Age effects in statistical learning of Japanese:
Evidence from the cross-situational learning paradigm

Patrick Rebuschat, Helena Farrimond, and Padraic Monaghan
IASCL, July 19, 2017

Outline

Background

- What is cross-situational learning?
- Rebuschat, Schoetensack, & Monaghan (in prep): CSL of everything in adults

Rebuschat, Farrimond, and Monaghan (in prep)

- Part of larger study on individual differences in language learning
- Children and adolescents
- CSL of Japanese

Background

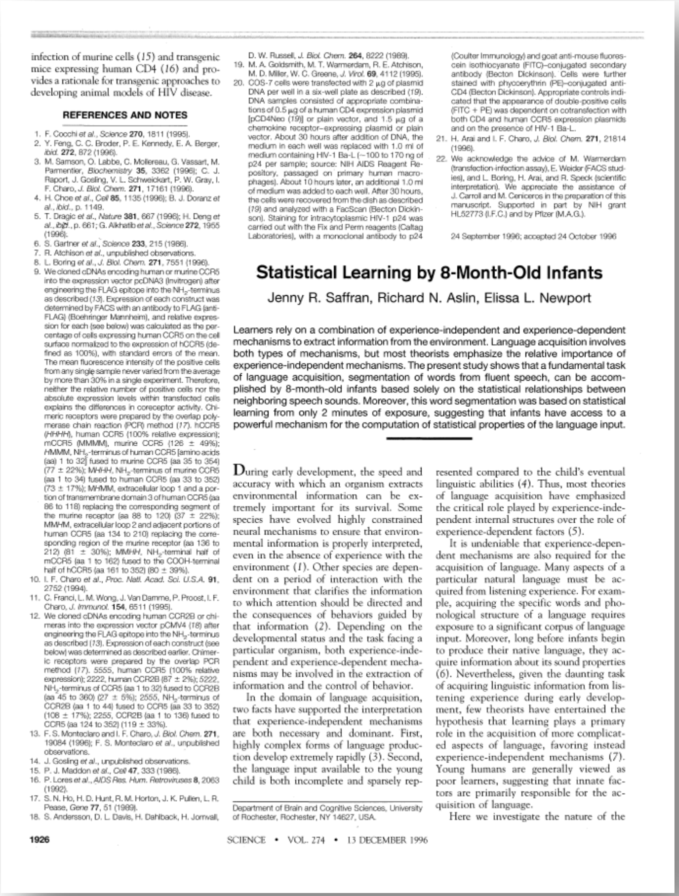
Statistical learning

Our ability to make use of statistical information in the environment to acquire (linguistic) knowledge.

SL well attested in infants, children, adults, and non-human primates.

We can use SL to succeed in a wide variety of linguistic tasks:

- word segmentation and word learning
- phonological development
- syntactic development



Statistical Learning by 8-Month-Old Infants

Jenny R. Saffran, Richard N. Aslin, Elissa L. Newport

Learners rely on a combination of experience-independent and experience-dependent mechanisms to extract information from the environment. Language acquisition involves both types of mechanisms, but most theorists emphasize the relative importance of experience-independent mechanisms. The present study shows that a fundamental task of language acquisition, segmentation of words from fluent speech, can be accomplished by 8-month-old infants based solely on the statistical relationships between neighboring speech sounds. Moreover, this word segmentation was based on statistical learning from only 2 minutes of exposure, suggesting that infants have access to a powerful mechanism for the computation of statistical properties of the language input.

During early development, the speed and accuracy with which an organism extracts environmental information can be extremely important for its survival. Some species have evolved highly constrained neural mechanisms to ensure that environmental information is properly interpreted, even in the absence of experience with the environment (*1*). Other species are dependent on a period of interaction with the environment that clarifies the information to which attention should be directed and the consequences of behaviors guided by that information (*2*). Depending on the developmental status and the task facing a particular organism, both experience-independent and experience-dependent mechanisms may be involved in the extraction of information and the control of behavior.

In the domain of language acquisition, two facts have supported the interpretation that experience-independent mechanisms are both necessary and dominant. First, highly complex forms of language production develop extremely rapidly (*3*). Second, the language input available to the young child is both incomplete and sparsely represented compared to the child's eventual linguistic abilities (*4*). Thus, most theories of language acquisition have emphasized the critical role played by experience-independent internal structures over the role of experience-dependent factors (*5*).

It is undeniable that experience-dependent mechanisms are also required for the acquisition of language. Many aspects of a particular natural language must be acquired from listening experience. For example, acquiring the specific words and phonological structure of a language requires exposure to a significant corpus of language input. Moreover, long before infants begin to produce their native language, they acquire information about its sound properties (*6*). Nevertheless, given the daunting task of acquiring linguistic information from listening experience during early development, few theorists have entertained the hypothesis that learning plays a primary role in the acquisition of more complicated aspects of language, favoring instead experience-independent mechanisms (*7*). Young humans are generally viewed as poor learners, suggesting that innate factors are primarily responsible for the acquisition of language.

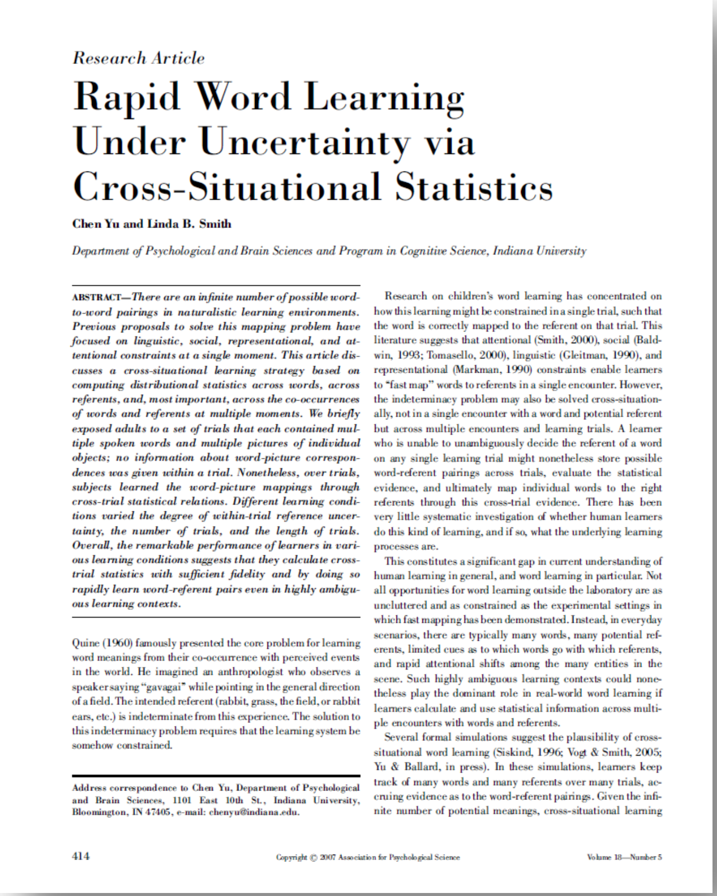
Here we investigate the nature of the

Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY 14627, USA.

Cross-situational learning

Yu and Smith (2007); Smith and Yu (2008)

Our ability to keep track of information across many learning trials (situations) and to make use of this information to learn language.



Research Article

Rapid Word Learning Under Uncertainty via Cross-Situational Statistics

Chen Yu and Linda B. Smith

Department of Psychological and Brain Sciences and Program in Cognitive Science, Indiana University

ABSTRACT—*There are an infinite number of possible word-to-word pairings in naturalistic learning environments. Previous proposals to solve this mapping problem have focused on linguistic, social, representational, and attentional constraints at a single moment. This article discusses a cross-situational learning strategy based on computing distributional statistics across words, across referents, and, most important, across the co-occurrences of words and referents at multiple moments. We briefly exposed adults to a set of trials that each contained multiple spoken words and multiple pictures of individual objects; no information about word-picture correspondences was given within a trial. Nonetheless, over trials, subjects learned the word-picture mappings through cross-trial statistical relations. Different learning conditions varied the degree of within-trial reference uncertainty, the number of trials, and the length of trials. Overall, the remarkable performance of learners in various learning conditions suggests that they calculate cross-trial statistics with sufficient fidelity and by doing so rapidly learn word-referent pairs even in highly ambiguous learning contexts.*

Quine (1960) famously presented the core problem for learning word meanings from their co-occurrence with perceived events in the world. He imagined an anthropologist who observes a speaker saying "gavagai" while pointing in the general direction of a field. The intended referent (rabbit, grass, the field, or rabbit ears, etc.) is indeterminate from this experience. The solution to this indeterminacy problem requires that the learning system be somehow constrained.

Address correspondence to Chen Yu, Department of Psychological and Brain Sciences, 1101 East 10th St., Indiana University, Bloomington, IN 47405, e-mail: chenyu@indiana.edu.

