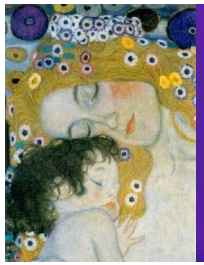
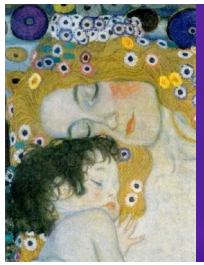


CLAN pour les Néophytes

Découvrir, comprendre et utiliser un nouvel outil de transcription et d'analyse des interactions

Stéphanie Caët
Université Sorbonne Nouvelle
stephanie.caet@gmail.com



CLAN utilisé pour ….

Linguistique interactionnelle

Linguistique clinique

Acquisition du langage

Audio/vidéo voire textes écrits

Sémantique

Morphosyntaxe

Phonologie

Analyse des gestes

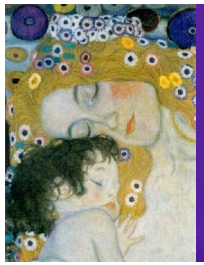
Interactions adultes-adultes

Interactions en classe

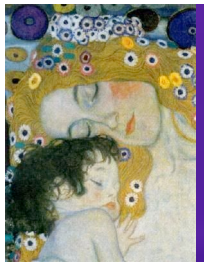
Interactions professionnelles

Interactions adultes-enfants

…



Processus de collecte de données

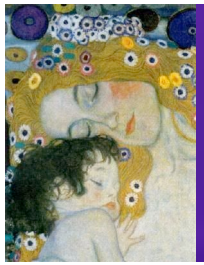


Processus de collecte de données

Enregistrement de la séance	1h voire 1/2 journée !
Numérisation et Compression du film	Quelques heures
Transcription de la vidéo et correction de la transcription	30h+10h
Analyse de la transcription et de la vidéo	Toute la vie d'un chercheur !

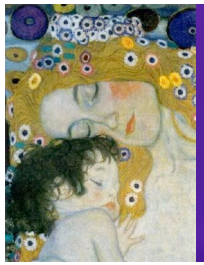
Processus long et coûteux

Données partiellement exploitées



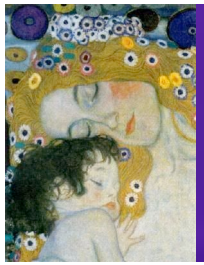
Dans cette présentation …

1. Le Projet CHILDES
2. Que peut-on faire avec CLAN ?
    1. Lire et analyser des transcriptions de la base de données CHILDES
    2. Transcrire
    3. Aligner la transcription avec la vidéo ou le son
    4. Faire des analyses automatiques sur les transcriptions



Dans cette présentation …

1. **Le Projet CHILDES**
2. Que peut-on faire avec CLAN ?
    1. Lire et analyser des transcriptions de la base de données CHILDES
    2. Transcrire
    3. Aligner la transcription avec la vidéo ou le son
    4. Faire des analyses automatiques sur les transcriptions



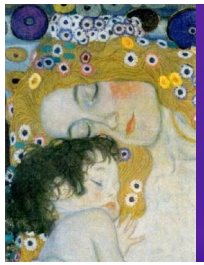
1. Le Projet CHILDES

CHIld Language Data Exchange System

Brian MacWhinney

Catherine Snow

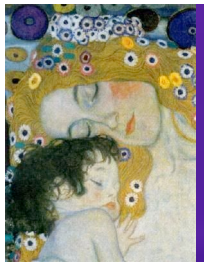
1er corpus : Brown Corpus (Adam, Sarah & Eve)
Au début, pas d'audio …



1. Le Projet CHILDES

CHIld Language Data Exchange System

- **Child Language Data …**
  - Interactions naturelles ou semi-guidées
  - 130 corpora partagés
  - dans 26 langues différentes
- **… Exchange System**
  - Base de données CHILDES avec fichiers vidéo/son et transcriptions
  - Logiciel de lecture et d'analyse : CLAN
  - Conventions de transcriptions communes : format CHAT



1. Le Projet CHILDES

Site Internet : http://childes.psy.cmu.edu

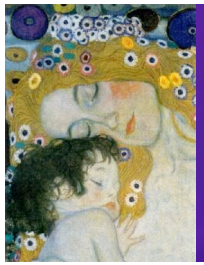
CHILDES Child Language Data Exchange System

CHILDES is the child language component of the TalkBank system. TalkBank is a system for sharing and studying conversational interactions.

System	Programs and Database
Ground rules	Downloadable Database
Guidelines for Contributors	Browsable Database
Overviews and Introductions	The CLAN Program
Membership list	Derived Corpora and Counts
How to subscribe to Mailing Lists	
Links	**Manuals**
The TalkBank Database	CHAT Transcription
IASCL information	CLAN Programs
Other Child Language sites	Database Manuals

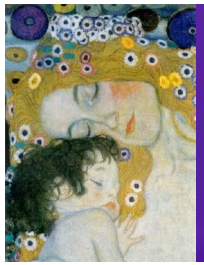
Informations Base de données

Informations CLAN et CHAT



Dans cette présentation …

1. Le Projet CHILDES
2. **Le logiciel CLAN**
    1. Lire et analyser des transcriptions de la base de données CHILDES
    2. Transcrire
    3. Aligner la transcription avec la vidéo ou le son
    4. Faire des analyses automatiques sur les transcriptions



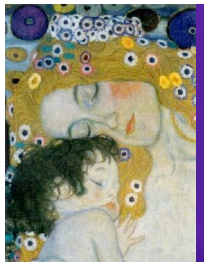
2. Le logiciel CLAN

Téléchargements sur le site de CHILDES

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 ← Logiciel
Membership list	Derived Corpora and Counts
How to subscribe to Mailing Lists	
Links	**Manuals**
The TalkBank Database	CHAT Transcription
IASCL information	CLAN Programs
Other Child Language sites	Database Manuals



2. Le logiciel CLAN

Téléchargements sur le site de CHILDES

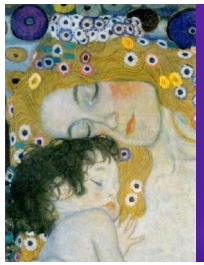
CHILDES Using CLAN

Windows	Mac	Unix
CLANWin	CLAN	UnixCLAN
V 18-Dec-2009 16:00	V 20-Dec-2009 12:00	V 18-Sep-2009 10:00

For Windows:

CLANWin is for Windows XP/2000/NT/Vista. Windows 95, 98, or ME are no longer supported. Windows installation involves clicking on the installation file and following the directions given by InstallShield. If you have an older version of CLAN on your machine, InstallShield will overwrite it with the newer version.

Also, you will need to install QuickTime and Unicode fonts.

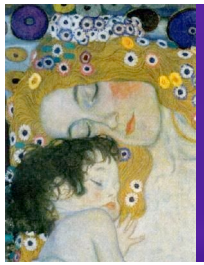


2. Le logiciel CLAN

Téléchargements sur le site de CHILDES

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	Training Videos
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	MRC lexical dictionary
Working with digitized audio	Syntactic analysis by GRASP
The Computerized Comprehension Task	ChildFREQ Site and Paper

Télécharger la grammaire de la langue étudiée



2. Le logiciel CLAN

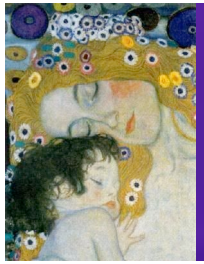
Téléchargements sur le site de CHILDES

CHILDES MOR Grammars

The MOR program provides a method for automatic tagging of corpora in the CHAT format. To make this work, it is necessary to construct a separate MOR grammar for each language. After analysis with MOR, users can then use the POST program to disambiguate the %mor line. We provide a POST disambiguation database for English, but for other languages, users will need to do the work of training a POST database for themselves. This whole system is described in a recent article on morphosyntactic analysis in CLAN.

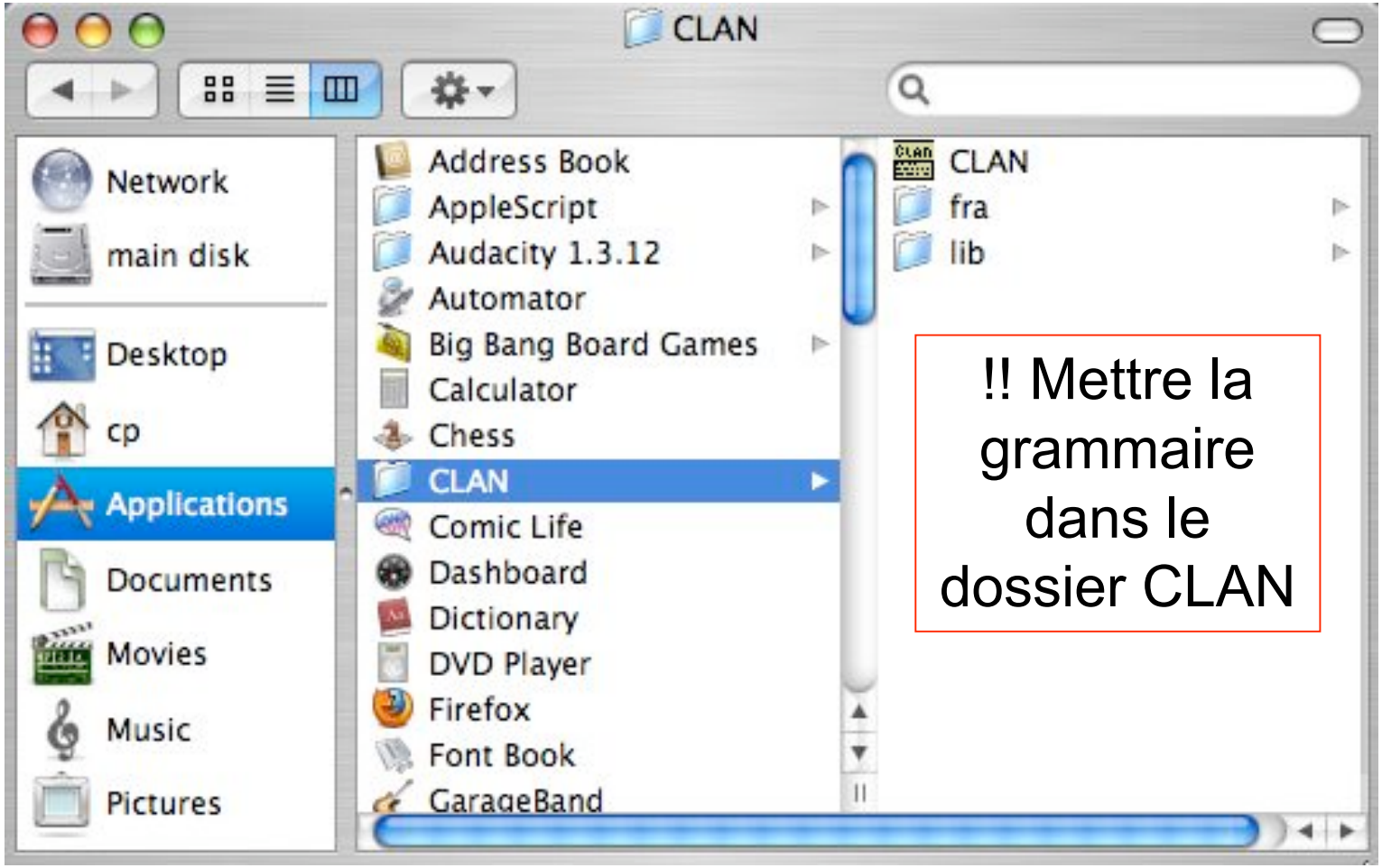
We have working MOR grammars for these languages:

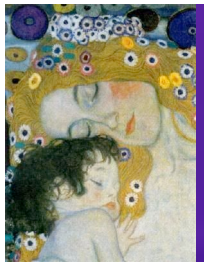
- Cantonese (yue): This grammar was built by Brian MacWhinney relying on a Cantonese-English lexicon provided by K. K. Luke.
- Chinese (zho): This grammmar was built by Brian MacWhinney and Twila Tardif. Thanks to K. J. Chen and the CKIP Group of the Academica Sinica for providing an Excel listing of the 20,000 highest frequency forms of Putonghua along with their English translations and romanizations.
- Danish (dan): This grammar was written by Brian MacWhinney
- Dutch (nld): This grammar was contributed by Steven Gillis.
- English (eng): This grammar was built initially by Brian MacWhinney and Mitzi Morris. It covers all the forms in the CHILDES English database.
- French (fra): This grammar was contributed by Christophe Parisse.
- German (deu): This grammar was contributed by Heike Behrens.
- Hebrew (heb): This grammar was contributed by Sigal Uziel-Karl and Bracha Nir-Sagiv of Tel-Aviv University in Israel.
- Japanese (jap): This grammar was constructed by Norio Naka and Susanne Miyata. The Wakachi system is helpful for reference.
- Italian (ita): This grammar was built by Livia Tonelli and Brian MacWhinney.
- Spanish (spa): This grammar was built by Brian MacWhinney and Monica Sanz



2. Le logiciel CLAN

Téléchargements sur le site de CHILDES

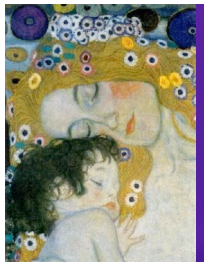




2. Le logiciel CLAN

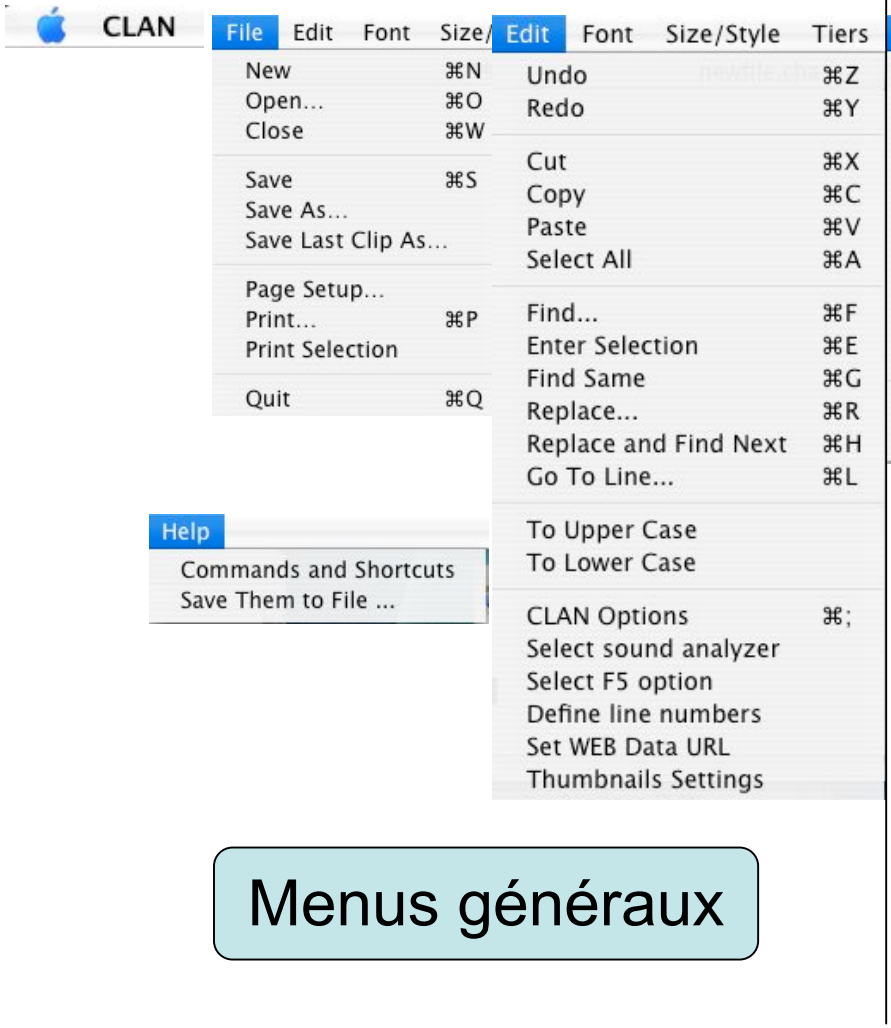
Une interface d'éditeur de texte

Quand on ouvre CLAN …



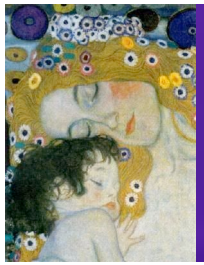
2. Le logiciel CLAN

Les menus



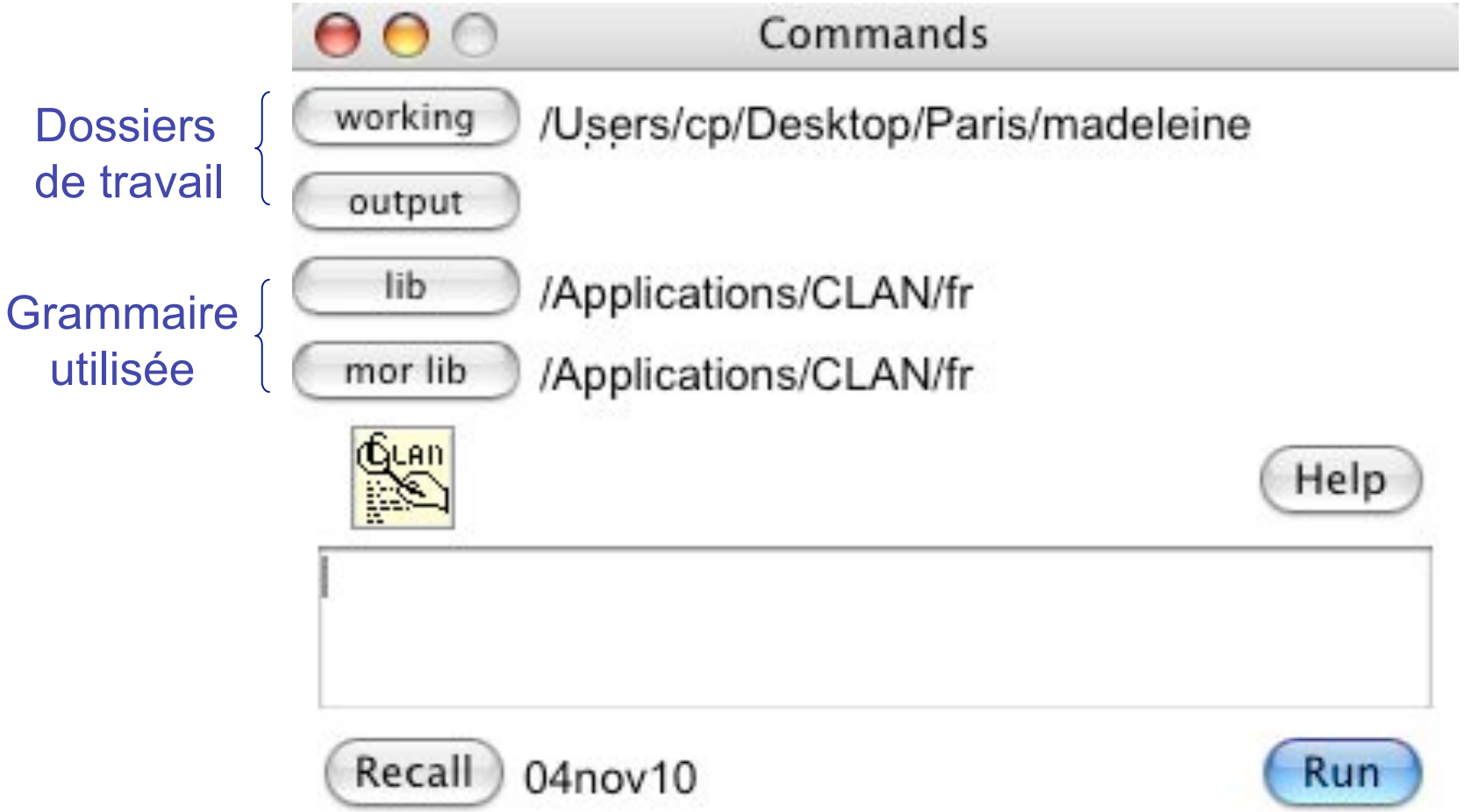
Menus généraux

Menus spécifiques à CLAN



2. Le logiciel CLAN

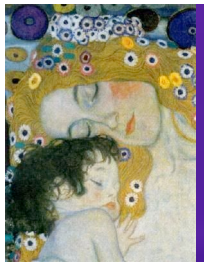
La fenêtre de commandes



Dossiers de travail

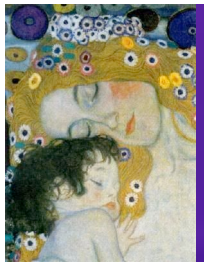
Grammaire utilisée

Pour ouvrir la fenêtre de commandes : Windows>Commands ou ctrl/ d



Dans cette présentation …

1. Le Projet CHILDES
2. **Le logiciel CLAN**
   1. **Lire et analyser des transcriptions de la base de données CHILDES**
   2. Transcrire
   3. Aligner la transcription avec la vidéo ou le son
   4. Faire des analyses automatiques sur les transcriptions

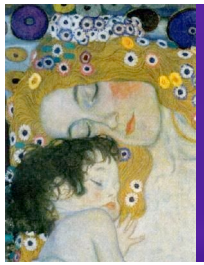


2.1 Lire des transcriptions de la base de données CHILDES

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	Derived Corpora and Counts
How to subscribe to Mailing Lists	
Links	**Manuals**
The TalkBank Database	CHAT Transcription
IASCL information	CLAN Programs
Other Child Language sites	Database Manuals ← Descriptif

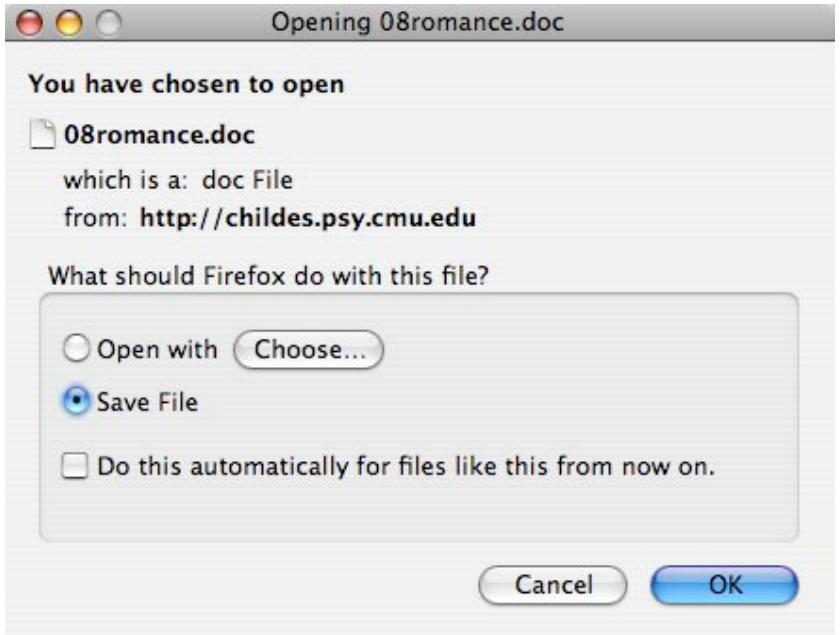


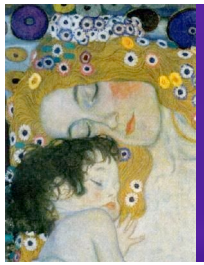
2.1 Lire des transcriptions de la base de données CHILDES

CHILDES The Database Manuals

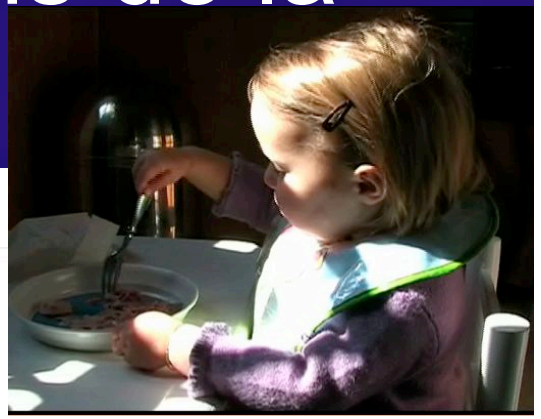
The guides to the CHILDES database are divided by subject matter into these files.

