

CHILDESワークショップ

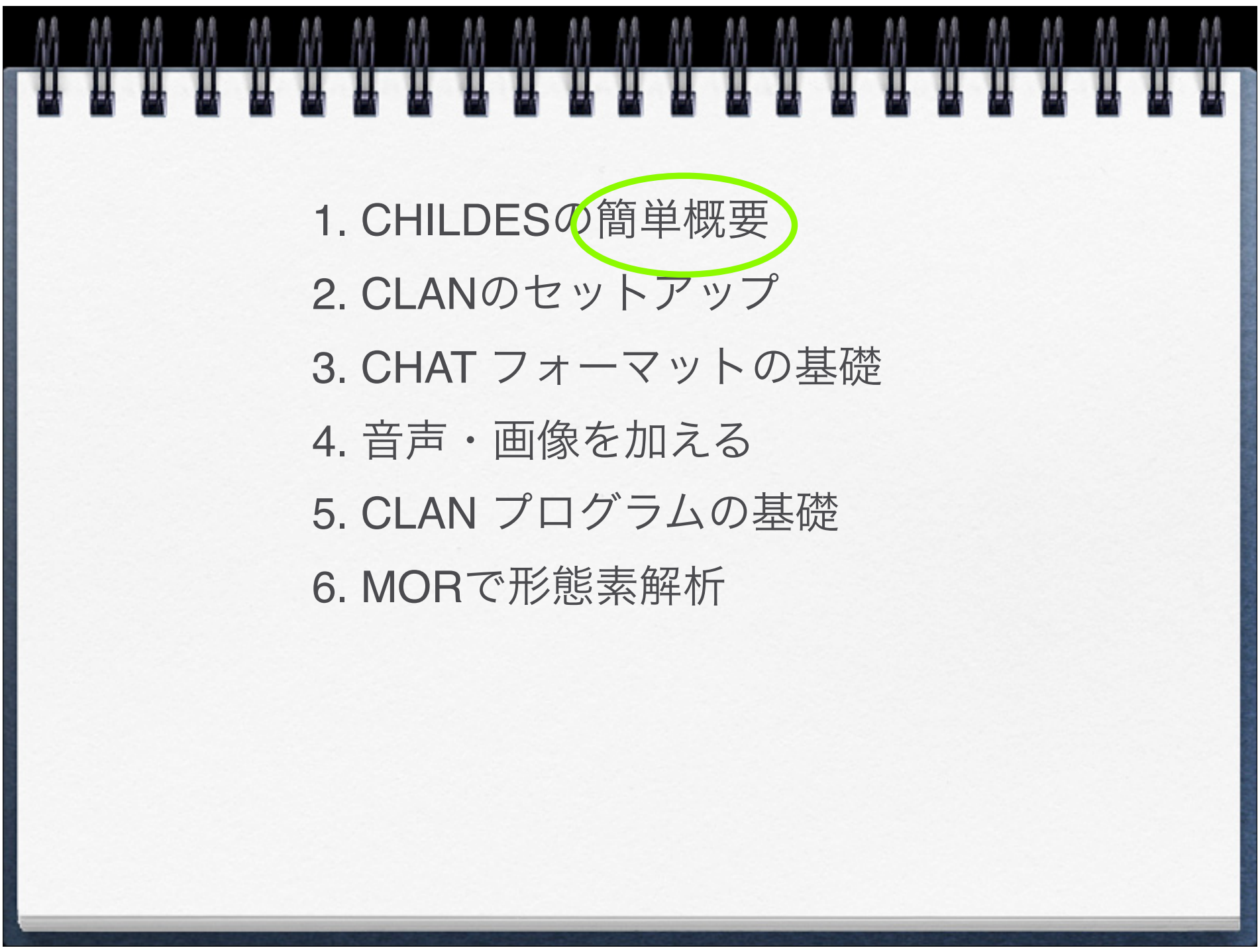


宮田 Susanne

(愛知淑徳大学医療福祉学部)

smiyata@asu.aasa.ac.jp

この資料は、英語コーパス学会（JAECs）第28大会（北海道大学 2006/10/7）のワークショップ
および特別講演のために作ったものにもとづきます。感謝を込めて編集したものを公開します。

- 
- A spiral-bound notebook with a black cover and silver spiral binding. The pages are white and slightly aged. The text is written in black on the first page.
1. CHILDESの簡単概要
 2. CLANのセットアップ
 3. CHAT フォーマットの基礎
 4. 音声・画像を加える
 5. CLAN プログラムの基礎
 6. MORで形態素解析

CHILDES

(チャイルズ、CHILd Language Data EXchange System)

非営利: 各種研究費とボランティア活動

-> 使用は無料・一般公開 4500メンバー

1984年 Brian MacWhinney & Catherine Snow



CHILDES

29ヶ国語の発話データ

第一言語獲得データ（英語:1049人のデータ）

- ・ 第二言語習得データ
- ・ バイリンガルデータ
- ・ 言語障害データ
- ・ 手話データ
- ・ story telling データ
- ・ オーディオ・ビデオ付きのデータ

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

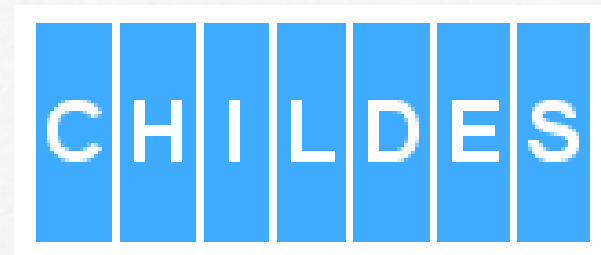
データ

+

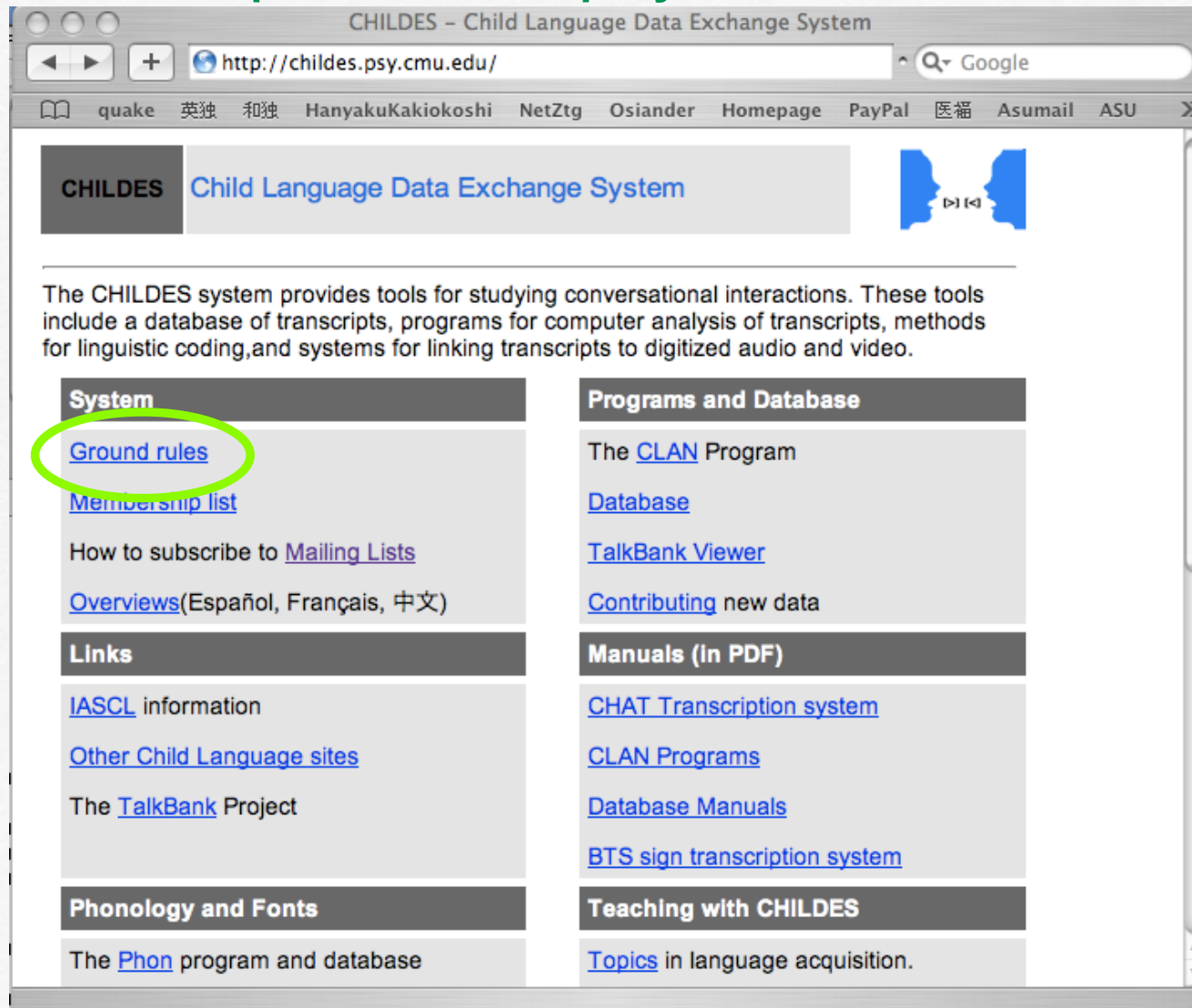
解析プログラム

+

入力フォーマット



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



The screenshot shows a web browser window with the title "CHILDES - Child Language Data Exchange System". The address bar displays "http://childes.psy.cmu.edu/". The browser's toolbar includes a search bar with "Google" and a list of bookmarks: "quake", "英独", "和独", "HanyakuKakiokoshi", "NetZtg", "Osiander", "Homepage", "PayPal", "医福", "Asumail", and "ASU". The website header features the "CHILDES" logo and the text "Child Language Data Exchange System" next to a graphic of two blue silhouettes of heads facing each other. Below the header, a paragraph describes the system: "The CHILDES system provides tools for studying conversational interactions. These tools include a database of transcripts, programs for computer analysis of transcripts, methods for linguistic coding, and systems for linking transcripts to digitized audio and video." The main content area is organized into two columns. The left column has a "System" header and contains links for "Ground rules" (circled in green), "Membership list", "How to subscribe to Mailing Lists", and "Overviews (Español, Français, 中文)". Below this is a "Links" header with links to "IASCL information", "Other Child Language sites", and "The TalkBank Project". At the bottom of the left column is a "Phonology and Fonts" header with a link to "The Phon program and database". The right column has a "Programs and Database" header with links to "The CLAN Program", "Database", "TalkBank Viewer", and "Contributing new data". Below this is a "Manuals (In PDF)" header with links to "CHAT Transcription system", "CLAN Programs", "Database Manuals", and "BTS sign transcription system". At the bottom of the right column is a "Teaching with CHILDES" header with a link to "Topics in language acquisition."

CHILDES - Child Language Data Exchange System

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

quake 英独 和独 HanyakuKakiokoshi NetZtg Osiander Homepage PayPal 医福 Asumail ASU

CHILDES Child Language Data Exchange System

The CHILDES system provides tools for studying conversational interactions. These tools include a database of transcripts, programs for computer analysis of transcripts, methods for linguistic coding, and systems for linking transcripts to digitized audio and video.

System	Programs and Database
Ground rules	The CLAN Program
Membership list	Database
How to subscribe to Mailing Lists	TalkBank Viewer
Overviews (Español, Français, 中文)	Contributing new data
Links	Manuals (In PDF)
IASCL information	CHAT Transcription system
Other Child Language sites	CLAN Programs
The TalkBank Project	Database Manuals
	BTS sign transcription system
Phonology and Fonts	Teaching with CHILDES
The Phon program and database	Topics in language acquisition.

基本ルール

- 倫理的使用（インフォームド・コンセント）
- アクセス許可レベル（完全公開～完全非公開9レベル）
- 参加者に対する尊敬（プライバシー）
- データ提供者に対する尊敬
- 引用義務（データ、フォーマット、プログラム使用の際）
- データ提供（データベースに貢献する）

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

CHILDES – Child Language Data Exchange System

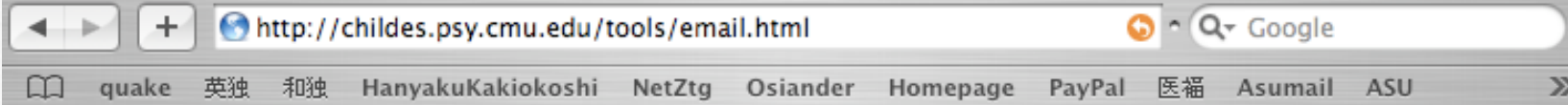
<http://childes.psy.cmu.edu/> Google

quake 英独 和独 HanyakuKakiokoshi NetZtg Osiander Homepage PayPal 医福 Asumail ASU >>

CHILDES Child Language Data Exchange System 

The CHILDES system provides tools for studying conversational interactions. These tools include a database of transcripts, programs for computer analysis of transcripts, methods for linguistic coding, and systems for linking transcripts to digitized audio and video.

System	Programs and Database
Ground rules	The CLAN Program
Membership list	Database
How to subscribe to Mailing Lists	TalkBank Viewer
Overviews (Español, Français, 中文)	Contributing new data
Links	Manuals (In PDF)
IASCL information	CHAT Transcription system
Other Child Language sites	CLAN Programs
The TalkBank Project	Database Manuals
	BTS sign transcription system
Phonology and Fonts	Teaching with CHILDES
The Phon program and database	Topics in language acquisition.



