Grammatical Influences in a Bayesian Speech Production Framework

Christo Kirov
Johns Hopkins University, ckirov at gmail dot com

Grammatical Influences

- Speech planning is known to be affected by competition. Choosing between multiple similar options results in longer latencies and hyper-articulated (carefully pronounced) speech [1].
- But, language-specific knowledge and biases (grammar) *also* affect the production of both known and novel utterances.
  - Phonotactic knowledge affects the speed and accuracy of word-form repetition for both real and nonce words [2].
- Most production models do not take the latter into account.

Bayesian Production Framework

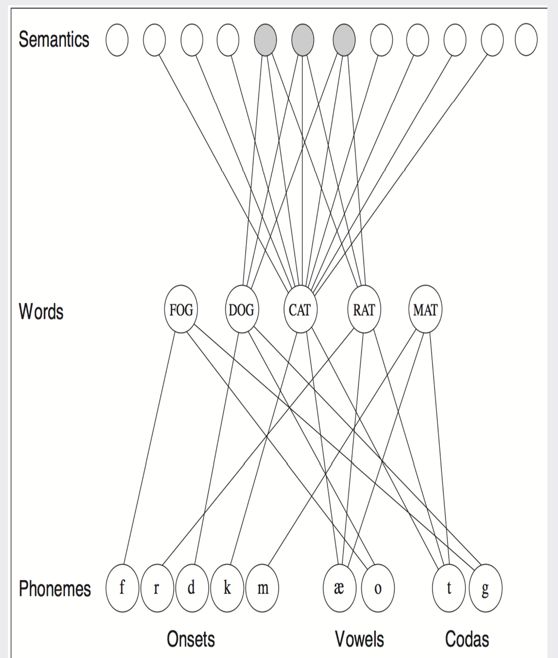
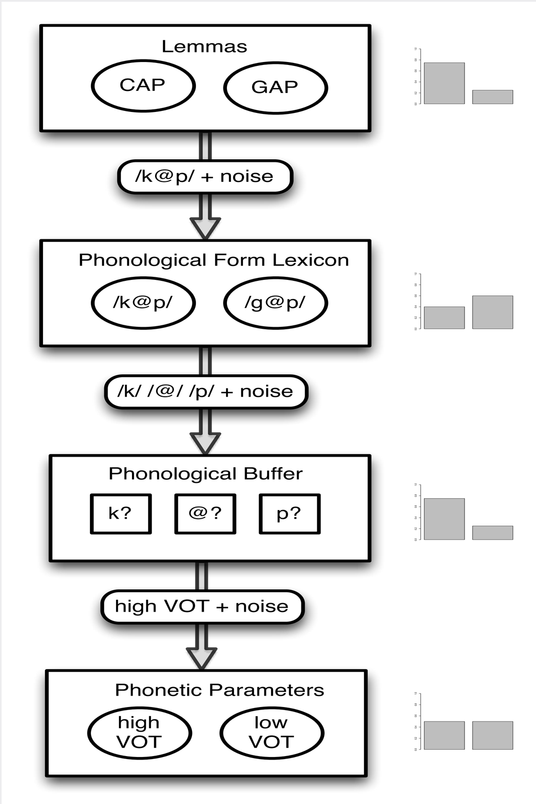


Fig. 1. Fragment of Dell's interactive lexical network. The nodes in the upper layer represent semantic features. The nodes in the middle layer represent words or lemmas. The nodes in the bottom layer represent onset, nucleus and coda phonemes (in particular consonants and vowels). All connections are bi-directional and there are only facilitatory, no inhibitory, connections. Activation spreads throughout the network without constraints; there is full cascading. It is always the most highly activated word or lemma node that gets selected. The moment of selection is determined externally, by the developing syntactic frame of the utterance. Upon selection the node receives an extra jolt of activation, which triggers its phonological encoding. The computational model has many more features than represented in the present figure. There is a further layer representing phonological features (such as 'voiced' or 'nasal') and there are versions of the model with a layer of syllable nodes. (Adapted from Dell *et al.*[12])



- Most production models involve communication between multiple levels of processing.
- Bayesian model assumed here [1] treats communications as evidence to update a probability distribution over what forms to produce. Updates repeat until some form reaches threshold probability.

$$p(r|E) = \frac{p(E|r)p(r)}{Z}$$

$$\text{posterior} = \frac{\text{likelihood} * \text{prior}}{\text{normalizing constant}}$$

- **Prior** encodes which representations are possible or expected — allows inclusion of task-specific knowledge (including **grammar**), and traces of previous productions.