- Introduction to the Database
- American English
- British English
- Bilingual Acquisition
- Language Disorders
- Narratives
- Germanic Languages
- Romance Languages
- Slavic Languages
- East Asian Languages
- Celtic Languages
- Other Languages





2.1 Lire des transcriptions de la base de données

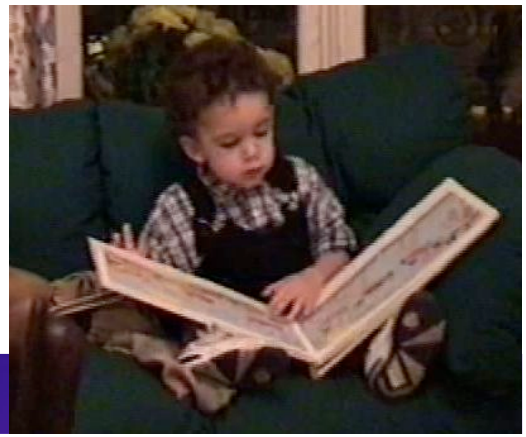


Quand on ouvre le fichier 08Romance.doc …

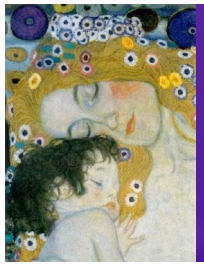
Romance Corpora

This is a guide to CHILDES data on the acquisition of Romance languages. For a general introduction to the CHILDES database, please consult intro.pdf. The links in the table below are clickable, as are the thumbnails to the left.

Corpus	Age Range	N	Comments
French – Champaud on page 4	1;9.18–2;5.27	1	Longitudinal study of language development in French with recordings made over a 28-month period
French – Geneva page 9		1	Case study of the child Marie
French – Hammelrath page 11	3;6 -5;6	489	Test samples from a wide preschool population
French – Hunkeler page 12	1-3	2	Dizygotic twins
French – Kern page 13	0;7-2;0	4	Phonological development
French - Leveille page 14	2;1.19–3;3.12	1	Longitudinal study with weekly recording sessions
French – Lyon on page 16	1-3	6	Longitudinal study with linked media for studying phonological development
French – Montreal page 17	5;6–11;6	36 s	Cross sectional study focusing on speech acts in Kindergarten, third- and fourth-grade children playing "veterinarian" and "assistant"
French – Paris on page 21	0;7-3;5	3	Longitudinal video study of three children in Paris
French – Pauline page 21	1;2-2;6	1	Case study



Télécharger les manuels de base de données



2.1 Lire des transcriptions de la base de données CHILDES

Références de CHILDES
(MacWhinney, 2000)
+
Références corpus

Références du corpus

English – USA

31. Providence

Katherine Demuth
Department of Linguistics
Macquarie University
Sydney, NSW 2109 Australia
katherine.demuth@ling.mq.edu.au

Katherine Demuth and her research assistants at Brown University compiled the Providence corpus from 2002-2005. The corpus contains longitudinal audio/video recordings of 6 monolingual English-speaking children's language development from 1-3 years during spontaneous interactions with their parents (usually their mothers) at home. The aim of the study was to provide a corpus of phonetically transcribed data, with linked acoustic files, for the purpose of studying early phonological and morphological development. Collection and transcription of the Providence Corpus (and the similar Lyon Corpus for French) was supported by NIMH grant #1ROIMH60922.

Participants

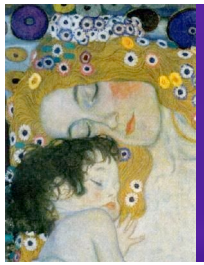
The participants included 3 boys (Alex, Ethan, William) and 3 girls (Lily, Naima, Violet). Each child was recorded for 1 hour every 2 weeks beginning at the onset of first words. Two of the children have denser corpora, with weekly recordings from 1;3-2;10 (Naima) and 2;0-3;0 (Lily). The three girls (Lily, Naima, and Violet) were also recorded monthly from 3-4 years. The total corpus consists of 364 hours of speech. Audio is available for all children and video is available for all children except Ethan, who was diagnosed with Aspergers Syndrome at the age of 5.

Transcription

Both adult and child utterances were orthographically transcribed using CLAN conventions, with the audio/video files linked. Trained transcribers then carried out a broad phonemic (SAMPA > Unicode) transcription of the child utterances. A second trained coder then retranscribed 10% of each recording. Reliability scores ranged from 80%-98% (discounting voicing errors) on this second segment.

Please cite the following reference when using the corpus:

Demuth, K., Culbertson, J. & Alter, J. 2006. Word-minimality, epenthesis, and coda licensing in the acquisition of English. *Language & Speech*, 49, 137-174.



2.1 Lire des transcriptions de la base de données CHILDES

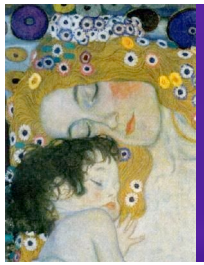
CHILDES Child Language Data Exchange System

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	Derived Corpora and Counts
How to subscribe to Mailing Lists	
Links	**Manuals**
The TalkBank Database	CHAT Transcription
IASCL information	CLAN Programs
Other Child Language sites	Database Manuals

Fichiers

Descriptif



2.1 Lire des transcriptions de la base de données CHILDES

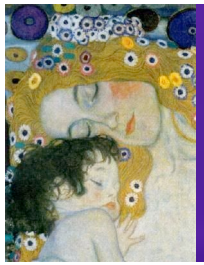
CHILDES The CHILDES Database

The CHILDES database contains transcript and media data collected from conversations between young children and their playmates and caretakers. Conversations with older children and adults at available from TalkBank. All of the data is transcribed in CHAT and CA/CHAT formats. The use of all CHILDES and TalkBank data is governed by the Gnu Public License (GPL). Please remember to read the database manuals and to follow the guidelines for data-sharing.

Downloading TalkBank Data

- **Downloadable Transcripts**
- **Downloadable Audio and Video**

!! Mettre transcriptions et vidéos dans le même dossier

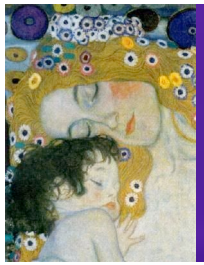


2.1 Lire des transcriptions de la base de données CHILDES

Using local transcripts and local media

1. You need to download the transcripts from "zipped transcripts" below directory of zipped XML data files.
2. If the corpus is linked to audio or video media, you need to download
3. You need to download and install the CLAN program.
4. To open a transcript, you double-click on it. If there is associated med

Zipped Transcripts
English - USA
English - UK
Celtic
East Asian
Germanic
Romance
Slavic
Other Languages
Bilinguals
Clinical



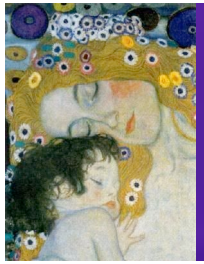
2.1 Lire des transcriptions de la base de données CHILDES

Index of /data/Romance

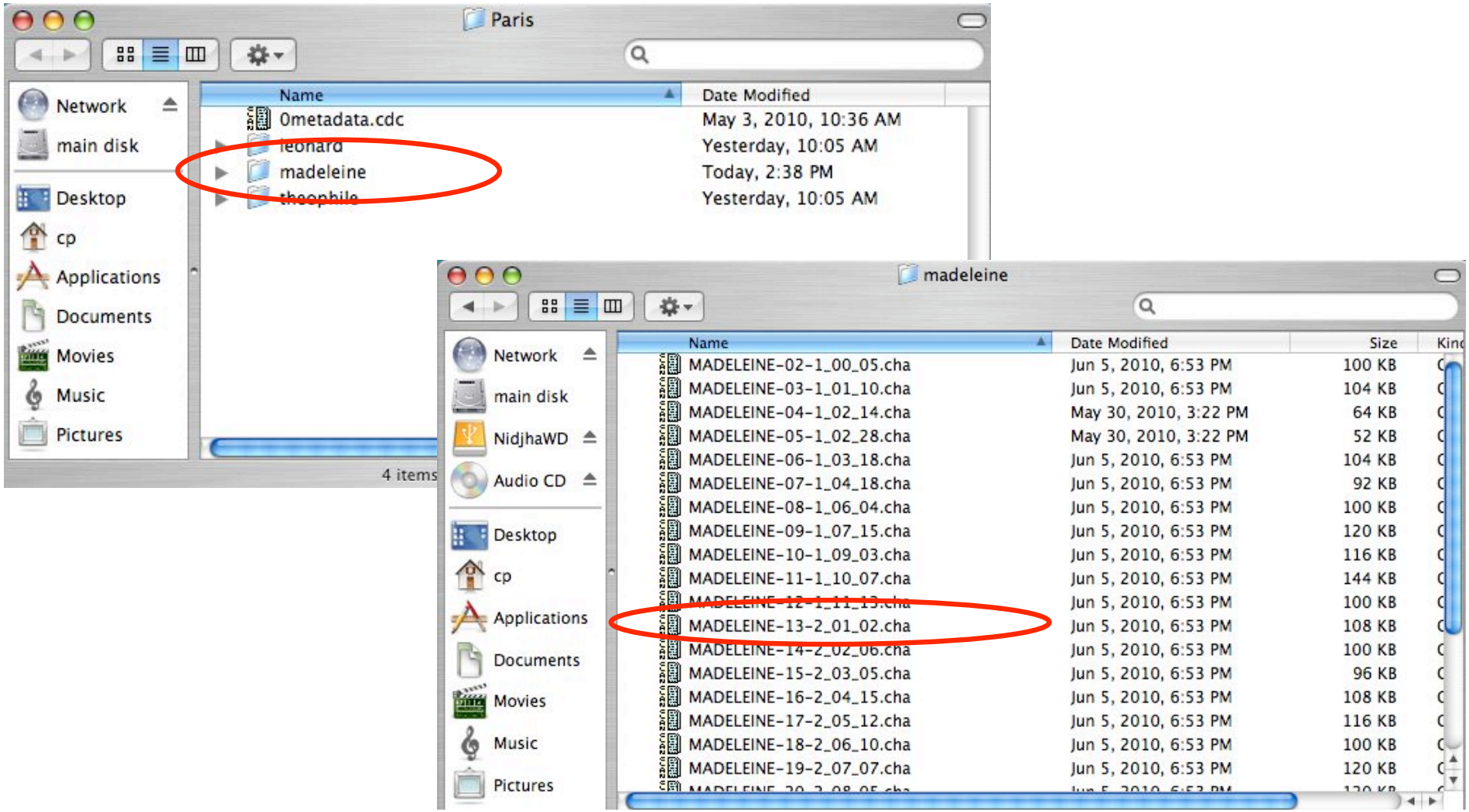
Name	Last modified	Size	Description
Parent Directory		-	
Catalan/	13-Jul-2010 20:49	-	
French/	23-Oct-2010 19:16	-	
Italian/	13-Jul-2010 19:53	-	
Portuguese/	13-Jul-2010 21:44	-	
Romanian/	20-May-2010 13:49	-	
Spanish/	22-Aug-2010 12:42	-	

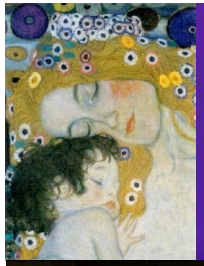
Index of /data/Romance/French

Name	Last modified	Size	Description
Parent Directory		-	
Champaud.zip	13-Jul-2010 19:53	144K	
Geneva.zip	13-Jul-2010 21:44	94K	
Hammelrath.zip	24-Jun-2010 05:18	490K	
Hunkeler.zip	13-Jul-2010 19:53	97K	
Leveille.zip	13-Jul-2010 21:44	390K	
Lyon.zip	23-Oct-2010 19:16	3.8M	
Montreal.zip	24-Jun-2010 05:17	108K	
Paris.zip	19-Jul-2010 23:52	1.9M	
Pauline.zip	30-May-2010 14:28	222K	
Rondal.zip	30-May-2010 14:28	674K	
York.zip	30-May-2010 14:28	1.6M	

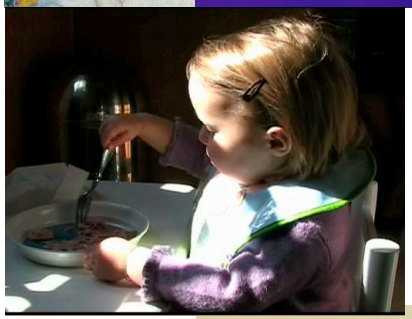


2.1 Lire des transcriptions de la base de données CHILDES

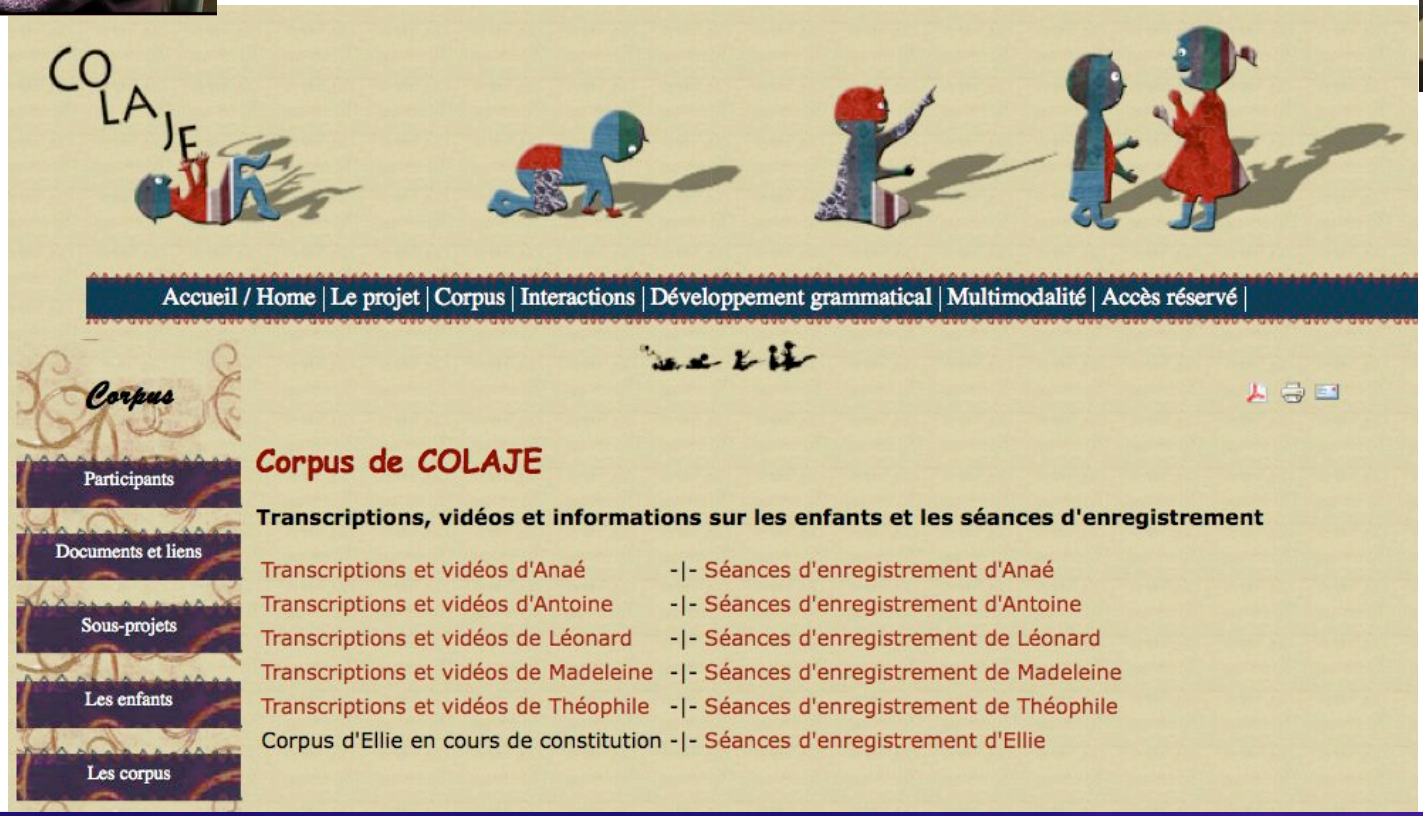


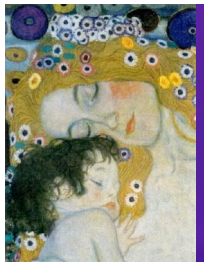


2.1 Lire des transcriptions de la base de données CHILDES



Madeleine + Anaé bientôt disponibles
jusqu'à 4 ans sur CHILDES
Déjà disponibles sur http://colaje.risc.cnrs.fr :





2.1 Lire des transcriptions de la base de données CHILDES

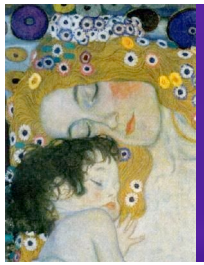
CHILDES The CHILDES Database

The CHILDES database contains transcript and media data collected from conversations between young children and their playmates and caretakers. Conversations with older children and adults at available from TalkBank. All of the data is transcribed in CHAT and CA/CHAT formats. The use of all CHILDES and TalkBank data is governed by the Gnu Public License (GPL). Please remember to read the database manuals and to follow the guidelines for data-sharing.

