A Brazilian Portuguese Phonological-prosodic Algorithm Applied to Language Acquisition: A Case Study

Vera Vasilévski
Federal University of Santa Catarina (UFSC-CAPES)
Florianópolis, Brazil
sereiad@hotmail.com

Helena Ferro Blasi
Federal University of Santa Catarina (UFSC)
Florianópolis, Brazil
helena.blasi@ufsc.br

Márcio José Araújo
Federal University of Technology of Paraná (UTFPR)
Curitiba, Brazil
marciomjapr@gmail.com

Abstract

The paper presents a system for transcribing and annotating phonological information in Brazilian Portuguese, including syllabification. An application of this system for the assessment of language understanding and production is described, following a child longitudinally, comparing expected production with observed production.

1 Introduction

We present an application of a phonological-prosodic algorithm which converts Brazilian Portuguese graphemes to phonological symbols. For a better understanding, a brief report about the origin of the algorithm, altogether with some theoretical comments are presented, before the case study of the phonological processes found in the speech samples of a child. Sessions were recorded, until the complete acquisition of all Portuguese phonemes by the child, which occurred in the fifth session.

In 2008, we created the first version of a phonological-prosodic algorithm for Brazilian Portuguese. Actually, it is the functional algorithm of the grapheme to phoneme converter Nhenhém (Vasilévski, 2008, 2012a, 2012b). It has all written Portuguese spelling rules, and also the entire Portuguese prosodic system. When that algorithm was built, we kept in mind its usefulness to different fields deeply related to phonology, such as speech therapy, allowing the study of children phonological disorders, and language acquisition.

We focus here on its application to language acquisition, allowing the study of children phonological acquisition processes. Hence, our objective is to show the phonological-prosodic algorithm usefulness to language survey, from a practical point of view, by showing the process involved in the last stages of the acquisition of Portuguese phonology.

For a better understanding of the application and of the case study, the paper starts with some theory on phonological acquisition, then, some aspects of Brazilian Portuguese acquisition are presented.

2 Phonological Development

Studies of first language acquisition tend to support the view that the ability for language is innate in healthy human beings, and that its appearance can be predicted as part of the normal development of any child, given the right environment (Beaken, 1971).

The greatest expansion of the phonological system is observed from 1 year and six months old up to 4 years old, when there is an increase of the phonetic inventory of most complex syllable structures and, therefore, a period characterized by the occurrence of omissions, substitutions, as well as other phonological processes (Wertzner, 2004).

A phonological process is a mental operation that applies in speech to substitute, for a class of sounds or sound sequences presenting a specific common difficulty to the speech capacity of the individual, an alternative class identical but lacking the difficult property (Stampe, 1973).

It is worth remembering that, at the richest stage of normal language development (1 year and a half to 4 years old, as said), inappropriate sound gestures are expected phonological processes that relate to children's adaptations, until they automate the adult speech patterns. Thus, the phonological processes – that are natural and inborn – guide the facilitation of complex vocal

gestures and their planning, until children reach the adult performance.

Moreover, the early-acquired competence is filtered through an increasing number of phonological transformations to produce, finally, a mature performance. Although the mature phonemic system is acquired at an early stage, articulation may not be completely mature until after 7 years. Even though most children can be said to have mastered the complete set of potential phonemic oppositions of adult language by the age of 4 years – in other words, their phonological competence is established – yet, in adult terms, their performance falls short of their competence, in that they are unable to produce many of the gestures of mature articulation of the phonemes. Development after this stage takes place in the maturing of articulation, and in the acquisition of the complex transformations which operate on the basic acquired competence, to produce forms of speech similar to those heard from mature speakers (Beaken, 1971).

2.1 Some Aspects of Brazilian Portuguese Phonological Acquisition

Regarding phonemes, Brazilian Portuguese has 21 consonants (/p/, /b/, /f/, /v/, /m/, /n/, /t/, /d/, /s/, /z/, /l/, /ʎ/, /r/, /ʁ/, /ʃ/, /ʒ/, /ɲ/, /k/, /g/, and the archiphonemes |R|, |S|), and 14 vowels (/a/, /e/, /ɛ/, /i/, /o/, /ɔ/, /u/, /j/, /w/, /ã/, /ẽ/, /ĩ/, /õ/, /ũ/) (Scliar-Cabral, 2003a; Câmara Jr., 1986, 1977; Vasilévski, 2012a).

