Exemplar-Modelling Vowel Harmony with Recurrent Neural Networks

Irfan S

Fall 2018

What is vowel harmony?

- Vowel harmony (VH) is a long-distance assimilatory phonological process
- long-distance: involves non-contiguous vowels in fixed morphological/prosodic domains

Approaches to Vowel Harmony

- traditional: understood in terms of feature-spreading or sharing (Anderson, 1980)
- coarticulatory approach, (Ohala, 1994)
  - "fossilised result of purely phonetic between-vowel assimilations"
  - diachronic sound change arising out of listeners' mis-perceiving sounds due to coarticulatory perturbation

An exemplar modelling approach?

- Cole (2009) suggests an exemplar-approach that models transitional probabilities between vowels co-occuring within words
  - harmonic classes would arise as emergent structure
  - recurrent neural networks are suggested as a possible implementation of this approach

VH in Finnish and Bangla

- Finnish: backness harmony, proceeds left-to-right
  - front or back vowel in the initial syllable spreads that feature to vowels in non-initial syllables
  - three harmonic classes - front [ä ö y]; back [a o u]; neutral [e i][1]
  - e.g., pos+ahta+(t)a → posahtaa (back)
  - räj+ahta+(t)a → räjähtää (front)
- Bangla: ATR harmony, proceeds right-to-left
  - ATR feature of suffixal vowels spreads to stem vowels
  - e.g., p tr+ika → potrika
  - kh l+i → kheli

[1] neutral vowels are unchanged by harmony

Dataset(s)

- Aalto University DSP Course Conversation Corpus
  - transcribed, force-aligned recordings of Finnish conversations
  - 5200 utterances, 9.7hrs of audio from 218 male and 24 female speakers
  - `http://urn.fi/urn:nbn:fi:lb-2017092133`
- SHRUTI Continuous ASR Speech Corpus
  - 7383 sentences spoken by 35 native speakers of Bengali
  - force-aligned phoneme and word-level annotations
- Both corpora publicly available, no IRB clearance required (yay!)

Data

- sequences of acoustic features (MFCCs) extracted from word-level segments
- potentially, word segments extracted from the corpus could also be sorted and separated by number of syllables (how is this useful?)

Model

- uni- or bi-directional recurrent neural networks
- model architecture similar to as described in Alishahi et al. (2017) (recurrent highway network with attention)
  - this would be advantageous in that we can use their findings to inform our understanding of what phonological representations different layers of the model are likely to learn
  - potential pitfall: this model relies on visual features to ground learning, which these datasets lack
  - substitute with word-vectors or augment corpora with images corresponding to lexical meaning (?)

Experiment 1

- following the methodology described by Alishahi et al. (2017), the following experiment is proposed
- from all possible CV combinations from our corpora, we construct tuples of the form (A,B,X) where
  - A, B, and X are syllables
  - B and X contain vowels drawn from the same harmonic class, while A is from another class or neutral
  - for each tuple, we calculate $sign(dist(A, X) - dist(B, X))$, where $dist(i, j)$ is euclidean distance between vector representations of syllables *i* and *j* output by our model
  - a positive value will indicate that representations of vowels sharing harmonic class are closer
- prediction: the performance of the model's output vectors should surpass baseline acoustic features at this task if the model has learnt phonological representations corresponding to harmonic features

Relevance

- the proposed experiment aims to verify the prediction made by Cole (2009)
- if successful, an exemplar-theoretic approach employing neural networks is validated as a viable alternative to formal phonological models based on phonemic features
- neural network models such as proposed here are capable of learning to represent data in multiple modalities
- a fully exemplar-theoretic approach, such as advocated by Port (2007), proposes that a rich variety of non-linguistic and linguistic formation is exploited by language learners and users in the processes of production and perception. statistical neural networks seem well-suited to modelling this kind of rich variety of inputs

Timeline

- Comprehensive literature review (mid-September)
- Data preparation for training and analysis (October, 1st week)
- Write model code, train models (October-end)
- Run experimental analysis (November)
- Write things up (November-...)

References I

Alishahi, A., Barking, M., and Chrupała, G. (2017). Encoding of phonology in a recurrent neural model of grounded speech. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, page nil.

Anderson, S. R. (1980). Problems and perspectives in the description of vowel harmony. *Issues in Vowel Harmony*, page 1.

Cole, J. (2009). Emergent feature structures: harmony systems in exemplar models of phonology. *Language Sciences*, 31(2-3):144–160.

Ohala, J. J. (1994). Towards a universal, phonetically-based, theory of vowel harmony. In *Third International Conference on Spoken Language Processing*.

References II

Port, R. (2007). How are words stored in memory? beyond phones and phonemes. *New Ideas in Psychology*, 25(2):143–170.