Downloading TalkBank Data

- **Downloadable Transcripts**
- **Downloadable Audio and Video**

!! Mettre transcriptions et vidéos dans le même dossier



2.1 Lire des transcriptions de la base de données CHILDES

Index of /media

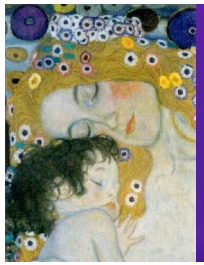
Name	Last modified	Size	Descriptio
Parent Directory		-	
Biling/	02-Jul-2009 12:52	-	
Clinical/	07-Feb-2006 13:31	-	
EastAsian/	18-Sep-2010 13:10	-	
Eng-UK/	30-Jan-2010 12:35	-	
Eng-USA/	19-Aug-2010 12:23	-	
Germanic/	02-Jun-2008 18:30	-	
Narrative/	20-May-2005 14:26	-	
Other/	11-Nov-2010 16:09	-	
PDF/	08-Dec-2007 16:12	-	
Password/	30-Jan-2010 13:13	-	
PhonBank-wav/	16-Sep-2010 11:34	-	
PhonBank/	16-Sep-2010 12:21	-	
Romance/	10-Oct-2010 15:01	-	
Slavic/	11-Oct-2005 17:06	-	
Unlinked/	10-Oct-2010 15:00	-	

Index of /media/Romance

Name	Last modified	Size	Description
Parent Directory		-	
French/	05-Mar-2009 15:45	-	
Italian/	19-Nov-2006 15:19	-	
Portuguese/	20-May-2005 15:17	-	
Spanish/	15-Sep-2008 17:33	-	

Index of /media/Romance/French

Name	Last modified	Size	Description
Parent Directory		-	
Lyon/	20-Sep-2007 19:05	-	
Paris/	05-Mar-2009 14:15	-	



2.1 Lire des transcriptions de la base de données CHILDES

Index of /media/Romance/French/Paris

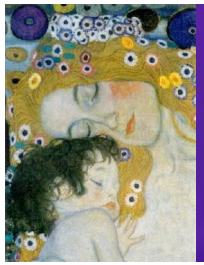
Name	Last modified	Size	Description
Parent Directory		-	
Leonard/	12-Mar-2009 17:45	-	
Madeleine/	12-Mar-2009 17:45	-	
Theophile/	12-Mar-200		

!! Mettre transcriptions et vidéos dans le même dossier

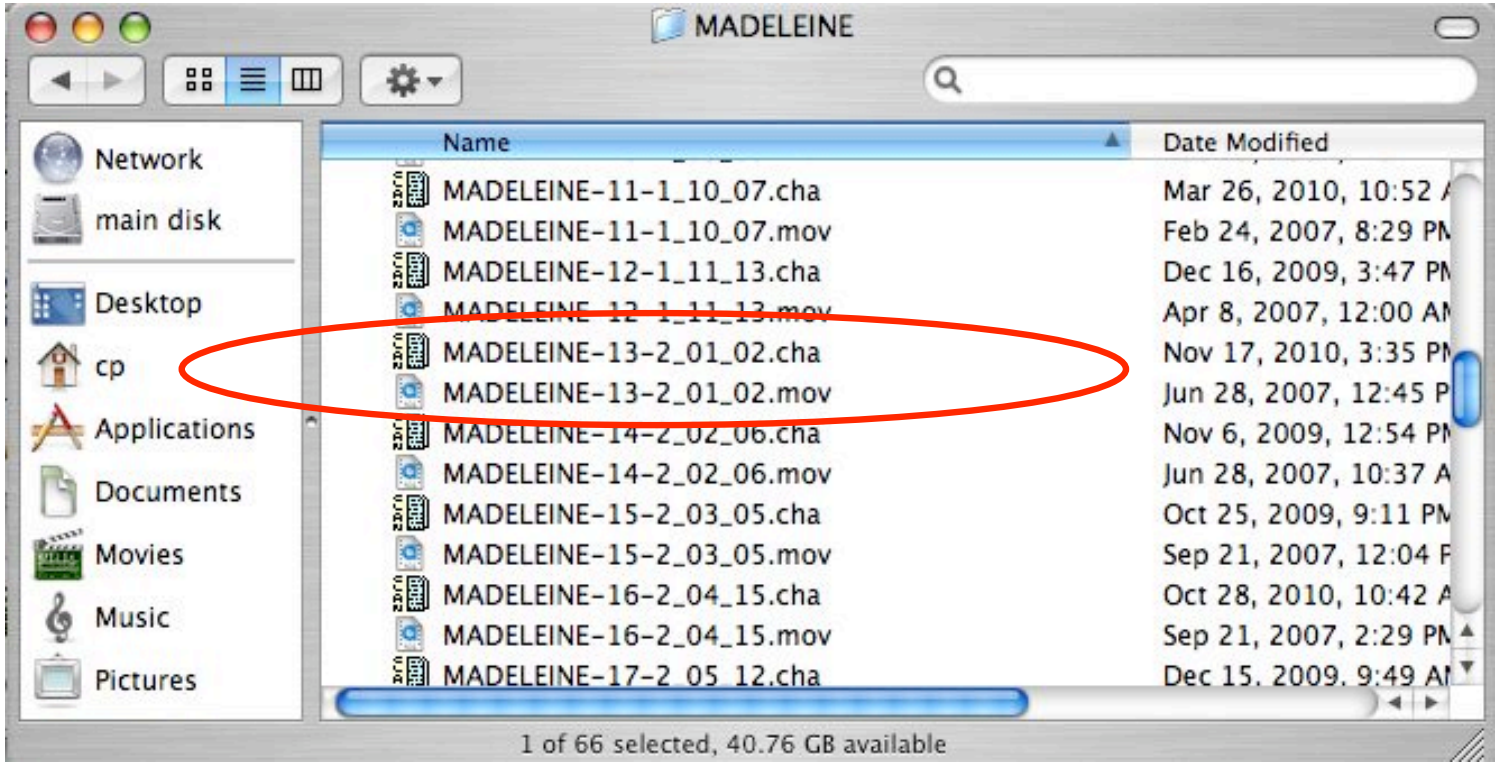
Index of /media/Romance/French/Paris/Madeleine

Name	Last modified	Size	Description
Parent Directory		-	
01.mov	05-Apr-2006 10:08	2.1G	
02.mov	21-Dec-2008 23:59	915M	
03.mov	22-Dec-2008 00:00	927M	
04.mo	08 00:06	804M	
05.mo	08 08:21	848M	
06.mo	08 00:11	915M	
07.mo	08 00:07	742M	
08.mo	08 00:09	882M	
09.mo	08 00:10	890M	
10.mo	07 11:51	939M	

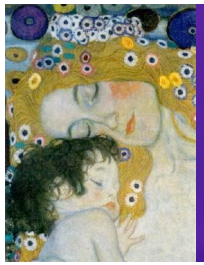
Open Link in New Window
Open Link in New Tab
Bookmark This Link
Save Link As...
Send Link...
Copy Link Location
Properties



2.1 Lire des transcriptions de la base de données CHILDES



!! Mettre transcriptions et vidéos dans le même dossier



2.1 Lire des transcriptions de la base de données CHILDES

Quand on ouvre un fichier CLAN …

CLAN File Edit Font Size/Style Tiers Mode Windows Help

/Users/cp/Desktop/Paris/madeleine/MADELEINE-13-2_01_02-

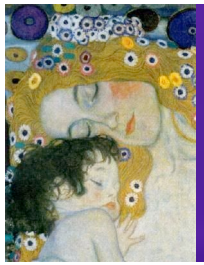
```
@Begin
@Languages:    fra
@Participants:  CHI Madeleine Target_Child, MOT Mother, OBS Martine Observer, UNI Unidentified
@ID:    fra|Paris-Corpus_Madeleine|CHI|2;01.02|female|||Target_Child||
@ID:    fra|Paris-Corpus_Madeleine|MOT||female|||Mother||
@ID:    fra|Paris-Corpus_Madeleine|OBS||female|||Observer||
@ID:    fra|Paris-Corpus_Madeleine|UNI|||||Unidentified||
@Birth of CHI:  14-APR-2005
@Media: MADELEINE-13-2_01_02, video
@Date:  16-MAY-2007
@Time Duration: 00:00:00-01:02:30
@Location:      Madeleine's home
@Comment:       code'"r - Stéphanie Caët
@G: dans la cuisine
*CHI:   0.
%act: se dirige vers la table
*MOT:   et ta table elle est propre ? •
*CHI:   oui . •
%pho:   wi
*OBS:   délicieux c(e) café (.) merci beaucoup ! •
%add: OBS s'adresse à MOT
%sit:   CHI nettoie sa table
*CHI:   moi j(e) nettoie . •
%pho:   mwa ʒ etwa
*CHI:   moi je nettoie moi . •
%pho:   mwa ʒə letwa mwa
*MOT:   tu nettoies ? •
04nov10[E|CHAT] * 13
```

En-têtes (headers)

* Lignes principales

% Lignes secondaires

Balises de temps



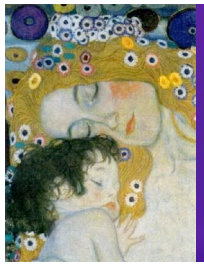
2.1 Lire des transcriptions de la base de données CHILDES

CLAN File Edit Font Size/Style Tiers Mode Windows Help

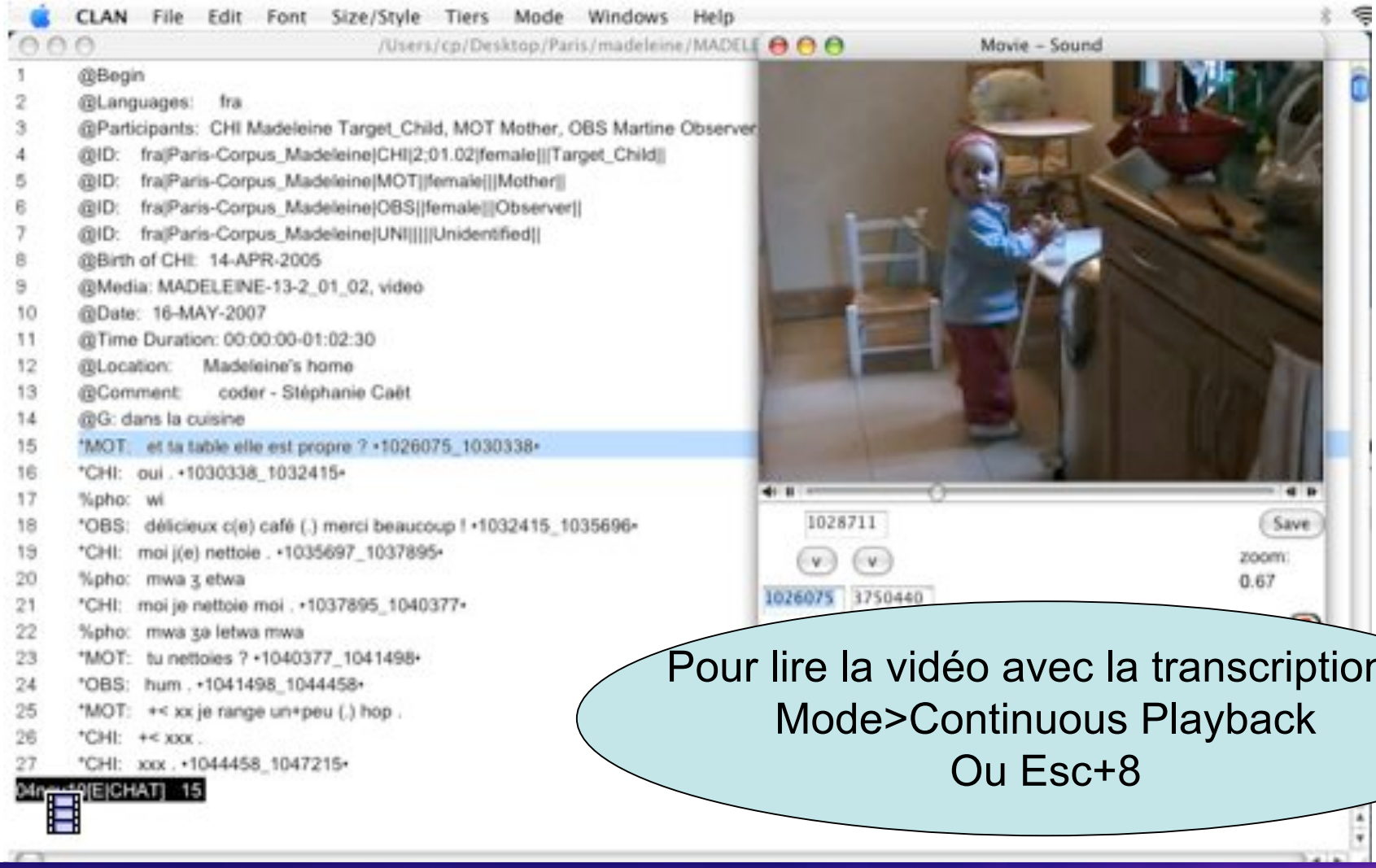
/Users/cp/Desktop/Paris/madeleine/MADELEINE-13-2_01_02-

@Begin
@Languages: fra
@Participants: CHI Madeleine Target_Child, MOT Mother, OBS Martine Observer, UNI Unidentified
@ID: fra|Paris-Corpus_Madeleine|CHI|2;01.02|female|||Target_Child||
@ID: fra|Paris-Corpus_Madeleine|MOT||female|||Mother||
@ID: fra|Paris-Corpus_Madeleine|OBS||female|||Observer||
@ID: fra|Paris-Corpus_Madeleine|UNI|||||Unidentified||
@Birth of CHI: 14-APR-2005
@Media: MADELEINE-13-2_01_02, video
@Date: 16-MAY-2007
@Time Duration: 00:00:00-01:02:30
@Location: Madeleine's home
@Comment: code"r - Stéphanie Caët
@G: dans la cuisine
*CHI: 0.
%act: se dirige vers la table
*MOT: et ta table elle est propre ? •10
*CHI: oui . •1030338_1032415•
%pho: wi
*OBS: délicieux c(e) café (.) merci b ucoup ! •1032415_1035696•
%add: OBS s'adresse à MOT
%sit: CHI nettoie sa table
*CHI: moi j(e) nettoie . •1035697_1 895•
%pho: mwa ʒ etwa
*CHI: moi je nettoie moi . •1037895_1040377•
%pho: mwa ʒə letwa mwa
*MOT: tu nettoies ? •1040377_1041498•
04nov10[E|CHAT] * 13

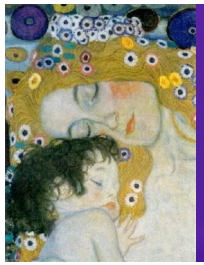
Pour afficher les balises de temps:
Mode>Expand bullets
Ou Esc+a



2.1 Lire des transcriptions de la base de données CHILDES

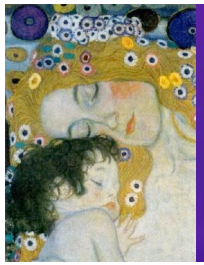


Pour lire la vidéo avec la transcription :
Mode>Continuous Playback
Ou Esc+8



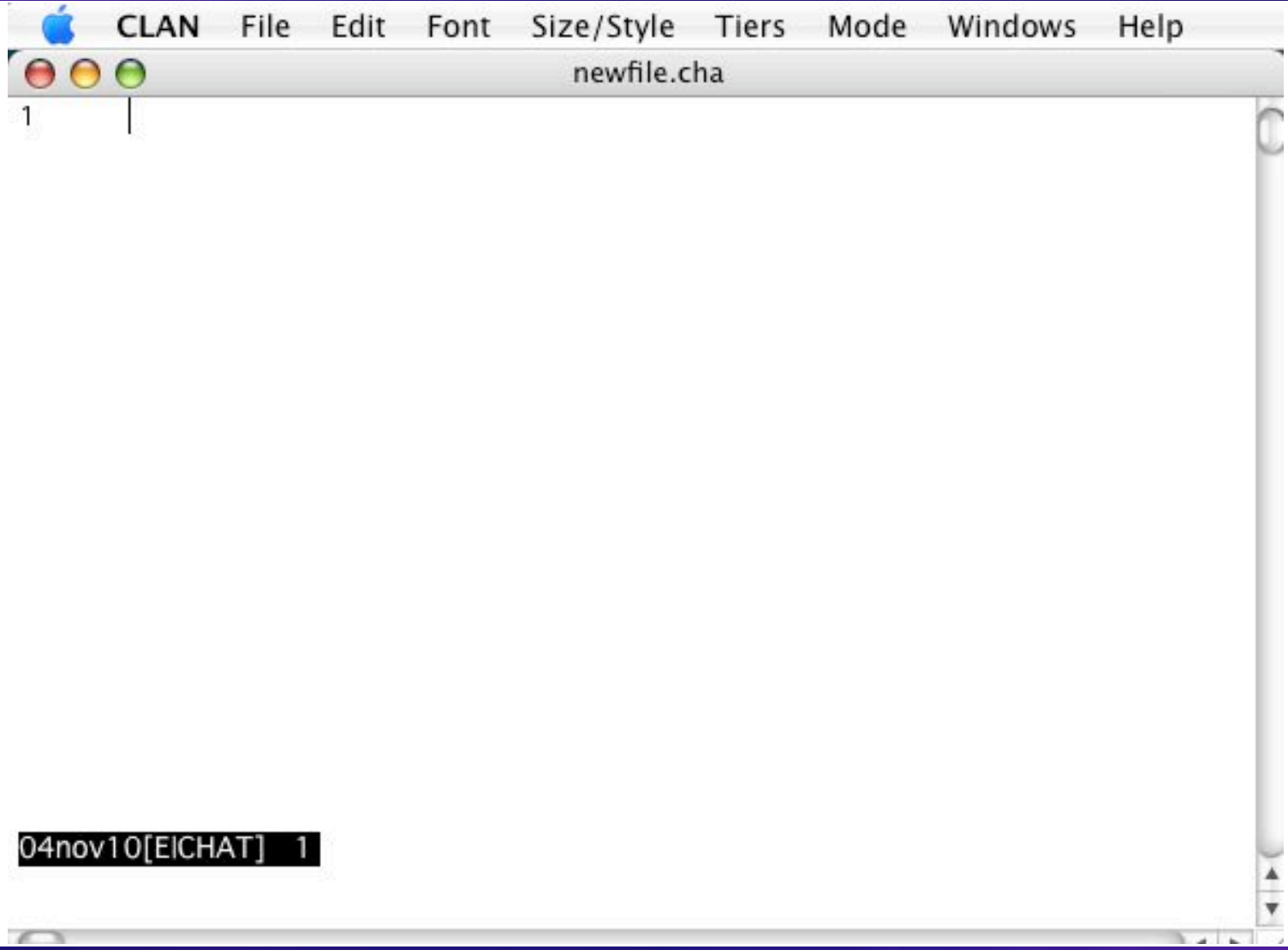
Dans cette présentation …

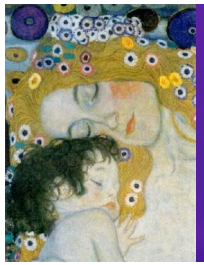
1. Le Projet CHILDES
2. **Le logiciel CLAN**
    1. Lire et analyser des transcriptions de la base de données CHILDES
    2. **Transcrire**
    3. Aligner la transcription avec la vidéo ou le son
    4. Faire des analyses automatiques sur les transcriptions



2.2 Transcrire avec CLAN

La feuille de texte





2.2 Transcrire avec CLAN

Les en-têtes (méta-données)

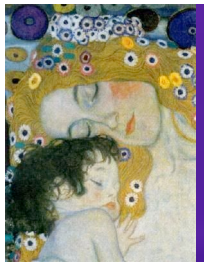
Mettre une tabulation après les :

@Begin
@Languages: fr
@Participants: CHI Madeleine Target_Child, MOT Mother, OBS Martine Observer
@ID: fr|Paris-Corpus_Madeleine|CHI|2;01.02|female|||Target_Child||
@ID: fr|Paris-Corpus_Madeleine|MOT||female|||Mother||
@ID: fr|Paris-Corpus_Madeleine|OBS||female|||Observer||
@Birth of CHI: 14-APR-2005
@Media: MADELEINE-13-2_01_02, video

Lien Vidéo

@Date: 16-MAY-2007
@Time Duration: 00:00:00-01:02:30
@Location: Madeleine's home
@Comment: coder - Stéphanie Caët

Pour insérer les ID rapidement :
« check » : Esc+I



2.2 Transcrire au format CHAT

La ligne principale : *COD:

*CHI: maman ə@fs veux ça s'i(l)+te+plaît [=! pleure]. •

- Pour les interactions adulte-enfant : découpage en énoncés => MLU

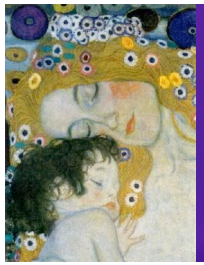
  Parfois problématique …

  *CHI: moi je veux pas jouer +…

  *CHI: parce que moi je veux lire un livre.

  OU

  *CHI: moi je veux pas jouer (.) parce que moi je veux lire un livre.



2.2 Transcrire au format CHAT

La ligne principale : *COD:

*CHI: maman ə@fs veux ça s'i(l)+te+plaît [=! pleure]. •

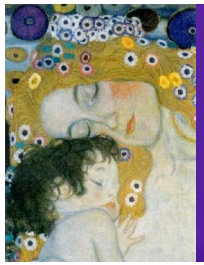
- Pour les interactions adulte-enfant : découpage en énoncés
- Transcription en orthographe => analyses

  Problématique aussi …

  *CHI: tombé / tombée / tombés / tomber.

  %pho: tɔ̃be

  Mais parfois nécessaire ….



2.2 Transcrire au format CHAT

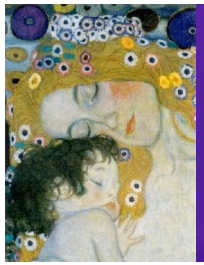
La ligne principale : *COD:

*CHI: tigre.
%pho: ti:lʁ

*CHI: tigre.
%pho: ti:g

*CHI: un tigre.
%pho: ɛ̃ ti:lg

*CHI: faut trouver du tigre.
%pho: fo kruve dy ti:ʁ



2.2 Transcrire au format CHAT

La ligne principale : *COD:

*CHI: maman ə@fs veux ça s'i(l)+te+plaît [=! pleure]. •

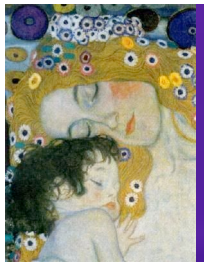
- Pour les interactions adulte-enfant : découpage en énoncés
- Transcription en orthographe
- Transcription adaptée à l'oral et à CLAN

  (.) pauses

  (élisions)

  mots+composés

  +< chevauchements



2.2 Transcrire au format CHAT

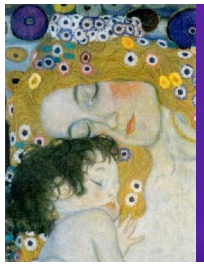
La ligne principale : *COD:

*CHI: maman ə@fs veux ça s'i(l)+te+plaît [=! pleure]. •

- Pour les interactions adulte-enfant : découpage en énoncés
- Transcription en orthographe
- Transcription adaptée à l'oral et à CLAN
- Adaptations au langage de l'enfant

  ə@fs veux

  un tiktik@c



2.2 Transcrire au format CHAT

Les lignes secondaires : %cod

*CHI: maman ə@fs veux ça s'i(l)+te+plaît [=! pleure].

