

CHAT

ממשק התכתוק ב-CHILDES

סיגל עוזיאל-קרל
סמינר הקיבוצים ואוניברסיטת חיפה
Sigal@alum.mit.edu

<http://semtalk.talkbank.org>

1 **ישראל** (מתיישב במפגש כשהוא מחזיק שקית ביד.)

2 גננת שים את השקית שם בצד מתוק.

האם אני מרשה לשבת במפגש עם דברים ביד?

3 **ביחד** לא

4 גננת לא מרשה, נכון?

שירית עוד מעט תבוא.

5 ילד אני רוצה רגע לצבוע פה

6 גננת רגע, מתוקי.

7 ילד גם לי

8 גננת כל יום עד מסיבת פורים, עד ש...

סליחה מתוקי שב שם ממול, ישראל (התיישב עם השקית ביד במקום אחר)

אני אזכור לתת לך את זה אל תדאג, זה שלך ואתה תקח את זה היום

הביתה. (לוקחת ממנו את השקית)

כל יום עד שיגיע חג הפורים נאפר **אותכם**, ואתם תתחפשו כל יום.

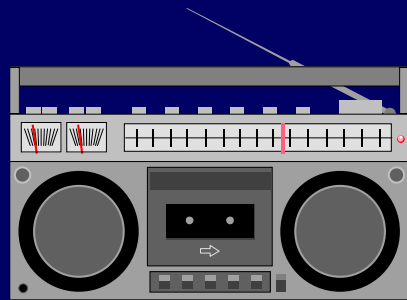
פעם תהיו ל - חתולים

ופעם אתם תהיו ל -

9 **נינג'ות** יאיר

10 גננת נינג'ות.

CHILD LANGUAGE DATA EXCHANGE SYSTEM



עורך הטקסטים
CED

כלי ניתוח הנתונים
CLAN

מנגבון קידוד
חצי-אוטומטי

מאגר קטגוריות
תעתוק
CHAT

כלי בקרה
CHECK

עורך הטקסטים

- העורך מאפשר לבנות קבצי תעתיק חדשים
- מקל על עריכתם של קבצים קיימים
- מאפשר בקרת איכות על התעתיקים באמצעות הרצת תוכנית הבדיקה (CHECK) בתוך חלון העריכה
- כמערכת קידוד – מספק למקודד דרך שיטתית לקידוד טקסטים מתוך תפריטי קטגוריות מוגדרים מראש

ארגון סביבת העבודה

סביבת העבודה

Clan - [sample.cha]

File Edit View Tiers Mode Window Help

New File Ctrl+D

commands

```
@Begin
@Filename:      lio105a
@Version:      orig+
@Coding:      CHILDES
@Age of LIO:   1;5.19
@Sex of LIO:   Female
@Date:        26-Jan-1990
@Situation:    At home
@Participants: LIO Lid
              TAL Tal
@Utterances:  LIO: 8:
              ADU: 200
@Cassette:     9a
@Comment:     Transcriber hears first two-word combination _ze savta_,
              but participants do not seem to notice it. The formulaic
              combination _od pa'am_ is also used.
              Participants tend to pronounce _et ha_ as _ta_
              _savta_ is always pronounced _safta_

*MOT:  Lior, boi kxi et ha^matate.
%lex:  $N $V:bwa1 $V:lqx1 $ACC $Det $N
%mor:  $N:S:FM $V:2S:FEM:IMP $V:2S:FEM:IMP $ACC $Det $N:S:MASC
*TAL:  ma ze, ma ze?
*TAL:  matate.
*MOT:  sapri lahem ma axalt, axalt avukado?
%lex:  $V:spr3 $PN&PR $WH $V:akl1 $V:akl1 $N
*LIO:  kado [: avukado] [*].
*MOT:  ve axalt gam yogurt?
%lex:  $CONJ $V:akl1 $FN $N
@End
```

Commands

working	C:\Documents and Settings\Sigal\My
output	
lib	C:\Documents and Settings\Sigal\My
mor lib	C:\Documents and Settings\Sigal\My

CLAN FILE IN Help

```
mlu -f sample.cha
modrep
modrep -f +*MOT +b%lex +c%mor sample.cha
```