Lateral liquid phonemes /l/ and /ʎ/ and non-lateral liquid /r/ and /ʁ/ are the latest to be acquired in Brazilian Portuguese. Furthermore, such acquisition is marked by intense use of diversified phonological processes. What perhaps justifies this late acquisition, in Brazilian Portuguese as well as in other systems, is that this class is very complex, both in articulation and phonological aspects (Lamprecht, 2004).

Within this group of sounds, lateral phonemes are acquired before the non-lateral ones. The first lateral phoneme to be stabilized by children is /l/, which is subdued before the emergence of the first non-lateral liquid phoneme /ʁ/. This occurs with the phonemes /ʎ/ – graphically lh – and /r/, being the first acquired before the second (Hernandorena and Lamprecht, 1997). In Portuguese, the phoneme /r/ occurs: 1) forming a syllable with an oral or nasal vowel (simple onset); 2) in second position of inseparable consonant clusters, preceding oral or nasal vowel (complex onset); and 3) in syllable ending (coda, when it is the archiphoneme |R|). See Tab.1 for examples.

In most cases, the acquisition of the phoneme /r/ happens initially in the position of simple onset (by 4 years old) and then in the position of complex onset (by 5 years old), the acquisition of the phoneme /r/ in coda position (that is, |R|) occurs by 4 years old (Lamprecht, 2004) either.

Another linguistic phenomenon to be taken into account is diphthongization. It happens when one vowel breaks into two segments, where the first one matches the original vowel and the second (/j/ or /w/) is harmonic with the nature of the triggering vowel. In Brazilian Portuguese, one of the conditions when diphthongization occurs, and that matters here, is thus defined: a stressed vowel, followed by a devoiced alveolar fricative [s], in the ending syllable of a word, becomes diphthongized by the addition of a second segment, an [i] (Cagliari, 2002). Since diphthongization is a strengthening process, it occurs preferentially with strong vowels, and, in Romance languages, /a/ is the strongest vowel, and /i/, the weakest (Foley, 1977). The semivowels of stressed syllables can be either produced or not in speech, both options belonging to Portuguese language system (Vasilévski, 2012a). From the linguistic point of view, diphthongization is strongly related to the geographical dialectal variation (Leiria, 2000).

3 A Program for Helping Language Acquisition Research

By using Nhenhém phonological-prosodic algorithm, we built Nhenhém Fonoaud – NhFonoaud –, an application for assisting speech therapy, and so language acquisition. We began covering just one phonological process, called "unvoicedness": a substitution of a voiced sound for an unvoiced one (e.g. /b/ → /p/) (Blasi and Vasilévski, 2011). Soon, we realized that the phonological-prosodic algorithm could cover much more.

One of the motivations for creating such a system is that many Brazilian language acquisition researchers record their collected data using orthographic representation. As a result, those transcriptions are idiosyncratic and cannot be properly generalized, since they lack patterns. Data must be recorded in a phonologic-phonetic format, essential for these studies, since they address phonological processes.

According to researchers and speech therapists, there is no similar work in Brazilian Portuguese. Probably, there are similar initiatives for

4

other languages, and we expect to make comparisons soon.

3.1 The decoder Nhenhém Phonological-prosodic Algorithm

Nhenhém (/ɲẽ.ˈɲẽj/) is a computational program that decodes Brazilian's official writing system into phonological symbols and marks prosody (Vasilévski, 2008, 2012a). In 2010, we augmented its main algorithm, so the system became able of providing the phonological syllabic division and the spelling syllabic division, with at least 99% of accuracy (see Vasilévski, 2012a, 2012b for more details). Then we developed an automatic syllable parsing (Vasilévski, 2010).

In 2012, we made some adjustments regarding morphology, and solved the unpredictable situations brought, for example, by the prefix "trans-" that can be either decoded as /trãz/ or /trãs/, in consequence of resyllabification (see Vasilévski, 2012a). NhFonoaud benefits of all improvements obtained by the basic algorithm.

3.2 Nhenhém Fonoaud

The application of Nhenhém phonological-prosodic algorithm to language acquisition and speech therapy has been presented (Blasi and Vasilévski, 2011, Vasilévski, 2012a, 2012b), but this is the first time that a case study is discussed.

The first major challenge of working with phonemic transcription is the consistency of data. Different research questions require different levels of representation (Albert et al., 2013). In this regard, relying on an orthographic representation of speech, when dealing with language acquisition, does not make sense.

The program supports the analysis of processes that occur in the child's phonological system, through the automatic phonological transcription simultaneously to samples of the child speech recording. Thus, data relies on a phonemic representation of speech, automatically done by the algorithm, through Nhenhém Fonoaud.

