Grammar Competition in Neutral Learning: A Reply to Han et al. (2016)

Henri Kauhanen[1], Caroline Heycock[2] & Joel C. Wallenberg[3]

[1]University of Manchester & Universität Konstanz,
henri.kauhanen@manchester.ac.uk
[2]University of Edinburgh, caroline.heycock@ed.ac.uk
[3]Newcastle University, joel.wallenberg@ncl.ac.uk

FWAV 2017, University of York

Introduction

- Grammar Competition (Kroch, 1989, 1994; Pintzuk and Taylor, 2006, inter alia) and Variational Learning (VL; Yang, 2000, 2002)
  - Is grammar competition a generally expected property of (syntactic) acquisition?
- Poverty of the Stimulus
- **Test Case**: verb-raising in Korean, where the input radically underdetermines a parameter setting.

Introduction

(1) sua sal ye **yure sinnes** les.
so shal you your sins lose
"In this way, you will let go of your sins."
(*Rule of St. Benet*, Yorkshire, date: 1425)

(2) þabbes sal quaintelike drahe **hir** to hir
the-abbess shall wisely draw her to herself
(*Rule of St. Benet*, Yorkshire, date: 1425)

Introduction

- Grammar Competition (Kroch, 1989, 1994; Pintzuk and Taylor, 2006, inter alia) and Variational Learning (VL; Yang, 2000, 2002)
  - Is grammar competition a generally expected property of (syntactic) acquisition?
- Poverty of the Stimulus
- **Test Case**: verb-raising in Korean, where the input radically underdetermines a parameter setting.
  - **Hypothesis (Han et al., 2016)**: if competing grammars is a last resort conclusion for acquirers, and then they will not acquire competing verb-raising grammars in Korean.

