Analyses of Child-Directed Speech

Sonja Eisenbeiss, University of Essex

Properties of Child-Directed Speech (O'Grady 1997)

- slow, with longer pauses between utterances and after words
- high and varied pitch
- exaggerated intonation and stress patterns
- fewer disfluencies than in adult-adult speech (1 vs. 4.5 disfluencies per 1000 words)
- restricted vocabulary
- reference mostly restricted to here and now
- short, typically not syntactically complex, but mostly correct and complete utterances
- more imperative and questions than in adult-adult speech
- more repetitions than in adult-adult speech (~16% of utterances repeated within 3 turns)

For a general introduction, see Gallaway and Richards (1994).

The Child Language Data Exchange System

See MacWhinney, B. (2000) and http://childes.psy.cmu.edu/

The Format of Transcripts

- transcript data from monolingual and bilingual children with and without impairments
- a broad range of languages, but a strong focus on English data
- longitudinal and crossesctional data partially with corresponding media-files
- naturalistic samples and elicited data (e.g. picture book descriptions)
- encoding of linguistic information for some files

Format of a CHAT transcript

- @BEGIN
- @Participants
- [other headers]
- *JOE: [spoken material]
- %mor: [morpho-syntactic coding]
- *INT: [spoken material]
- %mor: [morpho-syntactic coding]
- @End

Tiers

- three letters followed by a colon and a tab
- obligatory: @ Participants, on the second line of the transcript; e.g.:
  @Participants: JOE Joe child, INT Interviewer
- optional; e.g.:
  - @Birth of Learner: …
  - @Age of Learner: …
  - @Date: …

Main Tier

- what was actually said, one utterance per tier
- introduced by "*", the three-letter code for the participant and a tab; e.g.:
  *JOE: the boy put the leash on the cat.
- orthographical transcription in lower case Latin letters; except for proper nouns (e.g. *John*) and "I"
- numbers spelled out (*ten*, not *10*)
- normalisation of phonetically deviant forms (phonetic information about forms can be presented on a %pho dependent tier)

Dependent Tiers with further annotations

%mor [morphosyntactic coding]

%pho [phonological coding]

%syn [syntactic coding]

%err   [errors]

%com [comments]

%spa   [speech acts]

Main Tier Markers

- unfilled pauses: #
- filled pauses: eh@fp
- interruption: +/.
- self-interruption: +//.
- repetition w/o correction: [/]
- repetition with correction: [//]
- unintelligible speech: xxx
- material coded on phonol. tier: yyy
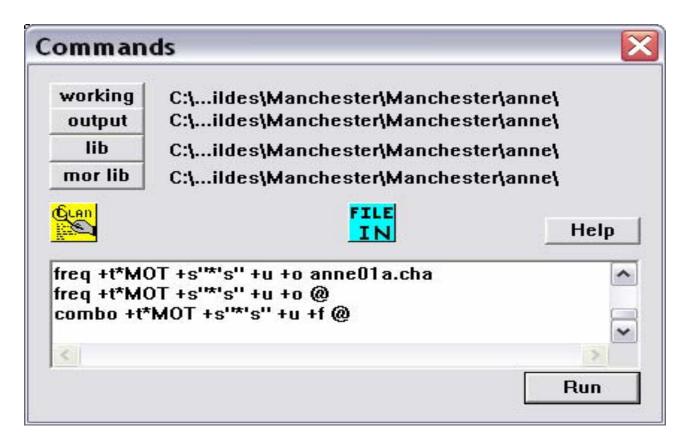- doubtful material: [?] or [=? text]
- omitted parts of words: ( )
- to refer to more than one word: < >

Using CLAN to Analyse CHAT Transcripts

See the CHILDES webpage for a CLAN manual: http://childes.psy.cmu.edu/

CLAN Windows

- the commands window where you specify the folders, files, and commands you want to use
- the CLAN output window, where you will see the results of your searches. If you have not specified an output file, your results will be displayed in this window. If you have saved your outputs into a file (as you will be asked to do for this exercise), you will not be able to see it in the output window, but the name and location of the output file will be displayed in the output window.

CLAN Steps

- specify your WORKING DIRECTORY, where the files you will be working with are stored
- specify your output directory, where any output files will be stored
- select a command (type of select from CLAN)
- select one or more transcription files for analysis (type name or select from FILE IN)
- optionally use some so-called switches to modify the commands.

