

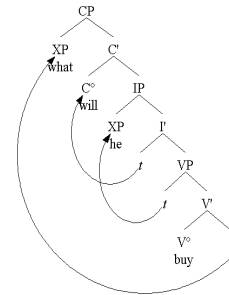
GRS LX 865

Topics in Linguistics

Week 1. CHILDES, root infinitives, and null subjects

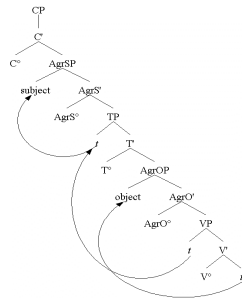
Syntax

- Recall the basic structure of adult sentences.
- IP (a.k.a. TP, INFLP, ...) is the position of modals and auxiliaries, also assumed to be home of tense and agreement.
- CP is where wh-words move and where I moves in subject-aux-inversion



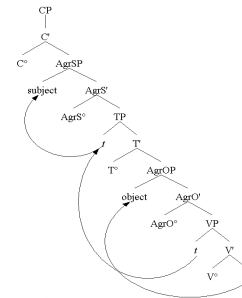
Splitting the INFL

- Syntax since 1986 has been more or less driven by the principle "every separable functional element belongs in its own phrase."
- Various syntactic tests support these moves as well (cf. CAS LX 523).



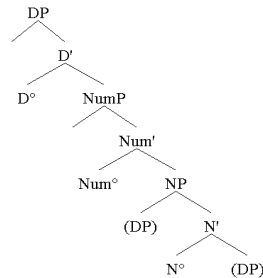
Splitting the INFL

- Distinct syntactic functions assigned to distinct functional heads.
 - T: tense/modality
 - AgrO: object agreement, accusative case
 - AgrS: subject agreement, nominative case
 - Neg: negation
- Origins: Pollock (1989) (split INFL into Agr and T), Chomsky (1993) (split INFL into AgrS, T, AgrO).



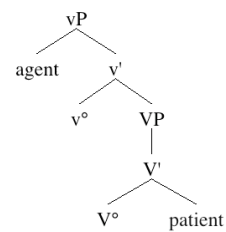
Functional heads

- The DP, CP, and VP all suffered a similar fate.
- DP was split into DP and NumP
 - Origin: Ritter 1991 and related work



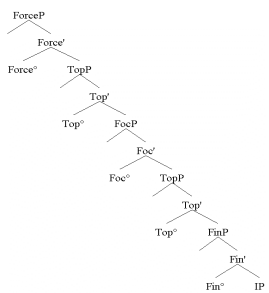
Functional heads

- VP was split into two parts, vP where agents start, and VP where the patient starts. V and v combine by head movement.
 - Origins: Larson (1988) proposed a similar structure for double-object verbs, Hale & Keyser (1993) proposed something like this structure, which was adopted by Chomsky (1993).



Functional heads

- CP was split into several “discourse-related” functional heads as well (topic, focus, force, and “finiteness”).



- Origins: Rizzi (1997)

Functional structure

- Often, the “fine structure” of the functional heads does not matter, so people will still refer to “IP” (with the understanding that under a microscope it is probably AgrSP, TP, AgrOP, or even more complex), “CP”, “DP”, etc.
- The heart of “syntax” is really in the functional heads, on this view. Verbs and nouns give us the *lexical content*, but functional heads (TP, AgrSP, etc.) give us the syntactic *structure*.

How do kids get there?

- Given the structure of adult sentences, the question we’re concerned about here will be in large part: how do kids (consistently) arrive at this structure (when they become adults)?
- Kids learn it (patterns of input).
 - Chickens and eggs, and creoles, and so forth.
- Kids start out assuming the entire adult structure, learning just the details (Does the verb move? How is tense pronounced?)
- Kids start out assuming some subpart of the adult structure, complexity increasing with development.

Testing for functional structure

- Trying to answer this question involves trying to determine what evidence we have for these functional structures in child syntax.
- It’s not very easy. It’s hard to ask judgments of kids, and they often do unhelpful things like repeat (or garble) things they just heard (probably telling us nothing about what their grammar actually is).

