Asymmetrical Generalisation of Harmony Triggers

Wendell Kimper // University of Manchester

wendell.kimper@manchester.ac.uk

10 October 2015 // AMP // Vancouver

Trigger Asymmetries in Vowel Harmony

High and non-high vowels exhibit a **typological asymmetry** with respect to their ability to trigger ROUND (and BACK) harmony — non-high vowels are **better** triggers than their high counterparts.

- e.g. Yakut, where non-high vowels trigger rounding harmony across a broader range of contexts than their high counterparts (Krueger, 1962; Kaun, 1995).

Phonetic grounding: non-high vowels are less articulatorily and perceptually extreme along these feature dimensions (Terbeek, 1977; Linker, 1982; Kaun, 1995).

- These vowels are the ones with the most need for the increase in perceptual salience that harmony confers (Suomi, 1983; Kaun, 1995; Kimper, 2011).

Substantive Bias

Wilson (2006): learners are *biased* in favour of patterns that are phonetically natural.

- Triggering asymmetry in velar palatalisation: higher vowels are better triggers (greater coarticulatory influence).
- Subjects trained on [e] generalised to [i], but not vice versa.

Moreton and Pater (2012a,b) reviewed the evidence for systematic biases in learning.

- Robust evidence for *structural* bias (toward simpler patterns).
- Mixed evidence for *substantive* bias (toward phonetically natural patterns).

Prediction

In BACK/ROUND harmony, **mid** (non-high) vowels are **better** triggers than high vowels. Therefore...

- If we know that high vowels trigger,
  - *we also know that mid vowels trigger.*
- If we know that mid vowels trigger
  - *we don't know what high vowels will do.*

Prediction:

- Subjects trained on a harmony pattern with **high** vowels should tend to make **broad** generalisations
- Subjects trained on a pattern with **mid** vowels should tend to make both **narrow** and **broad** generalisations.

The Experiment

Poverty of the Stimulus: given a partially ambiguous pattern, which generalisation (broad or narrow) do subjects make?

- Similar in design to (Finley, 2008, Experiment 8).

Subjects: 67 native speakers of British English, recruited from the University of Manchester and surrounding community.

- Subjects reported no speech, hearing, or learning disabilities and gave null responses on no more than 10% of trials.
- Subjects received either course credit or a £10 Amazon voucher for their participation.

Experimental groups:

- 33 subjects in the **mid-only** group
- 34 subjects in the **high-only** group

Stimuli: Vowel Sequences

Target pattern: root-controlled BACK/ROUND harmony.

CVCV stems: vowels agreed for both height and colour, but ATR features varied freely.

- Inclusion of the ATR feature dimension was intended to promote feature-based generalisations (cf. Finley 2008).

	Mid		High		
Front/Unround	ɛ e	e ɛ	ɪ i	i ɪ	ATR-disharmonic
	e e	ɛ ɛ	i i	ɪ ɪ	ATR-harmonic
Back/Round	ɔ o	o ɔ	ʊ u	u ʊ	ATR-disharmonic
	o o	ɔ ɔ	u u	ʊ ʊ	ATR-harmonic

Stimuli Formation

Stems: Vowels paired with randomly selected non-identical consonants from the set {p, t, k, s, b, d, g, z, m, n}.

- Items which closely resembled existing English words were replaced by re-sampling.
- 144 stems: 14 for each ATR-harmonic vowel sequence, and 4 for each ATR-disharmonic vowel sequence.

Suffixes: -ge/-go for singular forms, -gi/-gu for plural forms.

Stems and suffixes were recorded separately, then spliced together.

- Stems were recorded with a dummy -gə suffix, and suffixes were recorded with a dummy CəCə stem.
- Initial primary stress, final secondary stress.

Training

Explicit learning: subjects were instructed that they would be learning how to form plurals in Martian.

During training...

- The **mid-only** group only encountered mid-vowel stems.
- The **high-only** group only encountered high-vowel stems.
- **Both** groups encountered both mid and high vowels in suffixes (the -ge/-go singular and -gi/-gu plural).

Subjects were given two rounds of training, each round consisting of passive listening followed by responses with feedback.

Experiment was administered using E-Prime Professional 2.0, using a Serial Response Pad and circumaural headphones, in a sound-attenuated room at the University of Manchester Phonetics Lab.

Training

Part 1 (passive listening): 48 singular/plural pairs, presented auditorily and accompanied by images of various fruits.

