

English Pronunciation: Issues and Practices (EPIP)

Proceedings of the First
International Conference

Alice Henderson (ed.)

DIRECTEUR DU LABORATOIRE

Christian Guilleré

COMITÉ DE RÉDACTION

Alice Henderson

COMITÉ SCIENTIFIQUE

Una Cunningham, University of Dalarna, Sweden

Heather Hilton, University of Savoie, France

David Levey, University of Cadiz, Spain

Jan Majer, University of Łódź, Poland

John Osborne, University of Savoie, France

Linda Shockey, University of Reading,

Rias van den Doel, Utrecht University, The Netherlands

Ewa Waniek-Klimczak, University of Łódź, Poland

Cet ouvrage a été réalisé avec le concours de
l'Assemblée des Pays de Savoie,
la Région Rhône-Alpes



Rhône-Alpes Région

SOMMAIRE

Introduction to EPIP publication

Alice HENDERSON7

Pronunciation Preferences..... 11

A corpus-based study of phonological free variation in English

Jose A. MOMPEÁN 13

*English pronunciation preferences:
Research by 'indirect' questionnaire*

Mohamed BENRABAH.....37

Native & Non-native Speakers 55

*The perception of word stress in English and French:
Which cues for native English and French speakers?*

Dan FROST57

*The effect of task on the pronunciation of English.
High front vowels by Japanese learners*

Rika AOKI75

*An empirical study of individual differences in L2 oral proficiency:
What makes native-like speakers special?*

Tanja ANGELOVSKA95

Teaching Issues..... 117

*Speaking of speech: Developing metalanguage for effective communication
about pronunciation between English language teachers and learners*

Helen FRASER..... 119

*Phonology and Moodle:
Enhancing pronunciation through learning platform-based training?*

Angela HAHN 139

<i>Temporal parameters in the implementation of the voicing contrast in English spoken by Poles: A pedagogical perspective</i> Arkadiusz ROJCZYK	159
<i>The pedagogical implications of variability in transcription: The case of [i] and [u]</i> Sophie HERMENT	177
<i>Pronunciation teaching materials in finnish EFL textbooks</i> Elina TERGUJEFF	189
Future Developments	207
<i>How could English truly become a new Latin?</i> Sylvia SCHEUER	209
<i>The PhonBank initiative and second language phonological development: Innovative tools for research and data sharing</i> Yvan ROSE	223
Author contact details	243

THE PHONBANK INITIATIVE AND SECOND LANGUAGE PHONOLOGICAL DEVELOPMENT: INNOVATIVE TOOLS FOR RESEARCH AND DATA SHARING

YVAN ROSE¹

MEMORIAL UNIVERSITY OF NEWFOUNDLAND

ABSTRACT

In this paper, I first introduce the PhonBank initiative within the CHILDES project. I then offer a brief overview of the Phon software program, and show how it provides unprecedented support for research on the second language acquisition of phonology. Phon is a free, open-source software program with a user-friendly graphical interface that facilitates a number of tasks required for the analysis of phonology and phonological acquisition. Corpora managed within Phon can be queried and exported through a powerful searching and reporting system. Example analyses are provided in the context of second language development, loanword phonology and research on the phonological properties of language dialects. Throughout the discussion, I also emphasise the importance of data sharing for everyone involved in research on phonology and language development.

Keywords: Phonology, Second language, Acquisition, Development, Corpus, Software, Data sharing

1 I would like to thank everyone behind the organisation of EPIP, especially Alice Henderson and Heather Hilton, for their generous invitation to participate. I am also thankful to everyone behind the PhonBank initiative, including all the members of the Phon development team at Memorial University of Newfoundland and of the CHILDES project at Carnegie Mellon University for their tremendous support. The work discussed in this chapter has benefited from funding from the National Institute of Health, the National Science Foundation, the Canada Foundation for Innovation, the Social Sciences and Humanities Research Council of Canada, Memorial University of Newfoundland, and a Petro-Canada Young Innovator Award. All errors or omissions are my own.

Introduction

Since its inception in the early 1980s, the CHILDES project has provided researchers and students of child language with a wealth of systems geared towards the coding, analysis and sharing of corpora documenting language and language development from a number of different perspectives, including first and second language acquisition, speech disorders as well as conversational and discursive aspects of speech (e.g. MacWhinney 1991, 1996). This project has relied on two main pillars: first, the development of computational means to analyse speech data and, second, data sharing.

Up until recently, CHILDES offered extremely good support for research on units such as morphemes, words and even syntactic constructions or aspects of their conversational or interactional properties, while relatively little support was available for phonological investigations. However, this situation started to change a few years ago, through the PhonBank initiative. In the sections that follow, I briefly address the general goals of PhonBank. I then move on to a general description of the Phon software program, with an emphasis on new and improved features included in version 1.4 of the application, which was publicly released in the fall of 2009. Building on these descriptions, I illustrate how these tools can be used in the context of research on second language phonological systems, loanword phonology and dialectal variation.

