Lecture 4:
The communicative environment, input and uptake
Outline

• What is actually in the input?
• An example of how children might learn from it
  – A cross linguistic model of the optional infinitive error
• The development of communicative intention and intention-reading
• Cross-cultural and cross linguistic issues
The input
Frequency of units of 1-5 words in 1.72 millions words of CDS

[Bannard & Matthews, 2009]
• 45% of mothers’ utterances start with one of 17 words
• 52 ‘core frames’ account for 51% of all utterances

Cameron-Faulkner, Lieven & Tomasello, 2003
Implications

• We start by counting at the level of specific form and string:
  – *is/are*
  – *I’m X-ing/You’re Y-ing*
  – *What do X?/What can X?*

• We only count at more abstract level, when there is evidence for it

• We do not credit the child with pre-given, abstract linguistic categories from the outset
Do typological differences affect repetitiveness in CDS?

- English has very fixed word order
  - *The tiger ate the mouse*
  - *The mouse ate the tiger*
- German has more word order variants than English but has case inflections
  - *Der Tiger frisst den Hund*
  - *Den Hund hat der Tiger gefressen*
- Russian has ‘free word order’
  - *Ja videl svoju mašinu* (all 24 words orders possible)

Stoll, Abbot-Smith & Lieven, 2009
Two possible hypotheses

**H0:** Independent of language we expect item-specificity at the beginning of utterances.

**H1:** The rigid word order of English determines the highly predictable beginning of utterances. The degree of word-order determination will determine the degree of item-specificity.
Comparing input across languages

- **ENGLISH** (Manchester corpus):
  - 6 mothers
  - children between 1;9-2;6
  - M = 1400 utterances per mother

- **GERMAN** (Szagun corpus):
  - 6 mothers
  - children at 1;8 and 2;5 (+ part of file 1;4)
  - 1400 utterances per mother

- **RUSSIAN** (Stoll corpus):
  - 4 mothers
  - children between 1;8 – 2;4
  - 1400 utterances per mother
Lexical frames

• First one-three words of utterance
  – (omitting communicators, onomatopoeia, noises)

• Frame = 4+ tokens per mother
What counted as a ‘frame’?

Within one mother:

• That’s a dog
• That’s a girl
• That’s a flower
• That’s your pen
What counted as a ‘frame’?

Example utterances:

- That’s a dog
- That’s a girl
- That’s a flower
- That’s your pen

• FRAME = That’s …
What counted as a ‘frame’?

Example utterances:

• That’s a dog
• That’s a girl
• That’s a flower
• That’s your pen
• That’s a lorry

• FRAME = That’s …
What counted as a ‘frame’?

Example utterances:

• That’s a dog
• That’s a girl
• That’s a flower
• That’s your pen
• That’s a lorry

• FRAME = That’s a …
Input – English, German and Russian

Stoll, Abbot-Smith & Lieven, 2009

Percentage of utterances by individual mothers accounted for by frames and core frames

<table>
<thead>
<tr>
<th>Eng. Frames</th>
<th>Ger. Frames</th>
<th>Rus. Frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>122</td>
<td>79</td>
<td>63</td>
</tr>
</tbody>
</table>

60%
Number of one-, two- and three-word frames for individual mothers
When English needs three words, Russian often needs only one  
e.g. Wh-question, copulas.

=> Russian often drops arguments, has no articles, zero in present tense copula.

German has gender in the article, so there are more possibilities

=> 3 word frames are less likely than in English where there is no gender in the article
Conclusions

• Middle-class CDS is highly repetitive in initial sequences in three typologically different languages

• Typology makes a difference to the degree of repetitiveness

• We don’t yet know how this affects learning
Frequency effects in language acquisition

• Correlations between how much children are spoken to and the size of their lexicons
• Correlations between relative frequency of specific, lexically-based strings in the input and their order of emergence in children’s language e.g. copula constructions (Cameron-Faulkner, Lieven & Tomasello, 2003)
• Correlations between provision of complex syntax by children and relative frequency of complex constructions used by teachers and parents (Huttenlocher et al, 2002)
‘Frequency’ is short-hand for the number of times an event is experienced

- Experiencing an event repeatedly, changes its representation in the sense of how entrenched the response pattern is by comparison with other response patterns.
- How often an event is experienced makes a profound difference to all aspects of language development and use.
- We often don’t know the right level of analysis for a frequency-based prediction, so we have to test for it.
OI errors in German, Dutch and Spanish

Freudenthal, Pine, Aguado-Orea & Gobet, 2007
The Optional Infinitive [OI] error

The child uses non-finite verb forms in contexts where finite verbs forms are obligatory *That go there vs. That goes there* (3sg present)

**Wexler’s explanation**

- The child’s grammar identical to adult’s except the child is subject to a Unique Checking Constraint that can result in underspecification of **Tense** and/or **Agreement**

