour respecter la casse



CHILDES

CHILDES Child Language Data Exchange System



Exercice MODREP1.cha
Exercice MODREP2.cha

Éléments de correction pour l'exercice sur MODREP

➤ Commande :

➤ modrep ExerciceMODREP2.cha +b* +c%syn +k

➤ Résultats:

```
From file <d:\WorkingFlo\ExerciceMODREP2.cha>
1 ah+mais+non                2 dans
   1 adv                      2 prep
1 ben@i                       2 de
   1 me                       2 prep
2 c'est                      1 fauteuil
   2 va                       1 nc
2 chambre                    4 la
   2 nc                       4 det
1 comme                       4 le
   1 conj                     3 pro
                               1 det
```



➤ Outils pour l'analyse linguistique

➤ Phonétique, phonologie

➤ Fréquences phonèmes (PHONFREQ, MODREP)

➤ Morphologie, lexique, Syntaxe

➤ Fréquences (FREQ, STATFREQ, VOCD)



Calculs de Fréquences

Commande :

➤ freq [cN oN dN fS k l pF rN re sS tS u xN yN zN 2] filename(s)

➤ Options :

- +c : travailler seulement sur les mots comportant des majuscules
- +c1: trouver les mots qui ont des majuscules au milieu seulement
- +c2: trouver les correspondants aux chaînes spécifiées dans l'option +s
- +s inclure le mot S ou les mots dans le fichier @S
- -sS: exclure le mot S ou les mots dans le fichier @S
- +o : trier par fréquence descendante
- +o1: trier par concordance inverse
- +d : sort les mots avec fréquence et numéro de ligne
- +d1: les mots sans les info de fréquence. format KWAL ou COMBO
- +d2: format STATFREQ. Must include speaker specifications
- +d3: seulement les info type/token format STATFREQ. Must include speaker specifications
- +d4: seulement info type/token
- +d5: tous les mots, même ceux qui ont 0 pour fréquence
- +fS: fabrique un fichier
- -f : envoie les résultats dans la fenêtre output
- +k : considère la casse des caractères



- +pF: permet de définir un fichier de ponctuation F
- +rN: if N = 1 then "get(s)" devient "gets", 2- "get(s)", 3- "get" 4- reconnaît les symboles prosodiques dans les mots, 5- pas de remplacement de texte: [: *] 6- exclure les répétitions: </>, <///>, </-> et </?>, 7- ne pas effacer les caractères '/', '~' et ':'
- +re: cherche les fichiers dans les sous-répertoires
- +tS: inclure la ligne code S
- -tS: exclure la ligne code S
- +u : rassembler tous les fichiers spécifiés.
- -u : calculer les résultats pour chaque tour séparément
- +xN: inclure seulement les énoncés plus longs que N
- -xN: inclure seulement les énoncés égaux à ou moins longs que N
- +/-2:
 - +2 do not create different versions of output file names
 - -2 create them
- File names can be "*.cha" or a file of list of names "@:filename"



Freq sur ligne %mor (ajouts 2008)

- > +s@ Followed by file name or morpho-syntax
- > Morpho-syntax search pattern
- > # prefix marker
- > | part-of-speech
- > r stem of the word
- > - suffix marker
- > & nonconcatenated morpheme
- > = English translation for the stem
- > followed by - and <--does not apply to contractions (~)
- > * find any match
- > % erase any match
- > word -find "word"

- > Exception:
- > ~ contractions
- > followed by - and
- > * find ONLY contractions
- > + find contractions and other matches
- > % erase any match

- > For example:
- > +t%mor -t* +s"@r-*,o-%"
- > find all stems and erase all other markers
- > +t%mor -t* +s"@r-*,|-adv,o-%"
- > find all stems of all "adv" and erase all other markers
- > +t%mor -t* +s"@r-be"
- > find all forms of "be" verb
- > +t%mor -t* +s"@r-*,|-*,--%,o-%"
- > find all stems, parts-of-speech followed by any suffix and erase suffix and other markers



```

> freq newfile.cha +o
freq newfile.cha +o
Sat Mar 29 18:03:12 2003
freq (21-Sep-2001) is conducting analyses on:
  ALL speaker tiers

```

*****From file <newfile.cha>

3 de	1 isolée	1 s'
3 nous	1 les	1 société
2 il	1 liste	1 sont
2 la	1 mais	1 tous
2 pouvons	1 manière	1 un
2 que	1 ne	1 verbale
1 aborder	1 orale	1 vie
1 agissent	1 ou	1 violence
1 allons	1 pas	1 vivre
1 certainement	1 pouvez	1 vous
1 compléter	1 problèmes	1 à
1 en	1 qu'	
1 est	1 qui	41 Total number of different word types used
1 fait	1 rejet	49 Total number of words (tokens)
1 inhérents	1 rencontrer	0.837 Type/Token ratio



CHILDES

CHILDES

Child Language Data Exchange System



ExerciceFREQ.cha
ExerciceFREQ2.cha

Correction exercice FREQ

- freq @
- freq @ -t* +t%coi
- freq @ -t* +t%coi +s"A"
- freq @ -t* +t%coi +s"A«

- freq exerciceFREQ2.cha -t* +t%mor +s"det|%"
- freq exerciceFREQ2.cha -t* +t%mor +s"det*|%"
- freq exerciceFREQ2.cha -t* +t%mor +s"v*|%"
- freq exerciceFREQ2.cha -t* +t%mor +s"%|%&*"
- freq exerciceFREQ2.cha -t* +t%mor +s"%|%&*" +s"%|%-*" +s"%|%_*"
- freq exerciceFREQ2.cha -t* +t%mor +s"-"