Testing for functional structure

- We do know what various functional projections are *supposed* to be responsible for, and so we can look for evidence of their effects in child language.
- This isn’t foolproof. If a child fails to pronounce the past tense suffix on a verb that was clearly intended to be in the past, does this mean there’s no TP? Does it mean they simply made a speech error (as adults sometimes do)? Does it mean they haven’t figured out how to pronounce the past tense affix yet?

Helpful clues kids give us

- Null subjects
 - Kids seem to drop the subject off of their sentences a lot. More than adults would. There’s a certain crosslinguistic systematicity to it as well, from which we might take hints about kids’ functional structure.
- Root infinitives
 - Kids seem to use nonfinite forms of main (root) clause verbs where adults wouldn’t. Again, there’s a certain crosslinguistic systematicity to it that can provide clues as to what’s going on.

Null subjects

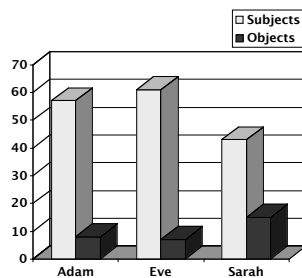
- Lots of languages *allow* you to drop the subject.
 - Italian, Spanish: the verb generally carries enough inflection to identify the person, number of the subject.
 - Chinese: where the subject is obvious from context it can be left out.
 - Not in English though: Let's talk about Bill. *Left. *Bought groceries. *Dropped eggs.
- On the view that kids know language, but are just trying to figure out the specific details (principles and parameters), one possibility is that they always start out speaking Italian (or Chinese) until they get evidence to the contrary.
 - (Hyams 1986 made a very influential proposal to this effect)

Null subjects

- Kids do tend to speak in short sentences. There seem to in fact be identifiable stages in terms of the length of the kids' sentences (one-word stage, two-word stage, multi-word stage...), often measured in terms of MLU (mean length of utterance) which roughly corresponds to linguistic development.
- Perhaps the kid's just trying to say a three-word sentence in a two-word window, so something has to go.
- That is, some kind of *processing* limitation.

Subject vs. object drop

Percentage of missing subjects and objects from obligatory contexts



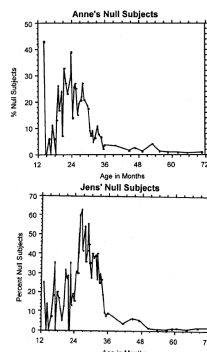
	A	E	S
Subject	57	61	43
Object	8	7	15

Null subjects

- Subjects (in a non-null subject language like English) are *way* more likely to be dropped than objects. There's something special about subjects.
- Makes a processing account more difficult to justify.
- Bloom (1990) made some well-known proposals about how the null subject phenomenon could be seen as a processing issue, and tried to explain why subjects are the most susceptible to being dropped. See also Hyams & Wexler (1993) for a reply.

Null subjects vs. time

- Null subjects seem to be pretty robustly confined to a certain portion of linguistic development. There's a pretty sharp dropoff at around 2.5 or 3.
- Hamann's Danish kids illustrate this well.



Why can't English kids really be speaking Italian?

- In Italian, subjects can be dropped (but *need not* be), in English, they can't be dropped at all.
- So since *having* subjects is consistent with Italian, what's going to signal to the kid that they've got the wrong kind of language?
- A "subset" problem.
 - Possible solution? Expletive *it* and *there*.
- In Italian, null subjects are allowed wherever a subject pronoun would be, including embedded finite clauses ("I know that [he] has left") and finite root questions ("What has [he] bought?").
- In Kid English, null subjects never show up in these environments. It doesn't seem so much like Italian.

Optional/root infinitives

- Kids around the age of 2 also sometimes use infinitives instead of finite verbs in their main clauses.
- It's "optional" in that sometimes they get it right (finite) and sometimes they get it wrong (nonfinite), at the same developmental stage.
- French:
 - Pas manger la poupée
not eat[inf] the doll
 - Michel dormir
Michel sleep[inf]
- German:
 - Zahne putzen
teeth brush[inf]
 - Thorstn das haben
Thorsten that have[inf].
- Dutch:
 - Ik ook lezen
I also read[inf.]

