

# Crosslinguistic language development: Corpus, experimental and modelling studies

## Lecture 3

# Outline

- Productivity
- Theoretical issues
- Naturalistic corpora: analyses
  - Person and number marking on Spanish verbs
- Experimental studies
  - Case-marking in Polish
  - Transitive word order in English and French
  - Relative clauses in English and German
- Modelling studies
  - Starting small
  - Head direction

# Productivity – why does it matter?

## Methodologically

- Unless they are clear errors and unattested in the adult language, utterances and accompanying morphology could have been rote-learned
- If its rote-learned, it should not be compared with a productive form in another language
- So we have to have ways of assessing productivity before we can make comparisons:
  1. Using corpora
  2. Introducing novel items and seeing what the child does with them
  3. Comparing children's comprehension and production of utterances with contrasting forms
  4. Modelling to see if productivity can develop as a function of input

# Theoretically

## Full competence model:

Children bring abstract linguistic categories to the learning of language

Why do children make errors?

- performance limitations
- late maturation
- language-specific features

Predictions:

- Ø Early abstract knowledge
- Ø Few errors with forms that they already know
- Ø Rapid development
- Ø Relatively minimal role of input

# Theoretically

## Constructivist model:

Children build abstract categories as they learn languages

Why do children make errors?

- Their representations are initially 'low-scope' and 'item-based'
- They extend forms they know to the wrong contexts
- They have mis-analysed the input

Predictions:

- Ø Item-based generalisations
- Ø Limited productivity, even with forms they know, at younger ages
- Ø Piecemeal development
- Ø Important role of input

# Studies with naturalistic corpora

## 1. Spanish verb inflections:

- Does productivity develop?
- Are children less productive than adults?

# Using corpora: Spanish verb inflections

## Nottingham corpus

- Lucia: 22 hours: 2;2.25 – 2;7.14
- Juan; 31 hours: 1;1-.21 – 2;5.28
- Only verbs used by both adult and child
  - stem
  - agreement properties
- Adult sample of verb tokens randomly reduced to number found in child's speech

# Number of inflections per stem

- No significant difference between parents
- Significant difference between children and parents at both tested ages
- For Juan, significant difference between first and second half of the corpus



## 2. Are errors just 'noise' in the data?

- Does productivity develop?
- Are children less productive than adults?

## Rates of Subject-Verb Agreement Error in Juan and Lucía's data broken down by Person and Number

	Juan		Lucía	
Inflection	Contexts	Error Rate	Contexts	Error Rate
Overall	3152	4.5%	1672	4.4%

Aguado-Orea & Pine, under review,  
Aguado-Orea, PhD.

