Higher order inference in verb argument structure acquisition.

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

1. Abstract

Successful language learning combines generalization and the acquisition of lexical constraints. The conflict is particularly clear for verb argument structures, which may generalize to new verbs (*John gorped the ball to Bill ->John gorped Bill the ball),* yet resist generalization with certain lexical items *(John carried the ball to Bill -> *John carried Bill the ball).* The resulting learnability "paradox" (Baker 1979) has received great attention in the acquisition literature. Wonnacott, Newport & Tanenhaus 2008 demonstrated that adult learners acquire both general and verb-specific patterns when acquiring an artificial language with two competing argument structures, and that these same constraints are reflected in real time processing. The current work follows up and extends this program of research in two new experiments. We demonstrate that the results are consistent with a hierarchical Bayesian model, originally developed by Kemp, Perfors & Tenebaum (2007) to capture the emergence of feature biases in word learning.

2. Background: Wonnacott, Newport & Tanenhaus (2008)

(Henceforth WNT)

WNT conducted a series of **Artificial Language Learning experiments** in which adult participants were exposed to miniature languages with two competing transitive constructions:

VSO and ***VOS-ka*** (where 'ka' is a particle with no references).
(Note: the two structures are **synonymous)**

For example:
flugat = BEE, *blergen* = LION, *glim* = RAM so

glim blergen flugat (VSO) = LION RAM BEE
glim flugat blergen ka (VOSka) = LION RAM BEE

Participants **learned the language aurally** (i.e. viewed video clips and heard sentences) in 5 short sessions **over 5 days.**

The **distribution of verbs and constructions was manipulated** across various experimental conditions.

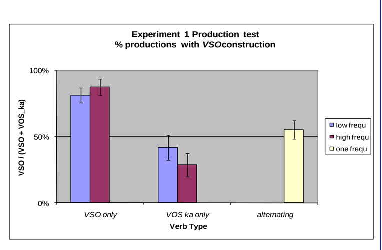
<u>WNT Results</u>

Learners found to acquire both verb-specific and verb-general patterns .

Verb-specific patterns:

- Participants learned that certain verbs were (arbitrarily) constrained to occur in one of the two structures.

- These constraints also affected real-time processing.



Verb-general patterns:

- Participants were also able to *generalize* verbs to structures with which they did had not occurred in the input.

- Tendency to generalize could be manipulated:

1) More likely to generalize with ***low frequency*** verbs (see above figure).

2) Tendency to generalize also affected by higher level distributional information: the extent to which verbs across the language alternated between structures. This was particulary seen by comparing the treatment of very low frequency 'minimal exposure' verbs in the context of different linguistic environments.

Result (2) was tentative due to small subject numbers, but is a potentially important finding. Such higher level learning – i.e. here about variability across the language -- has proved important in other domains, and is the focus of recent computational research (Kemp, Perfors & Tenenbaum 2007). Our new work replicates and extends the critical experiments from WNT and explores constraints on structure variability both experimentally (2 .– across) and computationally (3. – below).

2. Experiments

Method.

Input to Learning

Two groups of learners each learn one of two new Semi-Artificial Languages.

Both Languages:
Vocabulary
8 verbs (monosyllabic nonsense words referring to transitive actions – as in WNT)
5 nouns: (using *English* vocabulary "bee" "lion" – a methodological change from WNT)
1 particle ("ka" – as in WNT)

Structures:
VSO and *VOS-Particle* (as in WNT).

Example sentences: *glim lion bee* *glim bee lion ka* both mean LION RAMS BEE

NB: Assignment of particular verb to verb type randomized across subjects

<u>Languages differ in distributional structure:</u>

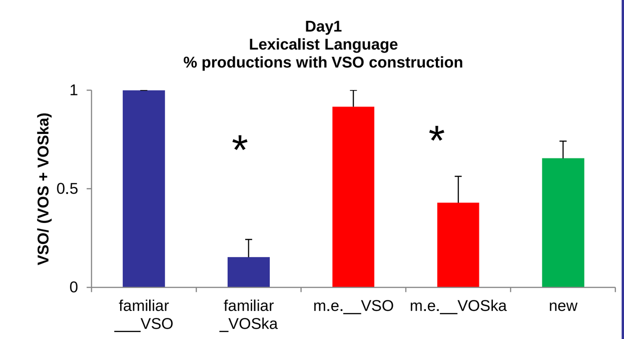
Lexical Language: 4 VSO-only verbs; 4 VOS-ka only verbs.
Generalist Language: 8 biased alternating verbs: 4 VSO 'biased' , 4 VOS-ka 'biased'
(where a 'biased' verb occurs 85% of time in in biased structure)

Exposure: aural exposure (as in WNT) but with testing at end of one 40 minute session
(i.e. 1 day procedure, compared to 5 days in WNT)

