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

Lecture 4

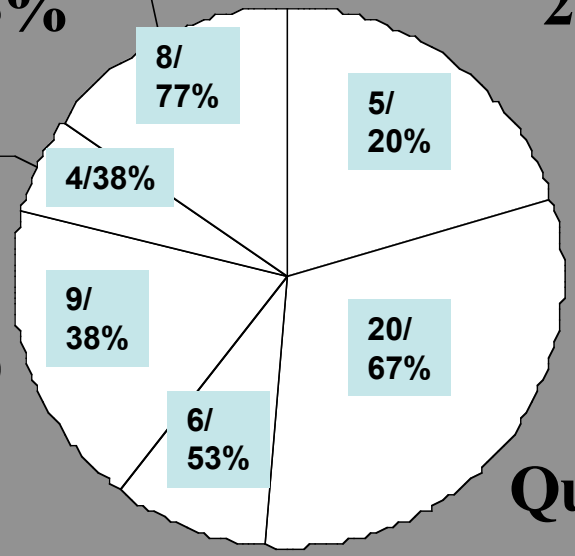
Outline

- 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



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 …

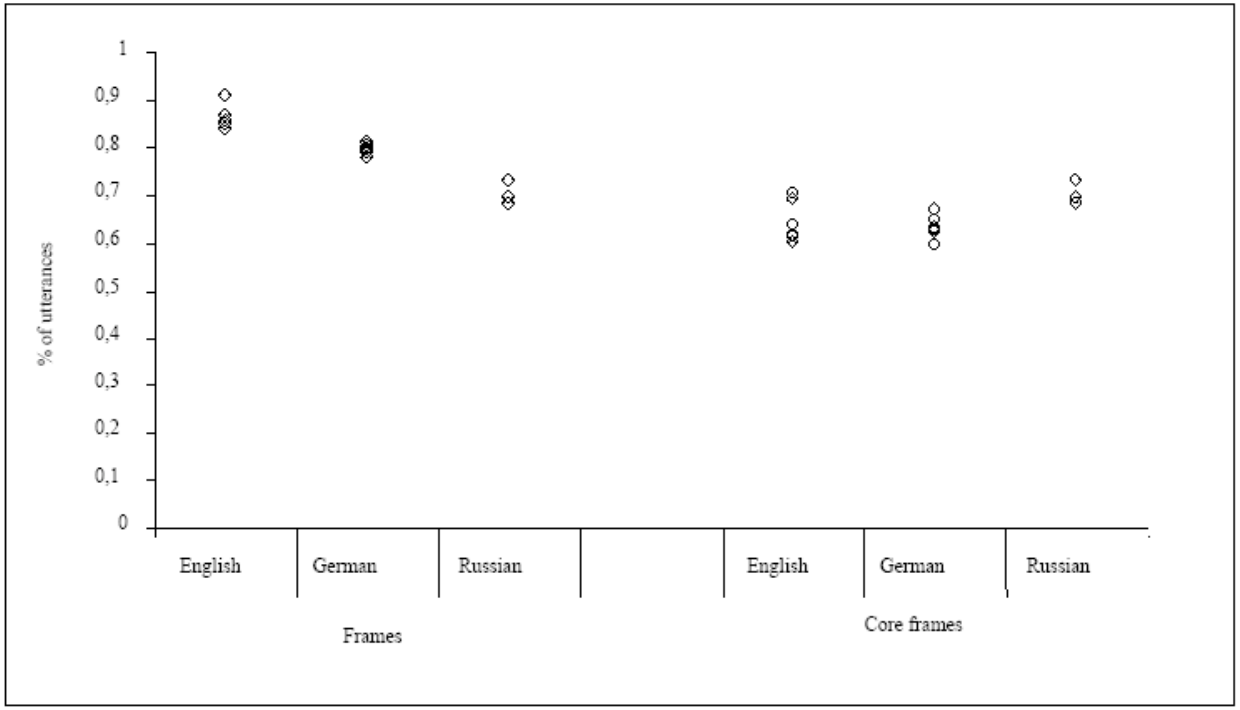
What counted as a ‘frame’?

Example utterances:

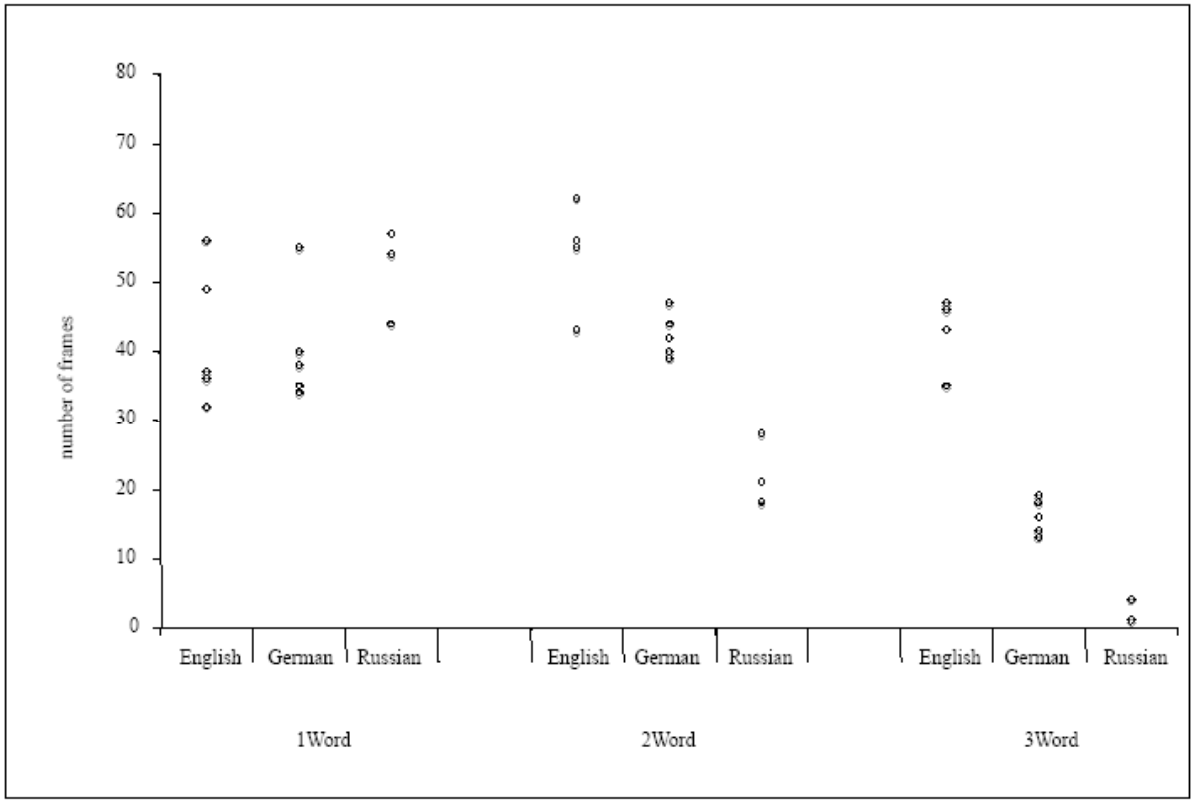
- **That’s a** dog
- **That’s a** girl
- **That’s a** flower
- **That’s** your pen
- **That’s a** lorry

- FRAME = That’s a …

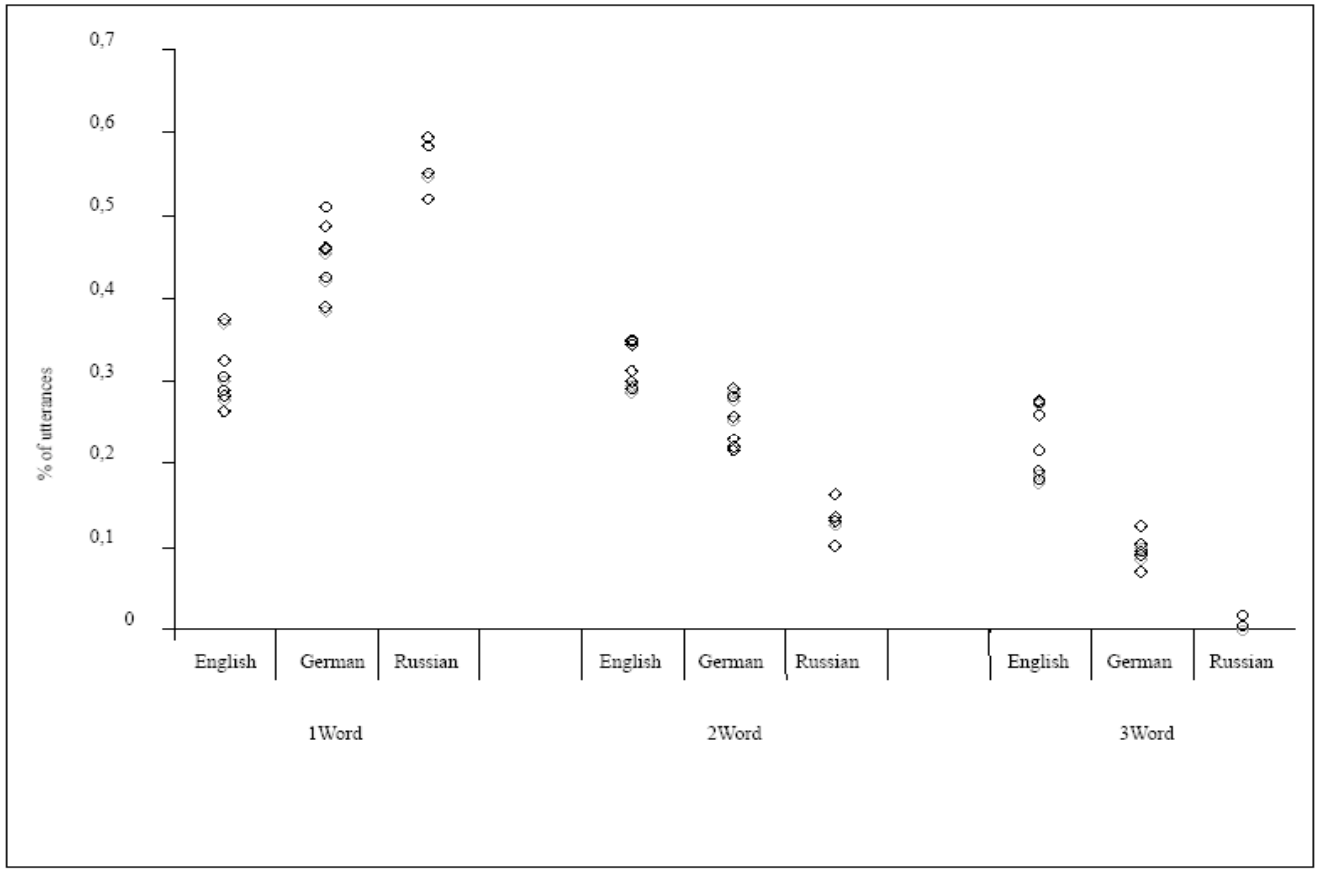
Percentage of utterances by individual mothers accounted for by frames and core frames



Number of one-, two- and three-word frames for individual mothers



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

Rowland & Pine, 2000, Rowland (2007)
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

High frequency words		Low frequency words	
Frames	Non-frames	Frames	Non-frames
2.05	13.09	11.27	11.71

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

WITH	**he goes**
WITHOUT	**he go/he going**

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]