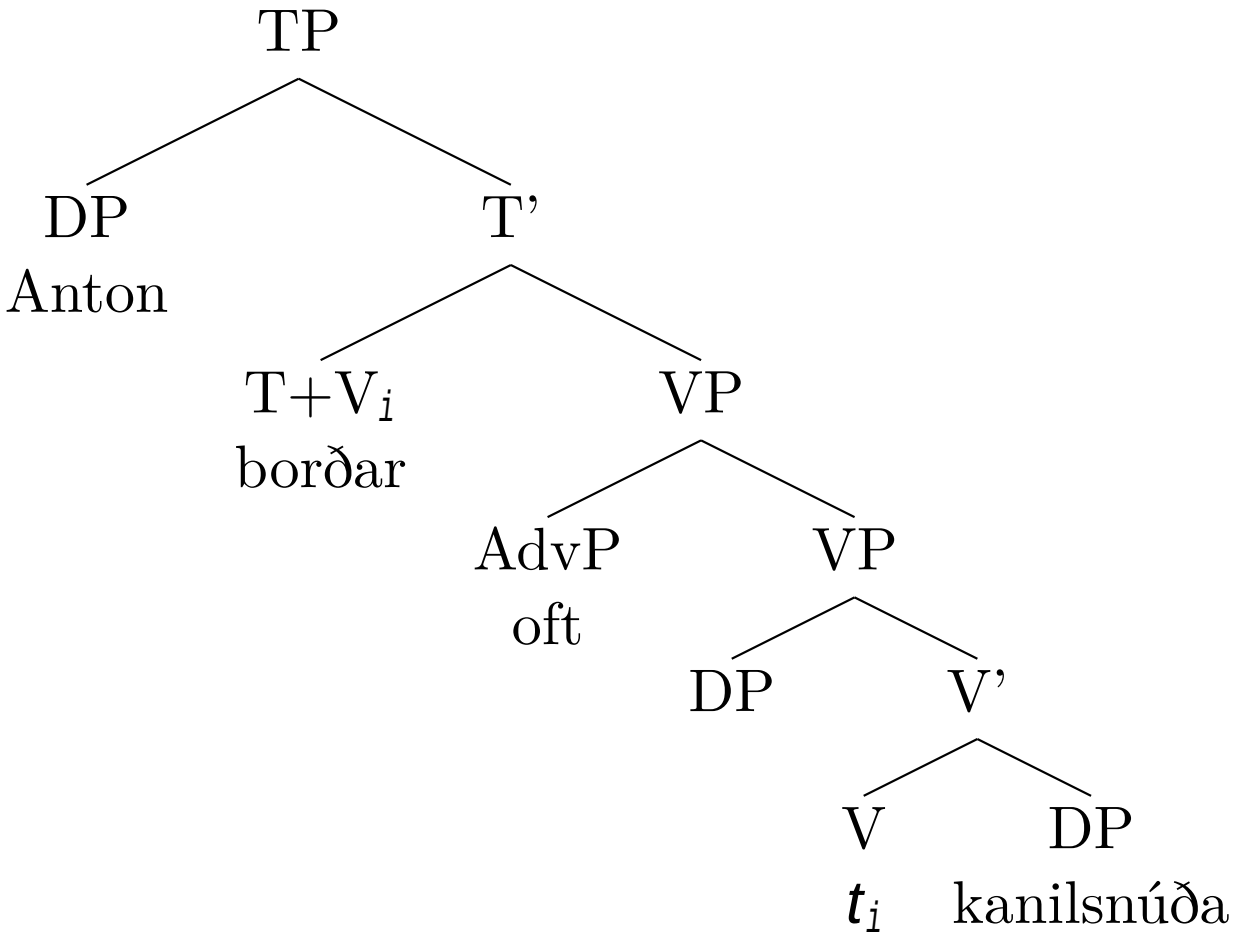
Introduction

We show that – given Variational Learning and finite population dynamics –

- Competing grammars can and do arise **under neutral conditions**.
- This has consequences for the **actuation problem** (Weinreich et al., 1968).

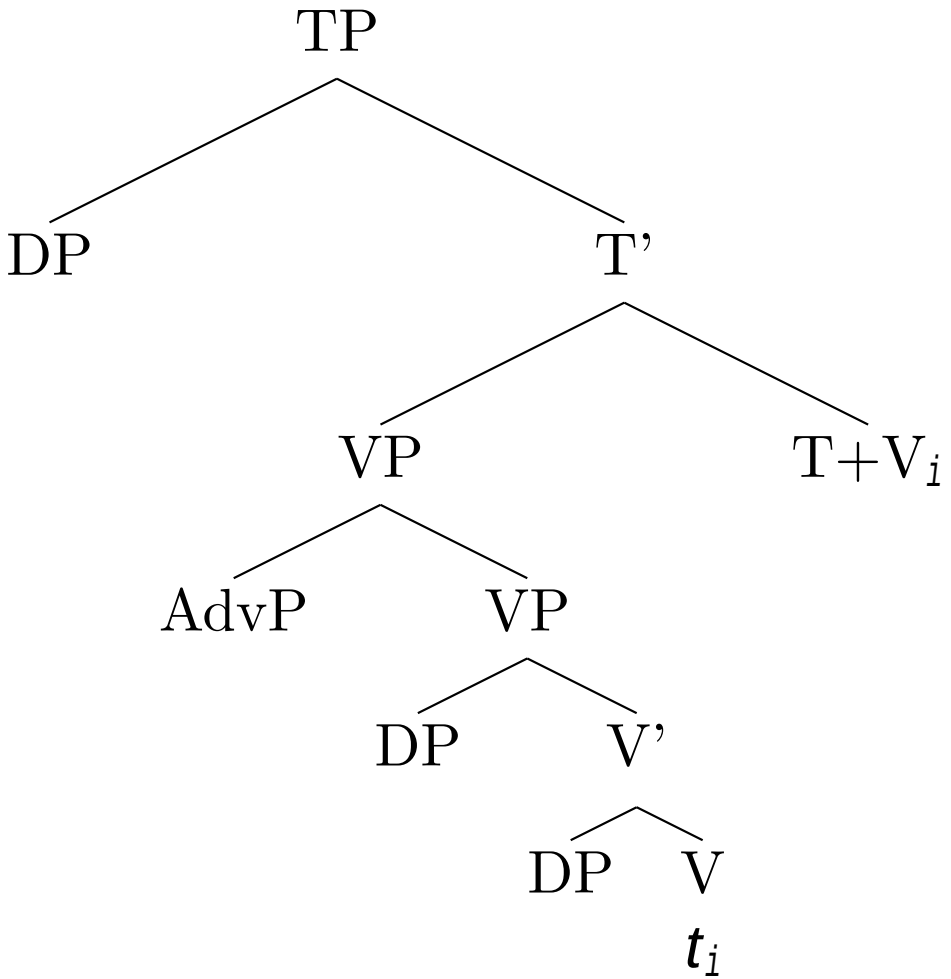
Outline

Left-headed TP



(3) . . . Ulla ofta äter kanelbullar.
Ulla often eats cinnammon rolls (*Swedish*)

