A Testbed for Developing Multilingual Phonotactic Descriptions

Simone Ashby, Julie Carson-Berndsen and Gina Joue

Department of Computer Science
University College Dublin
Belfield, Dublin 4
Ireland

{Simone.Ashby,Julie.Berndsen,Gina.Joue}@ucd.ie

Abstract

This paper presents a testbed for developing multilingual phonotactic descriptions that employs finite state methods to represent the phonotactics of one or more languages. The motivation for this work is to make an extensive range of phonotactic descriptions of varying granularity available for speech technology applications. We discuss the design of the phonotactic testbed and how various modules may be used to generate finite state phonotactic descriptions. We provide an example multilingual application drawn from a partial sample of onset clusters spanning four language families, demonstrating how the commonalities of a broad spectrum of languages can be expressed using individual and generic phonotactic automata. We then discuss how these representations are extended via a three-tiered model to provide the basis for the feature- and event-based phonotactic automata.

1. Introduction

This paper presents an approach to multilingual phonotactic description based on finite state techniques. Development and evaluation of multilingual descriptions is performed with the help of a phonotactic testbed which has been specifically designed for this purpose. In what follows, we introduce the testbed together with an example application of this testbed in the development of multilingual phonotactic knowledge bases. The example application covers phoneme-based descriptions of subsections of onset phonotactics for 11 different languages. It is demonstrated how language similarities can be expressed using individual and generic phonotactic automata. The motivation for this work is twofold. On the one hand, we aim to make an extensive range of phonotactic descriptions of varying granularity available for speech technology applications. These levels of granularity range from phoneme-based descriptions to event-based descriptions which can be used to interpret multilinear (autosegmental) representations [1]. On the other hand, we believe that in order to find the most suitable phonetic feature set for use in speech recognition applications, it is necessary to examine the systematicities and the idiosyncrasies of many languages. For this reason, we demonstrate in an example application how the common groups of languages are classified with respect to their phonotactic structures and are mapped to these classes using a finite state transducer.

In the next section, the phonotactic testbed is introduced. For this, and for the example application in section 3, discussion is restricted to phonemic representations. However, in section 4, we demonstrate how these representations are extended via a three-tiered model to provide the basis for feature-based phonotactics and the event-based phonotactic automata assumed by the Time Map model of [1,2].

2. The Phonotactic Testbed

The phonotactic testbed has been designed specifically for developing and evaluating multilingual phonotactic descriptions. The technology is based on finite state techniques and has a web interface for platform independent development. The current version of the testbed consists of four modules: an input and display module, a generation module, a comparison module and a parsing module.

The input and display module allows a user to input phonotactic descriptions in terms of a finite state machine, defining the number of tapes, the initial state, the final state and the set of transitions. The number of tapes is restricted in this application to either one or two tapes, i.e. an automaton or a transducer, but in principle, this could be extended to cater for multi-tape processing should it be required. All representations in the phonotactic database of the testbed are available for display to the user. A simple example description for stop + \{l, r\} combinations in English phonotactics is used for illustrative purposes in figure 1. The complete phonotactic descriptions are considerably larger.

<table>
<thead>
<tr>
<th>Tapes:</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial:</td>
<td>0</td>
</tr>
<tr>
<td>Final:</td>
<td>3</td>
</tr>
</tbody>
</table>

0 1 p, t, k, b, d, g
0 2 p, t, k, b, g
1 3 r
2 3 1

Figure 1: Finite state description of stop + \{l, r\} combinations in English syllable onsets

The generation module uses the information contained in the phonotactic automaton or transducer to generate all the phonotactic combinations. Since this is possible only if the automaton is free of any loops, a check is performed before generation is attempted. The set of combinations generated for the example description of figure 1 would be as follows:

1 SAMPA rather than IPA fonts are used in the testbed.
The **comparison module** allows different descriptions to be compared, highlighting where they differ. This is done by generating all the combinations of one phonotactic automaton and determining whether these represent acceptable strings for the other phonotactic automaton. Strings which are not accepted are highlighted. A subsection of the output of the comparison of an English syllable onset with a German syllable onset is depicted in figure 3.

