Generalizing French schwa deletion: the role of indexed constraints

Aleksei Nazarov (Utrecht University) and Brian Smith

Intro

- Two possible problems for learned models:
  - Not capturing training data (underfitting)
  - Not generalizing from training data (overfitting)
  - Tendency: tradeoff between these (Hastie et al. 2001)
- Indexed constraints designed to capture training data (Pater 2000)
  - Will this hurt generalization?
- Tested with 4 procedures for indexed constraint discovery
  - Learn from French schwa deletion corpus
  - Apply learned models to existing experimental data

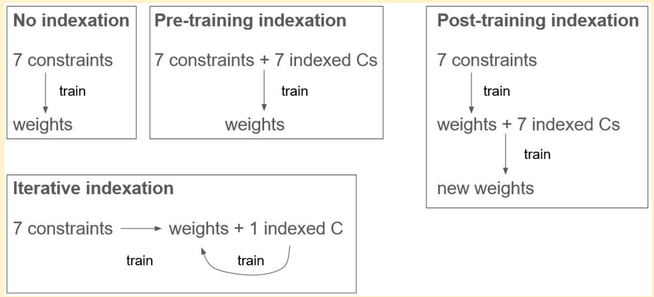
Ask question:

Download:

http://alekseinazarov.org/nazarov-smith-amp-poster/

Procedure

- MaxEnt (Goldwater & Johnson 2003)
  - batch-trained using hgR (Staubs 2011) from all-0 weights
- Start out without indexed constraints
  - Use word-specific gradients of constraint weights to find indexed constraints (cf. Becker 2009, Pater 2010)
- 4 increasingly involved indexation procedures
  - No indexation
  - Pre-training (use weight gradients before training)
  - Post-training (use weight gradients after training once)
  - Iterative (add one indexed constraint at a time while training in between, until convergence)



French schwa deletion (Dell 1985)

- Deletion of /ə/ (phonetically [œ]) optional in many contexts
- Rate of deletion depends on phonological context

e.g., VC_VC > #C_C > CC_C, C_CC
kas(ə)ʁɔl s(ə)ʁɛ̃ subʁ(ə)so
'pot' 'canary' 'jolt'

Examples from Racine (2008) corpus

- However, rate of deletion also depends on individual word

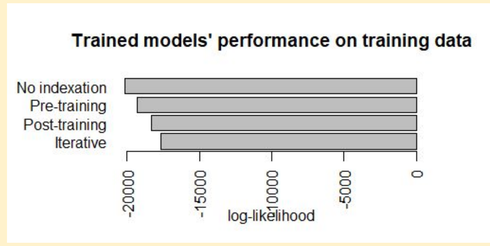
e.g. /sə̲mɛn/ 'week' (50% deletion)
/sə̲mɛstʁ/ 'semester' (14% deletion)

Training on corpus

- Corpus: 456 words × 12 France French speakers (Racine 2008)
  - Judgments on schwa-ful and schwa-less variants of same word
    - transformed into pseudo-frequencies
- Constraints:
  *ə, *ə/non-penult, Max, *#CC, *CCC, *CNC, *CTN (cf. Kaplan 2011)
- Evaluation: log-likelihood of training data (closer to 0 = better)

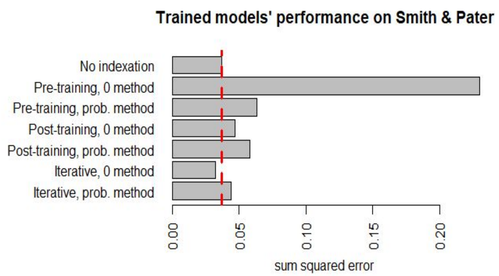
T = plosive
N = nasal

- As expected: data captured better as indexation gets more involved
- Performance not tied to number of indexed constraints:
  - Pre-training: 7
  - Post-training: 7
  - Iterative: 4



Generalization to Smith & Pater (2020)

- Smith & Pater (2020): Participants choose between schwa or no schwa in French phrases containing -CCə and -VCə words followed by CV or CVCV
- Use trained procedures to predict proportion of schwa responses per context
  - Constraints: *ə, *ə/non-penult, Max, *CCC, and any indexed versions
  - Generalization to experimental data:
    - 0 method: no indexed constraint violations
    - probabilistic method: indexed constraint violations determined by Maximum Likelihood (cf. Becker 2009)
- Evaluation: sum squared error (SSE; smaller = better)
  - More involved indexation = better
  - 0 method of generalization slightly better?
    - Except for pre-training indexation: definitely worse
  - Best indexation procedure on par with No indexation



References

- Barr, D.J., R. Levy, C. Scheepers & H.J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68.3:255-278.
- Becker, M. 2009. *Phonological Trends in the Lexicon: The Role of Constraints.* University of Massachusetts Amherst dissertation.
- Dell, F. 1985. *Les règles et les sons: Introduction à la phonologie générative*. Paris: Hermann.
- Goldwater, S. & M. Johnson. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. In J. Spenader, A. Eriksson, Ö. Dahl (eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*, 111–20.
- Hastie, T., R. Tibshirani, R. & J. Friedman. 2001. *The Elements of Statistical Learning.* New York: Springer.
- Kaplan, A. 2011. Variation through Markedness Suppression. *Phonology* 28.3:331–370.
- Pater, J. 2000. Non-uniformity in English secondary stress: The role of ranked and lexically specific constraints. *Phonology* 17.2:237–74.
- Pater, J. 2010. Morpheme-specific phonology: Constraint indexation and inconsistency resolution. In S. Parker (ed.), *Phonological argumentation: Essays on evidence and motivation.* London: Equinox.
- Racine, I. 2008. *Les effets de l'effacement du Schwa sur la production et la perception de la parole en français*. Université de Genève dissertation.
- Smith, B. & J. Pater. 2020. French schwa and gradient cumulativity. *Glossa* 5.1:24.
- Staubs, R. 2011. *Harmonic Grammar in R (hgR).* Software package. http://blogs.umass.edu/hgr/
- Zymet, J. 2018. *Lexical propensities in phonology: corpus and experimental evidence, grammar, and learning*. University of California Los Angeles dissertation.

Conclusions

- Indexed constraints improve account of training data, but do not have to hurt generalization!
  - Similar to how adding random effects improves generalization in Mixed Effects models (e.g., Zymet 2018; Barr et al. 2013)
- Best generalization behaviour when indexed constraints induced one by one, but not used in generalization to new items
  - Non-indexed constraint weights improved by precense of indexed constraints
- When evaluating on new data, method of generalization matters
  - For example, pre-training procedure fares much worse with 0 method