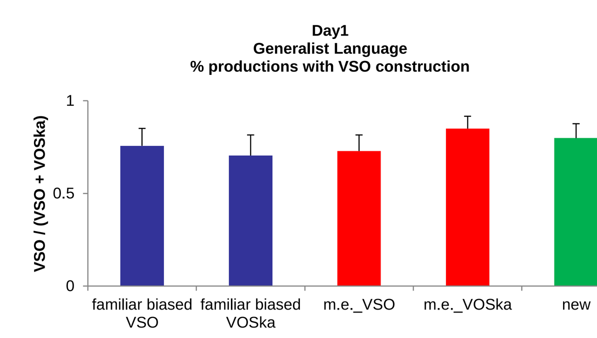
Testing (Production): Learners see scene, hear the first word (verb) and complete the sentence. Four different verb types tested

Familiar verbs : occurred in sentences presented during exposure
2 New verbs : did *not* occur in sentences presented during exposure
1 Minimal Exposure VSO verb: did *not* occur in sentences presented during exposure, just prior to test presented in 4 VSO sentences
1 Minimal Exposure VOSka verb: did *not* occur in sentences presented during exposure, just prior to test presented in 4 VOSka sentences

Results: ***Familiar Verbs***:
Lexical Lang.: strong lexical learning
Generalist Lang: verbs produced in both constructions, as in the input, but no learning of statistical verb biases (instead overall VSO bias)



Minimal exposure verbs:
Lexical Lang.: strong lexical learning (though only 4 instances per verb)
Generalist Lang: no lexical learning and verbs both used in both construction



Novel verbs:
BOTH LANGUAGES: overall VSO bias.
We also examined <u>individual subject responses</u>:

Novel verbs in Lexical Language
1/13 participants used both structures with both verbs
12/13 participants evidenced verb consistent pattern
•6 had one consistent VSO verb and one consistent VOSka verb
•5 used VSO consistently with both verbs
•1 used VOSka consistently with both verbs

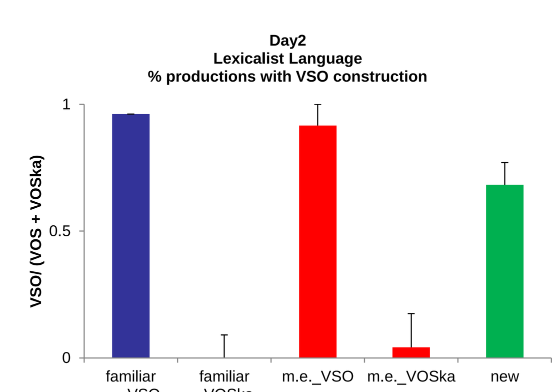
Novel Verbs in Generalist Language
7/12 participants use both structures with both verbs
6/12 participants evidence verb consistent pattern
• *all* 6 used consistent VSO with both verbs

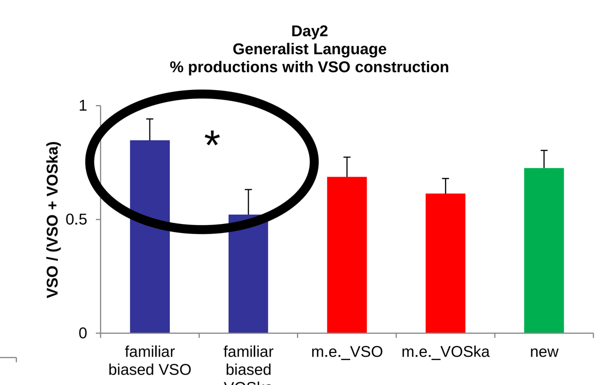
Summary:
• learners can acquire verb-specific restrictions
• also acquire higher level information about ***type of language being learned***

One surprise: participants didn't learn statistical verb bias in generalist language. However, there is clear evidence of the learning of such biases in natural languages (e.g. MacDonald et al 1994).

Participants invited to repeat entire experiment on a second day.

Critical result: familiar verbs in generalist language now show influence of statistical verb bias.