Right-headed TP



(4) Yuri-ka cacwu Toli-lul ttayli-n-ta
Yuri-NOM often Toli-ACC hit-PRES-DECL

Outline

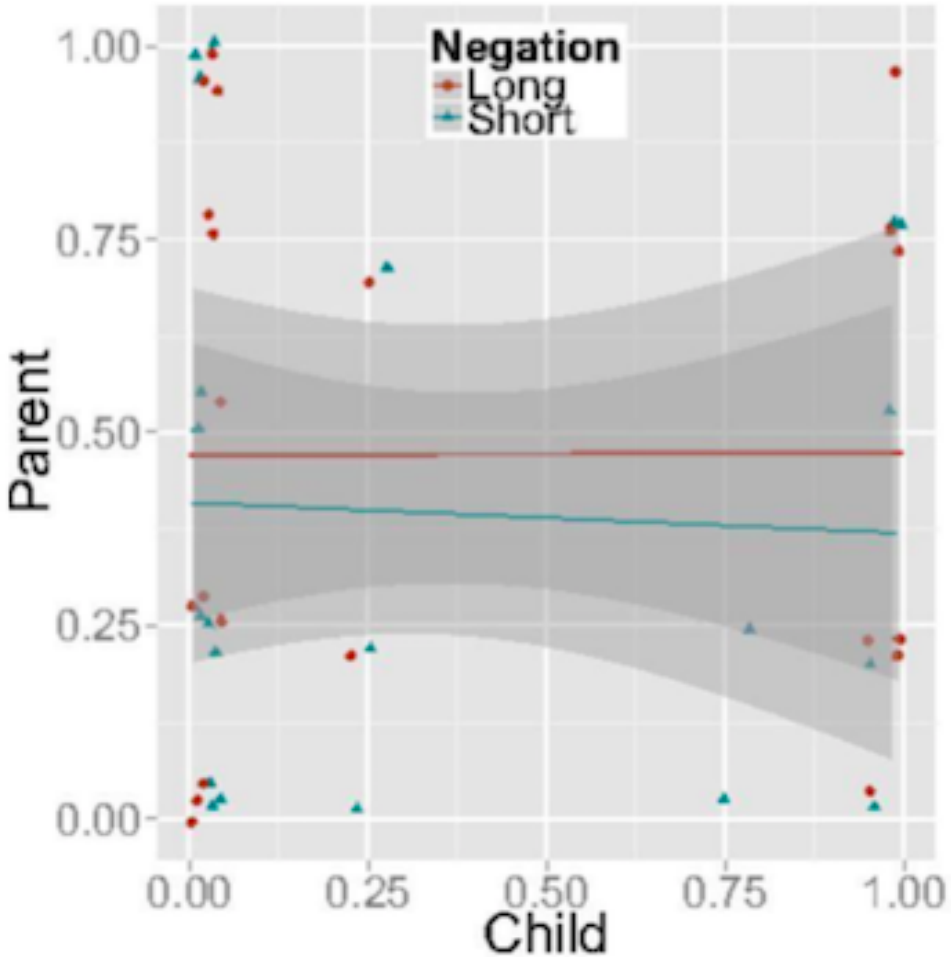
Experiments and Results

- Han et al. (2007): truth-value judgment task (Crain and Thornton, 1998) for adults and children.
- Han et al. (2016):
  1. TVJT for adults.
  2. TVJT for children and mothers of those children.

(5) Khwuki monste-ka motun khwuki-lul an
cookie monster-NOM every cookie-ACC NEG
mek-ess-ta.
eat-PST-DECL
"Cookie monster didn't eat every cookie."

