Generalising from ambiguous data

Karthik Durvasula   Adam Liter

Michigan State University

Introduction

- How do learners generalise from data that is ambiguous between multiple different generalisations?
  - Single *vs.* multiple compatible generalisations?
  - Simplest *vs.* most precise generalisation?
- The above questions lead to some of the possibilities that have been argued for.
  - *Subset Principle* (**SP**; Berwick 1985; Hale and Reiss 2003)
  - *Simplest Generalisation* (**SG**; Chomsky and Halle 1968)
  - *Multiple Simple(st) Generalisations* (**MSG**; Chomsky and Halle 1968; Hayes and Wilson 2008)
  - Learn all compatible generalisations, but simplest preferred (**PropSimple**; Linzen and Gallagher 2014; Linzen and O'Donnell 2015)
  - Learning proportional to specificity (**PropSpec**; Tenenbaum and Griffiths 2001)
- Previous work is inconsistent on this:
  - **SP** (Gerken 2006)
  - **PropSimple** (Linzen and Gallagher 2014)
- For ambiguous input, learners:
  - learn multiple generalisations.
  - don't seem to track more complex generalisations.

General Exp. Design

Training Phase

- Participants listened to and silently mouthed 100 CVCV nonce words (2 repetitions each).
- C=/p,b,t,d,f,v,s,z/   V=/a,i,u/.
- C obeyed both voicing and stop harmony simultaneously.
  - e.g., ✓[tipa, bida, fisa], *[tisa,bipa,fida].

Testing Phase

- Participants were asked if word was possible in the "language" they learned.
- CVCV nonce words of the following types:
  - 12 OldStims
  - 12 NewStims
  - 12 OnlyVoicing
  - 12 OnlyStop
  - 12 NoPattern

Experiment 1

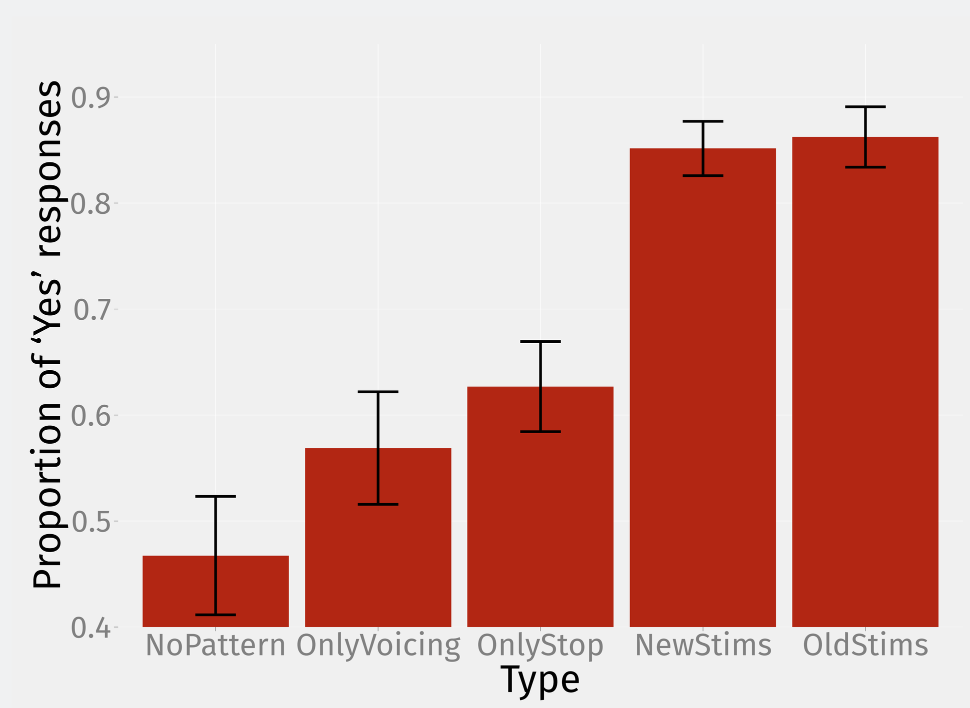
Methods

- 25 English-speaking undergraduates, and 3 were excluded due to non-learning.
- NewStims: C sequences possibly heard in *Training*.

Predictions

- **SP**: NewStims and OldStims preferred over other three (which are undifferentiated).
- **SG**: Some prefer OnlyVoicing, some OnlyStop; NewStims are as acceptable as either. Thus, all three are equally good.
- **MSG**: Both OnlyVoicing and OnlyStop are preferred; additive effect on NewStims.
- **PropSimple**: OnlyVoicing, OnlyStop, and StopVoicing are all learnt; therefore, interactive effect on NewStims. However, interaction effect smaller than either OnlyVoicing or OnlyStop.
- **PropSpec**: OnlyVoicing, OnlyStop, and StopVoicing are all learnt; therefore, interactive effect on NewStims. However, interaction effect larger than either OnlyVoicing or OnlyStop.