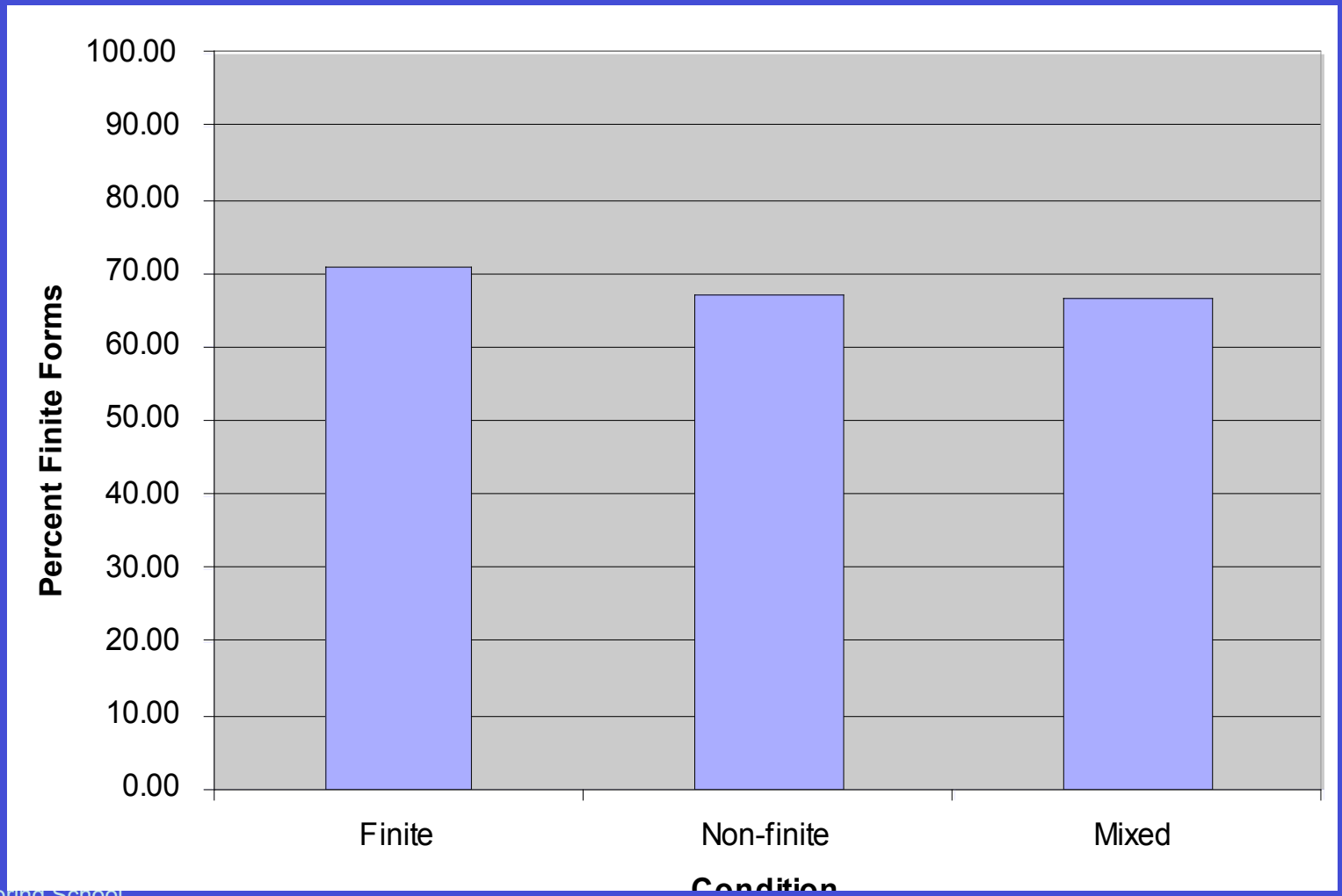
Theakston, Lieven & Tomsasello, 2003

Test questions: to elicit the use of verbs in FINITE contexts

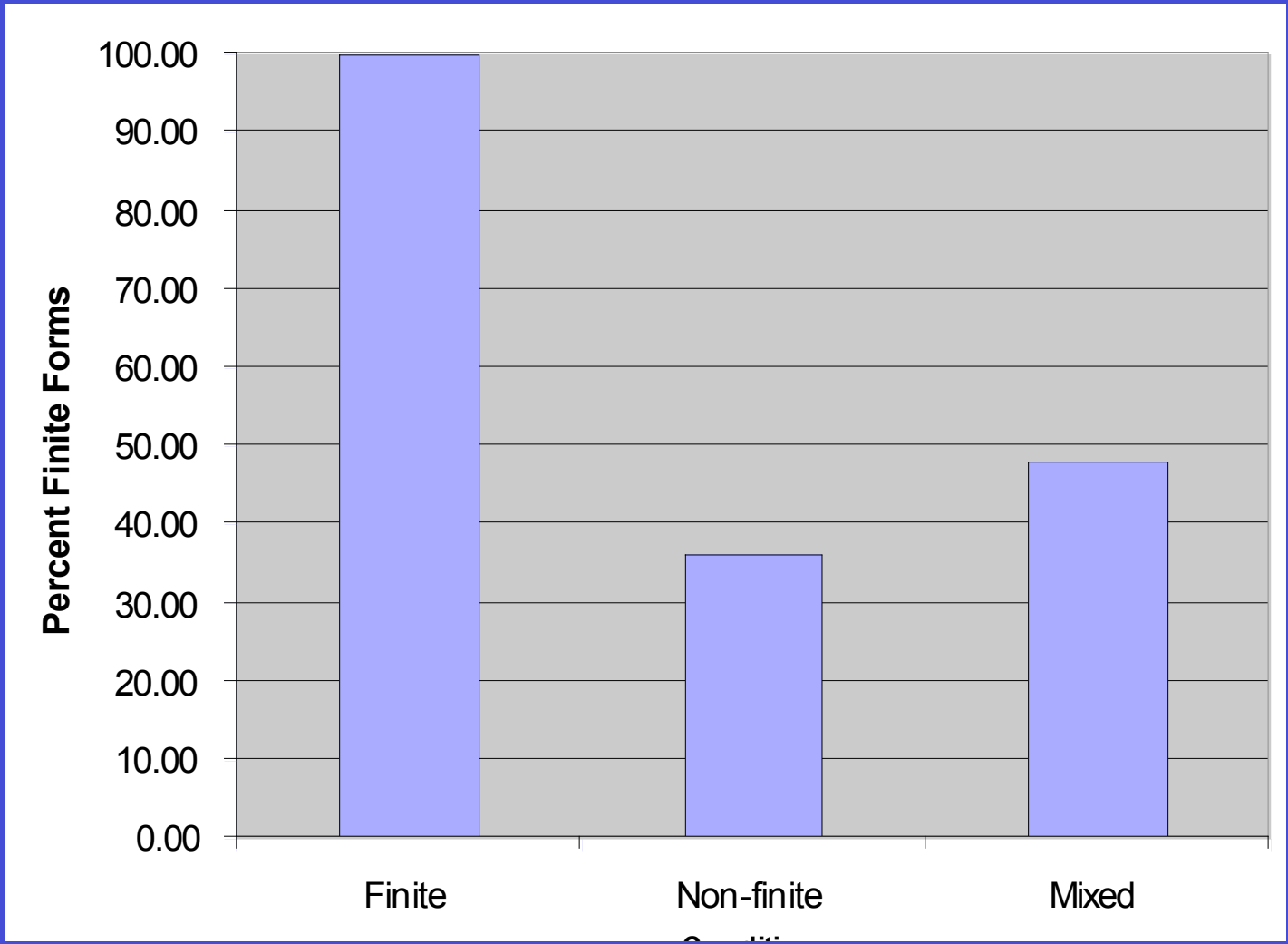
- **What does this one do?**
- **What does it do?**
- **It _____[s]?**

Participants:	24 children, mean age 2;8
Conditions:	Finite, Non-finite,Mixed between subjects