Root infinitives

- English kids do this too, it turns out, but this wasn't noticed for a long time.
 - It only write on the pad (Eve 2;0)
 - He bite me (Sarah 2;9)
 - Horse go (Adam 2;3)
- It *looks* like what's happening is kids are leaving off the -s.
- Taking the crosslinguistic facts into account, we now think those are nonfinite forms (i.e. *to write, to bite, to go*).

Root infinitives seem nonfinite

- Poeppel & Wexler (1993) looked at V2 in German (where finite verbs should be in second position, nonfinite verbs should be at the end)
- They concluded: the finiteness distinction is made correctly at the earliest observable stage.

	+finite	-finite
V2, not final	197	6
V final, not V2	11	37

CHILDES

- Child Language Data Exchange System
<http://chilDES.psy.cmu.edu>
 - Founded in 1984, Concord, MA.
Director Brian MacWhinney macw@mac.com.
- A source of, among other things, computerized—*searchable*—transcripts of child speech.
 - Note: When using data from CHILDES, you must always cite the original source of the data. See the CHILDES database manual for details on what to cite for each corpus.

Components

- CHAT: Chat is a transcription protocol common to most transcripts in the CHILDES database.
- CLAN: CLAN is a program (actually a collection of programs) used to transcribe data and analyze transcripts.
- CHILDES: The database itself consists of the transcripts (or other data, e.g., video, audio).

CHAT

- The CHAT format guidelines for coding your own transcripts are quite involved
 - see the 130-page manual for details.
- headers
 - @Participants
 - speaker "tiers"
 - *CHI, *PAT:
 - unintelligible speech
 - "xxx", ignored.
 - "xx", a word.

```
@UTF8
@Begin
@Languages:      en
@Participants:   CHI Peter Target_Child, MOTHER
                 PAT Patsy Investigator, LYN Lynn Investigator
                 Child
@ID:             enbloom701CHI12;1.1male|normal|1|Tar
@ID:             enbloom701MOT1111|Mother|1
@ID:             enbloom701ILO11111|Investigator|1
@ID:             enbloom701PAT11111|Investigator|1
@ID:             enbloom701LYN11111|Investigator|1
@ID:             enbloom701JEN11111|Child|1
@Tape Location:  Tape 16, side 1
@Comment:       MLU 2.39
@Time Start:    15:00
@Situation:     Peter is just waking up fr
adults talk about Jennifer who is now five
old
*PAT:          hey Pete # that's a nice new telepho
everything # it must ring and talk and .
%mor:          colhey n:prop|Pete pro:dem|that-v|be
n|look-PL v|like pro|it v:aux|must v|do pre
v|ring conj:cooland n|talk conj:cooland .
%exp:          Peter has a new toy telephone on tab
%com:          <def> untranscribed adult conversati
*CHI:          xxx telephone go right there .
%mor:          unk|xxx n|telephone v|go adv|right a
%fact:         <def> reaches out to lift phone rece
wire should connect receiver and telephone
*MOT:          the wire .
```

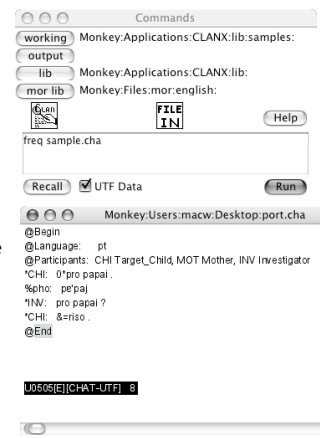
CLAN

- Analysis programs and transcript/text editor.
- Directories:
 - working: where it looks for transcript files to analyze
 - output: where it will put output files, default is working directory
 - lib and mor lib: where it looks for its own files, should be leave-able-as-is. If in doubt, set to lib in the same folder as the program file.



CLAN

- CLAN button: pops up command list.
- FILE IN: choose file(s) to analyze.
- Recall: get back previous command.
- Command window: where the real action is. We don't need no stinkin' buttons.
- Run: perform the action you asked for in the Command window.