%pho: mamɑ̃ ə vø sa sitəplɛ

%act: se dirige vers l'étagère

%xpnt: CHI pointe du chocolat

%sit: MOT le suit

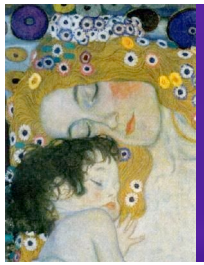
…

Transcription phonétique

Transcription du non verbal

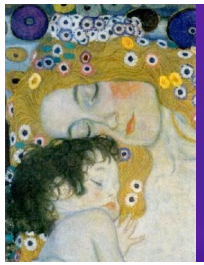
Lignes secondaires liées à la ligne principale

Lignes secondaires adaptables selon sujet de recherche



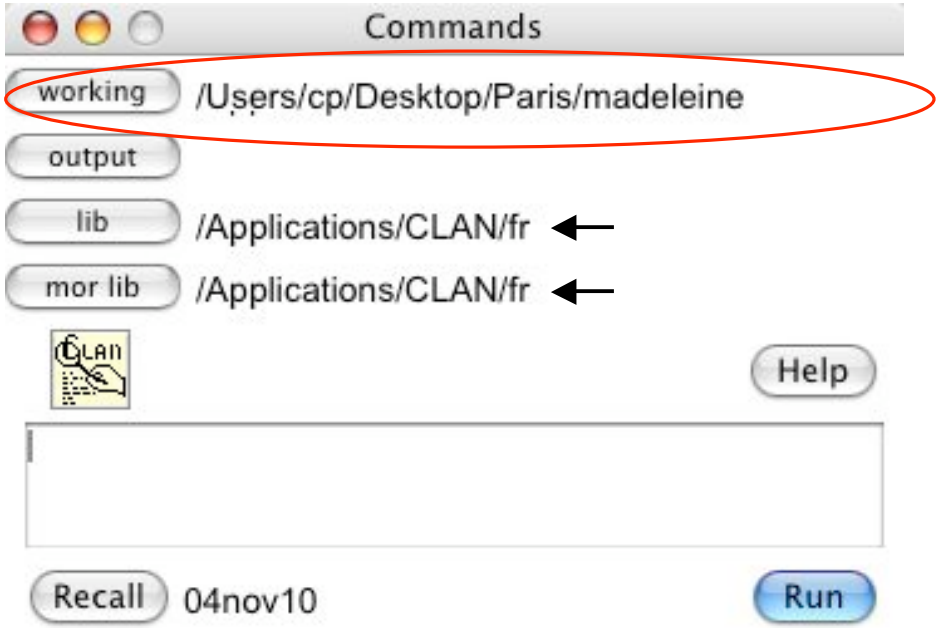
Dans cette présentation …

1. Le Projet CHILDES
2. **Le logiciel CLAN**
    1. Lire et analyser des transcriptions de la base de données CHILDES
    2. Transcrire
    3. **Aligner la transcription avec la vidéo ou le son**
    4. Faire des analyses automatiques sur les transcriptions

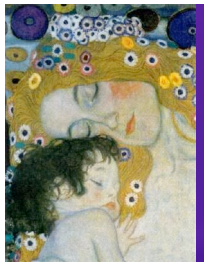


2.2 Aligner la vidéo et la transcription avec CLAN

1. Ouvrir CLAN. Le fenêtre de commandes s'ouvre (sinon : ctrl/ +d)
2. Dans la fenêtre de commandes, spécifier le dossier dans lequel se trouve la vidéo à transcrire et spécifier la grammaire



3. Fermer la fenêtre de commandes



2.2 Aligner la vidéo et la transcription avec CLAN

4. Ouvrir un nouveau document (File>New)
5. L'enregistrer aussitôt dans le dossier où se trouve la vidéo (Save as)
6. Remplir les en-têtes

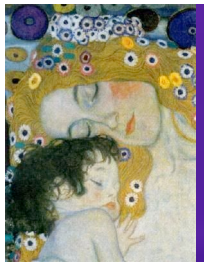
CLAN File Edit Font Size/Style Tiers Mode Windows Help

/Users/cp/Desktop/Paris/madeleine/MADELEINE-13-2_01_02.cha

```
1   @Begin
2   @Languages:    fra
3   @Participants:  CHI Madeleine Target_Child, MOT Mother, OBS Martine Observer, UNI Unidentified
4   @ID:    fra|Paris-Corpus_Madeleine|CHI|2;01.02|female|||Target_Child||
5   @ID:    fra|Paris-Corpus_Madeleine|MOT|||female|||Mother||
6   @ID:    fra|Paris-Corpus_Madeleine|OBS|||female|||Observer||
7   @ID:    fra|Paris-Corpus_Madeleine|UNI|||||Unidentified||
8   @Birth of CHI:  14-APR-2005
9   @Media: 13, video
10  @Date:  16-MAY-2007
11  @Time Duration: 00:00:00-01:02:30
12  @Location:     Madeleine's home
13  @Comment:      coder - Stéphanie Caët
14
15
16
17
18
```

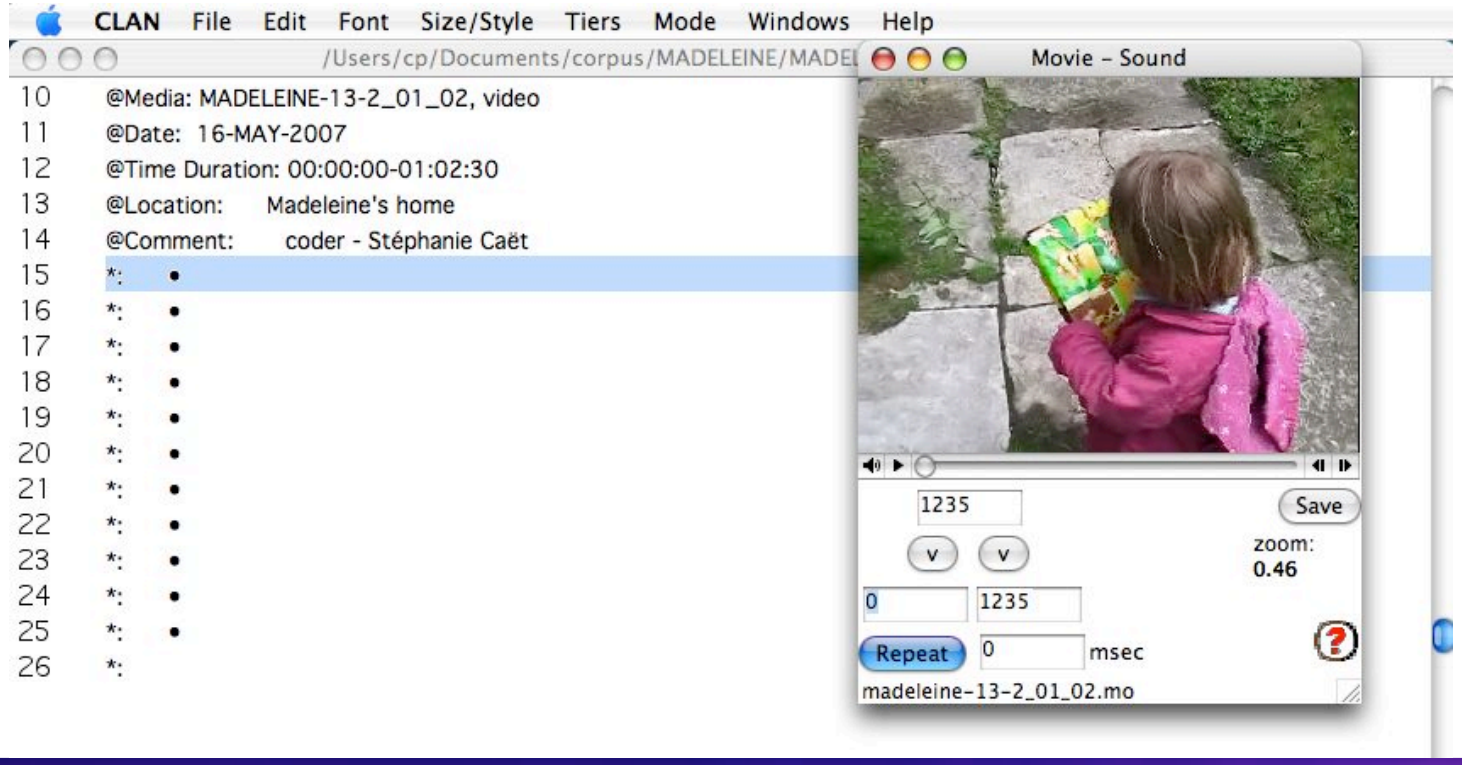
Mettre une tabulation après les :

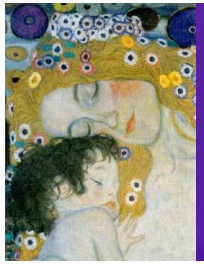
Pour insérer les ID rapidement : « check » : Esc+I



2.2 Aligner la vidéo et la transcription avec CLAN

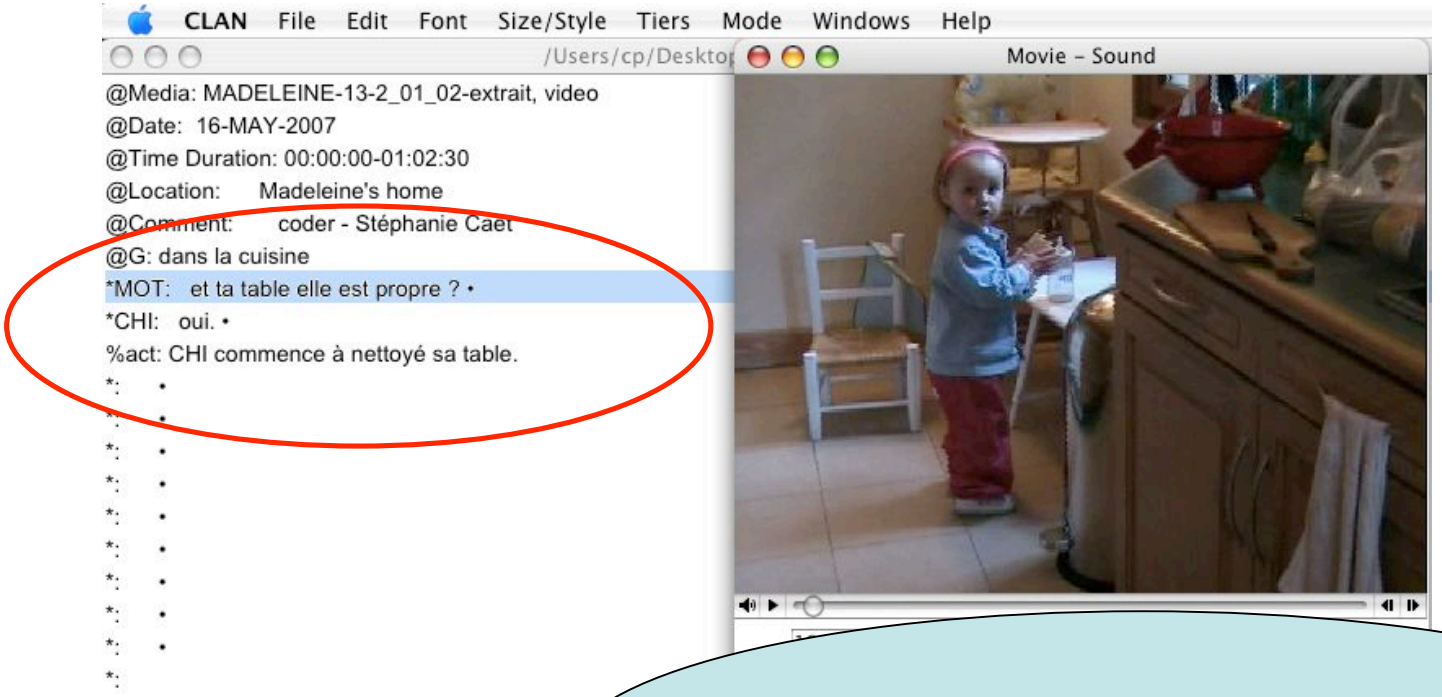
7. Lancer la vidéo et le processus d'alignement
   Mode>transcribe sound or movie (ou F5)
8. Appuyer sur Espace pour découper en énoncés et insérer les balises de temps



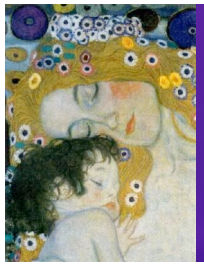


2.2 Aligner la vidéo et la transcription avec CLAN

9. Réécouter chaque balise (Mode>Play bullet Media ou F4 ou ctrl/ clic)
10. Insérer le nom du locuteur (ctrl+1 ou ctrl+2 …)
11. Et transcrire les énoncés juste devant les balises et le non verbal sous les balises

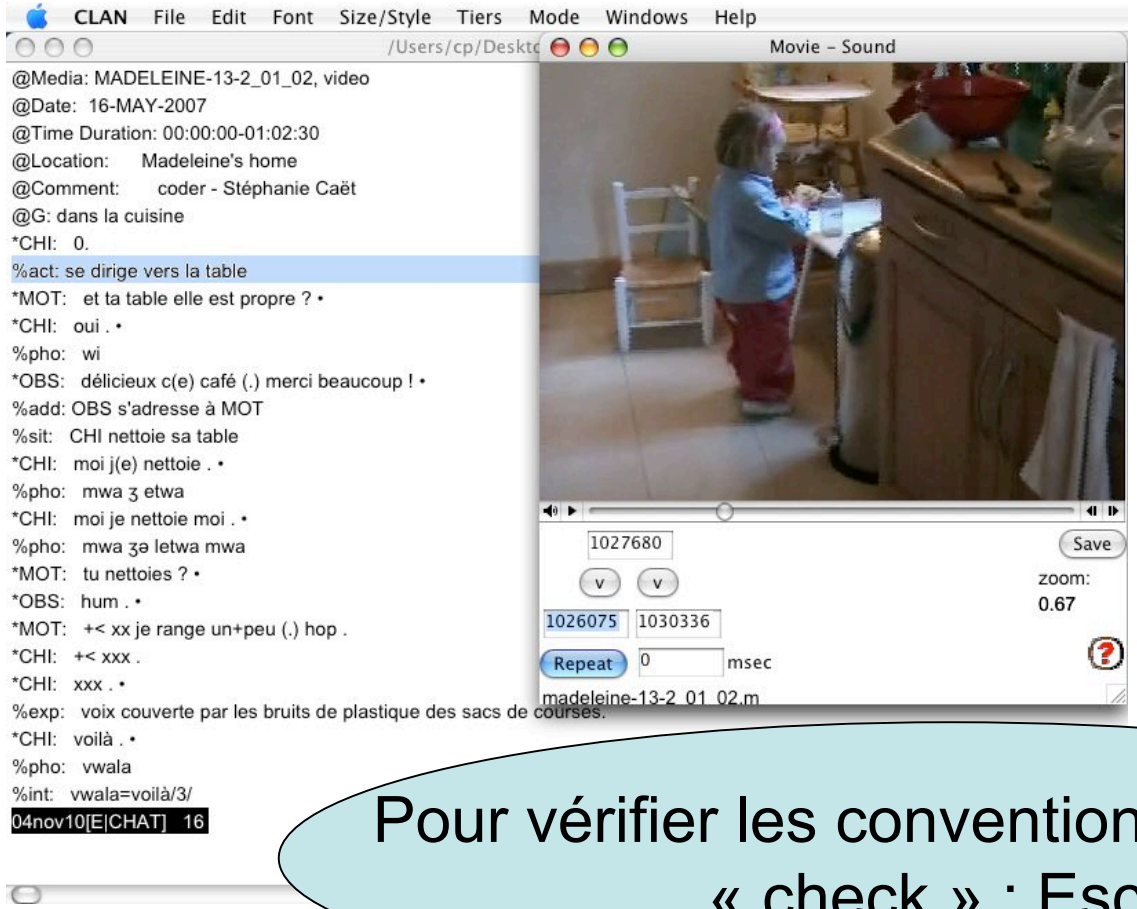


Raccourcis pour transcrire les participants : Tiers>Update puis ctrl+1/ctrl+2/ctrl+3 …

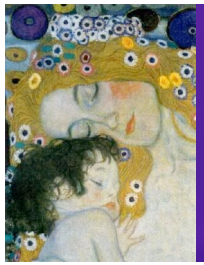


2.2 Aligner la vidéo et la transcription avec CLAN

12. Pour réécouter l'ensemble : Mode>Continuous Playback ou (Esc+8)
13. That's all !

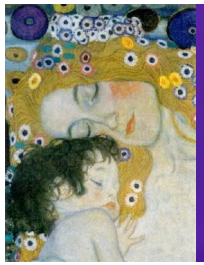


Pour vérifier les conventions pour CLAN:
« check » : Esc+l



Dans cette présentation …

1. Le Projet CHILDES
2. **Le logiciel CLAN**
    1. Lire et analyser des transcriptions de la base de données CHILDES
    2. Transcrire
    3. Aligner la transcription avec la vidéo ou le son
    4. **Faire des analyses automatiques sur les transcriptions**



2.3 Analyses automatiques
Les commandes de base

MLU

Nombre total d'énoncés
Nombre total de morphèmes ou de mots
MLU : Longueur Moyenne des énoncés

FREQ

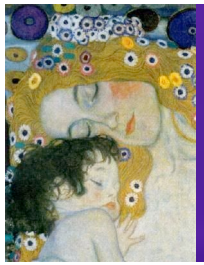
Nombre total de mots et de mots différents
Type/Token ratio
Fréquences d'items

COMBO

Combinaisons de critères de recherche
Fréquence des combinaisons

MOR|POST

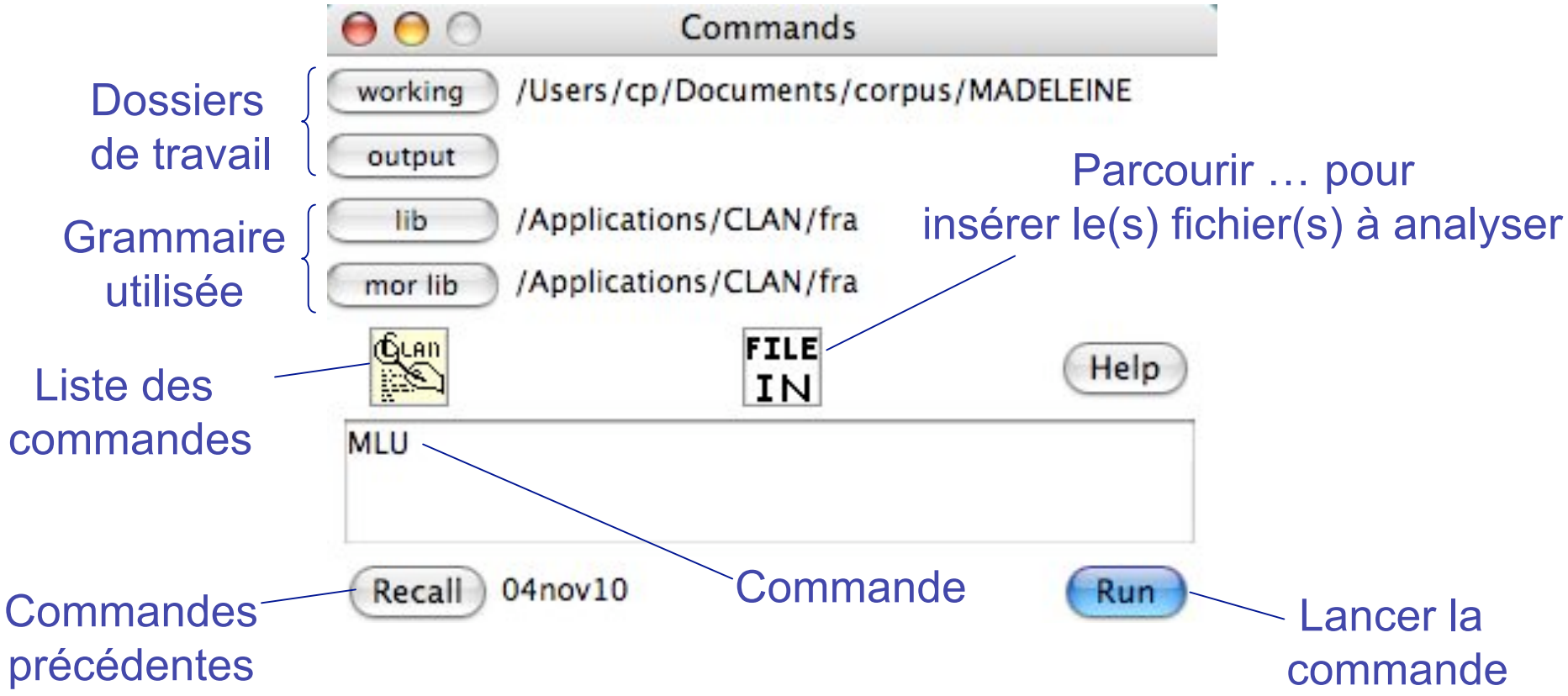
Analyse morphosyntaxique désambiguisée

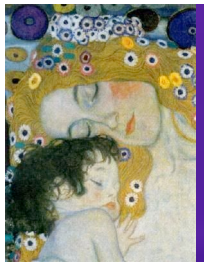


2.3 Analyses automatiques

La fenêtre de commandes

Ouvrir la fenêtre de commandes : Windows > Commands Ou ctrl/ +d





2.3 Analyses automatiques

Structure d'une commande simple

Nom de la commande

Requêtes supplémentaires (+s : string)

Nom du ou des fichier(s)

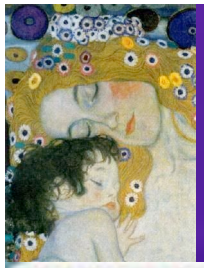
Espaces

```
MLU       +t*CHI    +t*MOT    MADELEINE-2_02_01.cha
FREQ      +t*CHI    +sje      MADELEINE-2_02_01.cha
COMBO     +t*CHI    +sje      MAD*.cha
MOR                           @
```

Participant (+t : tier)