% Finite verb forms produced with known verbs



% 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

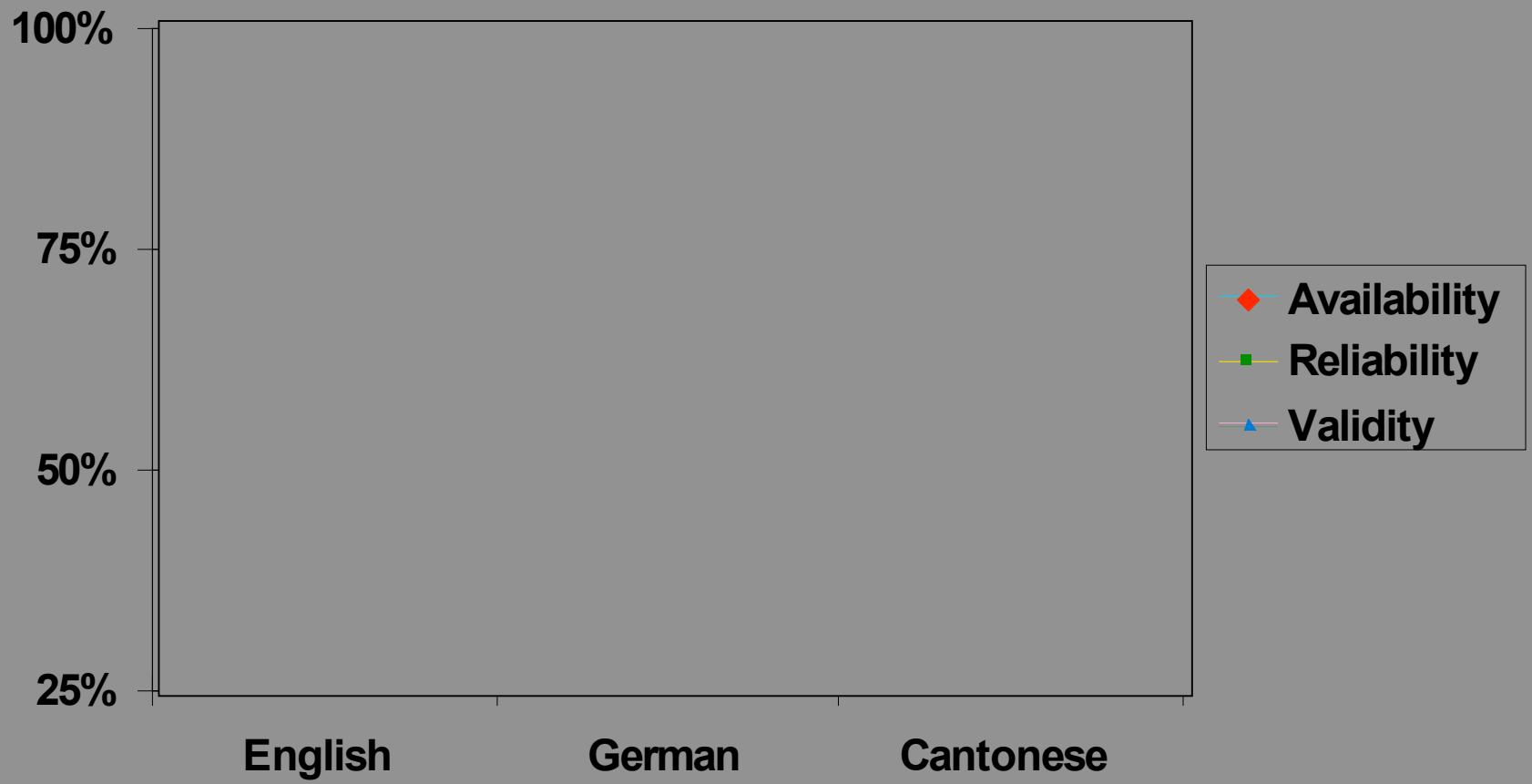
- chases

- Cue Reliability

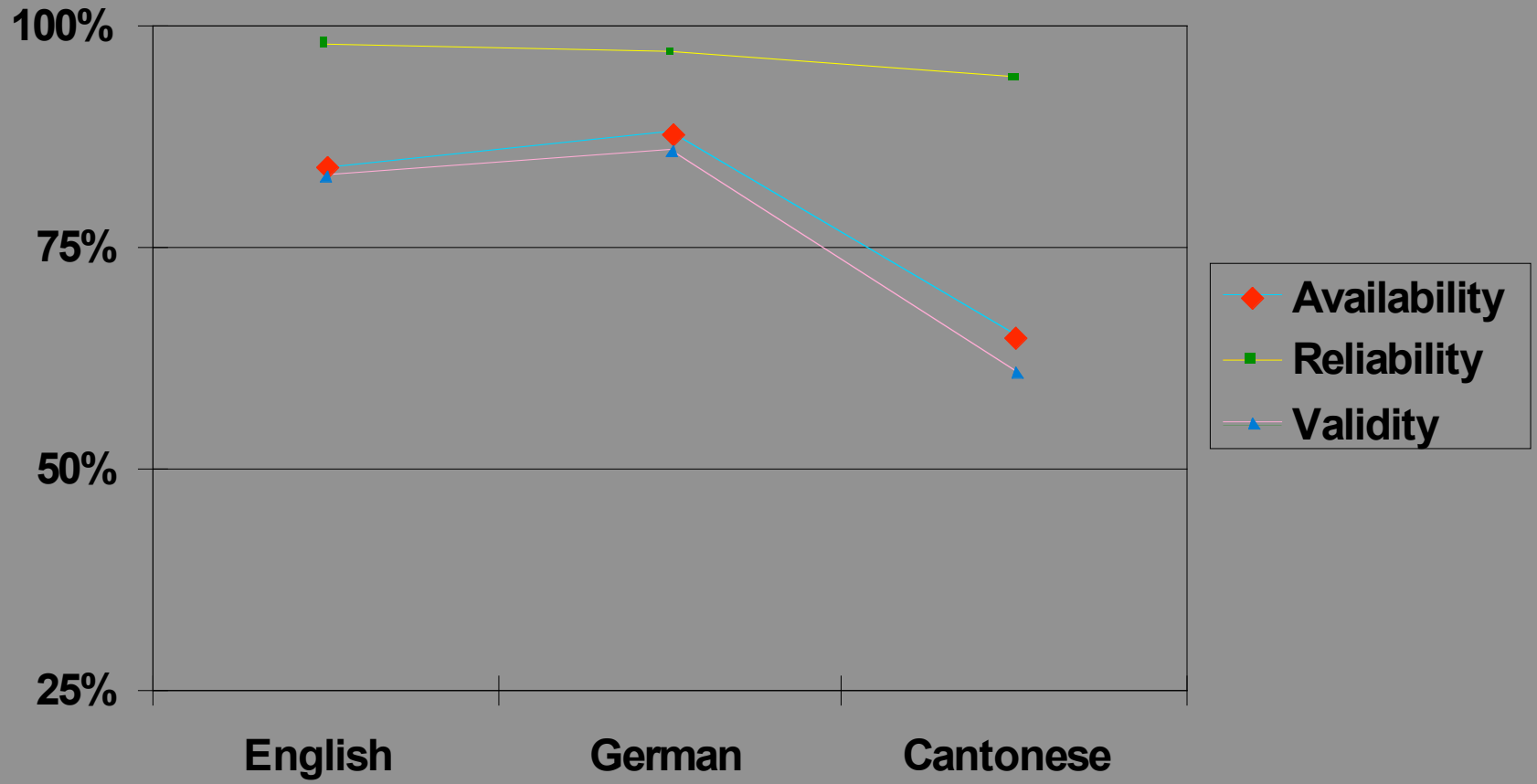
+ The man opens the door

- Den_{ACC} Hund schubst der_{NOM} Löwe

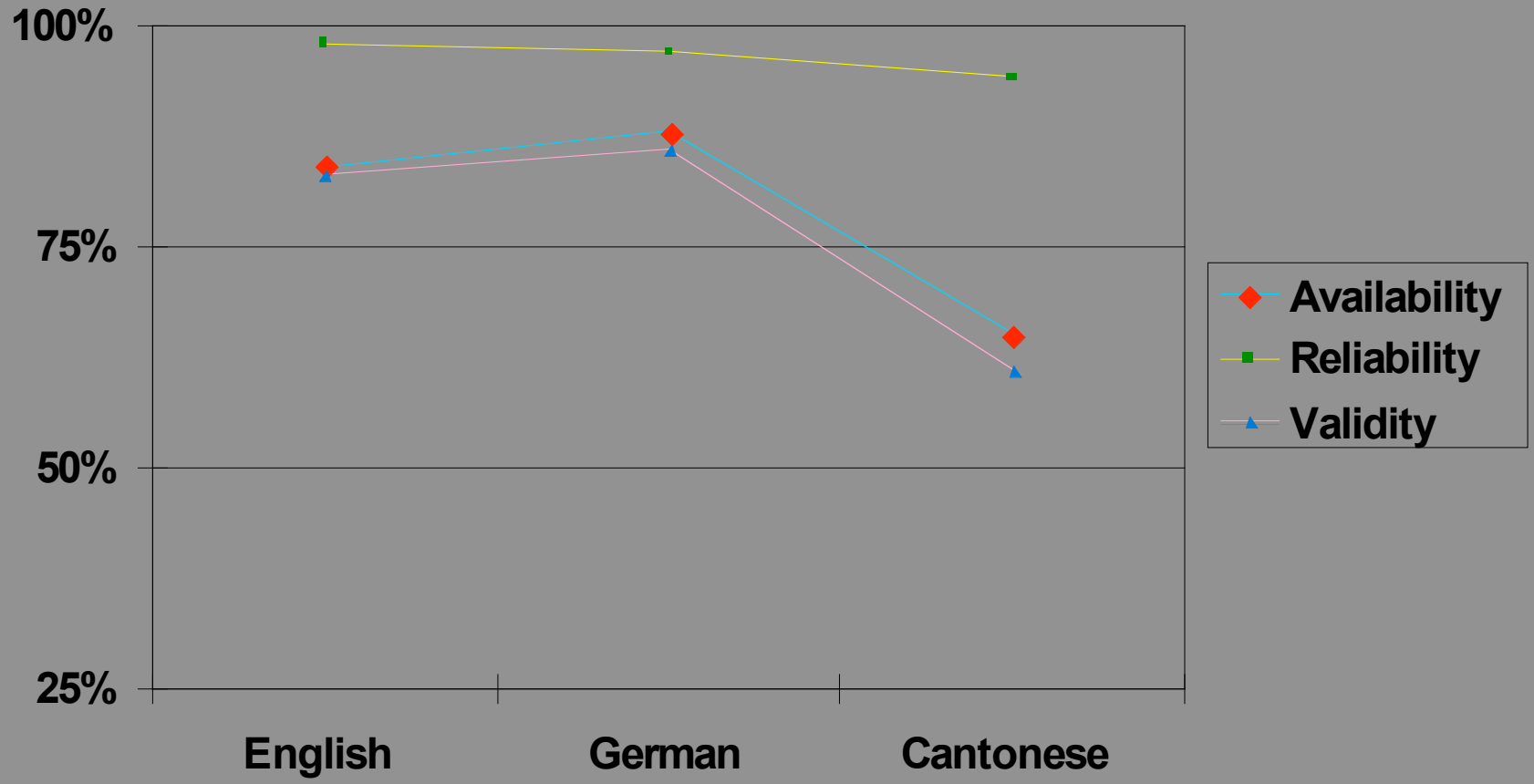
The animacy contrast cue



The animacy contrast cue

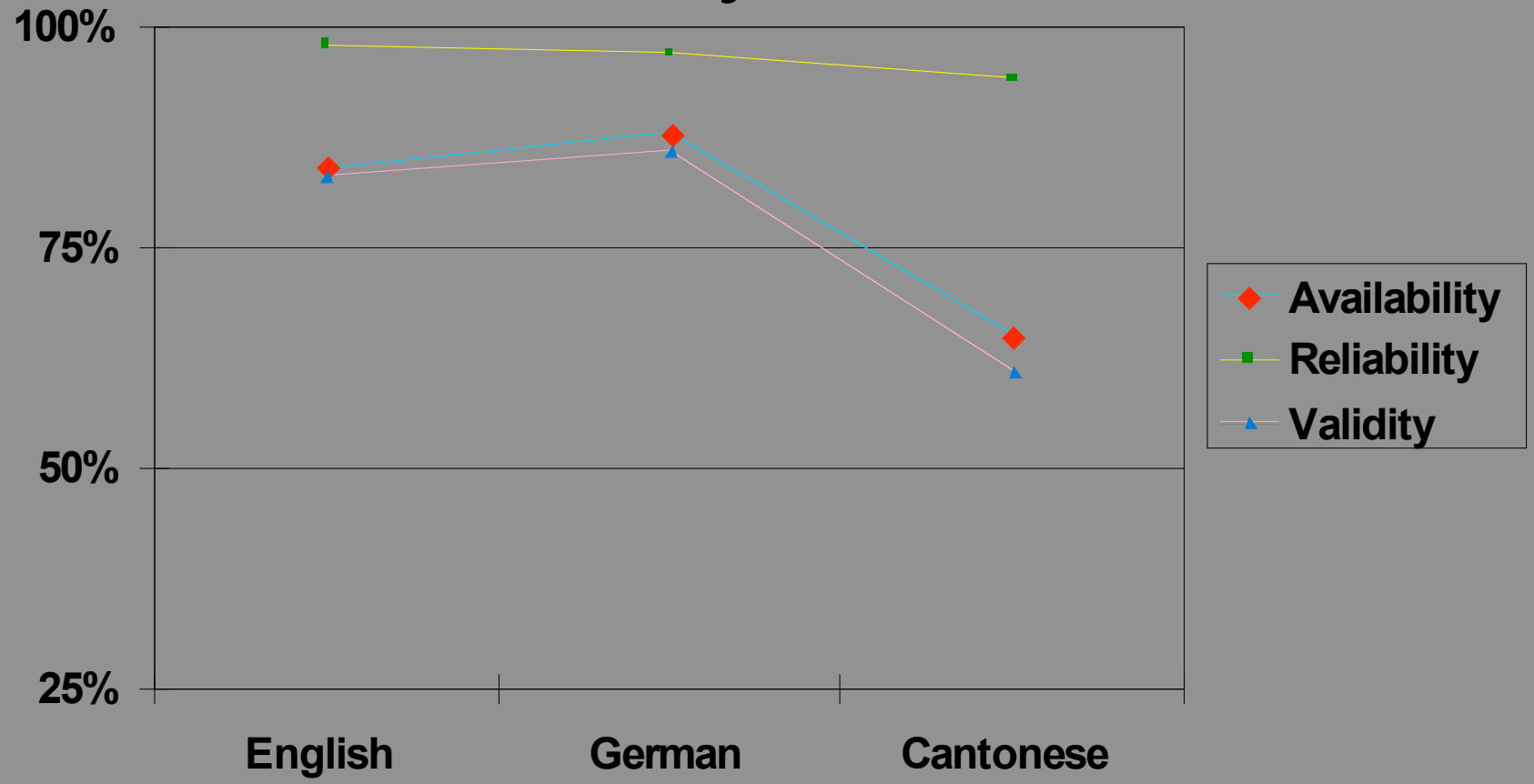


The animacy contrast cue



- highly reliable across languages

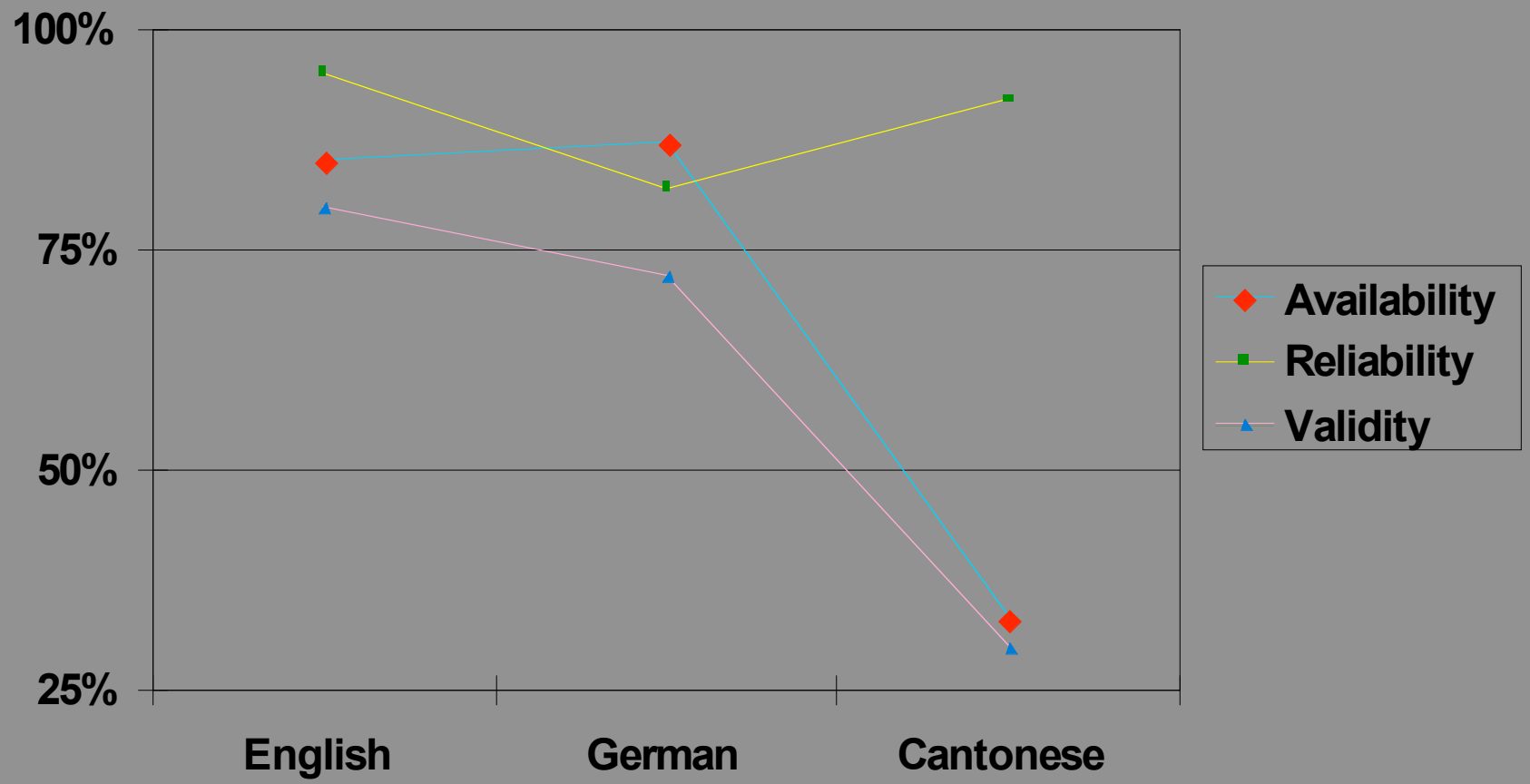