PhonBank

Background

The PhonBank project seeks to broaden the scope of the current CHILDES system to include the analysis of phonological development in first and second languages for learners with and without language disorders. To achieve this goal, Brian MacWhinney, professor of psychology at Carnegie Mellon University and I have endeavoured to develop a new phonological database called PhonBank and a program called Phon, whose functions support the elaboration and analysis of PhonBank data. Using Phon, and based on PhonBank data, we aim at a series of developmental, cross-linguistic, and methodological analyses, all targeting a better understanding of phonological development and related speech disorders.

All aspects of this work are accomplished in close collaboration with members of the PhonBank international consortium, a group of approximately 70 child language researchers and students whose input has proven invaluable to the progress made on the five overall goals of the project:

1. Building Phon: The construction of the Phon program is central to the project, since no software program was previously available to

support the needs of research specific to phonology and phonological development. The development of Phon has thus far supported the first strides in database development and preliminary hypothesis testing, most of the latter as part of Master's and Ph.D. theses recently completed or currently being written.

2. Hypothesis testing: Creation of this infrastructure will enable the testing of a series of important hypotheses regarding phonological development. Once the database has been constructed, consortium members can begin testing core theoretical claims regarding, e.g., prelinguistic development, segmental and prosodic development, acoustic patterns, individual differences, input effects, markedness effects, morphological effects, phonological disorders, and second language acquisition. In most cases, these predictions can be tested directly by examining variations in the production of sounds and sound sequences in different phonological environments across development.
3. Database development: In order to test the various hypotheses mentioned above, we need to compile accurately transcribed data from a large number of children learning many diverse languages. We currently have commitments from 30 members of our initial consortium group to contribute corpora from 16 languages.
4. Benchmark extraction for model testing: We will also use Phon to extract a set of fundamental descriptions of the database that will serve as benchmarks for the testing of generative systems based on learning algorithms.
5. Johnny Appleseed: Once we have initial versions of Phon and the database, we will begin to emphasise dissemination of these new tools through the currently existing CHILDES and TalkBank websites. We will also work to encourage the contribution of new datasets for under-represented language types and participant groups.

Current state of the project

During the first three and a half years of the project, we have focused mainly on goals 1, 3 and 5 above, as their accomplishment is required before we can tackle the other two goals. Version 1.4 of Phon was released in the fall of 2009. It offers improvements both in its functions and its graphical user interface. Phonological corpora made available by members of our consortium can be accessed from dedicated PhonBank sections of the CHILDES database.