Statfreq

➤ Résumé des fréquences pour 1 locuteur sur plusieurs dans un format lisible par les logiciels de stats

```
a cat chalk+chalk fine here just minute mommy neat no not
that the uhhuh what's white yeah
1.25.8.2 0 0 1 0 0 0 0 1 1 0 0 1 0 0
1 0 2
1.26.8.2 1 1 0 1 1 1 1 0 0 1 1 0 1 1
0 1 0
```



Statfreq

3 étapes

1. S'assurer que les en-têtes @ID sont OK
2. Utiliser `FREQ` avec les options `+t` (locuteur) et `+d2` (produit un fichier temporaire `stat.out`).

```
freq +d2 +t@ID="eng|samp|*|0110*CHI" sample.cha
```

3. `FREQ` demandera de lancer `STATFREQ`
`statfreq stat.out +f +d`



CHILDES

CHILDES

Child Language Data Exchange System



exercíce **STATFREQ**

Correction Exercice STATFREQ

Commandes :

- `freq +d2 +t*SBJ +s"*qu*" *.cha*`
- Ou `freq +d2 +t@ID="fr|Jisa_ReFlex|SBJ|*|*|gsd|*" +s"*qu*" *.cha*`
- `statfreq stat.out.cex`

Résultats :

language corpus speaker age sex group SES role situation *SBJ: lorsqu' marqué qu' que qui Types Token TTR									
fr Jisa_ReFlex SBJ 12;02.13 female gsd low Target_Child frgs01fewfad	9	0	1	0	1	1	3	3	1.000
fr Jisa_ReFlex SBJ 12;02.13 female gsd low Target_Child frgs01fnwfad	6	0	0	0	0	2	1	2	0.500
fr Jisa_ReFlex SBJ 11;00.19 female dsw low Target_Child frgs02fewfbw	8	0	0	0	0	1	1	1	1.000
fr Jisa_ReFlex SBJ 11;00.19 female gsw low Target_Child frgs02fnwfbw	8	0	0	0	0	0	0	0	-
fr Jisa_ReFlex SBJ 11;04.07 female gsw low Target_Child frgs03fewfaw	21	3	0	3	2	1	4	9	0.444
fr Jisa_ReFlex SBJ 11;04.07 female gsw low Target_Child frgs03fnwfaw	11	0	0	0	1	0	1	1	1.000



Utilisation du résultat dans Excel ou autres

- Fichier/ouvrir
- Format UTF8
- Texte avec séparateurs **espaces** (des séparateurs consécutifs doivent être considérés comme un seul)
- Vérifier (données/filtre automatique dans excel)



CHILDES

CHILDES

Child Language Data Exchange System



stat.out.sat.cex
vers Excel
vers Statview

VOCD

➤ VOCabulary Diversity (TTR sensible à la longueur du texte)

➤ Basé sur la probabilité que de nouveaux éléments soient ajoutés au cours du texte (modèle mathématique évolution TTR en fonction nb tokens. Cf. travaux de Malvern & Richards).

➤ Commande :

➤ `vocd [<options>] filenames`

➤ Options

+c : seulement les mots avec majuscules

+d : donne les énoncés et les infos type-token

+d4: donne seulement l'info type/token

+fS: send output to file (<infilename>.S)

+g : compute "limiting type-type ratio".

+k : prête attention à la casse des caractères

+r1: get(s) devient gets

+r2: get(s) devient get(s)

+r3: get(s) devient get

+r4: efface les symboles prosodiques dans les mots

+r5: ne fait pas les remplacements [: *]

+r6: exclure répétitions, reformulations: [/], [//]

+sS: mot à inclure

+s@S: liste de mots à inclure

-sS: mot à exclure

-s@S: liste de mots à exclure

+tS: lignes à inclure

+u : rassembler les fichiers



VOCD

- Seulement pour des corpus de plus de 50 items !
- Plus le corpus est petit, moins D est fiable
- Les fichiers doivent passer au check +g3 (ie. Maj, nb et apostrophes)
- Freq pour voir ce qui sera considéré comme word type
- Permet de construire une liste de mots à exclure (-s)
- Exemple:

➤ VOCD +t*CHI +r6 -s@MotsExclus.ext +s"*-%%%" file
Exclure les répétitions, reformulations, etc.



VOCD

➤ Résultat en 4 parties

- Liste des énoncés
- 3 tables pour les 3 échantillons aléatoires
- Résumé sous chaque table
- Résumé final

VOCD RESULTS SUMMARY

=====

Command line: vocd exerciceVOCD.cha

File name: exerciceVOCD.cha

Types, Tokens, TTR: <116,402,0.288557>

D_optimum values: <33.51, 33.68, 33.94>

D_optimum average: 33.71



CHILDES

kwal (02-001-2009) is conducting analyses on:
ALL speaker files

Ratio of morphemes over utterances

CHILDES

Child Language Data Exchange System



Exercice VOCD.cha

*MOT: is the bunny rabbit jumping ?

[>]

➤ Outils pour l'analyse linguistique

➤ Phonétique, phonologie

- Fréquences phonèmes (PHONFREQ, MODREP)

➤ Morphologie, lexique, Syntaxe

- Fréquences (FREQ, STATFREQ, VOCD)
- Longueur des mots (WDLEN, MAXWD) longueur des énoncés (MLU)

*NOT: is the bunny rabbit jumping ?