Part 2 (response with feedback): Same items as Phase 1, but only plurals presented (auditorily, no visual stimulus).

- Subjects were asked to indicate by pressing a button whether or not the item they just heard was correct.
- Feedback was given after each trial.
- The 'correct' items were previously heard plural forms.
- The 'incorrect' items had BACK/ROUND disharmonic suffixes.
- 50% correct items (counterbalanced).

Test Procedure

Test: Subjects were presented with potential plural forms and asked to indicate via a button press whether or not this was a well-formed Martian plural.

- Similar to Part 2 of training, but without feedback.

All subjects saw the same items in the test phase (the old and new items for one group were the novel items for the other).

- 24 old stems (items seen in training)
- 24 new stems (new items of the same type as training)
  - **Mid** for **mid-only**, **high** for **high-only**
- 48 novel stems (new items of a different type than in training)
  - **High** for **mid-only**, **mid** for **high-only**.

Data Processing

Subjects were categorised as **learners** if they performed better than chance on the **old** items in the test phase.

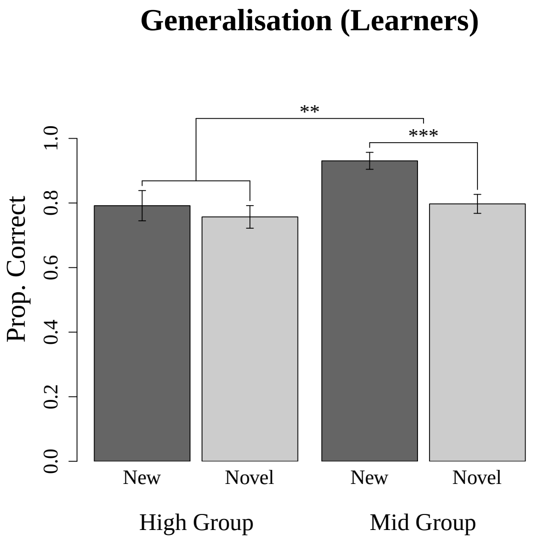
- High failure rate: only 15 learners in mid-only group and 12 learners in high-only.
- No statistically significant difference in performance overall.

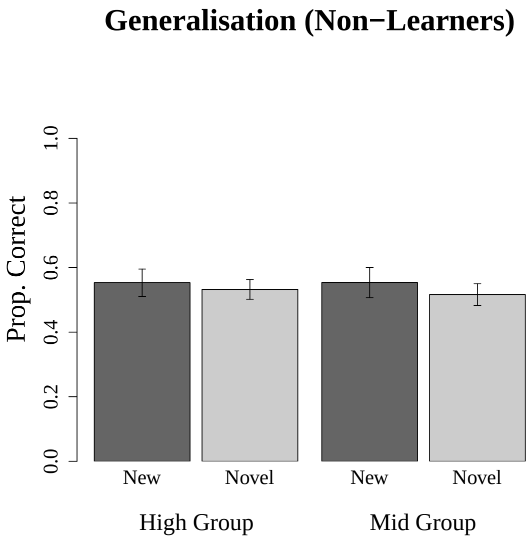
Non-learners serve as a de-facto control group.

- Any differences in baseline preferences between the two groups unrelated to learning should be apparent here.

A generalised linear mixed effects model was fitted on proportion correct for **new** and **novel** responses only, with item and subject as random effects. Sum coding was used for fixed effects.

- To get simple effects, model was re-run on subsets of the data.


Generalisation (Learners)
**
***
Prop. Correct
0.0
0.2
0.4
0.6
0.8
1.0
New
Novel
New
Novel
High Group
Mid Group


Generalisation (Non−Learners)
Prop. Correct
0.0
0.2
0.4
0.6
0.8
1.0
New
Novel
New
Novel
High Group
Mid Group

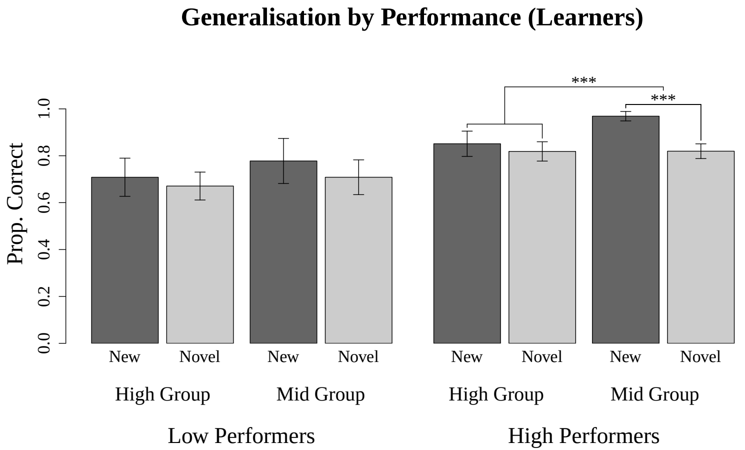
Could this be about performance differences?