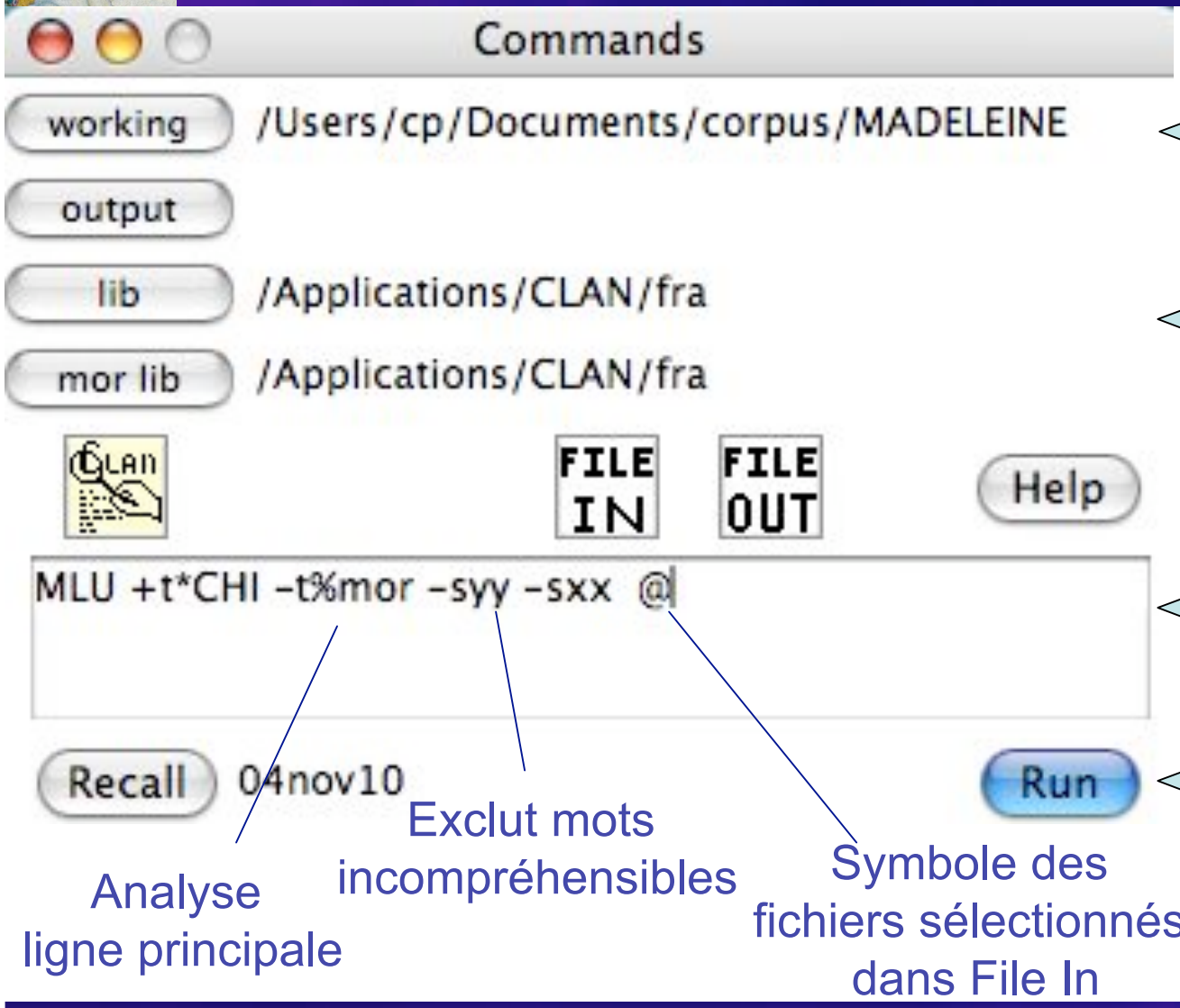
@ : symbole raccourci pour fichiers selectionnés

* : Tous les fichiers qui commencent par MAD et se terminent par .cha



2.3 Analyses automatiques

MLU : longueur moyenne des énoncés

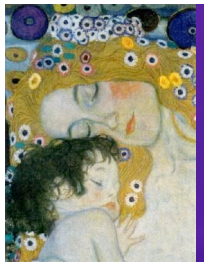


Dossier de travail

Grammaire utilisée

Nom de la commande

Lancer la commande



2.3 Analyses automatiques

MLU : longueur moyenne des énoncés

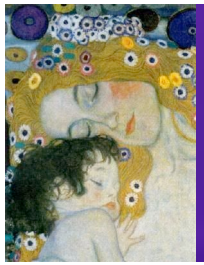
CLAN Output

```
> MLU +t*CHI -t%mor -syy -sxx  @
mlu +t*CHI -t%mor -syy -sxx @
Mon Nov 15 11:29:08 2010
mlu (04-Nov-2010) is conducting analyses on:
  ONLY speaker main tiers matching: *CHI;
****************************************
From file </Users/cp/Desktop/Paris/madeleine/MADELEINE-13-2_01_02.cha>
MLU for Speaker: *CHI
  MLU (xxx and yyy are EXCLUDED from the utterance and morpheme counts):
      Number of: utterances = 526, morphemes = 1511
      Ratio of morphemes over utterances = 2.873
      Standard deviation = 1.574
```

Rappel de la commande

Rappel du nom de fichier

MLU



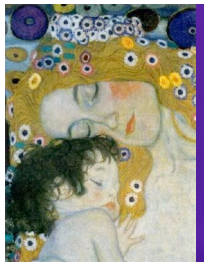
2.3 Analyses automatiques

MLU : longueur moyenne des énoncés

Après report des résultats dans Excel …

MLU-5enfants.xls

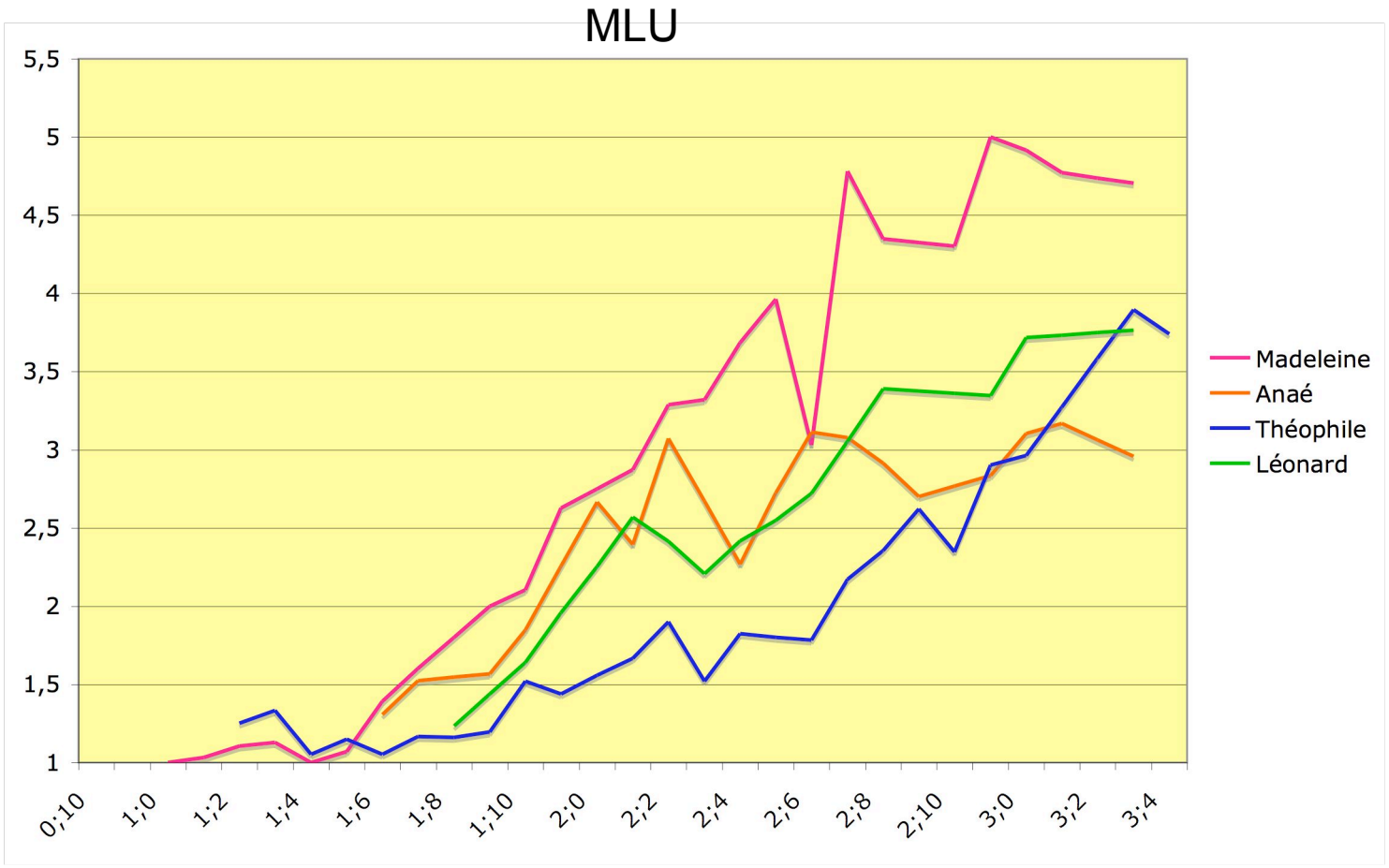
	A	B
1	Age	MLU Madeleine
2	1;0	1
3	1;1	1,032
4	1;2	1,105
5	1;3	1,128
6	1;4	1
7	1;5	1,07
8	1;6	1,393
9	1;7	1,603
10	1;8	1,8005
11	1;9	1,998
12	1;10	2,104
13	1;11	2,627
14	2;0	2,75
15	2;1	2,873
16	2;2	3,287
17	2;3	3,32
18	2;4	3,684
19	2;5	3,964
20	2;6	3,03
21	2;7	4,782
22	2;8	4,348
23	2;9	4,3255
24	2;10	4,303
25	2;11	5
26	3;0	4,916
27	3;1	4,772
28	3;2	4,7385
29	3;3	4,705

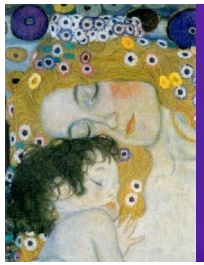


2.3 Analyses automatiques

MLU : longueur moyenne des énoncés

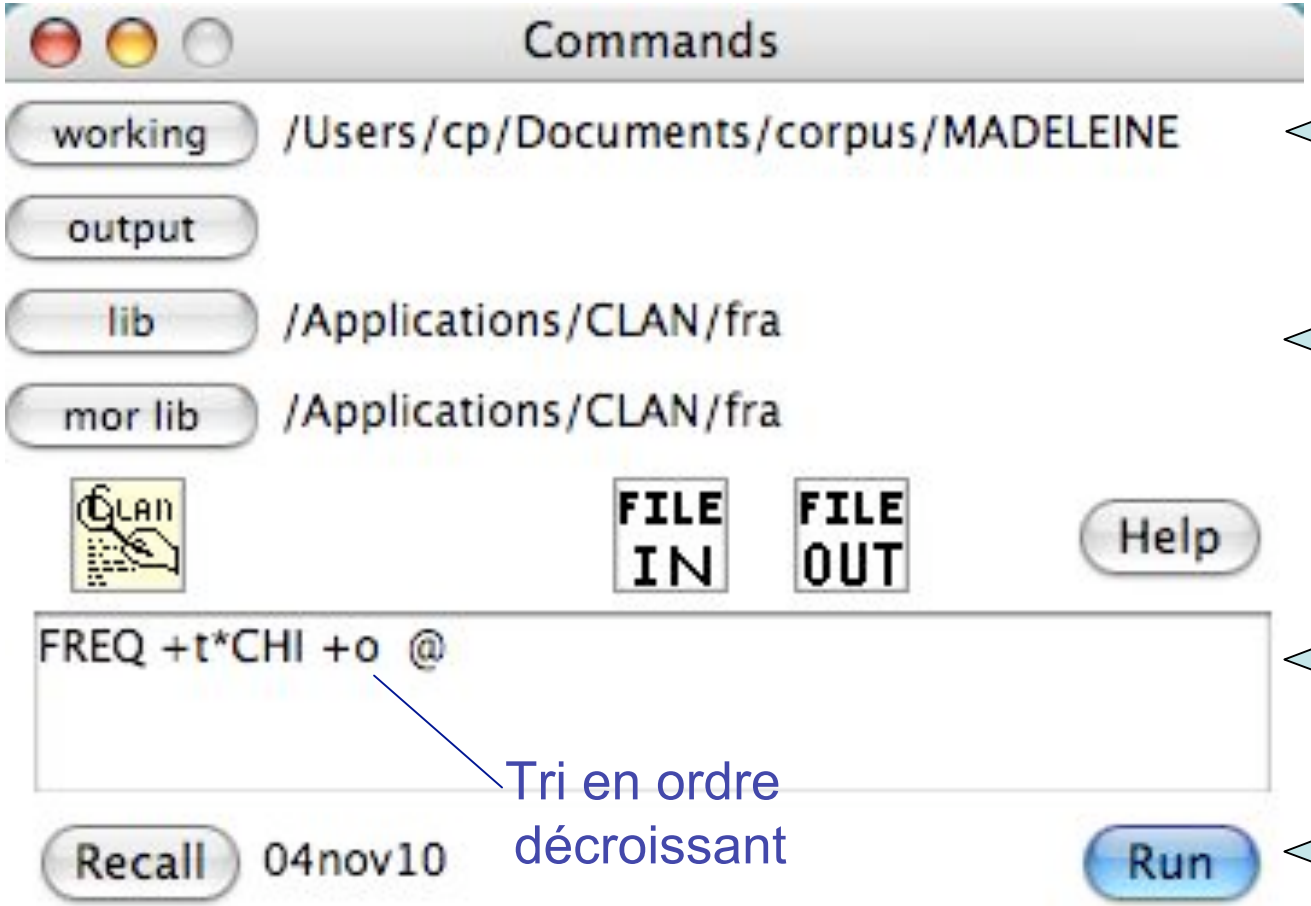
Permet de comparer avec d'autres enfants …





2.3 Analyses automatiques

FREQ 1 : liste des mots produits par l'enfant

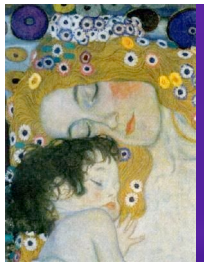


Dossier de travail

Grammaire utilisée

Nom de la commande

Lancer la commande



2.3 Analyses automatiques

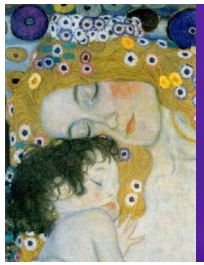
FREQ 1 : liste des mots produits par l'enfant

```
> freq  +t*CHI +o @
freq +t*CHI +o @
Sun Nov 14 20:05:46 2010

freq (04-Nov-2010) is conducting analyses on:
  ONLY speaker main tiers matching: *CHI;
****************************************
From file </Users/cp/Documents/corpus/MADELEINE/MADELEINE-13-2_01_02-mor.cha>
 94 moi
 85 c'est
 56 oui
 53 ça
 43 veux
 37 yy
 36 non
 32 le
 29 la
 27 on
 27 pas
[...]
  1 voudrais
------------------------------
  331  Total number of different word types used
 1745  Total number of words (tokens)
0.190  Type/Token ratio
```

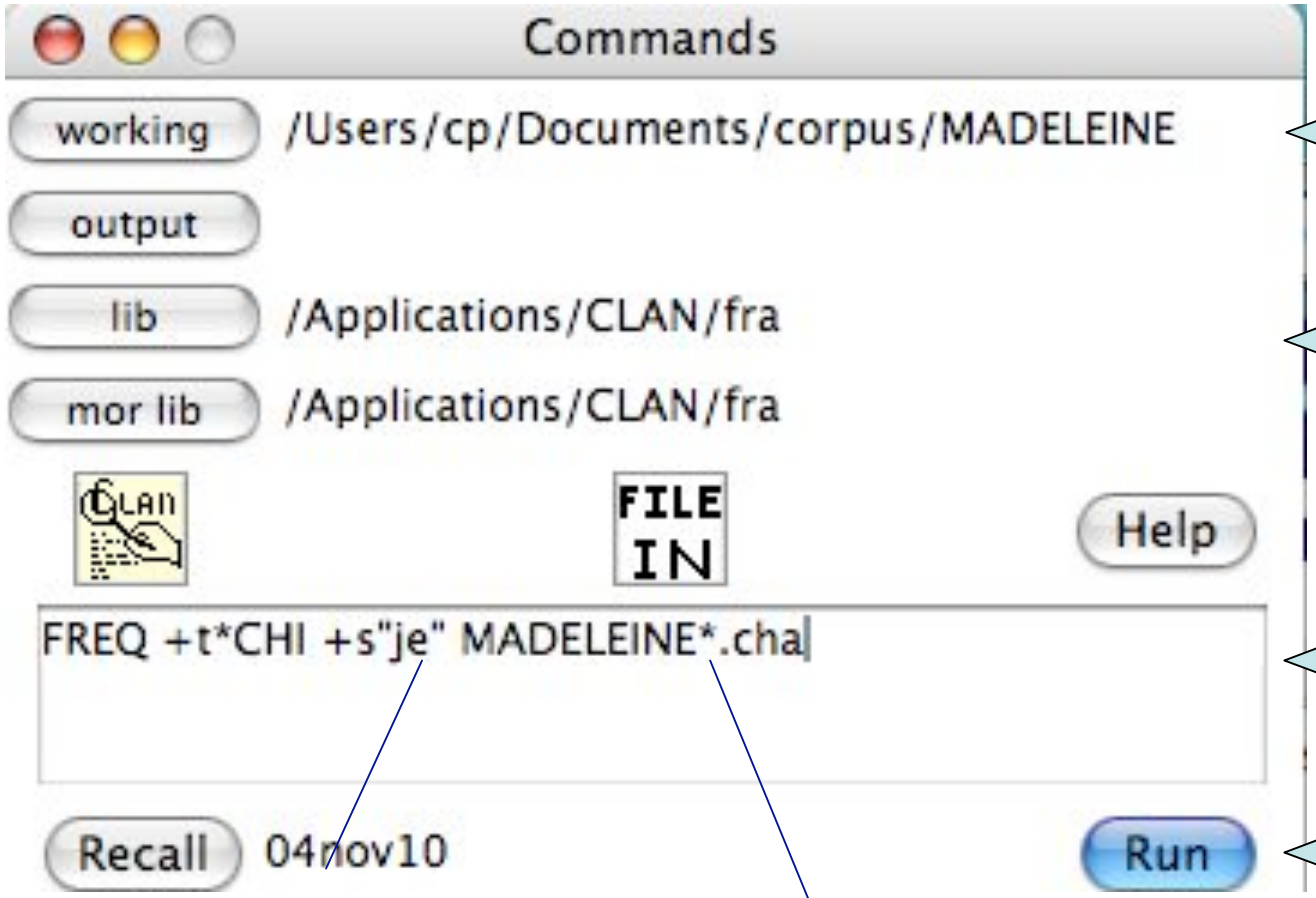
Rappel de la commande

Rappel du nom de fichier



2.3 Analyses automatiques

FREQ 2 : fréquences d'items



Dossier de travail

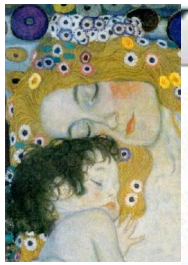
Grammaire utilisée

Nom de la commande

Lancer la commande

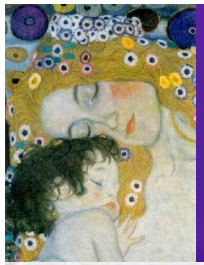
Cherche les « je »

Sur tous les fichiers commençant par « MADELEINE »
et situés dans le dossier de travail choisi



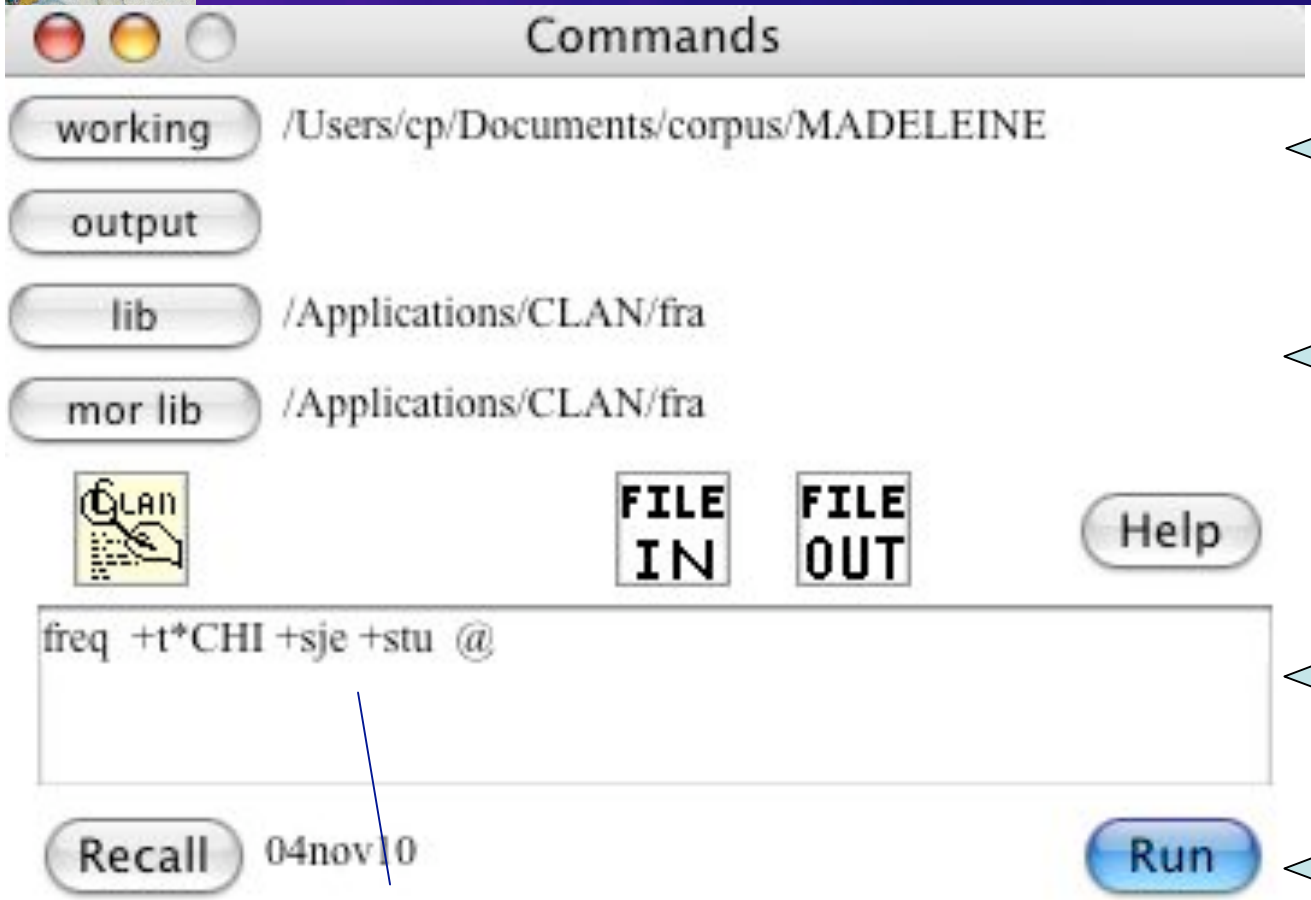
CLAN Output

```
> FREQ +t*CHI +s"je" MADELEINE*.cha
freq +t*CHI +sje MADELEINE*.cha
Mon Nov 15 11:41:04 2010
freq (04-Nov-2010) is conducting analyses on:
  ONLY speaker main tiers matching: *CHI;
****************************************
From file <MADELEINE-11-1_10_07-cor.cha>
  1 je
----------------------------
    1  Total number of different word types used
    1  Total number of words (tokens)
1.000  Type/Token ratio
From file <MADELEINE-12-1_11_13.cha>
----------------------------
    0  Total number of different word types used
    0  Total number of words (tokens)
From file <MADELEINE-13-2_01_02-bis.cha>
----------------------------
    0  Total number of different word types used
    0  Total number of words (tokens)
From file <MADELEINE-13-2_01_02-mor-pst.cha>
 19 je
----------------------------
    1  Total number of different word types used
   19  Total number of words (tokens)
```