The animacy contrast cue



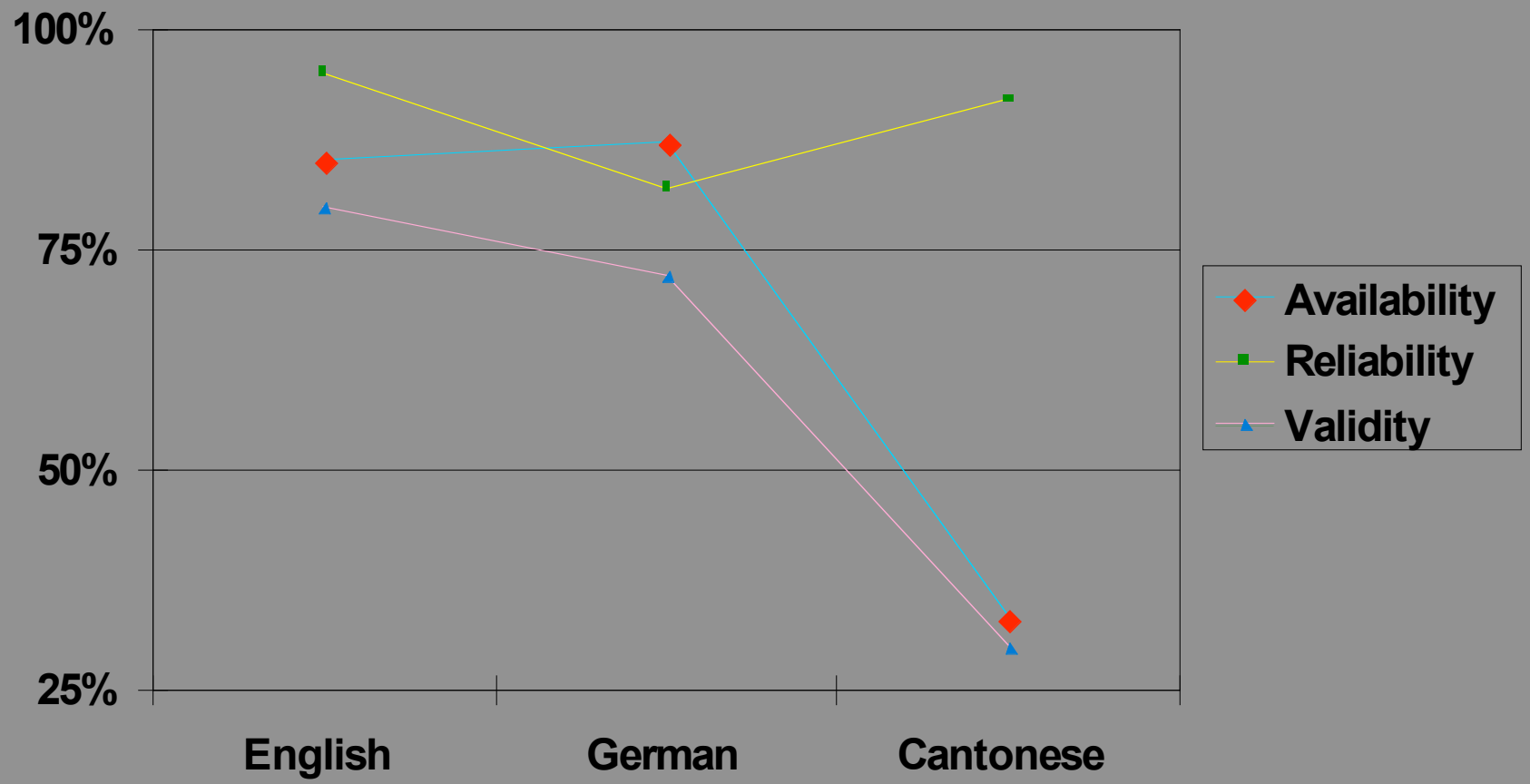
- highly reliable across languages
- availability is lower in Cantonese due to massive ellipsis and ambiguous pnouns

The word order cue

The word order cue



The word order cue



cue validity: English > German > Cantonese

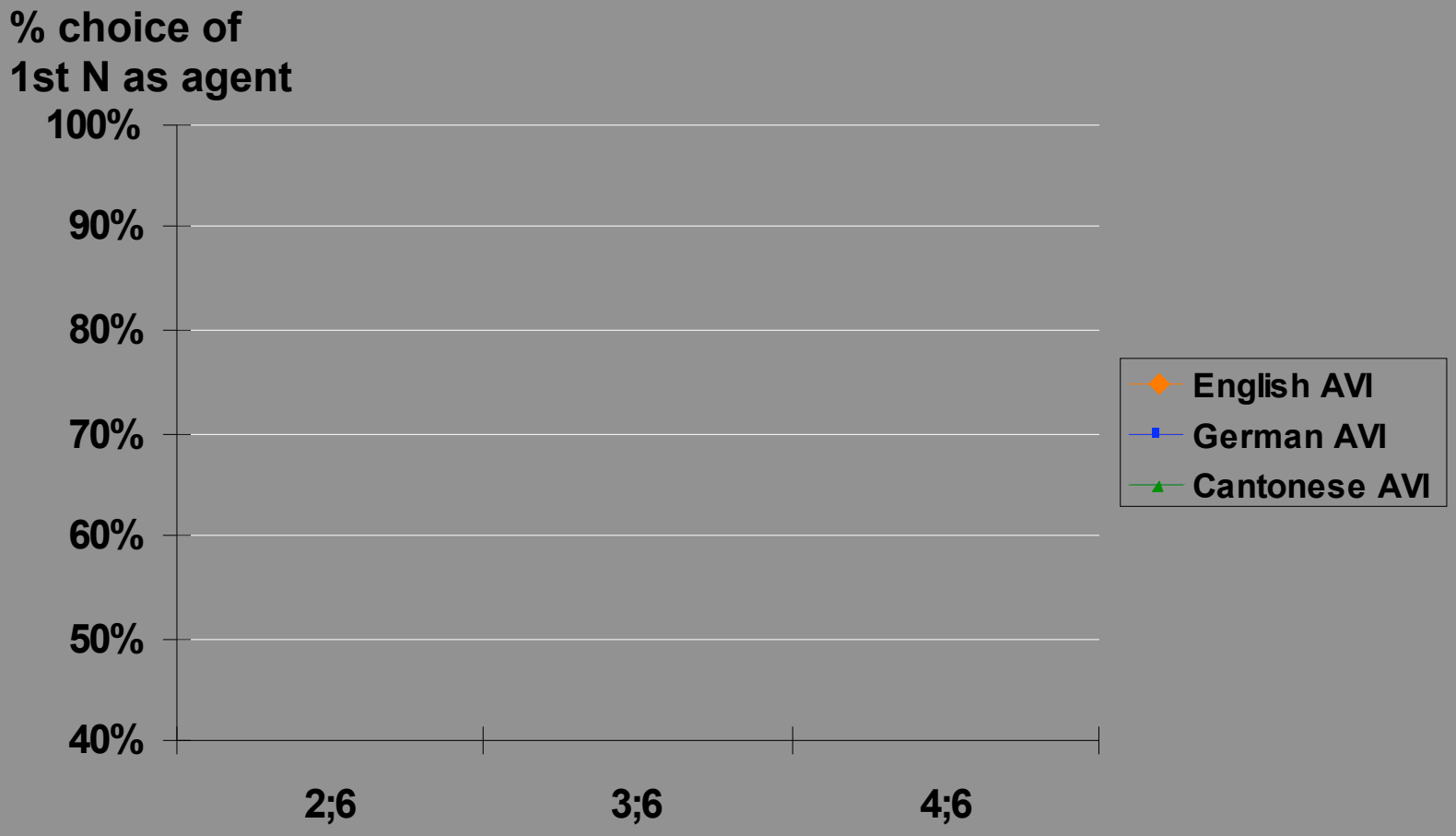
Developmental Findings

(i) Animate Noun - Verb - Inanimate Noun **(AVI)**

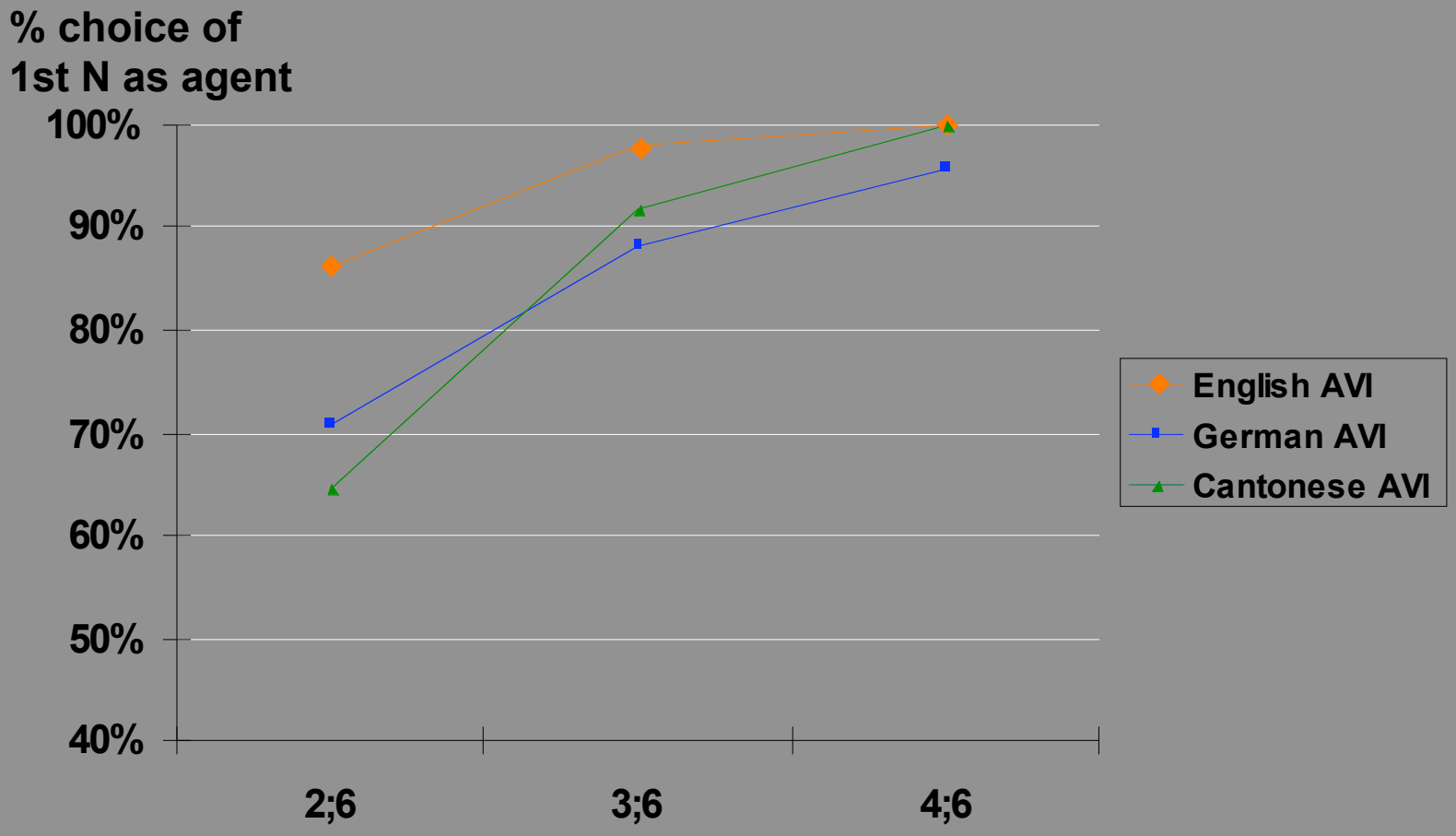
(ii) Inanimate Noun - Verb - Animate Noun **(IVA)**