Possible alternative explanation: the mid-only group happened to reach a more advanced stage of learning.

- While there were no significant differences overall, the mid-only group performed better ($p < 4.92e\text{-}05$) across crucial test items (new and novel).
- Consistent with the finding in Finley (2008) that mid-vowel learners performed better.
- Could narrow generalisation be a result of better learning?

Subjects categorised into **low performers** and **high performers**

- Low performers $< 80\%$ on **old** items $<$ high performers.


Generalisation by Performance (Learners)
Prop. Correct
0.0
0.2
0.4
0.6
0.8
1.0
***
***
New
Novel
New
Novel
New
Novel
New
Novel
High Group
Mid Group
High Group
Mid Group
Low Performers
High Performers

Explicit Rule Reporting

Subjects were asked during debriefing if they felt they identified the pattern, and to describe what they thought it was.

Was the subject aware of some kind of harmony rule?

Learners

	no	yes
high-only	3	9
mid-only	1	14

Non-Learners

	no	yes
high-only	21	1
mid-only	17	1

Possible explanation for poor learning overall:

- Many reports from non-learners referred to visual properties of the images of fruit (seeds, ‘masculine’ or ‘feminine’ fruit, etc.) rather than phonological properties of the words.

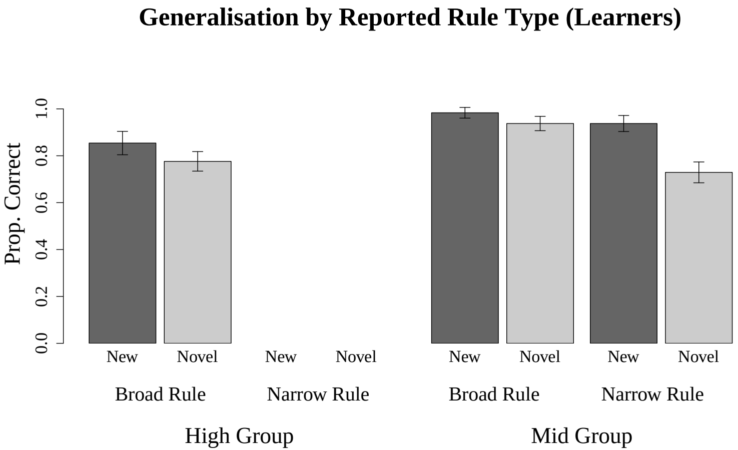
Explicit Rule Reporting

Generalisations described by the learners were classified as either **broad** (applicable to all vowels) or **narrow** (applicable to a restricted feature class).

	broad	narrow	no	vague
high-only	8	NA	3	1
mid-only	5	8	1	1

Some examples...

- **Broad**: *I think the word stems with low vowels - eg. "u" or "o" had the single suffix "o" and the plural suffix "u". Words with "i", "a", or "e" took the "i" plural ending.*
- **Narrow**: *I thought that "o" sounds in the word meant plural was "oo" e.g. nonogoo, & the same with "e" e.g. nenegee.*


Generalisation by Reported Rule Type (Learners)
Prop. Correct
0.0
0.2
0.4
0.6
0.8
1.0
New
Novel
New
Novel
New
Novel
New
Novel
Broad Rule
Narrow Rule
Broad Rule
Narrow Rule
High Group
Mid Group

Summary

The interaction between group (mid-only vs. high-only) and test item type (new vs. novel) was statistically significant ($p < 0.01$).

Learners in the **high-only** group formed **broad** generalisations.

- No difference between new and novel items ($p > 0.05$).
- No narrow rules reported.

Learners in the **mid-only** group formed **both broad and narrow** generalisations.

- Higher generalisation to new items than to novel items ($p < 0.001$), especially for high performers ($p < 0.01$).
- Both broad and narrow rules reported.

Evidence for Substantive Bias?

The results from the experiment are consistent with the predictions of substantive bias.

- Asymmetrical generalisation in the predicted direction, with no difference in complexity.

Why is this surprising?