Research on children's word learning has concentrated on how this learning might be constrained in a single trial, such that the word is correctly mapped to the referent on that trial. This literature suggests that attentional (Smith, 2000), social (Baldwin, 1993; Tomasello, 2000), linguistic (Gleitman, 1990), and representational (Markman, 1990) constraints enable learners to "fast map" words to referents in a single encounter. However, the indeterminacy problem may also be solved cross-situationally, not in a single encounter with a word and potential referent but across multiple encounters and learning trials. A learner who is unable to unambiguously decide the referent of a word on any single learning trial might nonetheless store possible word-referent pairings across trials, evaluate the statistical evidence, and ultimately map individual words to the right referents through this cross-trial evidence. There has been very little systematic investigation of whether human learners do this kind of learning, and if so, what the underlying learning processes are.

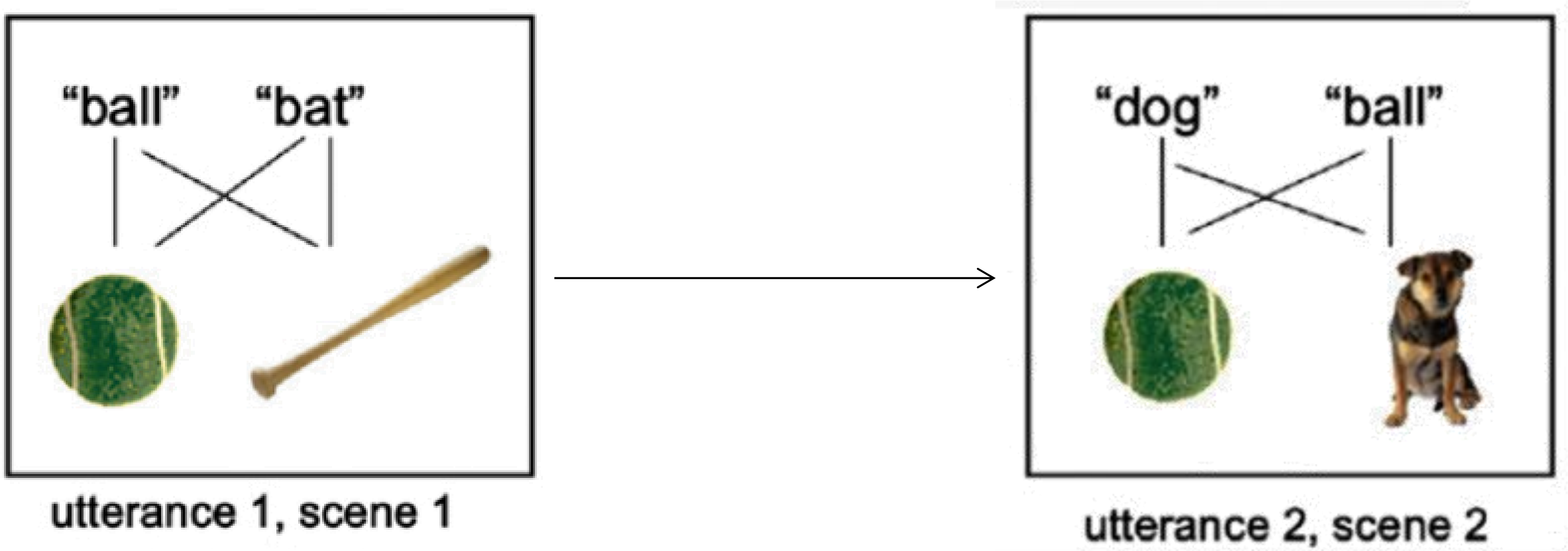
This constitutes a significant gap in current understanding of human learning in general, and word learning in particular. Not all opportunities for word learning outside the laboratory are as uncluttered and as constrained as the experimental settings in which fast mapping has been demonstrated. Instead, in everyday scenarios, there are typically many words, many potential referents, limited cues as to which words go with which referents, and rapid attentional shifts among the many entities in the scene. Such highly ambiguous learning contexts could nonetheless play the dominant role in real-world word learning if learners calculate and use statistical information across multiple encounters with words and referents.

Several formal simulations suggest the plausibility of cross-situational word learning (Siskind, 1996; Vogt & Smith, 2005; Yu & Ballard, in press). In these simulations, learners keep track of many words and many referents over many trials, accruing evidence as to the word-referent pairings. Given the infinite number of potential meanings, cross-situational learning

Cross-situational learning

Yu and Smith (2007); Smith and Yu (2008)

2x2 condition: two referents, two words



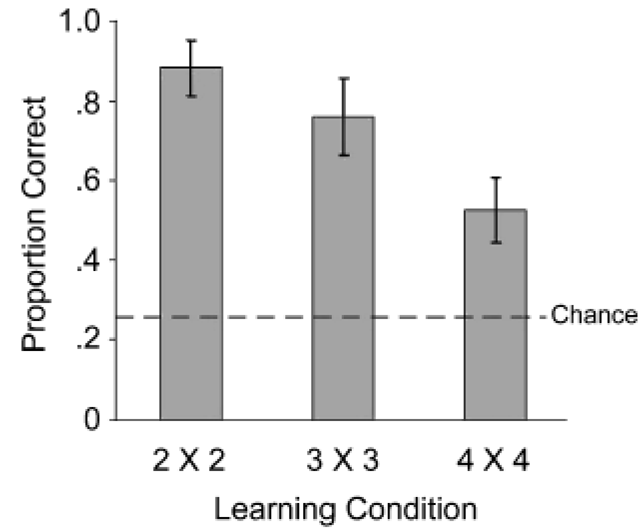
Cross-situational learning

Yu and Smith (2007); Smith and Yu (2008)

Adults can easily track cross-trial statistics and use this information to learn words (nouns).

Three conditions

- 18 words condition
- Each word occurs 6 times
- Exposure time: Less than 6 mins



Results:

- 2x2 condition = learn 16 words
- 3x3 condition = learn 13 words
- 4x4 condition = learn 10 words

Cross-situational learning

Yu and Smith demonstrate that infants (12 to 14 months) can use cross-situational learning to acquire novel nouns.

Scott and Fisher (2012) further showed that 2.5-year-old children can use CSL to acquire novel verbs.

Monaghan et al. (2015) confirms that adults can learn both nouns and verbs simultaneously from cross-situational statistics.

But what about other lexical items like function words? And what about syntax?

Rebuschat, Schoetensack, & Monaghan (in prep):

Can we learn words (nouns, verbs, adjectives, function words) and syntax simultaneously via cross-situational learning?

Part of larger project on individual differences in language learning across the lifespan.

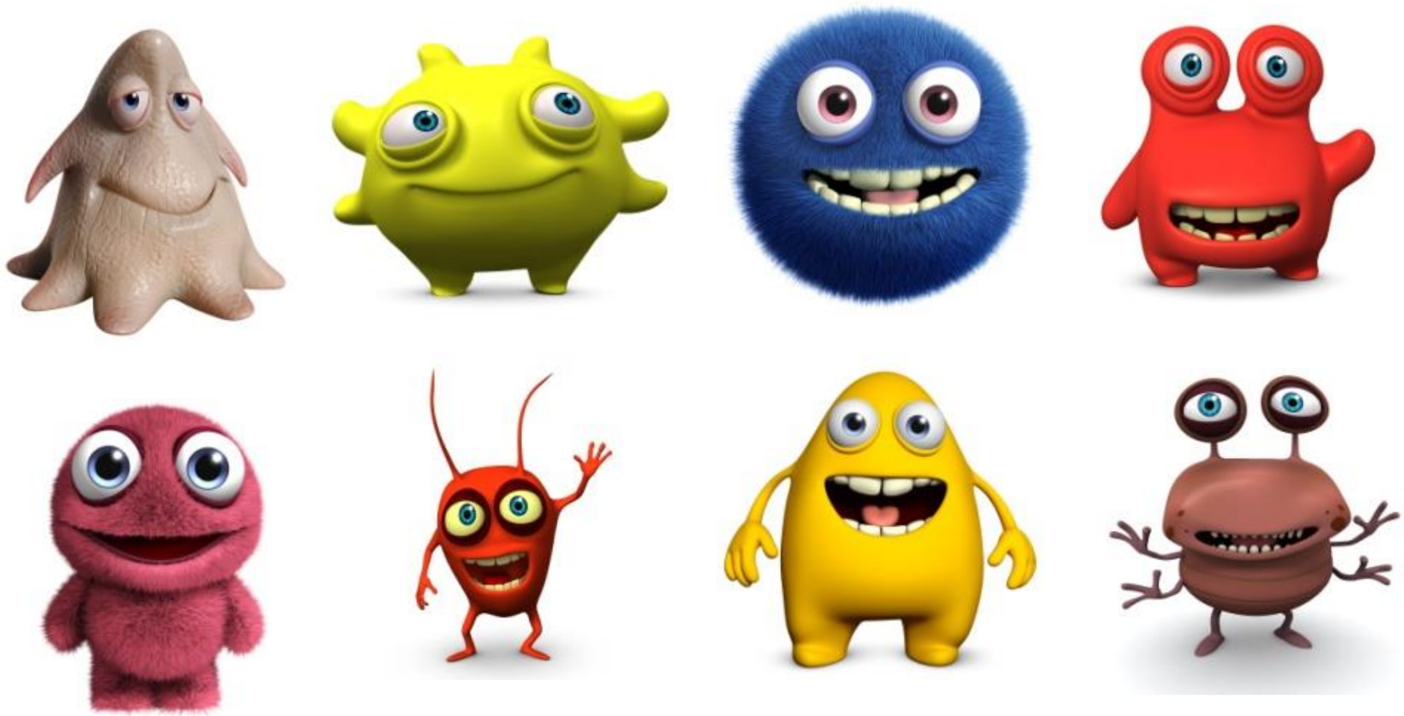
Participants:

- Twenty adult NS of English, no background in Japanese

Materials:

- Developed novel artificial language

Methods: Novel artificial language



Methods: New artificial language

Aliens are depicted performing one of four actions (hiding, jumping, lifting, pushing) in dynamic scenes.

