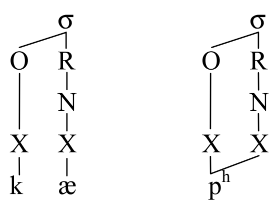
Onset-Nucleus Sharing in Interlanguage: Evidence from Brazilian Portuguese English

Walcir Cardoso and Denis Liakin
Concordia University

1. Introduction

The development of word-final English stops (codas) by Brazilian Portuguese (BP) speakers is characterized by three distinct stages: (1) i-Epenthesis, an early phonological phenomenon in which the coda syllabifies as the onset of the epenthetic vowel [i] (e.g. dog → do.g[i]; Cardoso 2007), and that results from the direct interference of BP phonology; (2) Stop Devoicing (e.g. dog → do[k]; Major 1986), a developmental process that is also found in the first language (L1) acquisition of English and across languages; and (3) native-like Coda Production (e.g. dog → do[g]), the final stage in the acquisition of this syllable constituent. Based on phonetic evidence and along the lines of Goad (2002), Goad and Brannen (2003), and Goad and Kang (2003), we provide evidence for another developmental stage in the acquisition of English codas: Onset-Nucleus sharing (ONS henceforth). Under this approach to the syllabification of word-final consonants, it is assumed that the to-be-acquired constituent syllabifies as an onset and, in order to be licensed, some of its features spread into the following empty nucleus in order to optimize the syllable shape of the intermediate grammars that characterize the interlanguage of these learners. This analysis reflects a minimal change from an L1 grammar that permits only CV syllables (i.e. BP), while still being faithful to the English input CVC melodic string, as illustrated in (1) below.

(1) Onset-Nucleus Sharing in interlanguage:



More generally, this investigation focuses on the variable syllabification of word-final voiceless stops (i.e. /p/, /t/ and /k/, to which we will occasionally refer as codas) in BP-based learner speech, in the context of data collected in a cross-sectional study involving 60 BP speakers learning English in a classroom environment. The results of the statistical analysis (conducted using VARBRUL 2, Pintzuk 1988) indicate that ONS is more likely to occur: (1) in intermediate stages of acquisition, with an increase and then decrease in frequency as the level of proficiency increases, as predicted by Major's

* We would like to thank the 60 students who participated in this study as well as the teachers and/or administrators Graça and Gilberto (*Beverly Curso de Idiomas*), and Renata and Paulo (*Yázigi Internexus*) for allowing us to recruit students in their classrooms and use their facilities to conduct the interviews. Valeu! For their insightful comments, criticisms and suggestions, we also thank the audience of the Generative Approaches to Language Acquisition North America Conference (GALANA 2, 2006), particularly Heather Goad (whose studies on Onset-Nucleus Sharing inspired us to conduct this research), Katherine Demuth and Joe Pater. Finally, we would like to acknowledge funding from the Fonds *québécois de la recherche sur la société et la culture* (FQRSC NC-96880) and from the *Canadian Social Sciences and Humanities Research Council* (SSHRC 410-2006-1920) to Walcir Cardoso. We alone are responsible for any remaining shortcomings.

(2001) Ontogeny Phylogeny Model; (2) in more formal stylistic environments, supporting the view that formal speech interactions require more faithful forms (e.g. Labov 1966, van Oostendorp 1997); and (3) when the coda's place of articulation is dorsal [k] or labial [p] (rather than coronal [t]), confirming the relative markedness of these classes of consonants when syllabified in coda position (e.g. Prince and Smolensky 1993).

For the analysis of the variable patterns that characterize learner speech, this paper adopts a stochastic version of the framework of Optimality Theory (Prince and Smolensky 1993): the Gradual Learning Algorithm, proposed by Boersma and Hayes (2001). The main advantage of this approach is that it allows the encoding of variability and its frequency effects (i.e. the quantitative values observed in the corpus) within a language by means of a single grammar. This investigation also promotes a multidisciplinary and integrative approach that combines theoretical and methodological tools from three linguistic disciplines (i.e. sociolinguistics, second language acquisition and formal phonology), in an attempt to develop a "socially realistic linguistics" (Wilson and Henry 1998).