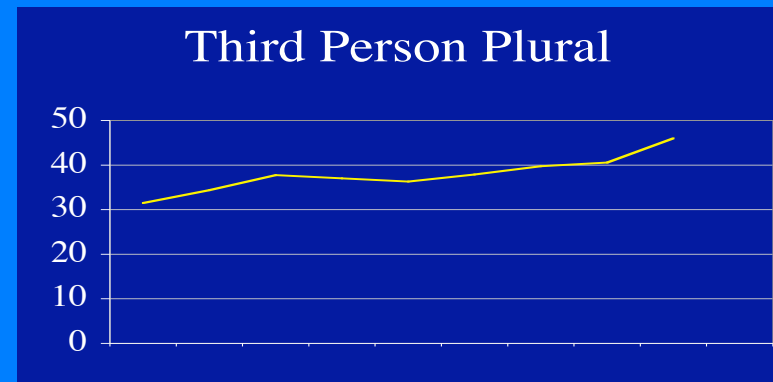
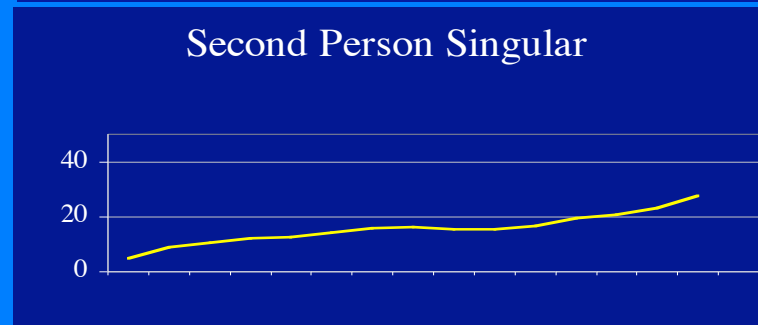
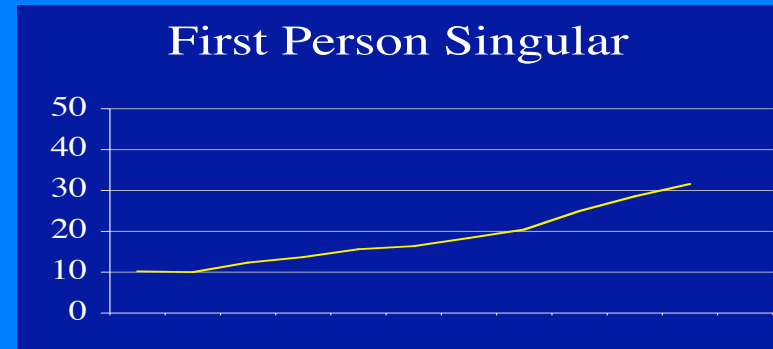
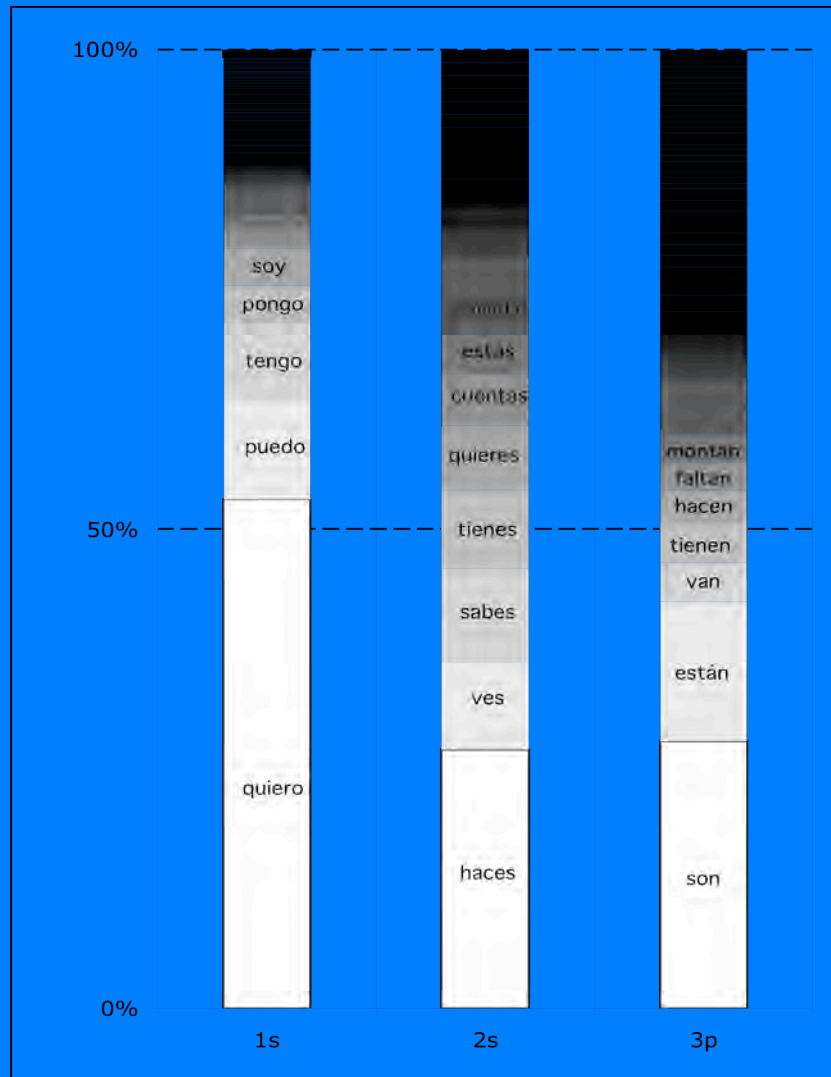
## Rates of Subject-Verb Agreement Error in Juan and Lucía's data broken down by Person and Number

	Juan		Lucía	
Inflection	Contexts	Error Rate	Contexts	Error Rate
Overall	3152	4.5%	1672	4.4%
1sg	693	4.9%	496	3.0%
2sg	147	10.2%	96	22.9%
3sg	1997	0.7%	1018	0.5%
1pl	61	0	14	0
3pl	251	31.5%	48	66.7%

**Pattern of error very similar across children ( $r = 0.99$ )**

**Not consistent with idea that errors can be disregarded as noise in the data**

# Rates of Subject-Verb Agreement Error in Juan's Data as a function of Lexical Frequency



# Summary

- Apparent sophistication of Spanish children's use of verb inflection an illusion
- Low overall error rate reflects
  - Children's knowledge of a relatively small number of high frequency forms
  - Children's use of most frequent inflection (3rd person singular) as a 'default' when they don't know what to do