2.3 Analyses automatiques

FREQ 2 : fréquences d'items



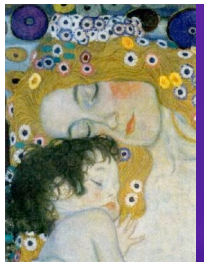
Dossier de travail

Grammaire utilisée

Nom de la commande

Lancer la commande

Calcule les « je » et les « tu »



2.3 Analyses automatiques

FREQ 2 : fréquences d'items

CLAN Output

```
> freq  +t*CHI +sje +stu  @
freq +t*CHI +sje +stu @
Wed Nov 17 12:29:18 2010
freq (04-Nov-2010) is conducting analyses on:
  ONLY speaker main tiers matching: *CHI;
****************************************
From file </Users/cp/Documents/corpus/MADELEINE/MADELEINE-13-2_01_02.cha>
 19 je
  6 tu
------------------------------
    2  Total number of different word types used
   25  Total number of words (tokens)
0.080  Type/Token ratio

>
```

Madeleine

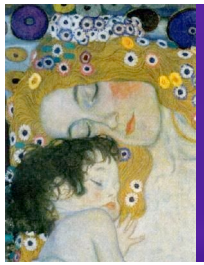
CLAN File Edit Font Size/Style Tiers Mode Windows Help

CLAN Output

```
> freq  +t*MOT +sje +stu  @
freq +t*MOT +sje +stu @
Wed Nov 17 17:05:04 2010
freq (04-Nov-2010) is conducting analyses on:
  ONLY speaker main tiers matching: *MOT;
****************************************
From file </Users/cp/Desktop/Paris/madeleine/MADELEINE-13-2_01_02.cha>
 49 je
157 tu
------------------------------
    2  Total number of different word types used
  206  Total number of words (tokens)
0.010  Type/Token ratio

>
```

La mère de Madeleine

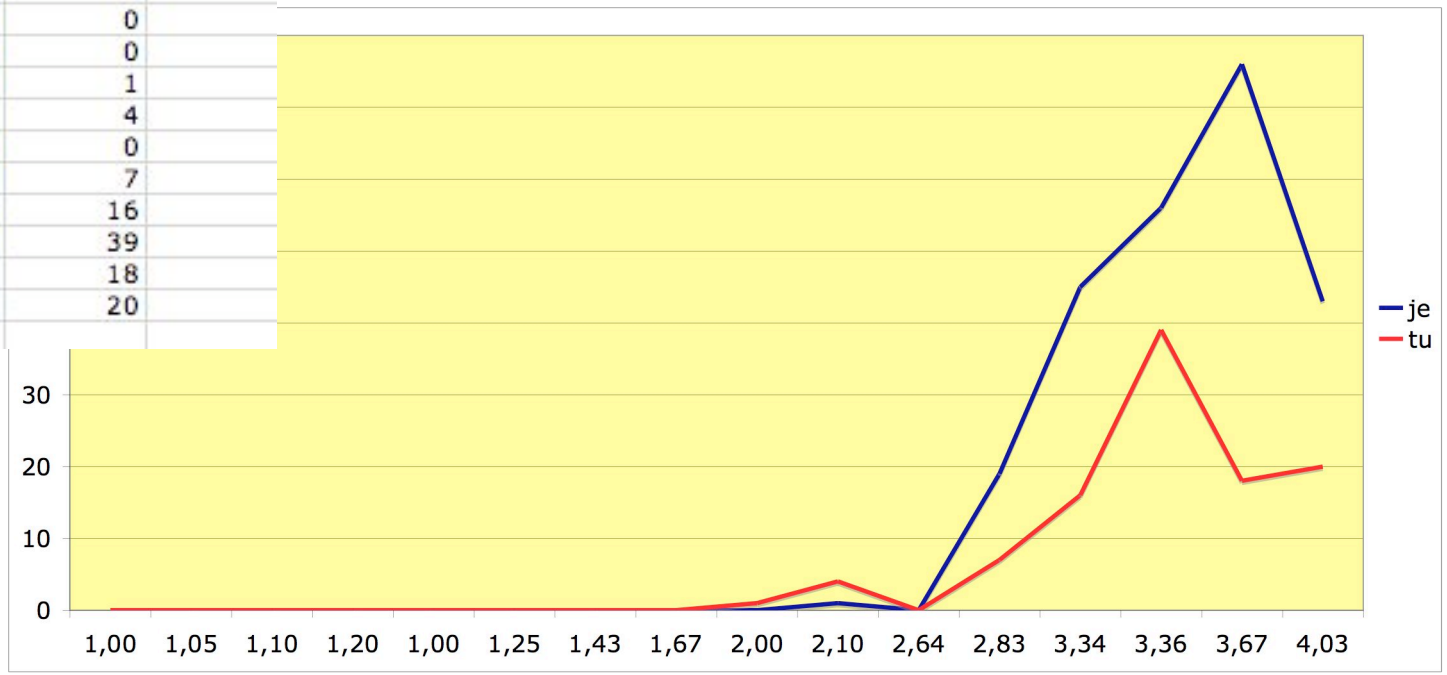


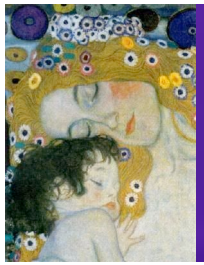
2.3 Analyses automatiques

FREQ 2 : fréquence d'items

Après report des résultats dans Excel …

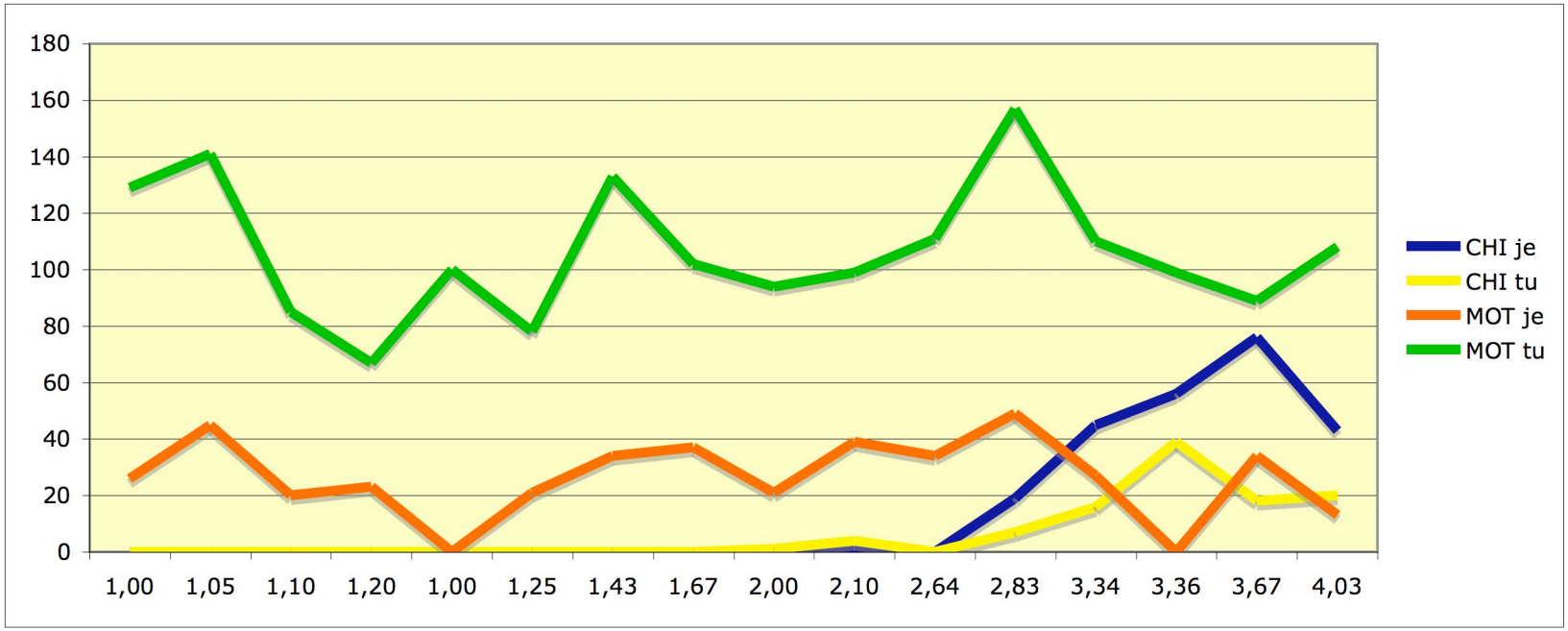
	A	B	C	D	E
1	Age	MLU	je	tu	
2	1;00	1,00	0	0	
3	1;01	1,05	0	0	
4	1;02	1,10	0	0	
5	1;03	1,20	0	0	
6	1;04	1,00	0	0	
7	1;05	1,25	0	0	
8	1;06	1,43	0	0	
9	1;07	1,67	0	0	
10	1;09	2,00	0	1	
11	1;10	2,10	1	4	
12	1;11	2,64	0	0	
13	2;01	2,87	19	7	
14	2;02	3,34	45	16	
15	2;03	3,36	56	39	
16	2;04	3,67	76	18	
17	2;05	4,03	43	20	
18					

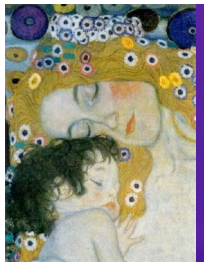




2.3 Analyses automatiques
FREQ 2 : fréquence d'items

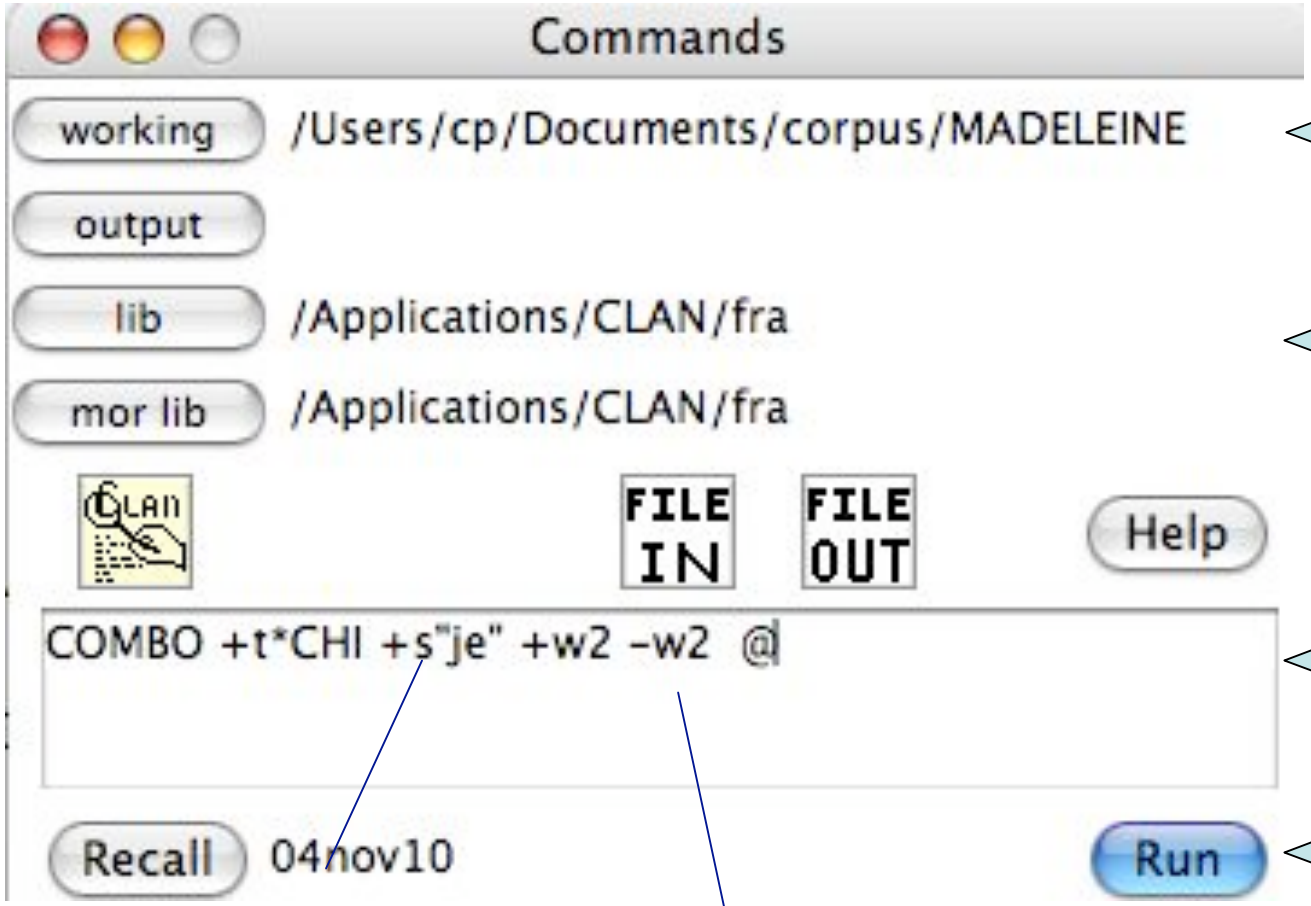
Permet de comparer la mère et l'enfant …





2.3 Analyses automatiques

COMBO 1 : contexte de production



Dossier de travail

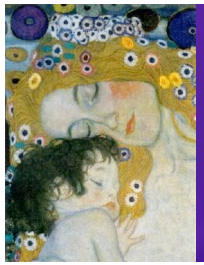
Grammaire utilisée

Nom de la commande

Lancer la commande

Cherche les « je »

Montre les deux énoncés précédents et suivants



2.3 Analyses automatiques

COMBO 1 : contexte de production

CLAN File Edit Font Size/Style Tiers Mode Windows Help

CLAN Output

```
> COMBO +t*CHI +s"je" +w2 -w2 @
(je)
combo +t*CHI +sje +w2 -w2 @
Mon Nov 15 11:56:01 2010
combo (04-Nov-2010) is conducting analyses on:
  ONLY speaker main tiers matching: *CHI;
****************************************
From file </Users/cp/Desktop/Paris/madeleine/MADELEINE-13-2_01_02.cha>
----------------------------------------
*** File "/Users/cp/Desktop/Paris/madeleine/MADELEINE-13-2_01_02.cha": line 235.
*CHI:   0 . •
*CHI:   et ton manteau ? •
*CHI:   <moi (1)je veux enlever ton manteau> [>] . •
*OBS:   <le manteau aussi tu veux> [<] ?
*OBS:   attends . •
----------------------------------------
*** File "/Users/cp/Desktop/Paris/madeleine/MADELEINE-13-2_01_02.cha": line 644.
*CHI:   oh moi z@fs veux fermer tout seul la porte ! •
*OBS:   qu' est+ce+qui t'arrive ? •
*CHI:   moi <(1)je veux ouv(rir)> [///] z@fs veux fermer yy +... •
*OBS:   fermer [>] ? •
```