Wdlen

- Nb de mots de chaque longueur (en morphèmes, en caractères)
- Nb d'énoncés de chaque longueur (en morphèmes, en mots)
- Nb de tours de parole de chaque longueur (en morphèmes, en mots)
- Commande
 - wdlen *nom_du_fichier* [options]



Wdlen

```
> wdlen @
wdlen @
Fri Nov 30 11:56:08 2007
wdlen (07-Mar-2007) is conducting analyses on:
  ALL speaker tiers
*****
From file <d:\workingflo\formation_childes\FC_Test_fs01feway.tr.cha>

Number of words of each length in characters
  1  2  3  4  5  6  7  8  9 10 11 12 13 14  Mean
-----
*SBJ:  16 47 24 24 20  5 13 14 11 13  5  5  2  1  4.97

Number of words of each of these morpheme lengths
  1  Mean
-----
*SBJ:  200  1.00

Number of utterances of each of these lengths in morphemes
  1  2  3  4  5  6  7  8  9 10 11 12  Mean
-----
*SBJ:   0  2  4  4  6  6  1  3  3  1  2  1  6.06

Number of utterances of each of these lengths in words
  1  2  3  4  5  6  7  8  9 10 11 12  Mean
-----
*SBJ:   0  2  4  4  6  6  1  3  3  1  2  1  6.06
```

Ratio of morphemes over utterances



CHILDES

kwai (02-001-2009) is conducting analyses on:
ALL speaker diaries

Ratio of morphemes over utterances

CHILDES

Child Language Data Exchange System



Exercice WDLEN.cha

*MOT: is the bunny rabbit jumping ?

[>]

kwai (03-Jui-2009) is conducting analyses on ALL speaker tiers

```

> wrlen exercicewdlen.cha
wrlen exercicewdlen.cha
Wed Jun 10 17:48:52 2009
wrlen (09-Jun-2009) is conducting analyses on:
  ALL speaker tiers
*****
From file <ExerciceWDLEN.cha>

Number of words of each length in characters
  1  2  3  4  5  6  7  8  9 10 Mean
-----
*MOT:  0  7  2  4  1  1  3  2  0  1  4.48

Number of words of each of these morpheme lengths
  1 Mean
-----
*MOT:  21  1.00

Number of utterances of each of these lengths in morphemes
  1  2  3  4  5  6  7 Mean
-----
*MOT:  0  1  0  0  0  2  1  5.25

Number of utterances of each of these lengths in words
  1  2  3  4  5  6  7 Mean
-----
*MOT:  0  1  0  0  0  2  1  5.25

Number of single turns of each of these lengths in utterances
  1  2  3  4 Mean
-----
*MOT:  0  0  0  1  4.00

Number of single turns of each of these lengths in words
  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 Mean
-----
*MOT:  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1  21.00

```



Mot le plus long

➤ Commande

- maxwd *nom_du_fichier* [options]

➤ Options

- +bS: make all S characters morpheme delimiters (default: -#~)
- -bS: do not consider all S characters to be morpheme delimiters
- +cN: display N longest utterances/words
- +gN: look for longest utterance instead of longest word 1 - number of morph; 2 - number of word; 3 - number of chars
- +j : consider ONLY unique length utterances/words
- +d : outputs one line for the length level and the next line for the word
- +d1: produces longest words, one per line, in legal CHAT format



Maxwd

```
> maxwd @
maxwd @
Fri Nov 30 11:59:28 2007
maxwd (07-Mar-2007) is conducting analyses on:
  ALL speaker tiers
*****
From file <d:\workingflo\formation_childes\FC_Test_fs01feway.tr.cha>
*** File "d:\workingflo\formation_childes\FC_Test_fs01feway.tr.cha": line 39: 13 characters long:
personnalités
*** File "d:\workingflo\formation_childes\FC_Test_fs01feway.tr.cha": line 46: 13 characters long:
apprentissage
```



CHILDES

CHILDES

Child Language Data Exchange System



Exercice MAXWD.cha

➤ Outils pour l'analyse linguistique

➤ Phonétique, phonologie

- Fréquences phonèmes (PHONFREQ, MODREP)

➤ Morphologie, lexique, Syntaxe

- Fréquences (FREQ, STATFREQ, VOCD)
- Longueur des mots (WDLEN, MAXWD) longueur des énoncés (MLU)
- Lemmatisation (MOR)
- Désambiguïisation du balisage morphologique (POST) -
Parisse, C. et Le Normand, M. T., 2000



