GRS LX 865
Topics in Linguistics

Week 2. Optional infinitives and subject case

Subject case errors

- Various people have observed that kids learning English sometimes will use accusative subjects.
  - Her play.
- It turns out that there's a sort of a correlation with the finiteness of the verb as well. Finite verbs go with nominative case, while nonfinite verbs seem to go with either nominative or accusative case.
- But why *can* a nonfinite verb's subject be nom?

Finiteness vs. case errors

	Schütze & Wexler (1996) *Nina* 1;11-2;6		Loeb & Leonard (1991) 7 representative kids 2;11-3;4	
subject	Finite	Nonfinite	Finite	Nonfinite
he+she	255	139	436	75
him+her	**14**	120	**4**	28
% non-Nom	**5%**	46%	**0.9%**	27%

EPP and missing INFL

- If there were just an IP, responsible for both NOM and tense, then they should go together (cf. "IP grammar" vs. "VP grammar")
- Yet, there are many cases of root infinitives with NOM subjects
- And, even ACC subjects seem to raise out of the VP over negation (*me not go*).
- We can understand this once we consider IP to be split into TP and AgrP; tense and case are separated, but even one will still pull the subject up out of VP. (ATOM:+Agr –Tns)

What to make of the case errors?

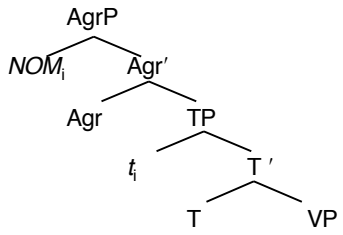
- Case is assumed to be the jurisdiction of AgrSP and AgrOP.
- So, nominative case can serve as an unambiguous signal that there is an AgrSP.
- Accusative case, conversely, may signal a missing AgrSP.
- Why are non-AgrSP subjects accusatives?
- Probably a default case in English:
  - Who's driving? Me. Me too. It's me.
- Other languages seem not to show this "accusative subject" error but also seem to have a nominative default (making an error undetectable).

"ATOM"

- Schütze & Wexler propose a model of this in which the case errors are a result of being able to either omit AgrSP or Tense.
- For a subject to be in nominative case, AgrSP must be there (TP's presence is irrelevant).
- For a finite verb, *both* TP and AgrSP must be there. English inflection (3sg present *–s*) relies on both. If one or the other is missing, we'll see an infinitive (i.e. bare stem).
- Thus, predicted: finite (AgrSP+TP) verbs show Nom (AgrSP), but only half of the nonfinite verbs (not both AgrSP and TP) show Nom (AgrSP). We should *not* see finite+Acc.

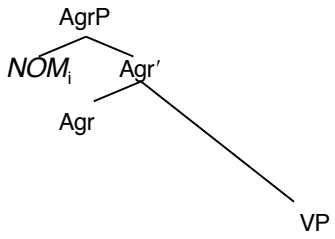
Agr/T Omission Model (ATOM)

- Adult clause structure:



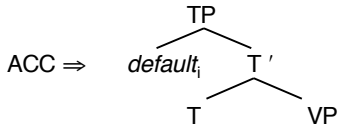
ATOM

- Kiddie clause, missing TP (—TNS):



ATOM

- Kiddie clause, missing AgrP (—AGR):



Pronunciation of English

- T+AgrS(+V) is pronounced like:
  - /s/ if we have features [3, sg, present]
  - /ed/ if we have the feature [past]
  - Ø otherwise
- Layers of "default", most specific first, followed by next most specific ("Distributed Morphology", Halle & Marantz 1993).
- Notice: 3sg present –s requires both TP and AgrSP, but past –ed requires only TP (AgrSP might be missing, so we might expect some accusative subjects of past tense verbs).

One prediction of ATOM

- +AGR+TNS: NOM with inflected verb (*-s*)
- +AGR–TNS: NOM with bare verb
- –AGR+TNS: *default* (ACC) with bare verb
- –AGR–TNS: GEN with bare verb
  (the GEN case was not discussed by Wexler 1998, but see Schütze & Wexler 1996)
- *Nothing* predicts Acc with inflected verb.

Finite pretty much always goes with a nominative subject.