Results



Fixed Effect	MeanYes (%)	Estimate	z-value	Pr(>z)
(Intercept)	0.4674	−0.1223	−0.534	0.2968
OnlyVoicing	0.5688	0.4801	2.485	0.0065 **
OnlyStop	0.6268	0.7664	3.897	<0.0001 ***
NewStims	0.8514	2.142	9.05	<0.0001 ***
OldStims	0.8623	2.2292	9.331	<0.0001 ***

Table 1: Logistic mixed-effects models

Fixed Effect	Estimate	z-value	Pr(>z)
(Intercept)	−0.1231	−0.544	0.2934
Voicing	0.4758	2.513	0.0059 **
Stopping	0.7574	3.920	<0.0001 ***
Voicing:Stopping	0.8881	3.032	0.0012 **

Table 2: Logistic mixed-effects model—Interaction effects for new test stimuli

Discussion

- Multiple simple generalisations are learnt for ambiguous data.
- Furthermore, interaction effect suggests potential support for **PropSimple** and **PropSpec**.
- However:
  - Perhaps phonological generalisations can also directly access segmental representations (i.e., segmental primitives) without making reference to the featural content.
  - If so, a generalisation based on a single segment might be as "simple" as a generalisation based on a single feature.
  - Therefore, **MSG** could also account for the interaction effect.

Conclusion

- We found that learners keep track of multiple compatible generalisations in the face of ambiguous data.
- Furthermore, there is evidence that participants keep track of segmental generalisations.
- However, there's no evidence that participants keep track of the "complex" featural generalisations.
- This suggests that "simple":
  - has to be seen as representationally simple (Chomsky and Halle 1968; Hayes and Wilson 2008).
  - has to include both featural and segmental representations.

Experiment 2

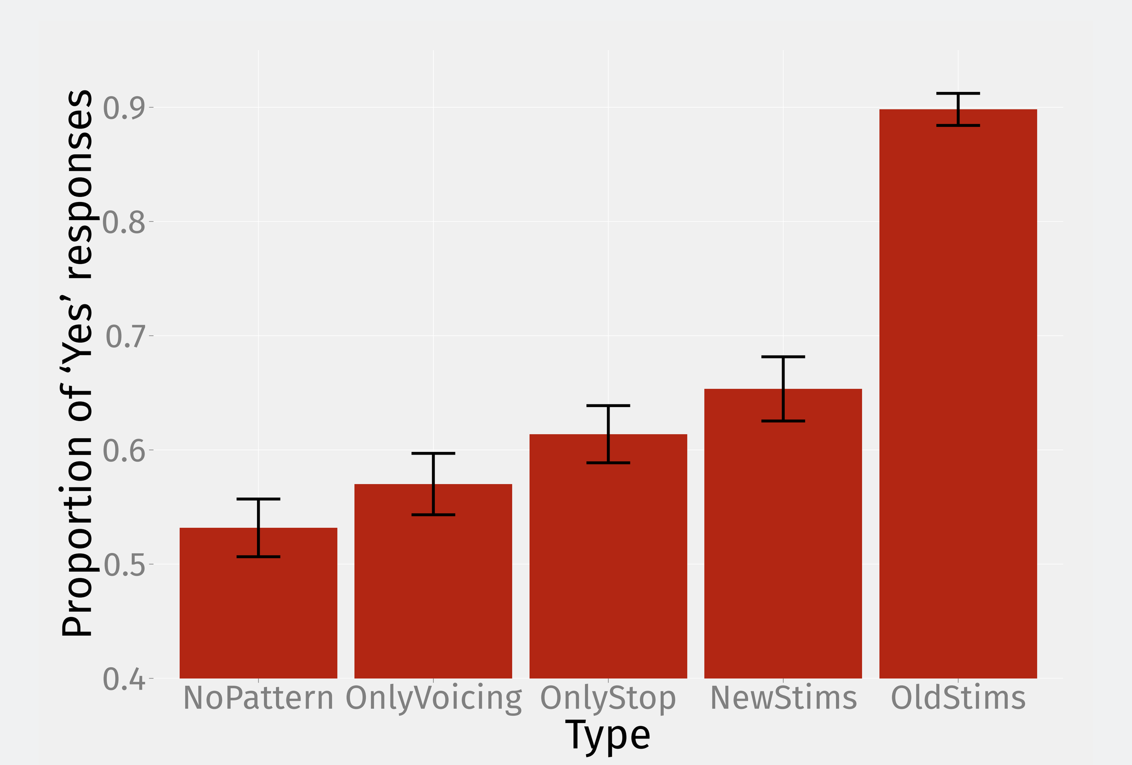
Methods

- 78 English-speaking undergraduates, and 15 were excluded due to non-learning.
- NewStims: C sequences **not** heard in *Training*.
  - Therefore, segmental generalisations will not help with the harmony patterns.

Predictions

- Similar to Exp. 1, but:
  - All three (**MSG**, **PropSimple**, and **PropSpec**) predict a drop in NewStim preference.
  - However, **PropSimple** and **PropSpec** still predict an interactive effect.