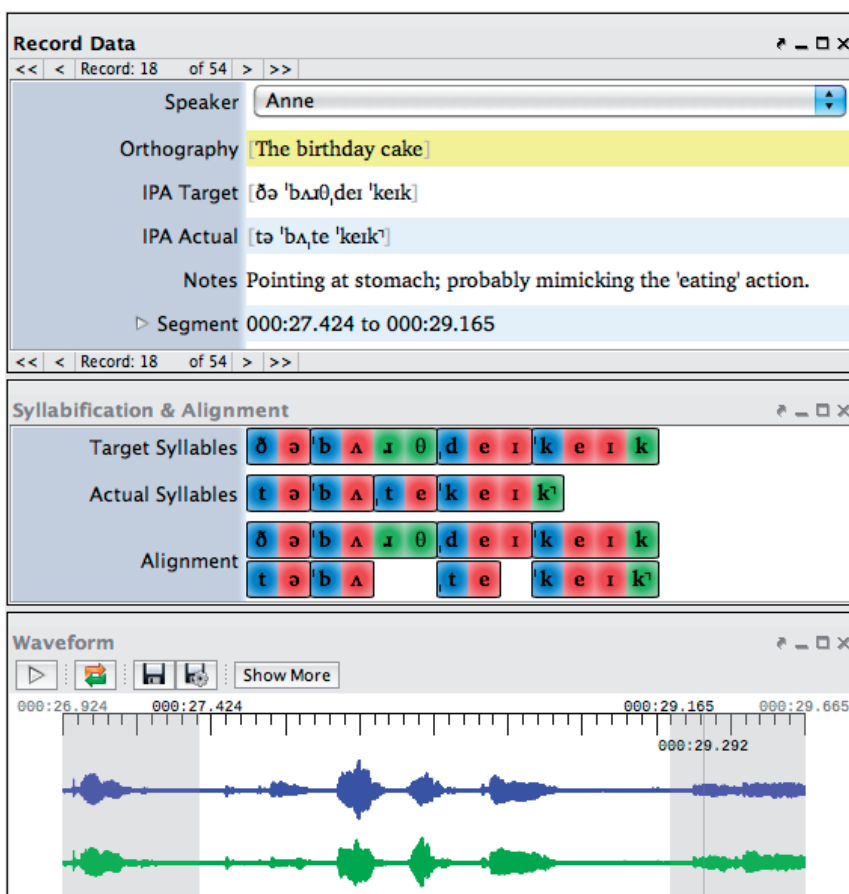
Phon: a general overview

Phon, the software program that supports most of the needs of researchers involved in the contribution and analysis of PhonBank data, is designed to facilitate a number of tasks related to the transcription and analysis of phonological data. Phon supports multimedia data linkage, multiple-blind transcription, unit identification (e.g. utterance, phrase, word), automatic labelling of data (features, syllabification), and systematic comparisons between target (model) and actual (produced) phonological forms. All of these functions are accessible through a user-friendly graphical interface. Databases managed within Phon can also be queried using a powerful search system specifically created for the needs of phonologists.

This software program is freely available as open-source software and runs on both Mac OS X and Windows platforms. The development of Phon is currently supported through a grant from the National Institute of Health. Many scholars as well as several of their students and research associates have joined together to provide their input for the development of Phon and contribute their data to PhonBank. Through this consortium, which has been growing at a steady pace, a large number of people have already been involved in the development and testing of Phon and the development of PhonBank.

Most of Phon's key features have been described in previous publications (e.g. Rose et al. 2007; Rose 2008). In the lines below, I will thus limit the discussion to basic descriptions of these features and refer the interested reader to these works. The descriptions that follow will serve as a useful preamble to a more in-depth discussion of some of the most recent additions and improvements currently available in version 1.4 of the application. This discussion will focus mostly on recent developments in the areas of data searching and reporting. However, many other improvements were made to the application, including a streamlined transcript editor, illustrated in (1).

Researchers already acquainted with the application's interface will notice a move of the tier labels to the left hand side of the interface, which offers more vertical room for data visualisation and editing. Other improvements include less obvious but crucially relevant 'under-the-hood' changes that bring more efficient memory management to the application, which is particularly useful for the processing and searching of large datasets.



1. Transcript editor

Media linkage and segmentation

Using media capabilities in Phon, the user can associate a transcript to a digital audio/video file, and identify the time intervals on the recorded media that are relevant for research (typically, words or utterances). Each time interval identified corresponds to a record in Phon (as in (1)). This time interval can be played back for transcription and related annotation or exported in different formats for further analysis or presentation purposes.

Phonetic transcription and validation

The task of phonetic transcription is an onerous one, but also one of prime importance, since many phonological analyses depend on the quality of the transcriptions contained in the database. It is with such important facts in mind that we endeavour to create a fully-integrated system for multiple-blind transcriptions. Using this system, a virtually unlimited

number of transcribers can perform their transcriptions without access to other transcribers' work. Phon also supports an optional password protection system for blind transcriptions.

The use of blind transcriptions implies the need for validation. The user (or, ideally, team of users) responsible for transcript validation can visualise and compare all blind transcriptions for each utterance. The transcription deemed the most accurate is selected, and can be further modified if needed (e.g. supplemented with details noted in other blind transcriptions or noticed upon playback of the recorded media during validation). While the use of multiple-blind transcription and associated validation systems is optional, only validated transcriptions are available for further data processing and searching. If the user decides not to use the multiple-blind option, the transcriptions are entered directly into the session transcript. In this case, no validation is required; the transcriptions are immediately available for further data processing.

Phon also provides useful functions to streamline the transcription process. These include a built-in IPA map, which helps the selection of IPA symbols, the full set of which is supported by the application. Phon also supports built-in dictionaries of pronounced forms, which provide quick access to generic target IPA forms. Of course, the types of generic citation forms found in these dictionaries do not provide accurate fine-grained characterisations of the target language (e.g. dialect-specific pronunciation variants; phonetic details such as degree of aspiration in obstruent stops). However, when used with the appropriate care, they can help streamline many of the steps involved in phonetic transcription.

Orthography	I play	[(with)]	my sister
IPA Target	['aɪ 'pleɪ]		['maɪ 'sɪstə]
IPA Actual	[ə'pe:]		['mæ 'sɪtə]

2. Word Groups

Word Groups for domain-based coding and analysis

Orthographically or phonetically transcribed utterances in Phon can be broken into smaller domains such as phrases, clitic groups or words, all of which are potential domains for phonological analysis. This further division of the utterance is automatically reflected in the IPA Target and IPA Actual tiers (or other, user-defined tiers). As we can see in (2), the preposition 'with'

is marked as truncated (through parenthetical bracketing) in the Orthography tier. While this coding indicates something about the child's omission of a function word, this information, contained in its own word group, does not interfere with the phonological analysis of the utterance in the IPA Target and IPA Actual tiers.

Searches on aligned Word Groups enable the quick identification of such word omissions on the learner's part. Word Group alignment thus effectively provides a system for 'vertical' analysis across target and actual forms within each utterance, which can also be used to narrow down searches within particular domains.