Movie – Sound

318426 Save

v v zoom: 0.46

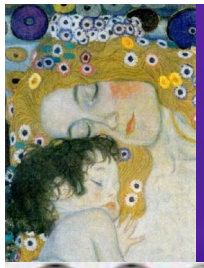
318120 321220

Repeat 0 msec

13.mov

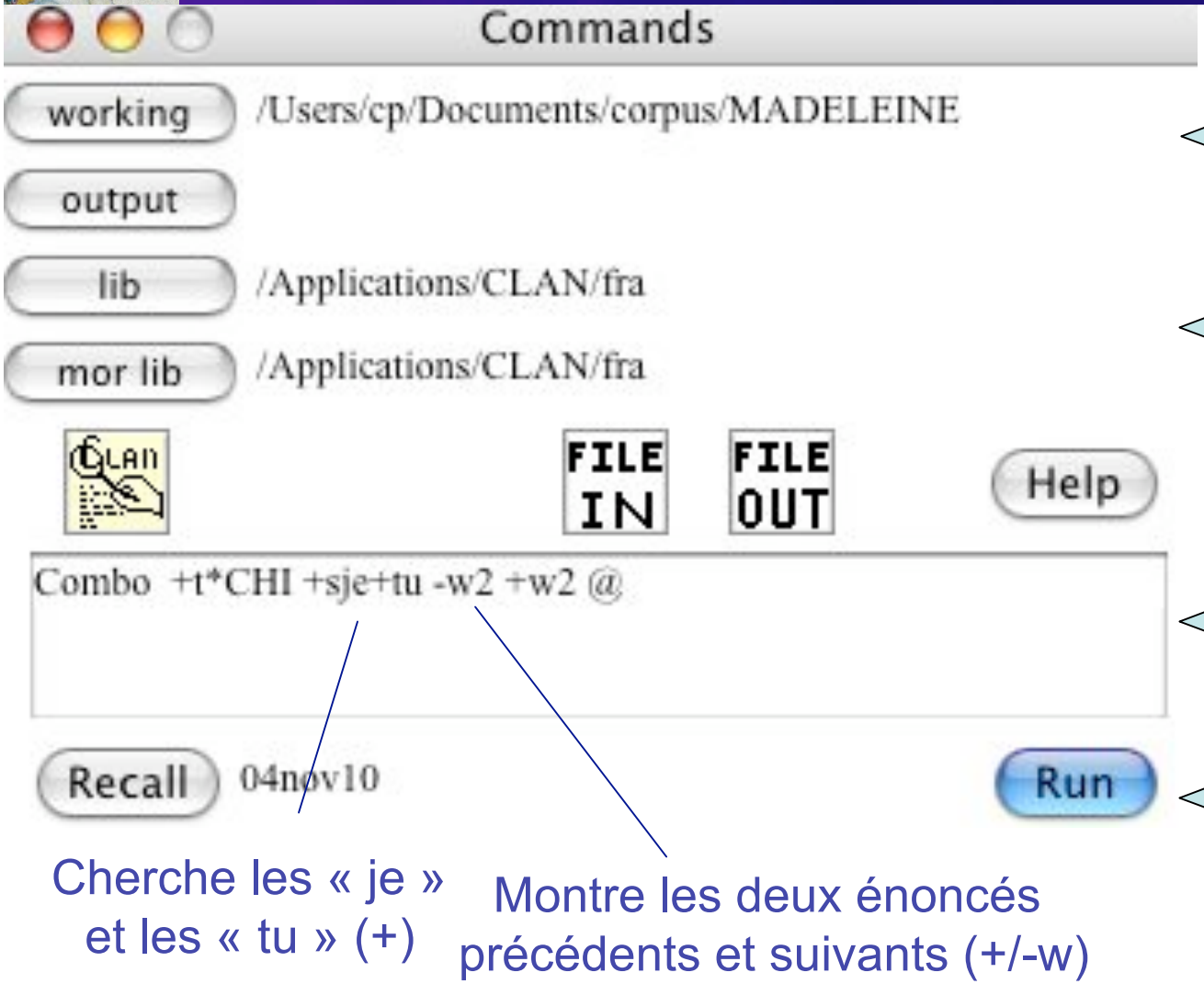
Discours précédent

Accès direct à la vidéo



2.3 Analyses automatiques

COMBO 2 : recherche double

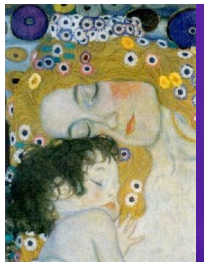


Dossier de travail

Grammaire utilisée

Nom de la commande

Lancer la commande



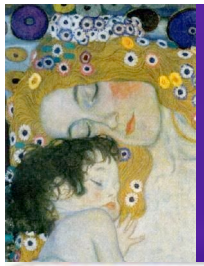
2.3 Analyses automatiques

COMBO 2 : recherche double

CLAN Output

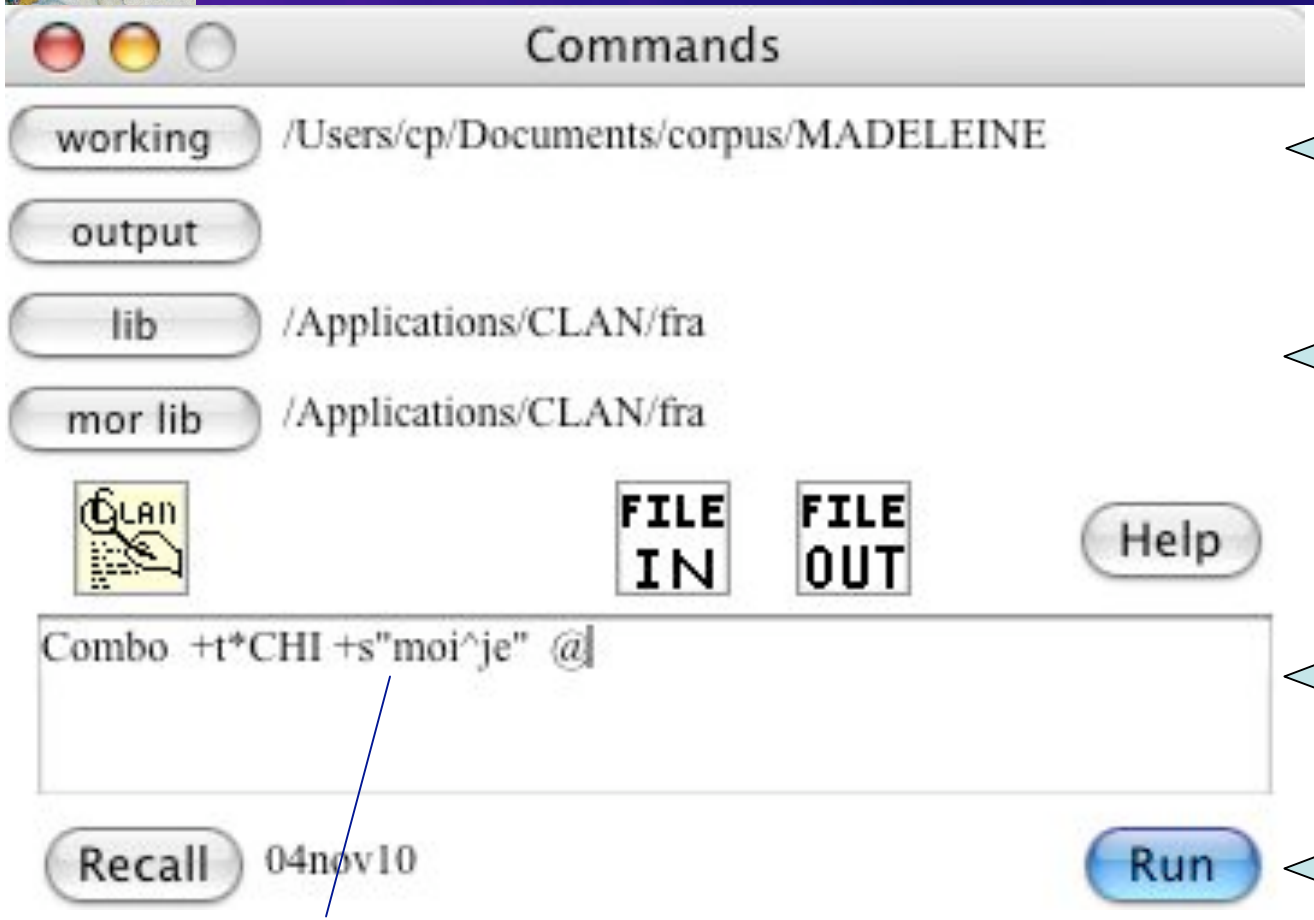
```
*** File "/Users/cp/Documents/corpus/MADELEINE/MADELEINE-13-2_01_02.cha": line 1989.
*CHI:   voilà [>] . •
*MOT:   <il en a bien> [<] besoin .
*CHI:   <(1)je v(eux)> [/] (2)je veux pas le chandail . •
*CHI:   maman prend le chandail . •
*MOT:   bah oui ben j(e) vais le mettre à la poupée . •
----------------------------------------
*** File "/Users/cp/Documents/corpus/MADELEINE/MADELEINE-13-2_01_02.cha": line 2300.
*MOT:   le col c'est important qu' i(l) soit bien r(e)passé . •
*MOT:   0 [=! imite bruit de la vapeur] . •
*CHI:   (1)t(u) as mis des [/] des +/. •
*MOT:   voilà ça y+est . •
*CHI:   t(u) as fini ? •
----------------------------------------
```

Triple clic sur nom de fichier : Accès au passage dans la transcription



2.3 Analyses automatiques

COMBO 3 : collocations



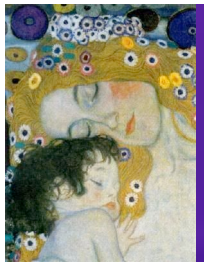
Dossier de travail

Grammaire utilisée

Nom de la commande

Lancer la commande

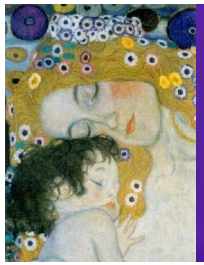
Cherche les « moi » immédiatement suivis de « je » (^)



2.3 Analyses automatiques
COMBO 3 : collocations

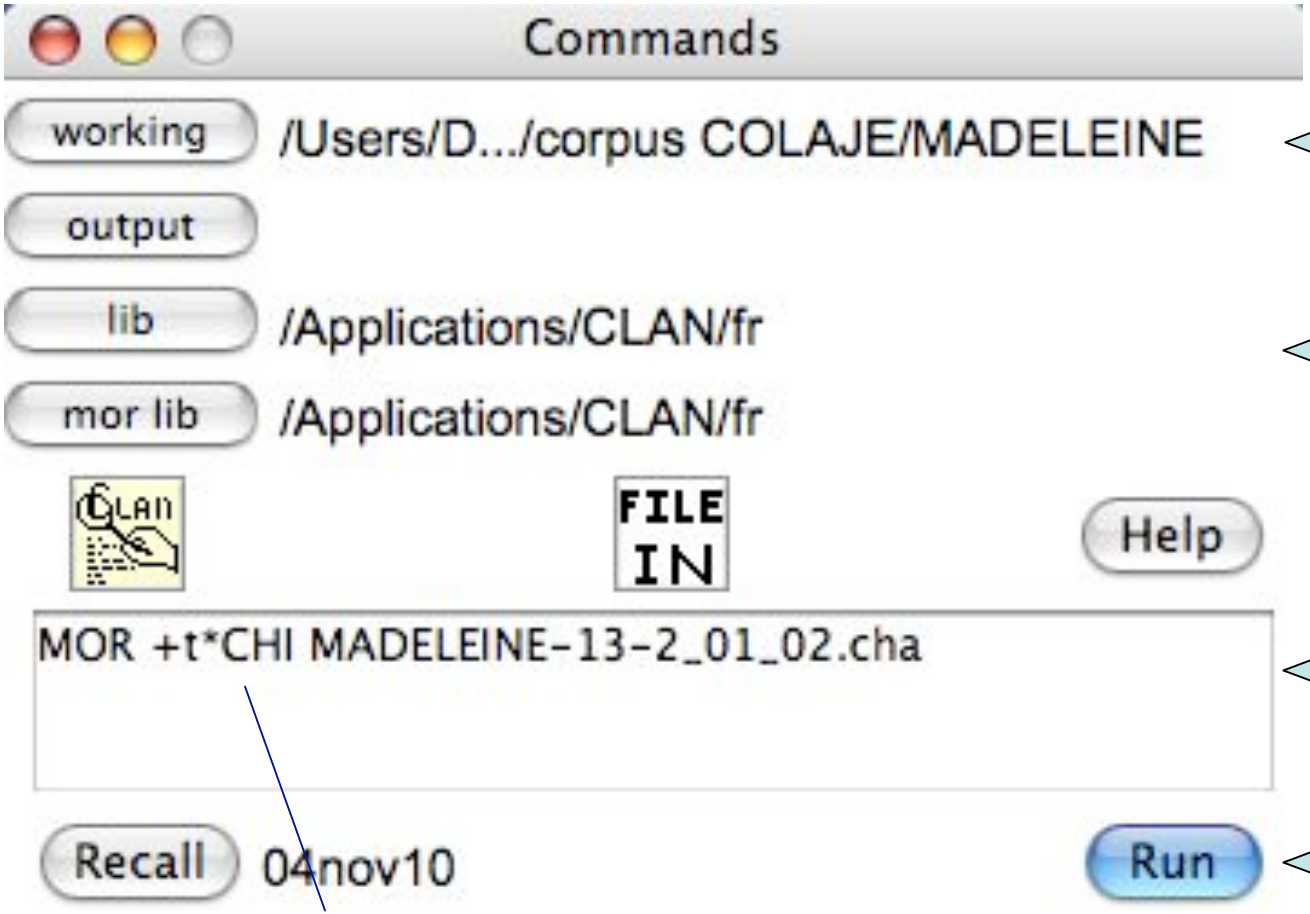
CLAN Output

```
> Combo  +t*CHI +s"moi^je"  @
((moi^je))
combo +t*CHI +smoi^je @
Wed Nov 17 12:45:32 2010
combo (04-Nov-2010) is conducting analyses on:
  ONLY speaker main tiers matching: *CHI;
****************************************
From file </Users/cp/Documents/corpus/MADELEINE/MADELEINE-13-2_01_02.cha>
----------------------------------------
*** File "/Users/cp/Documents/corpus/MADELEINE/MADELEINE-13-2_01_02.cha": line 275.
*CHI:   <(1)moi (1)je veux enlever ton manteau> [>] . •
----------------------------------------
*** File "/Users/cp/Documents/corpus/MADELEINE/MADELEINE-13-2_01_02.cha": line 704.
*CHI:   (1)moi <(1)je veux ouv(rir)> [///] zə@fs veux fermer yy +... •
----------------------------------------
*** File "/Users/cp/Documents/corpus/MADELEINE/MADELEINE-13-2_01_02.cha": line 793.
*CHI:   (1)moi (1)je veux nettoyer . •
----------------------------------------
*** File "/Users/cp/Documents/corpus/MADELEINE/MADELEINE-13-2_01_02.cha": line 821.
*CHI:   (1)moi (1)j(e) nettoie . •
----------------------------------------
*** File "/Users/cp/Documents/corpus/MADELEINE/MADELEINE-13-2_01_02.cha": line 823.
*CHI:   (1)moi (1)je nettoie (1)moi . •
```



2.3 Analyses automatiques

MOR : analyse morphosyntaxique



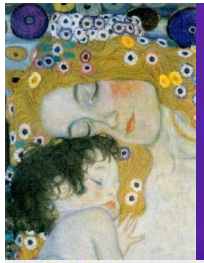
Dossier de travail

Grammaire utilisée

Nom de la commande

Lancer la commande

Fait l'analyse morphosyntaxique
sur les énoncés de l'enfant uniquement

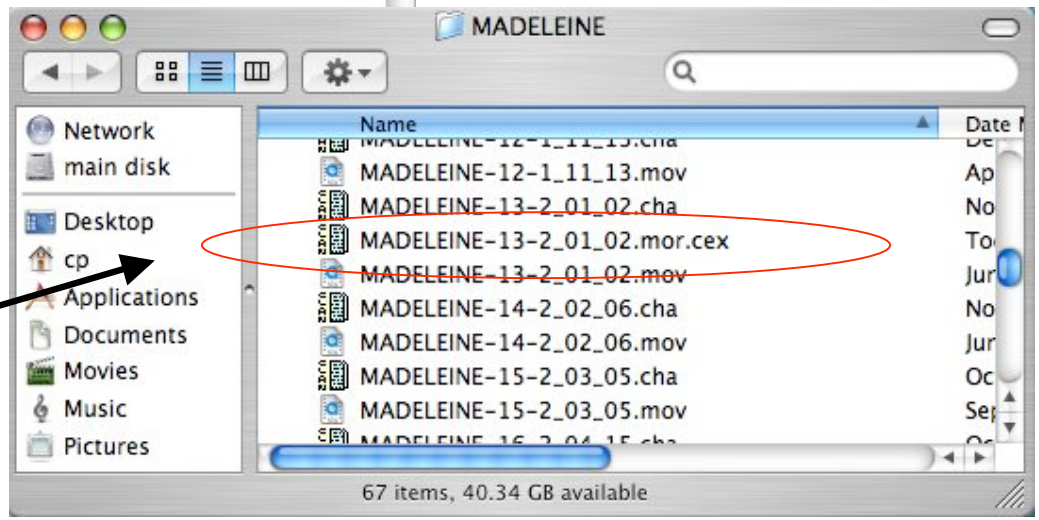


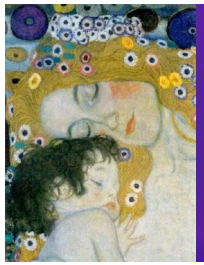
2.3 Analyses automatiques

MOR : analyse morphosyntaxique

CLAN Output

```
> MOR +t*CHI MADELEINE-13-2_01_02.cha
mor +t*CHI MADELEINE-13-2_01_02.cha
Mon Nov 22 16:02:31 2010
mor (04-Nov-2010) is conducting analyses on:
  ONLY speaker main tiers matching: *CHI;
    and those speakers' ALL dependent tiers
  and ALL header tiers
****************************************
From file <MADELEINE-13-2_01_02.cha> to file <MADELEINE-13-2_01_02.mor.cex>
Using sf-rule: /Applications/CLAN/fr/sf.cut.
Using a-rules: /Applications/CLAN/fr/ar.cut.
Using lexicon: /Applications/CLAN/fr/lex/0affix-n.cut.
Using lexicon: /Applications/CLAN/fr/lex/0affix-v-irreg.cut.
Using lexicon: /Applications/CLAN/fr/lex/0affix-v.cut.
[...
Using lexicon: /Applications/CLAN/fr/lex/zero.cut.
Loaded lexicon: 44572
Using c-rules: /Applications/CLAN/fr/cr.cut.
warning:  crule "adv-ment" referenced but not defined
warning:  crule "adv-ment" referenced but not defined
warning:  crule "adv-ment" referenced but not defined
warning:  crule "adv-ment" referenced but not defined
2600
Done with file <MADELEINE-13-2_01_02.mor.cex>
```





2.3 Analyses automatiques
MOR : analyse morphosyntaxique

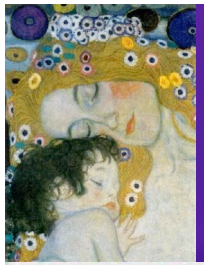
Fichier .mor.cha créé par CLAN :

/Users/cp/Documents/corpus COLAJE/MADELEINE/MADEL...

```
*CHI:   0 .
%act:   CHI jette une regard à OBS puis reprend son ménage .
*CHI:   moi I@fs a yy (.) nettoyé . •
%mor:   pro|moi&SING^n|moi&_MASC fs|I
        v:poss|avoir&PRES&3SV^v:aux|avoir&PRES&3SV unk|yy
        v|nettoyer-PP-_MASC .
%pho:   mwa I a tõ letaje
*OBS:   tu te vois dedans . •
04nov10[E|CHAT] * 1150
```

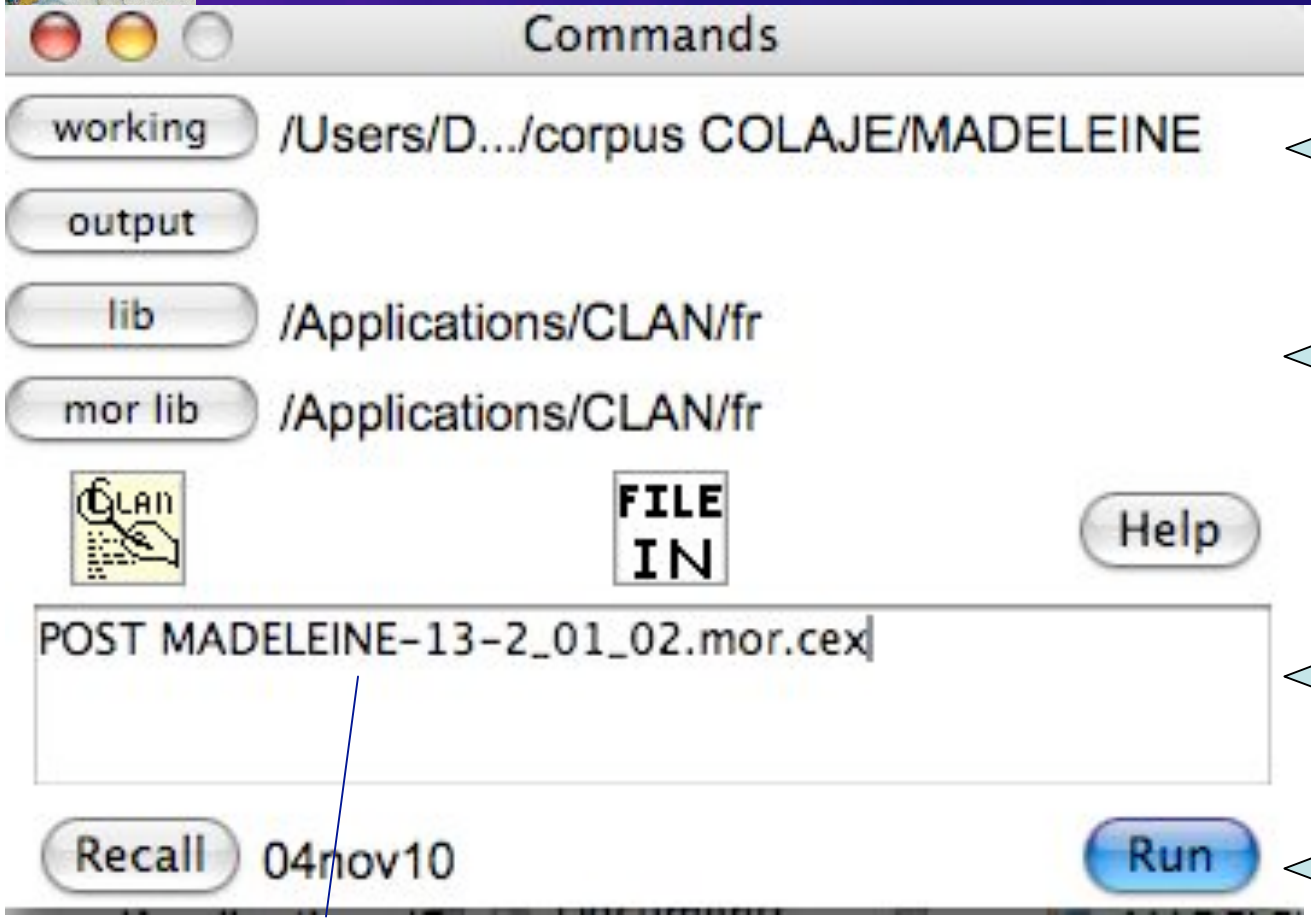
^ = OU

« a » est le verbe possessif avoir (v:poss|avoir&PRES&3SV)
OU l'auxiliaire avoir (v:aux|avoir&PRES&3SV)



2.3 Analyses automatiques

POST : désambiguïse l'analyse de MOR



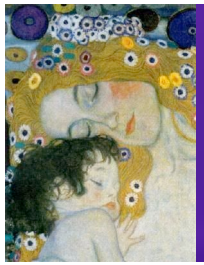
Dossier de travail

Grammaire utilisée

Nom de la commande

Lancer la commande

Désambiguïse les résultats de MOR sur le fichier .mor.cex



2.3 Analyses automatiques

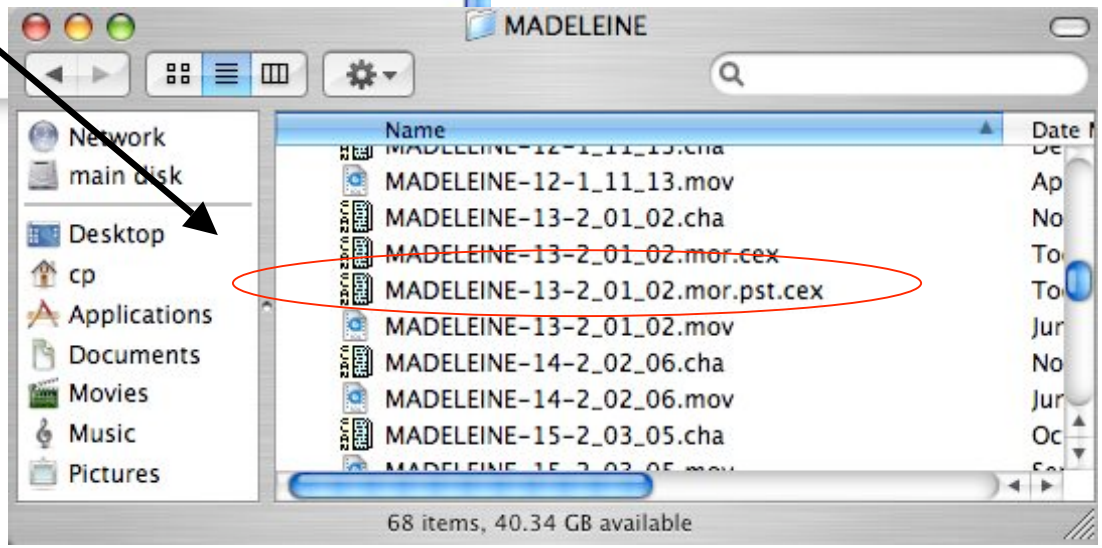
POST : désambiguïse l'analyse de MOR

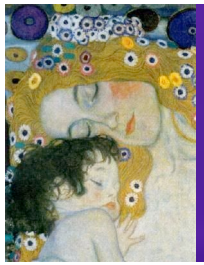
CLAN Output

```
> POST MADELEINE-13-2_01_02.mor.cex

Using file:     /Applications/CLAN/fr/post.db.
post MADELEINE-13-2_01_02.mor.cex
Mon Nov 22 16:03:49 2010
post (04-Nov-2010) is conducting analyses on:
  ALL speaker tiers
****************************************
From file <MADELEINE-13-2_01_02.mor.cex> to file <MADELEINE-13-2_01_02.mor.pst.cex>
Done with file <MADELEINE-13-2_01_02.mor.pst.cex>

04nov10[E|TEXT]  12
```





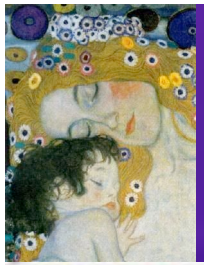
2.3 Analyses automatiques
POST : désambiguïse l'analyse de MOR

Fichier .mor.pst.cex créé par CLAN :