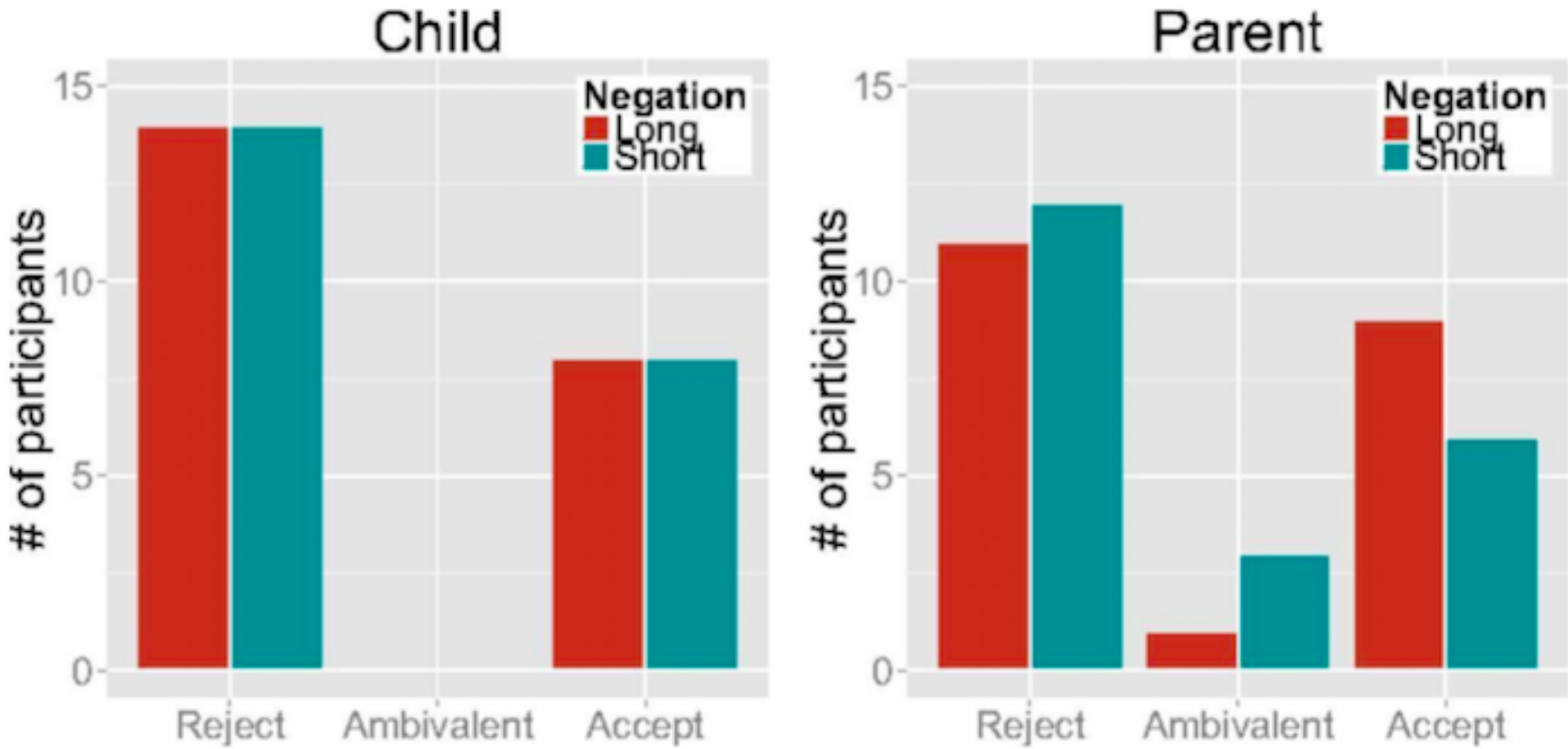
- Main Findings: in the absence of clear input...
  - **Parents' grammar doesn't predict children's.**
  - **The Korean population is split wrt to this syntactic parameter.**

Result (2016): No generational transmission



Chung-hye Han et al. PNAS 2016;113:942-947

Result (2016): competing grammars?