Core CLAN Commands for Analyses of Children's Input
- FREQ: will provide you with type and token frequency information
- COMBO: will find utterances matching a given set of criteria

Useful Switches

+f   saves output to file. For each transcription that you have chosen to analyse, an output file will be generated. By default, this output file will have the name of the transcription file and an extension that will show you which command was used to create the output (e.g. frq, mlu or cmb). NOTE: if you use this option, the output will not appear on the screen. You will only see the folder and file name for the resulting file.

+s   searches for a string in a file.

+t   restricts the search to a particular tier – e.g. the tier of a particular speaker.

+u   treats all files together.

+o   orders FREQ lists according to token frequency

+w   –w1 and +w1 provide one preceding/following   line, -w2 and +w2 will provide two preceding/following lines, etc.

Search Strings

^   immediately followed by

+   inclusive OR

!   logical NOT

*   "joker" or "wildcard"

""   strings including blanks, etc. should be put in quotes

Some Examples

freq sarah009.cha
(frequency list of all words for the given transcript file)

freq sarah*.cha
(frequency list of all words for all Sarah transcript files in the working directory)

freq *.cha
(frequency list of all words for all transcript files in the working directory)

NOTE: you do not have to type the file names, you can also select them using FILE IN. However, if all your files have a similar name, typing can be faster, due to the use of the "joker".

combo +sthank* +t*CHI sarah134.cha
(all child utterances with "thank")

combo +s"*'s*" +t*MOT sarah134.cha
(all maternal utterances with "'s": John**'s** hat/here)

combo +s"aux|**" +t*CHI +t%MOR sarah134.cha
combo +s"aux|**" +t%MOR sarah134.cha
(all auxiliaries (produced by child)), auxiliaries coded on the %mor tier for morpho-syntactic codes)

Frequency Analyses

Many studies on children's input investigate the role of frequency in children's input, distsinguishing between:

- **Type Frequency:** How many different types of this element do occur in the corpus (e.g.: How many different verbs does the corpus contain?)
- **Token Frequency:** How many individual instances of this element do occur in the corpus (e.g.: How many individual occurrences of these verbs can we find in the corpus ?)
- **Word form:** An individual form of a lexical item (e.g. *gives*)
- **Lemma:** All the word forms of a lexical element taken together (e.g. GIVE: *give, gave, gives, given*,…)

Example Analysis

You find the following word forms in the child's input: *give, give, gives, gave, take, takes, takes, taken, car, cars, cars*

- Lemma:
  - 2 verb types (GIVE, TAKE) and 1 noun type (CAR)
  - 8 verb tokens (*give, give, gives, gave, take, takes, takes, taken*) and 3 noun tokens (*car, cars, cars*)
- Word Form:
  - 8 word form types:
    - 3 word form types of GIVE (*give, gives, gave*),
    - 3 word form types of TAKE (*take, takes, taken)*,
    - 2 word form types of CAR (*car, cars*)
  - 11 word form tokens:
    - 4 word form tokens of GIVE (2 x *give,* 1 x *gives,* 1 x *gave*),
    - 4 word form tokens of TAKE (1 x *take,* 2 x *takes,* 1 x *taken)*,
    - 3 word form tokens of CAR (1 x *car,* 2 x *cars*)

NOTE: Check your calculations if you end up with more types than tokens. The number of types should never be higher than the number of tokens. If type and token frequencies are the same, this means that each type occurred once.

Using CLAN

You can use the FREQ command to produce frequency counts. The switches allow you to target the output to your needs.

freq 2-02-19.cha
(This provides you with a list of words and their frequencies, for all speakers, organised **alphabetically**)

11 .
273 a
 25 a@l
 92 a@p
  4 aah
  1 able
 10 about
  2 across ….

NOTE: The programme also counts the full stop if it is separated from the text by a blank/space.

freq 2-02-19.cha +f
(This provides you with a list of words and their frequencies, for all speakers, organised **alphabetically**, stored in a file)

freq 2-02-19.cha +o +f
(This provides you with a list of words and their frequencies, for all speakers, organised according to **frequency**, stored in a file)
319 the
273 a
191 you
189 and
110 no
103 it …..

freq 2-02-19.cha +t*CHI +f
(This provides you with a list of words and their frequencies, for the child only (+t*CHI), organised **alphabetically**, stored in a file)

freq 2-02-19.cha +t*MOT +f
(This provides you with a list of words and their frequencies, for the mother only (+t*MOT), organised **alphabetically**, stored in a file)

NOTE: If you do not know what the labels for the different speaker tiers are, you can find them at the beginning of each transcript.

freq 2-02-*.cha +t*MOT +u +f
(This provides you with a list of words and their frequencies, for the mother only (+t*MOT), organised **alphabetically**, combined for all transcripts specified in the command line, stored in a file)

Distributional Analyses

Many studies on child-directed speech investigate how particular words, word forms or morphemes are distributed in the child's input and how this affects the acquisition of these elements. In order to investigate the distribution of words, word forms or morphemes in children's input, one can use the COMBO command.