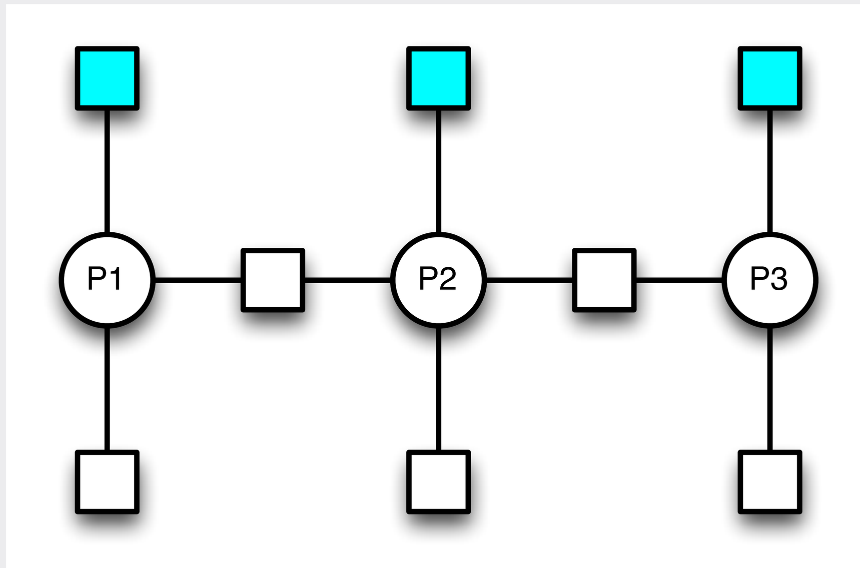
- At phonological levels, need to structure **prior** so that:
  - It can assign probabilities to arbitrary strings to allow for production of novel utterances.
  - It can incorporate grammatical knowledge.

Factor Graphs

- Denote a string s as $x_1 x_2 x_3 ... x_n$, where x_1 is the phone at position 1 in the string, and so on.
- $P(s)$ is a joint distribution over several random variables, X_1 to X_n, one for each position in the string
- Represent distribution as a product of factors.

$$P(s) = \frac{1}{Z} \prod_i f_i(\chi_i(s))$$

- Graphs connect variable nodes (circles) with factor nodes (squares).
- Unitary factors represent likelihood of a certain variable value. Some (blue) can represent evidence aggregated over time.
- N-ary factors represent co-occurrence constraints.
- There are efficient inference algorithms for values of interest such as the most probable arrangement of variable states (e.g., phonemes).



Harmonic Grammars as Graphs

- Harmonic grammars are a similar formalism to Optimality Theory.
- Harmonic grammars represent the *goodness* or a form by its *harmony*.
- Harmony is calculated as the sum of weighted constraint violations, where each constraint applies e.g., to a subset of positions in a phonological string.

$$h(x_1 x_2 ... x_n) = \sum_i w_i C_i$$

- By taking the exponent of the harmony, we get a proportional MaxEnt score [3]:

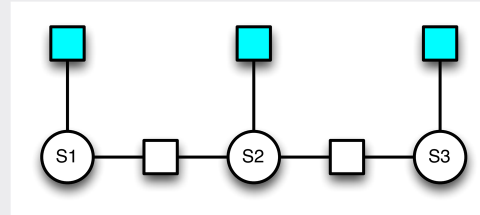
$$e^{h(x_1 x_2 ... x_n)} = e^{\sum_i w_i C_i}$$

- Thanks to the properties of exponentiation, we can represent this as a *product* of factors, which is all we need to generate a factor graph!:

$$e^{\sum_i w_i C(i)} = e^{w_1 C_1} e^{w_2 C_2} ... e^{w_N C_N} = f_1 f_2 ... f_N$$

Berber Syllabification

Base harmonic grammar [4] based on Dell and Elmedlaoui [5]. Variables have two possible states, **nucleus** and **non-nucleus**:



Unitary factors (blue):

- e^0 for non-nucleus state
- $e^{2^s - 1}$ for nucleus state, where s is the sonority of the most likely segment in position one.

Binary Factors:

- e^0 for when the two adjacent positions are not both nuclei
- e^{-2^8} when the two adjacent positions are both nuclei

Phonotactics

- Learn phonotactic factors based on three-segment words in the Hoosier Mental Lexicon.
- CVC words greatly outnumber all other structures such CCV (1338 to 275).
- In simulations, words with optimal phonotactics (i.e., CVCs) were produced in an average of 68.5 time steps. Words with other structures were produced in an average of 76.35 time steps.

Future Directions

- Moving beyond proof-of-concept simulations to test predictions of more detailed grammars - e.g., the full-scale phonotactic grammars induced by the Hayes/Wilson MaxEnt Learner [3].
- Thinking about plausibility as a neural mechanism:
  - Inference at each simulated time step is currently treated as instantaneous - should it have a temporal dimension? Do abstract grammatical processes such as allophonic variation have one?
  - Some neurons appear to perform probabilistic calculations, including normalization [6].

References

1. Kirov, C. & Wilson, C. (2012). The specificity of online variation in speech production. CogSci 2012. Sapporo, Japan.
2. Vitevitch, M. S. & Luce, P. A. (2005). Increases in phonotactic probability facilitate spoken nonword repetition. Memory and Language, 52(2), 193Ð204.
3. Hayes, B. & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. Linguistic Inquiry, 39(3), 379Ð440.
4. Smolensky, P. & Legendre, G. (2006). The Harmonic Mind, volume 2. The MIT Press.
5. Dell, F. & Elmedlaoui, M. (1985). Syllabic consonants and syllabification in Imdlawn Tashlhiyt Berber. Journal of African Languages and Linguistics, 7, 105Ð130.
6. Carandini & Heeger. (2012). Normalization as a canonical neural computation. Nature Reviews Neuroscience.