CHILDES Info-CHILDES

info-childes@mail.talkbank.org is an Internet list which promotes general discussion of issues relating to child language learning. It is not moderated, and only subscribers can post to the list. There are currently over 1600 subscribers and there are about six or seven messages posted to the list each week.

info-chibolts@mail.talkbank.org provides more technical discussions of the the CHILDES programs.

phon@mail.talkbank.org is a discussion list for the Phon program.

All of these lists can be received as weekly digests. To subscribe to any of these lists, send a request to Mary MacWhinney at mary-at-cmu-dot-edu.

To unsubscribe from info-childes send the message "unsubscribe info-childes" to requests@mail.talkbank.org.

To unsubscribe from info-childes-digest, info-chibolts, info-chibolts-digest, or phon, follow the same steps, but substitute the list name for info-childes.

If you would like to join the CHILDES system formally and be added to our larger database, please also send your postal address, phone number, institution, and other information. If possible, please give us your institutional/university address rather than your home postal address.

Past info-childes message and past [info-chibolts](#) messages are stored on the LinguistList server.

1. CHILDESの簡単概要
2. CLANのセットアップ
3. CHAT フォーマットの基礎
4. 音声・画像を加える
5. CLAN プログラムの基礎
6. MORで形態素解析

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

CHILDES - Child Language Data Exchange System

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

quake 英独 和独 HanyakuKakiokoshi NetZtg Osiander Homepage PayPal 医福 Asumail ASU >>

CHILDES Child Language Data Exchange System

The CHILDES system provides tools for studying conversational interactions. These tools include a database of transcripts, programs for computer analysis of transcripts, methods for linguistic coding, and systems for linking transcripts to digitized audio and video.

System	Programs and Database
Ground rules	The CLAN Program
Membership list	Database
How to subscribe to Mailing Lists	TalkBank Viewer
Overviews (Español, Français, 中文)	Contributing new data
Links	Manuals (In PDF)
IASCL information	CHAT Transcription system
Other Child Language sites	CLAN Programs
The TalkBank Project	Database Manuals
	BES sign transcription system
Phonology and Fonts	Teaching with CHILDES
The Phon program and database	Topics in language acquisition.

CHILDES Using CLAN

Windows	Macintosh	Mac 10.4	Unix
CLANWin	CLANX	CLANXu	UnixCLAN

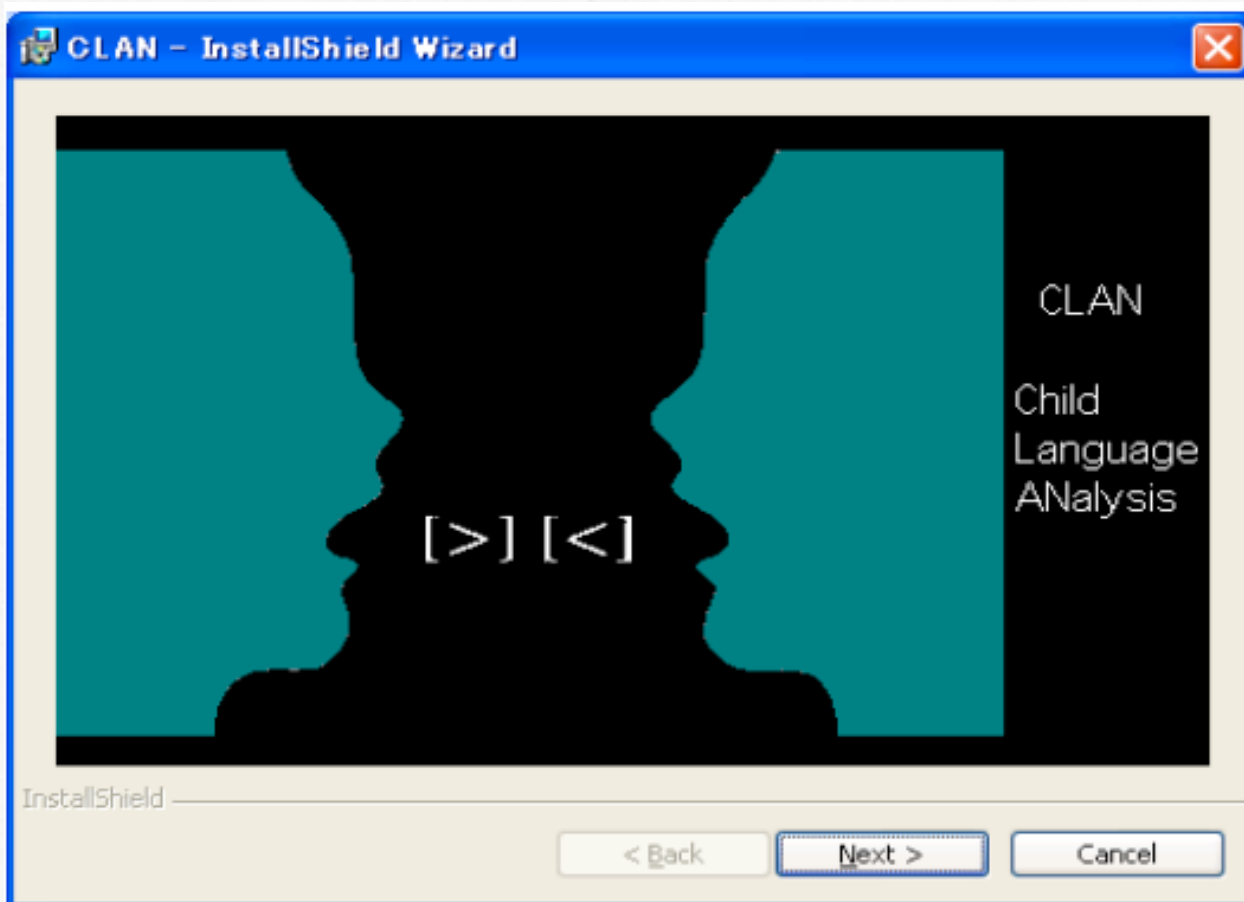
For Windows:

CLANWin is for Windows XP/2000/NT. 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, you must remove that version before installing the new version. Also, you will need to install [QuickTime](#) and [Unicode fonts](#).

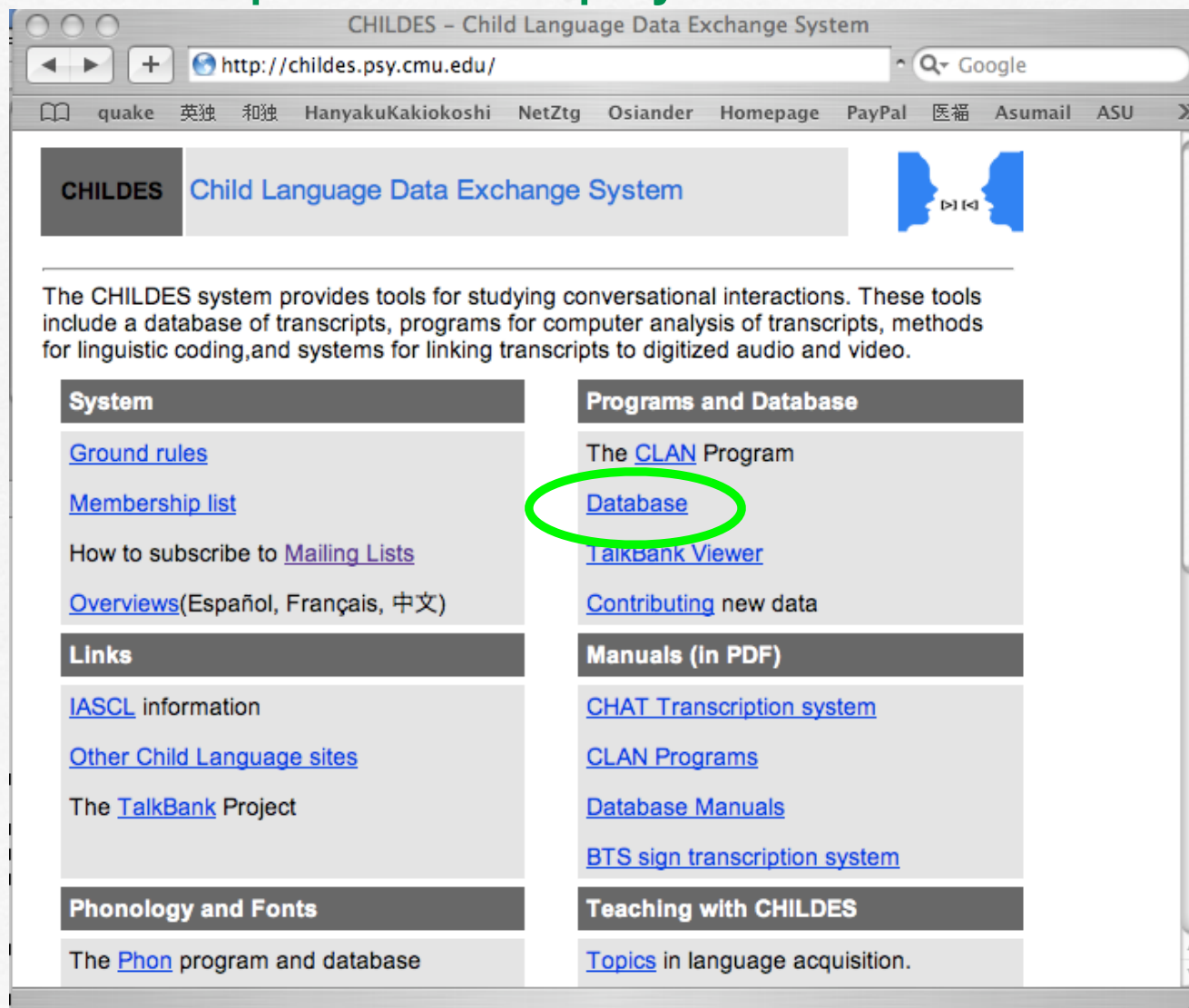
Japanese installation instructions are here [日本](#)

For Macintosh:

CLANXu is for Mac OS X users from 10.4 and up. Users with older Mac OS X should use CLANX. Mac OS 9 is no longer supported. Macintosh installation involves just unstuffing the file and double-clicking on the application. If you have an older version of CLAN on your machine, drag it into the trash and empty the trash before starting up the new version. Also, you will need to install [QuickTime](#) and [Unicode fonts](#).



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



The screenshot shows a web browser window with the title "CHILDES - Child Language Data Exchange System". The address bar displays "http://childes.psy.cmu.edu/". The browser's toolbar includes a search bar with "Google" and a list of bookmarks: "quake", "英独", "和独", "HanyakuKakiokoshi", "NetZtg", "Osiander", "Homepage", "PayPal", "医福", "Asumail", and "ASU". The website header features the "CHILDES" logo and the text "Child Language Data Exchange System" next to a graphic of two blue silhouettes of heads facing each other. Below the header, a paragraph describes the system: "The CHILDES system provides tools for studying conversational interactions. These tools include a database of transcripts, programs for computer analysis of transcripts, methods for linguistic coding, and systems for linking transcripts to digitized audio and video." The main content area is organized into two columns. The left column contains sections: "System" with links for "Ground rules", "Membership list", "How to subscribe to Mailing Lists", and "Overviews (Español, Français, 中文)"; "Links" with links for "IASCL information", "Other Child Language sites", and "The TalkBank Project"; and "Phonology and Fonts" with a link for "The Phon program and database". The right column contains sections: "Programs and Database" with links for "The CLAN Program", "Database" (circled in green), "TalkBank Viewer", and "Contributing new data"; "Manuals (In PDF)" with links for "CHAT Transcription system", "CLAN Programs", "Database Manuals", and "BTS sign transcription system"; and "Teaching with CHILDES" with a link for "Topics in language acquisition".

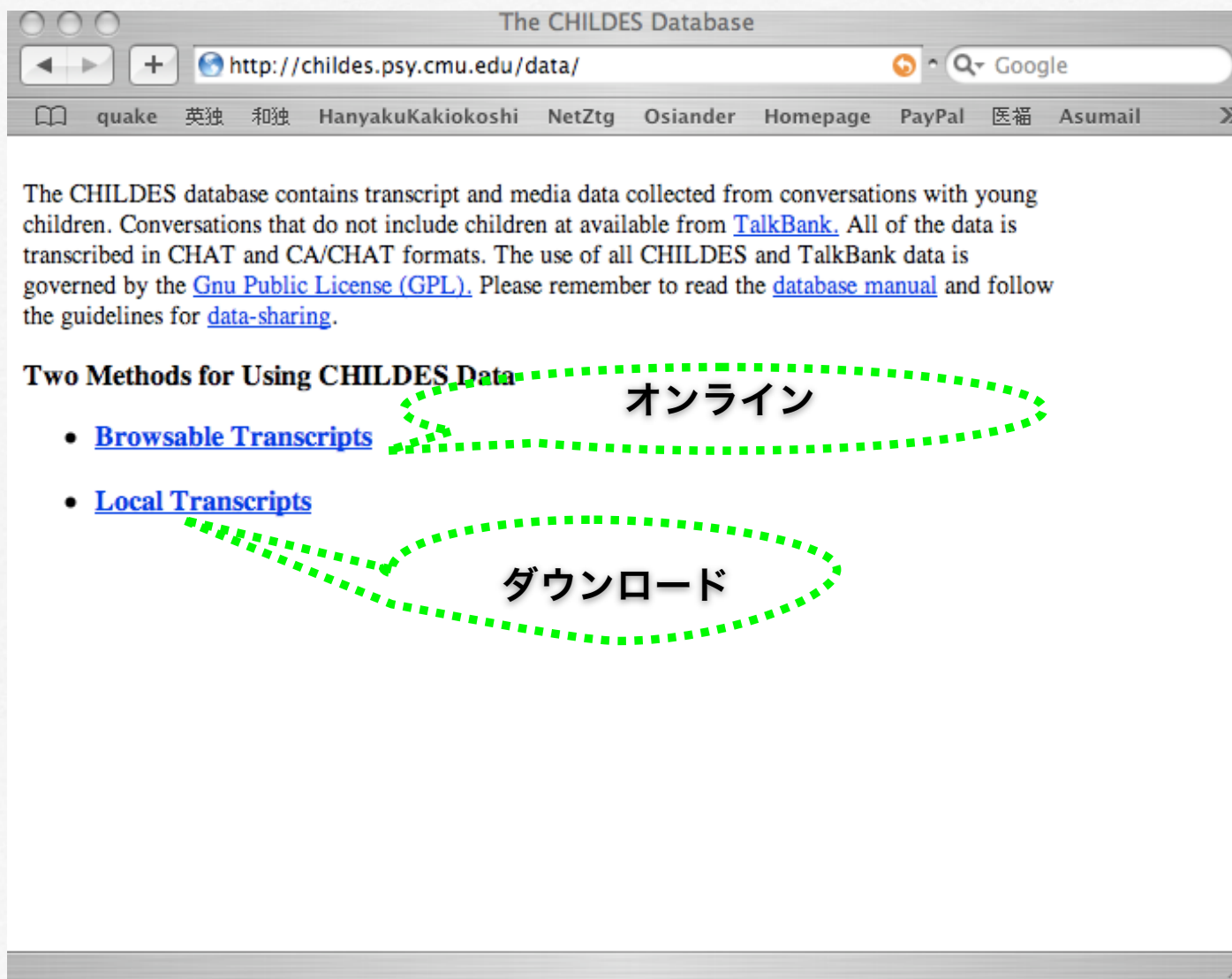
CHILDES Child Language Data Exchange System

The CHILDES system provides tools for studying conversational interactions. These tools include a database of transcripts, programs for computer analysis of transcripts, methods for linguistic coding, and systems for linking transcripts to digitized audio and video.