(iii) Animate Noun - Verb - Animate Noun **(AVA)**

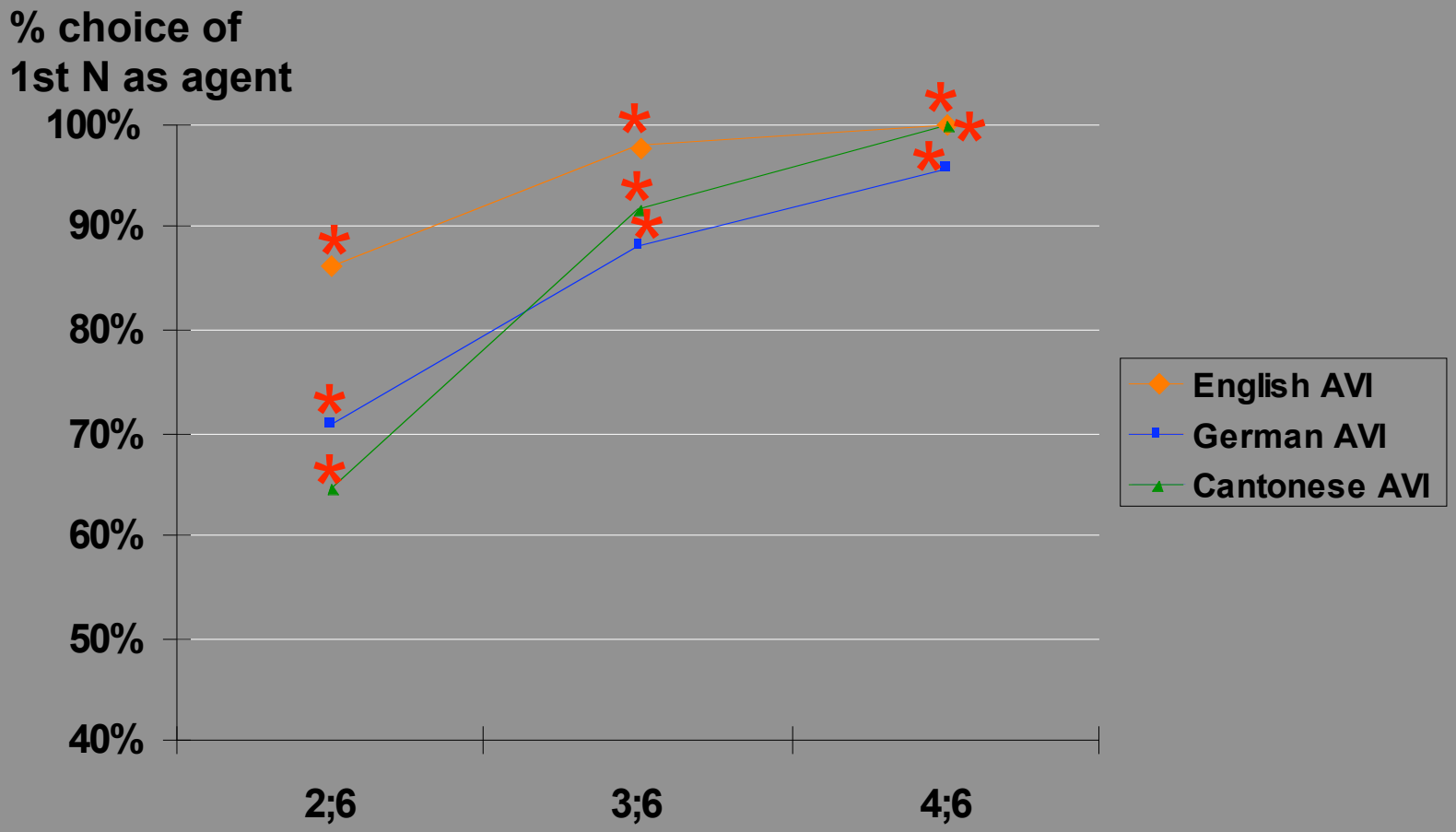
AVI: *The horse tams the telephone*



AVI: *The horse tams the telephone*

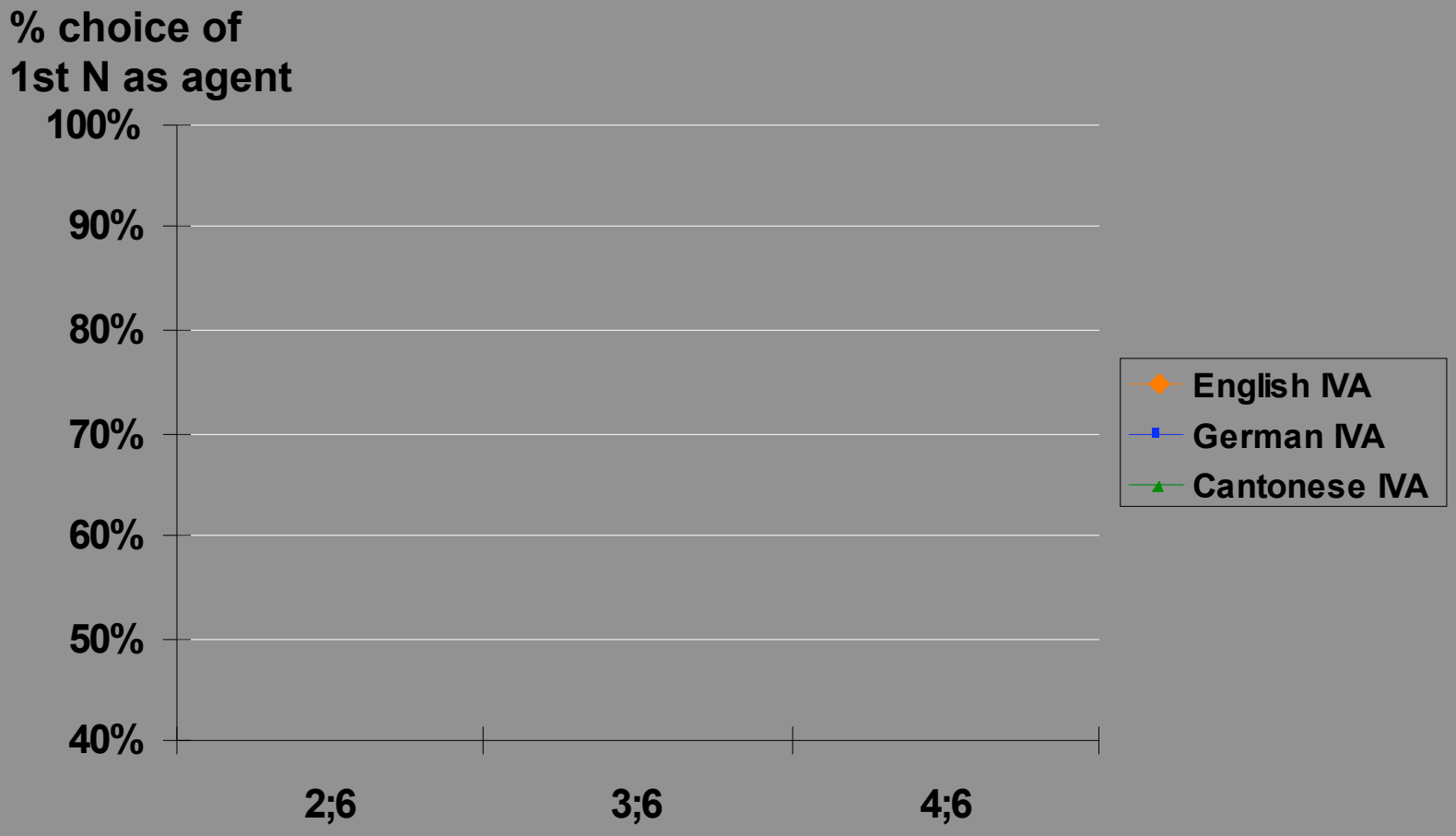


AVI: *The horse tams the telephone*

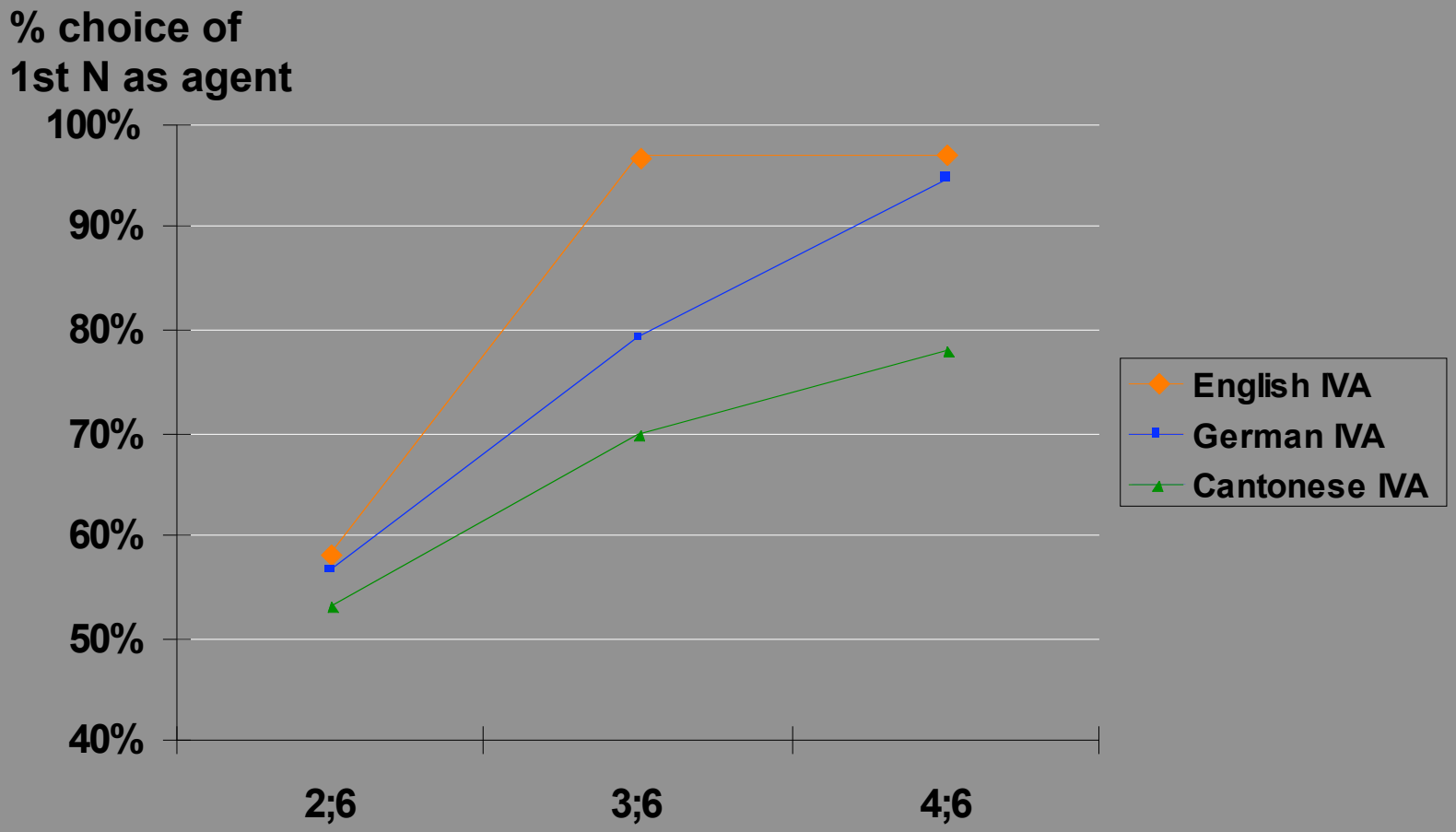


Across language groups, even the youngest 2-year-olds were above chance in choosing the 1st Animate Noun as the agent

