Constraining Generalisation in Artificial Language Learning: Children are Rational Too

Elizabeth Wonnacott,[1] Amy Perfors.[2] University of Oxford [1], University of Adeleaide[2]
elizabeth.wonnacott@psy.ox.ac.uk

1. Abstract

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*). Such lexical exceptions have been considered problematic for acquisition: if learners generalize abstract patterns to new words, how do they learn that certain specific combinations are restricted? (Baker, 1979). Certain researchers have proposed domain-specific procedures (e.g. Pinker, 1989 resolves verb sub-categorization in terms of subtle semantic distinctions). An alternative approach is that learners are sensitive to distributional statistics and use this information to make inferences about when generalization is appropriate (Braine, 1971).
A series of Artificial Language Learning experiments have demonstrated that adult learners can utilize statistical information in a rational manner when determining constraints on verb argument-structure generalization (Wonnacott, Newport & Tanenhaus, 2008). The current work extends these findings to **children** in a different linguistic domain (learning relationships between *nouns* and *particles*). We also demonstrate computationally that these results are consistent with the predictions of domain-general *hierarchical Bayesian model* (cf. Kemp, Perfors & Tenebaum, 2007).

2. Background

Wonnacott, Newport & Tanenhaus (2008) (henceforth WNT) conducted a series of **Artificial Language Learning experiments** in which adult participants were exposed to miniature languages with two competing synonymous transitive constructions. Verbs in these languages were arbitrarily constrained as to whether they occurred in just one or both structures, with no semantic or phonological cues to verb-type.

WNT Central questions*:*

Do learners acquire verb-specific and verb-general statistical patterns?
What factors affect the tendency to generalize a verb to a new construction not encountered in the input?

WNT Central Findings:

Learners acquire both verb-specific and verb-general statistical patterns (i.e. learned the likelihood of encountering a particular structure both with a given verb and with verbs in general).

The tendency to use a verb in a new structure was affected by:
- Verb frequency (less likelihood of generalizing a more frequent verb
- Extent to which construction usage was lexically determined across the language as a whole

The last was particularly obvious when comparing the treatment of very low frequency *'minimal exposure'* verbs by learners of different languages.

WNT argued that learners were utilizing statistical information in accordance with its utility/ relevance in the past – i.e. showing *rational statistical learning*.

The need for Artificial Language Learning experiments with children
-First language acquisition primarily occurs in early childhood.
- Language learning (first and second) is generally more successful when it begins in early childhood (Newport 1990)
- Adults may use conscious learning strategies unavailable to children.

The *difficulty* of Artificial Language Learning experiments with children
-Generally fewer and shorter sessions learning sessions are practical.
-Children are *slower* than adults in early stages of second language learning (Snow & Hoefnagle-Hohle, 1978)
-Pilot work suggests WNT video paradigm inappropriate for learning mappings between event structure and word-order. (Ongoing work explores alternative methodologies - e.g. live act out. Watch this space!)

Aim of current work: To explore factors affecting balance between generalization and lexically-specific learning with children in a new linguistic domain (similar to that used in previous Artificial Language Learning experiments with children – see Hudson-Kam & Newport, 2005).

Questions:

Are the rational statistical learning procedures in WNT also relevant to child learning?
Critically, is the tendency to generalize affected by:
(a) lexical frequency
(b) the extent to which the language as a whole exhibits lexically based patterns

Note – initial aim is to explore *purely distributional learning* - will use languages with no relevant semantic/phonological cues

3. Experiments

Participants:

44 children recruited from Year 1 classrooms (mean age 6 years).
11 children assigned to learn each of 4 input languages (below)

All monolingual native English speakers.
Pseudo-random assignment matching ages across conditions.

Language Paradigm

Vocabulary

8 nouns ("borrowed" from English) (4 in input, 4 reserved for testing)	*cat, giraffe, pig, dog, cow, crocodile, mouse*
1 verb	*moop* "THERE ARE TWO…."
2 particles	*dow, tay* NO SEMANTICS BUT OBLIGTORY IN NP

Sentences

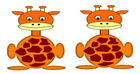
moop + noun + particle — e.g. *moop giraffe dow* / *moop giraffe tay*

noun-particle co-occurrences can be manipulated.

"THERE ARE TWO GIRAFFES *particle*"

Training Procedure
(in 2 * 15 minute sessions over 2 consecutive days)

Each session began with picture labelling:
e.g. *see* 
say : "cat"  "giraffe"

Then *sentence practice:*
e.g. *see*  *hear:* *"moop giraffe tay"*

Experimenter encourages children to repeat aloud.
No other instruction.

4 Input languages (one for each of four groups)

Generalist Language	Lexicalist Language
- 4 *'alternating' nouns* 75% dow; 25% tay	-3 *dow*-only nouns -1 *tay*-only nouns
Mixed language 1	**Mixed language 2**
- 2 *'alternating' nouns* 50% dow; 50% tay -1 *dow-only* noun -1 *tay-only noun* all nouns equally frequent	- 2 *'alternating' nouns* 50% dow; 50% tay -1 *dow-only* noun -1 *tay-only noun* constrained nouns 3x as frequent

No semantic or phonological cues to noun type.