Lexicon = 16 pseudowords

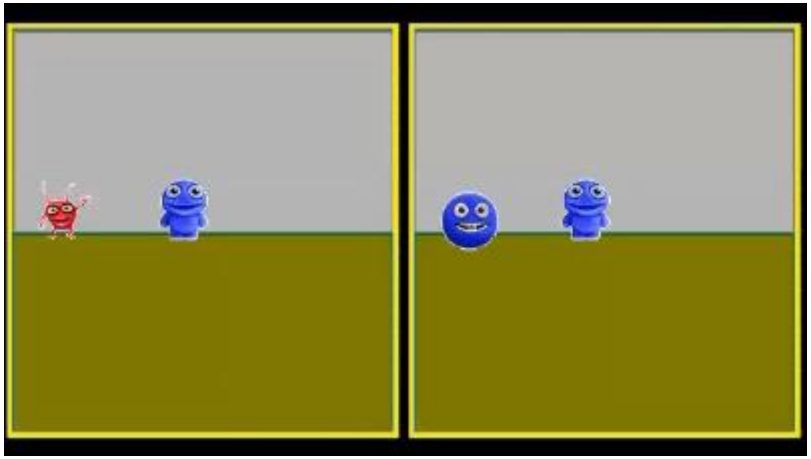
- Eight nouns, four verbs, two adjectives, two function words

Grammar = Japanese (SOV, OSV)

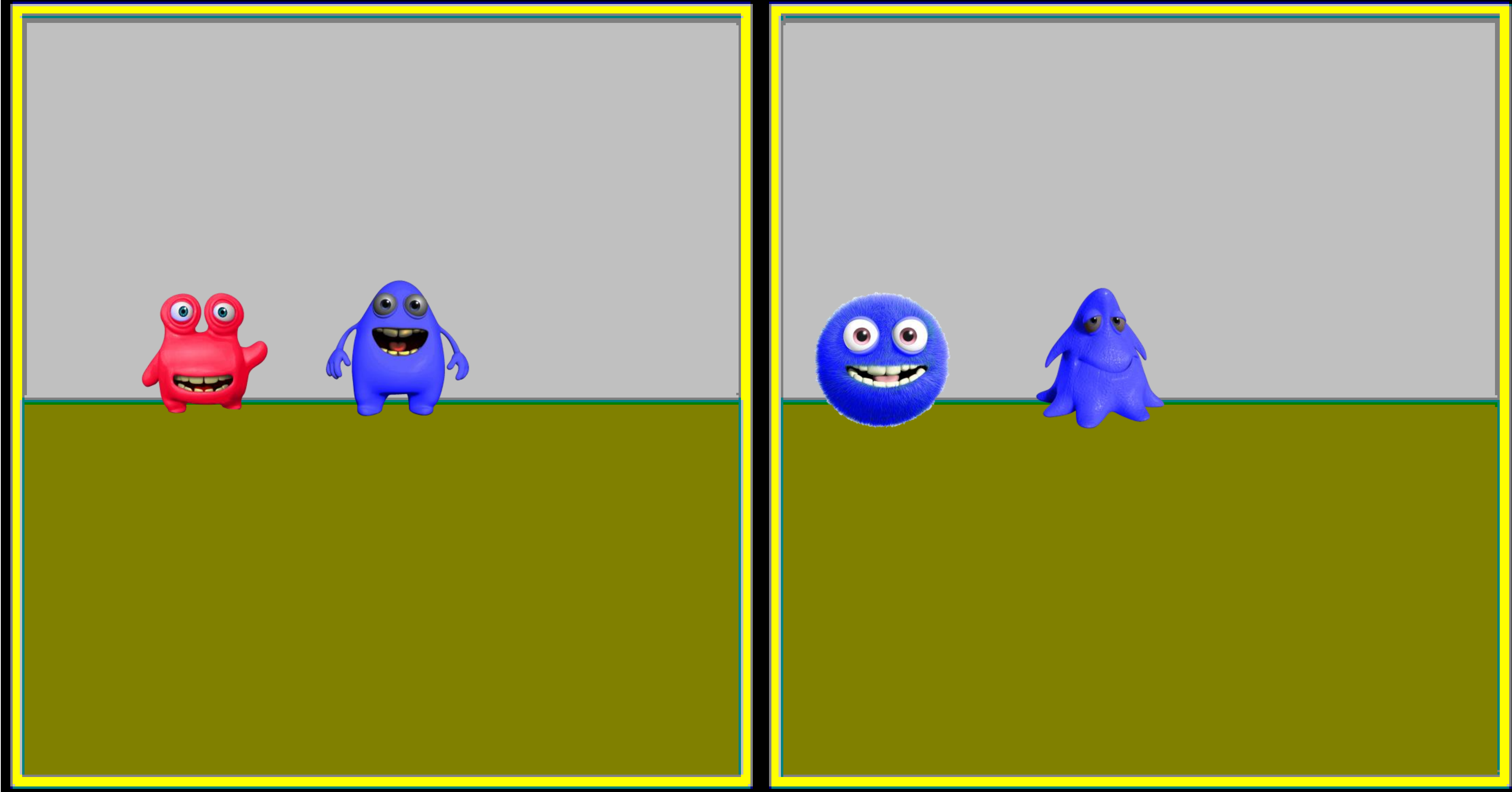
Sixteen training and test blocks:

- 192 training items, 92 test items
- Subjects are tested four times throughout experiment → Allows to check what is learned first and to later shorten exposure phase.

Cross-situational learning task

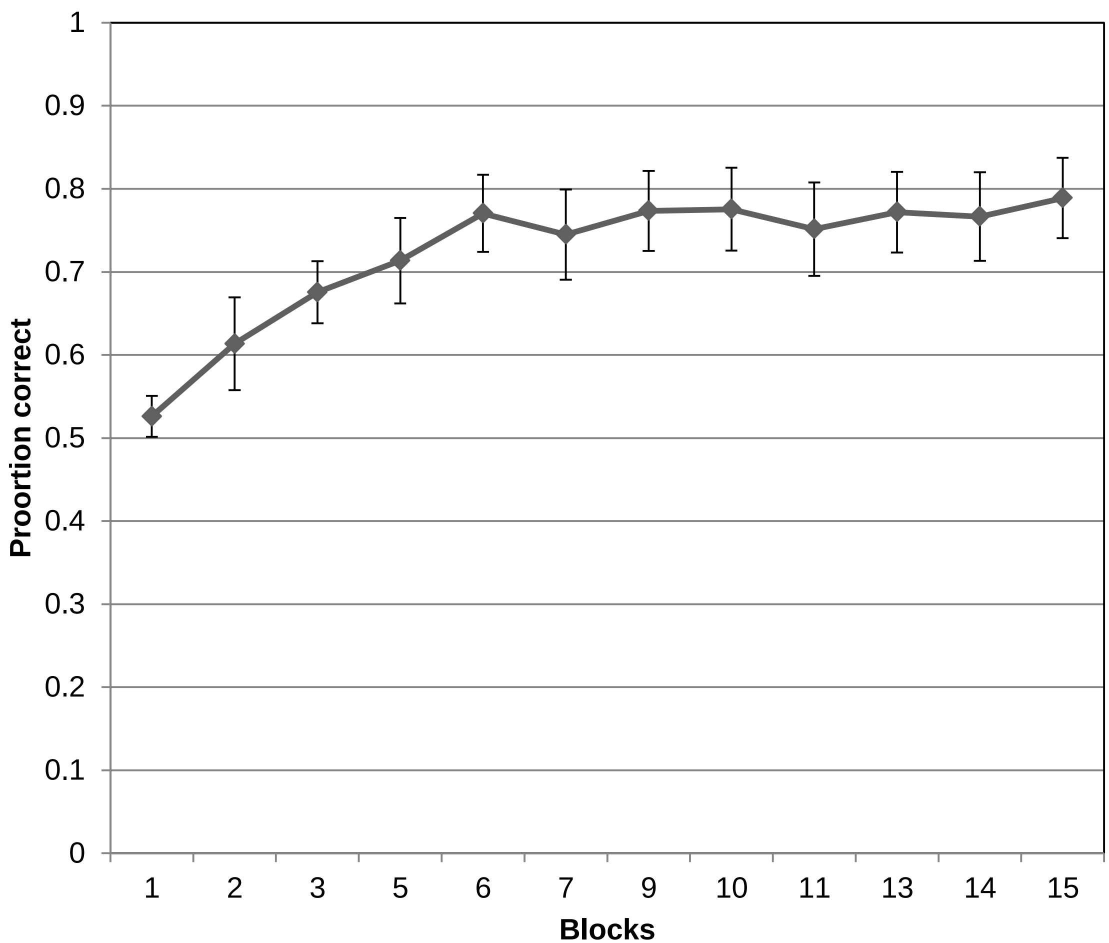


Example cross-situational learning trial



No feedback!