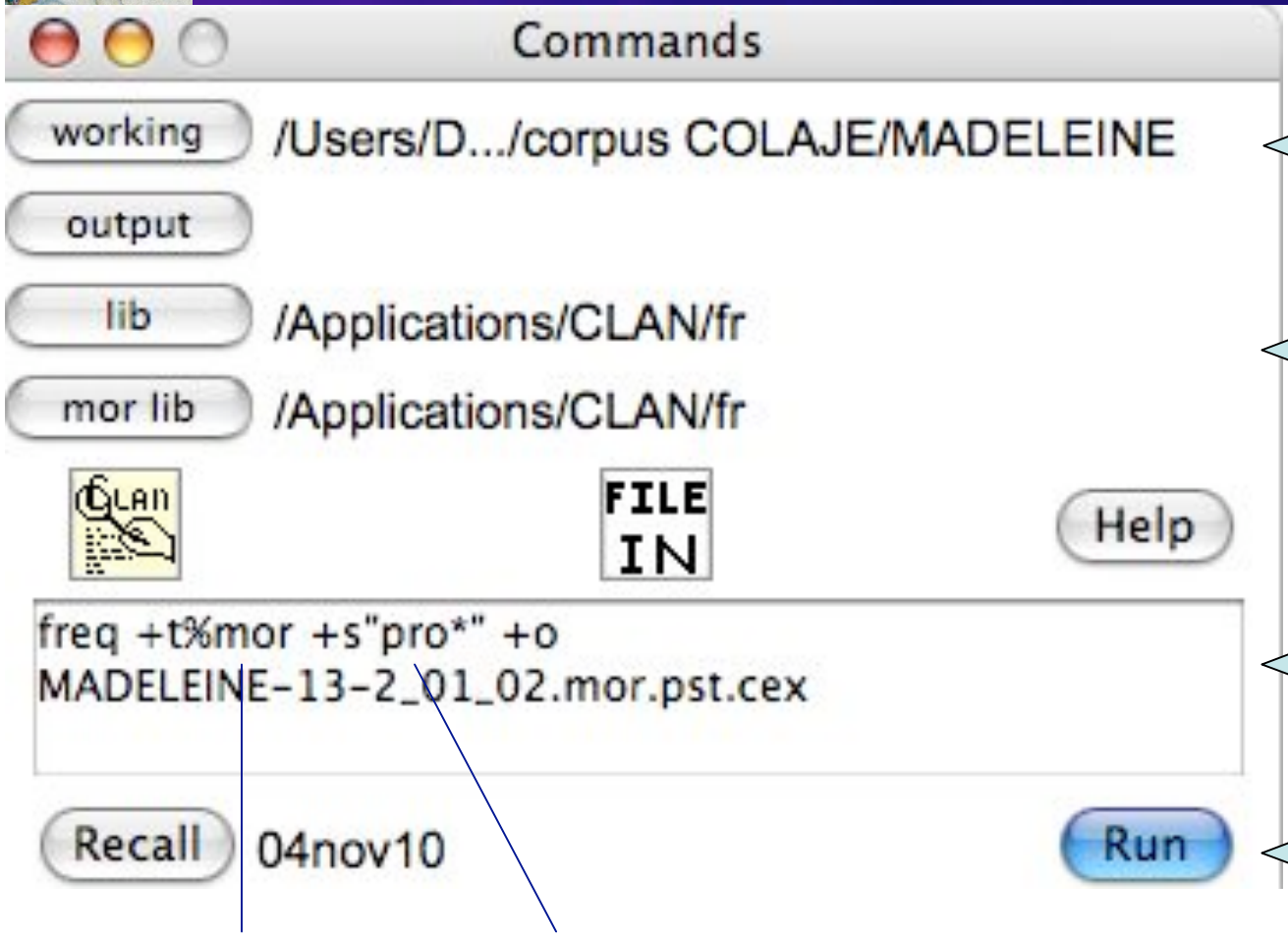
/Users/cp/Documents/corpus COLAJE/MADELEINE/MADELEINE-...

```
*CHI:   0 .
%act:   CHI jette une regard à OBS puis reprend son ménage .
*CHI:   moi I@fs a yy (.) nettoyé . •
%mor:   pro|moi fs|I v:aux|avoir unk|yy v|nettoyer-PP .
%pho:   mwa I a tɑ̃ letaje
*OBS:   tu te vois dedans . •
04nov10[E|CHAT]  980
```

Les ^ (OU) ont disparu.
« a » est analysé comme étant
l'auxiliaire avoir (v:aux|avoir)



2.3 Analyses automatiques
MOR|POST et FREQ



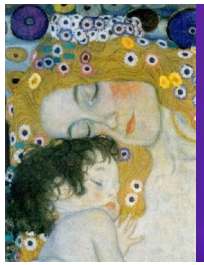
Dossier de travail

Grammaire utilisée

Nom de la commande

Lancer la commande

Sur la ligne %mor …Cherche la chaîne qui commence par « pro » suivi de « n'importe quoi » (*)



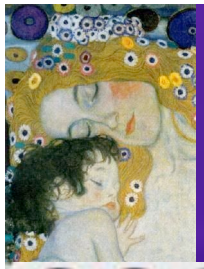
2.3 Analyses automatiques
MOR|POST et FREQ

CLAN Output

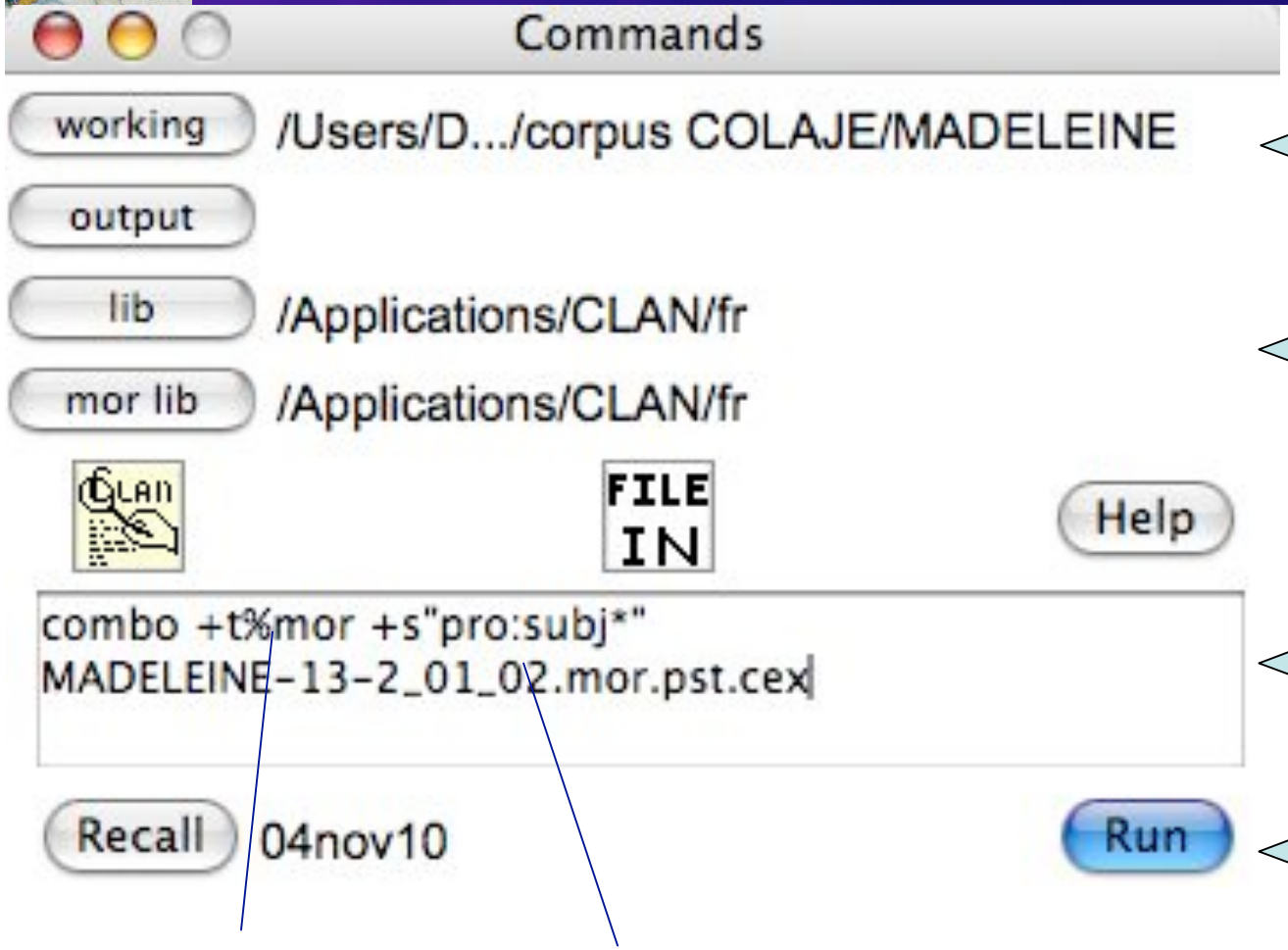
```
> freq +t%mor +s"pro*" +o MADELEINE-13-2_01_02.mor.pst.cex
freq +t%mor +spro* +o MADELEINE-13-2_01_02.mor.pst.cex
Mon Nov 22 16:32:21 2010
freq (04-Nov-2010) is conducting analyses on:
  ALL speaker tiers
      and those speakers' ONLY dependent tiers matching: %MOR;
****************************************
From file <MADELEINE-13-2_01_02.mor.pst.cex>
 89 pro|moi
 49 pro:dem|ça
 22 pro:subj|on
[...]
  1 pro:y|en
  1 problème
  1 promener
  1 propre
  1 pro|mêmes
---------- ----------
   26  Total number of different word types used
  261  Total number of words (tokens)
0.100  Type/Token ratio

04nov10[E|TEXT] * 7
```

!! attention



2.3 Analyses automatiques
MOR|POST et COMBO



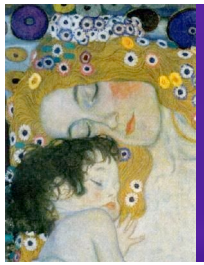
Dossier de travail

Grammaire utilisée

Nom de la commande

Lancer la commande

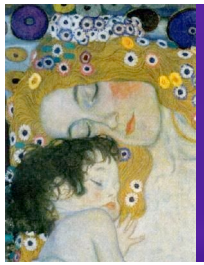
Sur la ligne %mor … Cherche la chaîne qui commence par « pro:subj » suivi de « n'importe quoi » (*)



2.3 Analyses automatiques
MOR|POST et COMBO

CLAN Output

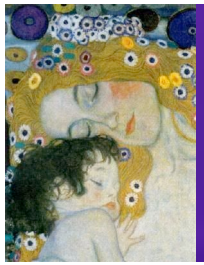
```
%mor:   (1)pro:subj|il v|tomber .
----------------------------------------
*** File "MADELEINE-13-2_01_02.mor.pst.cex": line 381.
*CHI:   (.) <i(l) tombe> [/] <i(l) tombe> [<] . •
%mor:   (1)pro:subj|il v|tomber .
----------------------------------------
*** File "MADELEINE-13-2_01_02.mor.pst.cex": line 711.
*CHI:   on va faire un café . •
%mor:   (1)pro:subj|on v:md|aller v:md|ex|faire-INF det|un n|café .
----------------------------------------
*** File "MADELEINE-13-2_01_02.mor.pst.cex": line 923.
*CHI:   moi je veux nettoyer . •
%mor:   pro|moi (1)pro:subj|je v:md|vouloir v|nettoyer-INF .
----------------------------------------
*** File "MADELEINE-13-2_01_02.mor.pst.cex": line 958.
*CHI:   moi j(e) nettoie . •
%mor:   pro|moi (1)pro:subj|je v|nettoyer .
----------------------------------------
*** File "MADELEINE-13-2_01_02.mor.pst.cex": line 961.
*CHI:   moi je nettoie moi . •
%mor:   pro|moi (1)pro:subj|je v|nettoyer pro|moi .
----------------------------------------
04nov10[E|TEXT]  38
```



2.3 Analyses automatiques

et beaucoup d'autres commandes encore ...

Command	Page	Function
CHAINS	50	Tracks sequences of interactional codes across speakers.
CHECK	53	Verifies the correct use of CHAT format.
CHIP	56	Examines parent-child repetition and expansion.
COMBO	63	Searches for complex string patterns.
COOCUR	71	Examines patterns of co-occurence between words.
DIST	71	Examines patterns of separation between speech act codes.
DSS	72	Computes the Developmental Sentence Score.
FREQ	81	Computes the frequencies of the words in a file or files.
FREQMERG	90	Combines the outputs of various runs of FREQ.
FREQPOS	90	Tracks the frequencies in various utterance positions.
GEM	91	Finds areas of text that were marked with GEM markers.
GEMFREQ	94	Computes frequencies for words inside GEM markers.
GEMLIST	94	Lists the pattern of GEM markers in a file or files.
KEYMAP	95	Lists the frequencies of codes that follow a target code.
KWAL	96	Searches for word patterns and prints the line.
MAXWD	98	Finds the longest words in a file.
MLT	100	Computes the mean length of turn.
MLU	103	Computes the mean length of utterance.
MODREP	108	Matches the child's phonology to the parental model.
PHONFREQ	111	Computes the frequency of phonemes in various positions.
RELY	112	Measures reliability across two transcriptions.
STATFREQ	113	Formats the output of FREQ for statistical analysis.
TIMEDUR	114	Uses the numbers in sonic bullets to compute overlaps.
VOCD	115	Computes the VOCD lexical diversity measure.
WDLEN	120	Computes the length of utterances in words.



Pour aller plus loin
avec CLAN …

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

Programs and Database

Downloadable Database

Browsable Database

The CLAN Program

Derived Corpora and Counts

Links

The TalkBank Database

IASCL information

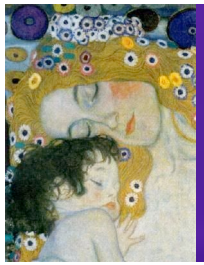
Other Child Language sites

Manuals

CHAT Transcription

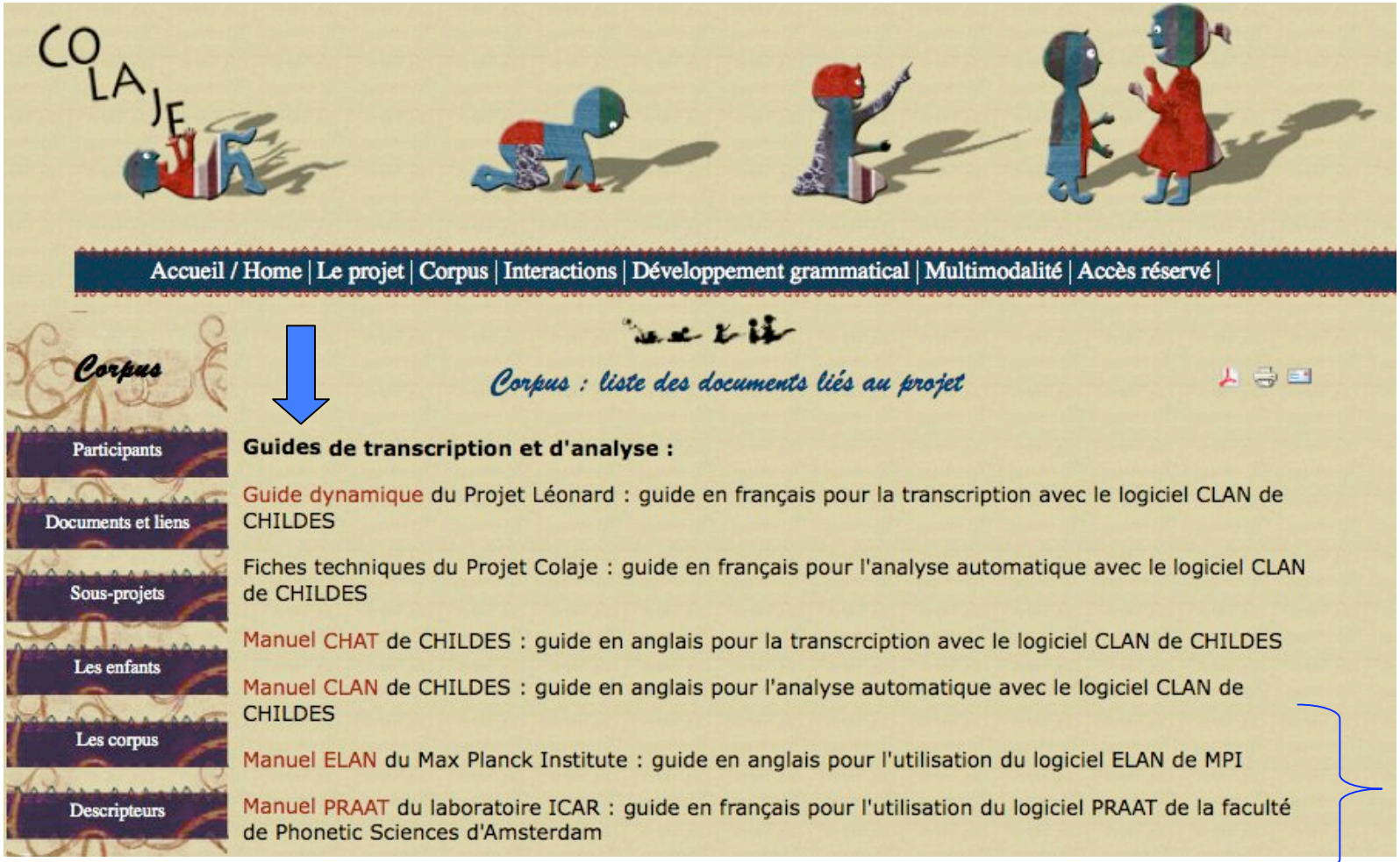
CLAN Programs

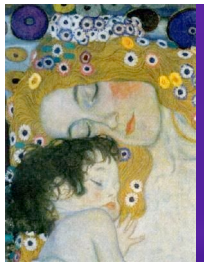
Database Manuals



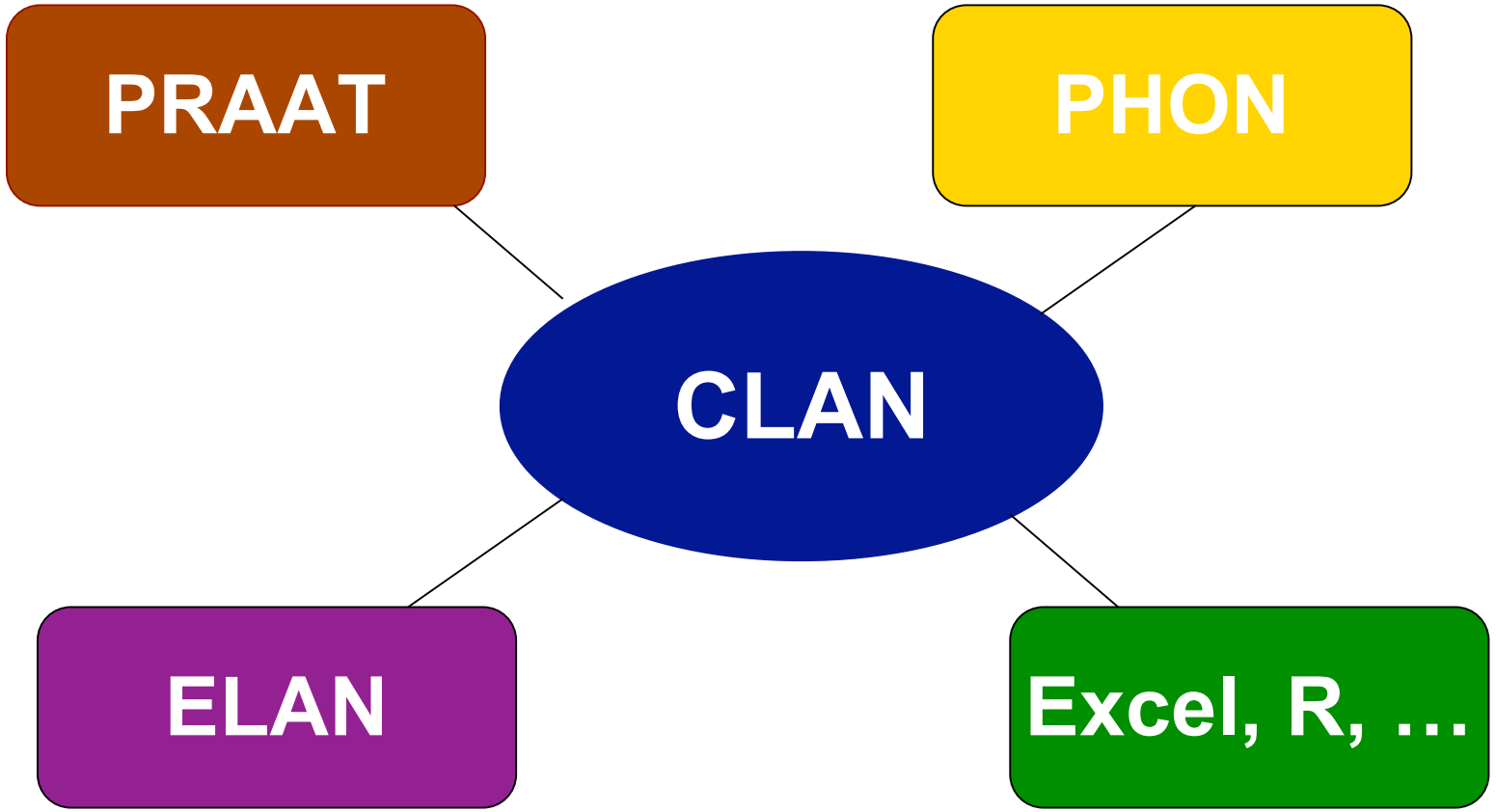
Pour aller plus loin

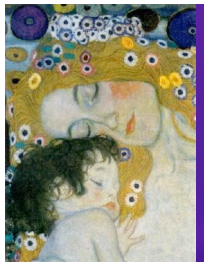
avec CLAN et d'autres logiciels…





Pour aller plus loin
avec CLAN et d'autres logiciels…





Merci !

Cette présentation a été créée à partir du matériel disponible sur
le site Internet de CHILDES (http://childes.psy.cmu.edu/),
le manuel CLAN de L. Balthasar (Laboratoire ICAR),
le manuel CLAN du Projet LEONARD dirigé par A. Morgenstern,
et constamment enrichie grâce aux questions des étudiants !

stephanie.caet@gmail.com