NhFonoaud is designed for dealing with phonological tests, using words wittingly grouped to analyze specific aspects of speech and phenomena involved in its development. One of the tools of the program was the tests battery called Reception and Production of Spoken Language Assessment (Scliar-Cabral, 2003b). These tests were elaborated for assessing overt symptoms of spoken language reception and production problems. The first step is assessing phonetic features perception, namely, the ability of distinguishing

minimal pairs, what means distinguishing Brazilian Portuguese words.

The battery is composed by 81 pictures that represent specific words in Portuguese. The pictures are grouped into cards of six elements each. There are 15 cards, and some pictures appear in more than one. Each card is assembled to address the perception and production of one specific phonetic feature: 1) /v/-/f/, /p/-/b/, /t/-/d/; 2) /k/-/g/, /ʃ/-/ʒ/, /s/-/z/; 3) /m/-/n/, /t/-/d/, /s/-/f/; 4) /b/-/g/, /f/-/r/, /k/-/p/; 5) /t/-/k/, |R|-/s/, /l/-/ʎ/; 6) /t/-/s/, /k/-|R|, /p/-/f/; 7) /d/-/t/, /t/-/r/, /d/-/s/; 8) /m/-/b/, /n/-/r/, /z/-/n/; 9) /ɲ/-/ʎ/, /d/-/n/, /n/-/l/; 10) /r/-/l/ (in three different contexts); 11) /ɛ/-/ɔ/, /i/-/u/, /e/-/o/; 12) /i/-/e/, /ĩ/-/ẽ/, /u/-/o/; 13) /e(j)/-/ɛ/, /o/-/ɔ/, /ɔ/-/o(w)/; 14) /ɔ/-/a/, /m/-/b/ (in two different contexts)/; 15) /ẽ/-/e/, /o/-/õ/, /a/-/ã/.

In the reception battery, the speech therapist, behind the child, says the word and the child must point to one of the six pictures in each card. In the production battery, the speech therapist points to one of the six pictures in each card and the child must label it.

While the child labels the picture, the researcher can edit the canonical transcription provided by the program to match the child's production. For example, writing the lateral phoneme, when the child produces it, instead of the vibrating one.

In principle, four situations may happen during the assessment (Fig.1): the child does not recognize the picture (NR); the child gives the
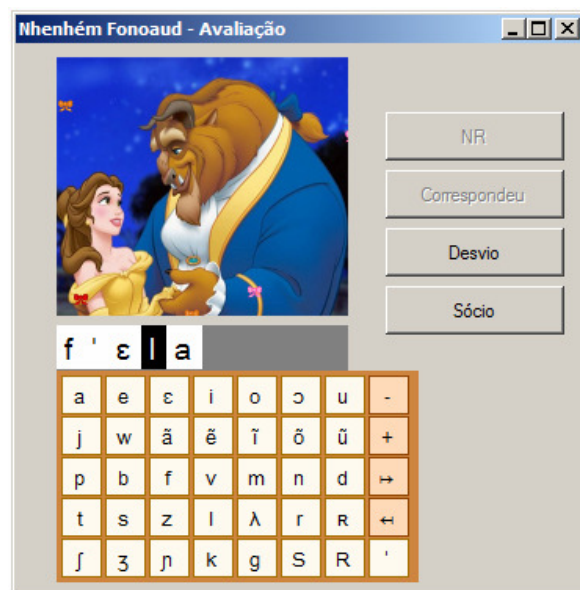

Figure 1. A register screen of Nhenhém Fonoaud
Picture: Beauty and the Beast. Disney©.

expected response (Correspondeu); the child gives an unexpected response (deviation – Desvio); the child translates the word into his/her sociolinguistic variety (not deviation – Sócio).

NhFonoaud stores the records and compares them with the transcription expected, for generating reports. Hence, it is possible to build a corpus, to retrieve it, grouping it according to date, situation, child's age, type of card (test); then it is possible the conversion into numbers, using different formats, comparing the phonological transcription and the correspondent audio, and the recorded sessions. Therefore, it facilitates child's progress monitoring.

In spite of working with words, NhFonoaud can be adjusted to work with bigger texts, formed by many sentences. For the purpose of assessing the child phonemic system, using minimal pairs and small sentences is enough.

3.3 Testing Nhenhém Fonoaud

The data analysis that we now present refers to a child in a clear process of language acquisition. It is based on oral emissions of a girl that we will call Inês. The 15 cards (Scliar-Cabral, 2003b) were applied, covering all the Brazilian Portuguese phonemes. Five sessions were recorded, starting when Inês was 2 years, 11 months, and 8 days; until she was 3 years, 8 months and 29 days.