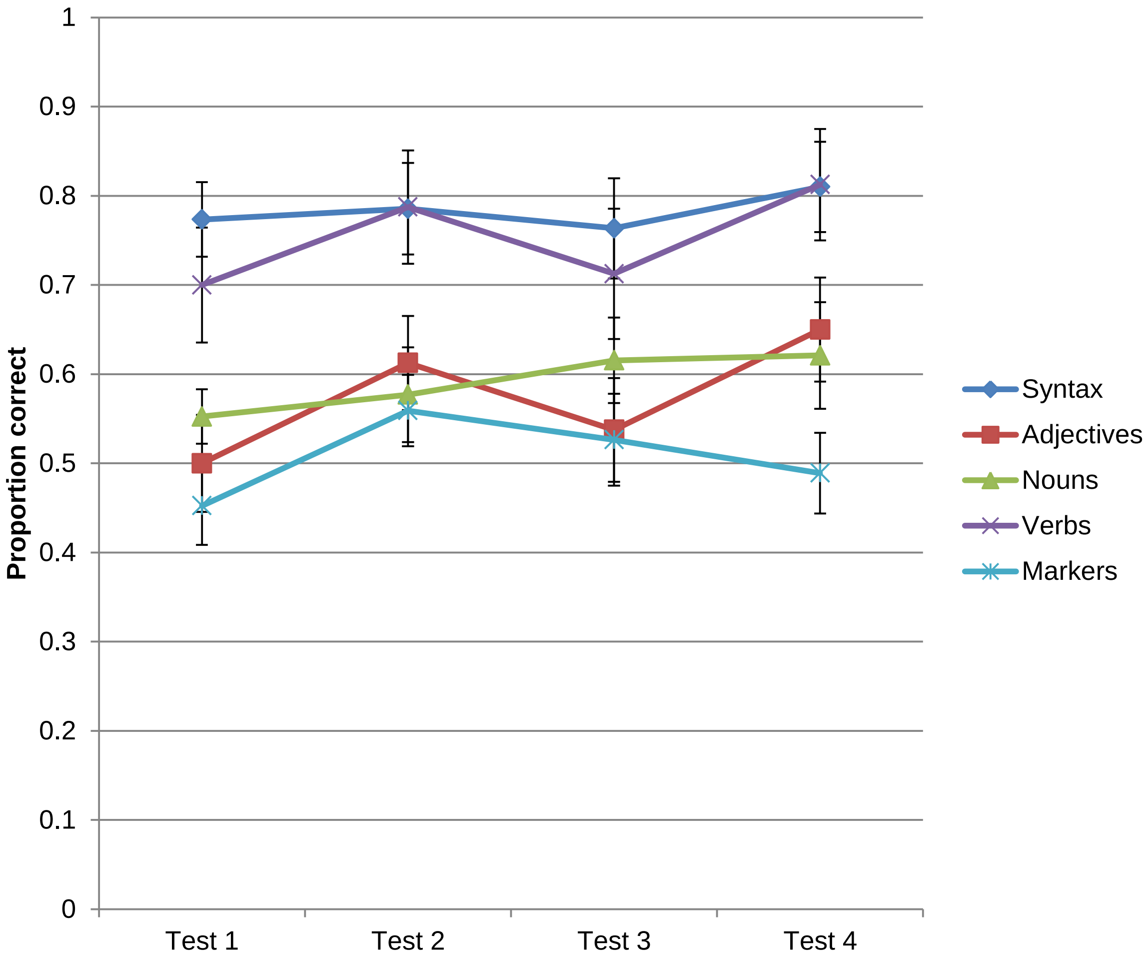
Results: Exposure trials



- Performance above chance from block 3 onwards.
- 48 exposure trials enough to reach above-chance performance

Results: Lexical and syntactic tests

Lancaster University



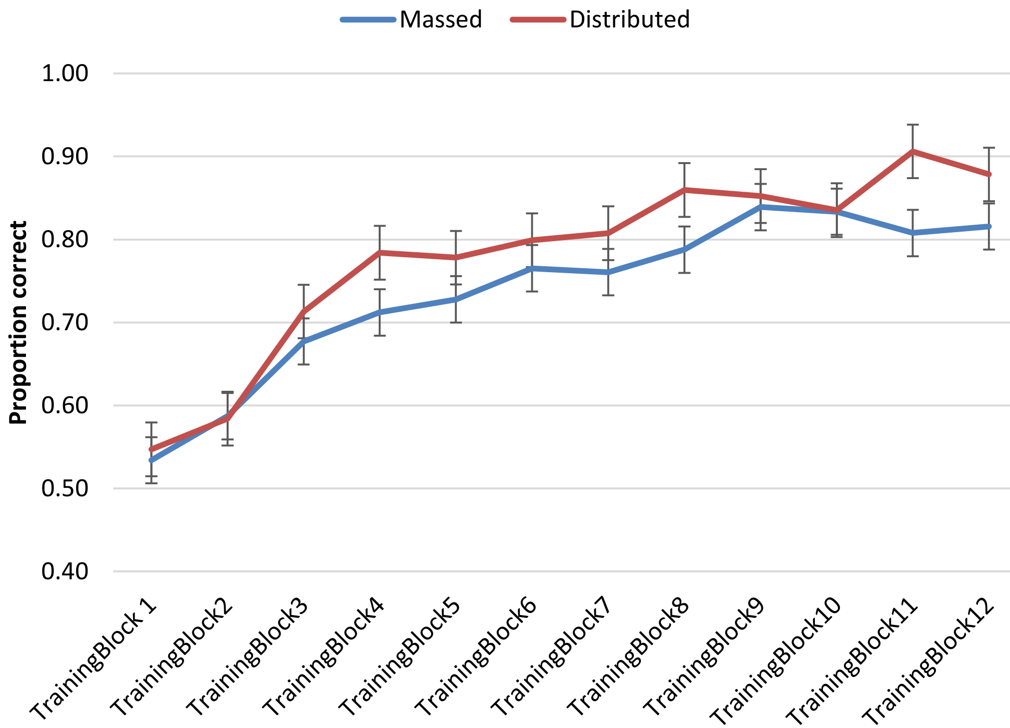
Performance strongest for...

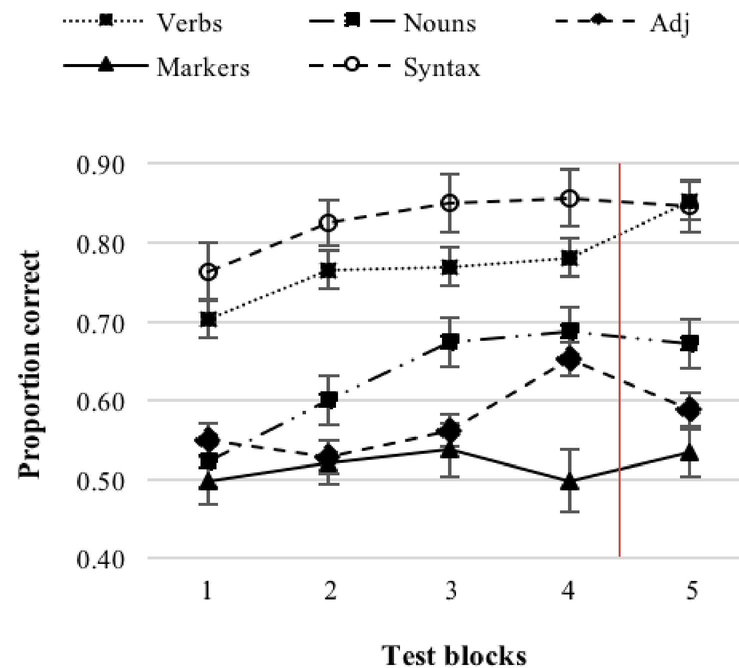
1. Syntax and verbs
2. Nouns
3. Adjectives
4. Marker words

Rapid learning of word order, nouns, verbs, adjectives and markers without feedback.