  - The unique checking constraint may prevent the child from checking the D feature of the Subject DP against more than one D feature (tense and agreement). So either can be optionally unspecified: either no tense or wrong agreement (*Him naughty, Her coming*)

  - Explains OI in obligatory subject languages (English, Dutch, German) Explains few OI errors in optional subject languages (Spanish, Italian) where only one feature need usually be checked (*Tense*)
Strengths of the ATOM

• Explains statistical patterns of error in English
  – He goes and He go, but few I goes
  – He goes, He go and Him go but few Him goes

• Explains why children learning other obligatory subject languages (e.g. Dutch, French) use infinitives in main clauses
  – Hij lopen (He to walk) Il faire (He to do)

• Explains why children learning optional subject languages (e.g. Spanish) do not use infinitives in main clauses
  – (El) habla (He speaks) not *(El) hablar (He to speak)
The MOSAIC model [Pine, Gobet & Freudenthal, 2005]

MOSAIC is a simple distributional learner that:

• Learns utterance final words and sequences
  – Do you want a biscuit? Biscuit
     A biscuit
     Want a biscuit

• Generates novel utterances by linking together words that have been preceded and followed by overlapping sets of words and substituting them in utterance final sequences
  – a linked to the on basis of: Want a biscuit
     Want the ball

  – allows: Want the biscuit
     Eat a biscuit
     Eat the biscuit
MOSAIC-Speak

ROTE LEARNED
• DOESN’T FALL OUT
• CHEEKY FACE
• WHERE DO YOU WANT THEM TO GO?
• HOLD THE CASE THEN
• TELL GRANDMA THEN
• IT’S THE PHONE
• WHICH FRIENDS ARE THEY THEN?
• GONNA WEE IN THE POTTY

GENERATED
• MIGHT FALL OUT
• CHEEKY FOOT
• WHERE DO YOU WANT HIM TO GO?
• TAKE THE CASE THEN
• SHOW GRANDMA THEN
• IT’S A PHONE
• WHICH FRIENDS IS HE THEN?
• GONNA WEE IN THE BALLOON
# OI errors

<table>
<thead>
<tr>
<th></th>
<th>OI errors at lowest MLU point (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>75</td>
</tr>
<tr>
<td>German</td>
<td>61</td>
</tr>
<tr>
<td>Spanish</td>
<td>18</td>
</tr>
</tbody>
</table>
At the same MLU for the child and the MOSAIC output, measure the proportions of:

- Compound finites  
  Er hat es gesehen
  He has it seen

- Optional infinitives  
  Er es gesehen*
  He it seen

- Simple finites  
  Er sieht es
  He sees it
Pattern of finiteness marking as a function of MLU for Leo and MOSAIC-Leo (German)

MOSAIC simulates the moderately high proportion of OI errors in German (and low proportion of compound finites)
Pattern of finiteness marking as a function of MLU for Juan and MOSAIC-Juan (Spanish)

MOSAIC simulates the low proportion of OI errors in Spanish (and high proportion of simple finites)
When compared to children at the same MLU, MOSAIC simulates:

- the high proportion of OI errors in Dutch and low proportion of compound finites
- the moderately high proportion of OI errors in German and low proportion of compound finites
- the low proportion of OI errors in Spanish and high proportion of simple finites
### Why?

<table>
<thead>
<tr>
<th>Language</th>
<th>OI errors at lowest MLU point (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>75</td>
</tr>
<tr>
<td>German</td>
<td>61</td>
</tr>
<tr>
<td>Spanish</td>
<td>18</td>
</tr>
<tr>
<td>Language</td>
<td>OI errors at lowest MLU point (%)</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Dutch</td>
<td>75</td>
</tr>
<tr>
<td>German</td>
<td>61</td>
</tr>
<tr>
<td>Spanish</td>
<td>18</td>
</tr>
<tr>
<td>Language</td>
<td>OI errors at lowest MLU point (%)</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Dutch</td>
<td>75</td>
</tr>
<tr>
<td>German</td>
<td>61</td>
</tr>
<tr>
<td>Spanish</td>
<td>18</td>
</tr>
</tbody>
</table>
Some claims made about language learning

Cultures:
• in which children are not spoken to before they speak
• in which babies are not interacted with much

- Preverbal intention reading and communicative interaction is a prerequisite for language development

How similar/different is the communicative behaviour and environment of children from different cultures?