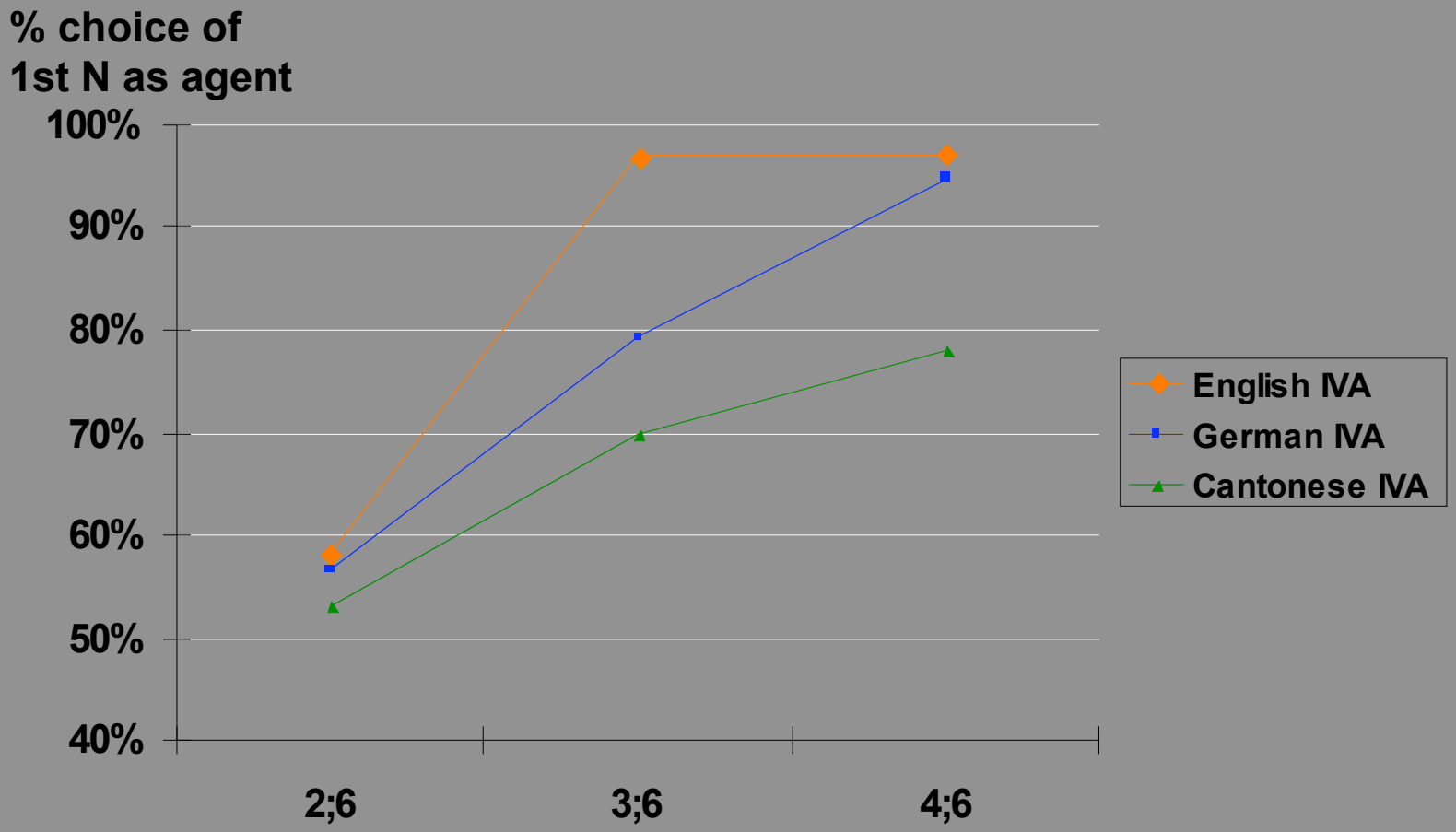
IVA: *The present meeks the chicken*



IVA: *The present meeks the chicken*

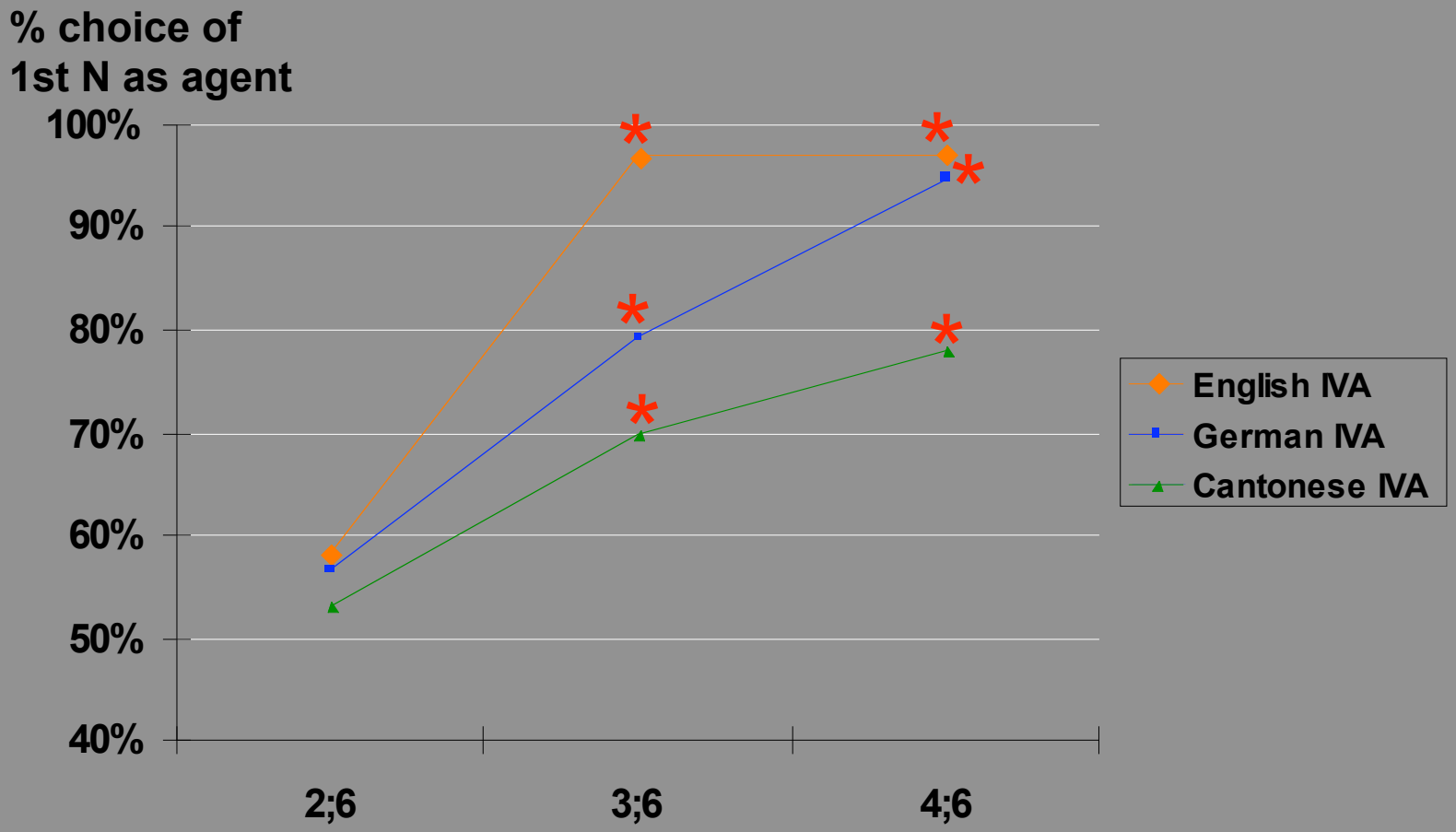


IVA: *The present meeks the chicken*



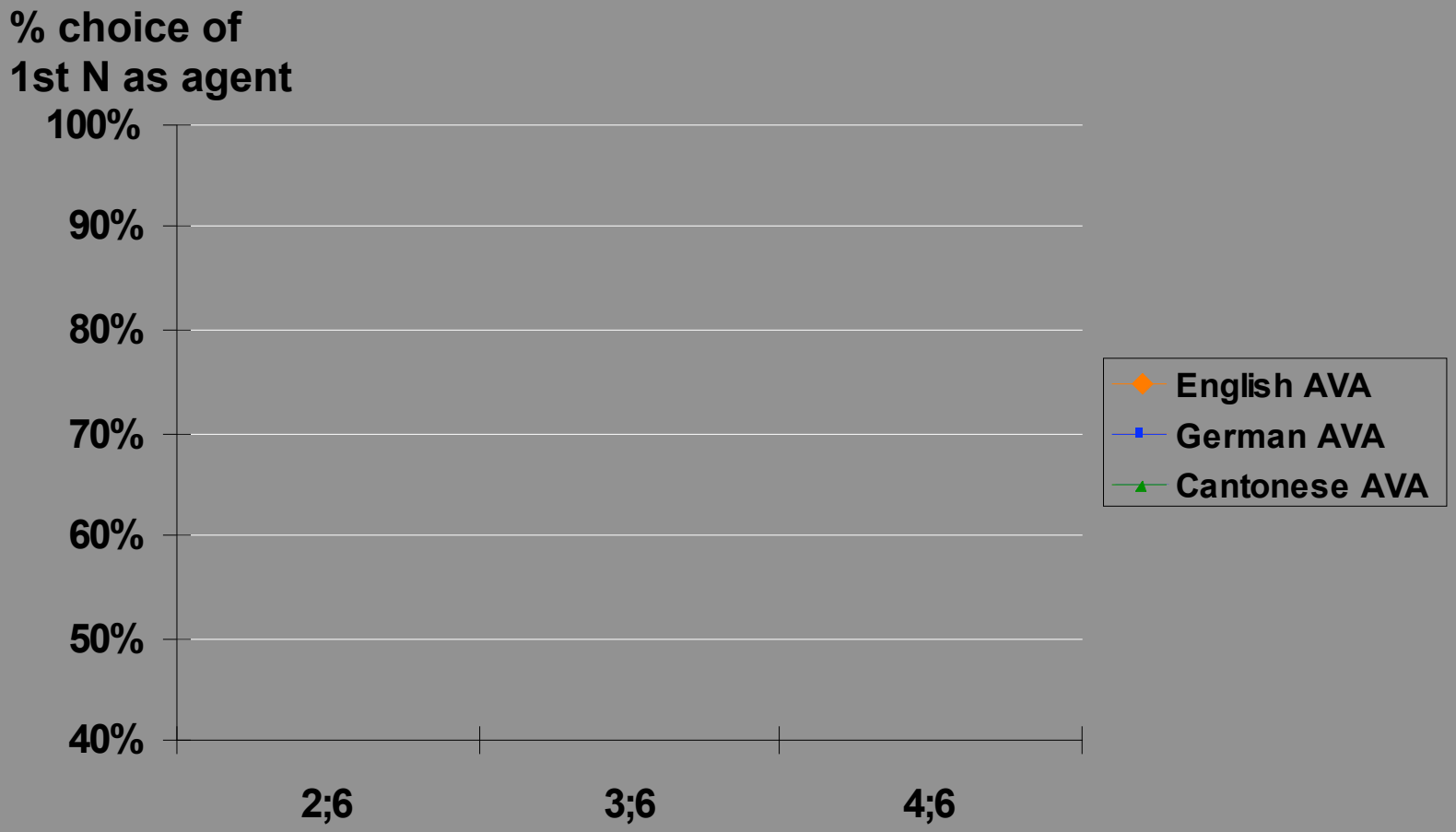
- Across language groups, 2-year-olds were at chance group performance