Spacing effect and IDs

Neil Walker (PhD student, Lancaster)





- Performance on massed condition replicates previous expt
- Delayed PT confirm acquired knowledge robust after 24 hrs
- Confirms learning sequence: Syntax and verbs > nouns > adj > markers

Summary

Rebuschat, Schoetensack, & Monaghan (in prep)

- Adults can use cross-situational statistics to learn words and grammar simultaneously.

Two questions

- What about children?
  - Dunn, Belteki, Rebuschat & Monaghan (in prep)
- Why not just use a natural language?

Rebuschat, Farrimond, & Monaghan (in prep)

Rebuschat, Farrimond, & Monaghan (in prep)

- Can we use a natural language to explore statistical learning of words and syntax?
- Does age make a difference in cross-situational learning?
- (Select stimuli and age cohort for subsequent studies)

Methods

Methods: Participants

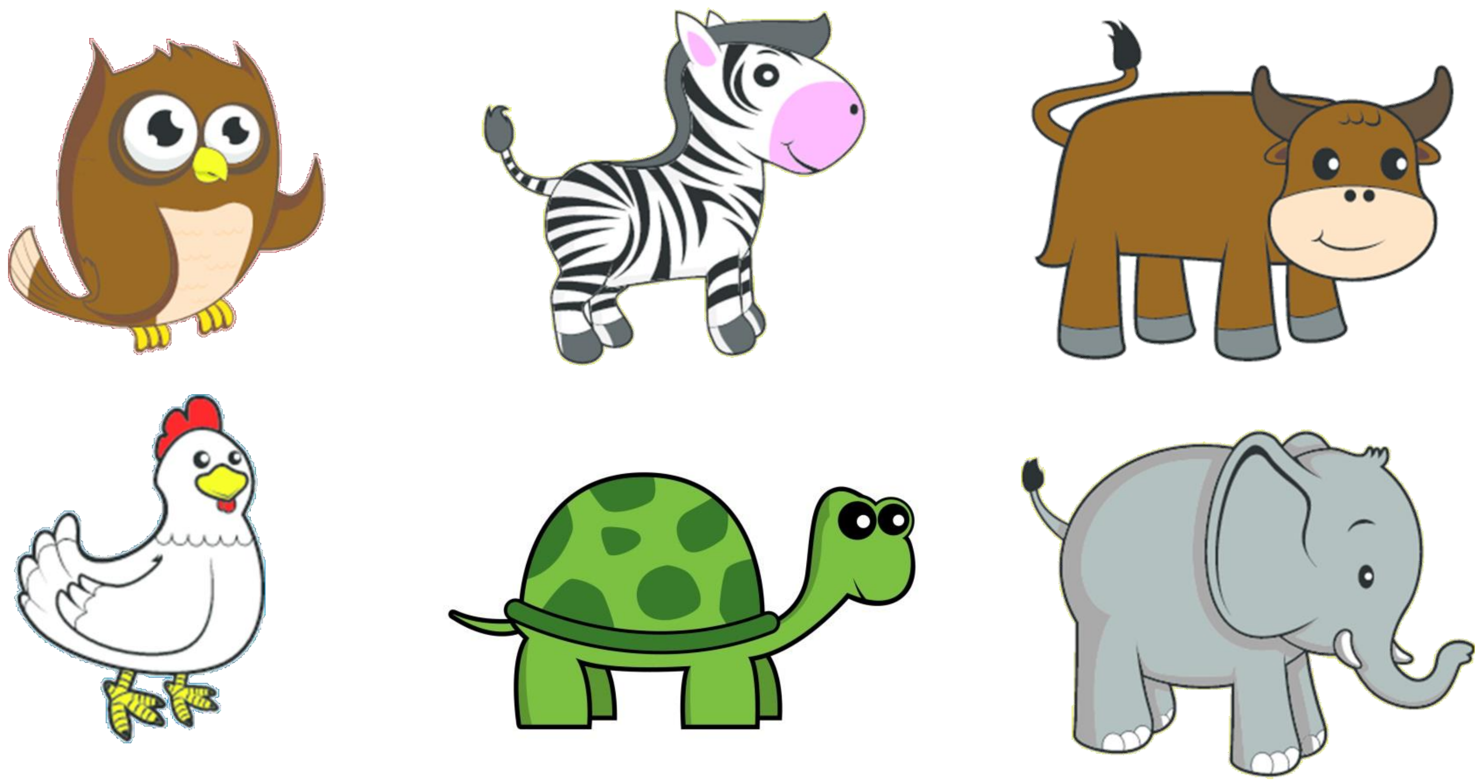
Participants:

- Forty-five NS of English across three age cohorts (each n = 15) :
  - 8-9 years
  - 11-12 years
  - 17-18 years

No background in Japanese or any other VF language.

Participants were recruited and tested at local schools.

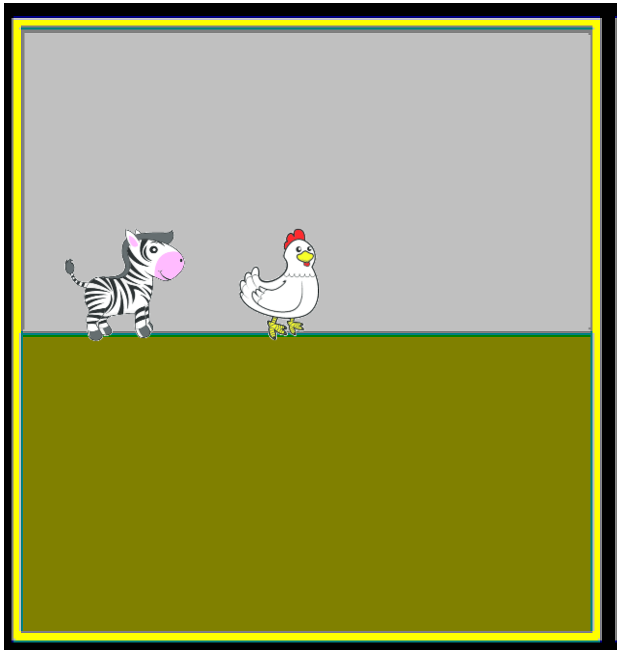
Methods: Mini-Japanese



- Six animal cartoon characters used in experiment

Methods: Mini-Japanese

Animals are depicted performing one of four actions (hiding, jumping, lifting, pushing) in dynamic scenes generated by E-Prime.



Methods: Mini-Japanese

Lexicon = 12 words

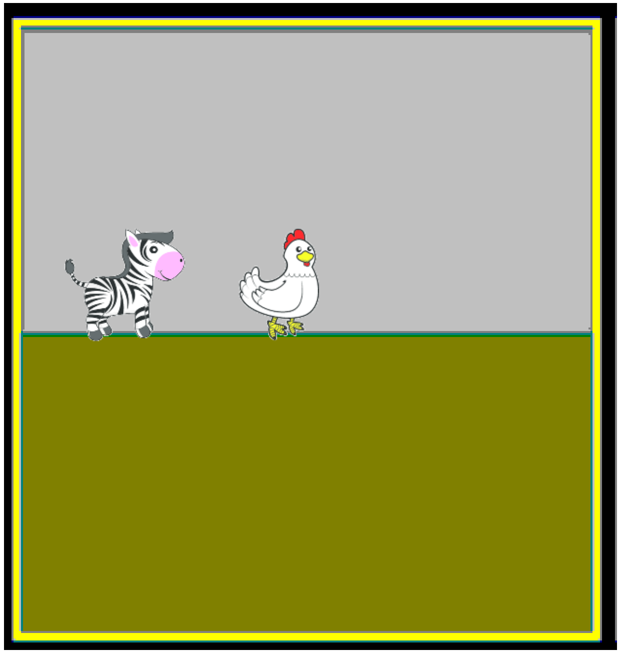
- Six nouns (one per animal)
  - *niwatori*, chicken; *ushi*, cow; *zou*, elephant; *kame*, turtle; *shimauma*, zebra; *fukuoru*, owl
- Four verbs (one per action)
  - *kakusu*, to hide; *tobikueru*, to jump; *mochiageru*, to lift; *taosu*, to knock down
- Two morphological markers
  - *ga* = subject marker; *o* = object marker

Japanese words controlled for length: Half the nouns and verbs three morae in length, the other half five morae.

Methods: Mini-Japanese

Syntax based on Japanese:

- Sentences either SOV or OSV
- Noun phrases have noun as head, followed by obligatory case marker, attached to noun.



Example:

Scene: Zebra jumping over chicken.

Possible descriptions:

- “Shimaumaga niwatorio tobikueru“(SOV)
- “Niwatorio shimaumaga tobikueru“(OSV

Methods: Mini-Japanese

- Generated 72 training sentences and 72 test sentences.

