Crosslinguistic language development: How does what the child hears affect what is learned?

Lecture 4
Outline

• Is the input chaotic?
  • Studies of Child Directed Speech (CDS)
• Does the input affect learning?
  • Corpus studies
  • Experimental studies
  • Modelling studies
• What characteristics of input do children need?
  • Is CDS universal?: Anecdotal evidence from other cultures
  • A study of the communicative environment of children in a non-technological culture
Studies of child directed speech

• Most studies of CDS show:
  – Exaggerated prosodic contours
  – Mostly about the here-and-now
  – Mostly grammatical utterances, though quite a lot of single words and fragments
  – Repetitive
English Child Directed Speech

- Copulas: 15%
- Complex: 6%
- SV(X): 18%
- Imperatives: 9%
- Questions: 32%
- Fragments: 20%

- 12 mother-child dyads
- 4 half-hour recordings
- Mean of 1,400 per dyad

- 45% of mothers’ utterances start with one of 17 words
- 52 ‘core frames’ account for 51% of all utterances

- A X
- It’s a X
- What do X …?
- Are you X…?
- Lets X

Cameron-Faulkner, Lieven & Tomasello, 2003
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, in press
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.
DATA

• 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
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 …
Percentage of utterances by individual mothers accounted for by frames and core frames
Number of one-, two- and three-word frames for individual mothers

![Graph showing the number of one-, two- and three-word frames for individual mothers in English, German, and Russian.]
Percentage of utterances by individual mothers accounted for by one-, two-, and three-word frames
When English needs three words, Russian often needs only one
e.g. Wh-question, copulas.

=> Russian is pro-drop, 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 reptitiveness

• We don’t yet know how this affects learning
Relationships between input and learning

Corpus studies
CDS and language learning: English

• Effects on the rate of development of:
  – The amount of talk to children
  – Mothers who elaborate on the child’s focus of attention
  – Mothers who elaborate on what the child has just said
  – Mothers and teachers who use more complex syntax to preschool children

• Strong correlations at every level with frequency of forms, constructions etc in CDS and the order of emergence of these forms in the child’s speech

• But can we explain errors from the input?
Errors in inverted questions

- Omission
  Where he go?
- Double marking
  Can he can go?
- Non-inversion
  Where he does go?
- Agreement errors
  Does you go?
- Case errors
  What does her want?

Explanations:
- Cognitive complexity
- Arguments vs. adjuncts
- DO-support
- BE inversion
- Main vs. modal auxiliaries
Errors based on frames?

**Non-inversion**

M. You don’t throw things
C. *Why* you don’t throw things?

**Double marking**

*Why don’t* you don’t like cakes?  →  *Why don’t* + X  You don’t like cakes

**Agreement errors**

*Where does* you go?  →  *Where does* X go?  You

Ambridge, Rowland, Theakston & Tomasello (in press)
• The error rate is low because children are learning constructions with slots
• High frequency frames should be protected from error
• Errors will occur when there isn’t a frame
## Error rates in syntactic questions

<table>
<thead>
<tr>
<th></th>
<th>High frequency words</th>
<th></th>
<th>Low frequency words</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Frames</td>
<td>Non-frames</td>
<td>Frames</td>
<td>Non-frames</td>
</tr>
<tr>
<td></td>
<td>2.05</td>
<td>13.09</td>
<td>11.27</td>
<td>11.71</td>
</tr>
</tbody>
</table>

Rowland, 2007
Relationships between input and learning

Experimental studies
Do omission errors derive from what children hear?

‘Optional’ stage: the same verb appears with and without 3rd person

<table>
<thead>
<tr>
<th>WITH</th>
<th>WITHOUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>he goes</td>
<td>he go/he going</td>
</tr>
</tbody>
</table>

Hypothesis 1: Children have abstract categories from the beginning including an **innate knowledge of tense** but think its optional (Wexler & Rice)

Hypothesis 2: Children **learn about tense-marking**. Before this they have learned both forms of some verbs but will only use a novel verb as they hear it (Pine et al.)
Optional infinitives: Input-based hypothesis

- Children will produce what they hear
- They hear many verbs with both finite and non-finite forms with adjacent Subjects

Can it go there?  
It goes here

This one jumps  
Does that one jump?
3rd person marking experiment

GAME 1: [Condition 1: all verbs unmarked]

- Will this one spin (known verb 1)
- Will this one swing (known verb 2)
- Will this one tam? Should it tam? Will it tam? (novel verb)

GAME 2: [Condition 2: all verbs 3rd person sing]

- This one jumps (known verb 1)
- This one rolls (known verb 2)
- This one mibS, Look, it mibS, it mibS (novel verb)

GAME 3: [Condition 3: mixed]

Theakston, Lieven & Tomsasello, 2003
Test questions: to elicit the use of verbs in FINITE contexts

• What does this one do?
• What does it do?
• It ______[s]?