The main objectives of this study are: (1) Following a sociolinguistic methodology for data collection, to investigate the variable development of English (L2) word-final voiceless stops /p t k/ by Brazilian Portuguese speakers learning English in a classroom environment; (2) Along the lines of Goad (2002) and Goad and Kang (2003), to provide evidence for an intermediate stage in the acquisition of these codas by Brazilian ESL learners: Onset-Nucleus Sharing; and (3) Using theoretical tools provided by current research in phonology, to provide support for the proposal that variation can be accounted for in Optimality Theory (OT; Prince and Smolensky 1993) via a stochastic version of the framework and its Gradual Learning Algorithm (Boersma 1998, 2000, Boersma and Hayes 2001). More generally, this study attempts to demonstrate that the variation observed in the development of word-final codas is inherently systematic (i.e. it is similar to what is observed in the development of first and second languages, and in the grammar of "full-fledged" adult languages), and is triggered by a combination of linguistic (e.g. place of articulation of the potential coda constituent) and extralinguistic factors (e.g. proficiency and style).

The paper is organized as follows: in section 2, we provide an introduction to codas in Brazilian Portuguese and in BP-based interlanguage, and describe the relevant initial and subsequent states in the phonology of these learners. Section 3 addresses the data collection procedures, introduces the statistical package used in the analysis (VARBRUL 2), as well as presents and discusses the quantitative results obtained. This is followed by section 4, in which we provide a stochastic analysis for the variable results within Boersma and Hayes' (2001) approach to variation via the Gradual Learning Algorithm. Finally, in section 4, we present our concluding remarks for the study.

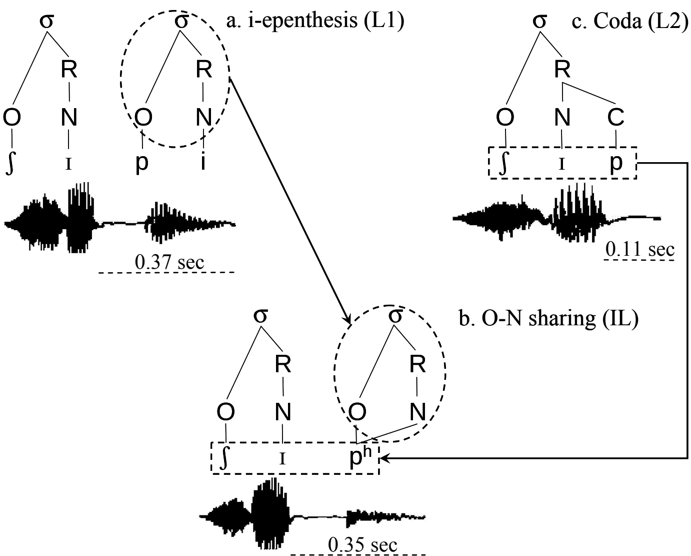
2. Word-final stops in BP English (BPE): The initial and subsequent states

In Brazilian Portuguese, only four consonants can potentially appear in coda position: /l N r s/. However, /l/ and /N/ (an underspecified nasal consonant) never surface as codas in most varieties of BP, triggering either diphthongization in the case of /l/ (e.g. /brazil/ → [bra.ziw] 'Brazil' – cf. [bra.zi.lej.ru] 'Brazilian', where "." indicates syllable boundaries), or the nasalization of the preceding vowel in the case of /N/ (e.g. /boN/ → [bõ] 'good' – cf. [bõ.na.ʃãw] 'good person'). /r/ and /s/, on the other hand, are the only consonants that syllabify as codas in the language. Word-finally, however, these two phonemes may variably undergo deletion in informal styles (e.g. /falar/ → [fa.lar] ~ [fa.la_] 'to speak' and /kazas/ → [ka.zas] ~ [ka.za_] 'houses' respectively; see Cardoso (1999) for a variable investigation of /r/-deletion in the same variety of Portuguese under investigation). If the syllable-final consonant is an obstruent stop (i.e. [p, t, k b, d, g]), BP opts for the syllabification of the illicit segment as the onset of the default epenthetic vowel in the language: [i] – i-epenthesis henceforth. This applies to both native (e.g. /sub+ʃɛfe/ → [su.bi.ʃɛ.fi] 'sub-chief' – cf. [su.baw.ter.nu] 'subordinate') and loan words (/pɑp/ → [pɔ.pi] 'pop').