Chung-hye Han et al. PNAS 2016;113:942-947

Interpretation

- **Their interpretation:** in the absence of unambiguous input, acquirers choose one grammar or the other at random.
  - They do not learn both, so no competing grammars in this situation.
  - Competing grammars only arises in the presence of unambiguous evidence for multiple syntactic-parameter settings.
- **Our interpretation:** these learners *are* showing competing grammars in the experiments.
  - They just *tend* to cluster in states of dominance for one grammar as a result of variational learning on finite populations of neutral utterances.

Outline

Yang (2000, 2002)

- Variational Learner (VL): syntactic acquisition consists of learning a set of grammar probabilities.
- A formalization of learning with "competing grammars" (i.e. syntactic parameter-settings), within individuals and across populations.
  - Here we assume just two competing settings G_1 and G_2 (i.e. one binary syntactic parameter)
- Reinforcement learning (the linear reward–penalty learner of Bush and Mosteller, 1958)
  - γ: learning rate
  - N: number of input sentences heard
  - c_i: prob. encountering a sentence not parsed by G_i
- ICBS: with small γ and $N \to \infty$, learner ends up with

$$p_1 = \frac{c_2}{c_1 + c_2} \quad \text{and} \quad p_2 = \frac{c_1}{c_1 + c_2} \tag{1}$$

VL in a neutral setting

- If learner's input contains no unambiguous evidence, $c_1 = c_2 = 0$

  Equation (1) won't work!

- How does VL behave in such a **neutral** setting?
- Insight: the learner becomes a random walk whose characteristics will be given by the two model parameters γ and N.
- Here, we explore neutral VL with simulations assuming finite numbers of neutral utterances (iterations of learning).

Random walks

Random walks

Random walks

Random walks

Random walks

Learning outcome varies from learner to learner.

- Finite iterations of learning means the stop point is crucial.

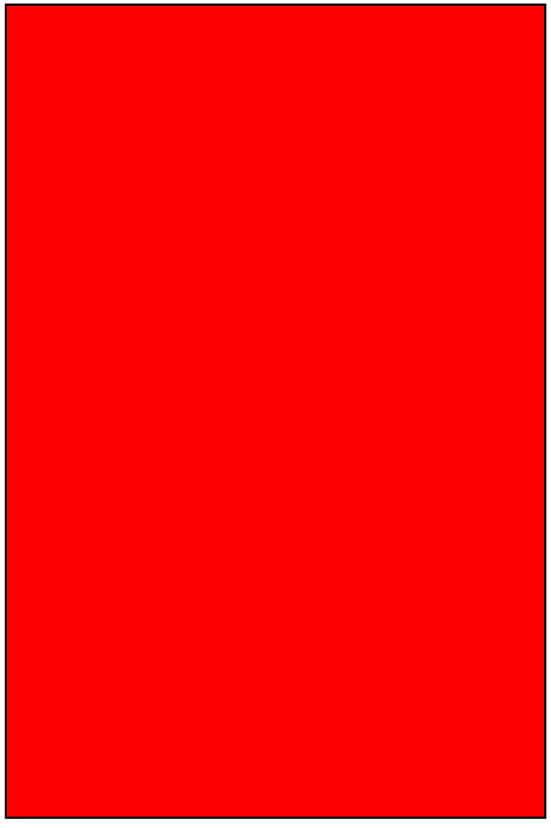
Want to find out the overall average behaviour:

- What is the expected learning outcome?
- How much variance is there about this expectation?