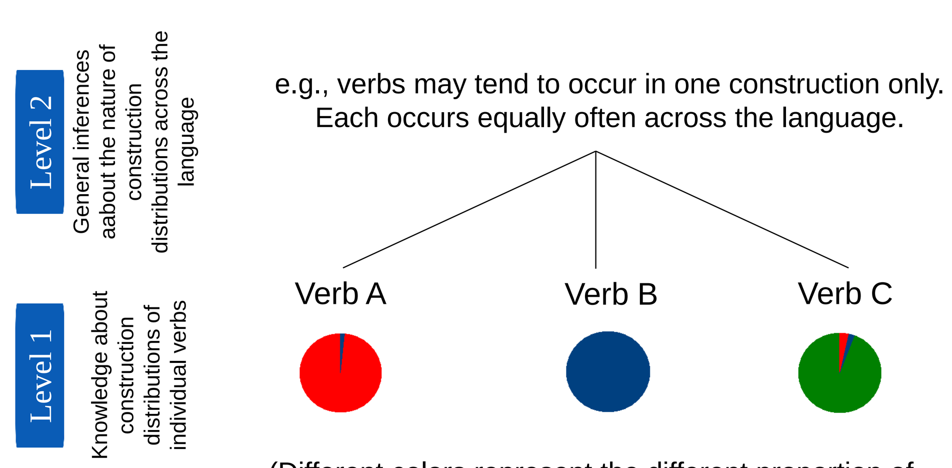




3. Modelling structure variability

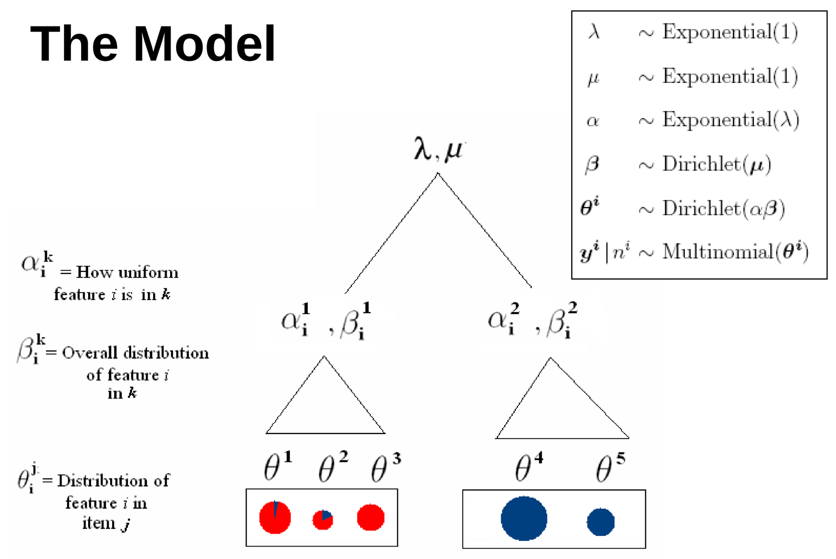
Hierarchical Bayesian Models (HBMs) can explain the computational principles that allow structure variability to be learned

HBMs learn on multiple levels simultaneously



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

The Model

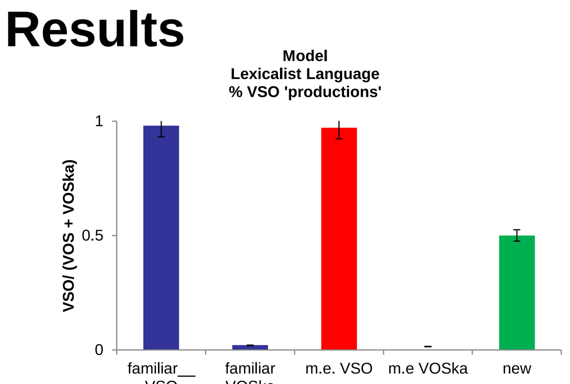


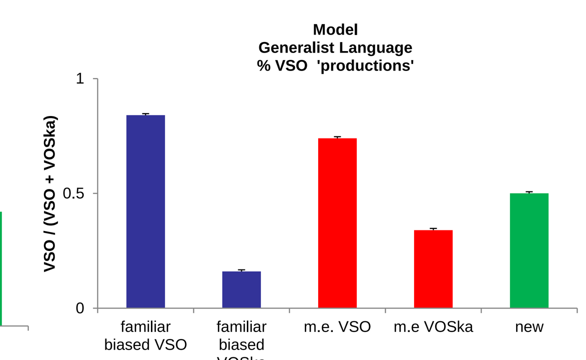
Inference is performed on multiple levels simultaneously: Level 1 knowledge about the construction distribution of specific verbs (represented by the θs); Level 2 knowledge about the nature of constructions in the language as a whole (represented by α and β); and Level 3 priors about the nature of that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

Input to the model

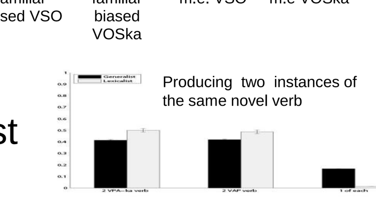
Model given the equivalent input to human participants on day 1 except that minimal exposure verbs heard only **once**.

Results





novel verbs: like humans, model more likely to be lexically consistent in lexicalist language



Summary: Model qualitatively replicates critical aspects of human performance, i.e. the difference in the generalization of minimal exposure verbs across Lexicalist and Generalist languages and the different treatment of novel verbs.

Conclusions

A hierarchical Bayesian model capable of learning about structure variability on several levels simultaneously can capture human performance in this artificial language learning task

Both humans and the model make inferences about construction variation about languages as a whole, and apply those inferences when faced with verbs for which they have very little data

Unlike the model, humans appear to have a bias favoring a VSO construction over a VOS(ka) ordering. We are currently exploring the extent to which this can explain the slight divergences between model predictions and human behaviour

References : - Baker, C. L. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry* 10, 533-581.
- Kemp, C., Perfors, A., Tenenbaum, J.B. (2007) Learning overhypotheses with hierarchical Bayesian models. *Developmental Science* 10, 307-321
- MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101, 676–703
- Wonnacott, E., Newport, E.L., & Tanenhaus, M.K. (2008). Acquiring and processing verb argument Structure: Distributional learning in a miniature language. *Cognitive Psychology* 56, 165-209.

Acknowledgements: Supported by a grant from the Economic and Social Research Council and fellowship from the British Academy. Many thanks to Jennifer Thomson and Julia Florentine for their help collecting the data.