Many morphophonological patterns can be analyzed with binary or multi-way classification. Allomorph selection, declension/conjugation assignment, and lexical stratification are canonical examples. Participation in phonological alternation (e.g., Gouskova & Becker, 2013) and grammatical category prediction are additional cases. Indeed, classification can be applied whenever a ‘diacritic’ feature of an item is (partly) predictable from its other phonological or morphological properties. While some previous research has explicitly adopted a classifier approach to such patterns (e.g., Ernestus & Baayen, 2003; Hayes, in press), existing proposals leave open the problem of how predictive properties are learned from examples. This talk introduces a probabilistic classifier model that induces weighted constraints from positive data, and demonstrates its usefulness and success on benchmark cases from the literature.

§1. Morphophonological classifier. The model assigns a probability distribution over classes conditioned on other phonological and morphological properties of a stem (or other item such as a word or phrase). Class probabilities are given by the equations that define logistic regression models (aka log-linear models or maximum entropy classifiers; see Jurafsky & Martin, 2009):

\[
p(c | x) = \frac{\exp(-\sum_k w_k f_k(x,c))}{Z(x)}
\]

where \( x \) is the item to be classified, \( c \) is a variable over classes, \( C \) is the set of possible classes, and \( w_k \) is the weight of the \( k \)th constraint. Each constraint \( f_k \) is a function from \((x,c)\) pairs to numerical values (assumed for convenience to be integer violations \( \geq 0 \)). This is essentially the same type of model that was considered by Ernestus & Baayen (2003) and Hayes (in press).

Applications of the model differ in the set of possible classes, the items that are classified, and the constraints and weights that determine classification probabilities. For example, the Russian masculine diminutive suffix has several allomorphs, of which \([-ok] / [-ik] / [-tʃʲik]\) are the most productive (Gouskova et al., 2015). While certain aspects of stem phonology are predictive of the choice among these allomorphs, the choice is essentially an idiosyncratic property of each noun: Gouskova et al. formalize it as a stem diacritic (e.g., /dom+ik/ ‘house’ permits /dom-ik/ ‘house-dim’ but not */dom-ok/, */dom-tʃʲik/). For the purpose of modeling this allomorphic distribution, the phonological forms of masculine nouns such as /dom/ would be values of \( x \), the diacritic +ik would be one value of \( c \), and \{+ok, +ik, +tʃʲik\} would be the set \( C \) of possible classes. The relevant constraints would evaluate pairs of the form (/dom/, +ok), (/dom/, +ik), and (/dom/, +tʃʲik).

§2. Learning weighted constraints. Given a data set of lexical items with their observed classes, the constraint weights can be learned by (regularized) maximum-likelihood fitting. Only positive examples are required, as maximizing the probability of an observed class entails minimizing the probabilities of possible but unobserved classes, and the data can be gradient (e.g., items can occur with multiple classes weighted by relative frequency).

A more difficult challenge is to induce the constraints themselves from the data. The present proposal adopts an approach related to that of Hayes & Wilson (2008), though significantly modified for the purpose of learning classification constraints rather than phonotactic restrictions. Along with the learning data, the model is provided with a feature matrix for the phonological segments and a set of phonological tiers (or ‘projections’). Each possible constraint takes the form \( f_k = *\langle \text{tier}_k, \text{pattern}_k, c_k \rangle \) where the \text{pattern} is a sequence of adjacent natural classes on the designated \text{tier}. In evaluating pairs of the form \((x,c)\), the constraint assesses one violation for each instance of \text{pattern}_k on \text{tier}_k in \( x \) iff \( c = c_k \) (i.e., for all other values of the class variable \( c \) the constraint assesses zero violations).
Holding the current (initially empty) grammar of constraints and weights fixed, the model calculates how much the total probability of observed classifications would increase as the result of adding each potential new constraint, and then adds the constraint with the highest value. This process iterates until no constraint affords an improvement above a specified threshold, which is the one free parameter of the constraint induction component and interpretable as a complexity penalty. All constraints discussed below apply on the default, segmental tier.

§3. Model evaluation. When provided with the lexical diminutive data of Gouskova et al. (2015), the model induced a compact grammar of 28 constraints, many of which accord with known generalizations (e.g., default tier constraints *[–son, +cor]#, +tʃik 10.27 and *[+dor, –palatal]#, +ik 7.91). When repeatedly trained and tested on disjoint subsets of the data, the model assigned highest probability to the observed allomorph for 70% (±02%) of the held-out forms. The model also correctly predicted 57% of the judgments in the wug-test experiment reported by Gouskova et al. The original paper did not provide a binary accuracy value, but it was estimated that their alternative constraint-based model (see §4) achieved a somewhat lower level of performance (~47% correct).

As indicated at the outset, the classifier approach to morphophonology extends far beyond allomorph selection. In a seminal paper, Ernestus & Baayen (2003) identified phonological factors of the final syllable that are relevant for predicting whether a Dutch stem-final stop undergoes voicing alternation (which is demonstrably idiosyncratic, e.g., verwij[t] ~ verwij[d]en ‘widen’ ~ ‘widen-inf’ but verwij[t] ~ verwij[t]en ‘reproach’ ~ ‘reproach-inf’). When trained on the same data considered by Ernestus & Baayen, the model learned 7 constraints that correctly predict the alternation behavior of 79% of the lexical items and match the majority human response for 85% of the nonce items in their wug-test. The latter value compares favorably with the 72% - 91% wug accuracy reported by Ernestus & Baayen for a range of models with hand-written constraints.

§4. Summary. Identifying the factors that predict allomorph selection and related patterns is important for the empirical study of morphophonological patterns, both in individual languages and typologically, and presents a clear learning challenge. The model proposed here faces this challenge directly, by inducing weighted constraints that penalize pairings of tier-based phonological patterns with classes. The only previous inductive constraint-based approach instead employs the model of Hayes & Wilson (2008) to learn class-specific phonotactics, which are then ‘inverted’ to yield predictions of classes given forms (Allen & Becker, submitted; Gouskova et al., 2015). The grammars learned by the present model are more compact and interpretable, and preliminary results suggest that they empirically outperform those of the alternative by a small margin. This motivates further exploration, across a wide range of cases, of the idea that morphophonological knowledge involves probabilistic classification.