# A note on the generation of allophonic rules

Luc Boruta

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

# *A note on the generation of allophonic rules*

Luc Boruta

**N° 0401**

Janvier 2010

Audio, Speech, and Language Processing

Rapport technique

ISSN 0249-0803 ISRN INRIA/RT-0401--FR+ENG

INSTITUT NATIONAL
DE RECHERCHE
EN INFORMATIQUE
ET EN AUTOMATIQUE

INRIA

centre de recherche
PARIS - ROCQUENCOURT

# A note on the generation of allophonic rules

Luc Boruta *

Theme : Audio, Speech, and Language Processing
Perception, Cognition, Interaction
Équipe-Projet Alpage

Rapport technique n° 0401 — Janvier 2010 — 10 pages

**Abstract:** Algorithms modeling the acquisition of allophonic rules by infants need a 'standard' allophonic grammar against which to evaluate their results. Because no wide-covering grammar exist in the literature and, most importantly, because allophonic complexity is a parameter that needs to be controlled, a common workaround has been to apply grammars of artificial allophonic rules to phonemically-transcribed corpora. We present a new algorithm to generate such allophonic rules, enforcing their linguistic plausibility using a description of phonemes in terms of distinctive features. Controlling the size of the allophonic grammar, we are able to generate transcriptions of various phonetic granularities.

**Key-words:** allophonic variation, computational linguistics, early language acquisition, phonology.

* The author would like to thank Benoît Crabbé, Emmanuel Dupoux and Sharon Peperkamp for the valuable insights they provided as this algorithm was developed.

Centre de recherche INRIA Paris – Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex
Téléphone : +33 1 39 63 55 11 — Télécopie : +33 1 39 63 53 30

# Note sur la génération de règles allophoniques

**Résumé :** Les algorithmes modélisant l'acquisition précoce des règles allophoniques ont besoin de grammaires allophoniques de référence pour évaluer leurs résultats. Cependant, aucune grammaire à large couverture n'a encore été décrite et, surtout, la complexité allophonique est un paramètre qu'il est souhaitable de contrôler. Pour pallier ces problèmes, il est d'usage d'appliquer des grammaires de règles allophoniques artificielles à des corpus de parole adressée aux enfants transcrits phonémiquement. Nous présentons un nouvel algorithme permettant de générer de telles règles. La plausibilité linguistique des règles est garantie par l'utilisation d'une description en traits distinctifs des phonèmes. En contrôlant la taille de la grammaire allophonique, nous sommes capables de de générer des transcriptions présentant des granularités phonétiques variées.

**Mots-clés :** acquisition précoce du langage, linguistique informatique, phonologie, variation allophonique.

# 1 Introduction

Algorithms modeling the acquisition of allophonic rules by infants need a 'standard' allophonic grammar against which to evaluate their results. Even if such grammars may be compiled from the linguistic literature for some well-described languages, we do not know how many phonetic and phonological variants infants actually process when they learn the phonemic categories of their native language. Therefore, to evaluate the algorithms' robustness without a precise idea of the true complexity of the task, we will want to vary the number of phoneme variants as a parameter. As no database of infant-directed speech[1] containing rich phonetic transcriptions or high-quality aligned audio from which such rules could be derived has been released yet, a common workaround in the absence of such resources has been to apply grammars of (attested or artificial) allophonic rules to phonemically-transcribed corpora of infant-directed speech [5, 6, 9, 10]. The automatic application of the rules thus systematizes contextual variation between adjacent segments. Furthermore, controlling the number of rules in the grammar allows for the creation of a continuum of possible inputs, ranging from coarse-grained to fine-grained phonetic transcriptions.

