November 18, 2004

Chapter 4.5–4.8

Pronunciation dictionaries & TTS

Overview

- OT: Recap
- Machine learning of phonological rules
- TTS overview
- Pronunciation dictionaries
- FST-based pronunciation lexicon
- Prosody

Recap: Optimality Theory (OT)

- GEN takes an underlying form and produces all possible surface forms.
- EVAL consists of a set of ranked constraints and an algorithm for choosing the best candidate.
- The best candidate is the one who's highest constraint violation is lower than any of the others. In the case of a tie, the next constraint violations are considered.
- Example of Yokuts resyllabification (p. 116)
- Low-ranked constraints can still be effective
- Implicit typologies

Implementing OT

- Explicit interpretation of constraints
- GEN: a regular relation (FST)
- EVAL: Cascade the constraints, in order of ranking (highest to lowest).
- If all candidates violate a constraint, ordinary composition would leave us with the empty set.
- Apply ‘lenient composition’ instead (Karttunen 1998):

Lenient Composition (1/2)

- Lenient composition: the composition of Q and R (Q ∘ R) plus all elements of the domain of Q which don't map to anything in Q ∘ R.
- Priority union of Q and R: all pairs from Q, plus R applied to all elements not in the domain of Q.
  - macro(priority_union(Q,R), {Q, !domain(Q) ∘ R}).
  - macro(lenient_composition(S,C), priority_union(S ∘ C,S)).

Lenient Composition (2/2)

- Example:
  - Q: { < b, bb >, < a, bb >, < a, bbbb >, < a, bbbbbb > . . . }
  - R: [bbb]*
  - What is the lenient composition of Q and R?
- How does this help with the OT violable constraints problem?

Learning Rankings

- Tesar & Smolensky (1993, 1998): Error-Driven Constraint Demotion, learns ordinal rankings.
- Boersma (1997, 1998, 2000): Gradual Learning Algorithm learns stochastic rankings, can handle optionality and variation, as well as noisy training data.

Learning Rules (1/2)

- Machine learning systems automatically induce a model for some domain, given some data and potentially other information.
- Supervised algorithms are given correct answers for some of the data and use the answers to induce generalizations to apply to further data.
- Unsupervised algorithms works only from data, plus potentially some learning biases.

Learning Rules (2/2)

- Johnson (1984)/Touretzky et al (1990) learn SPE-style rules from a corpus of input/output pairs.
- Ex: Gildea & Jurafsky (1996) specialize a learning algorithm for a subtype of FSTs to learn two-level phonological transducers from a corpus of input/output pairs.
- Required learning biases: Faithfulness and Community
- If SPE-style rules can be implemented as FSTs automatically, why learn the FSTs directly?

Text-To-Speech

- Map orthography to phonetic transcription
- Add in prosody
- Map phonetic transcription + prosody to acoustic signal

Pronunciation dictionaries

- List words and their pronunciations
- No morphological or phonological rules
- PRONLEX: 90,694 wordforms
- CMUdict: 100,000 wordforms
- CELEX: 160,595 wordforms
- Designed for ASR, but can be adapted for speech synthesis
- In what way do the requirements on dictionaries differ between these two applications?
- What problems might arise for this approach?

Problems for simple listing

- Highly variable pronunciations (*and*, *I*, *the*, *of* etc.)
- Names:
  - 21% of 33 million words of AP newswire were names (Liberman & Church 1992).
  - Includes not only people's names but also company names and product names.
  - ... named entity recognition
- Morphological productivity
- Number names, with different possible pronunciations:
  - Serial, combined, paired, hundreds, trailing unit, (trailing unit with a decimal)

FST-based approach

- Components:
  - large morpheme pronunciation dictionary, encoded as an FST
  - FSAs for morphotactics
  - FSTs for morphophonology (like spelling change rules)
  - heuristics and LTS rules/transducers for names and acronyms
  - default LTS rules/transducers for other unknown words
  - (Named-entity recognizer)

Architecture (1/2)

- Lexical, intermediate and surface levels all contain two tapes, one for pronunciation and one for orthography.
- Lexicon-FST: composed of two-level lexicon plus FSAs/FSTs for morphology (+PL| ϵ:s|z) [4.21–23]
- FST_1 ... FST_n: orthographic and phonological rules, run in parallel

Architecture (2/2)

Lexical:	f o:aa x:ks +N +PL
	LEXICON-FST
Intermediate:	f o:aa x:ks ^ s:z
	FST_1 ... FST_n
Surface:	f o:aa x:ks e:ix s:z

- Why have both orthographic and phonological representations at every level?

Names

- Donnelly marketing organization: 1.5 million name "tokens" (for 72 million US households)
- Liberman & Church (1992) attempt to handle most frequent 250,000 (1/6) of these
  - Dictionary of 50,000 names covers 59%
  - Stress-neutral suffixes (*-s*, *-son*, *-ville*): 84%
  - Name-name compounds and rhyming heuristics: 89%
  - Prefixes, stress-changing suffixes and suffix-exchanges: ??
  - LTS rules for the remainder.

Names: Your assignment

- Find a suitable set of “name stems” and two name suffixes (one stress-neutral and one stress-changing).
- Model (using xfst) the possible names made up of those stems and suffixes (at most one suffix per name)…
- … including the stress assignment.

Prosody

- Prominence: stress (lexical and sentential)
- Structure: intonational phrases/units, intermediate phrases
- Tune: F0 pattern, component parts include pitch accent

English pitch accents (Pierrehumbert 1980)

- **H***: high (on a stressed syllable)
- **L***: low (on a stressed syllable)
- **L*+H**: rise, starting on a stressed syllable
- **L+H***: rise, ending on a stressed syllable
- **H+L***: fall, ending on a stressed syllable
- (**H*+L**: apparently not needed)

Other components of the English system

- Phrase accents:
  - **L-**
  - **H-**
- Boundary tones:
  - **L%**
  - **H%**

Text-To-Speech

- Map orthography to phonetic transcription
- Add in prosody
- Map phonetic transcription + prosody to acoustic signal

Summary

- OT: Recap
- Machine learning of phonological rules
- TTS overview
- Pronunciation dictionaries
- FST-based pronunciation lexicon
- Prosody
- Next time: Reference resolution