CLAN

- Useful commands:
 - freq: calculate frequency of words in transcript(s) (page 71).
 - combo: search for things in the transcripts (page 56).
 - mlu: calculate mean length of utterance in the transcripts (page 94).



mlu

- The mlu command computes the mean length of utterance in morphemes. Used as a rough measure of the child's linguistic development.
 - Requires that CLAN can tell what the morphemes are.
- Many transcripts are tagged with %mor tiers for this purpose. Morphemes are delimited by, e.g., -, &, and ~ (see CHAT manual)
 - what' re...
pro:wh | what~v | be&PRES ...
 - ...brought...
...v | bring&PAST...

```
*LOI: why don't you bring your telephone to
%mor: adv:wh|why v:aux|do~neg|not pr|you
adv|down adv:loc|here n:prop|Peter ?
*LOI: why don't you put it on the floor ?
%mor: adv:wh|why v:aux|do~neg|not pr|you
?
%act: <aft> Peter puts it on floor <aft> P
to phone and receiver
%com: <aft> untranscribed adult conversati
what~re~you~bring_2
*LOI: what do you need ?
%mor: pro:wh|what~v|be&PRES pr|you partid
*CHI: xxx .
%act: <aft> Peter goes to hall closet, tri
what do you need ?
%mor: pro:wh|what v|do pr|you v|need ?
*CHI: xxx .
%mor: unk|xxx .
%mor: no # don't # see ?
colno v:aux|do~neg|not v|see ?
%gpx: pointing to hook which locks closet :
%com: <aft> untranscribed adult conversati
xxx .
*CHI: unk|xxx .
%mor: <def> goes to his room looking for t
well # they brought something too .
%mor: col|well v|bring&PAST pr:in
%act: <def> sends him back
*PAT: shall we take the ark ?
%mor: v:aux|shall pr|we v|take det|the n|
%act: <aft> goes to Peter's room, suggests
```

freq

- The freq command tallies up the number of times each word appears in the transcript.
 - Useful to figure out which words are most common (or which words are used at all) in a child's transcript.

```
> freq sample.cha
freq sample.cha
Sun Sep 12 19:48:56 2004
freq (10-Sep-2004) is conducting analyses on:
ALL speaker tiers
*****
From file <sample.cha>
1 a
1 any
1 are
3 chalk
1 chalk+chalk
1 delicious
1 don't
1 eat
[ . . . ]
1 toys+s
2 toys
3 want
1 what
2 what's
1 wonderful
2 yeah
2 you
-----
34 Total number of different word types us
50 Total number of words (tokens)
0.680 Type/Token ratio
```

combo

- The combo command is used to search for patterns in the transcripts.
- For all of the commands (including freq and mlu), there are certain options you should specify:
 - Tier +t*CHI
 - Input file(s) nina*
 - Output file > outfile.txt
- For example:
 - freq +t*CHI nina10.cha > freq-nina10.txt
 - mlu +t*CHI nina* > mlu-nina.txt

combo options

- In addition to those, combo has a couple of other options we care about:
 - `+s"eat"` search for...pattern in "..."
 - `+s@fname` search for...patterns in *fname*
 - `+w2` show 2 lines after a found result
 - `-w2` show 2 lines before a found result
- For example:
 - `combo +w2 -w2 +s"eat" nina10.cha > eatn10.txt`

Searches with combo

- `x^y`
finds x immediately followed by y (full words)
- `*`
finds anything
- `x+y`
finds x or y
- `!x`
finds anything but x
- `_`
finds any one character
- `x^*^y`
finds x eventually followed by y
- `*ing`
finds anything ending in ing
- `the^*^!grey^*^(dog+cat)`
finds the followed eventually by something other than grey, followed eventually by either dog or cat. Finds the black cat, the big red dog, but not the grey cat (though: why?)

Fabulous... now what does this have to do with root infinitives?