Participants: 24 children, mean age 2;8
Conditions: Finite, Non-finite, Mixed between subjects
% Finite verb forms produced with known verbs

![Bar chart showing percentage of finite forms in different conditions: Finite, Non-finite, and Mixed.](chart.png)
% Finite verb forms produced with novel verbs
The development of abstract argument structure

• Who does what to whom?
   The fox ate the chicken

• Cues:
   Animacy
   Word order
   Case marking
   Agreement

• Experiments with Novel verbs
Cue validity

Cue availability: number of times a cue is present

Cue reliability: number of times a cue marks the function

Cue validity = availability x reliability
Animacy and word word cues in English, German and Cantonese

Chan, Lieven & Tomasello, in press
The animacy contrast cue

• **Cue Availability**
  + The dog chases the ball
  - The dog chases the cat

• **Cue Reliability**
  + The man opens the door
  - The ball hits the man
The word order cue

• Cue Availability
  + The dog chases the ball
  - chases

• Cue Reliability
  + The man opens the door
  - Den_{ACC} Hund schubst der_{NOM} Löwe
The animacy contrast cue

<table>
<thead>
<tr>
<th>Language</th>
<th>Availability</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cantonese</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The animacy contrast cue

<table>
<thead>
<tr>
<th>Language</th>
<th>Availability</th>
<th>Reliability</th>
<th>Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cantonese</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The animacy contrast cue

- highly reliable across languages
The animacy contrast cue

- highly reliable across languages
- availability is lower in Cantonese due to massive ellipsis and ambiguous pronouns
The word order cue

<table>
<thead>
<tr>
<th></th>
<th>English</th>
<th>German</th>
<th>Cantonese</th>
</tr>
</thead>
<tbody>
<tr>
<td>100%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Availability

Reliability

Validity
The word order cue

- Availability
- Reliability
- Validity

Languages:
- English
- German
- Cantonese

Percentages:
- 100%
- 75%
- 50%
- 25%
The word order cue

cue validity: English > German > Cantonese
Developmental Findings

(i) Animate Noun - Verb - Inanimate Noun (AVI)
(ii) Inanimate Noun - Verb - Animate Noun (IVA)
(iii) Animate Noun - Verb - Animate Noun (AVA)
AVI: The horse tams the telephone

% choice of 1st N as agent
100%
90%
80%
70%
60%
50%
40%

2;6 3;6 4;6

English AVI
German AVI
Cantonese AVI
AVI: The horse tams the telephone

% choice of 1st N as agent

<table>
<thead>
<tr>
<th>Language</th>
<th>2;6</th>
<th>3;6</th>
<th>4;6</th>
</tr>
</thead>
<tbody>
<tr>
<td>English AVI</td>
<td>40%</td>
<td>50%</td>
<td>60%</td>
</tr>
<tr>
<td>German AVI</td>
<td>50%</td>
<td>60%</td>
<td>70%</td>
</tr>
<tr>
<td>Cantonese AVI</td>
<td>60%</td>
<td>70%</td>
<td>80%</td>
</tr>
</tbody>
</table>
Across language groups, even the youngest 2-year-olds were above chance in choosing the 1st Animate Noun as the agent.

---

AVI: *The horse tams the telephone*
IVA: *The present meeks the chicken*

% choice of 1st N as agent

- 100%
- 90%
- 80%
- 70%
- 60%
- 50%
- 40%

Years:
- 2;6
- 3;6
- 4;6

Languages:
- English IVA
- German IVA
- Cantonese IVA
IVA: The present meeks the chicken

% choice of 1st N as agent

- English IVA
- German IVA
- Cantonese IVA

Leipzig Spring School
April 2008
IVA: *The present meeks the chicken*

- Across language groups, 2-year-olds were at chance group performance
IVA: *The present meeks the chicken*

% choice of 1st N as agent

- Across language groups, 2-year-olds were at chance group performance.
- Older children at 3;6 and 4;6 preferred word order over animacy.
AVA: The cow tams the giraffe