# Experimental studies

<b>Structure</b>	<b>Method</b>	<b>Languages</b>
<b>Case-marking</b>	Elicited production	Polish
<b>Transitive</b>	Weird word order	English, French
<b>Relative clauses</b>	Sentence repetition	English, German

# Why novel forms?

- The only way to tell whether there is productivity
- Based on Berko's 'wug' test

# 1. Case-marking in Polish



# The effect of a **source form** on the production of a **target form**.

Six source-target pairs (conditions):

DAT MASC <b>-owi</b>	à	GEN MASC <b>-a</b>
LOC MASC <b>-u</b>	à	GEN MASC <b>-a</b>
INSTR MASC <b>-em</b>	à	GEN MASC <b>-a</b>
DAT FEM <b>-i</b>	à	GEN FEM <b>-i</b>
LOC FEM <b>-i</b>	à	GEN FEM <b>-i</b>
INSTR FEM <b>-ą</b>	à	GEN FEM <b>-i</b>

Krajewski et al., 2007  
Krajewski, PhD

# Method

- Two age groups:
  - younger (N = 24, mean = 2;8, median = 2;8)
  - older (N = 31, mean = 3;6, median = 3;6)
- 4 items in each condition
- Within-subject design: each child gets all 24 items
- Each item: 3 drawings featuring a funny creature, two drawings to model a source form of its name, the last drawing to test the target form

The boy is thinking about X-loc.  
(e.g. '*Doc*i**')

The boy is riding X-loc.  
(e.g. '*Doc*i')

The boy is cuddling up to \_\_\_\_\_ [X-gen].  
(e.g. '*Doc*i**')

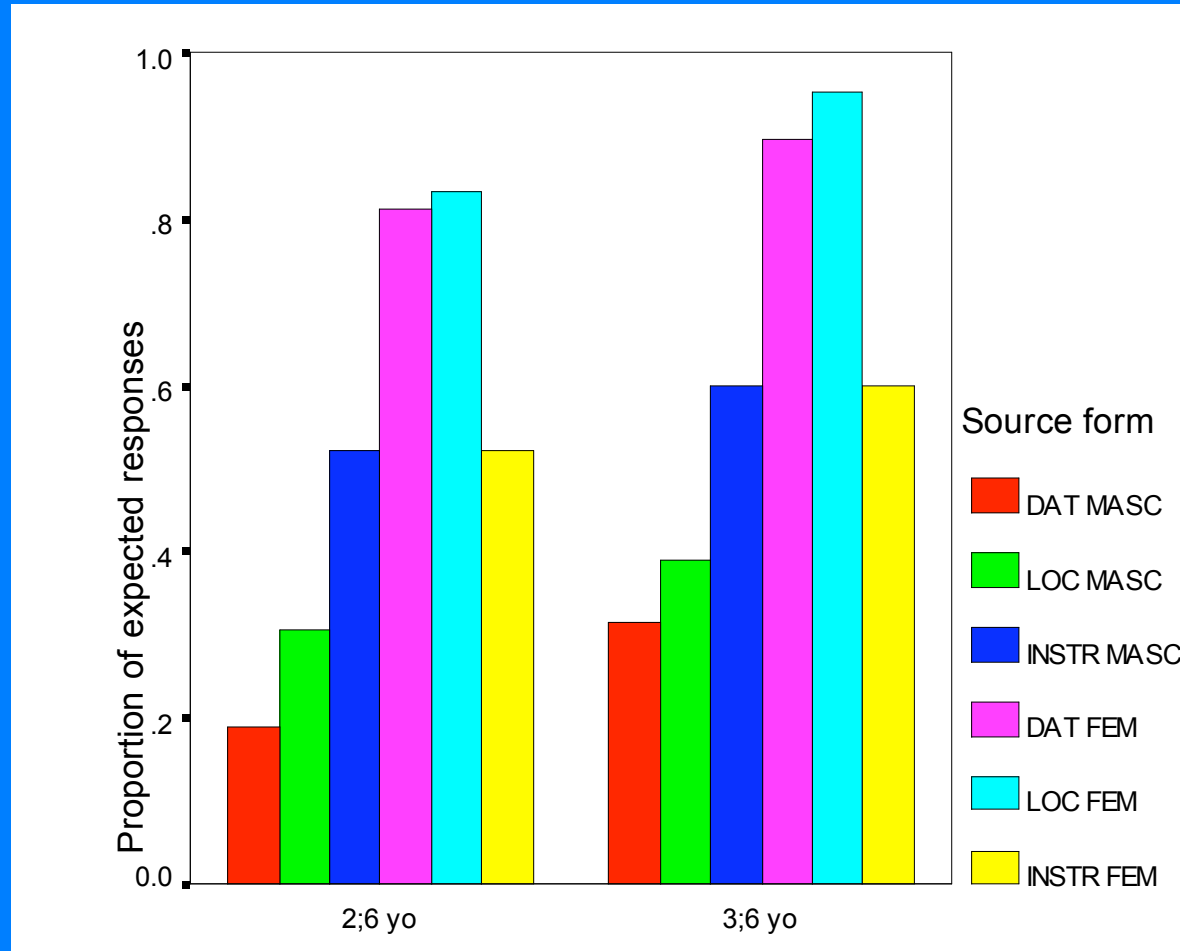
The doctor takes X's temperature [X-dat].  
(e.g. '*Puniow*o*wi*')