IVA: *The present meeks the chicken*

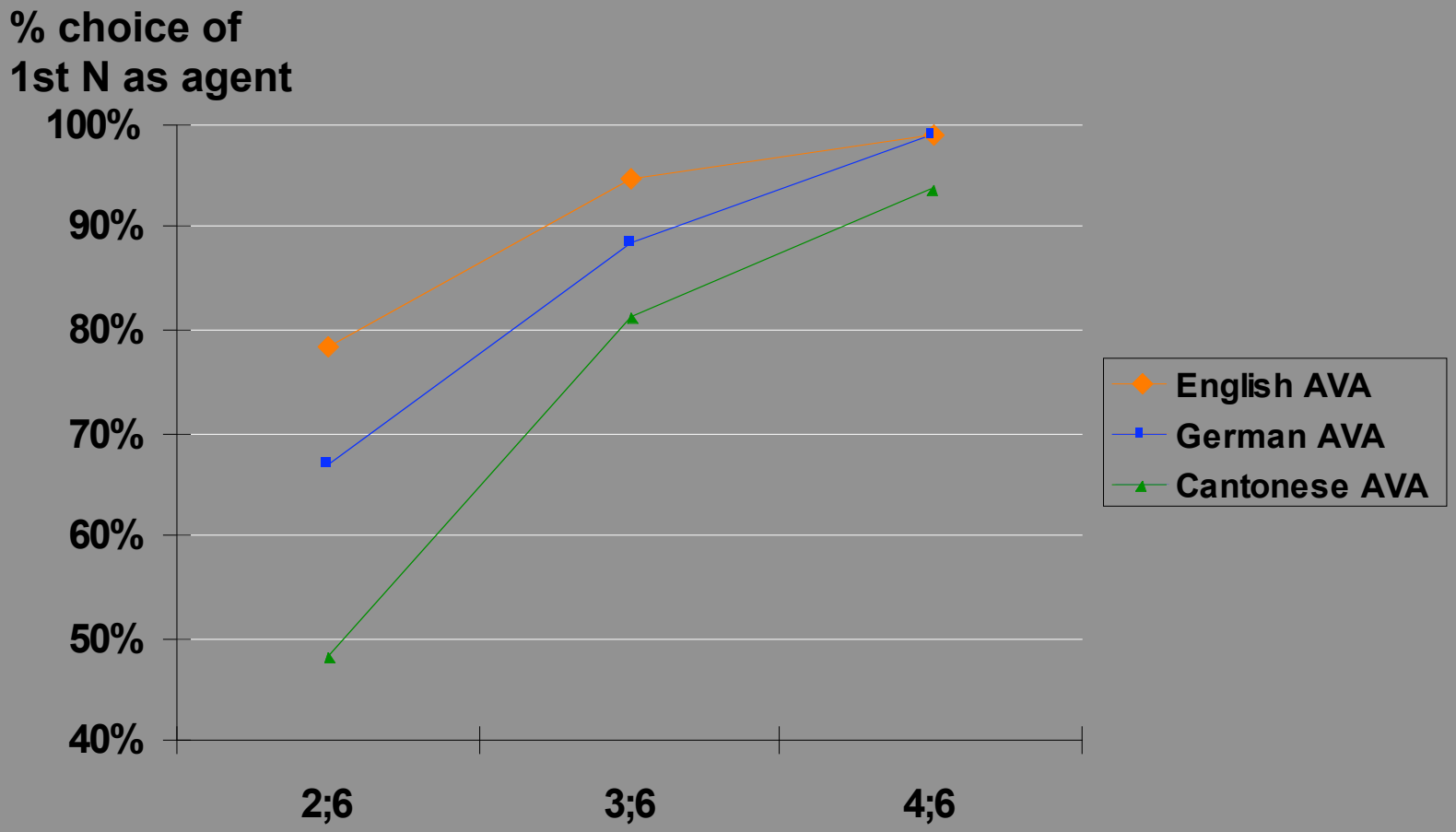


- Across language groups, 2-year-olds were at chance group performance
- Older children at 3;6 and 4;6 preferred word order over animacy

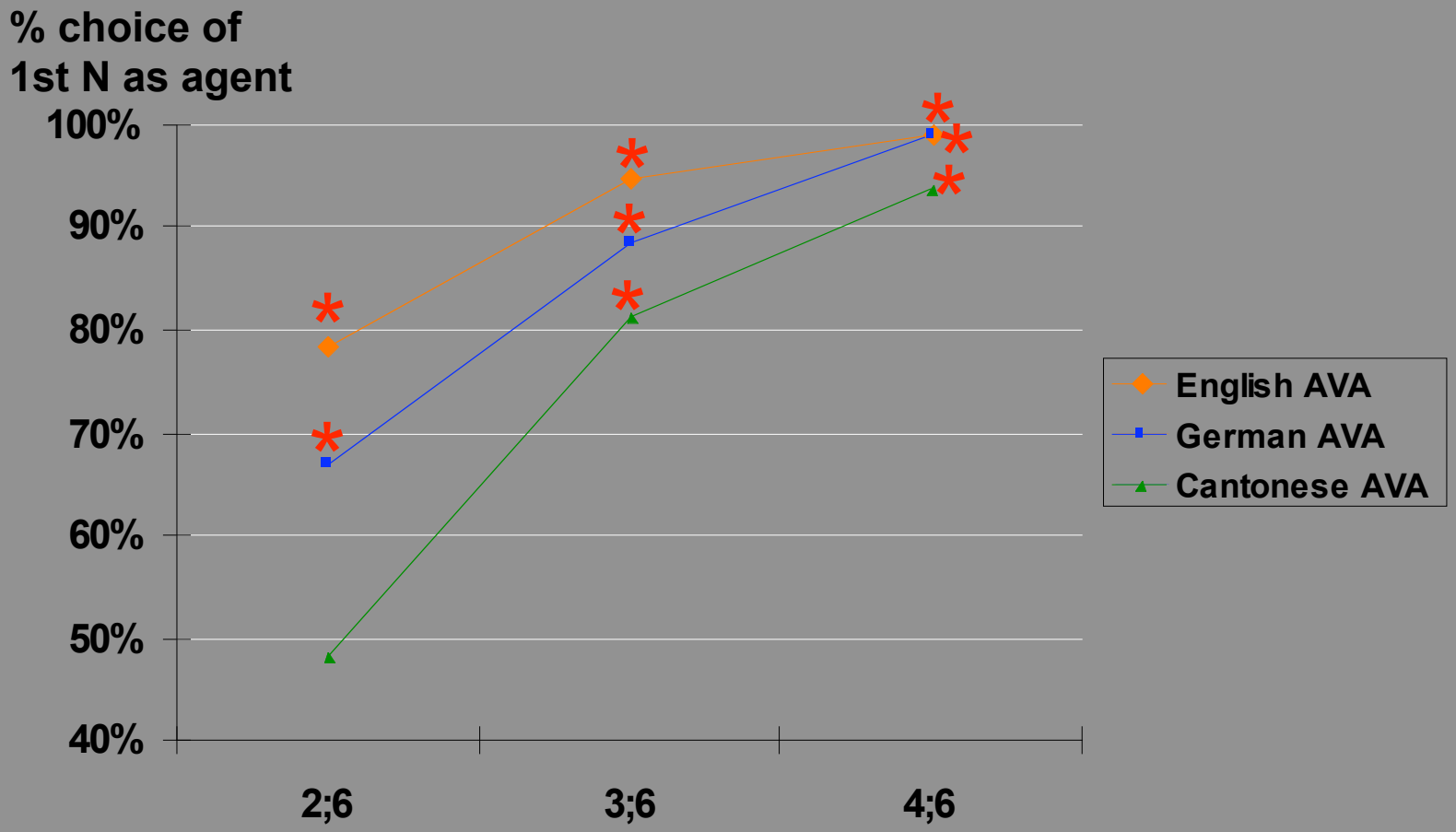
AVA: *The cow tams the giraffe*



AVA: *The cow tams the giraffe*



AVA: *The cow tams the giraffe*



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

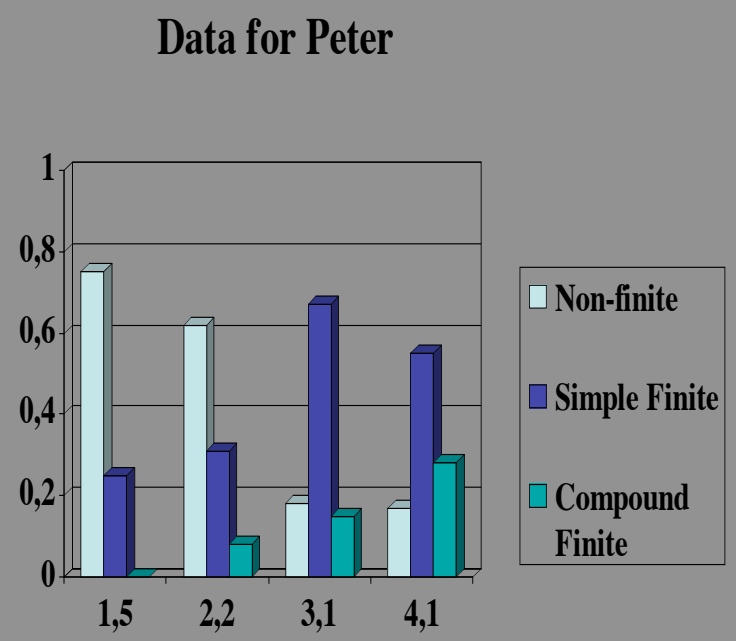
MOSAIC: Key Features

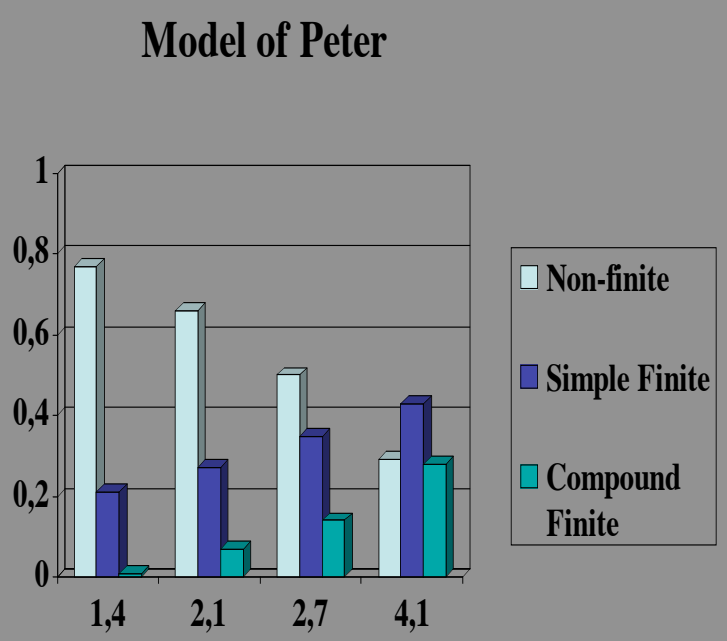
- Takes as input (orthographically transcribed) samples of Child-Directed Speech
- Produces output in the form of 'utterances' that can be compared with those of real children
- Learns to produce progressively longer utterances as a function of the amount of input it has seen

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)

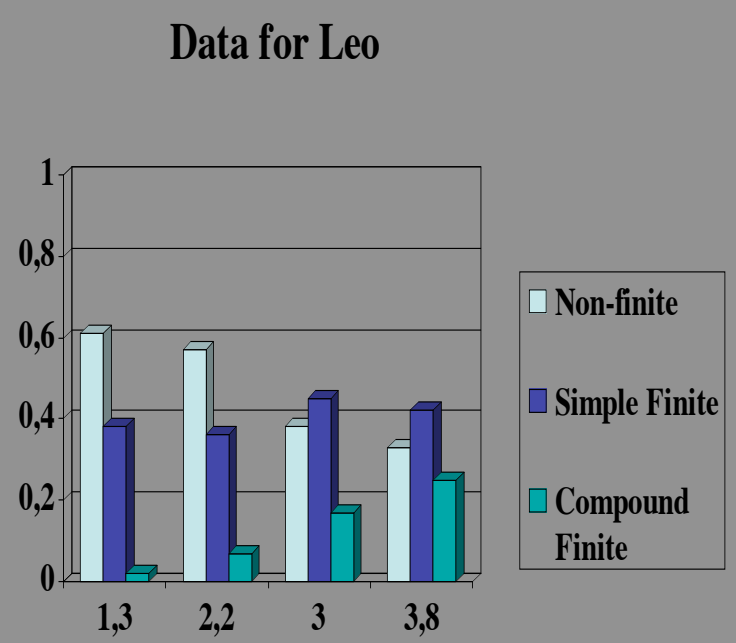
Pattern of finiteness marking as a function of MLU for Peter and MOSAIC-Peter (Dutch)