- The effect was more pronounced among learners who performed well on training data.
  - . . . but van de Vijver and Baer-Henney (2014) found that substantive bias is more relevant under uncertainty.
- Subjects were engaged in explicit learning, and learners were able to articulate the rule they learned.
  - . . . but this should reflect conscious, domain-general strategy (see Moreton and Pertsova 2015 and references therein).

A closer look at the task...

What is the explicit analytical task? In the training data, subjects need to notice that...

- The singular/plural alternation affects the final vowel.
- The singular and plural each have more than one form.
- The form can be predicted based on the preceding vowel.
- The relevant vowels agree in BACK/ROUND.

What subjects don't have access to in explicit training is information governing how **broad** the generalisation should be:

- *Any* natural class containing *all* the segments seen in training?
- The natural class containing *all and only* the segments seen in training?

A closer look at the task...

Determining the scope of the generalisation is not necessarily part of the explicit task.

- Subjects don't know that they will be asked to generalise to stems containing different vowels.

Assumption: subjects differ in their baseline tendency to form broad or narrow generalisations.

- Subjects exposed to the exact same training data formed different generalisations (at least for the mid-only group).

What remains to explain the divergent behaviour of the mid-only and high-only groups? **Substantive bias**.

- Likelihood of narrow generalisation = baseline + bias.

Modelling Biased Learning

Simulation of biased larning using the **Maxext Grammar Tool** (Wilson and George, 2009).

- **Input**: a schematic version of the training data (mid-only and high-only stems separate).

With substantive bias for mid triggers...

- A learner trained on mid vowels formed broad generalisations with a high generality bias, and narrow generalisations with a moderate generality bias.
- A learner trained on high vowels formed broad generalisations *regardless generality bias*.

With no substantive bias...

- Broad generalisations across the board.

Conclusion

The behaviour of learners in the lab reflected typological asymmetries in harmony triggers.

- Subjects trained on only high triggers formed broad generalisations; subjects trained on only mid triggers formed both broad and narrow generalisations.

Embedded in an explicit learning task was a somewhat implicit decision (the scope of the feature-based generalisation).

A Maxent learner with substantive bias in favour of mid-vowel triggers mimicked the behaviour of the human learners better than an unbiased learner.

Thank You!

Acknowledgements

Many thanks to Anie Thompson for lending her voice to the stimuli, and to Research Assistants Michaela Hejná, Ziyuan Pan, and Isabel Maccabee for help in running subjects.

This work has benefited from insightful discussions with Elliott Moreton, Joe Pater, Andrew Stewart, the University of Manchester Phon-Phon Group, and the UMass Amherst Sound Workshop. All errors and oversights remain my own, but typos should be blamed on my imaginary coauthor.

This research was funded by a grant from the School of Arts, Languages and Cultures Research Support Fund at the University of Manchester.

References I

Bates, Douglas, Martin Maechler, Benjamin M. Bolker, and Steven Walker. 2014. lme4: Linear mixed-effects models using eigen and s4. ArXiv e-print; submitted to *Journal of Statistical Software*.

Finley, Sara. 2008. Formal and cognitive restrictions on vowel harmony. Doctoral Dissertation, Johns Hopkins University.

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic In* 39:379–440.

Kaun, Abigail. 1995. The typology of rounding harmony: An Optimality Theoretic approach. Doctoral Dissertation, University of California, Los Angeles.

Kimper, Wendell. 2011. Competing triggers: Transparency and opacity in vowel harmony. Doctoral Dissertation, University of Massachusetts Amherst.

Krueger, John. 1962. *Yakut manual*. Indiana University Press.

References II

Linker, Wendy. 1982. Articulatory and acoustic correlates of labial activity in vowels: A cross-linguistic study. Doctoral Dissertation, UCLA.

Moreton, Elliott, and J Pater. 2012a. Structure and Substance in Artificial Phonology Learning, Part I: Structure. *Language and Linguistics Compass* 686–718.

Moreton, Elliott, and J Pater. 2012b. Structure and Substance in Artificial Phonology Learning, Part II: Substance. *Language and Linguistics Compass* .

Moreton, Elliott, and Katya Pertsova. 2015. Implicit and explicit phonology: What are artificial-language learners really doing? Paper presented at the 23rd Manchester Phonology Meeting.

R Core Team. 2014. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org/`.

Suomi, Kari. 1983. Palatal harmony: A perceptually motivated phenomenon? *Nordic Journal of Linguistics* 6:1–35.