Input languages differ in extent to which usage of particles is lexically determined

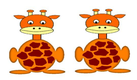
Noun-general usage of two particles matched in lexical and generalist languages (75% dow bias)

dow-only and tay-only nouns are:

matched in frequency in Lexicalist and Mixed Language 1

**3 more frequent in Mixed Language 2.*

Testing Procedure : Production Test

e.g. See  hear.: *"moop"*

Children asked to say the whole sentence.
Score – use of *dow* versus *tay*

NB – break mid testing for more sentence practice which provided the 'minimal exposure'

Input Languages and Test Types follow WNT . Note: new nouns explore generalization of noun-general patterns (entirely novel nouns) and how learners deal with very small amount of language specific input (minimal exposure nouns)

Test items for Production Test (identical for all groups)

Familiar nouns (from the input)

New nouns 3 types: 2 Entirely novel.
1 Minimal-exposure *-dow* not in input but presented in four sentences just before test, always dow in each sentence
1 Minimal-exposure *–tay* not in input but presented in four sentences just before test, always tay in each sentence

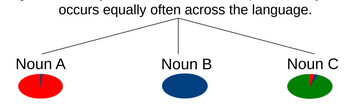
4. Computational Model

Hierarchical Bayesian Models (HBMs) can explain the computational principles that allow structure variability to be learned

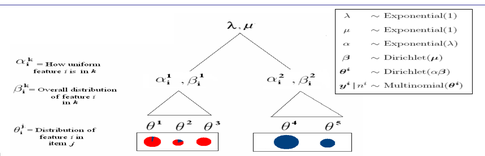
HBMs learn on multiple levels simultaneously

Level 2: General inferences about the range of particle distributions across the language
Level 1: Knowledge about particle distributions of individual nouns

e.g., Nouns may tend to occur with one particle only. Each occurs equally often across the language.



(Different colors represent the different proportion of time each noun occurs in a different construction)



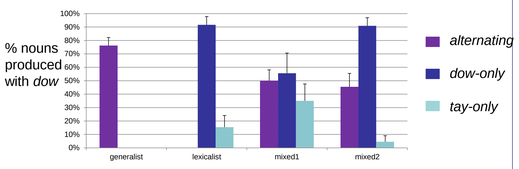
Inference is performed on multiple levels simultaneously: Level 1 knowledge about the construction distribution of specific nouns (represented by the θs); Level 2 knowledge about the nature of constructions in the language as a whole (represented by α and β); and Level 3 priors about the nature of that knowledge (represented by λ and μ).

In current work: the model was given the equivalent input to human participants *except* that minimal exposure nouns are heard only **once**.

References :
Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry* 10, 533-581.
Hudson-Kam, C.L., Newport, E.L. (2005) Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development, 1*, 151-195.
Kemp, C., Perfors, A., Tenenbaum, J.B. (2007) Learning overhypotheses with hierarchical Bayesian models. *Developmental Science* 10, 307-321
Newport, E. (1990) Maturational constraints on language learning. *Cognitive Science, 14,* 11-28
Snow, C.E., & Hoefnagel-Hohle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development,* 49, 1114-1128
Wonnacott, E., Newport, E.L., & Tanenhaus, M.K. (2008). Acquiring and processing verb argument Structure: Distributional learning in a miniature language. *Cognitive Psychology* 56, 65-209.

5. Results

Experimental Data: Productions probabilities with familiar nouns



Note: Sentences with incorrect nouns or no particle are excluded (these make up approx 10% of the data) so 0% = 100% tay.

Significant effect of noun type in all languages.

***alternating* nouns:**

Production probabilities **match input statistics**
(aprox. 75% *dow* in Generalist language 50 % *dow* in Mixed Language 1 and 2)

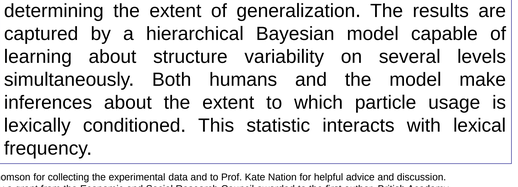
***dow-only* and *tay-only* nouns:**

Production probabilities **reflect lexical constraints** *but*

Significantly more lexical learning in Lexical Language than in Mixed Language 1 → **influence of presence of alternating nouns**

Significantly more lexical learning in Mixed Language 2 than in Mixed Language 1→ **influence of lexicial (noun) frequency**

Productions probabilities with novel nouns

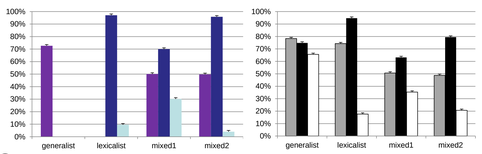


Significant effect of noun-type in lexical language, marginal effect in other languages

***Entirely novel* nouns:**
Production probabilities match input statistics – note *not* associated with these particular nouns (noun-general statistic).

Minimal exposure* nouns*:
In Generalist and Mixed languages usage of particles primarily influenced by noun-general statistic (little influence of 4 exposures).
In Lexicalist language primarily influenced by 4 noun-specific exposures (little influence of noun-general statistic)

Modeling data: "Production" probabilities with familiar and novel verbs.



Summary:
Model qualitatively replicates critical aspects of human performance. (N.b slightly more influence of 'lexical' constraints *but* model not subject to memory limitations – to be followed up!)

6. Conclusions

Like adults in previous studies, children in these experiments show *rational statistical learning* when determining the extent of generalization. The results are captured by a hierarchical Bayesian model capable of learning about structure variability on several levels simultaneously. Both humans and the model make inferences about the extent to which particle usage is lexically conditioned. This statistic interacts with lexical frequency.

Acknowledgements:
Many thanks to Jennifer Thomson for collecting the experimental data and to Prof. Kate Nation for helpful advice and discussion.
This work was supported by a grant from the Economic and Social Research Council awarded to the first author, British Academy Fellowship awarded to the first author and a NSF Graduate Fellowship awarded to the second author.