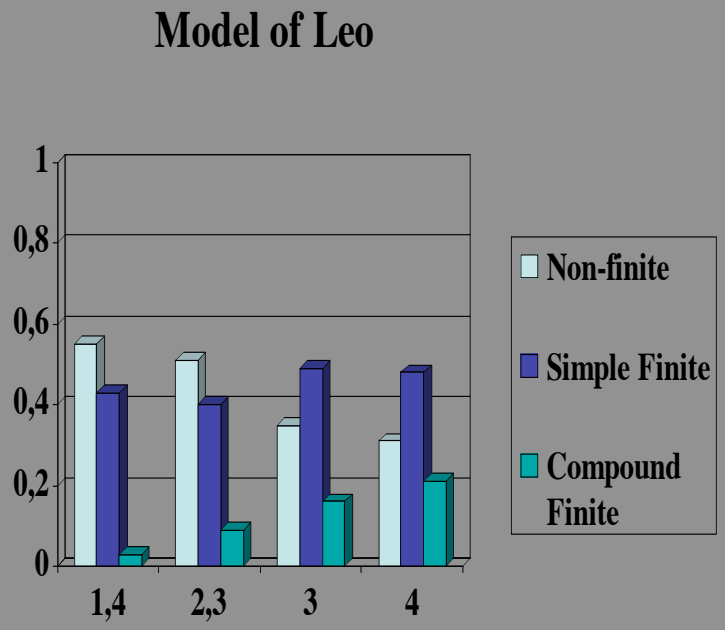




MOSAIC simulates high proportion of OI errors in Dutch (and low proportion of compound finites)

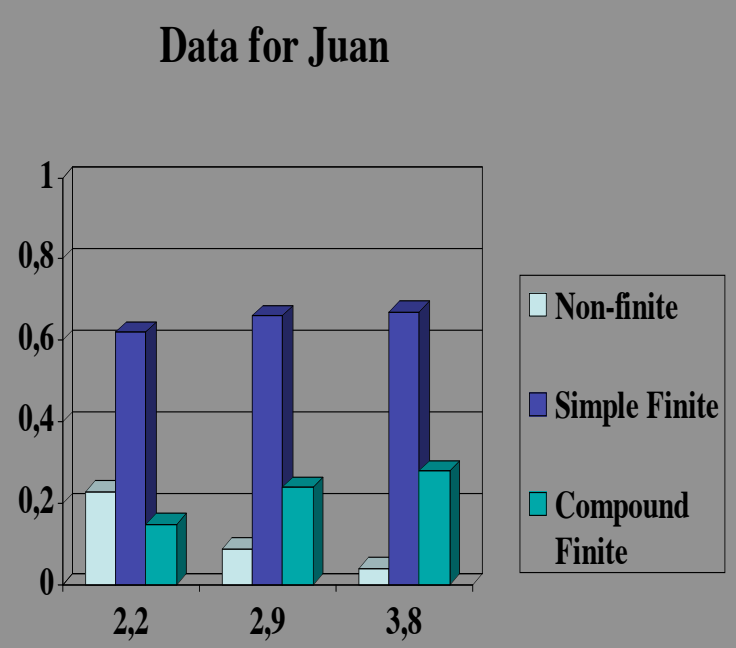
Pattern of finiteness marking as a function of MLU for Leo and MOSAIC-Leo (German)

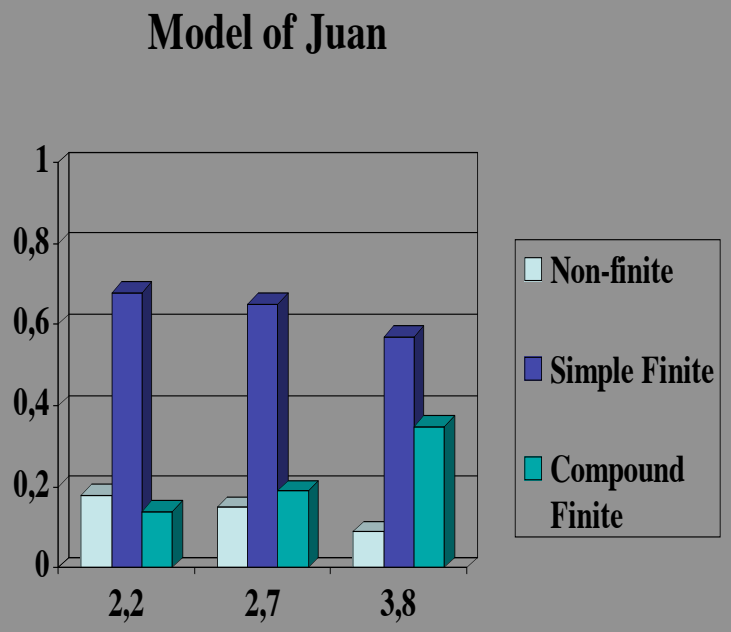




MOSAIC simulates the moderately high proportion of OI errors in German (and low proportion of compound finites)

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

	OI errors at lowest MLU point (%)	Compound Finites in Input (%)	Utterance-final finite verbs (%)
Dutch	75		
German	61		
Spanish	18		

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

	OI errors at lowest MLU point (%)	Compound Finites in Input (%)	Utterance-final finite verbs (%)
Dutch	75	31	
German	61	22	
Spanish	18	25	

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

	OI errors at lowest MLU point (%)	Compound Finites in Input (%)	Utterance-final finite verbs (%)
Dutch	75	31	18
German	61	22	35
Spanish	18	25	74

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.

Data collection

'BABIES' **2-3 hours per cycle**	6m	8m	10m	12m	15m	18m	21m	24m
Dipkala	X	X	X	X	X	X	X	X
Saphal	X	X	X	X	X	X	X	X

'TWO'-S **3-4 hours per cycle**	2;2 – 3;2	3;4 – 3;8
Khem	Monthly	Bi-monthly
Kamala	Monthly	Bi-monthly

'THREE'-S **3-4 hours per cycle**	3;2 – 4;2	4;4 – 4;8
Kalpana	Monthly	Bi-monthly
Man Kumar	Monthly	Bi-monthly

Data collection

'BABIES' **2-3 hours per cycle**	6m					18m	21m	24m
Dipkala	X					X	X	X
Saphal	X					X	X	X

'TWO'-S **3-4 hours per cycle**	2;2 – 3;2	3;4 – 3;8
Khem	Monthly	Bi-monthly
Kamala	Monthly	Bi-monthly

2;2
2;6
2;10

'THREE'-S **3-4 hours per cycle**	3;2 – 4;2	4;4 – 4;8
Kalpana	Monthly	Bi-monthly
Man Kumar	Monthly	Bi-monthly

3;0
3;4
3;9

What to compare with?

The Rigol corpus

‘Babies’:	Johanna	Lars
‘Two’-s	Pauline	Sebastian
‘Three’-s	Corinna	Niklas

Saphal: 0;7


joh000827
00:13:51:00

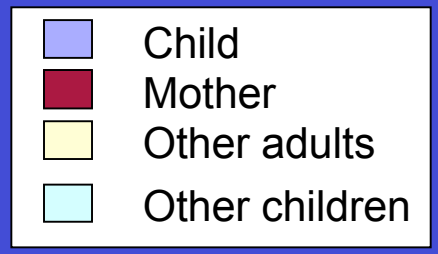
Man Kumar 3;0+


nik030022
00:25:50:00
5 2 1993

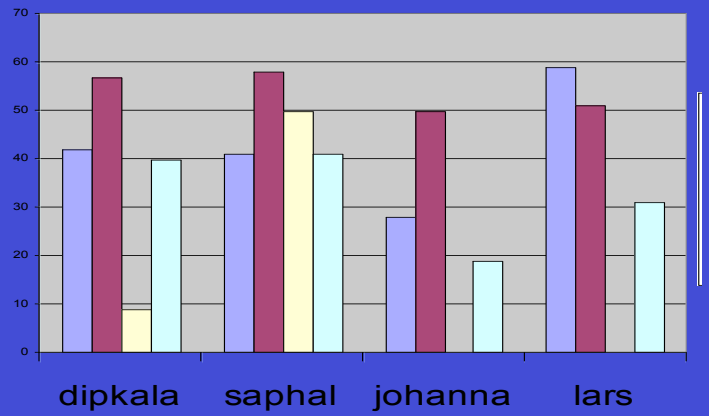
Categories for characterising the communicative environment

Proportions per hour	Child	Mother	Other adults	Other children
Minutes with utterances				
Pointing				
Offering				
Imitation				
Teasing				
Object handling				
Mutual gaze				
Attention getting				
Showing				
Affection				
Playing				

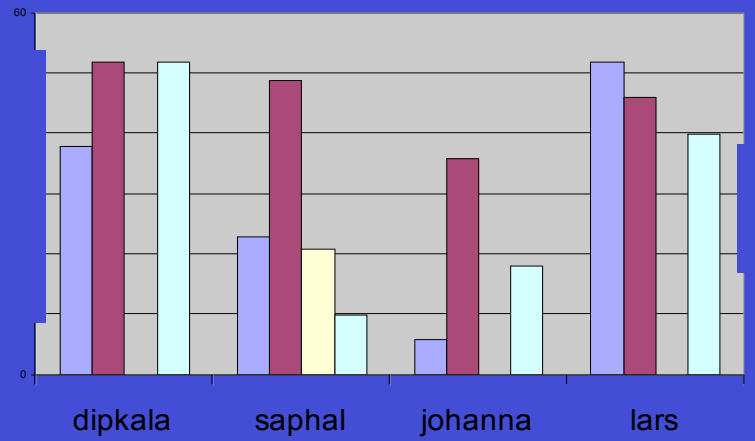
Minutes with utterances/vocalisations: Babies



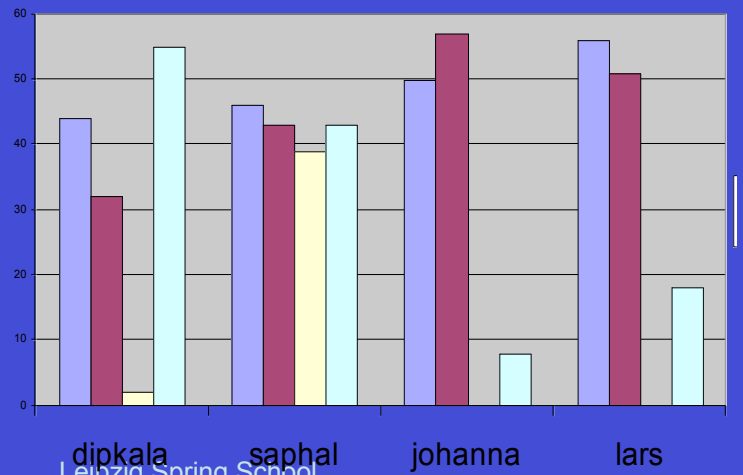
0;8



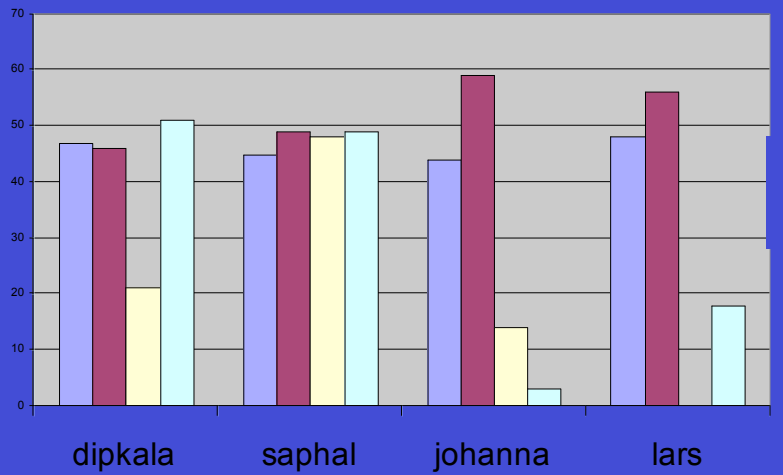
0;10



1;0



1;3



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

Our study

- Mostly outside
- Many different situations
- Mother often absent
- Many other children

Most previous studies

- Inside the house
- Mother and child playing
- Only mother present
- No other children

The end

Thank you!