	Schütze & Wexler (1996) *Nina* 1;11-2;6		Loeb & Leonard (1991) 7 representative kids 2;11-3;4	
subject	Finite	Nonfinite	Finite	Nonfinite
he+she	255	139	436	75
him+her	**14**	20	**4**	28
% non-Nom	**5%**	46%	**0.9%**	27%

ATOM and morphology

- [+3sg +pres] = -s
- [+past] = -ed
- — = Ø
- [+masc +3sg +nom] play+[3sg+pres]
  - he plays.
- [+2sg +nom] play+[2sg +past]
  - you play.
- But is this knowledge built-in? *Hint*: no.
- [+masc, +3sg, +nom] = he
- [+masc, +3sg, +gen] = his
- [+masc, +3sg] = him
- [+fem, +3sg, +nom] = she
- [+fem, +3sg] = her
- [+1sg, +nom] = I
- [+1sg, +gen] = my
- [+1sg] = me
- [+2, +gen] = your
- [+2] = you

ATOM and morphology

- What if the child produces a lot of utterances like
  - her sleeping
  - her play
- and even
  - her sleeps
  - her goes to school
- but never uses the word *she*?
- ATOM predicts that agreement and nominative case should correlate.
- *Her goes to school* is predicted never to occur.
- So does this child's use of *her goes to school* mean ATOM is wrong?

Schütze (2001, *inter alia*)

- No.
- *Her goes to school* is not *necessarily* a counterexample to ATOM (although it is a candidate).
- Morphology must be learned and is crosslinguistically variable.
- *She* is known to emerge rather late compared to other pronouns.
- If the kid thinks *her* **is** the nominative feminine 3sg pronoun, *her goes to school* is perfectly consistent with ATOM.
- Hence, we should really only count *her*+agr correlations from kids who have demonstrated that they know *she*.

ATOM and morphology

- Morphology (under "Distributed Morphology") is a system of defaults.
- The most specified form possible is used.
- Adult English specifies *her* as a feminine 3sg pronoun, and *she* as a *nominative* feminine 3sg pronoun.
- If the kid doesn't know *she*, the result will be that all feminine 3sg pronouns will come out as *her*. That's just how you pronounce nominative 3sg feminine, if you're the kid.
  - Just like adult *you*.
- [+masc, +3sg, +nom] = he
- [+masc, +3sg, +gen] = his
- [+masc, +3sg] = him
- [+fem, +3sg, +nom] = she
- [+fem, +3sg] = her
- [+1sg, +nom] = I
- [+1sg, +gen] = my
- [+1sg] = me
- [+2, +gen] = your
- [+2] = you

Rispoli (2002, *inter alia*)

- Rispoli has his own theory of *her*-errors.
- Pronoun morphology is organized into "tables" (paradigms) basically, where each form has a certain weight.
- When a kid is trying to pronounce a pronoun, s/he attempts to find the entry in the table and pronounce it.
- The kid's success in finding the form is affected by "gravity". "Heavier" forms are more likely to be picked when accessing the table, even if it's not quite the right form. If it's close and it's heavy, it'll win out a lot of the time.
- *Her* by virtue of being both acc and gen is extra-heavy, and pulls the kid in fairly often.

Her plays

- ATOM and Rispoli make different predictions with respect to *her plays*.
- ATOM says it should never happen (up to simple performance error)
- Rispoli says case errors are independent of agreement, *her plays* is perfectly possible, even expected.
- Rispoli's complaints about Schütze's studies:
- Excluding kids who happen not to produce *she* in the transcript under evaluation is not good enough. The assumption is that this learning is monotonic, so if the kid ever used *she* (productively) in the past, the *her* errors should not be excluded.

Monotonicity

- Schütze assumes that use of *she* is a matter of *knowledge* of *she*. Once the kid knows it, and given that the adult version of the kid will know it, it's there, for good.
- Rispoli claims that the "weight" of *she* can fluctuate, so that it could be "known" but mis-retrieved later if *her* becomes too heavy.
- Rispoli (2002) set out to show that there is a certain amount of "yo-yo'ing" in the production of *she*.
- We'll focus on Nina, for whom we can get the data.

Nina *she* vs. *her*

- Rispoli's counts show Nina using *she* from basically the outset of her use of pronouns, and also shows a decrease of use of *she* at 2;5.

	she	her
2;2 13-15	2 4%	43 96%
2;3 16-19	1 8%	12 92%
2;4 20-23	1 14%	6 86%
2;5 24-31	7 9%	73 91%

Checking Rispoli's counts

- 2;2
  - *CHI: she have hug a lady .
  - *CHI: she have jamas@f on .
- 2;3
  - *MOT: does she like it ?
  - *CHI: she drink apple juice .
  - *CHI: her like apple juice .
- 2;4
  - *MOT: he's up there ?
  - *CHI: no # she's not up there .
  - *CHI: he's up there .