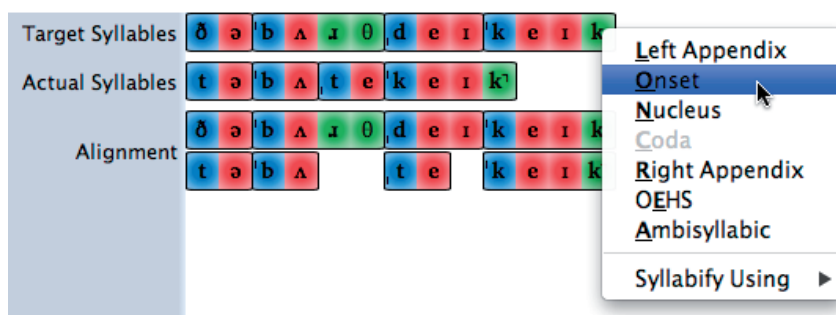
Automatic labelling of features and syllabification

All symbols of the IPA (including diacritics) entered in the IPA Target and IPA Actual tiers are automatically associated with a set of descriptive features. The feature set incorporated into Phon is meant to be as theory-neutral as possible, at least for the phonologist assuming a featural level of phonological encoding (researchers rejecting this theoretical construct do not need to assume it to take advantage of other Phon functions). As such, the feature set supported in Phon does not assume any theory of feature underspecification or implicational relations between redundant features. For example, nasal consonants are redundantly specified for features such as {Nasal}, {Sonorant} and {Voiced}.² The features associated with IPA symbols can be used in many types of queries, independent of the symbols themselves, which facilitates the identification of natural classes. For example, researchers interested in high, front vowels ([i, y, ɪ, ʏ]) do not need to list all the individual symbols; they can access such phones through a simple listing of two features: {High, Front}.

Following the same theory-neutral philosophy as much as possible, we also implemented two automatic algorithms for data labelling. The first consists of a system for the encoding of syllable-level annotations through a deterministic cascade of rules. Syllabification rules can be defined to encode the syllable structure of virtually any language. Several languages are already supported in Phon; we are able to incorporate new syllabification algorithms upon user request. The second algorithm consists of a dynamic programming system that makes 'best-guesses' about the alignment of target and actual word forms. Examples of syllabification and alignment codings can be seen in (3), where the syllable labelling is represented in colours (e.g. onsets in blue, nuclear segments in red, and codas in green)³

2 Feature sets in Phon are specified between braces, with each feature beginning with a capital letter.

3 Other syllable constituents such as appendices and onsets of empty-headed syllables are also supported. They are invoked in algorithms that assume the validity of such



3. Syllabification and alignment

and the alignment is displayed using a vertical organisation between IPA Target and IPA Actual tiers.

Of course, as is the case with all algorithms included in the program, the user is able to perform manual adjustments of the computer-generated labelling whenever necessary. Note also that these two screen shots come from a single record, but with different tier visualisation options. This partly illustrates the fact that Phon users can control the visualisation of each tier as well as the ordering of these tiers in the transcript editor.

Query and reporting functions

After all transcription and data-coding tasks are performed, the user is in a position to start taking full advantage of the application, through its searching and reporting functions. Using specialised functions such as those described in more detail in the next section, the user can build textual strings, regular expressions and ‘phonex’ expressions⁴ to query transcribed data, feature- or syllabification-related information as well as other information relative to the transcribed data (e.g. participants’ names, ages or age ranges). Another feature, crucial in the context of phonological development, is that the user can perform systematic comparisons between IPA Target and IPA Actual forms in order to assess learner’s phonological performance within and across recording sessions. In addition, Phon incorporates support for the detection of consonant and vowel harmony as well as consonant metathesis. The data returned by the query functions can be visualised directly in the application (and further edited if necessary) or saved as reports in many different formats.

constituents.

⁴ *Phonex* expressions use the specialised language created for the need of queries of phonological properties that transcend strings of phonetic symbols. For example, using phonex, the user can specify a search for voiceless coronal stops in onsets, as follows: “{Voiceless, Coronal, Stop}: Onset”.

The study of second language phonological phenomena, and beyond

As the above description suggests, the PhonBank initiative can facilitate several aspects of research into the second language acquisition of phonology. Similarly, specialists of loanword phonology can take advantage of many functions available in Phon to assess patterns of loanword adaptation within and across languages. Finally, the PhonBank project, through its natural connection to the TalkBank database, can facilitate research on the phonological characterisation of language dialects. In the subsections that follow, I address these topics in turn. Before going into the heart of these matters, I begin in the next subsection with a brief but significant observation about the field of second language research in phonology.