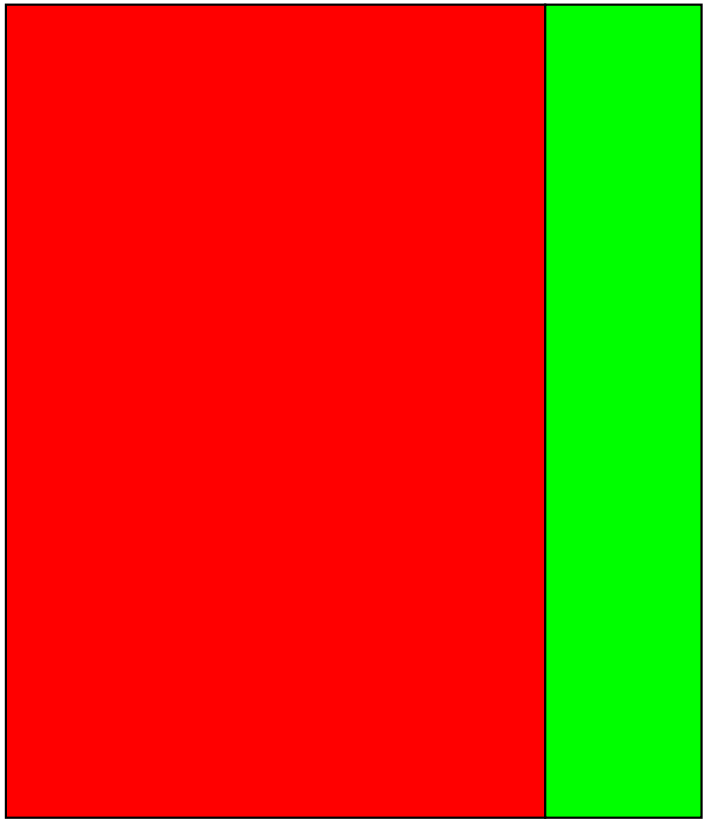
Strategy: set up a large number ($100 \sim 1000$) of such learners; perform a sweep over the N-parameter space; observe learners' terminal states.

Variation in (N = 10^5; 1000 learners per)

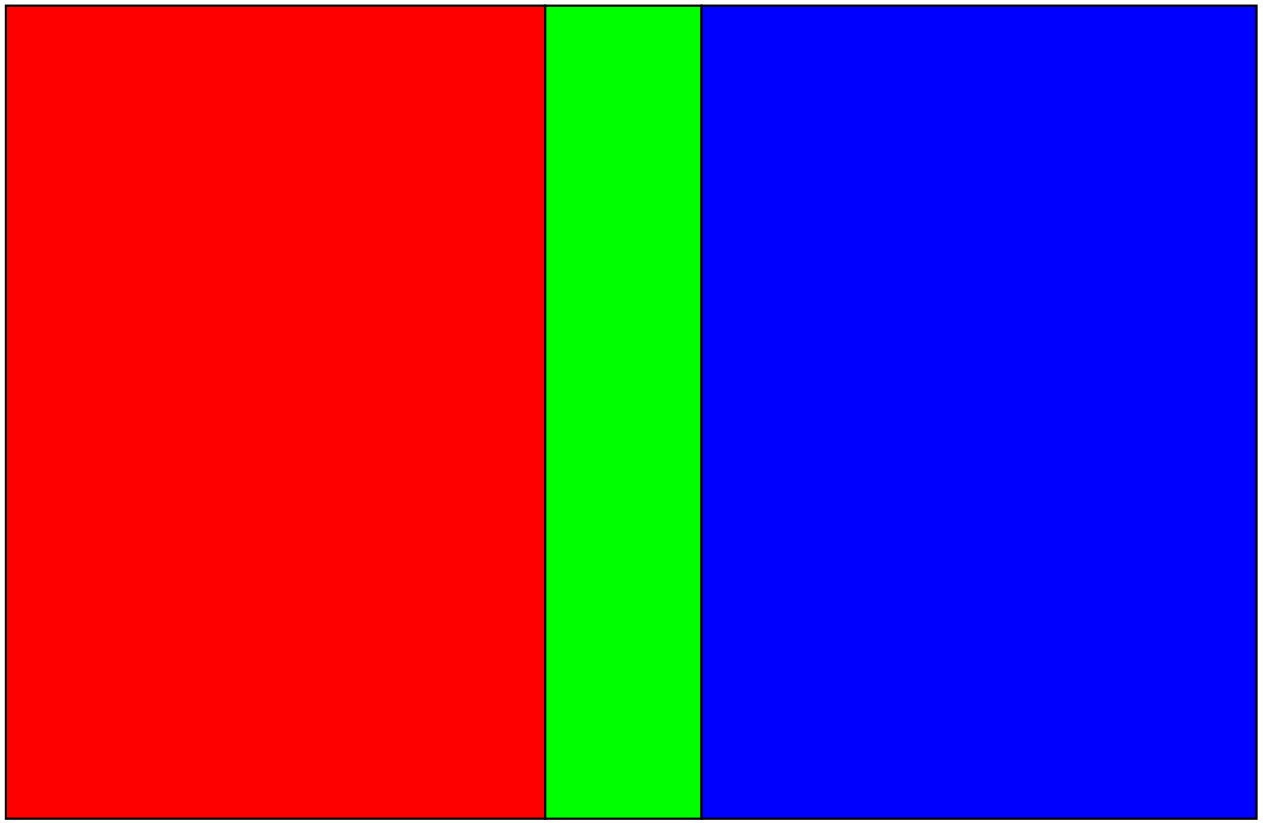
Variation in (N = 10^5; 1000 learners per)