Grammaires

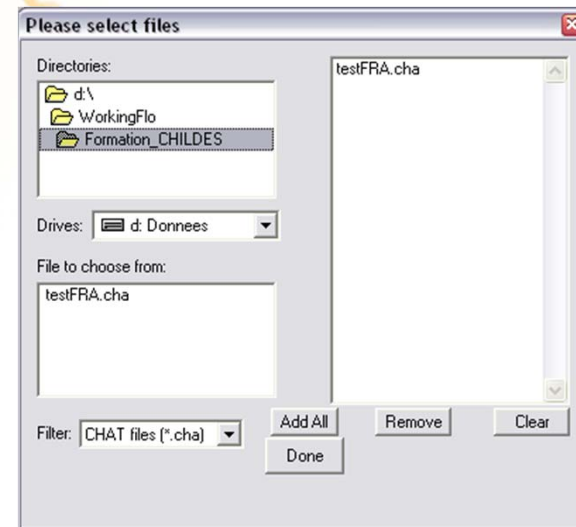
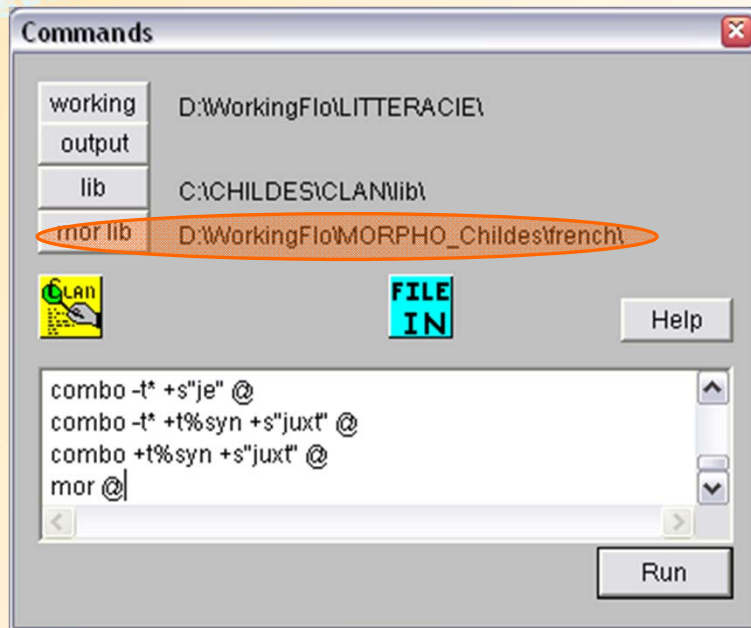
- allemand, anglais, cantonais, chinois, espagnol, français, hébreu, hollandais, italien, japonais
 - <http://chilides.psy.cmu.edu/morgrams/>
- Télécharger la grammaire, la décompresser, indiquer à Clan dans quel dossier se trouve la grammaire (par l'intermédiaire du bouton mor lib de la fenêtre de commandes)



Commande MOR

kwai (03-Jui-2009) is coming analyses on:
ALL speaker tiers

Ratio of morphemes over utterances



*MOT: is the bunny rabbit jumping ?



> mor @
mor @
Wed Nov 28 15:14:59 2007
mor (07-Mar-2007) is conducting analyses on:
ALL speaker tiers
and those speakers' ALL dependent tiers
and ALL header tiers

From file <d:\workingflo\formation_childes\testFRA.cha> to file <d:\workingflo\formation_childes\testFRA.mor.cex>
Using sf-rule: D:\WorkingFlo\MORPHO_Childes\french\sf.cut.
Using a-rules: D:\WorkingFlo\MORPHO_Childes\french\ar.cut.
Using lexicon: D:\WorkingFlo\MORPHO_Childes\french\lex\ADJ.CUT.
Using lexicon: D:\WorkingFlo\MORPHO_Childes\french\lex\CLO.CUT.
Using lexicon: D:\WorkingFlo\MORPHO_Childes\french\lex\COMPOSES.CUT.
Using lexicon: D:\WorkingFlo\MORPHO_Childes\french\lex\N.CUT.
Using lexicon: D:\WorkingFlo\MORPHO_Childes\french\lex\NPROP.CUT.
Using lexicon: D:\WorkingFlo\MORPHO_Childes\french\lex\SUP.CUT.
Using lexicon: D:\WorkingFlo\MORPHO_Childes\french\lex\SUPMARIE.CUT.
Using lexicon: D:\WorkingFlo\MORPHO_Childes\french\lex\V.CUT.
Using lexicon: D:\WorkingFlo\MORPHO_Childes\french\lex\VAUX.CUT.
Using lexicon: D:\WorkingFlo\MORPHO_Childes\french\lex\VINFL.CUT.
Using lexicon: D:\WorkingFlo\MORPHO_Childes\french\lex\VPP.CUT.
Using lexicon: D:\WorkingFlo\MORPHO_Childes\french\lex\VPROG.CUT.
Loaded lexicon: 89189
Using c-rules: D:\WorkingFlo\MORPHO_Childes\french\cr.cut.
Done with file <d:\workingflo\formation_childes\testFRA.mor.cex>

>



MOR

kwai (03-Jui-2009) is conducting analyses on:
ALL speaker tiers

Ratio of morphemes over utterances

*SBJ: nous allons aborder les problèmes .

```
%mor: pro:refl|nous&1P^pro:subj|nous&1P^pro:obj|nous&1P^pro|nous&PL  
v:mdllex|aller&IMP&1PV^v:mdl|aller&IMP&1PV^v:mdllex|aller&PRES&1PV^v:mdl|aller&PRES&1PV  
v:inf|aborder det|les&PL^pro:obj|les&PL n|problème&_MASC-_PL .
```

%cpx: \$Lon4:Sem3:Mod2:Dep3:Var1:nsu:FF

%nsl: \$nous|suj:pro:per:ani:FF \$problèmes|aut:lex:lex:lex:FF

*SBJ: que nous pouvons tous rencontrer .

```
%mor: adv:neg|que^adv|que^adv:int|que^pro:rel|que^conj|que^prep|que  
pro:refl|nous&1P^pro:subj|nous&1P^pro:obj|nous&1P^pro|nous&PL  
v:mdllex|pouvoir&IMP&1PV^v:mdl|pouvoir&IMP&1PV^v:mdllex|pouvoir&PRES&1PV^v:mdl|pouvoir&PRES&1PV  
det|tous&MASC&_PL^pro|tous&MASC&_PL v:inf|rencontrer .
```

%nsl: \$que|aut:pro:per:ina:FF \$tous|aut:pro:imp:gen:FF

*SBJ: qui sont inhérents à la vie en société .

```
%mor: pro:rel|qui^pro:int|qui v:exist|être&PRES&3PV^v:aux|être&PRES&3PV  
adj|inhérent&MASC-_PL adv|à^prep:art|à^prep|à  
det|la&FEM&SING^pro:obj|la&FEM&SING n|vie&_FEM  
adv|en^pro:y|en^prep:art|en^prep|en n|société&_FEM .
```