- These are the times when Nina used *she* (twice at 2;2, once at 2;3, once at 2;4).
- Rispoli found 7 at 2;5, we'll deal with them later.

Checking

- 2;2
  - *CHI: helping her have a yellow blanket .
  - *MOT: she has a yellow blanket ?
  - *CHI: yeah [= yes] .
  - *CHI: her's ok .
  - *CHI: her ok .
  - *MOT: she's ok ?
  - *CHI: ok .
  - *CHI: her's ok .
  - *CHI: her ok .
  - *CHI: her's ok .
  - *MOT: she's ok .
- These three and one other time Nina said *her's ok* are the only candidate counterexamples at 2;2.
- At 2;2, 45 *her*+bare verb.
  - (R got 43, possibly including *her's ok*)
- At 2;3, no candidate counterexamples, 14 *her*+bare verbs.
  - (R got 12)
- At 2;4 none, 7 *her*+bare.
  - (R got 6)

Checking

- *MOT: what happened when I shampooed Miriam yesterday ?
- *CHI: her was cried .
- *MOT: oh # there's the dolly's bottle .
- *CHI: her's not going to drink it .
- *MOT: I'll start washing it .
- *MOT: see how clean it comes ?
- *MOT: you want to use the pot ?
- *CHI: a little bit .
- *CHI: her don't .
- *CHI: her's not dirty .
- *CHI: not dirty .

- 2;5:
  - I found about 76 *her*+bare/past verbs.
  - I found 3 potential counterexamples.

Bottom line?

- It doesn't seem like anything was particularly affected, even if Nina's early files were fully included.
- The number of possible counterexamples seems within the "performance error" range.
- The point about variation in usage of *she* is valid, worth being aware of the assumptions and being sure we're testing the right things.
- Rispoli was trying to make the point that if we'd accidentally missed a *she* in the early files, we might have excluded counterexamples there.
- Yet, even including *everything*, the asymmetry is strong.

Implementing ATOM

- The basic idea: In adult clauses, the subject needs to move *both* to SpecTP *and (then)* to SpecAgrP.
- This needs to happen because T "needs" something in its specifier (≈EPP) and so does Agr.
- The subject DP can "solve the problem" for both T and for Agr—*for an adult*.

Implementing ATOM

- Implementation: For adults:
  - T needs a D feature.
  - Agr needs a D feature.
  - The subject, happily, *has* a D feature.
  - The subject moves to SpecTP, takes care of T's need for a D feature (the subject "checks" the D feature on T). The T feature loses its need for a D feature, but the subject still has its D feature (the subject is still a DP).
  - The subject moves on, to take care of Agr.

Implementing ATOM

- Implementation: For kids:
  - Everything is the same except that the subject can only solve *one* problem before quitting. It "loses" its D feature after helping out either T or Agr.
  - Kids are constrained by the *Unique Checking Constraint* that says subjects (or their D features) can only "check" another feature once.
  - So the kids are in a bind.

Implementing ATOM

- Kids in a pickle: The only options open to the kids are:
  - Leave out TP (keep AgrP, the subject can solve Agr's problem alone). Result: nonfinite verb, nom case.
  - Leave out AgrP (keep TP, the subject can solve T's problem alone). Result: nonfinite verb, default case.
  - Violate the UCC (let the subject do both things anyway). Result: finite verb, nom case.
- No matter which way you slice it, the kids have to do something "wrong". At that point, they choose randomly (but cf. Legendre et al.)

Minimalist terminology

- Features come in two relevant kinds: interpretable and uninterpretable.
- Either kind of feature can be involved in a "checking"—only interpretable features survive.
- The game is to have no uninterpretable features left at the end.
- "T needs a D" means "T has an uninterpretable [D] feature" and the subject (with its normally interpretable [D] feature) comes along and the two features "check", the interpretable one survives. UCC=D uninterpretable on subjects?