The **parser module** allows user input to be parsed according to the units defined in the finite state phonotactic description. That is to say, if the phonotactic descriptions only cover onsets, then only permissible onsets can be parsed; if the phonotactic description covers syllables, then permissible syllables are parsed. Starting in the initial state, the input string is processed symbol by symbol until a final state in the description is reached. The parser then sets a boundary marker ($) and returns to the initial state if the input string has not been completely consumed. This is equivalent to a standard finite state acceptor which outputs a boundary marker every time it reaches a final state. The parser returns either the parsed string, including the path through the automaton or transducer, or a message stating that parsing has failed. The structural information (e.g. syllable structure with respect to onset, peak and coda) is implicit in the path information.

Parser output for the English word *relate*, given a complete syllable phonotactic automaton for English, is as follows:

\[\text{pr} \rightarrow \text{kr} \rightarrow \text{dr} \rightarrow \text{pl} \rightarrow \text{bl}\]

\[\text{tr} \rightarrow \text{br} \rightarrow \text{gr} \rightarrow \text{kl} \rightarrow \text{gl}\]

**Figure 2:** Output stop + \{l, r\} combinations in English syllable onsets

There are two possible parses for this form; the second corresponds to a form such as *real eight*. The paths through the phonotactic description indicate the structure of the syllable. For example, the second path corresponds to the following structure:

\[\text{b} \rightarrow \text{p} \rightarrow \text{k} \rightarrow \text{e} \rightarrow \text{a} \rightarrow \text{t} \rightarrow \text{e} \rightarrow \text{r} \rightarrow \text{e} \rightarrow \text{t} \rightarrow \text{e}\]

\[\text{b} \rightarrow \text{n} \rightarrow \text{t} \rightarrow \text{e} \rightarrow \text{r} \rightarrow \text{e} \rightarrow \text{t} \rightarrow \text{e}\]

**Figure 3:** Subsection of English syllable onset vs. German syllable onset

The phoneme-based phonotactic descriptions define the base model. This is then extended to incorporate commonalities between the phonemes through the use of features. These features may be selected from a given feature classification (i.e. binary distinctive features, multivalued phonetic features, cognitive features such as those proposed by [3]) or they may be defined by the user. We discuss this issue further below. The feature-based model then provides the basis for the event-based phonotactic descriptions used in speech recognition within our research group [1,4]. We envisage that the three-tiered approach will enhance the multilingual functionality of the *Time Map* model.

### 3. An Example Multilingual Application

As discussed in the previous section, the user can input phonotactic descriptions for multiple languages. We performed a multilingual analysis over a partial data sample representing four different language families: Celtic (including Irish [Ir] and Welsh[W]); Germanic (including Dutch [D], English [E] and German [G]); Romance (including French [F], Italian [IT] and Spanish [S]); and Slavic (including Bulgarian [B], Czech [C] and Russian [R]) based on data from [5]. For the purposes of illustration, our example application is limited to a subsection of onset phonotactics, namely the cluster obstruent + \{l, r\}.

We used the input and display module to enter phonotactic descriptions for each language. The following example reveals a combined finite state description for Irish and Welsh, whereby the number of tapes has been set to 2 for the purpose of generating a phonotactic transducer; the initial state has been set to 0, representing the start of the syllable onset; and the final state has been set to *vo* (vowel onset) path for the path from onset to vowel.

**Figure 4:** Parse of [r1l $ e r t]

In the current example, our output tape is the string of languages (represented as small cap abbreviations) whose phonotactics allow for the corresponding path combination.