System	Programs and Database
Ground rules	The CLAN Program
Membership list	Database
How to subscribe to Mailing Lists	TalkBank Viewer
Overviews (Español, Français, 中文)	Contributing new data

Links	Manuals (In PDF)
IASCL information	CHAT Transcription system
Other Child Language sites	CLAN Programs
The TalkBank Project	Database Manuals
	BTS sign transcription system

Phonology and Fonts	Teaching with CHILDES
The Phon program and database	Topics in language acquisition.



Using Local Transcripts

http://chilides.psy.cmu.edu/data/local.html

quake 英独 和独 HanyakuKakiokoshi NetZtg Osiander Homepage PayPal 医福 Asumail

Using local transcripts and local media

1. You need to download the transcripts from "zipped transcripts" below and unzip them. If you want to work with XML versions of the data, rather than CHAT versions, please go to the directory of [zipped XML data files](#).
2. If the corpus is linked to [audio or video media](#), you need to download those files and place the media into the transcript folders.
3. You need to download and install the [CLAN](#) program.
4. To open a transcript, you double-click on it. If there is associated media, you can play the media using escape-8 for continuous playback or command-click for playing single utterances.

Zipped Transcripts

[English - USA](#)

[English - UK](#)

[Celtic](#)

[East Asian](#)

[Germanic](#)

[Romance](#)

[Slavic](#)

[Other Languages](#)

[Bilinguals](#)

Index of /data/EastAsian

http://chldes.psy.cmu.edu/data/EastAsian/

quake 英独 和独 HanyakuKakiokoshi NetZtg Osiander Homepage

Index of /data/EastAsian

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Descript</u>
 Parent Directory	23-Aug-2006 12:00	-	
 Cantonese/	16-Sep-2006 12:34	-	
 Chinese/	16-Sep-2006 12:34	-	
 Japanese/	16-Sep-2006 12:34	-	
 Korean/	16-Sep-2006 12:34	-	
 Thai/	16-Sep-2006 12:34	-	

Apache/1.3.33 Server at chldes.psy.cmu.edu Port 80

Index of /data/EastAsian/Japanese

http://chilides.psy.cmu.edu/data/EastAsian/Japanese/

quake 英独 和独 HanyakuKakiokoshi NetZtg Osiander Homepage Payf

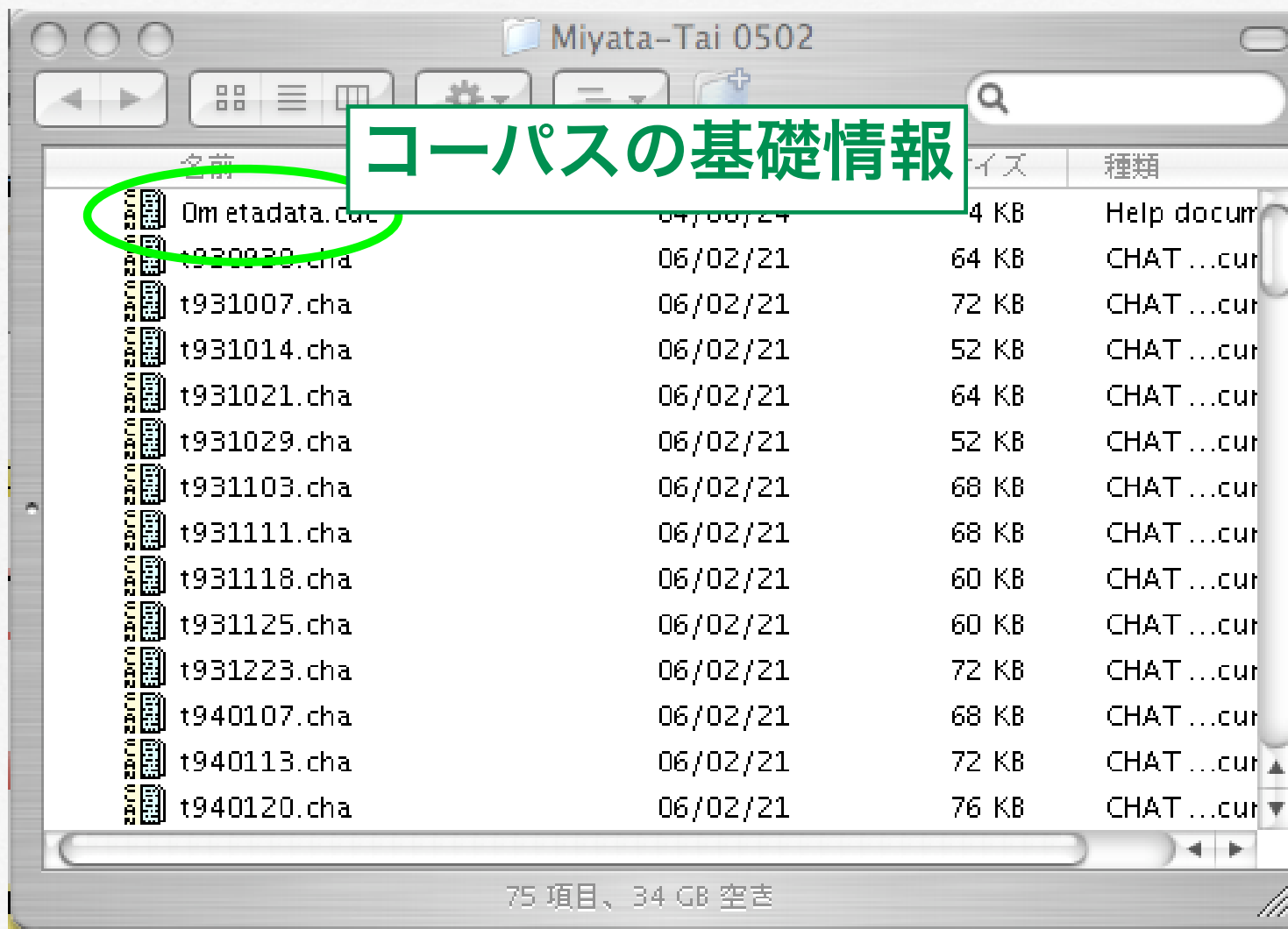
Index of /data/EastAsian/Japanese

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
 Parent Directory	19-Jun-2006 18:49	-	
 Hamasaki.zip	16-Sep-2006 12:34	1.3M	
 Ishii.zip	16-Sep-2006 12:34	1.3M	
 Miyata-Aki.zip	16-Sep-2006 12:34	418k	
 Miyata-Ryo.zip	16-Sep-2006 12:34	181k	
 Miyata-Tai.zip	16-Sep-2006 12:34	1.4M	
 Noji.zip			

Apache/1.3.33 Server at cmu.edu

ZIPファイルとしてダウンロード

コーパスの基礎情報



名前	サイズ	種類
0m etadata.doc	4 KB	Help docum
t930039.cha	64 KB	CHAT...cur
t931007.cha	72 KB	CHAT...cur
t931014.cha	52 KB	CHAT...cur
t931021.cha	64 KB	CHAT...cur
t931029.cha	52 KB	CHAT...cur
t931103.cha	68 KB	CHAT...cur
t931111.cha	68 KB	CHAT...cur
t931118.cha	60 KB	CHAT...cur
t931125.cha	60 KB	CHAT...cur
t931223.cha	72 KB	CHAT...cur
t940107.cha	68 KB	CHAT...cur
t940113.cha	72 KB	CHAT...cur
t940120.cha	76 KB	CHAT...cur

75 項目、34 GB 空き

Macintosh HD:Users:sm:SU:d a t a:TAI:Tai.jmor+KANA+snd2006/3:0metadata.cdc
Title: Japanese - Miyata - Tai Corpus

Creator: Miyata, Susanne

Subject: child language development

Subject.olac:linguistic-field: language_acquisit

Subject.olac:language: ja

Subject.chldes:participant: age="1;5.20 - 3;1.29" sex="ma

Description: longitudinal study of Japanese boy (Tai) in h
Wakachi2002 v.2.1, JMOR02 (2005, January)

Publisher: TalkBank

Contributor:

Date: 2004-03-30

Type: text

Type: sound

Type.olac:linguistic-type: primary_text

Type.olac:discourse-type: dialo

Format:

Identifier: 1-59642-057-X

Language:

Relation:

書籍のように引用

Miyata, S. (2004) *Japanese - Miyata - Tai Corpus*
Pittsburgh:Talkbank. ISBN 1-59642-057-X

ISBN番号

C:\CHILDRES

CLAN

CLAN.exe

lib

...

MOR

english

JMOR03

work

hokkaido.cha

ダウンロードしたCLANフォルダー

CLANプログラム

C:\CHILDES

CLAN

CLAN.exe

lib

形態素解析用のフォルダー

言語別にダウンロード

MOR

english

JMOR03

work

hokkaido.cha

C:\CHILDES

CLAN

CLAN.exe

lib

...

MOR

english

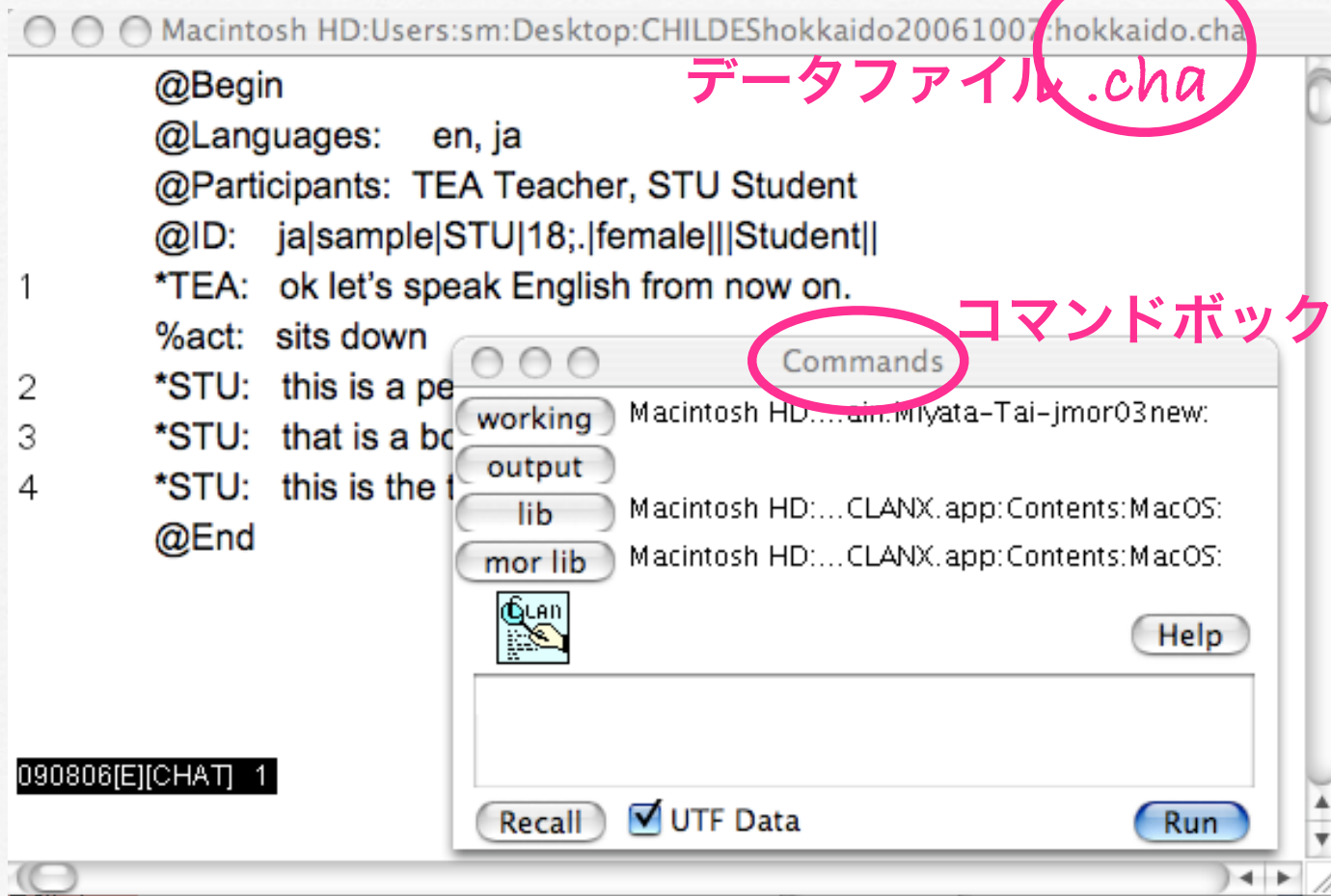
JMOR03

データファイルのフォルダー

work

hokkaido.cha

1. CHILDESの簡単概要
2. CLANのセットアップ
3. CHAT フォーマットの基礎
4. 音声・画像を加える
5. CLAN プログラムの基礎
6. MORで形態素解析



Macintosh HD:Users:sm:Desktop:CHILDEShokkaido20061007:hokkaido.cha

@Begin

@Languages: en, ja

@Participants: TEA Teacher, STU Student

@ID: ja|complex|T|C|S|E|||Student||

*TEA: ok let's speak English from now on.