% choice of 1st N as agent

100%
90%
80%
70%
60%
50%
40%

2;6 3;6 4;6

English AVA
German AVA
Cantonese AVA
AVA: *The cow tams the giraffe*

% choice of 1st N as agent

- English AVA
- German AVA
- Cantonese AVA

Leipzig Spring School
April 2008
AVA: *The cow tams the giraffe*

% choice of 1st N as agent

Reliance on word order (as a marker of the agent-patient relations): English > German > Cantonese children
• Young children show differential and restricted competence in comprehension early on
  - ‘the horse tams the telephone’ versus ‘the present tams the chicken’

• The nature of the early transitive construction is locally-structured
  – around particular semantic types of participants

• The acquisition of the transitive construction is
  – protracted rather than instantaneous

• Children’s linguistic productivity is
  – tied closely to their linguistic experience
Relationships between input and learning

Modelling

Optional Infinitive errors

Freudenthal, Pine, Aguado-Orea, & Gobet (2007)
The AGR/TNS Omission Model

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

• The child uses non-finite verb forms in contexts where finite verbs forms obligatory
  – That go there v That goes there (3sg present)

• Since AGR assigns NOM, child also produces Non-NOM subjects when AGR absent
  – Him naughty, Her coming
The unique checking constraint

[Wexler]

- 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
- Child produces infinitives where finites required
- 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)
Can a model replicate the patterns of finite/non-finite marking in different languages?

- Model is trained repeatedly on speech addressed to a particular child
- Output generated after each run through input
- Output files selected on basis of MLU
- Compared with samples of child speech matched as closely as possible for MLU
- Data from child and model coded for non-finites, simple finites and compound finites using same (automated) coding procedures
The MOSAIC model

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: Key Features

- Takes as input (orthographically transcribed) samples of Child-Directed Speech
- Produces output in the form of ‘utterances’ that can be compared with those of real children
- Learns to produce progressively longer utterances as a function of the amount of input it has seen
Simulating differences in patterns of finiteness marking in Dutch, German and Spanish

• Children modelled:
  – Peter - Gronigen Dutch corpus (Bols, 1995)
  – Leo - MPI German corpus (Behrens, in press)
  – Juan - Nottingham Spanish corpus (Aguado-Orea, 2004)
Pattern of finiteness marking as a function of MLU for Peter and MOSAIC-Peter (Dutch)

Data for Peter

Model of Peter

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

Data for Leo

Model of Leo

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)
**OI errors as a function of compound finites in the input and percentage of utterance final verbs in the input that were finite vs. non-finite**

<table>
<thead>
<tr>
<th>Language</th>
<th>OI errors at lowest MLU point (%)</th>
<th>Compound Finites in Input (%)</th>
<th>Utterance-final finite verbs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td>61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**OI errors as a function of compound finites in the input and percentage of utterance final verbs in the input that were finite vs. non-finite**

<table>
<thead>
<tr>
<th>Language</th>
<th>OI errors at lowest MLU point (%)</th>
<th>Compound Finites in Input (%)</th>
<th>Utterance-final finite verbs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>75</td>
<td>31</td>
<td>61</td>
</tr>
<tr>
<td>German</td>
<td>61</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>18</td>
<td>25</td>
<td></td>
</tr>
</tbody>
</table>
OI errors as a function of compound finites in the input and percentage of utterance final verbs in the input that were finite vs. non-finite

<table>
<thead>
<tr>
<th></th>
<th>OI errors at lowest MLU point (%)</th>
<th>Compound Finites in Input (%)</th>
<th>Utterance-final finite verbs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dutch</td>
<td>75</td>
<td>31</td>
<td>18</td>
</tr>
<tr>
<td>German</td>
<td>61</td>
<td>22</td>
<td>35</td>
</tr>
<tr>
<td>Spanish</td>
<td>18</td>
<td>25</td>
<td>74</td>
</tr>
</tbody>
</table>
Learning language in different cultures
Some claims made about language learning

• There are cultures in which children are not spoken to before they speak
  à Children only require minimal input to learn language
  OR
  à Children can learn language through overhearing

• There are cultures which believe children have to be taught language and corrected from ‘babytalk’
  à Children can learn language from a highly didactic interactive style
Ideologies of childhood

• Status in Samoa (Ochs)

• Children learn independently (Brice Heath)

• Children need protection (Pye)

• Children have to be taught (Schieffelin)
What do children need from their input?

- Children have to learn form-meaning mappings from what they hear
- They have to learn the distributional information from the input

Either
- Children need minimal amounts of this à triggering parameter setting

or
- Children are getting this information though not necessarily in the same way as children in advanced technological societies
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
Cross-cultural studies of what children hear

What is the nature of preverbal communication?

How much speech is addressed to children?