A word about data sharing

The field of study in second language phonological development currently suffers from many of the symptoms that affected the field of first language development prior to the inception of the PhonBank initiative, namely that it is extremely difficult at the moment to access second language phonological data from public sources. The issue is not that data corpora do not exist; they do, and are promoted on the websites of various research groups and within the scientific literature. While these data are at times being shared, the practice is too often constrained by conditions that prevent any type of real public dissemination. This situation, both unfortunate and unnecessary, could, and should, be improved — as it has, and continues to, in first language phonology— through the active contribution of researchers specialising in second language acquisition phonology who administer existing corpora, and through the elaboration of new, publicly-accessible corpora. Such a positive change, especially if effected through the PhonBank initiative, would offer second language phonology researchers unprecedented technological support for their corpus-based studies as well as a ready infrastructure for the data sharing. Bringing about these important improvements to the current state of affairs would also open the possibility of new and exciting research, enabling systematic comparisons between first and second language phonological development and between various populations of learners, from children being raised with two or more mother tongues, to early or later bilingual development, for example. Whichever the field of empirical or theoretical interest, everyone involved would gain from such a positive change. In the subsections that follow, I illustrate a few types of investigations enabled by PhonBank, with the hope of enticing second language specialists to join in this collaborative endeavour.

Some illustrations of Phon's query and reporting functions

Second language phonological development

As Eckman, Elreyes & Iverson (2003) put it, in order to learn a target language, the second language learner must develop a lexicon, including the various phonological and morphological properties regulating the target language's lexical items, and learn to combine the lexical items into multiple-word utterances. Second language learners must also acquire the pronunciation rules of the target language, a task that will most inevitably be influenced by the phonological properties of the language(s) already acquired. Such influences are often referred to in terms of phonological transfer effects (e.g. Lado 1957; Flege 1987; Major & Kim 1999). Each aspect of transfer comes with its own set of theoretical questions, for example about the learnability of phonological contrasts (e.g. Escudero & Boersma 2002), several of which call for the types of phonological investigations that PhonBank facilitates.

Concretely, from an empirical perspective, we observe in second language phonological development many of the same processes that are observed in first language development, although the sets differ both qualitatively and quantitatively. For example, while processes such as consonant harmony and velar fronting are documented in vast bodies of literature on first language development, they are virtually absent from the literature on second language development. Also in contrast to first language development, we generally observe what appears to be significantly higher rates of vowel epenthesis in second language phonological productions, especially in the speech of learners whose first language(s) display syllable properties that are descriptively simpler than those of the target language. Details such as these aside, however, studies of second language phonological patterns often rely on target-actual comparisons similar to those that must be performed for first language development. The methods and tools developed within the PhonBank initiative are thus readily available to researchers and students of second language phonological acquisition. This is illustrated in (4) where we can see that Phon enables the identification of segmental discrepancies between the target (model) French word *citrouille* [sitruj] 'pumpkin' and its actual production by a Japanese learner of French.

As we can see in this example, the software correctly aligns all IPA Target phones with their IPA Actual counterparts, thereby setting the stage for investigations of the segmental properties of the second language productions relative to the target model (e.g. target [s] produced as [ʃ]). Similarly, the automatic alignment enables a ready identification of word-medial and word-final vowel epenthesis sites, each of which can be uniquely identified and extracted using the application's query system. For example,

SpeakerJapanese Informant

Orthographycitrouille

IPA Targetsitruj

IPA Actualsituruju

<<<Record: 63 of 63>>>

Syllabification & Alignment

Target Syllables

Actual Syllables

Alignment

4. Japanese pronunciation of French word *citrouille* [sitruj] ‘pumpkin’ focussing on patterns of vowel epenthesis that break up target onset consonant clusters, one can use the Aligned Phones search system to identify all of the relevant cases present in the database, as shown in (5).

Aligned Phones

Target phonex{}:Onset {}:Onset+

Actual phonex{} {}*

☐ Search actual before target

5. Aligned Phones query: realisation of target onset clusters in actual productions

In this example, the user employs phonex to search for two or more consecutive onset consonants in IPA Target (the ‘more’ part of this expression is encoded with the “+” operator), and requests a report on the types of phone sequences that result from it in the corpus. As we can see in (6) from a tiny set of representative examples, the application identifies three matching patterns, all of which involve vowel epenthesis, identified by the addition of a segment in position 2 of the aligned string (“+P2”).

The screenshot shows a transcript editor interface. On the left, the 'Speaker' is 'Japanese Informant', 'Orthography' is 'citrouille', 'IPA Target' is [sitruj], and 'IPA Actual' is [ʃitruju]. Below this is a 'Syllabification & Alignment' section showing 'Target Syllables' (s i t r u j), 'Actual Syllables' (ʃ i t u r u j u), and 'Alignment' (s i t r u j / ʃ i t u r u j u). On the right, a 'Search 7' panel shows 'Query # 3', 'Search # 7', 'Number of results 1', and 'Records with res... 1/63'. Below this is a table with 'Record #' and 'Result' columns. The first record is '63' with the result 'tr ↔ tur +P2'. Below the table, a list of phonetic changes is shown: 'Deletion', 'Epenthesis P2', 'Match P1', and 'Substitution P3'.