%act: sits*down

*STU: this is a pe

*STU: that is a br

*STU: this is the

@End

@ 基本情報

% 発話

% その発話に関する情報



090806[E][CHAT] 1

@ 基本情報

@Begin

@Languages: ja

@Participants: CHI Kentakun Child, MOT Mother,
INV Investigator

@ID: jalsampleIChII2;6.12ImaleIIIChildII

@Comment: overall time 75'00"

* 発話

*MOT: moo tsunagenai .

*CHI: iya da .

* 発話

*MOT: moo tsunagenai .

*CHI: iya da .

3文字発話者コード

% その発話に関する情報

*MOT: moo tsunagenai .

%ort: もうつなげない。

*CHI: iya da .

%ort: 嫌だ。

%com: 泣き出す

@Begin

@Languages: en, ja

@Participants: TEA Teacher, STU Student

@ID: jalsample|STU|18;.lfemale|||Student||

*TEA: ok let's speak English from now on.

%act: sits down

*STU: iya da .

%ort: いやだ。

*STU: that is a book

*STU: this is the ta

@End

hokkaido.cha
(サンプルファイル)

入力フォーマット

Chapter 4



@Begin

@Languages: en, ja

@Participants: TEA Teacher, STU Student

@ID: jalsample|STU|18;.Ifemale|Student|

*TEA: ok let's speak English from now on.

%act: sits down

*STU: iya da.

%ort: いやだ。

*STU: that is a book.

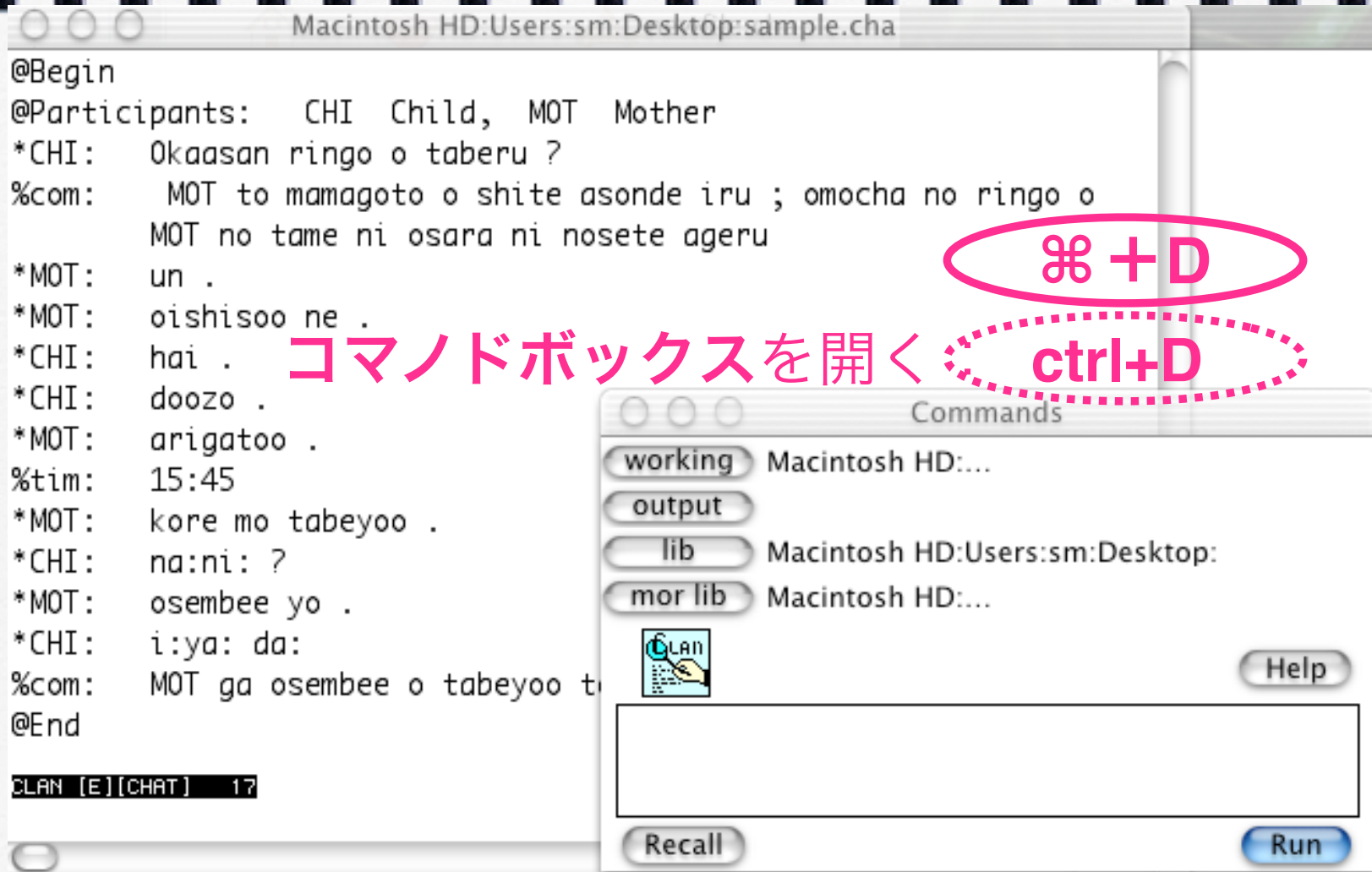
*STU: this is the table.

@End

とりあえずCLANを走らせよう

Chapter 5





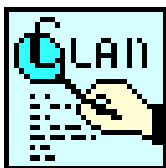
作業フォルダを設定

working

output

lib

mor lib



C:¥CHILDRES

└ CLAN

└ CLAN.exe

└ lib

└ ...

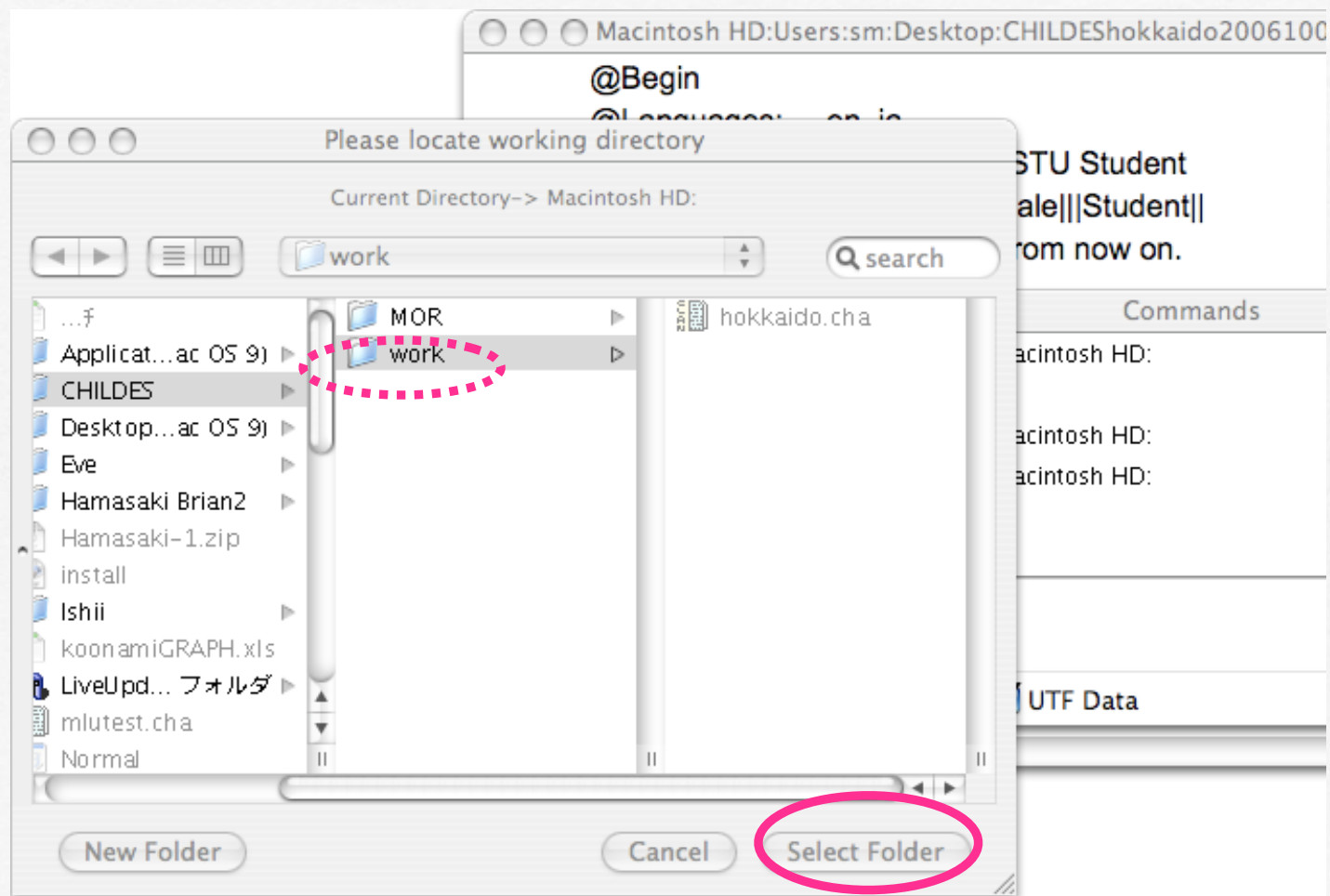
└ MOR

└ ...