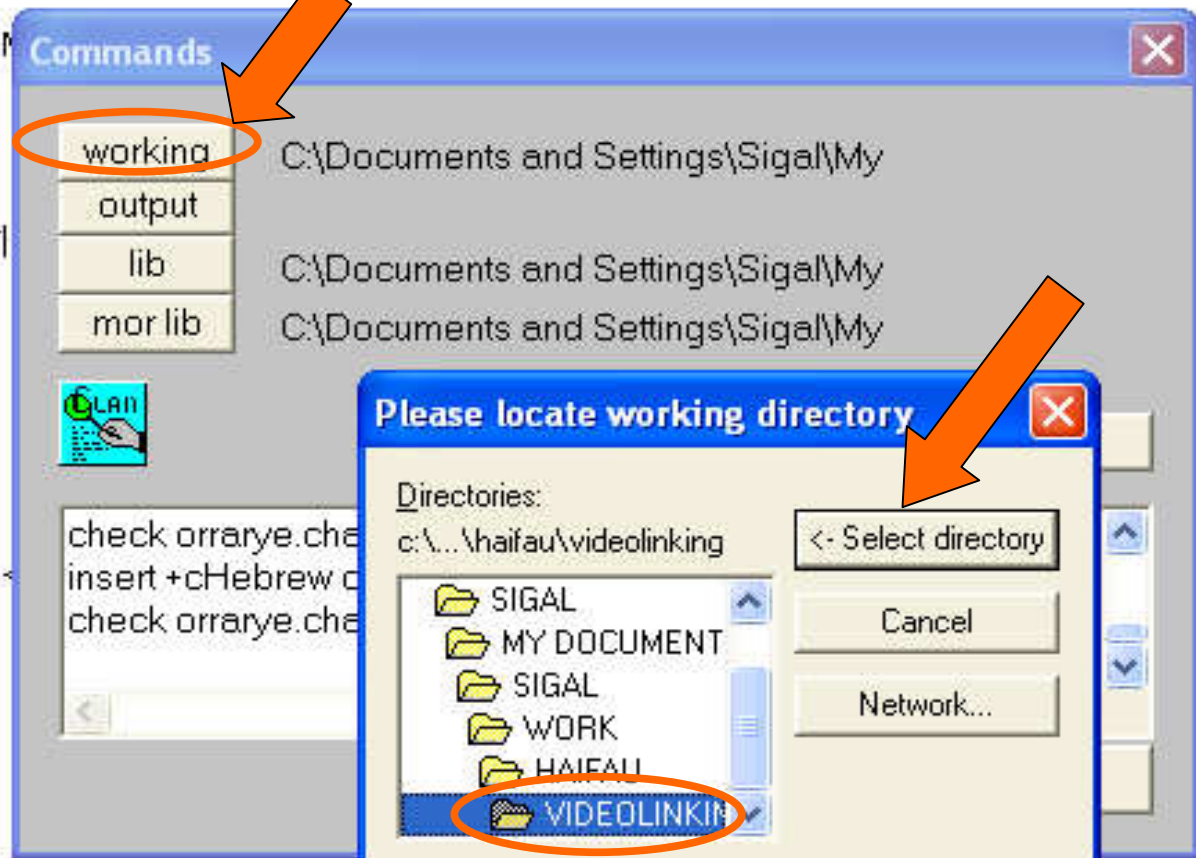
Run

0412[E][CHAT] 23 : File "C:\Documents and Settings\Sigal\My Documents\Sigal\RESEARCH\BERMAN\CODING\LIB\sample.cha" written.

סביבת העבודה

```
@Begin
@Languages: he, lecgacy
@Participants: ORR Orr Child, LIL Lilach M
              Cousin
@ID: he,lecgacy|Hebrew|ORR||||Child|
@ID: he,lecgacy|Hebrew|LIL||||Mother|
@ID: he,lecgacy|Hebrew|OMR||||Brother|
@ID: he,lecgacy|Hebrew|NIR||||Cousin|

*OMR: tagid <arye>["].
*OMR: tagid <arye>["], nu.
*LIL: tagid <arye>["].
*OMR: aye [: aye][*].
*OMR: aye [: aye][*], hu mer [: omer][*]
*LIL: pagashti aye mehudar u malxuti.
*LIL: eyfo ha^arye?
*LIL: tar'e le^ima et ha^arye.
*ORR: aye [: aye][*].
*LIL: aye.
*LIL: aye sho'eg, naxon.
*NIR: ani yaxol lish'og yotem mi^namer.
*LIL: dov, tavi dov.
*LIL: Orr, tavi dov.
*OMR: gam ani xxx.
@End
```