The doctor gives an injection to X-dat.  
(e.g. '*Puniow*i')

The doctor is cuddling up to \_\_\_\_\_ [X-gen].  
(e.g. '*Punia***a**')



# Study 1: Results



- Small but significant effect of age

$F(1, 53)=4.109$ ,  
 $p=.048$ , partial  
 $\eta^2=.072$

- The same pattern of results in both age groups

- Main effect of source form

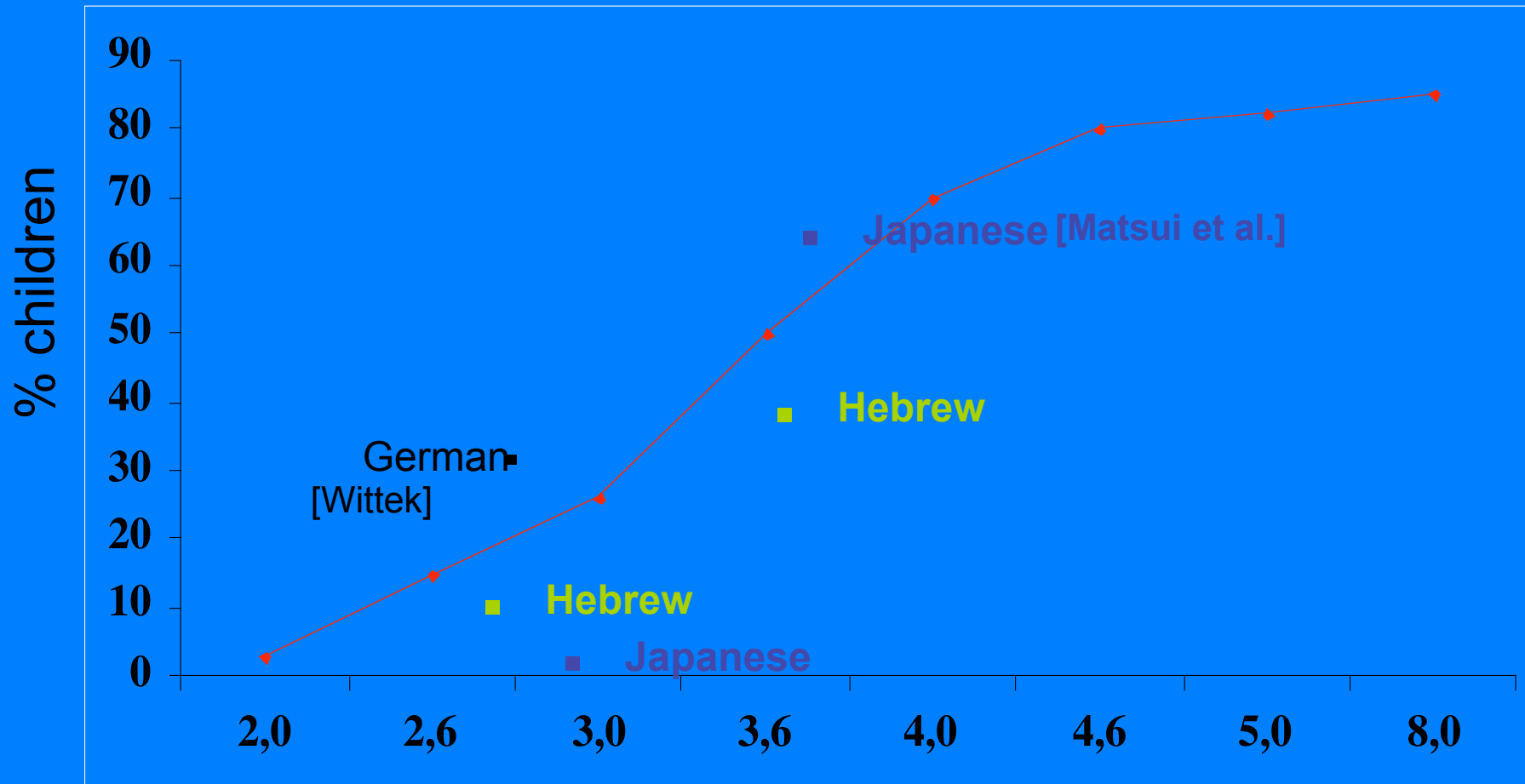
$F(3.83,$   
 $203.008)=60.184$ ,  
 $p<.001$ , partial  
 $\eta^2=.532$