%nsl: \$qui|suj:pro:per:ina:FF \$vie|aut:lex:lex:lex:FF

\$société|aut:lex:lex:lex:FF

MOR

➤ Commande

➤ mor [options] filename(s)

➤ Options

- +b : use disambiguation rules
- +c : treat clitics as separate words
- +eS: show result of a-rules on either a stem S or stems in file @S
- +p : use pinyin lexicon format
- +q : don't treat quoted material as one object
- +xi: interactive test mode
- +xl: check lexicon mode
- +xa: check lexicon for ambiguous entries
- +xc: check lexicon mode, including capitalized words
- +xb: check lexicon mode, include word location in data files

- +fS: send output to file (program will derive filename)
- -f : send output to the screen or pipe+l : add language tag to every word
- +re: run program recursively on all sub-directories.
- +sS: search for word S or words in file @S in an input file (+s@ for more info).
- -sS: exclude word S or words in file @S from a given input file (-s@ for more info)
- +tS: include tier code S
- -tS: exclude tier code S
- +/-2: +2 do not create different versions of output file names / -2 create them
- File names can be "*.cha" or a file of list of names "@:filename"



CHILDES

kwal (02-001-2009) is conducting analyses on:
ALL speaker files

Ratio of morphemes over utterances

CHILDES

Child Language Data Exchange System



Exercice MOR.cha

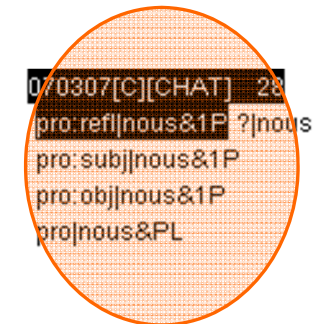
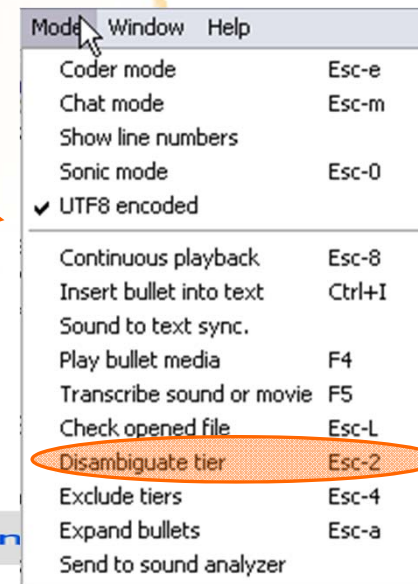
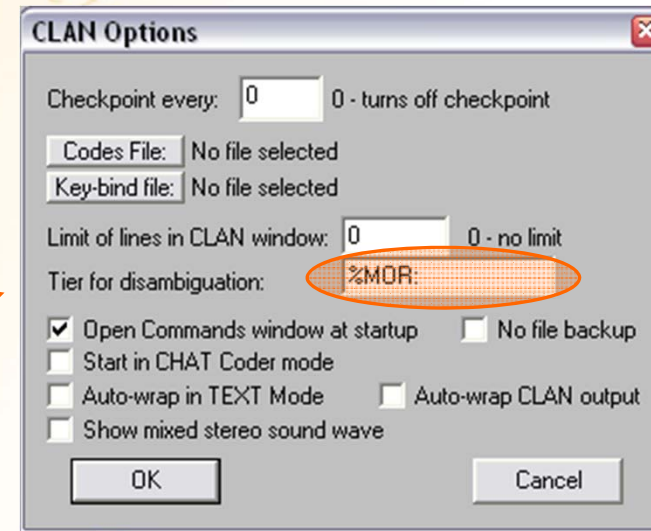
*MOT: is the bunny rabbit jumping ?

[>]

Désambiguiser

➤ Choix de la ligne à désambiguiser

➤ Mode « disambiguate tier »



CHILDES

CHILDES

Child Language Data Exchange System



Exercice Desamb.cha

➤ Outils pour l'analyse linguistique

➤ Phonétique, phonologie

➤ Fréquences phonèmes (PHONFREQ, MODREP)

➤ Morphologie, lexique, Syntaxe

➤ Fréquences (FREQ, STATFREQ)

➤ Lemmatisation (MOR, POST)

➤ Désambiguïisation du balisage morphologique (POST) -

Parisse, C. et Le Normand, M. T., 2000

➤ Longueur des mots (WDLEN, MAXWD)

➤ Contexte (KWAL, COMBO, COOCCUR)



Désambigüiser automatiquement

➤ MOR

***SBJ:** nous allons aborder les problèmes .

%mor: pro:refl|nous&1P^pro:subj|nous&1P^pro:obj|nous&1P
v:mdllex|aller&IMP&1PV^v:mdl|aller&IMP&1PV^v:mdllex|aller
&PRES&1PV^v:mdl|aller&PRES&1PV v:inf|aborder
det|les&PL^pro:obj|les&PL n|problème&_MASC-_PL .

➤ POST

***SBJ:** nous allons aborder les problèmes .

%mor: pro:subj|nous&1P
v:mdl|aller&PRES&1PV v:inf|aborder det|les&PL
n|problème&_MASC-_PL .