  → Less training than Rebuschat, Schoetensack, & Monaghan (in prep): 48 trials was enough to reach above chance performance

- Lexical frequencies, agent-patient assignment, and word order carefully counterbalanced.

Methods: Materials

Lancaster University

Cross-situational learning task

- Four exposure blocks [EXP] → exposure trials only
- Four mixed blocks [M] → exposure trials and lexical test trials
- Four test blocks [ST] → syntactic test trials

	Block							
	1	2	3	4	5	6	7	8
Block type	Exp	Exp	M	ST	Exp	Exp	M	ST
Nr of trials	12	12	12+14	12	12	12	12+14	12

Methods: Procedure

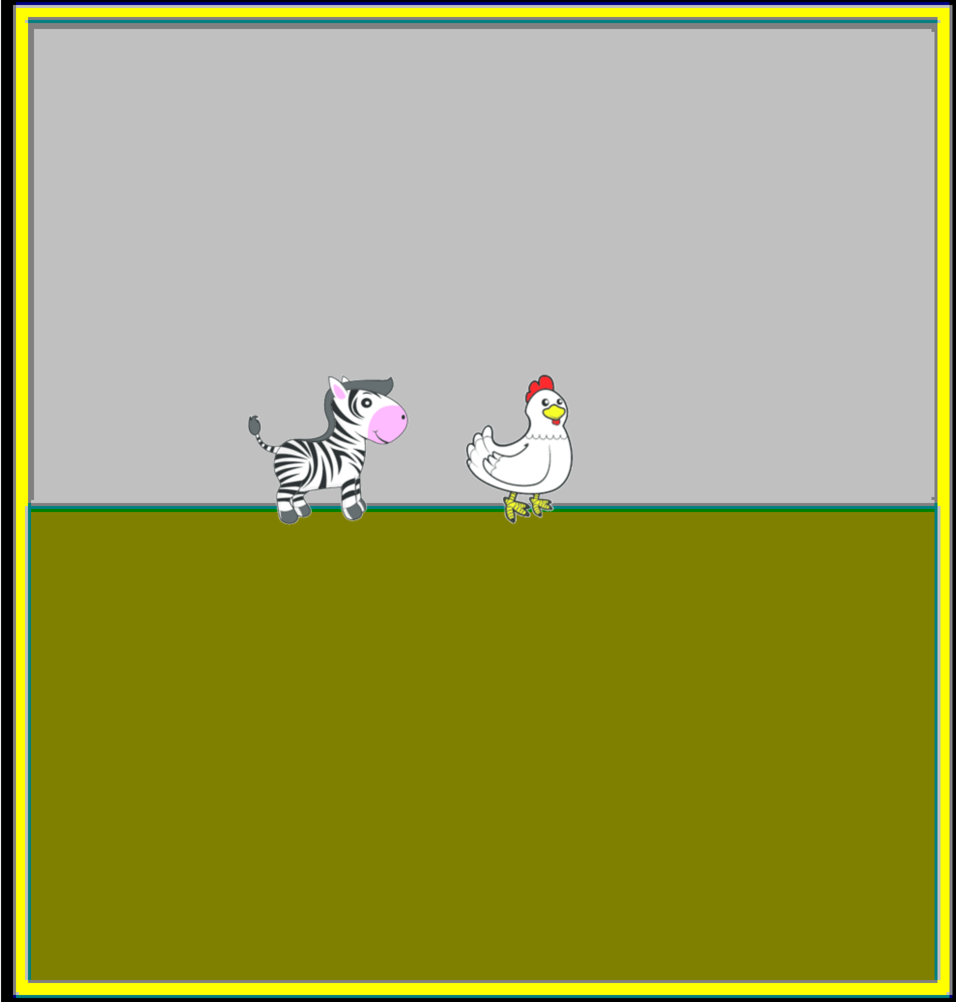
Exposure trials and lexical test trials

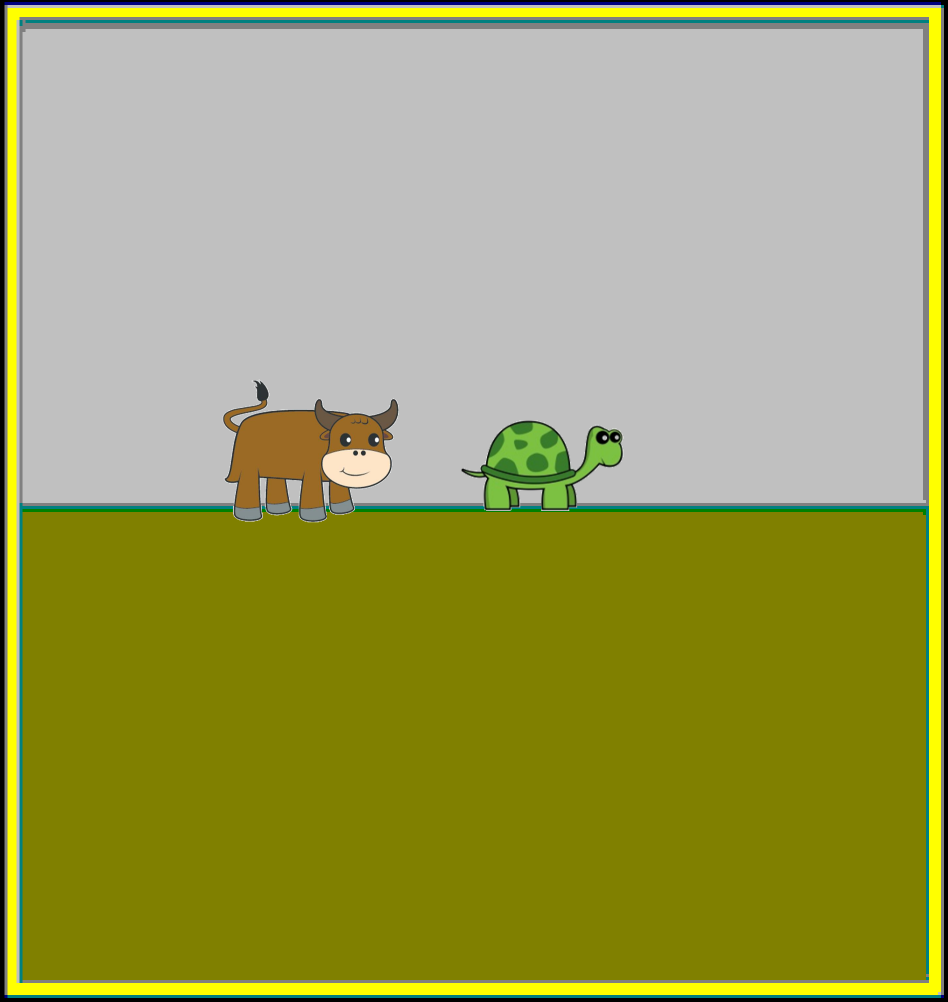
- Participants informed that they would learn a new language.
- They observed two dynamic scenes and hear a sentence in the new language over headphones.
- Task: Decide, as quickly and accurately as possible, which scene the sentence referred to.

No feedback!

	Block							
	1	2	3	4	5	6	7	8
Block type	Exp	Exp	M	ST	Exp	Exp	M	ST
Nr of trials	16	16	40	16	16	16	40	16

Example trial

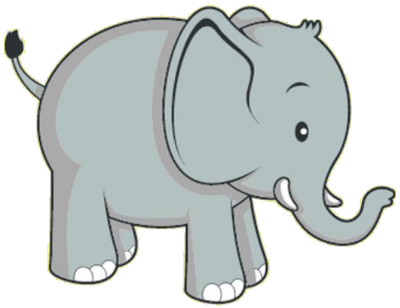




Methods: Procedure

Exposure trials and lexical test trials

- In the lexical test trials, the scenes were identical with one difference.



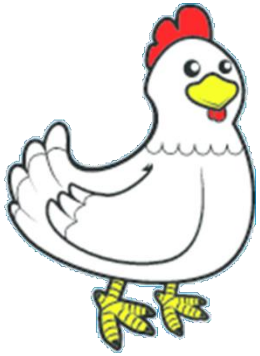
	Animals	Actions	Agent-patient assignment
Noun test	Different	Same	Same
Verb test	Same	Different	Same
Marker test	Same	Same	Different

No feedback!