# Conclusions

- Full system builds up, not all there from the outset – even some adults may not be fully productive
- Movement from one form to another may differ for both source and target
- Nominative not necessarily the ‘base’
- Frequency is important but so are other factors e.g. ‘phonological neighbourhoods’

## 2. Learning an abstract transitive construction

# Novel verb studies of Syntax



# Weird word order

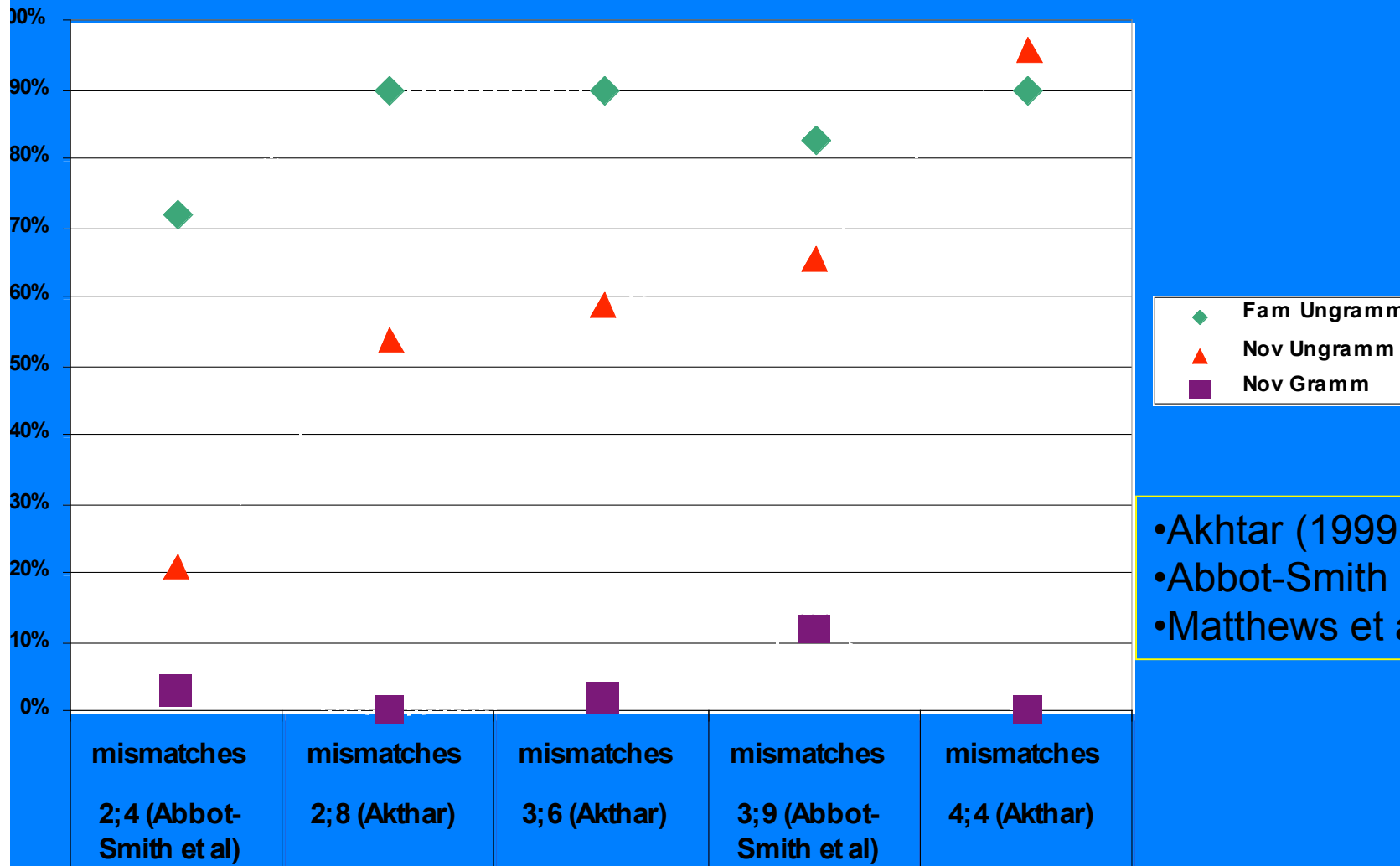
## Pull

Watch what Bear is going to do to Duck!

Bear Duck pulling!  
Look! Bear Duck pulling  
Oh, Bear Duck pulled  
Did you see what happened?  
Bear Duck pulled!



# % Mismatches as a function of condition and age group



- Akhtar (1999)
- Abbot-Smith et al.(2001)
- Matthews et al, (2005)

# Weird word order in French and English

- John pushes **Mary**      John pousse **Mary**
- He pushes her      Il **la** pousse

Matthews et al, 2007]

## SOV

*Oh! Regarde ce que fait Renard à Canard*

[Oh! Look what Fox is doing to Duck]

*Renard Canard pousse*

[Fox Duck pushes/is pushing]