Using CLAN

Typically one uses the COMBO command with the +f switch and the +s switch to save a list of utterances with a particular element or combination of elements to a file. If one wants to get more context, one can use the +w and –w switches to specify how many preceding and how many following utterances should be added to the target utterances. The +t switch can be used to restrict the search to particular tiers/speakers.

Example: Co-occurrences

If one studies formulas like *where-s-the*+X or *that-a*+X in early child language, it is important to check whether the particular word combinations that the child uses are common in the input – and more common than other combinations. For this, one has to find utterances that contain particular words or word combinations. Moreover, in order to investigate the basis for children's morphological generalizations, one might want to see with how many tokens and with how many different types of words a particular morpheme is combined in the child's input.

combo 2-02-*.cha +t*MOT +f +sthe
(This provides you with all maternal utterances containing the definite article, allowing you to calculate with how many types and tokens of nouns this grammatical morpheme is combined)

combo 2-02-*.cha +t*MOT +f +s*s
(This provides you with all maternal utterances containing words ending in *s*, allowing you to calculate with how many types and tokens of nouns this grammatical morpheme is combined. NOTE: this command will also output utterances with words where *s* is not a grammatical morpheme, e.g. *bus*. If the transcript has a mor-line, you might be able to search for s-morphemes there, adding +t%mor)

combo 2-02-*.cha +t*MOT +f +sun*

(This provides you with all maternal utterances containing words starting with *un*, allowing you to calculate with how many types and tokens of nouns the prefix *un* is combined. NOTE: this command will also output utterances with words where *un* is not a prefix, e.g. *under*. If the transcript has a mor-line, you might be able to search for un-morphemes there, adding +t%mor)

combo 2-02-*.cha +t*MOT +sthe^car +f

(This provides you with all maternal utterances containing the word *car* preceded by the defininte article, not separated by other words or not. NOTE: +sthe^*^car* would find utterances with *the* preceding *car*, even if they are separated by other words)

Example: Contrasts

Contrasts between word forms might help children to detect morphemes (e.g. *car* vs. *cars*). You can use FREQ to find out whether several forms of the same word appear in the same transcript. For suffixes, it is often useful to have the output in alphabetic order so that you can find forms like *car* and *cars* next/close to one another. For prefixes, you can use Word search functions to find prefixed forms. You might also want to see whether contrasting forms appear within the same sentence, which would make contrasts more salient (e.g. .*I have one blue car and two green cars.*). You can explicitly search for this, using the ^ symbol.

combo 2-02-*.cha +t*MOT +scar*^*^car* +f

(This provides you with all maternal utterances containing at least two forms of the word *car* words, whether they are separated by other words or not. NOTE: +scar*^car* would only find adjacent forms of *car*)

If you want to look whether other forms of the same word appear in preceding or following utterances, you can use +w or –w.

Example: Variation Sets

Child-directed speech is characterized by sequences of adult utterances that have a constant communicative intention, but show different types of variation in form; e.g.: lexical substitution and rephrasing, a shift from full noun phrases to pronouns, the addition, deletion or reordering of constituents (Küntay and Slobin 1996, 2002, Slobin et al. 2011, Waterfall ms.). Such "variation sequences", in combination with syntactic simplification and strong prosodic patterns, can highlight constituent boundaries as well as morphological contrasts and provide evidence for word order flexibility, syntactic processes and the optionality of particular constituent types (see Slobin et al. 2011 for overviews and discussion). According to Waterfall (ms.), nouns and verbs appearing in variation sets are picked up by the child more easily than others, even when word frequency is controlled for.

J (2;1) (see Slobin et al. 2011)

let's put J's bottles in the refrigerator
want to put them in the refrigerator with me
let's put J's bottles in the refrigerator
we'll put it in the refrigerator
let's put it in the refrigerator
we'll put it in the refrigerator
you can put it in
I'll let you put it in yourself
you put it right in
you put it in there
put it right in the refrigerator

⇒ optionality of adverbials like "with me"
⇒ link between pronouns and the correspoinding full noun phrases (e.g. *J's bottle* and *them/it*)
⇒ imperative vs. declarative (*you put it in there* vs. *put it right in the refrigerator*)
⇒ person contrasts (*let'(**u**)s* vs. *let **you***)
⇒ relationship between particles and prepositional phrases (*in* vs. *in the refrigerator*)
⇒ identifying auxiliaries and modals (*you **can** put it in* vs. *you put it in there*,…)
⇒ use of infinitive with vs. without *to* (*want to put them in the refrigerator with me* vs. *want **to** put them in the refrigerator with me*,…)
⇒ ….