שמו של קובץ התעתיק

- שמו של קובץ התעתיק ייסתיים תמיד בסיומת – .cha.
- שם הקובץ יכול להכיל אותיות ומספרים **ללא** סימנים נוספים (למשל, מקף)
- רצוי ששם הקובץ לא יהיה ארוך משמונה תווים
- שם הקובץ ישקף את תוכן התעתיק

Lio105a.cha

שם הילד

גיל הילד

מספר התעתיק

סיומת הקובץ

מבנה קובץ CHAT

עורך הטקסטים

@Begin

@Languages: he, ar, en

@Participants:.....

@ID:.....

[other headers]

*HAG: [spoken material]

%xxx: [coding of spoken material]

*FAT: [spoken material]

%xxx: [coding of spoken material]

@End

@Begin

@Languages: ar
@Participants: SMD Smadar Child, MOT Ziva Mother
@ID:
@Filename: smd106a.cha
@Age of SMD: 1;6.2
@Sex of SMD: female
@Date: 4-OCT-1988
@Situation: Looking at pictures, then playing with cubes
@Utterances: SMD: 263
@Comment: many two-word combinations.

קובץ התעתיק

כותרות

@End

עורך הטקסטים כותרות חובה

כותרות קבועות: מסומנות על ידי @

@Begin

@Languages: he, ar

@Participants: HAG Hagar Child, FAT Mushi
Father

@ID: language|corpus|code|age|sex|
group|SES|role|education|

.

.

@End

עורך הטקסטים כותרות חובה

כללי רישום המשתתפים:

@Participants: **Tab**HAG Hagar Child, FAT Mushi
Father

- הכותרת באנגלית כשלפניה @ ולאחריה נקודותיים ו-TAB, **לא רווח!**
- קוד המשתתף בשלוש אותיות גדולות
- רווח
- שם המשתתף באות ראשונה גדולה
- תפקיד המשתתף באות ראשונה גדולה (מתוך רשימת התפקידים הקיימת)
- בין כל שני משתתפים פסיק ורווח

תפקידי המשתתפים באינטראקציה

@Participants: Target_child, Child
Mother, Father, Brother, Sister, Sibling
Grandmother, Grandfather, Aunt, Uncle,
Cousin
Family_Friend, Playmate
Visitor Student Teacher
Babysitter, Housekeeper
Investigator, Examiner, Observer
Clinician, Therapist
Camera_Operator
Doctor, Nurse, Patient, Client
Subject, Unidentified
Adult Teenager Boy Girl
Non_Human, Toy, OffScript Text Narrator

"הכה את המומחה"

דוברים שונים בעלי שם דומה נסמן:

• באותו קוד עם מספרים עוקבים:

CH1, CH2, CH3 ...

• בקודים שונים:

אביב AVV, אבי AVI

• דובר כלשהו יסומן באותו קוד בקבצים שונים.

• בתעתיק יתחיל כל שם פרטי באות גדולה.

• בכל קובץ יאויית שמו של דובר מסוים באותו

אופן:

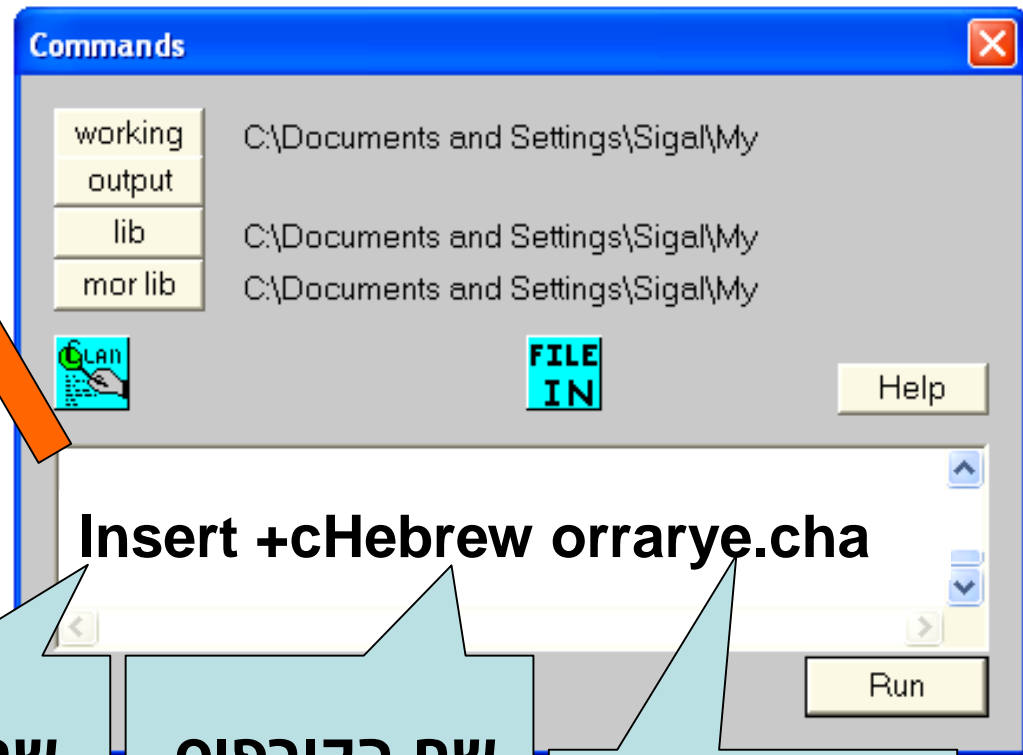


Muhamed, ~~Muhammad~~, ~~Mhamed~~

```
@Begin
@Languages: he, lecgacy
@Participants: ORR Orr Child, LIL Lilach Mother, OMR Omer Brother, NIR Nir
Cousin
```

```
@ID: he,lecgacy|Hebrew|ORR||||Child|
@ID: he,lecgacy|Hebrew|LIL||||Mother|
@ID: he,lecgacy|Hebrew|OMR||||Brother|
@ID: he,lecgacy|Hebrew|NIR||||Cousin|
```

```
*OMR: tagid <arye>["].
*OMR: tagid <arye>["], nu.
*LIL: tagid <arye>["].
*OMR: aye [: aye][*].
*OMR: aye [: aye][*], hu mer [: omer][*] <arye>["] [: aye][*].
*LIL: pagashti aye mehudar u malxuti.
*LIL: eyfo ha^arye?
*LIL: ta'e le^ima et ha^arye.
*ORR: aye [: aye][*].
*LIL: aye.
*LIL: aye sho'eg, naxon.
*NIR: ani yaxol lish'og yotem mi^namer.
*LIL: dov, tavi dov.
*LIL: Orr, tavi dov.
*OMR: gam ani xxx.
@End
```



שם הפקודה

שם הקורפוס

שם הקובץ שבו
יתבצע השינוי

עורך הטקסטים כותרות רשות

@ כותרות קבועות: מסומנות על ידי @

@Activities: activities involved in the situation

@Coder:

@Comment: any kind of general comment

@Font:

@Date: 01-JUL-1965

@New Episode

@Room Layout: room configuration and
positioning of furniture

עורך הטקסטים שורת התעתיק

שם הדובר בשורת התעתיק מסומן על ידי *
שלוש אותיות גדולות, נקודתיים, ולאחריהן **TAB**

קוד נפרד לכל נבדק:

- | | |
|-------|------------|
| *HAG: | – ילד/ה |
| *FAT: | – הורה |
| *INV: | – חוקר/ת |
| *CLI: | – קלינאי/ת |

CLAN - [sample.cha]

