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

Henri Kauhanen\textsuperscript{1}, Caroline Heycock\textsuperscript{2} & Joel C. Wallenberg\textsuperscript{3}

\textsuperscript{1}University of Manchester & Universität Konstanz, henri.kauhanen@manchester.ac.uk
\textsuperscript{2}University of Edinburgh, caroline.heycock@ed.ac.uk
\textsuperscript{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.
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

Introduction

Background

  V-to-T Raising

Han et al. (2007, 2016)

  Experiment and Results
  Interpretations

Neutral Variational Learning

  Model
  Estimating Model Parameters
  Comparison

Conclusions
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

Introduction

Background

V-to-T Raising

Han et al. (2007, 2016)

Experiment and Results

Interpretations

Neutral Variational Learning

Model

Estimating Model Parameters

Comparison

Conclusions
Experiments and Results


- Han et al. (2016):
  1. TVJT for adults.
  2. TVJT for children and mothers of those children.

    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
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

Introduction

Background
  V-to-T Raising

Han et al. (2007, 2016)
  Experiment and Results
  Interpretations

Neutral Variational Learning
  Model
  Estimating Model Parameters
  Comparison

Conclusions

• 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)
  - $\gamma$: learning rate
  - $N$: number of input sentences heard
  - $c_i$: prob. encountering a sentence not parsed by $G_i$

• ICBS: with small $\gamma$ and $N \to \infty$, learner ends up with

$$p_1 = \frac{c_2}{c_1 + c_2} \quad \text{and} \quad p_2 = \frac{c_1}{c_1 + c_2} \quad (1)$$
VL in a neutral setting

- If learner’s input contains no unambiguous evidence, $c_1 = c_2 = 0$
  - $\Rightarrow$ 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 $\gamma$ and $N$.
- Here, we explore neutral VL with simulations assuming finite numbers of neutral utterances (iterations of learning).
Random walks

The graph shows the probability $p_1$ over the learning step. The x-axis represents the learning step, ranging from $0e+00$ to $1e+05$, while the y-axis represents the probability $p_1$, ranging from $0.0$ to $1.0$. The data points fluctuate, illustrating the random walk behavior over time.
Random walks
Random walks

![Random walks graph](image-url)
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 ∼ 1000) of such learners; perform a sweep over the $\gamma, N$ parameter space; observe learners’ terminal states.
Variation in $\gamma$ ($N = 10^5$; 1000 learners per $\gamma$)
Variation in $\gamma$ ($N = 10^5$; 1000 learners per $\gamma$)
Variation in $\gamma$ ($N = 10^5$; 1000 learners per $\gamma$)
Variation in $\gamma$ ($N = 10^5$; 1000 learners per $\gamma$)
Variation in $\gamma$: 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 $\gamma$
  
  (1) with low $\gamma$, individual learners end up with $p_1 \approx 0.5$ (**random** speakers)
  (2) with intermediate $\gamma$, individual learners end up with some $p_1$ from the entire interval $[0, 1]$ (**variable** speakers)
  (3) with large $\gamma$, 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 $\gamma$ 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 $\gamma > 0$, there exists an $N$ such that learner ends up categorical with probability $1 - \epsilon$, for any small $\epsilon > 0$
Estimating Model Parameters

- To find out whether VL is consistent with the Korean data, we need empirical estimates of
  - the learning rate ($\gamma$)
  - the number of sentences children hear ($N$)
- $N$ is (in principle) not too difficult to estimate, $\gamma$ 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 fixation 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 ~223,213.8 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.
So, we run our final simulations with $5 \text{ million} \leq N \leq 12 \text{ million}$ sentences.
Estimating $\gamma$

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

- 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 $\approx 350$ years ($\approx 20$ generations) for the change to go to completion (Sundquist, 2002).
  - Run a simulation for a range of $\gamma$ values, starting from $p_1 = 0.01$ (1% use of V-in-situ at point of actuation)
  - Observe the final 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
Prob. of V-in-situ after 20 generations (100 runs per $\gamma$)
Proportion of "successful" simulations

learning rate $\gamma$

1e-08  1e-06  1e-04  1e-02  1e+00
Proportion of "successful" simulations

learning rate $\gamma$
Parameter Estimate Bounds

- From the above, we estimate:

\[
\begin{array}{c|cc}
\text{Parameter} & \text{lower bound} & \text{upper bound} \\
\hline
N & 5.0 \times 10^6 & 1.2 \times 10^7 \\
\gamma & 10^{-5} & 10^{-1} \\
\end{array}
\]

- 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
proportion categorical (75% cutoff)

- \( N \) makes little difference (makes sense, since \( 5.0 \times 10^6 \) and \( 1.2 \times 10^7 \) are roughly the same order of magnitude)
- Sharp transition from noncategoricity to categoricity in response to variation in \( \gamma \)
Comparison

- For definiteness, assume $0.0005 \leq \gamma \leq 0.005$
- And assume Han et al.’s (2016) criteria for categoricity:
  - reject: $\leq 25\%$ sentences
  - ambivalent: $25\% < x < 75\%$ sentences
  - accept: $\geq 75\%$ sentences
- How do our neutral VL learners compare to Han et al.’s empirical data?
Our learners compared to Han et al. (2016)’s

![Bar chart comparing rejection, ambivalence, and acceptance rates between different groups: children, long negation, children, short negation, adults, long negation, adults, short negation, neutral VL, N = 5,000,000, neutral VL, N = 12,000,000. The chart shows statistical significance with p-values ≤ 0.005.]
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 $\gamma$ (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 the actuation of new syntactic variants.
Conclusions

Further Work

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


References II


References III


References IV