*Oh la la, Renard Canard pousse!*

*or*

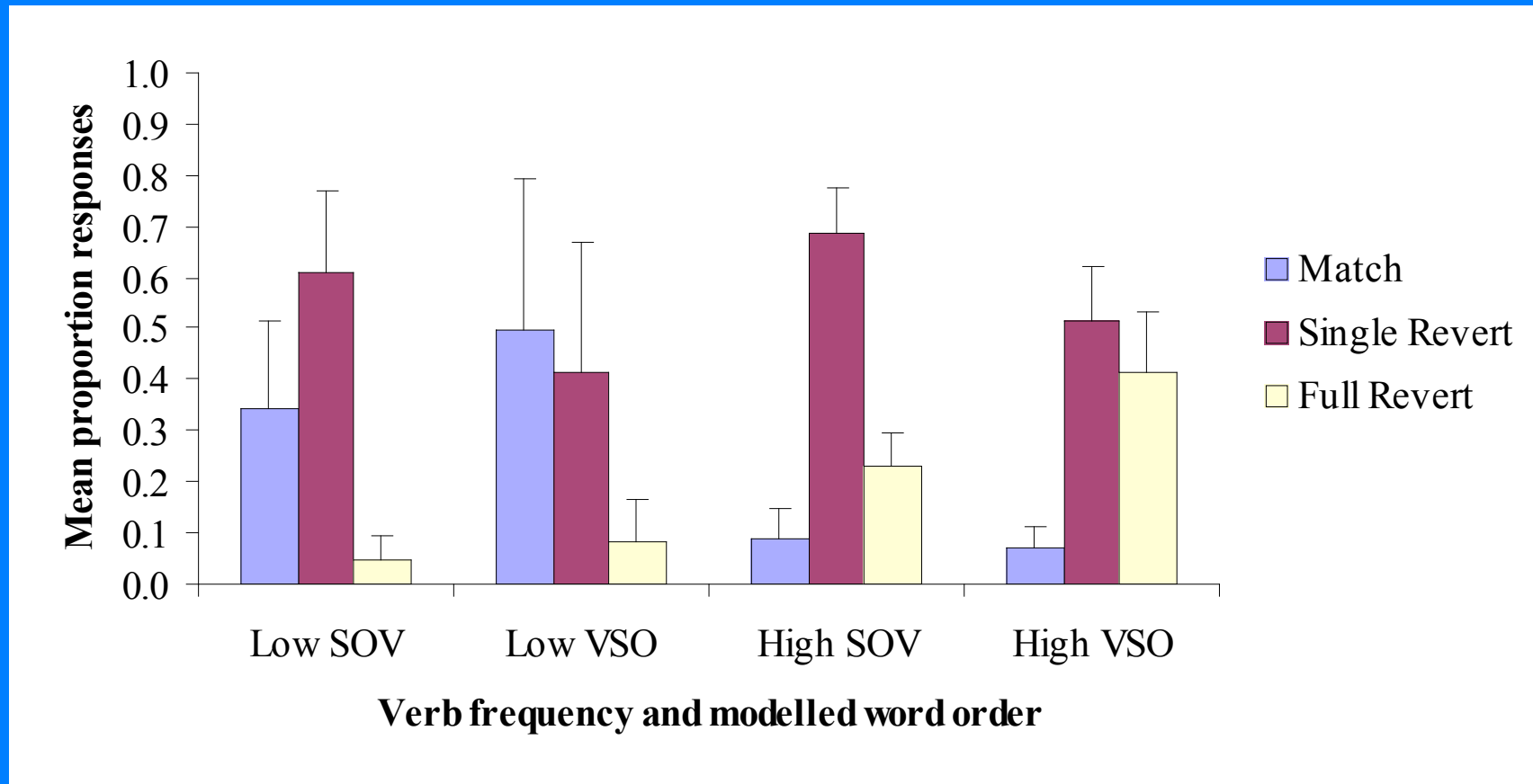
## VSO

*Pousse Renard Canard*

*Etc....*

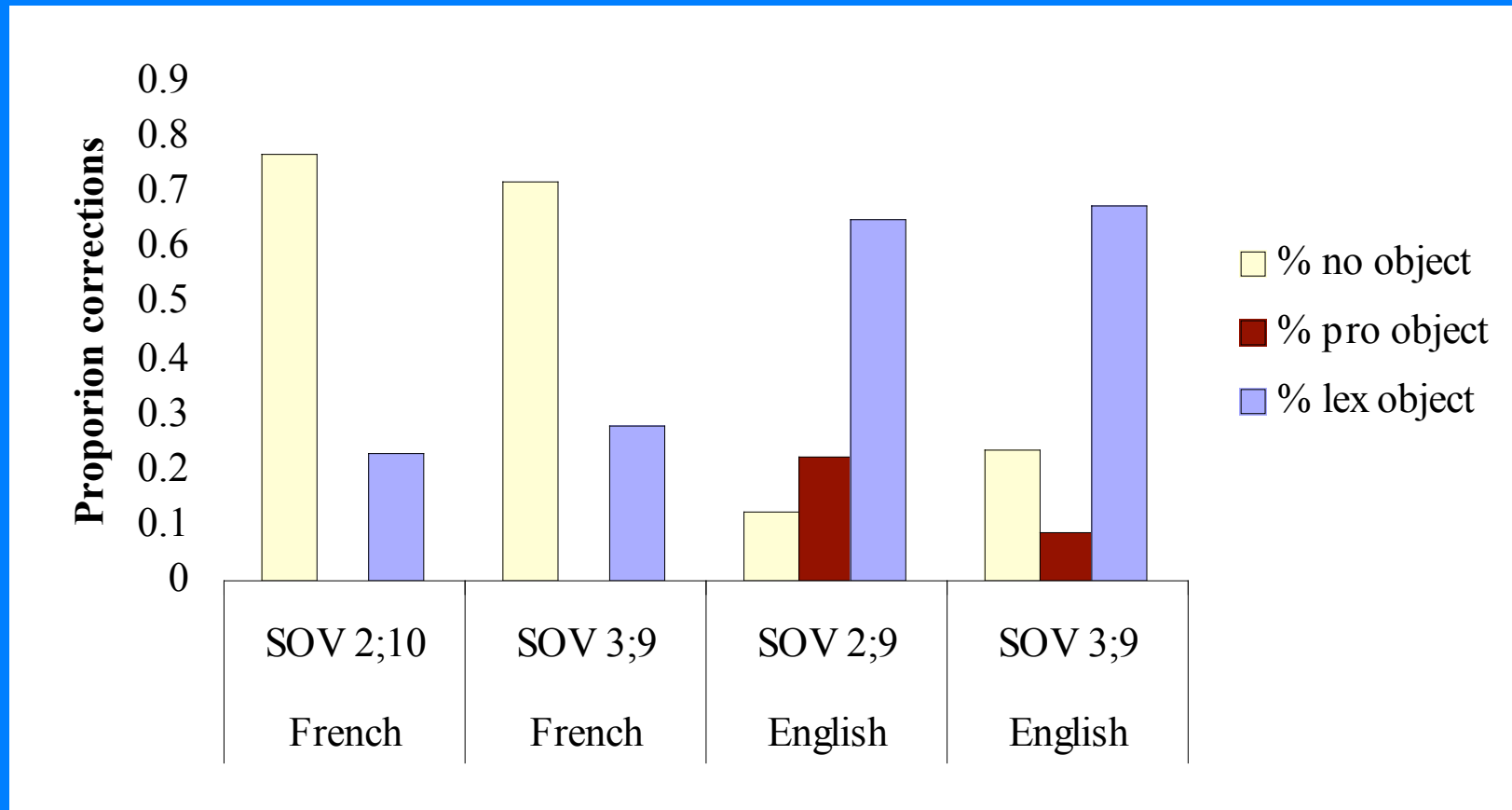


# Weird word order in French



Mean proportion of Matches, Single Argument Reversions and Full Reversions as a function of verb frequency and modelled word order (mean age 2;10).