- Harkening back, we talked about a couple of ideas about what's wrong with kids' trees.
- Each idea makes *predictions* about what kids will and won't say—and CHILDES can be used to see to what extent these predictions are met.
 - *Relatively* painless computerized searching
 - relative to pen and paper, at least
 - A lot of data available, a lot of kids available

Harris & Wexler (1996)

- Child English bare stems as "OIs"?:
 - In the present, only morphology is 3sg -s.
 - Bare stem isn't unambiguously an infinitive form.
 - No word order correlate to finiteness.
- OIs are clearer in better inflected languages. Does English do this too? Or is it different?
- Hypotheses:
 - Kids don't "get" inflection yet; *go* and *goes* are basically homonyms.
 - These are OIs, the -s is correlated with something systematic about the child syntax (e.g., a structure missing T).

Harris & Wexler (1996)

- Exploring a consequence of having T in the structure: *do* support.
- Rationale:
 - Main verbs do not move in English.
 - Without a modal or auxiliary, T is stranded: The verb -ed not move.
 - *Do* is inserted to save T.
 - Predicts: No T, no *do* insertion.

Harris & Wexler (1996)

- Empirically, we expect:
 - She go
 - She goes
 - She not go (no T no *do*)
 - She doesn't go (adult, T and *do*)
- but never
 - She not goes (evidence of T, yet no *do*).
- On the other hand: All should be valid options if kids just don't "get" inflection.

Harris & Wexler (1996)

- Looked at 10 kids from 1;6 to 4;1
 - Adam, Eve, Sara (Brown), Nina (Suppes), Abe (Kuczaj), Naomi (Sachs), Shem (Clark), April (Higginson), Nathaniel (Snow).
- Counted sentences...
 - with *no* or *not* before the verb
 - without a modal/auxiliary
 - with unambiguous 3sg subjects
 - with either *-s* or *-ed* as inflected.

Harris & Wexler (1996)

- Affirmative:
 - 43% inflected
- Negative:
 - < 10% inflected
- It not works Mom
- no N. has a microphone
- no goes in there
- but the horse not stand ups
- no goes here!

	aff	neg
-inflec	782	47
+inflec	594	5

Harris & Wexler (1996)

- Small numbers, but in the right direction.
- Generalization: Considering cases with no auxiliary, kids inflect about half the time normally, but almost never (up to performance errors) inflect in the negative.
- If *do* is an indicator of T in the negative, we might expect to see that *do* appears in negatives about as often as inflection appears in affirmatives.
- Also, basically true: 37% vs. 34% in the pre-2;6 group, 73% vs. 61% in the post-2;6 group.

Harris & Wexler (1996)

- Also, made an attempt to ascertain how the form correlated with the intended meaning in terms of tense. (Note: a nontrivial margin of error...)
- Inflected verbs are overwhelmingly in the right context.

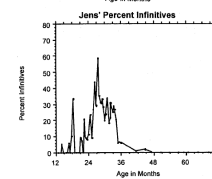
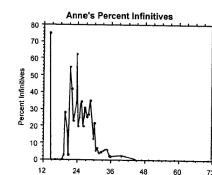
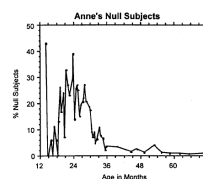
	present	past	future
bare stem	771	128	39
-s	418	14	5
-ed	10	168	0

NS/OI

- Some languages appear not to undergo the “optional infinitive” stage. Seems to correlate (nearly? perfectly?) with the target language’s allowance of null subjects. In principle, it would be nice to get this too, if it’s true. See, e.g., Wexler (1998).
 - **OI languages:** Germanic languages studied to date (Danish, Dutch, English, Faroese, Icelandic, Norwegian, Swedish), Irish, Russian, Brazilian Portuguese, Czech
 - **Non-OI languages:** Italian, Spanish, Catalan, Tamil, Polish

Root infinitives vs. time

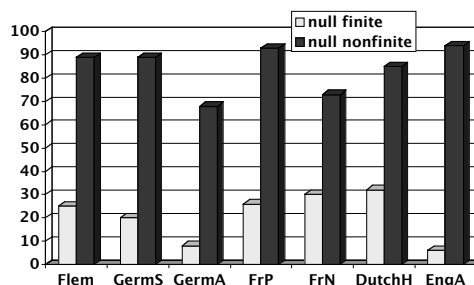
- The timing on root infinitives is likewise pretty robust, quitting around 3 years old. Cf. null subjects.