Methods: Procedure

Syntactic test trials

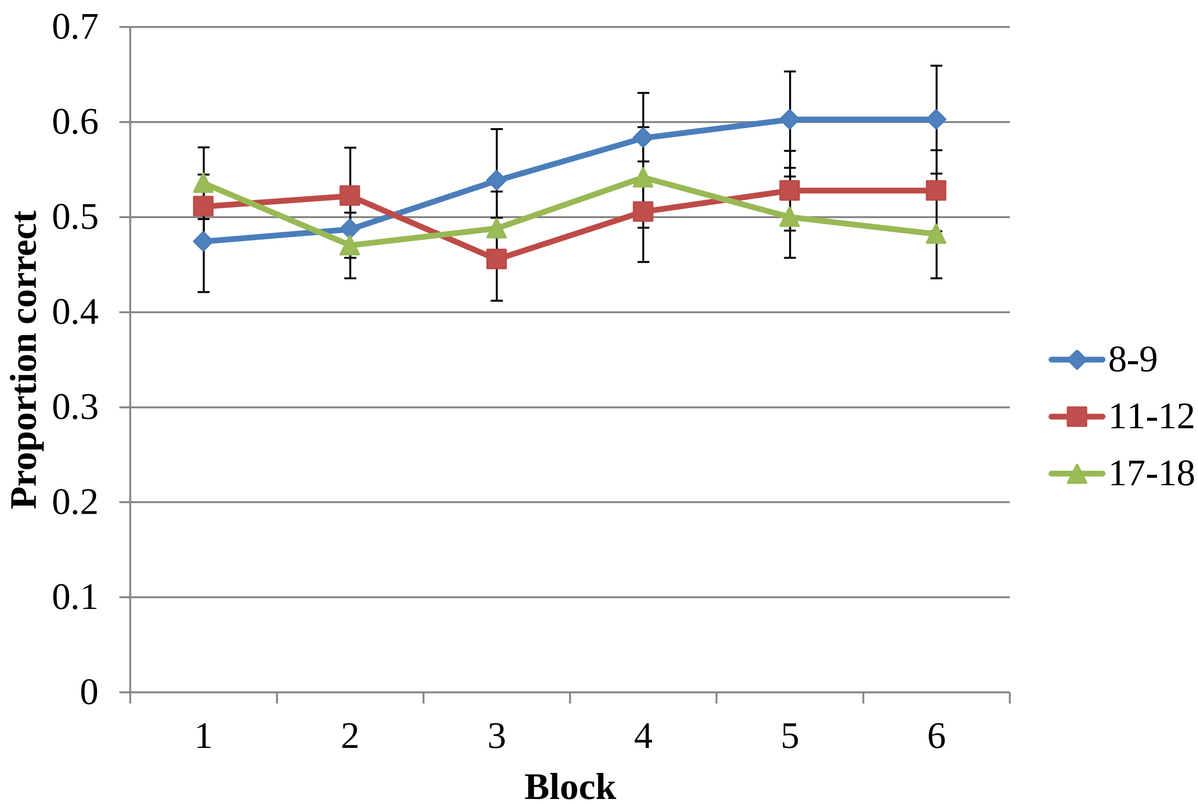
- Subjects see one dynamic scene and hear a sentence.
- Task: Decide, as quickly and accurately as possible, whether sentence sounds “good” or “bad” (in relation to the previous sentences).
- Patterns: SOV, OSV vs *SVO, *OVS, *VSO, *VOS



	Block							
	1	2	3	4	5	6	7	8
Block type	Exp	Exp	M	ST	Exp	Exp	M	ST
Nr of trials	16	16	40	16	16	16	40	16

Results

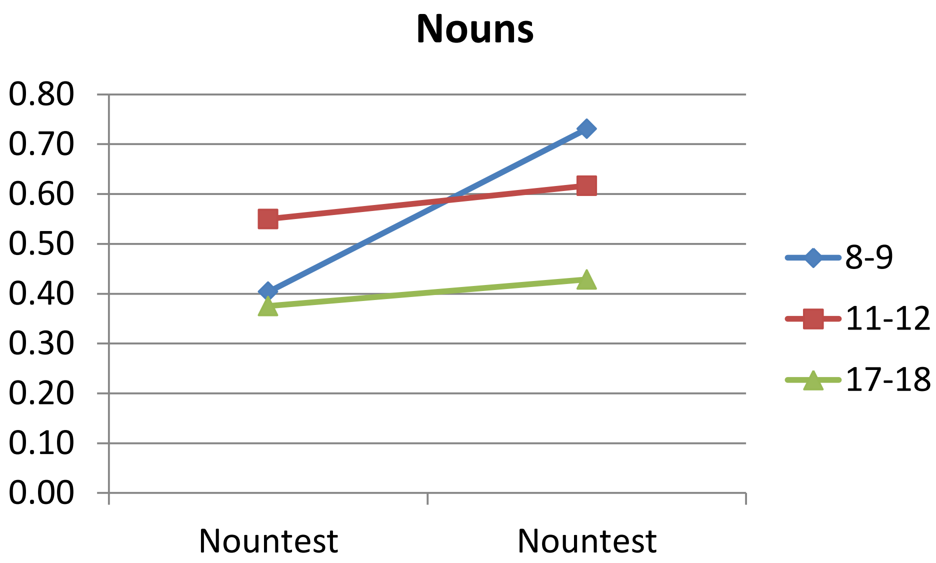
Performance on training trials



- Performance not sig above chance across blocks.
- No effect of training in CSL task. → More exposure necessary.

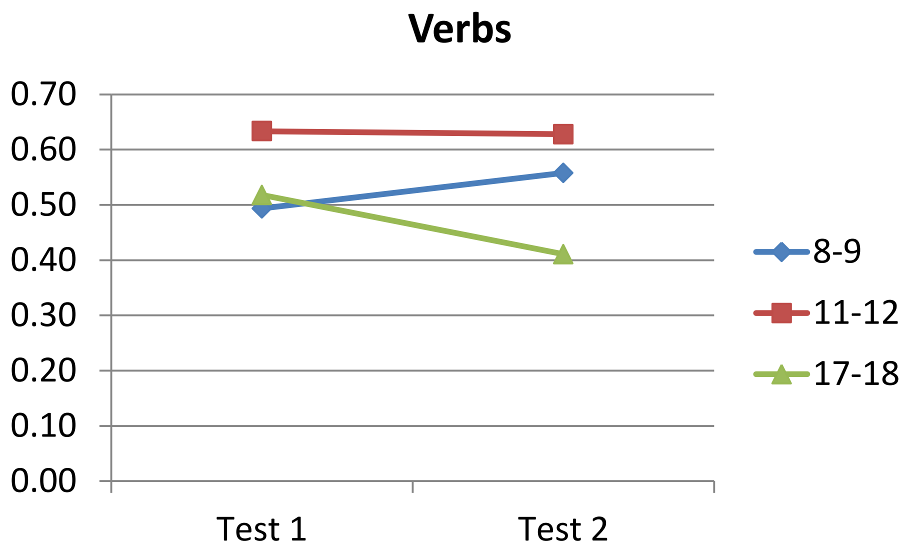
Note: In sequence learning, absence of training effect is common when exposure period brief (e.g. Destrebecqz, 2004).

Performance on noun tests



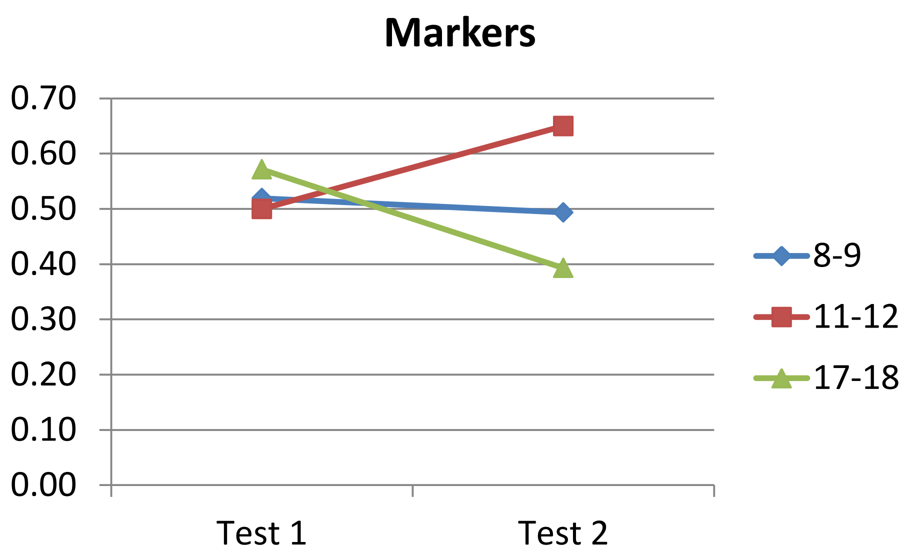
- Sig learning effect in 8-9 year olds and 11-12 year olds.
- Sig advantage for younger learners over 17-18 year olds.