Peperkamp, Le Calvez et al. [5, 6, 10] applied sets of hand-written allophonic rules, e.g. the palatalization rule of the velar consonants /k, g/ before the front vowels and glides /i, y, e, ɛ, ø, œ, j, ɥ, ɛ̃/ in French. This method has an obvious advantage as rules are not only linguistically plausible, but also attested in the language at hand. However, the small allophonic grammars used in these studies introduced a low number of contextual variants: Le Calvez et al. [6], for example, implemented only 11 allophonic rules. Martin et al. [9] recently circumvented this problem using a simple language-independent algorithm where allophones are nothing but numbered versions of the target phoneme, e.g. using [$\mathrm{p}_1$, $\mathrm{p}_2$, $\mathrm{p}_3$] as contextual realizations of /p/. Concretely, to generate $n$ allophones of a given phoneme, the set of all its possible contexts (i.e. the phonemic inventory and the utterance boundary) is split into $n$ partitions, and each partition is then randomly tied to one numbered allophone. However, as allophones are conditioned partitioning the set of all possible contexts, and not the set of all attested contexts, this algorithm may generate rules which cannot apply because of phonotactic constraints. If no member of a partition is attested as a context of the target phoneme, e.g. a rule describing the realization of /ɥ/ before /k/ in French, then the allophone will never be realized. The actual granularity of the emulated phonetic transcription may thus be far less than expected.

We present a new algorithm for generating allophonic rules, combining the linguistic plausibility of Peperkamp et al.'s rules and Martin et al.'s ability to emulate rich transcriptions. We first describe the algorithm, and then present experimental results showing we can outperform Martin et al.'s algorithm in increasing the granularity of the phonetic transcriptions we emulate.

[1] Infant-directed speech differs from adult-directed speech, and there is evidence that it has special properties that facilitate learning [4]. Though phonetically-transcribed databases of adult-directed speech exist, they are useless for our purposes. See, however, work by Dautriche using adult-directed speech, where allophones are defined in terms of acoustic similarity [2].

# 2 Feature-based generation of allophonic rules

The major contribution is to enforce the linguistic plausibility of the rules we generate. The plausibility of an artificial allophonic rule depends on the relation between the target phoneme, the application context and the allophone. As the vast majority of allophonic rules are assimilatory, we want to guarantee that the allophone shares some properties with both the target phoneme and the context. For example, a rule characterizing voicing assimilation describes a common phenomenon, even if it is not attested in the language at hand. By contrast, a rule where the allophone has nothing in common with neither the target phoneme nor the application context, such as a → ɥ / _ k, is unlikely.

The general idea is to create assimilatory rules whose application contexts span similar contexts of the target phoneme. To do so, we represent the internal structure of the phonemes in terms of distinctive features: similarity between contexts can thus be either articulatory or acoustic, depending on the features. In order to maximize the number of allophones, each attested context of the target phoneme must fall in the scope of exactly one rule. As a consequence, if the desired number of allophones is inferior to the number of contexts, the latter must be aggregated into as many exclusive clusters as desired allophones. We represent a context cluster by an under-specified phonological matrix whose features are the ones shared by the members of that cluster and, to make the rules as plausible as possible, maximize phonological similarity between members.

In line with previously reported experiments [5, 6, 9, 10], we make two simplifying assumptions about the nature of the allophonic rules we generate. First, all rules are of the type $p \to a$ / _ $c$ where a phoneme $p$ is realized as its allophone $a$ before context $c$. Second, we ensured that no two allophonic rules introduced the same allophone (as in English flapping, where both /t/ and /d/ have an allophone [ɾ]) using parent annotation: each phone is marked by the phoneme it is derived from (e.g. $[\text{ɾ}]^{/\text{t}/}$ and $[\text{ɾ}]^{/\text{d}/}$).

## 2.1 Algorithm

The description of the algorithm relies on a few operations with phonological attribute-value matrices:

- a matrix $m$ is an **extension** of another matrix $m'$ if and only if all attributes in $m'$ are specified in $m$ with equal values;
- the **fusion** of two matrices $m$ and $m'$, $m \circ m'$, is defined as follows: if an attribute is specified in only one of the two matrices, then it is specified in the resulting matrix with the same value; else, if it is specified in both, then it is specified in the resulting matrix with the value it has in $m$';
- the **fission** of a matrix $m$ by an attribute $a$, $m \star a$, yields two new matrices combining the information in $m$ and one of the values $a$ can take, i.e. $m \star a = \{m \circ [-a], m \circ [+a]\}$ in the case of traditional binary attributes.