Commande POST

Options :

```
> post
POST-ANALYZE: version - 2.5 - INSERM - PARIS - FRANCE
Usage: post [dF eNc f IN sN tF] filename(s)
+dF : use POST database file F (if omitted, default is fr.db).
+eNC: this option allows you to change the separator used.
      It is a complement to the option +s2 and +s3 only.
      +e1C- change separator used between different solutions to C (default '#').
      +e2C- change separator used before info on parsing process to C (default '/').
+f : send output to file derived from input file name (default).
+fF : send output to file F.
-f : send output to the screen.
+IN : line limit.
+sN : choose style N.
      0- replace ambiguous %mor lines with disambiguated ones (default).
      1- keep ambiguous %mor lines and add disambiguated %pos lines.
      2- output as in N=1, but with slashes marking undecidable cases.
      3- keep ambiguous %mor lines and add %pos lines with debugging info.
      4- outputs results in an internal format. It's useful for developer mostly.
      5- inserts a %nob line before the %mor/%pos line that presents the results of the analysis without using Brill rules.
-b : do not use Brill's rules to improve the results of the analysis (uses rules by default).
+bs : use a slower but more thorough version of Brill's rules analysis.
+t : keep all affixes.
+tF : keep the affixes listed in file F.
-t : keep only the affixes defined during training with POSTTRAIN (default).
-tF : omit the affixes in file F.
+lm : reduce memory use (but longer processing time).
+unk: tries to process unknown words.
+re : run program recursively on all sub-directories.
+1 : replace original data file with new one.
```



*kwal (03-Jui-2009) is conducting analyses on:
ALL speaker tiers*

Ratio of morphemes over utterances

> post @ +dfr.db

Using file: D:\WORKINGFLO\MORPHO_CHILDES\french2007\fra.db.

warning: there are no Brill's rules in this database.

post @ +dfra.db

Wed Nov 28 15:21:44 2007

post (07-Mar-2007)

From file <d:\workingflo\formation_childes\testFRA.cha> to file <d:\workingflo\formation_childes\testFRA.pst.cex>

Done with file <d:\workingflo\formation_childes\testFRA.cha>

From file <d:\workingflo\formation_childes\testFRA.mor.cex> to file <d:\workingflo\formation_childes\testFRA.mor.pst.cex>

Done with file <d:\workingflo\formation_childes\testFRA.mor.cex>

**MOT: is the bunny rabbit jumping ?*



POST

*SBJ: nous allons aborder les problèmes .

%mor: pro:subj|nous&1P v:mdl|aller&IMP&1PV^v:mdl|aller&PRES&1PV v:inf|aborder
det|les&PL n|problème&_MASC-_PL .

%cpx: \$Lon4:Sem3:Mod2:Dep3:Var1:nsu:FF

%nsl: \$nous|suj:pro:per:ani:FF \$problèmes|aut:lex:lex:lex:FF

*SBJ: que nous pouvons tous rencontrer .

%mor: conj|que pro:subj|nous&1P v:mdllex|pouvoir&IMP&1PV^v:mdllex|pouvoir&PRES&1PV
det|tous&MASC&_PL v:inf|rencontrer .

%nsl: \$que|aut:pro:per:ina:FF \$tous|aut:pro:imp:gen:FF

*SBJ: qui sont inhérents à la vie en société .

%mor: pro:int|qui v:exist|être&PRES&3PV adj|inhérent&MASC-_PL prep|à det|la&FEM&SING
n|vie&_FEM prep:art|en n|société&_FEM .

%nsl: \$qui|suj:pro:per:ina:FF \$vie|aut:lex:lex:lex:FF

\$société|aut:lex:lex:lex:FF



CHILDES

CHILDES

Child Language Data Exchange System



Exercice POST.cha

➤ Outils pour l'analyse linguistique

➤ Phonétique, phonologie

- Fréquences phonèmes (PHONFREQ, MODREP)

➤ Morphologie, lexique, Syntaxe

- Fréquences (FREQ, STATFREQ, VOCD)
- Longueur des mots (WDLEN, MAXWD) longueur des énoncés (MLU)
- Lemmatisation (MOR)
- Désambiguïsation du balisage morphologique (POST) -
Parisse, C. et Le Normand, M. T., 2000
- Contexte (KWAL, COMBO, COOCCUR)



Kwal

- Pour sélectionner des lignes en fonction de critères particuliers et fabriquer un fichier contenant cette sélection sur lequel on pourra lancer des analyses
- KWAL- Key Word And Line concordance program : recherche des mots-clef et produit ces mots clefs en contexte.



Kwal

➤ Commande :

➤ kwal [a b n dN fS k l oS pF rN re sS tS u wN xN yN zN 2] filename(s)

➤ Options

+a résultat par ordre alphabétique des mots-clefs
+b seulement si le mot-clef est le seul élément de la ligne
+nS inclut tous les énoncés du locuteur S quand +s
-nS exclut tous les énoncés du locuteur S quand +s
+d format CHAT legal
+d1 format CHAT legal plus noms de fichiers et n° de lignes
+d2 noms de fichiers une seule fois par fichier
+d3 seulement les éléments cherchés
+fS Résultat dans un fichier
-f Résultat dans la fenêtre output
+k prendre en compte la casse des caractères
+l ajouter le code langue pour tous les mots
+oS lignes à inclure dans l'output
-oS lignes à exclure de l'output
+pF prendre en compte la ponctuation définie dans F
+rN if N = 1 then "get(s)" goes to "gets", 2- "get(s)", 3- "get" 4- recognize prosodic symbols in words, 5- no text replacement: [: *] 6- exclude repetitions: </>, <//>, <///>, </-> and </?>, 7- do not remove '/', '~' and ':' characters