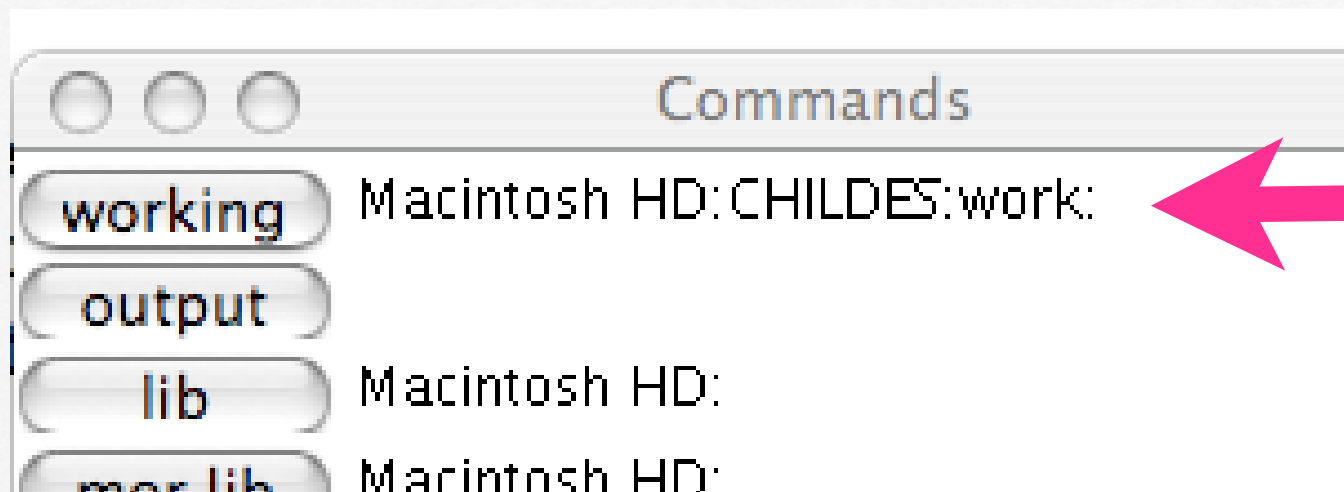
└ **work**

└ hokkaido.cha

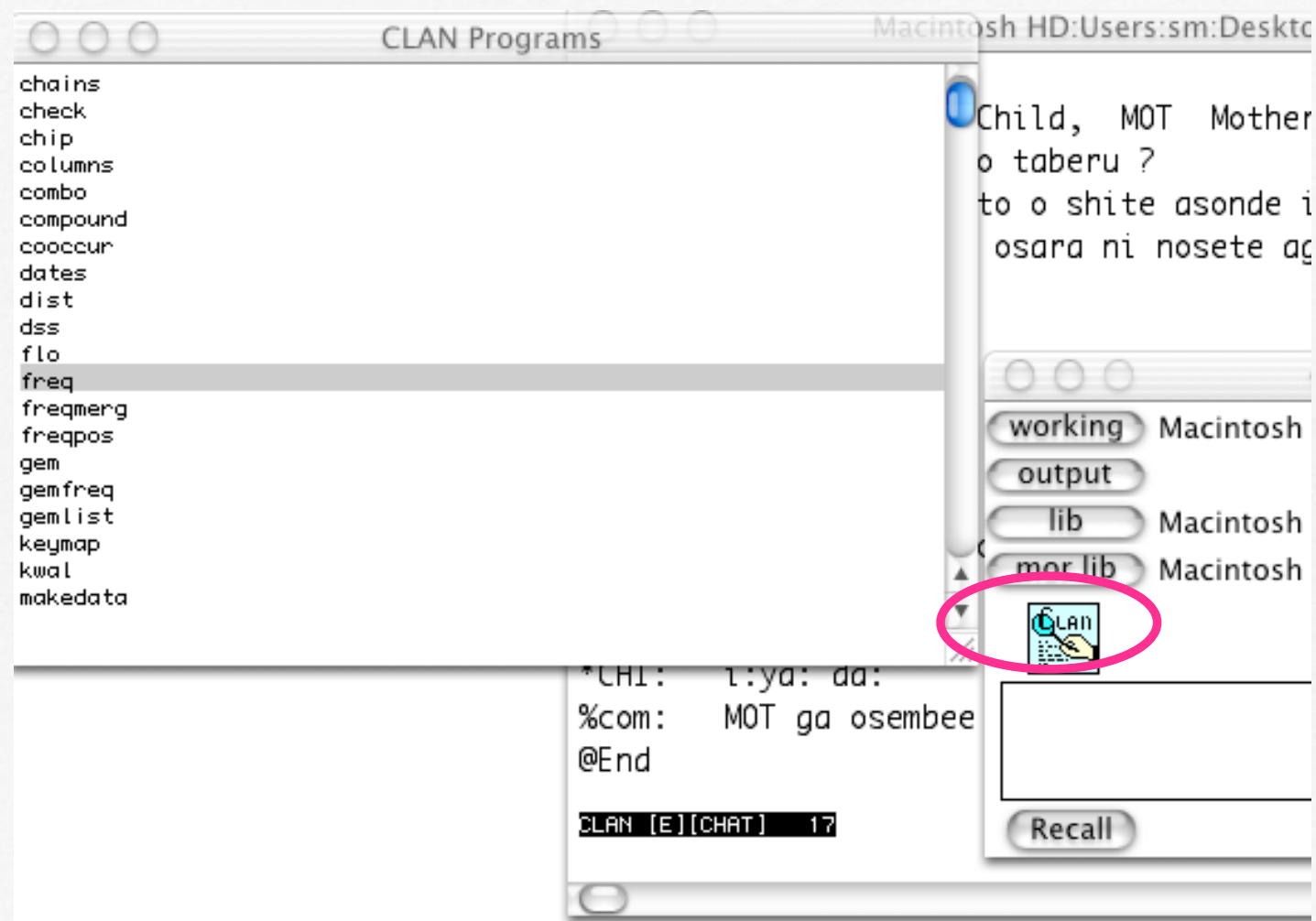
作業フォルダを設定



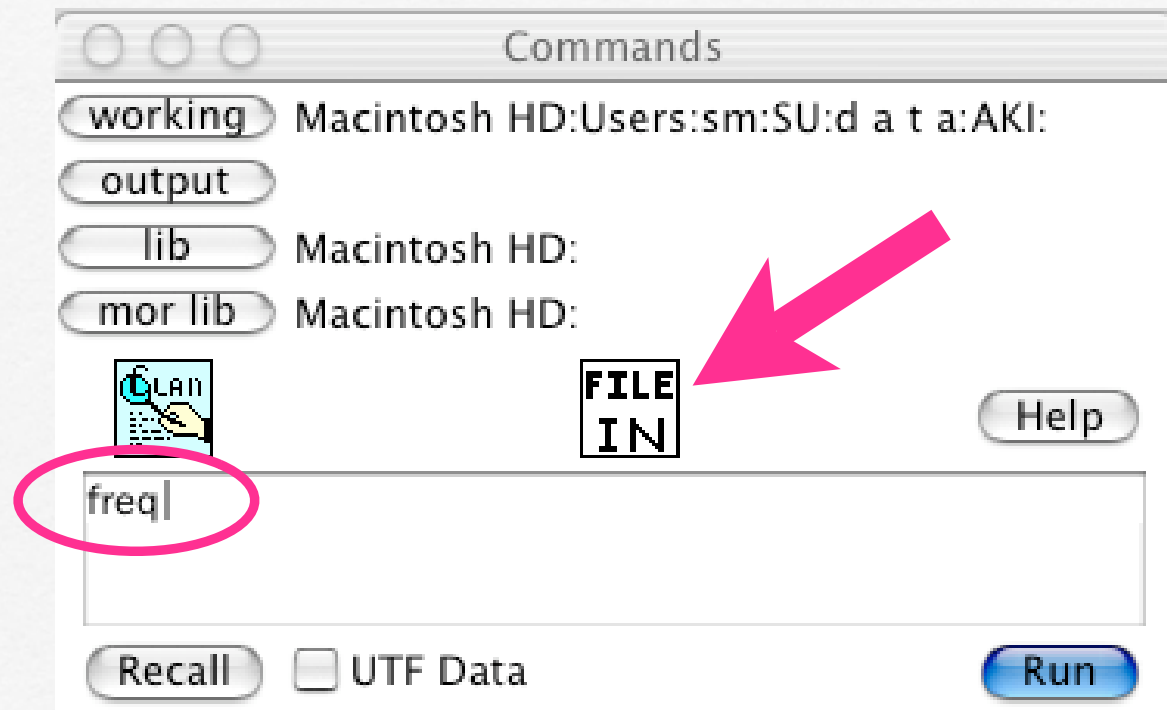
作業フォルダが設定されているかを確認



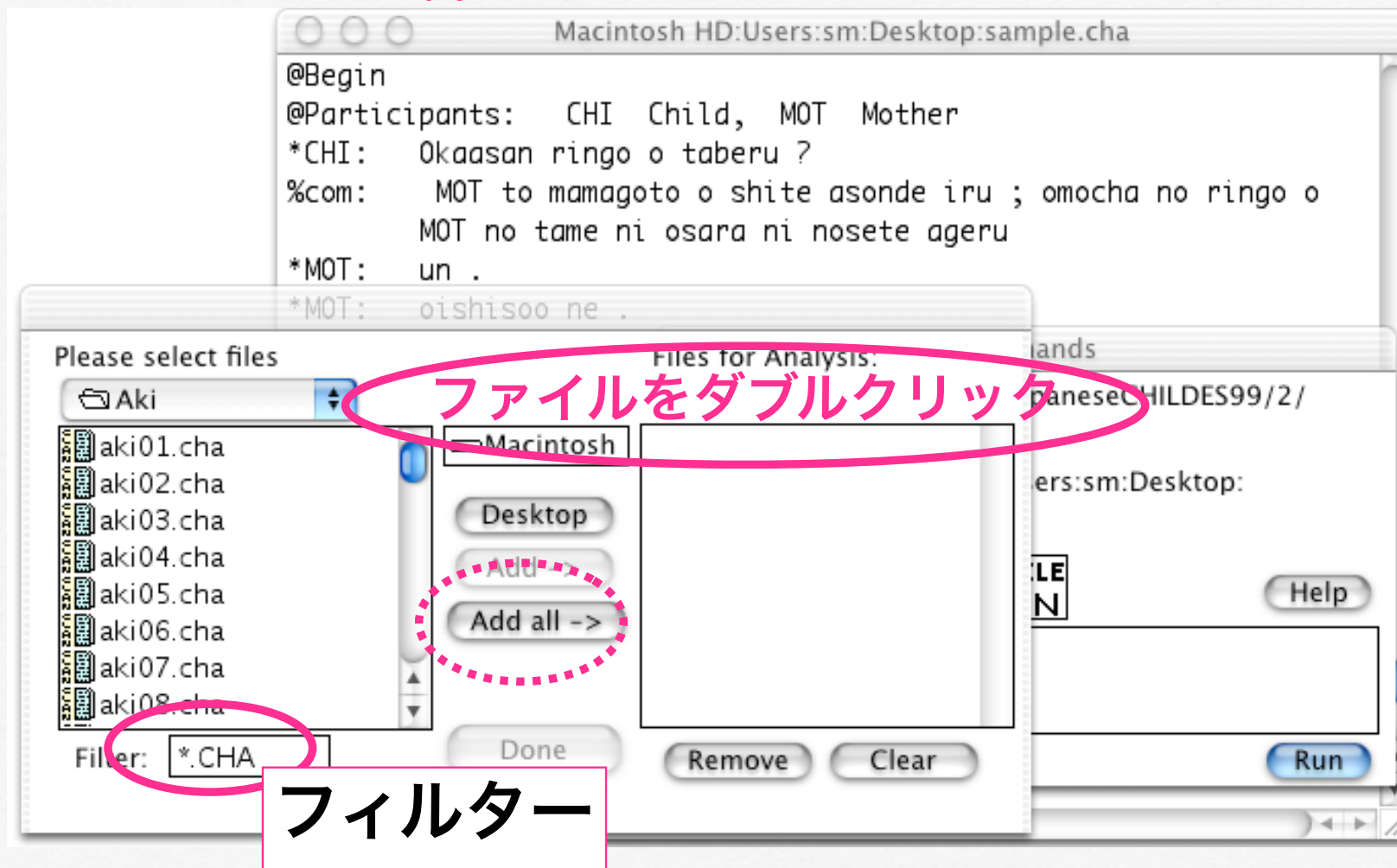
コマンドを選ぶ



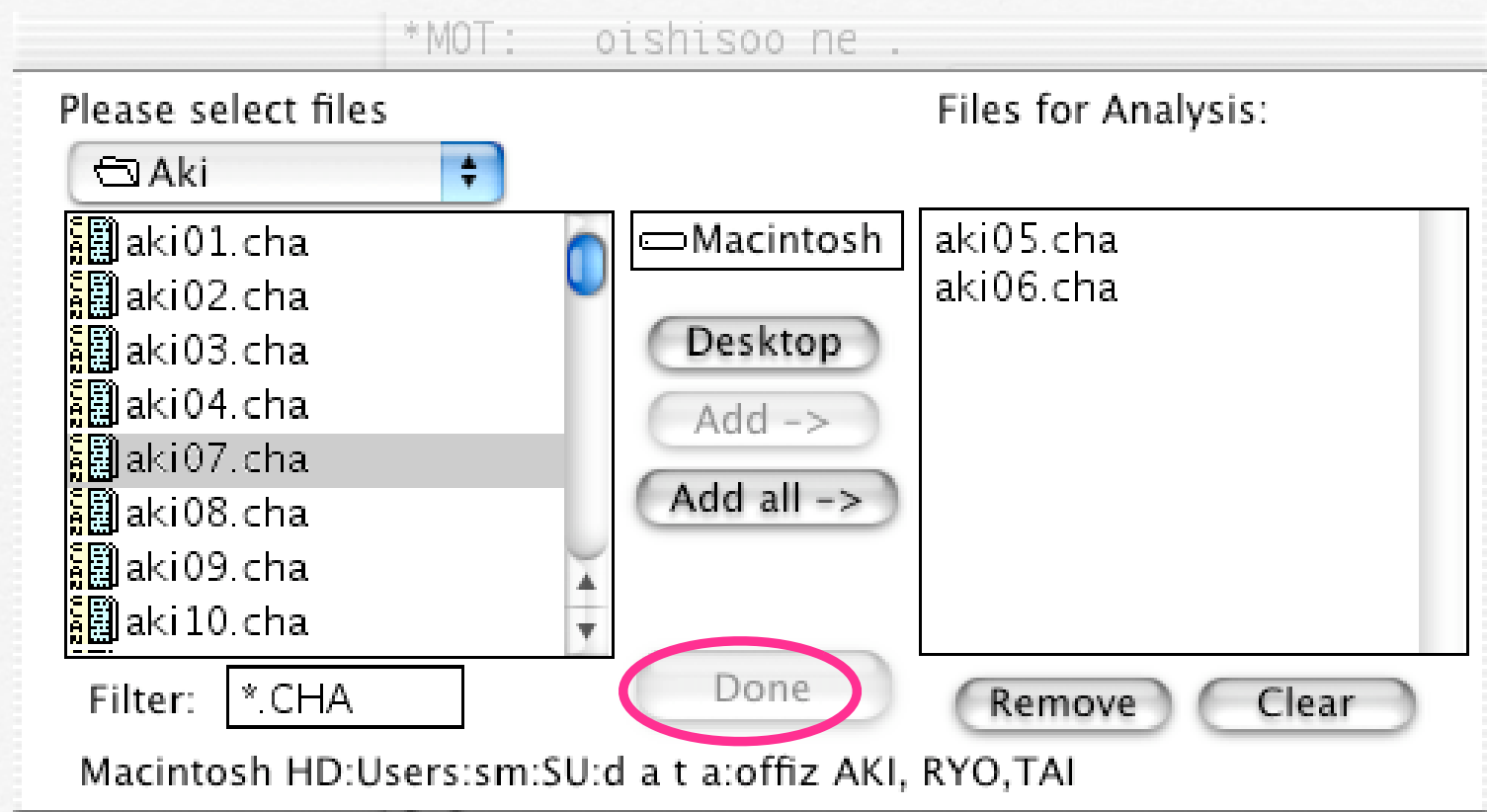
分析するファイルを選ぶ



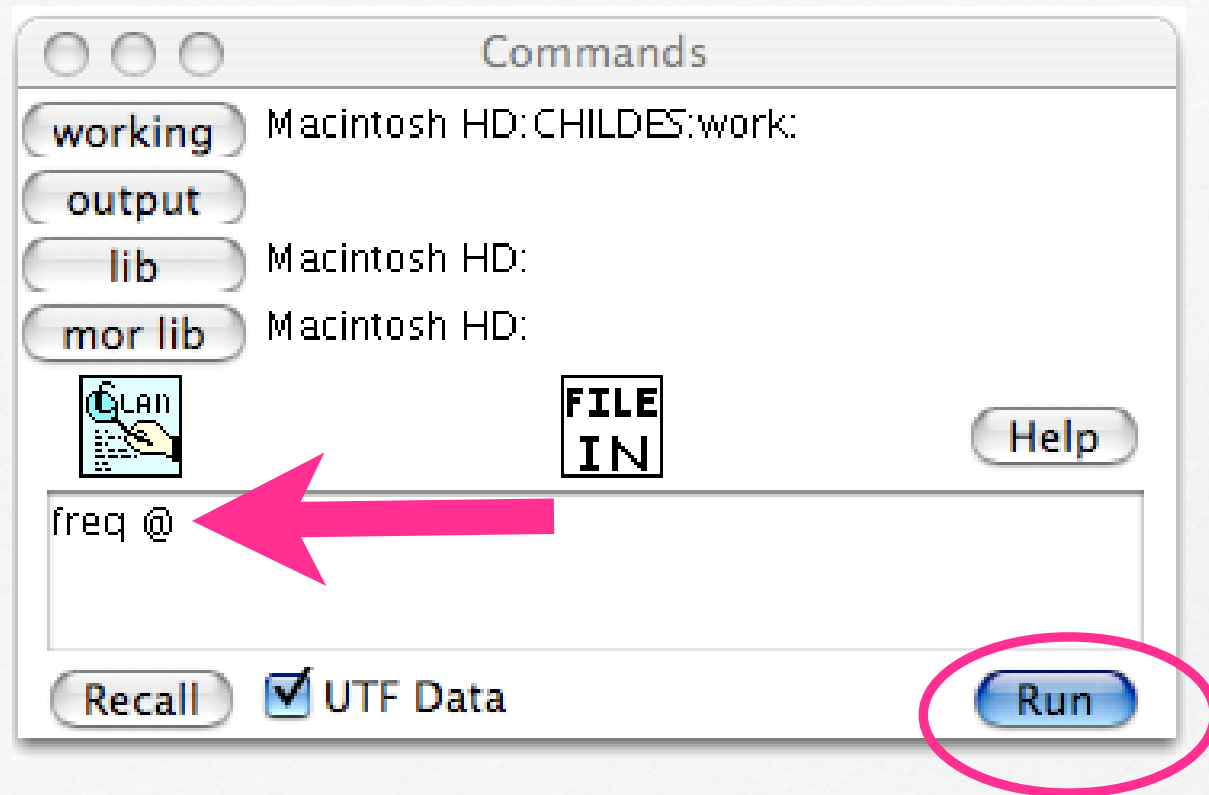
分析するファイルを選ぶ



分析するファイルを選ぶ



コマンドを打つ



1 yaroo
11 yatte
49 yo
1 yoisho
1 yoku
1 yonde
1 yoshi
4 yottsu
4 yubi
1 yubiwa
1 zembu
2 zoo
3 zoosan

FREQ

377 Total number of different word types used
2013 Total number of words (tokens)
0.187 Type/Token ratio

タイプ

トークン

TTR

コマンドの構造

コマンド

FREQ

FREQ

KWAL

KWAL

+t*CHI

+t*TEA

+s"this"

+s"this" +t*STU

ファイル名

hokkaido.cha

@

@

@

FREquency 頻度計算・単語一覧

コマンドの構造

コマンド

FREQ

FREQ

KWAL

KWAL

オプション

+t*STU

+t*TEA

+s"this"

+s"this" +t*STU

ファイル名

hokkaido.cha

@

@

@

+t この行を対象に (**t**urn)

コマンドの構造

コマンド

FREQ
FREQ
KWAL

KWAL

オプション

+t*CHI
+t*TEA
+s"this"

+s"this" +t*STU

ファイル名

hokkaido.cha
@
@

@

+s これを探せ(**s**earch)

KeyWord And Line ある単語を含む発話の一覧

コマンドの構造

コマンド

FREQ

FREQ

KWAL

KWAL

...

オプション

+t*CHI

+t*TEA

+s"this"

+s"this" +t*STU

...

ファイル名

hokkaido.cha

@

@

@

+s +t

1. CHILDESの簡単概要
2. CLANのセットアップ
3. CHAT フォーマットの基礎
4. 音声・画像を加える
5. CLAN プログラムの基礎
6. MORで形態素解析

Macintosh HD:Users:sm:SU:d a t a:MII PRO:3saidai fuer K...

%jpn: ーい。

MAJ: yappa Natchan ga yatta n da yo ,, hannin wa [] . .

%jpn: やっぱなっちゃんがやったんだよ、犯人は。

*MAJ: Miichan janai mon |.

%jpn: みーちゃんじゃないもん。

*MAJ: Natchin ga hannin da yo .

%jpn: なっちゃんが犯人だよ。

*NAT: O . .

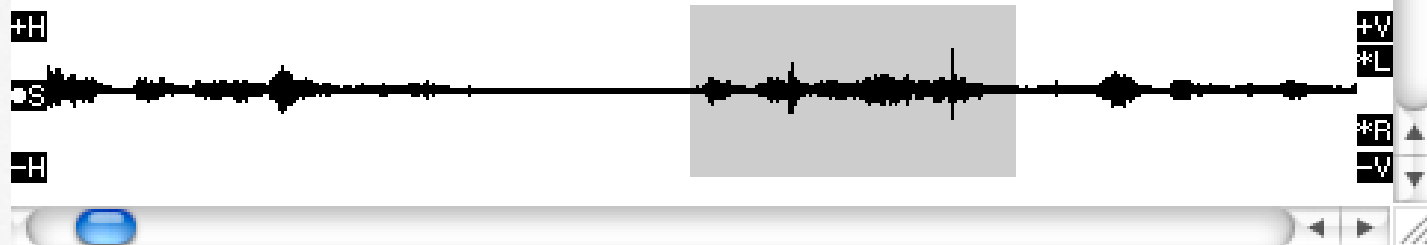
%jpn: うええ。

*MAJ: ne , Papa . .

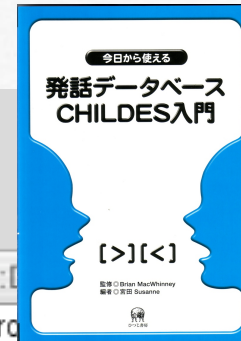
%jpn: ねーパパ。

CLAN [E][CHAT] 132

音声・画像の再生と作成



Chapter.10+11



Macintosh HD:Users:sm:1...cha

%ges: pointing with L hand toward atlas from
*NOR: if y- it's the bottom of this thing. #2.4 •
%ges: rises from chair and moves toward flipchart. •
*NOR: It's right in here. •
%ges: Places forefinger of R hand on chart and twice
trace the lower edge of a structure in the sagittal section,
right-most, second row. #.8 •
UNI: (Its un xxxxx xxxx) •
*MAR: I think it's un:der. •
%ges: Norman returns to seat. •
*NOR: It's under that? = •
*MAR: =I think it's on the inside. •
*COA: It's on the middle, (0.3) middle top. #0.4 •
U0903[E][CHAT-UTF] 29

1. CHILDESの簡単概要
2. CLANのセットアップ
3. CHAT フォーマットの基礎
4. 音声・画像を加える
5. CLAN プログラムの基礎
6. MORで形態素解析

検索の基礎

FREquency 頻度計算・単語一覧

コマンド

FREQ

FREQ

FREQ

FREQ

+t*CHI +o

+s"this"

+s"th*"

+s"th*" +d

+o 頻度順

ファイル名

hokkaido.cha

@

@

@

検索の基礎

FREquency 頻度計算・単語一覧

+0 頻度順

FREQ	+t*CHI	+0	hokkaido.cha
FREQ	+s"this"		@
FREQ	+s"th*		@
FREQ	+s"th*" +d		@

* ワイルドカード

検索の基礎

FREQuency 頻度計算・単語一覧

FREQ	+t*CHI	+o	hokkaido.cha
FREQ	+s"this"	* ワイルドカード	@
FREQ	+s"th*"		@
FREQ	+s"th*" +d	+d 行番号	@

その番号にトリプルクリックすると、ファイルが開いてその行まで飛ぶ

検索の基礎

FREQuency 頻度計算・単語一覧

+0 頻度順

FREQ	+t*CHI	+0	hokkaido.cha
FREQ	+s"this"	* ワイルドカード	@
FREQ	+s"th*"		@
FREQ	+s"th*" +d	+d 行番号	@

Chapter 8



Keyword And Line 特定単語を含む発話の一覧

検索の基礎

KWAL +t*CHI +s"this" @

KWAL +t*CHI +s"this" +d2 @

KWAL -w2 +w2 @

Keyword And Line 特定単語を含む発話の一覧

検索の基礎

KWAL +t*CHI +s"this"

KWAL +t*CHI +s"this" +d2

KWAL -w2 +w2

+d2 発話のみ

@

@

@

Keyword And Line 特定単語を含む発話の一覧