Starting from an empty matrix, of which any phoneme is an extension, the algorithm successively fission matrices, incorporating more and more features. Each of the final matrices is the application context of an allophonic rule of the type $p \to a$ / _ $c$. Such rules are assimilatory because we define the allophone as the fusion of the target phoneme and its context. Let $p$ be a phoneme and $C_p$ the set of its attested contexts, the procedure to generate $n \leq |C_p|$ rules can be detailed as follows:

1. initialize the set of context clusters $C$ with an empty matrix, and a buffer $B$ as an empty set;
2. if $B$ is empty, then:
   (a) pick at random and remove a matrix $m$ from the subset of the smallest matrices in $C$;
   (b) let $C'_p$ be the subset of $C_p$ whose members are extensions of $m$, and $a$ one of the attributes yet not specified in $m$ whose values partition $C'_p$ in subsets as evenly-sized as possible: fill $B$ with $m \star a$;
3. pick at random and remove a matrix from $B$, and add it to $C$;
4. repeat from step 2 while $|C| < n$;
5. for each context $c \in C$, create a new rule $p \to p \circ c$ / _ $c$.

As shown in Figure 1, the successive states of the set of contexts $C$ can be represented as a tree where the leaves are the actual members of the set, and the history of the fissions can be recollected by a walk up to the root.

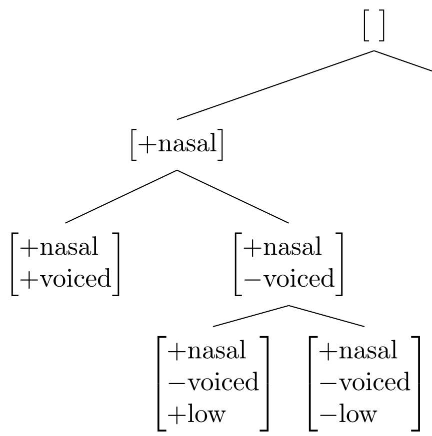


Figure 1: Tree representation of application contexts generated by successive fissions of under-specified phonological matrices.

## 2.2 Example outputs

Examples of artificial allophonic variation, using Dell's articulatory description of French phonemes [3], are presented in Figures 2 and 3 where the original phoneme is at the top of the matrix and feature values modified by rule application are circled.

| | k | a | n | a | r | f | u |
|---|---|---|---|---|---|---|---|
| front | − | − | + | − | − | + | − |
| back | + | + | − | + | + | − | + |
| low | − | + | − | + | − | − | − |
| continuant | ⊕ | + | + | + | + | + | + |
| consonantal | + | − | + | − | + | + | − |
| coronal | − | − | + | − | − | − | − |
| high | + | − | − | − | − | − | + |
| nasal | − | − | + | − | − | − | − |
| delayed release | − | + | − | + | + | + | + |
| round | − | − | − | − | − | − | + |
| sonorant | − | + | + | + | + | − | + |
| syllabic | − | + | − | + | − | − | + |
| voiced | ⊕ | + | + | + | ⊖ | − | + |

Figure 2: Artificial allophonic variation of the French phrase /kanarfu/ (*canard fou*, 'mad duck').

| | i | k | u | r | a | p | r | ɛ | l | ə | k | o | ʃ | ɔ̃ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| front | − | − | − | − | ⊕ | + | − | + | + | ⊖ | − | ⊖ | − | − |
| back | − | + | + | + | + | − | ⊖ | − | − | + | + | + | ⊕ | + |
| low | − | − | − | ⊕ | + | − | ⊕ | + | ⊕ | + | − | − | − | + |
| continuant | ⊖ | − | + | + | + | − | + | + | + | ⊖ | − | + | + | + |
| consonantal | ⊕ | + | ⊕ | ⊖ | ⊕ | + | ⊖ | ⊕ | ⊖ | − | + | ⊕ | + | − |
| coronal | − | − | − | − | − | − | − | ⊕ | + | − | − | ⊕ | + | − |
| high | + | + | + | − | − | − | − | − | − | − | + | ⊕ | + | − |
| nasal | − | − | − | − | − | − | − | − | − | − | − | − | ⊕ | + |
| delayed release | + | − | + | + | ⊖ | ⊕ | + | + | ⊖ | − | − | + | + | + |
| round | − | ⊕ | + | ⊖ | − | − | − | − | − | − | ⊕ | + | ⊕ | + |
| sonorant | + | − | + | + | + | ⊕ | + | ⊕ | + | − | − | − | − | + |
| syllabic | + | ⊕ | + | − | + | − | − | + | − | + | ⊕ | + | ⊕ | + |
| voiced | − | − | + | + | ⊖ | − | + | + | + | ⊖ | − | ⊖ | − | + |

Figure 3: Artificial allophonic variation of the French utterance /ikuraprɛləkoʃɔ̃/ (*il court après le cochon*, 'he runs after the pig').