6. Search results visualised in the transcript editor

As we can see in this example, the application returns two cases of [o] epenthesis and one of [u] epenthesis. According to Shinohara (2004), the difference between these two cases of epenthesis is regulated by an allophonic rule in Japanese that requires mid, as opposed to high, round vowels after coronal

7. Aligned Phones query: realisation of onsets clusters beginning with coronal stops

The screenshot shows the 'Aligned Phones' query window. It has two input fields: 'Target phonex' with the value '{Coronal, Stop}:Onset {}:Onset +' and 'Actual phonex' with the value '{} {}*'. Below these fields is a checkbox labeled 'Search actual before target' which is currently unchecked.

7a. Query string

The screenshot shows the transcript editor interface with the 'Aligned Phones' query results. The 'Speaker' is 'Japanese Informant', 'Orthography' is 'citrouille', 'IPA Target' is [sitruj], and 'IPA Actual' is [ʃitruju]. The 'Syllabification & Alignment' section is the same as in the previous screenshot. On the right, a 'Search 8' panel shows 'Query # 4', 'Search # 8', 'Number of results 4', and 'Records with res... 4/63'. Below this is a table with 'Record #' and 'Result' columns. The results are: '10 tr ↔ tr', '11 dr ↔ der +P2', '53 tl ↔ tar +P2', and '63 tr ↔ tur +P2'.

7b. Results

plosives. In order to check the reliability of this pattern in the corpus, one can search for either the realisation of branching onsets whose first consonant is a coronal stop, or else make an inventory of all coronal stops followed by non-low round vowels in the actual forms. The former option and its results are exemplified in (7) from the same mini-sample, which now contains only results from target onset clusters that begin with a coronal.

While the preceding examples provide a simple illustration of the query system, many more types of queries can be performed by Phon's search system. In the next subsection, I move the discussion to the phonological adaptation of loanwords, a phenomenon that is pervasive in societal contexts involving languages in contact.

Loanword Phonology

As discussed in Rose & Demuth (2006), many factors may influence the transformations that affect a foreign word which is incorporated into the lexicon of the language borrowing it. Many of these transformations are driven by phonological properties of the borrowing language such as the system of segmental contrasts it displays. Situations of languages in contact that involve a sufficiently high rate of bilingualism favour such phonological adaptations (e.g. Paradis & LaCharité 1997; see also Grosjean 1982).

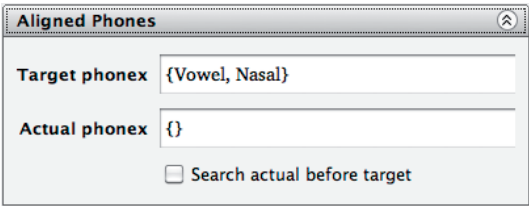
Of particular interest in this context are the actual patterns of adaptation, which often reveal properties of the borrowing language that would not otherwise be observed, because neither the morphology nor the syntax of the language provides the relevant contexts for the expression of these properties (e.g. Shinohara 2004). The systematic study of phonological patterns of loanword adaptation can thus shed light on intricate aspects of phonological systems.

Phon readily supports corpus-based investigations of loanwords which involve systematic comparisons between source words and their corresponding, adapted loanwords. In Phon, this translates into comparisons between target and actual forms. In the lines that follow, I illustrate how the search and reporting systems in Phon can be used to tackle the segmental adaptation of French nasal vowels in Kinyarwanda. For this demonstration, I use a subset of the corpus assembled by Rose (1995). This subset consists of 62 words that begin with the letter A. The full version of this corpus will soon become publicly available through PhonBank. The methods discussed on the basis of these examples can be extended to all other source and borrowing languages.

Aside from the fact that Kinyarwanda, in contrast to French, lacks any type of nasal vowel, the two languages have other properties relevant to the problem. Kinyarwanda, a strictly CV language, displays phonological vowel length contrast in word-medial positions (which is however neutralised at word

edges). Kinyarwanda also allows for prenasalised stops and fricatives, with the peculiarity that prenasalised voiceless stops are debuccalised to [h], with the place of articulation of the consonant expressed in the prenasal portion of the consonant (e.g. /mp/ → [m̥h]; /nt/ → [n̥h]; /ŋk/ → [ŋ̥h]; see Kimenyi, 1979 and Jouannet, 1983 for phonological descriptions of Kinyarwanda).