NS/OI via UCC

- An old idea about NS languages is that they arise in languages where Infl is "rich" enough to identify the subject.
- Maybe in NS languages, AgrS does not *need* a D (it may in some sense be nouny enough to say that it *is*, or already *has*, D).
- If AgrS does not need a D, the subject is free to check off T's D-feature and be done.

The spreadsheet

- A spreadsheet is fundamentally a big table, with rows and columns, and each **cell** can contain data of any sort.
- What's fancy about spreadsheet programs is they allow you to enter **formulae** into a cell, computing the value based on the values in other cells.

	A	B
1	width	4
2	height	2
3	area	8
4		

=B1*B2

The spreadsheet

- The most obvious applications of this are mathy: financial, statistical, etc.
- But this can be quite helpful in organizing our data as we search through CHILDES.
- This is much better than simply marking things down on paper, since it counts everything for you and makes changes easy.

	A	B	C
1	fin	non fin	utterance
2	0	1	he go
3	1	0	she went
4	1	1	

=SUM(B2:B3)

What CLAN (combo) gives us

- We get a text file with some information about the search at the top, and then groups of utterances and context, with the found child utterance in the middle.

```
combo +t*CHI +w2 -w2 +s@prons1.txt peter07a.cha
Sun Sep  8 00:08:11 2002
combo (02-Aug-2002) is conducting analyses on:
  ONLY speaker main tiers matching: *CHI;
****************************************
From file <peter07a.cha>
----------------------------------------
*** File "peter07a.cha": line 52.
*MOT:    the wire .
*PAT:    oh <the &te> [//] the wire's gone ?
*CHI:    xxx # need it # (1)my need it # xxx .
*CHI:    xxx .
*PAT:    uhhuh .
----------------------------------------
*** File "peter07a.cha": line 207.
*CHI:    xxx # xxx .
*PAT:    what ?
*CHI:    this is # (1)I'll show you # (2)I'll show
*LOI:    you'll show me ?
*LOI:    ok .
----------------------------------------
*** File "peter07a.cha": line 329.
```

The plan

- Not every child utterance is relevant.
- The first part of our plan is to isolate the child utterances from the context so we can narrow down on just the relevant ones.

```
combo +t*CHI +w2 -w2 +s@prons1.txt peter07a.cha
Sun Sep  8 00:08:11 2002
combo (02-Aug-2002) is conducting analyses on:
  ONLY speaker main tiers matching: *CHI;
****************************************
From file <peter07a.cha>
----------------------------------------
*** File "peter07a.cha": line 52.
*MOT:    the wire .
*PAT:    oh <the &te> [//] the wire's gone ?
*CHI:    xxx # need it # (1)my need it # xxx .
*CHI:    xxx .
*PAT:    uhhuh .
----------------------------------------
*** File "peter07a.cha": line 207.
*CHI:    xxx # xxx .
*PAT:    what ?
*CHI:    this is # (1)I'll show you # (2)I'll show
*LOI:    you'll show me ?
*LOI:    ok .
----------------------------------------
*** File "peter07a.cha": line 329.
```

The plan

- We'll start by making a formula that counts the number of lines that start with "*" since the last line of dashes.
- The child's utterance will be the fourth one.

```
combo +t*CHI +w2 -w2 +s@prons1.txt peter07a.cha
Sun Sep  8 00:08:11 2002
combo (02-Aug-2002) is conducting analyses on:
  ONLY speaker main tiers matching: *CHI;
****************************************
From file <peter07a.cha>
----------------------------------------
*** File "peter07a.cha": line 52.
*MOT:    the wire .
*PAT:    oh <the &te> [//] the wire's gone ?
*CHI:    xxx # need it # (1)my need it # xxx .
*CHI:    xxx .
*PAT:    uhhuh .
----------------------------------------
*** File "peter07a.cha": line 207.
*CHI:    xxx # xxx .
*PAT:    what ?
*CHI:    this is # (1)I'll show you # (2)I'll show
*LOI:    you'll show me ?
*LOI:    ok .
----------------------------------------
*** File "peter07a.cha": line 329.
```

Computing "stars"

- We'll do this with a fancy formula.
- LEFT(C4,3) gives us the first (leftmost) 3 characters of the transcript line in C4.
- (LEFT(C4,3)="---") will be 1 if those three characters are "---" and 0 otherwise.
- Subtracting that from 1 will be 0 for "---" lines, and 1 otherwise.