# 3 Evaluation

The amount of contextual variation the application of an allophonic grammar introduces in a corpus can be quantified by the average number of allophones per phoneme, i.e. the ratio between the number of attested phones in the now phonetically-transcribed corpus and the number of phonemes in the language. We refer to this quantity as the **allophonic complexity** of the corpus.

## 3.1 Method

We compare our algorithm to a variant of Martin et al.'s where numbered versions of the target phoneme are randomly tied to partitions of the set of its attested contexts. While preventing the generation of allophonic rules that cannot apply, this modification does not alter the core idea behind their algorithm. Moreover, and for both algorithms, we do not consider the utterance boundary as a possible application context as it can not be suitably described in terms of features, would they be articulatory or acoustic.

We use corpora of three typologically different languages: English, French and Japanese. All were extracted from transcribed adult-infant verbal interactions collected in the CHILDES database [7] and were derived for the purpose of previously published experiments [1, 5, 6, 10]. The English corpus is a subset of the *Bernstein-Ratner* corpus that was transcribed using an inventory of 50 phonemes [1]. To make the French corpus, the *Champaud*, *Leveillé* and *Rondal* corpora were automatically transcribed and then manually corrected to match a set of 35 phonemes [5]. The Japanese sample was derived by automatically transcribing the *Ishii* and *Noji* corpora from rōmaji, using an inventory of 49 phonemes [5]. All transcription choices made by the authors were respected.

Both algorithms have a single parameter to vary: $n$, the desired allophonic complexity. For each language, starting from the phonemically-transcribed corpus, we generate corpora of increasing allophonic complexity, testing the following values of $n$: from 1 to 5, with increments of 1, and from 5 to the numbers of phonemes (included) in the language at hand, with increments of 5. Then, for each desired complexity, we compare the attested allophonic complexity in each algorithm's output. Due to random picks, both algorithms are non-deterministic processes: all scores reported below are the average of the values obtained over three distinct runs for each model, corpus and desired number of allophones.

## 3.2 Results and discussion

Attested allophonic complexities for both algorithms and all three languages are presented in Figure 4 as functions of the desired complexities. The main observation is that the nature of the rules —i.e., whether contexts are aggregated at random or based on a linguistic description of the phonemes— does not influence the shape of the curves but their growth rate and, as a consequence, the maximal allophonic complexities that can be introduced. It is also worth noting that introducing linguistic constraints in the generation of artificial allophonic rules does not reduce the maximal allophonic complexity, quite to the contrary. Indeed, our feature-based algorithm outperforms Martin et al.'s in all three languages: on average, we were able to introduce up to 23 contextual variants per phoneme in the English corpus (vs. 18 with random ties), 22 (vs. 17) in the French one and 11 (vs. 9) in the Japanese sample.

Both algorithms generate rules with monolateral application contexts and can therefore theoretically produce as many rules as the square of the number of phonemes, describing in this way the realization of each phoneme before any phoneme. Obviously, such allophonic complexities cannot be reached as not all phonemes occur in all contexts: while some are subject to phonotactic sequencing constraints (e.g. /ɥ/ in French, which only occurs before a vowel), others simply occur very infrequently (e.g. /ʒ/ in English). Similarly, lower values for Japanese are not especially surprising as strict phonotactics constrain the maximal syllabic template to CVC, e.g. /hon/. By contrast, because consonant clusters are legal in these languages, syllables such as /strikt/ (CCCVCC) and /dɛkstr/ (CVCCCC) can be observed in French and, equivalently, /stɹɪkt/ (CCCVCC) and /twɛlfθs/ (CCVCCCC) in English. In other words, whereas the possible contexts of a Japanese consonant are almost exclusively vowels, the contexts of most English and French consonants include both vowels and