Chintang Puma Documentation Project
Bickel et al.
<table>
<thead>
<tr>
<th></th>
<th>‘BABIES’</th>
<th>‘TWO’-S</th>
<th>‘THREE’-S</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-3 hours per cycle</td>
<td>3-4 hours per cycle</td>
<td>3-4 hours per cycle</td>
</tr>
<tr>
<td></td>
<td>6m 8m 10m 12m 15m 18m 21m 24m</td>
<td>2;2 – 3;2 3;4 – 3;8</td>
<td>3;2 – 4;2 4;4 – 4;8</td>
</tr>
<tr>
<td>Dipkala Saphal</td>
<td>X X X X X X X X</td>
<td>Monthly</td>
<td>Bi-monthly</td>
</tr>
<tr>
<td></td>
<td>X X X X X X X X</td>
<td>Monthly</td>
<td>Bi-monthly</td>
</tr>
<tr>
<td>Khem Kamala</td>
<td></td>
<td>Bi-monthly</td>
<td>Bi-monthly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kalpana Man Kumar</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Data collection

<table>
<thead>
<tr>
<th>‘BABIES’</th>
<th>6m</th>
<th>18m</th>
<th>21m</th>
<th>24m</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-3 hours per cycle</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dipkala</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Saphal</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>‘TWO’-S</th>
<th>2;2 – 3;2</th>
<th>3;4 – 3;8</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-4 hours per cycle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khem</td>
<td>Monthly</td>
<td>Bi-monthly</td>
</tr>
<tr>
<td>Kamala</td>
<td>Monthly</td>
<td>Bi-monthly</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>‘THREE’-S</th>
<th>3;2 – 4;2</th>
<th>4;4 – 4;8</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-4 hours per cycle</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kalpana</td>
<td>Monthly</td>
<td>Bi-monthly</td>
</tr>
<tr>
<td>Man Kumar</td>
<td>Monthly</td>
<td>Bi-monthly</td>
</tr>
</tbody>
</table>
What to compare with?

The Rigol corpus

<table>
<thead>
<tr>
<th>‘Babies’:</th>
<th>Johanna</th>
<th>Lars</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘Two’-s</td>
<td>Pauline</td>
<td>Sebastian</td>
</tr>
<tr>
<td>‘Three’-s</td>
<td>Corinna</td>
<td>Niklas</td>
</tr>
</tbody>
</table>
Saphal: 0;7
Man Kumar 3;0+

Ia Kuluke tusande
Look, Kuluke, digged!
<table>
<thead>
<tr>
<th>Categories for characterising the communicative environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportions per hour</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>Minutes with utterances</td>
</tr>
<tr>
<td>Pointing</td>
</tr>
<tr>
<td>Offering</td>
</tr>
<tr>
<td>Imitation</td>
</tr>
<tr>
<td>Teasing</td>
</tr>
<tr>
<td>Object handling</td>
</tr>
<tr>
<td>Mutual gaze</td>
</tr>
<tr>
<td>Attention getting</td>
</tr>
<tr>
<td>Showing</td>
</tr>
<tr>
<td>Affection</td>
</tr>
<tr>
<td>Playing</td>
</tr>
</tbody>
</table>
Utterances addressed to the child

In these recordings:

• The number of minutes with at least one utterance were roughly equivalent across the two cultures

• Chintang children were hearing more language from other children and adults
Other communicative interactions:

- Pointing: Chintang children pointed later despite receiving more pointing interactions

- Imitation: Was established by 2;2 and low by 2;10 but individual differences were the most evident

- Offering: Seemed similar across cultures, maintained for the Chintang ‘two-s’ by other children

- Teasing: very little but when there by Chintang other children to late babies and ‘two-s’
Interim thoughts

- For babies, the main form of interaction seems to be dyadic, with the mother.
- Interacting with babies seems to afford the same types of interactions in both cultures.
- For Chintang children, the part played by other adults and children is always greater.
- We cannot assess the volume of talk to the baby from these results, but they are certainly being talked to.
- We need more fine-grained analyses to assess the culturally-specific content of these interactions.
- At least on these measures, individual differences can outweigh cultural differences.
# Comparing recording situations

<table>
<thead>
<tr>
<th>Our study</th>
<th>Most previous studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Mostly outside</td>
<td>• Inside the house</td>
</tr>
<tr>
<td>• Many different situations</td>
<td>• Mother and child playing</td>
</tr>
<tr>
<td>• Mother often absent</td>
<td>• Only mother present</td>
</tr>
<tr>
<td>• Many other children</td>
<td>• No other children</td>
</tr>
</tbody>
</table>
The end

Thank you!