Inês was born in Curitiba, Brazil, of Brazilian parents. She was not considered to have significant hearing loss. The child had developed some computer skills. Data was collected by her parents, by showing her the cards at the computer, during a daily conversation. Inês had already contact with the pictures, and had learned some names that were not part of her daily life. The sessions were recorded by using the audio resources of the same computer, and the records were clear enough to be used in this study.

3.3.1 Testing results

The results reveal the phonological processes used by Inês. Four were observed in her emissions: two of substitution, one of deletion, and one of adding. From the reports generated by Nhenhém Fonoaud, we created Tab 1.

So, at the age of 2y11m8d, three phonological processes relate to a single phoneme of her mother tongue, the non-lateral, vibrating /r/, and another one to diphthongization. Thus, Inês is only unable to produce the most complex Portuguese phoneme, in anyone of the three cases in which it occurs. The first session reveals that the child is able to produce all the vowels (14) of her mother tongue and 20 consonantal phonemes among the 21 of Brazilian Portuguese. No changes are observed for about 4 months, but this is expected, since progress in language is

Spelling	Meaning	Expected sounds	Emitted sounds				
			2y11m8d	3y2m23d	3y6m25d	3y7m24d	3y8m29d
porta	door	/pˈɔRta/	[pˈɔjta]	[pˈɔjta]	[pˈɔjta]	[pˈɔrta]	[pˈɔrta]
torta	pie	/tˈɔRta/	[tˈɔjta]	[tˈɔjta]	[tˈɔjta]	[tˈɔjta]	[tˈɔrta]
porco	pig	/pˈoRku/	[pˈojku]	[pˈojku]	[pˈojku]	[pˈorku]	[pˈorku]
		Process:	A	A	A	A	-
barata	cockroach	/barˈata/	[balˈata]	[balˈata]	[balˈata]	[barˈata]	[barˈata]
pera	pear	/pˈera/	[pˈela]	[pˈela]	[pˈela]	[pˈera]	[pˈera]
mureta	a low wall	/murˈeta/	[mulˈeta]	[mulˈeta]	[murˈeta]	[murˈeta]	[murˈeta]
perada	pear jelly	/perˈada/	[pelˈada]	[pelˈada]	[perˈada]	[perˈada]	[perˈada]
vara	fish rod	/vˈara/	[vˈala]	[vˈala]	[vˈala]	[vˈara]	[vˈara]
feira	street market	/fˈera/	[fˈela]	[fˈela]	[fˈera]	[fˈera]	[fˈera]
fera	beast	/fˈɛra/	[fˈɛla]	[fˈɛla]	[fˈɛra]	[fˈɛra]	[fˈɛra]
		Process:	B	B	B	-	-
traça	bookworm	/trˈasa/	[tˈasa]	[tˈasa]	[tˈlasa]	[tˈrasa]	[tˈrasa]
trança	braid	/trˈãsa/	[tˈãsa]	[tˈãsa]	[tˈlãsa]	[tˈrãsa]	[tˈrãsa]
trens	trains	/trẽjS/	[tẽjs]	[tẽjs]	[tlẽjs]	[tlẽjs]	[trẽjs]
três	three	/treS/	[tejs]	[tejs]	[tlejs]	[trejs]	[trejs]
		Process:	C/D	C/D	B/D	B/D	D
Phonological processes							
A	Substitution of the non-lateral, vibrating sound ǀRǀ or /r/ for the glide sound /j/ (semivowelization)						
B	Substitution of the non-lateral, vibrating sound /r/ for the lateral liquid sound /l/						
C	Reduction of the consonant cluster plosive+non-lateral /tr/ to the single sound /t/						
D	Diphthongization through insertion of a vowel in the last syllable of words ending with vowel+ǀSǀ.						

Table 1. Phonological processes used by Inês.

never regular; it may proceed at a fast rate for some periods while at others very little seems to be happening (Beaken, 1971). Then, the sound /r/ is emerging, only in simple onset, and she produces the cluster, but says /tl/ instead of /tr/ (3y6m25d). One month later (3y7m24d), she starts producing /r/ in coda position and the cluster /tr/, with some difficult yet. The sound /r/ in simple onset is naturally produced. One more month (3y8m29d), and she is able to naturally produce /r/ in all the contexts it happens in Portuguese, and keeps diphthongization.