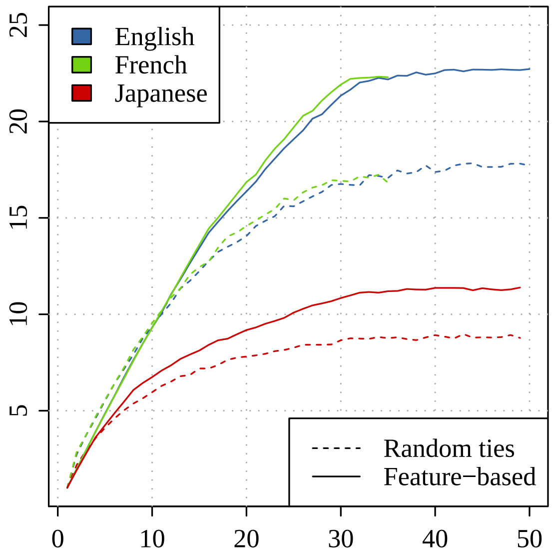


Figure 4: Attested allophonic complexity as a function of the desired complexity.

consonants. As a consequence, models of allophonic variation that only consider adjacent segments as contexts cannot create as many rules for languages with strict phonotactic constraints, such as Japanese, as they do with English or French.

# 4 Conclusion

We described a new algorithm to generate artificial assimilatory allophonic rules which, when applied to a phonemically-transcribed corpus, can emulate phonetic transcriptions. Moreover, controlling the size of the output allophonic grammar, we were able to generate transcriptions of various phonetic granularities. The only *a priori* linguistic knowledge the algorithm requires is a description of the phonemic inventory in terms of distinctive features. Most importantly, this algorithm satisfies two important properties: a significant number of contextual variants can be introduced for each phoneme and, though artificial, the rules are linguistically admissible.

Although we are looking forward to the setting up of phonetically-transcribed databases of infant-directed speech, this algorithm could be improved in various ways to better emulate phonetic transcriptions, for example by controlling the linguistic well-formedness of the phonological matrices during successive fissions. Furthermore, modeling 'sandwich' rules, of the type $p \to a \ / \ c_1 \_ c_2$, with bilateral contexts would increase the number of attested contexts for each phoneme and, as a consequence, help to reach greater allophonic complexities. Most of all, considering bilateral contexts would improve the plausibility of the output rules not only from a linguistic point of view, but also with respect to speech processing studies [2, 8] where various acoustic models represent contextual realizations as triphones.

# References


[1] M. R. Brent and T. A. Cartwright. Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, 61:93–125, 1996.

[2] I. Dautriche. Modélisation des processus d'acquisition du langage par des méthodes statistiques. Master's thesis, INSA, Toulouse, 2009.

[3] F. Dell. *Les règles et les sons.* Hermann, 1973.

[4] A. Fernald. The perceptual and affective salience of mothers' speech to infants. In L. Feagans, C. Garvey, and R. Golinkoff, editors, *The origins and growth of communication*, pages 5–29. Ablex, 1984.

[5] R. Le Calvez. *Approche computationnelle de l'acquisition précoce des phonèmes.* PhD thesis, UPMC, Paris, 2007.

[6] R. Le Calvez, S. Peperkamp, and E. Dupoux. Bottom-up learning of phonemes: A computational study. In *Proceedings of the Second European Cognitive Science Conference*, pages 167–172, 2007.

[7] B. MacWhinney. *The CHILDES Project: Tools for Analyzing Talk.* Lawrence Elbraum Associates, 2000.

[8] J. Makhoul and R. Schwartz. State of the art in continuous speech recognition. *PNAS*, 92:9956–9963, 1995.

[9] A. Martin, S. Peperkamp, and E. Dupoux. Learning phonemes with a pseudo-lexicon. Submitted.

[10] S. Peperkamp, R. Le Calvez, J.-P. Nadal, and E. Dupoux. The acquisition of allophonic rules: statistical learning with linguistic constraints. *Cognition*, 101(3):B31–B41, 2006.

# Contents

Centre de recherche INRIA Paris – Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-0803