1. Contributions

2. Present a model which generalizes on the basis of phonological features. This model:
   (a) is provably correct, and provably efficiently estimable
   (b) integrates into Strictly Local (n-gram) or Strictly Piecewise models
   (c) assumes statistical independence of individual features
   (d) captures intuition that sounds with like features have like distributions

2. Expositional Feature Chart

We demonstrate with this feature system, but nothing hinges on it. This proposal accommodates privative and multi-valued feature systems.

<table>
<thead>
<tr>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>+</td>
</tr>
<tr>
<td>b</td>
<td>+</td>
</tr>
<tr>
<td>c</td>
<td>+</td>
</tr>
</tbody>
</table>

3. Expressivity of maxent models

Theorem 1. Every maxent model with featural constraints which describes a distribution is describable by one with segmental constraints.

Proof sketch.

<table>
<thead>
<tr>
<th>Grammar 1</th>
<th>Grammar 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>constraint weight</td>
<td>constraint weight</td>
</tr>
<tr>
<td>[+F][+G]</td>
<td>$w_1$</td>
</tr>
<tr>
<td>[+G][-F]</td>
<td>$w_2$</td>
</tr>
<tr>
<td>*ab</td>
<td>$w_1$</td>
</tr>
<tr>
<td>*ac</td>
<td>$w_1 + w_2$</td>
</tr>
<tr>
<td>*bb</td>
<td>$w_1$</td>
</tr>
<tr>
<td>*bc</td>
<td>$w_1$</td>
</tr>
</tbody>
</table>

For each constraint $C$ with weight $w$ (e.g. *+[X] or *[+X]+[Y]), add $w$ to the weight of all segmental sequences violating $C$, (adding more segmental constraints with weight $w$ if needed). This procedure ensures maxent grammars $G_1$ and $G_2$ assign the same maxent scores to all words.


The table shows the correlation (Spearman’s r) between Hayes & Wilson’s maxent grammars obtained with their learner on CMU English onsets and Scholze’s (1966) experimental results. Are their results due to features or use of complement classes?

<table>
<thead>
<tr>
<th>Hayes and Wilson maxent models</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>features &amp; complement classes</td>
<td>0.95</td>
</tr>
<tr>
<td>no features, complement classes</td>
<td>0.94</td>
</tr>
<tr>
<td>no features &amp; no complement classes</td>
<td>0.88</td>
</tr>
</tbody>
</table>

5. Feature-based Strictly 2-Local (Bigram) Probability Distributions

Let $w = a_1a_2...a_n$ and let $\mathcal{F}$ be a feature system. Then

$$
P(w) = P(a_1 | \#) \times P(a_2 | a_1) \times \cdots \times P(a_n | a_{n-1}) \times P(\# | a_n)
$$

Equations 1 and 2 define a well-formed probability distribution over $\Sigma^*$.

Theorem 2. Equations 1 and 2 define a well-formed probability distribution over $\Sigma^*$.

Corollary 1. There are $|\Sigma| \times |\mathcal{F}|$ parameters of the distribution. They are, for all $a \in \Sigma$ and $f \in \mathcal{F}, P(a | f)$.

Corollary 2. These parameters can be estimated by finding the Maximum Likelihood Estimate using standard techniques for probabilistic finite-state machines (de la Higuera, in press).

Proof sketch of Theorem 2.

$$
P(a | b) \overset{\text{def}}{=} P(a | F_b) = \frac{\Pi_{f \in \mathcal{F}} P(a | f)}{\sum_{a' \in \Sigma \setminus \{\#\}} \Pi_{f \in \mathcal{F}} P(a' | f)}
$$

6. Feature-Based Bigrams

Table shows the correlation (Spearman’s r) between Hayes & Wilson’s maxent grammars obtained with their learner on CMU English onsets and Scholze’s (1966) experimental results. Are their results due to features or use of complement classes?

7. Word Initial Velar Nasals

Since nasals like *[m,n] and velars like *[k,g] begin words, the model infers [y] ought to as well.

$$
\text{Expected}(X) = P(X) \times N
$$

Expected(\#y) = 0.0005 \times 31,641 = 15.8

Observd(\#y) = 0.

This is instructive!

References and Acknowledgments


This work was supported by a University of Delaware Research Fund grant for 2008-2009 and a General University Research grant from the University of Delaware for summer 2009. We thank James Rogers and Robert Wilder for valuable discussion.

[There are] …two stages of evaluation: a preliminary initial assessment of probability of segment combinations and subsequent grammatical evaluation… Albright (2009)

We expect comparing expected values given by the feature-based distributions to observed values provides a platform for the inference of constraints.