File Edit View Tiers Mode Window Help

*CHI: Ctrl+1
Ctrl+2
Ctrl+3
Ctrl+4
Ctrl+5
Ctrl+6
Ctrl+7
Ctrl+8
Ctrl+9

Update

*CHI: @Begin
@Languages: h
@Participants: Ch
*CHI: Hagar holex
*CHI: Dan halax.
*CHI: Smadar lok
*CHI: kaftor.
*CHI: masmerim.
*CHI: Dan shar shirim be-kol ram ba-yam.
*CHI: Dan ve Tamar korim be kolot ramim ba yamim.
*CHI: Dan golesh ba yamim.
*CHI: kshe-halax Dan.
*CHI: kshe-holexet Hagar.
*CHI: Dan halax lagan.
*CHI: Dan halax la gan.
*CHI: Dan halax la-gan.
*CHI: Dan halax ba-gan.
*CHI: ganim.
CHI: nanu [: gamarnu] [].
CHI: eynanu [: gamarnu] [].
*CHI: salim.
*CHI: gan.
*CHI: hagar.

working C:\CHILDES\lib\heb\heb
output C:\CHILDES\lib\heb\heb
lib C:\CHILDES\lib\heb\heb
mor lib C:\CHILDES\lib\heb\heb

CLAN FILE IN

סימני פיסוק

- המילה הראשונה במבע מופרדת משם הדובר באמצעות TAB
- מילים בתוך מבע מופרדות זו מזו באמצעות רווח
- סימן הפיסוק המותר בתוך מבע הוא פסיק (,)
- כל מבע חייב להסתיים בסימן פיסוק כלשהו
- סימני הפיסוק המותרים בסוף מבע הם:
נקודה (.), סימן שאלה (?), סימן קריאה (!)

```
*ATA: mi ze ha^saya'at?  
*CHI: ha^akavish.  
*ATA: mi ze Mira?  
*ATA: ha^saya'at.  
*TCR: naxon.  
%elc: $EVL  
@Bg: reading  
*TCR: boker exad he'ifa ha^ruax et ha^akavish me^al ha^sade.  
*TCR: xut meshi dak nimtax mi^gufo, ha^akavish naxat al gader leyad  
ha^xava ve hitxil litvot et xut ha^meshi shelo she korim [: kor'im]  
[*] lo +...  
%elc:  
$PRM:COM  
*@Bg: reading  
*TCR: kur.  
%elc: $REV:REF:LIN  
*ATA: ma ze tur?  
*ATA: ma ze tur?  
%add: TCR, ASS.  
*TCR: saymu le'exol at pashut lokaxat oto, ki xaval, ha^yladim [: yeladim]  
[*] ha^axerim loh makshivim la^sipur.
```

בוקר אחד העיפה הרוח את העכביש
מעל השדה.
חוט משי דק נמתח מגופו...

קובץ התעתיק

הערות

```
*SMD: teyp.  
*MOT: tep [: teyp] [*]?  
*MOT: ken.  
*MOT: nasim oto po?  
%act: placing tape-recorder.  
%sit: a few seconds of silence.  
*MOT: ma?  
*SMD: hine teyp.  
%lex: $H $N  
%mor: IRV $N:S:MASC  
%src: $SF  
*MOT: hine teyp?  
*SMD: hine.  
*MOT: ken.  
*MOT: lexi tivni ba^migdal.  
*SMD: e.
```

עורך הטקסטים שורת ההערות

הערה תסומן בקובץ התעתיק ב- % ושלוש אותיות המציינות את סוג ההערה, שלאחריהן : ו- **TAB**

%sit: Israel is sitting in the circle, holding a plastic bag in his hand.

- שלוש האותיות בשם ההערה צריכות להיות **קטנות**
- אין להתחיל את שורת הטקסט בהערה. הערה חייבת להופיע לאחר שורת תעתיק.
- בתוך הערה ניתן להשתמש בפסיקים, ובסופה בנקודה.

עורך הטקסטים

שורת ההערות

קוד	סוג ההערה	הסבר	דוגמא
%act	action	פעולות של השומע או הדובר	*CHI: I do it! %act: runs to toy box
%add	addressee	מי הדובר ואל מי הוא פונה	*MOT: be quiet. %add: RON, LIO
%alt	alternate transcription tier	תעתיק אלטרנטיבי	

עורך הטקסטים

שורת ההערות

קוד	סוג ההערה	הסבר	דוגמא
%com	comment	שורת הערות כללית	*EVE: that's nasty # is it? %com: note tag question
%eng	English	תרגום לאנגלית	
%exp	explanation	להבהרה במקרה של דאיקסיס	*CHI: I want it! %exp: it refers to a book
%sit	situation	תאור סיטואציה מסוימת	

כתיבת הערות

הערות ניתן לרשום בתעתיק במספר אופנים בהתאם לכלליות ההערה.