+re run program recursively on all sub-directories.
+sS search for word S or words in file @S in an input file (+s@ for more info).
-sS exclude word S or words in file @S from a given input file (-s@ for more info)
+tS include tier code S
-tS exclude tier code S
+u merge all specified files together.
+wN display N number of utterance AFTER the given one
-wN display N number of utterance BEFORE the given one
+xN include only utterances which are N items in length or longer
-xN include only utterances which are N items in length or shorter
+y work on TEXT format files one line at the time
+y1 work on TEXT format files one utterance at the time
+zN compute statistics on a specified range of input data
+/-2 +2 do not create different versions of output file names / -2 create them

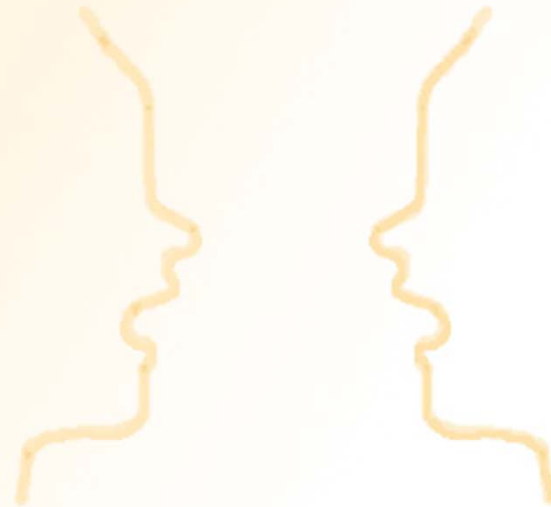
File names can be "*.cha" or a file of list of names "@:filename"

Kwal

kwal (03-Jui-2009) is conducting analyses on:
ALL tiers

➤ +o@ pour inclure les lignes d'en-tête

Ratio of morphemes over utterances



*MOT: is the bunny rabbit jumping ?



CHILDES

CHILDES

Child Language Data Exchange System



Exercice KWAL.cha

Combo

- COMBO permet de composer des recherches de chaînes booléennes pour trouver des patrons de lettres, mots, ou groupes de mots (syntactic analysis)
- Obligation d'inclure la ligne principale

Table 3: COMBO Strings

Meaning	Type	Symbol
immediately FOLLOWED by	Boolean	^
inclusive OR	Boolean	+
logical NOT	Boolean	!
repeated character	metacharacter	*
single character	metacharacter	_
quoting	metacharacter	\



Combo

- +bN: search a N number-cluster unit(s)
- +g1: do a string oriented search on a whole tier. (Default: word oriented)
- +g2: do a string oriented search on just one word.
- +g3: do not continue searching on a tier after first failure.
- +g4: perform additional filter on negative searches.
- +g5: exclude utterance delimiters from search.
- +g6: negative search must match at least one word.
- +sS: specify search pattern S
- +x : make search <s1>^<s2> identical to search <s2>^<s1>
- +d : outputs each matched sentence in a simple legal CHAT format
- +d1: outputs legal CHAT format with line numbers and file names
- +d2: outputs file names once per file only
- +d3: outputs ONLY matched words in the same format as +d1
- +d4: if string match is found, add codes and tiers to data file
- +fS: send output to file (program will derive filename)
- f : send output to the screen or pipe
- +k : treat upper and lower case as different
- +l : add language tag to every word
- +oS: include additional tier code S for output purposes ONLY
- oS: exclude tier code S from an additional inclusion in an output

- +pF: define punctuation set according to file F
- +rN: if N = 1 then "get(s)" goes to "gets", 2- "get(s)", 3- "get" 4- recognize prosodic symbols in words, 5- no text replacement: [:*] 6- exclude repetitions: </>, <//>, <///>, </-> and </?>, 7- do not remove '/', '~' and ':' characters
- +re: run program recursively on all sub-directories.
- +tS: include tier code S
- tS: exclude tier code S
- +u : merge all specified files together.
- +wN: display N number of utterance AFTER the given one
- wN: display N number of utterance BEFORE the given one
- +xN: include only utterances which are N items in length or longer
- xN: include only utterances which are N items in length or shorter
- +y : work on TEXT format files one line at the time
- +y1: work on TEXT format files one utterance at the time
- +zN: compute statistics on a specified range of input data
- +/-2: +2 do not create different versions of output file names / -2 create them
File names can be "*.cha" or a file of list of names "@:filename"



> **combo** +sregarde MAS05VR.cha
(regarde)
combo +sregarde MAS05VR.cha
Mon Mar 31 17:05:17 2003
combo (06-Mar-2003) is conducting analyses on:
ALL speaker tiers

From file <MAS05VR.cha>

*** File "MAS05VR.cha": line 17.

*VAL: tiens # (1)regarde .

*** File "MAS05VR.cha": line 50.

*MOT: tiens (1)regarde c(e) que j(e) voulais t(e) donner aujourd'hui .

*** File "MAS05VR.cha": line 51.

*MOT: (1)regarde .

*MOT: is the bunny rabbit jumping ?