Performance on verb tests



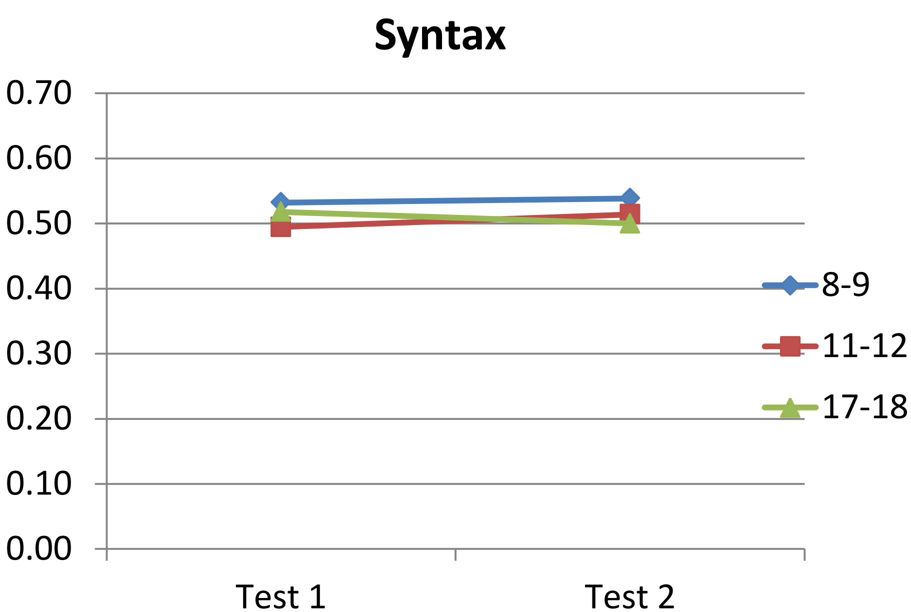
- Learning effect only for 11-12 year olds.
- 11-12 year olds sig outperform 17-18 year olds.

Performance on marker tests



- Learning effect only for 11-12 year olds.
- 17-18 year olds get sig worse.
- Sig difference btw 11-12 and 17-18 year olds.

Performance on syntax tests



- No evidence of learning across groups.

Discussion

Discussion

We investigated children, adolescents learning real Japanese via cross-situational learning paradigm (without feedback)

What have our participants learned?

- 8-9 year olds: Nouns
- 11-12 year olds: Verbs > nouns, marker words
- 17-18 year olds: (...)
- 11-12 year old outperformed older learners
- Nobody learns syntax...

Discussion

Rebuschat, Schoetensack, & Monaghan (in prep)

- Adults learning pseudowords and Japanese syntax
- Acquisition sequence:

  Syntax and verbs > nouns > adj > markers.

Rebuschat, Farrimond, & Monaghan (in prep)

- Children and adolescents learning real Japanese
- Only 11-12 year olds show same acquisition sequence as adults, except syntax

Discussion

Why are our participants doing worse?

- Lexicon was simpler (no adjectives, only 6 nouns) but they received less exposure.
- Difference could be due to reduced exposure.
  - 72 training trials (child study)
  - 192 training trials (adult study)

Discussion

Lancaster University

Why is the acquisition sequence different?

- Adult study: Syntax and verbs > nouns > adj, markers
- 11-12 year olds: Verbs > nouns, markers. No syntax!
- 8-9 year olds: Nouns only
- 17-18 year olds: No learning (in the right direction)

- Noun advantage well documented in developmental literature so there is expectation that they should do well with nouns.
- But: Verbs are very prominent in this language → Sentence final, associated with overt movement on screen.
- Surprising that only 11-12 year olds learn this.

Discussion

Lancaster University

Why is the acquisition sequence different?

- Absence of syntax learning effect surprising.
- Adults learn basic word order rapidly, typically strongest learning effect for verbs and syntax
- Here, 11-12 year olds show learning effect for verbs but chance performance on syntax tests.

- Perhaps use different test to measure syntactic development? → L2 grammaticality judgments could be more challenging for younger learners (literature suggests 3-5 year olds can do L1 grammaticality judgments).

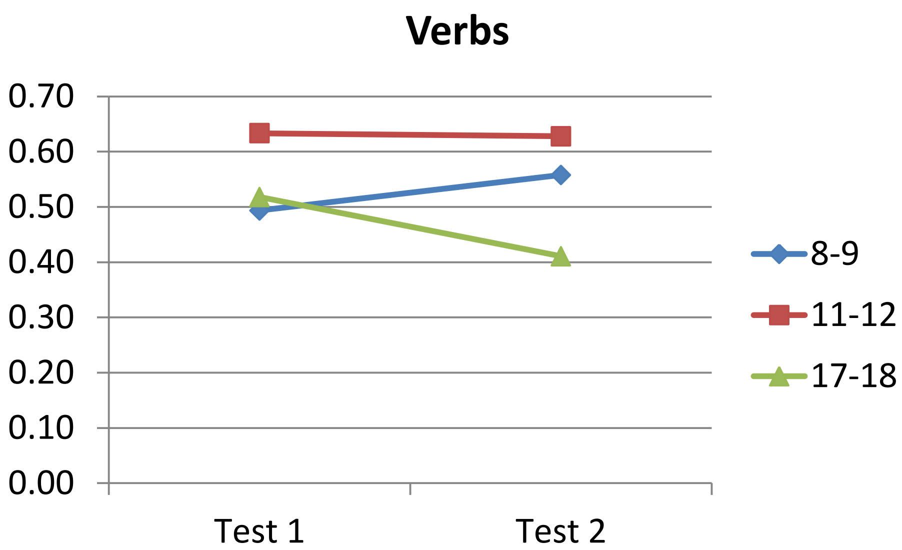
Discussion

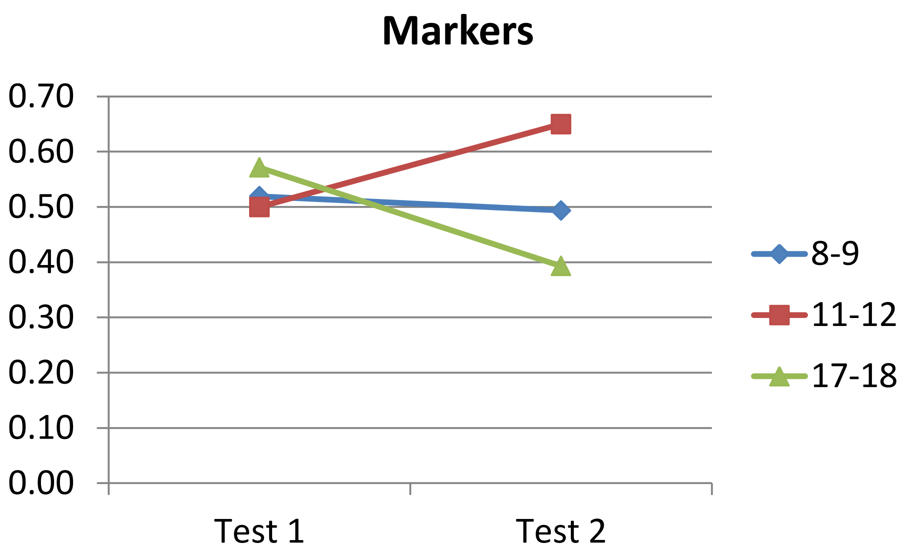
Why are the 17-18 year olds not outperforming the younger learners?

- Expectation was that 17-18 year group should perform very similar to adult subjects in previous studies (mean age ≈ 20)
- Instead: "Sweet spot" for performance in 11-12 year old children.

- Janacsek et al (2012): Sequence learning across the lifespan → Strongest performance between ages 4 and 12, decline afterwards.
- Interference from prior knowledge more likely in older learners, e.g. L1 and metalinguistic knowledge.

Different strategy use?





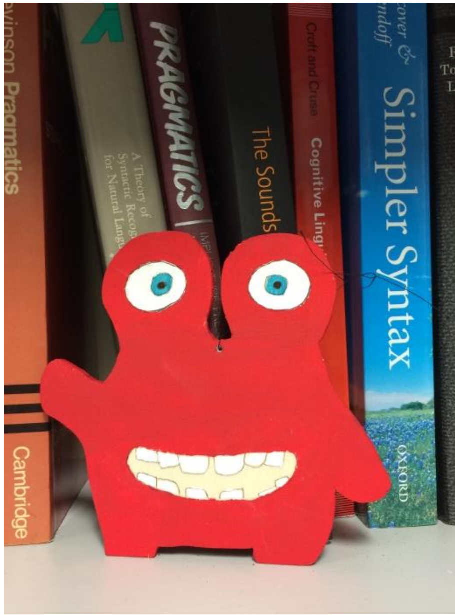
Could older learners use different strategies?
→Explicit hypothesis-testing

Explicit learning works well in simpler systems...

Next steps

- Follow-up studies with mini-Japanese
- Focus on ages 10-13 years
- Increased exposure: 18 training and testing blocks, over two days

Thanks!



Padraic Monaghan

Helena Farrimond,

Katharina Braungart , Christine Schoetensack,

Neil Walker.

AMLaP 2017

Lancaster 7-9 September 2017

http://wp.lancs.ac.uk/amlap2017

Invited Speakers:

- Jeff Elman (UCSD), Susan Goldin-Meadow (Chicago)
- Florian Jaeger (University of Rochester)
- Núria Sebastian-Galles (Universitat Pompeu-Fabra)