הערה שנוגעת לתעתיק בכללותו –

@comment:

הערה שנוגעת לחלק מסוים באינטראקציה –

%com:

הערה שנוגעת למילה מסוימת בתעתיק –

[= text]

כללי רישום מילים ושיח

סימון צורת מיוחדות

משמעות	דוגמה	קודים	קטגוריות
—	abame@b	@b	מלמול
פועל כלשהו	mangid@c	@c	מילה מומצאת
שם חיבה לניצן	nicanon@f	@f	צורה ייחודית למשפחה
התפרק	nifraq@n	@n	תחדיש
מכונית נוסעת	ananan@o	@o	אונומטופיאה
שיר	lalala@si	@si	מילות שיר
יצור קטן	wug@t	@t	פריט מבחן
טל-גל (חריזה)	talgal@wp	@wp	משחק מילים

כללי רישום מילים

• כאשר מילים בתמליל המתועתק אינן ברורות, הן תסומנה ברצף של שלושה Xים: XXX

*CHI: yesh li XXX sham.

• כאשר מילה בתמליל מופיעה באופן חלקי, אך ניתן לשחזר את המשכה בבירור, מתעתקים את ההמשך בתוך סוגריים עגולים:

*CHI: yashav(ti) kol ha^yom.

• לציון הפסקה בין הברות משתמשים ב # או ב ## לציון הפסקה ממושכת:

*MOT: ze hipo#potam?

כללי רישום מילים

• הטעמה מציינים באמצעות לוכסן לפני ההברה המוטעמת:

*CHI: Rina /raca, aval Dan ra/ca lalexet.

• הברה ארוכה נסמן בנקודתיים:

*MOT: baby wants bana:~nas?

• התייחסות מטה-לינגוויסטית למילה:

*CHI: ma ze <sushi>[“]?

• מצמד נסמן באמצעות + בין המילים שיוצרות אותו:

*CHI: yom+huledet

כללי רישום מילים

• מילה שהושמטה ניתן לסמן בשורת התעתיק באמצעות הספרה **0** ובצמידות לה המילה שהושמטה:

*CHI: **0**sefer nafal.

• שגיאה תסומן בתוך התעתיק ב- **[* text]**

*CHI: Aba yoshevet **[* yoshev]**

• החלפת מילה בשורת התעתיק תסומן ב- **[: text][*]**

*CHI: **ofo** **[: eyfo][*]** aba?

• הסבר בשורת התעתיק יסומן ב- **[= text]**

*CHI: hine Gonzo **[= name of a puppet]**

דוגמאות לכללי רישום שיח

• מידע אשר לא יודעים כיצד לתעתק או שלא רוצים לתעתק מסמנים ברצף של שלושה wים (www) ולאחריו שורת ההערות %exp:

*MOT: www.

%exp: mother is speaking on the phone.

• הפסקה עצמית בשטף הדיבור תסומן ב- +...

*TCR: pa'am tihyu le +...

• הפרעה עצמית לשטף הדיבור תסומן ב- +//.

*TCR : Dan roce lihyot +//.

*TCR : Rega, lama ata lo' makshiv?

• הפרעה לרצף הדיבור על ידי דובר אחר יסומן ב- +//.

*MOT: what did you +//.

*CHI: Mommy.

דוגמאות לכללי רישום שיח

• השלמה עצמית תסומן ב- +, :

*MOT: halaxti la +...

*MOT: +, la^gan.

• השלמה על ידי דובר אחר תסומן ב- ++ :

*TCR: pa'am tihyu le +...

*CHI: ++ ninjot.

• ציטוט בשור הבאה יסומן ב- +"/. :

• שורת הציטוט תסומן ב- +":

*TCR: ha^yeled sha'al +"/.

*TCR: +": ma yesh ba^sakit?

דוגמאות לכללי רישום שיח

• חפיפה בדיבור עם דובר אחר תסומן כך:

*MOT: no # Sarah # you have to <stop doing that> [>] !

*SAR: <Mommy I don't like this> [<].

*SAR: it is nasty.

[>] – חפיפה בשורה הבאה

[<] – חפיפה בשורה הקודמת

!10