Regarding diphthongization, it happens when the child inserts the semivowel /j/ between a vowel and the coda |S|, creating a diphthong. This circumstance advises that the child is adjusting her speech according to adult speech, since the region where Inês lives is where this phonological phenomenon occurs most, considering the South of Brazil (Leiria, 2000). It is a trait of the child's sociolinguistic variety, dependent upon geographic factor, and so she keeps saying it.

Hence, this research found that Inês completed the acquisition of the phonemes of her native language at 3 years and 8 months approximately, in normal development.

4   Conclusion and Outlooks

We briefly presented a system for dealing with phonological information in Brazilian Portuguese, and a case study from it, that is, the longitudinal speech recording of a child – the girl called Inês. Data allowed to know the last processes involved in the acquisition of the phonemes of her mother tongue.

From the preliminary results obtained, it is possible to conclude that Nhenhém Fonoaud can be helpful to language acquisition research. Nevertheless, the usefulness of the phonological prosodic algorithm has to be proven, by testing it in different situations, such as deviant language acquisition, speech therapy, and also other researches. This will be our next step.

Reference

Aviad Albert, Brian MacWhinney, Bracha Nir, Shuly Wintner. 2013. The Hebrew CHILDES Corpus Transcription and Morphological Analysis. *Language resources and evaluation*. Springer, Netherlands.

Beauty and the Beast. 2013. Disney© downloads. Available from http://disney.go.com/disneyvideos/animatedfilms/beauty/downloads.html

Carmen Lúcia M. Hernandorena and Regina R. Lamprecht. 1997. A aquisição das consoantes líquidas do português. *Letras de Hoje*, Porto Alegre, 32(4):7-22.

David Stampe. *A dissertation on natural phonology*. 1973. PhD Thesis, University of Chicago.

Haydée F. Wertzner. 2004. Fonologia: Desenvolvimento e Alterações. In: LP Ferreira, DM Befi-Lopes and SCO Limongi. *Tratado de Fonoaudiologia*. Roca, São Paulo, 772-786.

James Foley. 1977. *Foundations of theoretical phonology*. Cambridge University Press, Cambridge.

Helena F. Blasi and Vera Vasilévski. 2011. Programa piloto para transcrição fonética automática na clínica fonoaudiológica. *Documentos para el XVI Congresso Internacional de la ALFAL*, Universidad de Alcalá, Alcalá de Henares/Madrid.

Joaquim M. Câmara Jr. 1986. *Estrutura da língua portuguesa*. 16.ed. Vozes, Petrópolis, RJ.

Joaquim M. Câmara Jr. 1977. *Para o estudo da fonêmica portuguesa*. 2.ed. Padrão, Rio de Janeiro.

Leonor Scliar-Cabral. 2003a. *Princípios do sistema alfabético do português do Brasil*. Contexto, São Paulo.

Leonor Scliar-Cabral. 2003b. *Guia prático de alfabetização*. Contexto, São Paulo.

Lúcia Lovato Leiria. 2000. A ditongação variável em sílabas tônicas finais travadas por /S/. *Organon*, 14(28/29):133-141.

Luiz Carlos Cagliari. 2002. *Análise fonológica: introdução à teoria e à prática*. Mercado das Letras, Campinas, São Paulo, Brazil.

Michael Alan Beaken. 1971. *A study of phonological development in a primary school population of East London*. PhD Thesis. London University. http://discovery.ucl.ac.uk/1317623/1/261970.pdf

Nhenhém®. 2008-2011. Grapheme do phoneme converter for Brazilian Portuguese. INPI 1265-1.

Regina R. Lamprecht. 2004. *Aquisição fonológica do Português: Perfil de desenvolvimento e subsídios para terapia*. Artmed, Porto Alegre.

Vera Vasilévski. 2012a. *Descodificación automática de la lengua escrita de Brasil basada en reglas fonológicas*. Saarbrücken, Editorial Académica Española, Germany.

Vera Vasilévski. 2012b. Phonologic Patterns of Brazilian Portuguese: a grapheme to phoneme converter based study. *Proceedings of the EACL, Workshop on Computational Models of Language Acquisition and Loss*. University of Avignon, France.

Vera Vasilévski. 2011. O hífen na separação silábica automática. *Revista do Simpósio de Estudos Lingüísticos e Literários – SELL*, 1(3):657-676.

Vasilévski, Vera. 2008. *Construção de um programa computacional para suporte à pesquisa em fonologia do português do Brasil*. PhD Thesis. Universidade Federal de Santa Catarina, Florianópolis, Brasil.