Variation in (N = 10^5; 1000 learners per)



Variation in (N = 10^5; 1000 learners per)



Variation in : Recap

The average outcome (over an entire population) is $p_1 = 0:5$

It is equally likely that a randomly encountered sentence in a population was generated by either grammar

However, individual learner behaviour varies tremendously in response to variation in the learning rate parameter

(1) with low , individual learners end up with p_1 0:5 (random speakers)

(2) with intermediate , individual learners end up with some p_1 from the entire interval $[0; 1]$ (variable speakers)

(3) with large , individual learners end up with either $p_1 = 0$ or $p_1 = 1$ (categorical speakers)

Empirically, the results of Han et al. (2016) lie somewhere between (2) and (3)

Variation in N

Variation in N

Variation in N

Variation in N: Recap

The smaller is, the longer it takes for a learner to become categorical.

However, the states $p_1 = 0$ and $p_1 = 1$ are absorbing, and a random walk is guaranteed to visit one of these states at some point.

Thus, increasing N will have the effect of making more and more learners categorical.

> Conjecture: for any > 0, there exists an N such that learner ends up categorical with probability 1 , for any small > 0

Estimating Model Parameters

To nd out whether VL is consistent with the Korean data, we need empirical estimates of

- the learning rate ( )
- the number of sentences children hear ( )

N is (in principle) not too di cult to estimate, is trickier

Estimating N, Method 1

Shneidman and Goldin-Meadow (2012): ~900 utterances per hour (US, suburban, middle class) ~400 utterances per hour (Mayan, rural).

Han et al. (2016) claim xation on a grammar by age 4 years. Let's assume 3 years (esp. given production lag).

Assuming 12 hours waking time per day, we estimate 11,826,000 utterances for US-type and 5,256,000 for Mayan-type.

Estimating N , Method 2

Human Speechome Project: 12 million words of speech, continuously recorded around a single child from ages 9-24 months.

Vosoughi et al. (2010); Vosoughi and Roy (2012): 2.5 million utterances represent 70% of the child's input for that time range.

Therefore, total input for 9-24 months = 3,571,420 utterances, or ~223213.8 utterances per month.

Over three years = 8,035,697 utterances.

> Very close to the average of the Method 1 Mayan and US estimates, 8,541,000 .

Working Estimates of N

So, we run our final simulations with
5 million N 12 million
sentences.

Estimating

It is currently not known how large an update to p a VL makes per each learning event.

An indirect strategy:

1. Take a well-understood historical change which has been modelled with VL.
2. See what range of learning rates is consistent with that modelling.

   Too small and too large values of make learners fail to converge, and the change is predicted not to have happened, contra facts.
3. Assume true human must lie somewhere within that range.

Estimating

Heycock and Wallenberg (2013) apply VL to the loss of V-to-T in Faroese and Mainland Scandinavian

Strategy:

- Take the parsing advantage parameters estimated by Heycock and Wallenberg (2013) for the V-to-T and V-in-situ grammars.
- Assume 350 years (20 generations) for the change to go to completion (Sundquist, 2002).
- Run a simulation for a range of values, starting from $p_1 = 0{:}01$ (1% use of V-in-situ at point of actuation)
- Observe the nal state after 20 iterated learners; if this is $p_1 = 0{:}99$ (99% V-in-situ) or more, declare change has gone to completion
- Repeat 100 times for reliable statistics

Parameter Estimate Bounds

From the above, we estimate:

	lower bound	upper bound
N	$5{:}0 \quad 10^{6}$	$1{:}2 \quad 10^{7}$
	10^{5}	10^{1}

How does neutral VL behave within these bounds?