	A	B	C
1	0		----------
2	1		*** File "peter07
3	2		*MOT: the wire
4	3		*PAT: oh <the

=((LEFT(C4,1)="*")+A3)*
(1-(LEFT(C4,3)="---"))

Computing "stars"

- LEFT(C4,1)="*" will be 1 if the transcript line starts with "*".
- We add that (1 if there's a "*") to the previous number (in A3, for cell A4). That is, count the "stars".
- Finally, for "---" multiply by zero (restart the count).

	A	B	C
1	0		----------
2	1		*** File "peter07
3	2		*MOT: the wire
4	3		*PAT: oh <the

=((LEFT(C4,1)="*")+A3)* (1-(LEFT(C4,3)="---"))

Counting child utterances

- Column B will keep track of how many child utterances there have been.
- That is, how many times A registers 4.
- The formula copies the previous number and adds one if column A has 4 in it.

	A	B	C
3	2	0	*MOT: the wire
4	3	0	*PAT: oh <the
5	4	1	*CHI: xxx # need it # (1)my need it
6	5	1	*CHI: xxx

=B5+(A6=4)

Getting the kid utt's alone

- Then, we'll start a fresh sheet and copy in just the child utterances.
- The idea: in row 1, we'll want to find the utterance where column B in our previous spreadsheet is (first) 1, in row 2…
- The utterance is in column C (column 3). We can also refer to this as R*row*C*column*.

	A	B	C
3	2	0	*MOT: the wire
4	3	0	*PAT: oh <the
5	4	1	*CHI: xxx # need it # (1)my need it
6	5	1	*CHI: xxx

C6 *or* R6C4

Getting the kid utt's alone

- Our earlier spreadsheet is named "raw", so raw!A1 is the content of A1 on sheet "raw", raw!B1:B800 refers to the cells in column 2, rows 1 through 800.
- ROW(A4) is simply the row number of cell A4 (namely, 4).

	A	B
1	5	*CHI: my need
2	12	*CHI: I'll show
3	19	*CHI: xxx
4	26	*CHI: xxx

=MATCH(ROW(A4), raw!B1:B800, 0)

Getting the kid utt's alone

- MATCH(a, cells, sort) finds the first "a" in cells when *sort* is 0.
- In this case, we're looking for the first 4 between B1 and B800 on the "raw" spreadsheet.
- The resulting number is the row number (from "raw").

	A	B
1	5	*CHI: my need
2	12	*CHI: I'll show
3	19	*CHI: xxx
4	26	*CHI: xxx

=MATCH(ROW(A4), raw!B1:B800, 0)

Getting the kid utt's alone

- =INDIRECT("raw!R2C2", FALSE) will copy the contents of raw!B2 (FALSE means to use the R2C2 type reference, not the B2 type).
- What we're doing is using the row number we just found (in column A), and column 3 (where the utterances are).
- raw!R26C3

	A	B
1	5	*CHI: my need
2	12	*CHI: I'll show
3	19	*CHI: xxx
4	26	*CHI: xxx

=INDIRECT("raw!R" & A2 & "C3", FALSE)

The plan continues

- At this point, we'll have the child utterances alone, so we can look at them and see if they contain a subject pronoun (and see which one) or if they contain an irrelevant match.

- My need it.
- My pencil.
- I'll show you.
- Show me.
- …

The plan continues

- We'll do a coloring trick to "grey out" the things we marked as irrelevant.
- We'll code the utterances for finite verbs, nonfinite verbs, or ambiguous forms.

- my going
- you go
- I'll show you
- he go
- he runs
- …

The plan continues

- After that, we'll bring back the context with a similar method so we can make sure that we're not counting repetitions, etc.

- And finally, we'll count up how many nominative subjects come with finite verbs, how many accusative subjects come with nonfinite verbs, etc.

What to do next

- We'll try this out on the peter07 file.
- Later, you'll adapt this to look at the nina13.cha (with not a great deal of modification).

- Run through the steps on the web page (or printout), now that we know what it's doing.

Comments about nina13

- When I did it…
- I found about 70 relevant utterances (where there is a pronoun subject and the verb is unambiguous) to pass on to the "subjects" sheet.
- Of those I omitted around 10 as repetitions or otherwise uninformative.
- Be particularly careful about the lower bounds on these larger blocks—nina13 is a bigger file than peter07, and so you will occasionally need to increase some of the numbers to get all of the utterances in.