Comparing generalisation in children and adults learning an artificial language

Elizabeth Wonnacott [1], Helen Brown [1] , Kate Nation [2]
University of Warwick [1], University of Oxford [2]

INTRODUCTION

Successful language acquisition involves generalization, but learners must balance this against the acquisition of lexical constraints. Examples occur throughout language. For example, English native speakers know that certain noun-adjective combinations are impermissible (*e.g., strong winds, high winds, strong breezes, *high breezes*). Another example is the restrictions imposed by verb sub-categorization (*e.g., I gave/sent/threw the ball to him; I gave/sent/threw him the ball; I donated/carried/pushed the ball to him; * I donated/carried/pushed him the ball*; Baker, 1979). A central debate has been the extent to which learning such patterns depends on semantic cues (Pinker, 1989) and/or distributional statistics (Braine et al., 1990).

The current experiments extend previous work which used Artificial Language learning to demonstrate that adults (Wonnacott et al., 2008) and 6 year olds (Wonnacott, 2011) are able to learn lexically based restrictions on generalization using distributional statistics. Here we directly compare the two age groups learning the same artificial language, with a view to exploring maturational differences in language learning. In addition to manipulating frequency (across high and low frequency items) and quantity of exposure (across days), languages were constructed such that a word's semantic class was helpful for learning the restrictions for some types of lexical items, but potentially misleading for others.

QUESTIONS

1. Will children over-generalize trained nouns more than adults?
   Hudson-Kam & Newport (2009): children show more regularization,
   Boyd & Goldberg (2011): children are more conservative in learning a new structure
2. Will both groups nevertheless show the same distributional effects of *frequency* (across words) and *quantity of exposure* (across days)?
3. Will they show different learning and influence of semantic cues?

METHOD: DESIGN

Participants: 30 Year 1 children (6-years-old). 30 adults (undergraduates, U. of Warwick)
All monolingual English speakers

Language Paradigm (following Wonnacott, 2011)

<u>*Vocabulary*</u>

1 verb	8 nouns	2 particles
moop	*bee, frog, camel* *car, bus, helicopter*	*dow, tay*
"THERE ARE TWO…"	*(borrowed from English)*	*NO SEMANTICS BUT OBLIGATORY IN NP*

<u>*Sentences*</u>

moop + noun + particle e.g. *moop camel dow*
moop camel tay

some nouns alternate between the two others are restricted to occur with just one

Structure of Input Language

half of participants assigned to each version (vocabulary randomized)

half of the time particle1=dow particle2=tay, and vice versa

	VERSION 1	VERSION 2	
PARTICLE 1 - ONLY NOUNS (2)	VEHICLES	ANIMALS	***semantics consistent***
PARTICLE 2 – ONLY NOUNS(2)	ANIMALS	VEHICLES	***semantics inconsistent***
ALTERNATING NOUNS (4)	ANIMALS	VEHICLES	

1 high frequency
1 low frequency
(*3 less freq.)

METHOD: PROCEDURE

4 Day Procedure

Day 1	Training on input language	Copy back 24 sentences *e.g.,* see  hear "*moop car tay*" Experimenter encourages participants to repeat aloud. No other instructions
	Trained nouns test	Produce sentences with the nouns used in training *e.g.,* see  hear "*moop…*" Participant is asked to say the whole sentence
	Novel nouns test	Same procedure as familiar nouns Tests one unseen vehicle and one unseen animal
Days 2 and 3	Training *2 Novel nouns test (no trained nouns test)	
Day 4	Training Old nouns test Novel nouns test	

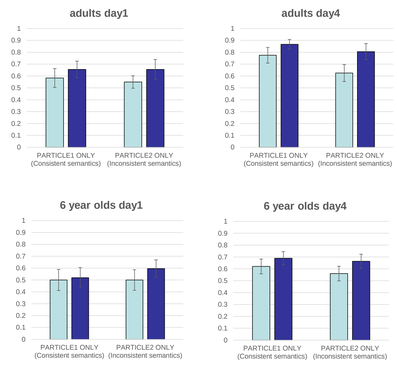
Note, different novel nouns were tested each day

RESULTS: TRAINED NOUNS

***Note*:** "error data" where participants didn't produce the correct noun followed by EITHER dow OR tay are excluded. (1% adult data, 10% child data).