Use criterion from Han et al. (2016): speaker is categorical if he/she uses one option at least 75% of the time

N makes little difference (makes sense, since $5{:}0 \quad 10^6$ and $1{:}2 \quad 10^7$ are roughly the same order of magnitude)

Sharp transition from noncategoricity to categoricity in response to variation in

Comparison

For de niteness, assume 0:0005 0:005

And assume Han et al.'s (2016) criteria for categoricity:

reject : 25% sentences
ambivalent : 25% < x < 75% sentences
accept : 75% sentences

How do our neutral VL learners compare to Han et al's empirical data?

Our learners compared to Han et al. (2016)'s

Conclusions

Han et al's results do not show an absence of competing grammars in Korean speakers.

Variational learning in a neutral setting of finite utterances produces speakers with competing grammars.

Realistic learning-parameter values lead to most speakers having a highly dominant grammar.

Han et al's results are compatible with neutral VL for a range of learning rates (over an order of magnitude), which are independently plausible based on non-neutral VL modelling.

The emergence of competing grammars in neutral settings can be thought of as theactuation of new syntactic variants.

Conclusions

Further Work

- Gather stronger empirical bases for estimating N and .
- Explore how robust our results are wrt larger variation in .
- Incorporate further factors, such as population structure, in the modelling, to look at the spread of actuated grammars.
- Analytical results, to con rm our simulation results.

Acknowledgements

Thanks to

Ella and Georg Ehrnrooth Foundation for funding (HK).

the University of Manchester HTCondor pool for CPU time.

Betsy Sneller and Charles Yang for helpful discussion.

References I

Bush, Robert, and Frederic Mosteller. 1958. Stochastic models for learning. New York: Wiley.

Crain, Stephen, and Rosalind Thornton. 1998. Investigations in universal grammar: a guide to research on the acquisition of syntax and semantics.

Han, Chung-hye, Je rey Lidz, and Julien Musolino. 2007. V-raising and grammar competition in korean: Evidence from negation and quanti er scope. Linguistic Inquiry 38:1 47.

Han, Chung-hye, Julien Musolino, and Je rey Lidz. 2016. Endogenous sources of variation in language acquisition. Proceedings of the National Academy of Sciences 113:942 947.

References II

Heycock, Caroline, and Joel C. Wallenberg. 2013. How variational acquisition drives syntactic change: the loss of verb movement in Scandinaviation. *The Journal of Comparative Germanic Linguistics* 16:127–157.

Kroch, Anthony S. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1:199–244.

Kroch, Anthony S. 1994. Morphosyntactic variation. In *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on Variation and Linguistic Theory*, ed. K. Beals et al.

Pintzuk, Susan, and Ann Taylor. 2006. The loss of OV order in the history of English. In *Blackwell handbook of the history of English*, ed. A. van Kemenade and B. Los, 247–278. Blackwell.

References III

Shneidman, Laura A., and Susan Goldin-Meadow. 2012. Language input and acquisition in a mayan village: how important is directed speech? *Developmental Science* 1–16.

Sundquist, John D. 2002. Morphosyntactic change in the history of the mainland Scandinavian languages. Doctoral Dissertation, Indiana University.

Vosoughi, Soroush, Brandon C Roy, Michael C Frank, and Deb Roy. 2010. Effects of caregiver prosody on child language acquisition. In *Proceedings of the 5th International Conference on Speech Prosody*. Speech Prosody Special Interest Group (SProSIG) of the International Speech Communication Association (ISCA).

References IV

Vosoughi, Soroush, and Deb K Roy. 2012. An automatic child-directed speech detector for the study of child language development. In *Proceedings of the 13th Annual Conference of the International Speech Communication Association*. http://hdl.handle.net/1721.1/82526: International Speech Communication Association.

Weinreich, Uriel, William Labov, and Marvin Herzog. 1968. Empirical foundations for a theory of language change. In *Directions for historical linguistics*, ed. W. Lehmann and Y. Malkiel, 95–195. Austin, TX: University of Texas Press.

Yang, Charles. 2000. Internal and external forces in language change. *Language Variation and Change* 12:231–250.

Yang, Charles D. 2002. *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.