検索の基礎

KWAL +t*CHI +s"this"

KWAL +t*CHI +s"this" +d2

KWAL -w2 +w2

+d2 発話のみ

@

@

@

+W ウィンドー

ウィンドーで前後の発話を
表示できる

Keyword And Line 特定単語を含む発話の一覧

検索の基礎

KWAL

+t*CHI +s"this"

+d2 発話のみ

@

KWAL

+t*CHI +s"this" +d2

@

KWAL

-w2 +w2

@

+W ウィンドー

Chapter 9



検索の基礎

MLU 平均発話長 で発話数・単語数

-t 形態素行%morを無視

MLU

-t%mor

@

MLU

-t%mor +d1

@

...無視というか、
%morが付いていないから
強制的に発話行を対象にする

CLAN Output

> mlu -t%mor @

mlu -t%mor @

Tue Sep 19 18:10:22 2006

mlu (09-Aug-2006) is conducting analyses on:

ALL speaker tiers

発話者

From file <Macintosh HD:User:smi.SU:Wak1:bri:EuerPosttrain:Miyata-Tai-jmor03new:hokkaido.cha>

MLU for Speaker: *STU:

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

Number of: utterances = 3, morphemes = 12

Ratio of morphemes over utterances = 4.000

Standard deviation = 0.000

単語数

MLU for Speaker: *TEA:

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

Number of: utterances = 1, morphemes = 7

Ratio of morphemes over utterances = 7.000

発話数

090806[E][TEXT] * 1

検索の基礎

MLU 平均発話長 で発話数・単語数

-t 形態素行%morを無視

MLU

-t%mor

@

+d1 数字のみ

MLU

-t%mor +d1

@

CLAN Output

```
> mlu -t%mor +d1 @
mlu -t%mor +d1 @
Tue Sep 19 18:20:13 2006
mlu (09-Aug-2006) is conducting analyses on:
  ALL speaker tiers
*****
From file <Macintosh HD:Users:sm:SU:Wak1:bri:FuerPosttrain:Mi
*STU:  3  12  4.000 .
*TEA:  1   7  7.000 .
>
```

発話者： 発話数 単語数 平均発話長

検索の基礎

MLU 平均発話長 で発話数・単語数

-t 形態素行%morを無視

MLU

-t%mor

@

+d1 数字のみ

MLU

-t%mor +d1

@

Chapter 7



COMBO combination 単語の組み合わせを探す

検索の基礎

+s ^ 組み合わせ

COMBO +s"this"^"is" @

COMBO +s"this"^*^"a" @

COMBO combination 単語の組み合わせを探す

検索の基礎

+s ^ 組み合わせ

COMBO +s"this"^"is" @

COMBO +s"this"^*"^"a" @

* ウアイルドカード

CLANで解析

FREQ frequency +s“word” +s”word*”
KWAL keyword and line
COMBO combination of keywords +s“word”^“word”
WDLEN word length in letters, morphemes & words
PHONFREQ phonetic frequency
MLU (TTR) mean length of utterance & type/token ratio
VOCD vocabulary diversity
DSS developmental sentence score
DATES birthday / date / age +b1-JAN-2000
MOR morphological analysis

1 yaroo
11 yatte
49 yo
1 yoisho
1 yoku
1 yonde
1 yoshi
4 yottsu
4 yubi
1 yubiwa
1 zembu
2 zoo
3 zoosan

FREQ

377 Total number of different word types used
2013 Total number of words (tokens)
0.187 Type/Token ratio

タイプ

トークン

TTR

CLANで解析

FREQ frequency +s“word” +s”word*”

KWAL keyword and line

COMBO combination of keywords +s“word”^“word”

WDLEN word length in letters, morphemes & words

PHONFREQ phonetic frequency

MLU (TTR) mean length of utterance & type/token ratio

VOCD vocabulary diversity

DSS developmental sentence score

DATES birthday / date / age +b1-JAN-2000

MOR morphological analysis

FREQ +s"core"

135 kore

1 Total number of different word types used
135 Total number of words (tokens)
0.007 Type/Token ratio

CLANで解析

FREQ **frequency** +s“word” **+s”word*”**

KWAL keyword and line

COMBO combination of keywords +s“word”^“word”

WDLEN word length in letters, morphemes & words

PHONFREQ phonetic frequency

MLU (TTR) mean length of utterance & type/token ratio

VOCD vocabulary diversity

DSS developmental sentence score

DATES birthday / date / age +b1-JAN-2000

MOR morphological analysis

FREQ +s"ko*"