8. French nasal vowels in Kinyarwanda: general overview



8a. Aligned Phon search

Result	Count	
ẽ ↔ e	1	The Aligned Phones search of the sample corpus illustrated in (8a) reveals two distinct patterns of adaptation: aside from issues pertaining to vowel place of articulation, out of the 31 occurrences of nasal vowels in the sample corpus, 21 are realised as short, while 10 vowels are realised as long. This is illustrated in (8b). While the distribution of long versus short vowels in the adapted forms may look random from the summary above, a closer examination of the results, however, suggests that position within the word may be influencing it. This is illustrated with the word <i>ambulance</i> in (9), which contains two nasal vowels in
ĩ ↔ o	10	
ĩ ↔ o:	1	
ã ↔ e:	2	
ã ↔ a	9	
ã ↔ a:	7	
ẽ ↔ a	1	

8b. Result summary
(as rendered in Phon
Report format)⁵

the source form. As we can see, while the two vowels yield prenasalisation of the following consonant, the vowel in word-initial position is realised as short [a], while word-medial vowel is long [aː].

A further exploration of this working hypothesis confirms that the pattern is indeed stable; all of the 21 French word-initial and word-final

5 Phon reports can also be exported in CSV, PDF, Excel, Word, OpenOffice and HTML formats.

Record Data		Record #	Result
Speaker	Japanese Informant	22	ã ↔ a
Orthography	ambulance	22	ã ↔ a:
IPA Target	[ãbylãs]	24	ã ↔ a:
IPA Actual	[ãmbira:nsi]	27	ã ↔ a
		30	ã ↔ a
		31	ñ ↔ n:

9. Example result: *ambulance* [ãbylãs] → [ãmbira:nsi]

nasal vowels found in the sample corpus are realised as short while the word-medial ones undergo lengthening, as illustrated in the screen shot in (10).⁶

As we can see from this short demonstration the use of searching and reporting capabilities built into Phon enables linguists to make efficient and systematic observations and empirical verifications of their working hypotheses.

Orthography	atlantique	44	ã ↔ e:
IPA Target	[atlatik]	44	ñ ↔ o
IPA Actual	[atareñhika]	47	ã ↔ a
		48	ñ ↔ o
		49	ñ ↔ o
		49	ñ ↔ o
		50	ã ↔ a:
		53	ã ↔ a:
		56	ã ↔ a:
		56	ñ ↔ o
		58	ã ↔ a:

Syllabification & Alignment	
Target Syllables	a t l ã t i k
Actual Syllables	a t a r a: ñ h i k a
Alignment	a t l ã t i k a t a r a: ñ h i k a

10. Sample search and results: nasal vowel adaption in word-medial position

The study of dialectal variation

The two preceding subsections focus on phonological issues that relate primarily to bilingualism and bilingual societies. The tools provided by PhonBank, however, enable studies in many more areas of investigation, for example in the documentation of dialectal varieties of languages. Research in such areas serves various purposes, for example the establishment of standards for education or speech-related diagnosis and treatment protocols. Also, while many aspects of the phonology of language dialects such as

⁶ Note as well in this example the debuccalisation of the prenasalised voiceless [t], which fits the Kinyarwanda phonotactics described above.

the several documented dialects of world Englishes can be considered to be relatively stable, corpus-based studies are especially relevant in cases where variation within or across sub-dialects is observed. Such studies can serve the identification of phonological and/or socio-linguistic factors governing the patterns of variation observed.

Thus far, the functions available in Phon limit support to transcription-based studies. However, given the fact that the devil typically hides in the details, acoustic investigations are often required in order to describe the actual properties of a given dialectal variant (e.g. Pillai, Mohd Don, Knowles & Tang, in press, on vowel place of articulation in Malaysian English). While, Phon does not currently handle acoustic measurement data, adding such support is high on the list of priorities for upcoming versions of the application. The vision behind this future development is that of providing support for the storage and compilation of acoustic data generated through speech analysis software such as Praat, Speech Filing System or Computerised Speech Lab. Also explicit in this vision is the important fact that this extension to Phon does not aim at replacing any of these applications but rather supplementing them within a better-supported workflow. Among other possibilities, this system will enable the compilation of acoustic measurements based on phonological criteria such as those exemplified above.

Conclusion

In this paper, I have discussed the PhonBank initiative in the context of second language development and the related contexts of loanword phonology and, to a lesser extent, the study of dialectal variation. In all cases, the tools available through PhonBank such as those implemented in Phon already offer a great deal of technological support for further methodological, empirical and theoretical advances in these areas. More support will come through the continuation of this project, including the management and compilation of acoustic measurement data, which will provide further support for acoustic analyses of speech. Combined with existing features of Phon such as those described in the preceding sections, this additional function will also provide technological support for studies integrating both phonological (e.g. features, syllable structure) and acoustic (e.g. formant structure) datasets.