Alternating nouns: dow/tay produced around 50:50 across days and groups (not included in analyses)

Analyses: Linear mixed effect model predicting whether the correct (i.e., attested) particle was used



Findings:

- children over-generalize more than adults z=3.3 p<0.005
- more-generalization with low frequency words z=2.9 p<0.005
- more-generalization on day 1 than day 3 z=4.5 p<0.001
- no sig. (or near sig.) interactions with age

- no main effect of semantic type (p>0.3) *but*
  sig. interaction type*frequ. z=2.44, p<0.05
  marginal three way interaction type*frequ.*age z=1.65, p<0.1

- adults overgeneralize inconsistent semantics items more with low frequency verbs (z=2.22 p<0.05; 12% difference)

low frequ *high frequ.*

RESULTS: NOVEL NOUNS

Analyses: linear mixed effect model predicting whether use the particle associated (consistently or inconsistently) with that semantics.

pooled across 4 days

	consistent	inconsistent
adults	67%	69%
children	51%	55%

Findings:

- adults use predicted semantics more than children z=3.36,p<0.001
- children's particle use not sig. different from chance z=0.86,p=0.39
- adult's particle use IS sig. different from chance z=4.18, p=2.93e-05
- no main effect of day z=0.74, p>0.2
- consistent semantics no more strongly associated with particle1 than inconsistent is with particle2 z=0.74, p>0.2

Acknowledgements:
This work was supported by a grant from the Economic and Social Research Council awarded to the first and third authors, and a British Academy Fellowship awarded to the first author..

Contact: the first author at *e.a.wonnacott@warwick.ac. uk*

SECOND EXPERIMENT

How does performance compare to a matched language with *no* semantic cues?

METHOD

	VERSION 1	VERSION 2	NEW LANGUAGE
particle1-only nouns (2) (1 high; 1 low frequ.)	VEHICLES	ANIMALS	ANIMALS
particle2-only nouns (2) (1 high; 1 low frequ.)	ANIMALS	VEHICLES	ANIMALS
alternating nouns (4)	ANIMALS	VEHICLES	ANIMALS

n = 15 adults
15 children

RESULTS

	ADULTS		CHILDREN	
	low freq.	high freq.	low freq.	high freq.
consistent semantics	68%	76%	56%	60%
inconsistent semantics	59%	73%	53%	62%
all animals	59%	69%	51%	58%

Analyses: LME models comparing productions of attested particles with one-particle nouns in the *all animals* language with (i) particle1-only nouns with *consistent semantics* and (ii) particle2-only nouns with *inconsistent semantics* from the the previous experiment.

No effect of condition in either analysis and no sig. interaction of age with any other factor (p's>0.2).

No evidence that learning of inconsistent semantics hinder learning.

CONCLUSIONS

1. Given matched input, adults show less overgeneralization than 6-year olds, and this effect remains true after four training sessions.
2. Both groups overgeneralize more with low frequency than high frequency items.
3. Both groups improve with increased exposure.
4. Adult show clear evidence of having generalized the semantic cues with novel words, children do not.
5. Semantic cues influence the learning of the constraints with low frequency items. However we do not see evidence that "unhelpful" semantics lead to worse learning compared to control language with no semantic cues.

FUTURE EXPERIMENTS

New experiments are exploring learning of a language with entirely consistent semantic cues

Will children now pick up on semantic cues? Will it aid their learning of trained nouns?

particle1-only nouns (4) (2 high; 2 low frequ.) — VEHICLES

particle2-only nouns (4) (2 high; 2 low frequ.) — ANIMALS

Data collection is ongoing