CHILDES

CHILDES

Child Language Data Exchange System



Exercice COMBO.cha

Cooccur

→ Cooccur cherche des co-occurrences de mots (clusters syntaxiques). Par défaut, les clusters ont une longueur de 2 mots (+n)



cooccur +t%mor -t* +s"*|nous*" fcorpus-4.mor.txt

cooccur +t%mor -t* +s*|nous* fcorpus-4.mor.txt

Fri Apr 04 10:03:48 2003

cooccur (21-Sep-2001) is conducting analyses on:

ONLY dependent tiers matching: %MOR;

From file <FCorpus-4.mor.txt>

- 1 adv:yn|oui pro|nous&1p
- 1 prep|avec pro|nous&1p
- 1 pro:dat|nous&1p v:mdl|faire&pres&2sv
- 1 pro:int|qui pro|nous&1p
- 1 pro:subj|tu&2s pro:dat|nous&1p



CHILDES

CHILDES

Child Language Data Exchange System



Exercice COOCCUR.cha

➤ Outils pour l'analyse linguistique

➤ Phonétique, phonologie

- Fréquences phonèmes (PHONFREQ, MODREP)

➤ Morphologie, lexique, Syntaxe

- Fréquences (FREQ, STATFREQ, VOCD)
- Longueur des mots (WDLEN, MAXWD) longueur des énoncés (MLU)
- Lemmatisation (MOR)
- Désambiguïsation du balisage morphologique (POST) -
Parisse, C. et Le Normand, M. T., 2000
- Contexte (KWAL, COMBO, COOCCUR)
- DSS (developmental sentence score) IPSYN (index of productive syntax)



➤ Analyse du discours, analyse conversationnelle

- Longueur moyenne des tours de parole, longueur moyenne des énoncés MLT/MLU



*MOT: is the bunny rabbit jumping ?



> **mlt sample.cha**

mlt sample.cha

Tue Apr 01 19:27:12 2003

mlt (21-Sep-2001) is conducting analyses on:

ALL speaker tiers

From file <sample.cha>

MLT for Speaker: *CHI:

MLT (xxx and yyy are INCLUDED in the utterance and morpheme counts):

Number of: utterances = 6, turns = 6, words = 8

Ratio of words over turns = 1.333

Ratio of utterances over turns = 1.000

Ratio of words over utterances = 1.333

MLT for Speaker: *MOT:

MLT (xxx and yyy are INCLUDED in the utterance and morpheme counts):

Number of: utterances = 8, turns = 7, words = 43

Ratio of words over turns = 6.143

Ratio of utterances over turns = 1.143

Ratio of words over utterances = 5.375



> mlu sample.cha

mlu sample.cha

Tue Apr 01 19:25:07 2003

mlu (21-Sep-2001) is conducting analyses on:

ALL speaker tiers

From file <sample.cha>

MLU for Speaker: *CHI:

MLU (xxx and yyy are EXCLUDED from the utterance and morpheme counts):

Number of: utterances = 5, morphemes = 7

Ratio of morphemes over utterances = 1.400

Standard deviation = 0.548

MLU for Speaker: *MOT:

MLU (xxx and yyy are EXCLUDED from the utterance and morpheme counts):

Number of: utterances = 8, morphemes = 46

Ratio of morphemes over utterances = 5.750

Standard deviation = 3.012



CHILDES

CHILDES

Child Language Data Exchange System



ExerciceMLU.cha

ExerciceMLT.cha

► Analyse du discours, analyse conversationnelle

► Longueur moyenne des tours de parole, longueur moyenne des énoncés MLT/MLU

► Décompte des séquences (GEM, GEMFREQ)



Gem Gemlist Gemfreq

➤ Marquer les Gem

- @Bg: *textecode*
- @Eg: *codegem*

➤ Lancer la commande Gemlist, ou Gemfreq

```
@Bg: PART1
*CHI: word1_1 word1_2 .
%err: $ERRO_1 $ERRO_2
@Bg: TU1
*CHI: word2_1 word2_2 .
*MOT: word3_1 word3_2 .
@Eg: TU1
@Bg: PART1
@Bg: PART2
@Bg: TU2
*CHI: word4_1 word4_2 .
%err: $BAD_ERR
@Eg: TU2
*MOT: word5_1 word5_2 .
%err: $ERR1_1 $ERR1_2
@Bg: TU3
*CHI: word6_1 word6_2 .
@Eg: TU3
*MOT: word7_1 word7_2 .
%err: $ERR1_2 $ERR2_1
@Eg: PART2
```

```
gemfreq +s"PART2" +t%err -t*
```



CHILDES

CHILDES

Child Language Data Exchange System



ExerciceGEMLIST.cha
ExerciceGEMFREQ.cha

kwai (03-Jui-2011) is conducting analyses on:
ALL speakers
Ratio of morphemes over morphemes

► Analyse du discours, analyse conversationnelle

► Longueur moyenne des tours de parole, longueur moyenne des énoncés MLT/MLU

► Décompte des séquences (GEM, GEMFREQ)

► Reprises et changements lexicaux dans l'interaction conversationnelle (adulte/enfant - CHIP)

► Fréquences, Cooccurrences (actes de langage...)

*NOT: is the bunny rabbit jumping?



Exercices

- VOCD sur lemmes
- Combo sur pro suivi de verbe
- Combo sur pro suivi de pro puis verbe
- Combo avec « veux » sur ligne principale et pro sur la ligne %mor.
- Cooccur sur det
- Cooccur sur n comme dernier élément d'un cluster de 3
- Remettre des lignes dans l'ordre



Exercices

➤ VOCD sur lemmes

- `vocd exercicevocd.cha +t%mor +s"%|*&%%" +s"%|*-%%" +s"%|*"`

➤ COMBO sur lignes différentes

- `combo exercicecombo.cha +t%mor +s"veux^*^%mor:^*^pro*|*"`

➤ Cooccur sur det

- `cooccur exerciceCOOCCUR.cha +s"det*|*" +t%mor`

➤ Cooccur sur n comme dernier élément d'un cluster de 3

- `cooccur exerciceCOOCCUR.cha +s"n:*|*" +s"n|*" +t%mor -b +n3`



kwai (03-Jui-2009) is conducting analyses on:
ALL speaker tiers

➤ Tierorder

- tierorder exercice TIEROrder.cha
- Cf tierorder.cut

*MOT: is the bunny rabbit jumping ?