Building on this discussion, I also emphasised the fact that the vast majority of phonological corpora documenting second language development that could be studied systematically using these tools are generally not available in any kind of public form, despite the discussion of many such corpora in the body of scholarly publications and other venues. This situation needs to change in the same way it has changed for first

language phonology since the inception of the PhonBank initiative, thanks to active members of our research consortium. Because of the data sharing philosophy behind PhonBank, the active contribution of a maximum number of interested members of the research community is crucial in order to bring this project to full fruition. Such contributions can for example take the form of feedback on existing versions of the Phon application and, central to the success of this communal endeavour, the provision of research data to PhonBank, which will directly benefit our understanding of the mechanisms and factors that affect phonological development.

REFERENCES

- CHILDES [On-line database]. <http://childes.psy.cmu.edu/>
 Computerized Speech Lab [Speech analysis system]. <http://kayelemetrics.com/>
 Eckman, Fred R., Elreyes, A., & Iverson, Gregory K. (2003). Some principles of second language phonology. *Second Language Research*, 19(3), 169-208.
 Escudero, P., & Boersma, P. (2002). The subset problem in L2 perceptual development: Multiple-category assimilation by Dutch learners of Spanish. In B. Skarabela, S. Fish & A. H.-J. Do (Eds.), *Proceedings of the 26th Annual Boston University Conference on Language Development* (pp. 208-219). Somerville, MA: Cascadilla Press.
 Flege, James E. (1987). The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15, 47-65.
 Grosjean, F. (1982). *Life with two languages: An introduction to bilingualism*. Cambridge, MA: Harvard University Press.
 Jouannet, F. (1983). Phonétique et phonologie: le système consonantique du kinyarwanda. In F. Jouannet (Ed.), *Le kinyarwanda: langue bantu du Rwanda* (pp. 55-73). Paris: SÉLAF.
 Kimenyi, A. (1979). *Studies in Kinyarwanda and Bantu phonology* (Vol. 33). Edmonton, Canada: Linguistic Research.
 Lado, R. (1957). *Linguistics across cultures*. Ann Arbor, Michigan: University of Michigan Press.
 MacWhinney, B. (1991). *The CHILDES project : Tools for analyzing talk*. Hillsdale, NJ: Lawrence Erlbaum.
 MacWhinney, B. (1996). The CHILDES system. *American Journal of Speech-Language Pathology*, 5(1), 5-14.
 Major, Roy C., & Kim, E. (1999). The similarity differential rate hypothesis. In J. Leather (Ed.), *Phonological issues in language learning: A supplement to language learning* (pp. 151-183). Ann Arbor, MI: Basil Blackwell.
 Paradis, C., & LaCharité, D. (1997). Preservation and minimality in loanword adaptation. *Journal of Linguistics*, 33, 379-430.
 Pillai, S., Mohd Don, Z., Knowles, G., & Tang, J. (in press). An instrumental analysis of vowel contrast in Some varieties of Malaysian English. *World Englishes*, 29 (2).
 Phon [Software program]. <http://phon.ling.mun.ca/phontrac/>
 Praat [Software program]. <http://www.fon.hum.uva.nl/praat/>
 Rose, Y. (1995). *Minimalité, préservation et tolérance dans les emprunts français en kinyarwanda*. Unpublished M.A. Thesis, Université Laval, Québec, Canada.
 Rose, Yvan. 2008. Phon 1.3: Current Features and Short Term Outlook. Child Language Bulletin 28. <http://iascl.talkbank.org/bulletins/bulletinV28N1.html>
 Rose, Y., & Demuth, K. (2006). Vowel epenthesis in loanword adaptation: Representational and phonetic considerations. *Lingua*, 116, 1112-1139.

- Rose, Y., Hedlund, Gregory J., Byrne, R., Wareham, T., & MacWhinney, B. (2007). Phon 1.2: A computational basis for phonological database elaboration and model testing. In P. Buttery, A. Villavicencio & A. Korhonen (Eds.), *Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition, 45th Annual Meeting of the Association for Computational Linguistics* (pp. 17-24). Stroudsburg, PA: ACL.
- Shinohara, S. (2004). Emergence of universal grammar in foreign word adaptation. In R. Kager, J. Pater & W. Zonneveld (Eds.), *Constraints in Phonological Acquisition* (pp. 292-320). Cambridge: Cambridge University Press.
- Speech Filing System [Software program]. <http://www.phon.ucl.ac.uk/resource/sfs/>

This book is the fruit of the first English Pronunciation: Issues & Practices (EPIP) conference, which took place at the University of Savoie, France, in June 2009. Researchers and teachers from sixteen different countries came together to discuss: phonetic variations and phonological changes; varieties, identity and their implications for teaching; and the use of new technologies in research and in the classroom.

The twelve papers selected for publication reflect EPIP's prime objective: the creation of bridges between researchers and teachers from various backgrounds (EFL, ESL, EAP, ESP, language acquisition, etc.). Such sharing is essential for the development of new theories and effective teaching methods.

Alice Henderson is responsible for coordinating ESP teaching within the Humanities faculty at the Université de Savoie.

20 €



ISBN: 978-2-915797-73-2
ISSN: 1952-0891