35 koko

2 kono

4 koo

1 kooen

9 kootai

135 kore

2 koronjatta

53 kotchi

1 kowarechatta

ko... (ワイルドカード)

9 Total number of different word types used

242 Total number of words (tokens)

0.037 Type/Token ratio

CLANで解析

FREQ frequency +s“word” +s”word*”

KWAL keyword and line

COMBO combination of keywords +s“word”^“word”

WDLEN word length in letters, morphemes & words

PHONFREQ phonetic frequency

MLU (TTR) mean length of utterance & type/token ratio

VOCD vocabulary diversity

DSS developmental sentence score

DATES birthday / date / age +b1-JAN-2000

MOR morphological analysis

KWAL +s"core"

*TMO: nani ,, kore ? •

*TMO: kore na:n da ? •

*TAI: kore ? •

音声リンク付き

*TMO: kore . •

*TMO: Taishoo, kore wa ? •

*TMO: ja kore wa ? •

*TMO: kore da yo ,, Taishoo no daisuki na no . •

CLANで解析

FREQ frequency +s“word” +s”word*”

KWAL keyword and line

COMBO combination of keywords +s“word”^“word”

WDLEN word length in letters, morphemes & words

PHONFREQ phonetic frequency

MLU (TTR) mean length of utterance & type/token ratio

VOCD vocabulary diversity

DSS developmental sentence score

DATES birthday / date / age +b1-JAN-2000

MOR morphological analysis

COMBO +s"kore"^^"da"

- *TMO: kore na:n da ? •
- *TMO: kore da yo , , Taishoo no daisuki na no . •
- *TMO: kore ne ii kaori ga suru n da yo , , oiru ga . •
- *TMO: kore da yo ? •
- *TMO: kore wa rappa da mon . •
- *TMO: kore kishapoppo da mon . •
- *TMO: un , kore mo issho da ne . •
- *TMO: kore nan da ? •
- *TMO: &ka [//] kore jangurujimu da ne . •
- *TMO: kore ne shinkonryokoo no shashin da yo . •

CLANで解析

FREQ	frequency	+s“word” +s”word*”
KWAL	keyword and line	
COMBO	combination of keywords	+s“word”^“word”
WDLEN	word length	in letters, morphemes & words
PHONFREQ	phonetic frequency	
MLU (TTR)	mean length of utterance & type/token ratio	
VOCD	vocabulary diversity	
DSS	developmental sentence score	
DATES	birthday / date / age	+b1-JAN-2000
MOR	morphological analysis	

Number of words of each length in characters

文字数

WDLEN

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Mean
*TAI:	14	35	95	75	145	43	39	11	4	6	2	1	0	0	4.56
*SUU:	16	28	22	32	8	12	6	6	2	2	0	1	0	0	4.40
*TMO:	46	325	172	183	168	78	94	37	20	21	4	3	0	2	4.19

Number of words of each of these morpheme lengths

	1	Mean
*TAI:	470	1.00
*SUU:	135	1.00
*TMO:	1153	1.00

Number of utterances of each of these lengths in morphemes

形態素数

	1	2	3	4	5	6	7	8	9	10	11	12	13	Mean
*TAI:	187	77	13	9	3	5	0	0	1	0	0	0	0	1.59
*SUU:	46	13	9	4	4	0	0	0	0	0	0	0	0	1.63
*TMO:	287	82	59	44	22	17	9	1	2	1	0	1	2	1.96

Number of utterances of each of these lengths in words

単語数

	1	2	3	4	5	6	7	8	9	10	11	12	13	Mean
*TAI:	187	77	13	9	3	5	0	0	1	0	0	0	0	1.59
*SUU:	46	13	9	4	4	0	0	0	0	0	0	0	0	1.63
*TMO:	287	82	59	44	22	17	9	1	2	1	0	1	2	1.96

CLANで解析

FREQ	frequency	+s“word” +s”word*”
KWAL	keyword and line	
COMBO	combination of keywords	+s“word”^“word”
WDLEN	word length	in letters, morphemes & words
PHONFREQ	phoneme frequency	
MLU (TTR)	mean length of utterance & type/token ratio	
VOCD	vocabulary diversity	
DSS	developmental sentence score	
DATES	birthday / date / age	+b1-JAN-2000
MOR	morphological analysis	

PHONFREQ

2	A	initial =	0,	final =	1,	other =	1
1	I	initial =	0,	final =	1,	other =	0
3	a	initial =	1,	final =	1,	other =	1
2	m	initial =	2,	final =	0,	other =	0
3	n	initial =	1,	final =	1,	other =	1
2	o	initial =	0,	final =	2,	other =	0
2	w	initial =	2,	final =	0,	other =	0

CLANで解析

FREQ	frequency	+s“word” +s”word*”
KWAL	keyword and line	
COMBO	combination of keywords	+s“word”^“word”
WDLEN	word length	in letters, morphemes & words
PHONFREQ	phoneme frequency	
MLU	mean length of utterance	
VOC	vocabulary diversity	
DSS	developmental sentence score	
DATES	birthday / date / age	+b1-JAN-2000
MOR	morphological analysis	

CLAN Output

> mlu -t%mor @

mlu -t%mor @

Tue Sep 19 18:10:22 2006

mlu (09-Aug-2006) is conducting analyses on:

ALL speaker tiers

発話者

From file <Macintosh HD:User:smi.SU.Wak1:bri:EucPosttrain:Miyata-Tai-jmor03new:hokkaido.cha>

MLU for Speaker: *STU:

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

Number of: utterances = 3, morphemes = 12

Ratio of morphemes over utterances = 4.000

Standard deviation = 0.000

単語数

MLU for Speaker: *TEA:

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

Number of: utterances = 1, morphemes = 7

Ratio of morphemes over utterances = 7.000

発話数

090806[E][TEXT] * 1

CLANで解析

FREQ	frequency	+s“word” +s”word*”
KWAL	keyword and line	
COMBO	combination of keywords	+s“word”^“word”
WDLEN	word length	in letters, morphemes & words
PHONFREQ	phoneme frequency	
MLU	mean length of utterance	& type/token ratio
VOC	vocabulary diversity	
DSS	developmental sentence score	
DATES	birthday / date / age	+b1-JAN-2000
MOR	morphological analysis	

ランダムな100単語を 50回 解析

VOCD

47	100	0.7694	0.055	60.311
48	100	0.7642	0.058	59.427
49	100	0.7469	0.051	54.015
50	100	0.7564	0.052	58.717

D: average = 59.309; std dev. = 2.399

D_optimum <59.15; min least sq val = 0.001>

VOCD RESULTS SUMMARY

=====

Types, Tokens, TTR: <385, 2019, 0.190688>

D_optimum values: <58.63, 58.71, 59.15>

D_optimum average: 58.83

CLANで解析

FREQ	frequency	+s“word” +s”word*”
KWAL	keyword and line	
COMBO	combination of keywords	+s“word”^“word”
WDLEN	word length	in letters, morphemes & words
PHONFREQ	phoneme frequency	
MLU	mean length of utterance	& type/token ratio
VOC	vocabulary diversity	
DSS	developmental sentence score	
DATES	birthday / date / age	+b1-JAN-2000
MOR	morphological analysis	

文法項目を点数化 (1 ~ 8 点)

DSS

"I like them."

I = Indefinite Pronouns Stage 1 -> 1 point
like = Main Verb Stage 1 -> 1 point
them = Indefinite Pronouns Stage 3 -> 3 point
Sentence point = adult-like sentence -> 1 point

=====

6 points

total score for 50 sentences:

366 points

average score (366/50)

7.32 points

Macintosh HD:Desktop Folder:projects\INDEX K.G.Hokokusho2001:0tomomiyaClinicalsample2.mor...

dss +bNIC

dss (21-Feb-2001) is conducting analyses on:

ONLY speaker main tiers matching: *NIC;

From file <Macintosh HD:Desktop Folder:sample2.mor.cex> to file <Macintosh HD:Desktop Fo>

Developmental Sentence Analysis

Sentence	IP	PP	PV	SV	NG	CNJ	IR	WHQ	S	TOT
I like this .	1	1	1						1	4
I like that .	1	1	1						1	4
I want hot dog .		1	1						0	2
I like it .	1	1	2	1					1	6

Developmental sentence score: 4.00.

CLAN [E][TEXT] 16

DSS

CLANで解析

FREQ	frequency	+s“word” +s”word*”
KWAL	keyword and line	
COMBO	combination of keywords	+s“word”^“word”
WDLEN	word length	in letters, morphemes & words
PHONFREQ	phoneme frequency	
MLU	mean length of utterance	& type/token ratio
VOCB	vocabulary diversity	
DSS	developmental sentence score	
DATES	birthday / date / age	+b1-JAN-2000
MOR	morphological analysis	

DATES +b1-JAN-2000 +d7-OCT-2006

```
> dates +b1-JAN-2000 +d7-OCT-2006
```

```
@Birth of CHILD: 1-JAN-2000
```

```
@Date: 7-OCT-2006
```

```
@Age of CHILD: 6;9.6
```

CLANで解析

FREQ	frequency	+s“word” +s”word*”
KWAL	keyword and line	
COMBO	combination of keywords	+s“word”^“word”
WDLEN	word length	in letters, morphemes & words
PHONFREQ	phoneme frequency	
MLU	mean length of utterance	& type/token ratio
VOC	vocabulary diversity	
DSS	developmental sentence score	
DATES	birthday / date / age	+b1-JAN-2000
MOR	morphological analysis	

など

などなどなど

オプションの概要は...

kwal

KWAL- Key Word And Line concordance program searches data for key words and produces the keywords in the context.

Usage: kwal [a n dN fS k oS pF rN re sS tS u wN xN yN zN 1 2] filename(s)

+a : display output in alphabetical order of keywords

+nS: include all utterances from Speaker S when they follow a match for +s

-nS: exclude all utterances from Speaker S when they follow a match for +s

+d : outputs legal CHAT format

+d1: outputs legal CHAT format plus file names and line numbers

+d2: outputs file names once per file only

+d3: outputs ONLY matched items

+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

+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 </?>,

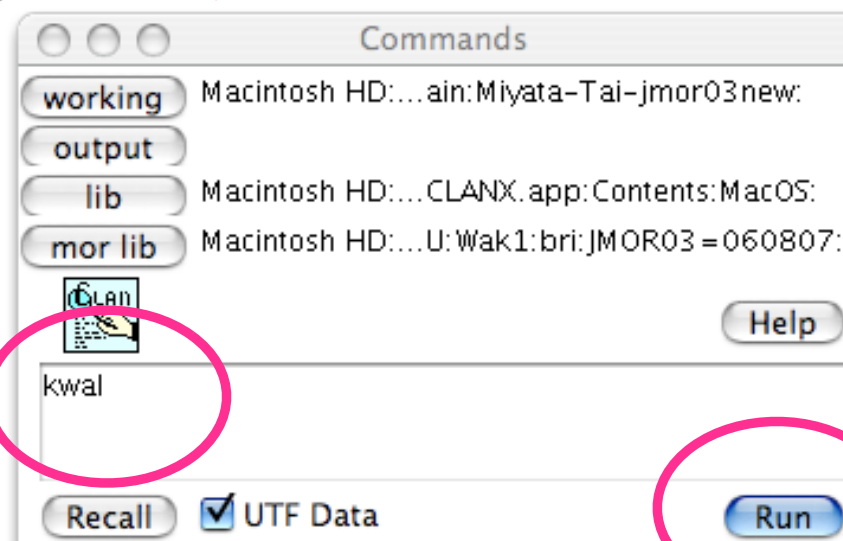
7- do not remove '/' and '-' characters

+re: run program recursively on all sub-directories.

+sS: search for word S or words in file @S in an input file (+s@ for more info).

090806[E][TEXT] * 1

コマンドだけ
打ち込む



CLAN Output

kwat

KWAL- Key Word And Line concordance program searches data for key words and produces the keywords in the context.

Usage: kwat [-a -dN -fS -k pS -pF -rN -re -sS -tS -u wN -xN -yN -zN 1 2] filename(s)

-a: display output in alphabetical order of keywords

+nS: include all utterances from Speaker S when they follow a match for +s

-nS: exclude all utterances from Speaker S when they follow a match for +s

+d: outputs legal CHAT format

+d1: outputs legal CHAT format plus file names and line numbers

+d2: outputs file names once per file only

+d3: outputs ONLY matched items

+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

+oS: include additional tier code S for output (uppercase ONLY)

-oS: exclude tier code S from an additional inclusion in an output

+pF: define punctuation set according to file F

+iN: 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).

090806[E][TEXT] * 1

コマンドの概要

オプションの使い方

Commands

working

Macintosh HD:...ain:Miyata-Tai-jmor03 new:

output

lib

Macintosh HD:...CLANX.app:Contents:MacOS:

mor lib

Macintosh HD:...U:Wak1:bri:JMOR03 = 060807:



Help

kwat

Recall

☒ UTF Data

Run

1. CHILDESの簡単概要
2. CLANのセットアップ
3. CHAT フォーマットの基礎
4. 音声・画像を加える
5. CLAN プログラムの基礎
6. MORで形態素解析

品詞 | 語幹-語尾

nlpen-PL

vlSpeak-3S

detlthis

pro:demlthis

日本語のMOR

ptl:case|ga

助詞:格|語

v:c|yar-PAST

動詞:子音|語幹-語尾

n|hannin

名詞|語幹

英語のMOR

pro|she

v|want-PAST

n|cookie-PL

MOR (形態素解析プログラム)

*CHI: she wanted cookies.

%mor: pro|she v|want-PAST n|cookie-PL.

*CHI: Makichan ga yatta .

%mor: n:prop|Maki-chan ptl:case|ga
v:c|yar-PAST .