References III

Terbeek, Dale. 1977. A cross-language multi-dimensional scaling study of vowel perception. Doctoral Dissertation, UCLA.

van de Vijver, Ruben, and Dinah Baer-Henney. 2014. Learners' little helper. Paper presented at the Annual Meeting of the Linguistics Association of Great Britain.

Wilson, Colin. 2006. Unbounded spreading is myopic. Bloomington, IN: PhonologyFest 2006.

Wilson, Colin, and Ben George. 2009. Maxent grammar tool. Available at http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/.

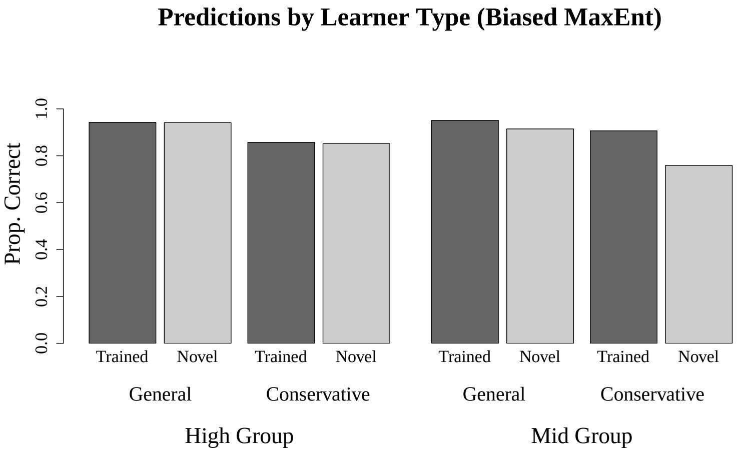
Appendix I: The Full Model

```
Fixed effects:
                        Estimate  Std. Error z value Pr(>|z|)
(Intercept)              1.95424    0.21226   9.207  < 2e-16 ***
group                   -0.92192    0.22710  -4.060 4.92e-05 ***
new.novel               -0.73092    0.17679  -4.134 3.56e-05 ***
performance             -0.84767    0.22239  -3.812 0.000138 ***
atrV2                    0.19250    0.07423   2.593 0.009508 **
colour                  -0.20479    0.07363  -2.781 0.005415 **
explicit.rule1           0.66988    0.24653   2.717 0.006583 **
explicit.rule2          -0.81856    0.31655  -2.586 0.009713 **
explicit.rule3           0.20819    0.36167   0.576 0.564867
group:new.novel          0.51557    0.17719   2.910 0.003619 **
group:performance        0.43164    0.20159   2.141 0.032261 *
new.novel:performance    0.43869    0.15673   2.799 0.005126 **
group:new.novel:perf.   -0.40171    0.15673  -2.563 0.010373 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Model was fitted using the lme4 package (Bates et al., 2014) in R (R Core Team, 2014), with fixed effects and their interactions added one at a time until there was no further improvement in the model fit.

Appendix 2: Biased Learning

	General		Conservative		Unbiased	
Constraint	μ	σ	μ	σ	μ	σ
AGREE (V-V)	0	**2**	0	**0.5**	0	**1**
AGREE (Mid-V)	0	**0.5**	0	**0.5**	0	**0.01**
AGREE (High-V)	0	0.01	0	0.01	0	0.01
AGREE (Front-V)	0	0.01	0	0.01	0	0.01
AGREE (Back-V)	0	0.01	0	0.01	0	0.01
AGREE (ATR-V)	0	0.01	0	0.01	0	0.01
AGREE (RTR-V)	0	0.01	0	0.01	0	0.01
AGREE (i-V)	0	0.01	0	0.01	0	0.01
AGREE (ɪ-V)	0	0.01	0	0.01	0	0.01
AGREE (u-V)	0	0.01	0	0.01	0	0.01
AGREE (ʊ-V)	0	0.01	0	0.01	0	0.01
AGREE (e-V)	0	0.01	0	0.01	0	0.01
AGREE (ɛ-V)	0	0.01	0	0.01	0	0.01
AGREE (o-V)	0	0.01	0	0.01	0	0.01
AGREE (ɔ-V)	0	0.01	0	0.01	0	0.01


Predictions by Learner Type (Biased MaxEnt)
Prop. Correct
0.0
0.2
0.4
0.6
0.8
1.0
Trained
Novel
Trained
Novel
Trained
Novel
Trained
Novel
General
Conservative
General
Conservative
High Group
Mid Group

Predictions (Unbiased MaxEnt)