Variation sets can be created by several speakers, see e.g. the following example from Sarah (CHILDES, Brown 1973, age of Sarah: 4;11). Here, the mother and the father are varying what the other person is saying:

*FAT: she was callin(g) them off .
*FAT: we were playin(g) Pokino the other night and she was callin(g) off the cards ?
*FAT: y(ou) know you got ta call em off just just like a bingo game ?
*GAI: yeah .
*FAT: and she called em all off .
*MOT: she called every one of those cards .
*FAT: she didn't know a jack .
*MOT: she knew an ace (.) a king (.) a queen .
*MOT: she knew all the &uh (.) diamonds (.) spades (.) clubs (.) and hearts .
*MOT: all the suits she knew .

Defining and finding variation sets is not that straightforward, see Waterfall (ms.) for discussion. You can use COMBO with the +w/-w switch if you are looking for the use of particular elements/strings in variation sets.

Analyses of Feedback

Many studies on child-directed speech investigate which types of feedback receive and whether this has an impact on their own production (Chouinard et al. 2003, Eisenbeiss 2009, Farrar 1990, 1992, Marcus 1993, Pinker 1989, Saxton 1997, Saxton et al. 1998). Some relevant feedback categories are:

- explicit corrections
- reformulations (of correct or non-target-like forms) and rephrasing
- positive responses like *yes*, *well done*, *exactly,...*
- negative responses like *no, this is not right, …*
- exact repetitions
- clarification questions
- ….

Using CLAN

You can see how parents respond by using the +w switch to add following utterances. Often, researchers are interested in responses to non-target-like forms. In many CHILDES transcripts, these are indicated by [*]. You can search for this using +s[*]. NOTE: You cannot search for this code +s[*]. Without the \ character, * is interpreted as a joker/wildcard.

References

Brown, R. (1973). *A First Language: The Early Stages.* London: Allen & Unwin.

Chouinard, M. M. & Clark, E. V. (2003). Adult reformulation of child errors as negative evidence. *Journal of Child Language* 30, 637–69.

Eisenbeiss, S. (2009) Generative approaches to language learning. *Linguistics* 47(2): 273–310.

Farrar, J. (1990). Discourse and the acquisition of grammatical morphemes. *Journal of Child Language* 17, 607-24.

Farrar, J. (1992). Negative evidence and grammatical morpheme acquisition. *Developmental Psychology* 28, 90-98.

Gallaway, C. & Richards, B. (eds.) (1994). Input and interaction in language acquisition. London: Cambridge University Press.

Küntay, A. and Slobin, D. I. 1996. Listening to a Turkish mother: some puzzles for acquisition. In: Slobin, D.I., Gerhardt, J., Kyratzis, A. and Guo, J. (eds.). *Social interaction, social context, and language*. Mahwah, NJ: Erlbaum, 265-286.

Küntay, A. and Slobin. D. I. 2002. Putting interaction back into child language: examples from Turkish. *Psychology of Language and Communication* 6:5-14.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk.* Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Marcus, Gary F. (1993). Negative evidence in language acquisition? *Cognition* 46, 53-85.

O'Grady, William. 1997. *Syntactic development.* Chicago: University of Chicago Press.

Pinker, Steven (1989). *Learnability and Cognition: The Acquisition of Argument Structure.* Cambridge, MA: MIT Press.

Saxton, M. (1997) The contrast theory of negative input. *Journal of Child Language 24,* 139-161.

Saxton, M.; Kulcsar, B.; Marshall, G., & Rupra, M. (1998) Longer-term effects of corrective input: an experimental approach. *Journal of Child Language 25,* 701-721.)

Slobin, D. I., Bowerman, M., Brown, P., Eisenbeiss, S. & Narasimhan, B. (2011). Putting Things in Places: Developmental Consequences of Linguistic Typology. (downloadable: http://essex.academia.edu/SonjaEisenbeiss/Papers/110813/Putting-things-in-places--Developmental-consequences-of-linguistic-typology

Waterfall, H.R. (ms.) *A little change is a good thing: The relation of variation sets to children's noun, verb and verb-frame development*