---

2 T and D are the SAMPA symbols for \[\emptyset\] and \[\emptyset\].
As one might expect, a number of interesting similarities and differences become apparent when we extend our example multilingual application to the additional 9 languages in our data sample (see figure 8). For instance, all 11 languages allow for the onset cluster \{p, b, k, g, f\} + \{l, r\}. Similarly all 11 languages allow for the onset cluster \{t, d\} + \{l, l\}. Yet only the Slavic and Celtic families in the language set allow for the onset cluster \{l\} + \{l, l\}, with only the Slavic group also allowing for the cluster \{d\} + \{l, l\}.

Figure 7: Finite state transducer of obstruent + \{l, r\} combinations in Irish and Welsh syllable onsets

This example application does not address feature phenomena and is solely for the purpose of demonstrating how, at a more granular level, the testbed might be used to generate multiple, parallel phonotactic descriptions for examining how different languages treat the distribution of phonemes and the various constraints that govern how they are organised (but cf. section 4).

\(^3\) Welsh distinguishes between an alveolar /s/ and a dental /l/.

4. Extending to a Three-Tiered Model

As demonstrated in the previous section, comparison of phoneme-level phonotactics can aid the process of examining multiple languages at a more granular level. However, we believe a three-tiered model, which allows phonotactic description at finer grained levels adds an important descriptive dimension. In addition to the phonemic level, we now move on to discuss this in the context of how one might extend such a phonemic representation to a more fine grained level of feature- and event-based modelling.
theories as well as comparisons of them. The event level treats these features as independent temporal events [1] which have temporal relations amongst themselves.

Allowing for feature-level phonotactics has a couple of advantages. It promotes phoneme classification, thus clarifying and allowing generalisations of the tendencies for certain classes of sounds to combine with other classes. For example, acoustic features may indicate which phonemes tend to share similar adjacent phonemes, and perhaps in another language, articulatory features dominate these tendencies.

The advantage of a feature-level phonotactic system can be seen in examining the transition from node 0 to 1 (see figures 9 and 10) of the combinatory transducer for the minidata set shown in Figure 8.

![Figure 9: Feature specification (from the IPA) for part of the combinatory network](image)

![Figure 10: Feature specification (from Phonology as Human Behavior) for part of the combinatory network](image)

As seen in figures 9 and 10, feature-level descriptions collapse the entities that need to be considered during the analysis of phonotactic patterns. The groupings around each of the phonemes in figure 8 reflect the natural classes which emerge when the feature-level, based on the IPA, is considered. Figure 10 further illustrates that different feature sets can create different “natural” classes or more encompassing classes, which may contribute to explaining phonotactic patterns.

Feature-level phonotactics also address some of the main issues in automatic language identification. Automatic language identification techniques have largely relied on the most frequent n-grams of phones as drawn from a corpus for the given language. This does, to an extent, reflect phonotactic structures but our three-tiered phonotactic model bypasses some of the weaknesses of this approach such as in dealing with unseen n-grams and requiring huge training/testing resources. As the phonotactics represent all possible combinations, the testbed can cope with unseen structures. As the model also facilitates classification of languages into families, it can identify the closest family of languages when data for the given language to be identified is not available.

Event-level phonotactics also aids language identification by integrating temporal and prosodic information. The syntax of prosodic structure, such as phone/feature/syllable duration and prosodic patterns are issues which have not been properly addressed in identification techniques.

5. Conclusions and Future Research

This paper has demonstrated the application of a phonotactic testbed as a means of highlighting the usefulness of finite state technology in conducting multilingual phonotactic analysis for theoretical and applied purposes. Work is in progress to extend this testbed to a more extensive multilingual toolkit that allows for more fine grained linguistic analyses such as feature- and event-based finite state analyses. Furthermore we are also researching the use of weighted automata to create phonotactic automata which reflect the frequencies of phonotactic combinations. By incorporating articulatory and acoustic similarity measurements into the phonotactic network, we hope to complete a multidimensional scaling of languages and thus create a first level of quantifying similarities between languages within and between language families. This can then be extended by a dedicated inclusion of prosodic information into the finite-state model, using the event level to capture the temporal events of pitch and stress variations.

6. References


* The authors gratefully acknowledge the support of the HEA under the MMRP programme.