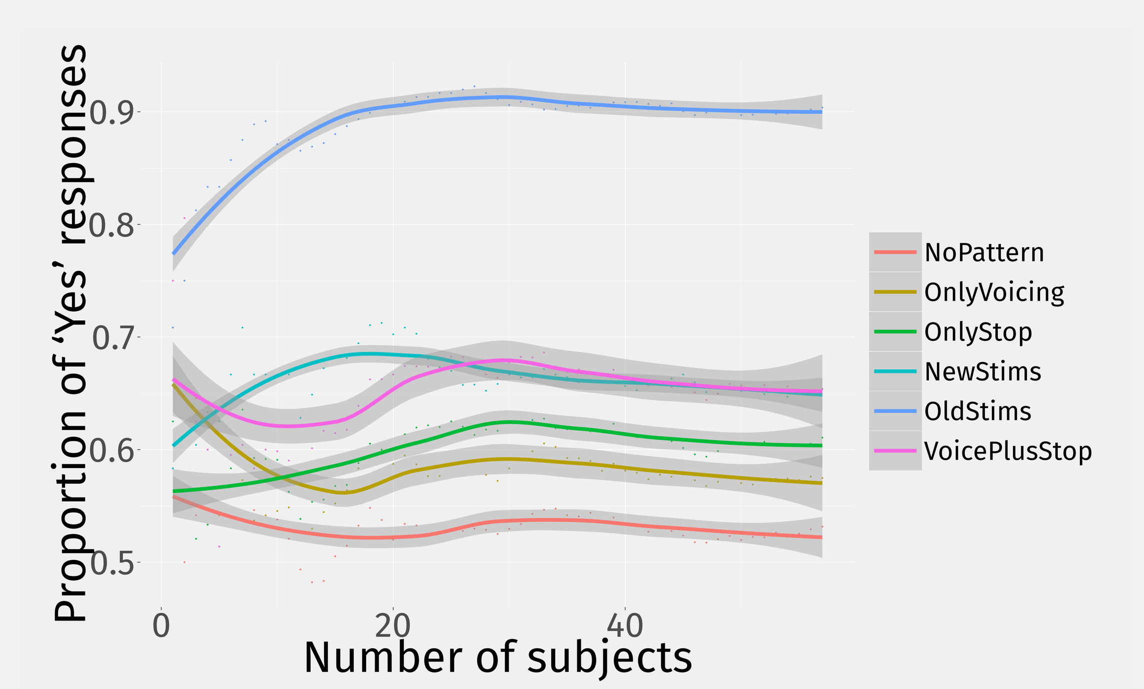
Results



Fixed Effect	MeanYes (%)	Estimate	z-value	Pr(>z)
(Intercept)	0.5317	0.1628	1.273	0.1014
OnlyVoicing	0.5701	0.1735	1.387	0.0825 .
OnlyStop	0.6138	0.3889	3.087	0.0010 **
NewStims	0.6534	0.5836	4.433	<0.0001 ***
OldStims	0.8981	2.2820	14.274	<0.0001 ***

Table 3: Logistic mixed-effects models

- OnlyVoicing+OnlyStop vs. NewStim responses (with increasing number of participants).



Discussion

- Noticeable drop in preference for NewStims.
- No evidence of interaction effect.
- The results are only consistent with **MSG**.

References

Berwick, Robert C. (1985). *The Acquisition of Syntactic Knowledge*. MIT Press.
Chomsky, Noam and Morris Halle (1968). *The Sound Pattern of English*. Harper and Row.
Gerken, LouAnn (2006). "Decisions, decisions: infant language learning when multiple generalizations are possible." *Cognition* 98, B67–B74.
Hale, Mark and Charles Reiss (2003). "The Subset Principle in phonology: why the tabula can't be rasa." *Journal of Linguistics* 39.02, pp. 219–244.
Hayes, Bruce and Colin Wilson (2008). "A maximum entropy model of phonotactics and phonotactic learning." *Linguistic Inquiry* 39, pp. 379–440.
Linzen, Tal and Gillian Gallagher (2014). "The Timecourse of Generalization in Phonotactic Learning." In: *Proceedings of the 2013 Meeting on Phonology*. Ed. by John Kingston et al. Washington D.C.: Linguistic Society of America, pp. 1–12. DOI: 10.3765/amp.v1i1.18.
Linzen, Tal and Timothy J. O'Donnell (2015). "A Model of Rapid Phonotactic Generalization." In: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, pp. 1126–1131.
Tenenbaum, J. B. and T. L. Griffiths (2001). "Generalization, similarity, and Bayesian inference." *Behavioral and Brain Sciences* 24, pp. 629–640.

Acknowledgements

Thanks to Mina Hirzel for help with recording the stimuli, Alan Munn and Cristina Schmitt for help with participant recruitment, Mike Kramizeh for help with use of the MSU CeLTA lab facilities, Yen-Hwei Lin, Suzanne Wagner, and the SoPhon group at MSU for valuable feedback.