So what allows null subjects?

- Subjects of infinitives can be null.
 - I want to win the lottery.
- Kids at the age where subjects are often missing often use infinitive verb forms.
- Perhaps that's the key: Since kids can use infinitives where adults can't (main clause main verb), this allows them to use null subjects in those sentences as a side effect.

Proportion of null subjects in finite and non-finite clauses



Null subjects and infinitives

- Perhaps we're on to something here.
- So null subjects are (for the most part—not *completely*) allowed by virtue of having infinitives.
- What allows the infinitives in child language?
- Generally taken as some kind of “disturbance of IP” (e.g., TP is missing), home of both tense and the EPP.

Null subjects...

- Null subject parameter(s) is/are *not* initially mis-set (kids don't all start off speaking Italian or Chinese—contra Hyams 1986, 1992); rather, child null subjects are (at least in part) due to the availability of non-finite verbs (the OI stage).
- Most null subjects are licensed by being the subject of a nonfinite verb (i.e. PRO)
- But there are still *some* null subjects with finite verbs... More on this in a moment.

Whence the infinitives?

- Two major types of syntactic proposals:
- Truncation
 - What the kids do not know is that trees go all the way to CP, so they sometimes stop early, sometimes short of TP (e.g., Rizzi). Or they don't know about higher functional structure at all (e.g., Radford).
- Optional tense
 - Kids will sometimes leave out a projection in their tree (e.g., TP and/or AgrP), but the rest of it is still there (e.g., Wexler).
- What do these predict?

Back to null subjects vs. ±Fin

- Bromberg & Wexler (1995) promote the idea that null subjects with finite verbs arise from a kind of “topic drop” (available to adults in special contexts).
- **Proposal** (Bromberg & Wexler)
Topic-drop applies to Very Strong Topics

Kids sometimes take (in reality) non-VS topics to *be* VS topics (a **pragmatic** error)

Prediction about NS

- RI's have *two* ways of licensing NSs:
 - PRO (regular licensing of null subject)
 - Topic drop
- Finite verbs have *one* way to license a NS:
 - Topic drop
- So: We expect more null subjects with root infinitives (which we in fact see).
 - Cf. Rizzi: Subject in highest specifier can always be dropped, and RI's also allow PRO. Same story, basically.

Bromberg, Wexler, *wh*-questions, and null subjects

- If *topic drop* is something which drops a topic in SpecCP...
- ...and if *wh*-words *also* move to SpecCP...
- ...we would not expect null subjects with non-subject (e.g., *where*) *wh*-questions where the verb is *finite* (so PRO is not licensed).
 - Cf. Rizzi: Same prediction; if you have a CP, a subject in SpecTP won't be in the highest specifier, so it can't be dropped. One difference: Rizzi predicts no nonfinite *wh*-questions at all, hence no null subjects at all.

Bromberg, Wexler, *wh*-questions, and null subjects

Finiteness of null/pronominal subjects, Adam's *wh*-questions (Bromberg & Wexler 1995)

	Finite	Nonfinite
Null	2	118
Pronoun	117	131

*Truncation

- Rizzi's "truncation" theory predicts:
- No *wh*-questions with root infinitives
 - *wh*-question \Rightarrow CP, but
 - CP \Rightarrow IP, and
 - IP \Rightarrow finite verb
- And of course we wouldn't expect null subjects in *wh*-questions if null subjects are allowed (only) in the specifier of the root.

Adult null subjects ("diary drop")

- Both Rizzi and Bromberg & Wexler appeal to properties of adult language to justify the child null subjects.
 - B&W suggest that topic drop is available in English, but only for Very Strong topics, and what kids are doing wrong is identifying far too many things as VS topics.
 - Rizzi suggests that the ability to drop a subject in the highest specifier is available in certain registers ("diary drop") (where presumably Root=CP is disregarded, or at least relaxed to allow Root=IP).
 - Saw John today. Looked tired.