- Children only require minimal input to learn language
- Children can learn language through overhearing

Are children receiving minimal input?
Communicating with other minds

- Children start to communicate intentionally

- They start to behave as if others were communicating intentionally

- They show developing skills at reading other minds

Bruner, Bates, Gergely, Tomasello, Liszkowski, Warneken, Moll
Misunderstanding

[Liszkowski, 2006]
Uninterested
Helping (Warneken, 2006)
Early social cognition in three cultures
Callaghan, Moll et al. (submitted)

- Peru, Junin province: rural villages; 3,000m
- India, Andra Pradesh: rural villages
- Canada, Nova Scotia: small rural town

<table>
<thead>
<tr>
<th>Imitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrumental helping</td>
</tr>
<tr>
<td>Declarative pointing</td>
</tr>
<tr>
<td>Joint attention</td>
</tr>
<tr>
<td>Pretence</td>
</tr>
<tr>
<td>Pictorial symbols</td>
</tr>
</tbody>
</table>
Results

• Understanding intentions and attention
  – Similar across cultures and at similar ages
    • Except for one task involving locomotion around a barrier

• Sharing intentions and attention
  – Similar across cultures
    • Except Indian children slightly more collaborative at a younger age but less pointing

• Comprehending and using symbols
  – Canadian children between 2.5 – 3.0
  – Peruvian and Indian children approaching 4.0
Interaction in two cultures
Brown, 2008

– Tzeltal (Mexico)
- Rossel (Papua New Guinea)

Similarities:
Small-scale traditional societies
Extended households, multiple caregivers, child caregivers
Multiparty interactions the norm

Tzeltal:
Physical restraints
Few interlocutors,
  restricted interactional space
Nonresponsiveness to
  infant’s preverbal ‘utterances’
Little scaffolding
Low interaction density

Rossel:
Physical freedom
Many interlocutors,
  large interactional space
Responsiveness to
  preverbal ‘utterances’
Ample scaffolding
High interaction density
Interim results

• Pace of interaction much higher in Rossel
  – Twice as many initiations per minute ($\approx 7:3$)
  – Due to other interactants

• Rossel infants initiate interaction only very slightly more than Tzeltal infants

• Pointing similar and develops at same age
Comparison with a ‘technological culture’

(Stoll, Lieven et al.,)
## Data collection

<table>
<thead>
<tr>
<th>'Babies’</th>
<th>6m</th>
<th>8m</th>
<th>10m</th>
<th>12m</th>
<th>15m</th>
<th>18m</th>
<th>21m</th>
<th>24m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dipkala</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Saphal</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>'Two’-S</th>
<th>6m</th>
<th>8m</th>
<th>10m</th>
<th>12m</th>
<th>15m</th>
<th>18m</th>
<th>21m</th>
<th>24m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Khem</td>
<td>Monthly</td>
<td>Bi-monthly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kamala</td>
<td>Monthly</td>
<td>Bi-monthly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>'Three’-S</th>
<th>6m</th>
<th>8m</th>
<th>10m</th>
<th>12m</th>
<th>15m</th>
<th>18m</th>
<th>21m</th>
<th>24m</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalpana</td>
<td>Monthly</td>
<td>Bi-monthly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Man Kumar</td>
<td>Monthly</td>
<td>Bi-monthly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Ia Kuluke tusande
Look, Kuluke, digged!
What to compare with?

The Rigol corpus (Biberthal)

‘Babies’: Johanna Lars

‘Two’-s: Pauline Sebastian

‘Three’-s: Corinna Niklas
# Categories for characterising the communicative environment

<table>
<thead>
<tr>
<th>Proportions per hour</th>
<th>Child</th>
<th>Mother</th>
<th>Other adults</th>
<th>Other children</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minutes with utterances</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pointing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Imitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Offering</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaching/Requesting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attention getting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Showing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Touching</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feeding</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Interim results: Pointing

• Pointing goes up between 0;8-1;3, particularly after 1;0
• Individual differences most obvious
• No obvious correlation with the amount of pointing by mothers or by everyone to babies
• No obvious cultural difference but we need much more fine-grained analysis
Interim results: Other

- For babies, the main form of interaction seems to be dyadic, often, though not always, with the mother.
- Interacting with babies seems to afford the same types of interactions in both cultures.
- For Chintang toddlers, the part played by other children is always greater and increases with age.
- We cannot assess the volume of talk to the children from these results, but they are certainly being talked to.
- At least on these measures, individual differences can outweigh cultural differences.
The contexts of language development

- Children must learn language, at least in part, from what they hear.
- Almost all our research is based on very intensive, dyadic conversations between mothers and children from middle-class backgrounds in urban, technological societies.
- We have no idea how much is enough
- And enough for what?????
What is ‘naturalistic data’?

**Our study:**
Chintang and Biberthal

- Outside and inside
- Different situations
- Mother not always present
- Other children present

**Most previous studies:**

- Inside the house
- Mother and child playing
- Only mother present
- No other children present
Possible ways of learning distributions and form-meaning mappings

- Children could learn from other children
- Children could learn from listening and looking
- Caretaker talk may not be closely tied to the child’s vocalisations but might be tied to the child’s attentional behaviour
- Children could learn by imitating adults and then starting to vary the imitations