CHILDES Child Language Data Exchange System



The CHILDES system provides tools for studying conversational interactions. These tools include a database of transcripts, programs for computer analysis of transcripts, methods for linguistic coding, and systems for linking transcripts to digitized audio and video.

System

[Ground rules](#)

[Membership list](#)

How to subscribe to [Mailing Lists](#)

[Overviews](#) (Español, Français, 中文)

Links

[IASCL](#) information

[Other Child Language sites](#)

The [TalkBank](#) Project

Phonology and Fonts

The [Phon](#) program and database

[Slicing and splicing](#) audio files with Praat

[Unicode fonts](#)

IPA for [Mac](#) and [Windows](#)

Special Procedures

Procedures and tools for [CA analysis](#)

Working with [digitized video](#)

Working with [digitized audio](#)

Programs and Database

The [CLAN](#) Program

[Database](#)

[TalkBank Viewer](#)

[Contributing](#) new data

Manuals (in PDF)

[CHAT Transcription system](#)

[CLAN Programs](#)

[Database Manuals](#)

[BTS sign transcription system](#)

Teaching with CHILDES

[Topics](#) in language acquisition.

Teaching [Tips](#) and [Resources](#).

[Child Language Bibliographies](#)

Morphology and Lexicon

[Part of Speech Analysis by MOR](#)

Parental frequency count: [view](#),
[download](#), [documentation](#)

MRC lexical dictionary

プログラム

データベース

マニュアル

MORファイル



The MOR program provides a method for automatic tagging of corpora in the CHAT format. To make his 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.

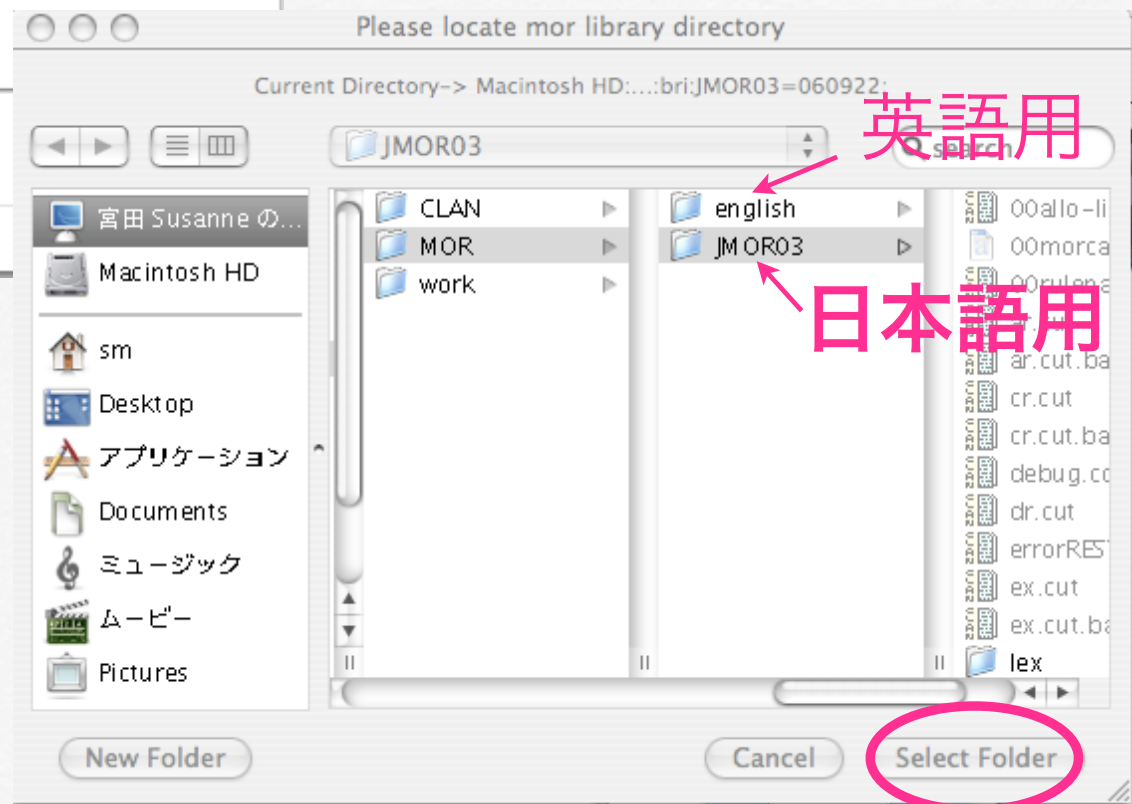
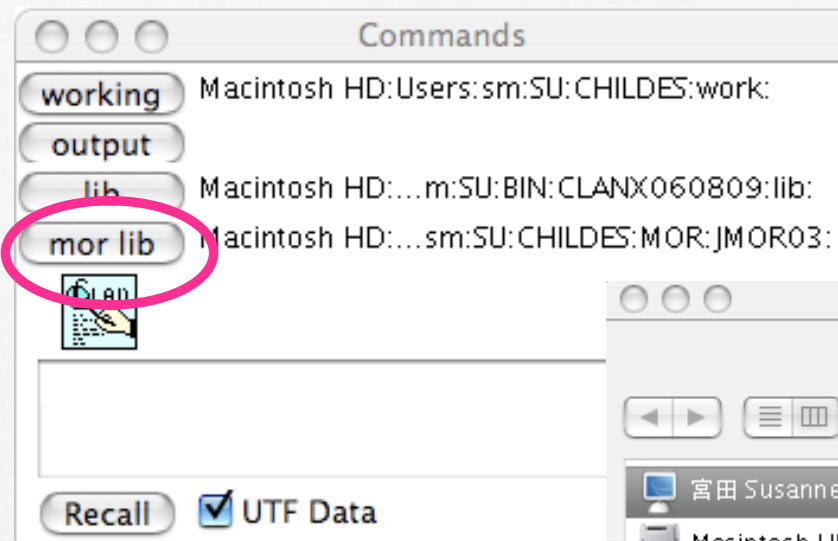
We have good working MOR grammars for these languages:

- [Cantonese](#): This grammar was contributed by Zehava Weizman, Emily Ma, Paul Fletcher, Samuel Leung, and Stephanie Stokes at the University of Hong Kong. Because disambiguation is built in to this system, there is no need to use POST.
- [Chinese](#): Our thanks to the K. J. Chen and the CKIP Group of the Academia Sinica for providing a listing of the 20,000 highest frequency forms of Puotonghua along with their English translations and romanizations. The MOR reformatting of their grammar is provided for both Simplified and Traditional character sets.
- [Dutch](#): This grammar was contributed by Steven Gillis.
- [English](#): This grammar was built initially by Mitzi Morris and refined by Brian MacWhinney. It covers all the forms in the CHILDES English database. The zip file also includes a POST eng.db database trained on the maternal utterances in the Eve corpus.
- [French](#): This grammar was contributed by Christophe Parisse, using the minMOR framework. It works well, but the coverage is still incomplete.
- [German](#): This grammar was contributed by the MPI-Leipzig.
- [Japanese](#): This grammar was constructed by Norio Naka and refined by Suzanne Miyata. The [Wakachi](#) system is helpful for reference.
- [Italian](#): This grammar was built by Brian MacWhinney.
- [Spanish](#): This grammar was built by Brian MacWhinney with input from Monica Sanz Torrent.

英語用

日本語用

形態素分析：%mor



英語用

日本語用

```
@Begin
@Languages:      en, ja
@Participants:   MOT Mother, CHI Tatoechan Child
@ID:             ja|sample|CHI|2;7.21|female||Child|
1 *CHI:          kore choodai .
%mor:            n:deic:dem|kore v:ir:sub|choodai&IMP^v:ir|choodai&IMP .
%ort:            これちょうだい。
2 *MOT:          nani ga hoshii ?
%mor:            n:deic:wh|nani
                 ptl:conj|ga^ptl:case|ga^onoma|ga^n|ga adj|hoshi-PRES ?
%ort:            何が欲しい？
3 *CHI:          hayaku.
%mor:            adj|haya-ADV.
%ort:            はやく。
4 *CHI:          e o kakitai kara .
%mor:            n|e=picture^ptl:case|e ptl:case|o
                 v:c|kak-DESID-PRES=scratch^v:c|kak-DESID-PRES=write
                 ptl:conj|kara=after^ptl:conj|kara=because^ptl:case|kara^onoma|kara^n|kar
%ort:            絵を描きたいから。
@End
```

```
000006[C][CHA7] 6
v:ir:sub|choodai&IMP
v:ir|choodai&IMP
?|choodai
```

自動解析を選択：esc-2

Macintosh HD:Users:sm:SU:CHILDES:work:hokkaido(.mor.pst.cex)

@Begin

@Languages: en, ja

@Participants: MOT Mother, CHI Tatoechan Child

@ID: ja|sample|CHI|2;7.21|female||Child||

*CHI: kore choodai .

%mor: n:deic:dem|kore v:ir|choodai&IMP .

%ort: これちょうだい。

*MOT: nani ga hoshii ?

%mor: n:deic:wh|nani ptl:case|ga adj|hoshi-PRES ?

%ort: 何が欲しい？

*CHI: hayaku.

%mor: adj|haya-ADV.

%ort: はやく。

*CHI: e o kakitai kara .

%mor: n|e=picture ptl:case|o v:c|kak-DESID-PRES=write

ptl:conj|kara=because .

%ort: 絵を描きたいから。

@End

FREQuency 頻度計算・単語一覧

+t 形態素で検索

(英語の場合)

FREQ +t%mor +s"*-3S"

第3人称の-sを探す

FREQ +t%mor +s"*-PL"

KWAL +t%mor +s"*-PL"

COMBO +t%mor +s"nl"*^"vl"*)

...

FREQuency 頻度計算・単語一覧

+t 形態素で検索

(英語の場合)

FREQ +t%mor +s"*-3S"

第3人称の-sを探す

FREQ +t%mor +s"*-PL"

複数形の-sを探す

KWAL +t%mor +s"*-PL"

COMBO +t%mor +s"*nl"*^"vl"*"

...

+t 形態素で検索

(英語の場合)

FREQ +t%mor +s"*-3S"

第3人称の-sを探す

FREQ +t%mor +s"*-PL"

複数形の-sを探す

KWAL +t%mor +s"*-PL"

複数形を含む発話の一覧

+t 形態素で検索

(英語の場合)

FREQ +t%mor +s"*-3S"

三人称の -s を探す

FREQ +t%mor +s"*-PL"

複数形の -s を探す

KWAL +t%mor +s"*-PL"

COMBO +t%mor +s"n|*"^"v|*"

...

名詞 + 動詞の組み合わせ

CLANで解析（日本語の場合）

FREQ +t%mor

+s“**ptl:case**l*”

格助詞の一覧表

CLANで解析

FREQ +t%mor

+s“ptl:case|*”

格助詞の一覧表

+s“*-PAST”

動詞&形容詞の過去形

CLANで解析

FREQ +t%mor

+s“ptl:case|*”

格助詞の一覧表

+s“*-PAST”

動詞&形容詞の過去形

KWAL +t%mor

+s “**V***”

動詞を含む発話の一覧表

CLANで解析

FREQ +t%mor +s“ptl:case|*”

格助詞の一覧表

+s“*-PAST”

動詞&形容詞の過去形

KWAL +t%mor +s “v*”

動詞を含む発話の一覧表

COMBO +t%mor +s “ptl:case|^v*”

格助詞と動詞を両方含む発話

CLANで解析

FREQ +t%mor	+s“ptl:case1*”	格助詞の一覧表
	+s“*-PAST”	動詞&形容詞の過去形
KWAL +t%mor	+s “v*”	動詞を含む発話の一覧表
COMBO +t%mor	+s “ptl:case1*^v*”	格助詞と動詞を両方含む発話

など

などなどなど

などなどなど

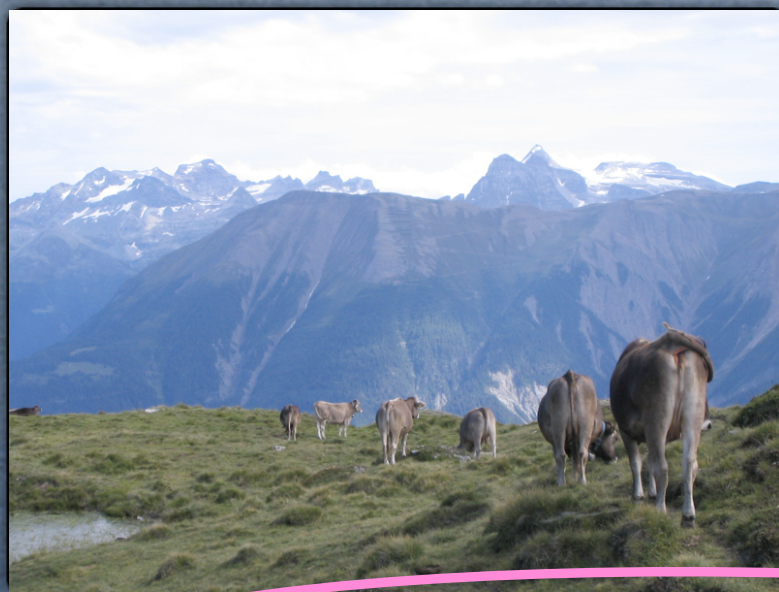
などなどなどなどなどなど

などなどなどなどなどなどなどなど

などなどなどなどなどなどなどなどなどなど

1. CHILDESの簡単概要
2. CLANのセットアップ
3. CHAT フォーマットの基礎
4. 音声・画像を加える
5. CLAN プログラムの基礎
6. MORで形態素解析

たいへんお疲れさまでした



たいへんお疲れさまでした