# Object expression in English and French



Mean proportion of canonically ordered responses that expressed no object, a pronominal object or a lexical object as a function of age and language.

## 2. Relative clauses in English and German

- Sentence repetition is relatively easy for children although they have to be able to say sentences of equivalent length
- Compare repetition on different structures
- Rank structures in terms of correct repetitions/types of errors
- Compare rankings between languages
- Children aged 4;3 – 4;9

# Examples of English test sentences

- S-relative: intransitive verb and subject gap
  - There's the boy **who** played in the garden yesterday
- A-relative: transitive verb and subject gap
  - This's the man **who** saw Peter on the bus this morning
- P-relative: direct object gap
  - This's the girl **who** the boy teased at school this morning
- IO-relative: indirect object gap
  - There's the girl **who** Peter borrowed a football from
- OBL-relative: oblique gap
  - This's the dog **that** the cat ran away from this morning
- GEN-relative: with a genitive relative pronoun
  - This's the woman **whose** cat caught a mouse yesterday

# Examples of German test sentences

- S-relative: intransitive verb and subject gap
  - Da ist die Katze, **die** gerade auf den Baum geklettert ist
- A-relative: transitive verb and subject gap
  - Da ist der Junge, **der** den Mann im Garten gesucht hat
- P-relative: direct object gap
  - Da ist der Mann, **dem** das Mädchen im Stall gesehen hat
- IO-relative: indirect object gap
  - Da ist der Junge, **dem** Paul die Mütze weggenommen hat
- OBL-relative: oblique gap
  - Da ist das Pferd, **auf dem** Gabi vorhin geritten ist
- GEN-relative: with a genitive relative pronoun
  - Da ist die Frau, **deren** Katze eine Maus gefangen hat

# Filler-gap hypothesis

## SUBJ-relative

- The boy [who \_\_\_\_ kissed Mary]

## OBJ-relative

- The boy [(who) Mary kissed \_\_\_\_]

ØS = A relatives

ØS and A relatives >> P and IO relatives

# Results

For both English and German:

- Copula main clauses better than transitives
- S-relatives best > A-relatives > P-relatives
- GEN-relatives almost always incorrect

à Filler-gap hypothesis will not account for these data

à Children building from earlier simple sentences



# Errors

P-, IO- and OBL-relatives à S/A relatives

When converting, the English children :

- often changed the word order
  - This is the girl who the boy teased at school this morning  
Ø This is the girl that teased the boy at school this morning
- often left out the subject
  - This is the girl who Peter borrowed a football from  
Ø This is the girl who borrowed a football from
- sometimes used resumptive pronouns
  - This is the girl who the boy teased this morning  
Ø This is the girl who, who the boy teased the girl this morning

Test:

Hier ist der Junge, dem die Frau ein Buch  
vorgelesen hat

Here is the boy whom (DAT) the woman read a book to

Child:

Hier ist der Junge, der der Frau ein Buch  
vorgelesen hat

Here is the boy who (NOM) read a book to the woman

# Conclusions

- Copula relatives easier because this is where children start with relative clauses
- Filler-gap hypothesis not true for either language
- Children working from word order in English and case-marking in German
- Different, typologically-specific requirements for dealing with the 'same' structure

# Modelling studies

- Analyses of real corpora using a particular learning model
- Using a subset of real sentences
- Using an artificial language

# Starting small

- Connectionist learning model is given a set of sentences
- Has to predict what word comes next
- Measure is whether it can distinguish ungrammatical from grammatical sentences
- Can do this with simple sentence but not when complex sentences added:
  - Boy chases dog
  - Boys chase dog
  - Mary walks
  - Boy who dogs chase feeds cat
  - Girls who chases dogs hits cat
- But if complex sentences added after simple sentences, then can do it

# Head direction

## Head direction consistency

- Artificial strings in a connectionist net
  - Sets with inconsistent head direction were harder to learn
- Used the FANAL database of languages (Dryer)
  - genera involving rule combinations that are harder for the network to learn, are rarer

Christiansen & Devlin, 1997

The end

Thank you  
and see you on  
Friday!