Assuming the standard view in the generative second language literature that the initial state for L2 acquisition coincides with the final state of the L1 grammar (e.g. Broselow and Finer 1991, Archibald 1993, Pater 1997, Davidson et al 2004), we predict that the initial grammar of the English learner will directly reflect the phonology of BP, and will therefore constitute a transitory system in which i-epenthesis will categorically operate in order to syllabify illicit word-final voiceless stops.

In subsequent stages during the acquisition of English codas, however, voiceless stops syllabify variably via one of the following variants: (1) i-epenthesis, which, as described above, directly mirrors the phonology of the L1 (2a); (2) onset-nucleus sharing, a developmental stage in which the potential coda syllabifies as both onset and nucleus ((2b); see forthcoming discussion below); and (3) coda syllabification, the target-like English form (2c).

(2) Intermediate states – Variable outcomes: i-epenthesis, ONS, and Coda:



Let us now concentrate on the syllabification of word-final stops via ONS, a phenomenon described and formalized by Goad (2002) and adopted later in Goad and Brannen (2003), and Goad and Kang (2003). As illustrated in (1) and (2b), under this approach to the syllabification of word-final consonants, it is assumed that the illicit constituent syllabifies as an onset. Because segments must be licensed by a head constituent within the syllable domain (e.g. a nucleus), some of the onset's features (e.g. the laryngeal feature *spread glottis*) spread into the following empty nucleus via a "homorganic burst of noise" (Goad and Brennan 2003:13, as illustrated in (2b)). In the context of the monosyllabic (and consequently stressed) words considered in the study (see forthcoming section 3 for a description of the data collection procedures), the phenomenon is characterized phonetically by both aspiration and lengthening of the voiceless stop segment. ONS could then be interpreted as a type of vowel deletion followed by a remedial compensatory lengthening (or feature *epenthesis*, Bernhardt and Stemberger 2003:378) in which the developing grammar reflects the CV structure of the L1, while attempting to emulate the input CVC melodic structure of the L2 English, as shown in (2b).

The aspiration and lengthening associated with ONS were also attested via an ad hoc acoustic analysis conducted in *Praat* (Boersma and Weenink 2000) of the word "ship", pronounced by 20 of the participants from different levels of proficiency. We computed the duration and analyzed the waveforms of the voiceless stop from its outset to its completion via one of the variants considered in this study. The results are illustrated in (2), where we show that the average length of voiceless stop [p] plus epenthetic [i] was .37 seconds, closely corresponding to the average duration of the form in which ONS has occurred, i.e. .35 seconds. Codas, on the other hand, are the shortest of the three variants, with an average of .11 seconds in slow, careful speech. In the waveforms in (2), the differences in the shape of the sound frequency for each variant's content, especially those that distinguish vowel epenthesis from aspiration in (2a) and (2b) respectively, confirm that ONS, i-epenthesis and coda production are different entities in the grammar of these learners.

The data and discussion provided above allow us to formulate some general questions that we will attempt to answer in this study. Firstly, within a variationist approach to language study, how systematic is the variation observed in the development of English word-final voiceless stops in interlanguage? More specifically, how do the phenomena involved in coda acquisition develop across proficiencies and how do they compare with those found in other languages and/or forms of language? Secondly, at what stage of proficiency is ONS acquired and when does the phenomenon develop into full codas? These questions will be addressed in forthcoming section 3. Finally, how can one explain the variation observed in light of current developments in phonological theory, namely within the framework of Optimality Theory? This will be the focus of section 4.

To summarize, during the acquisition of L2 English codas, the language learner progresses hypothetically from an initial state in which the grammar is merely a reflection of the L1 (i-epenthesis) to a full-fledged grammar that is strikingly analogous to that of the target language (coda production). This process, however, is mediated by a developmental phenomenon in the form of ONS, which combines features from both the L1 and the L2 in order to optimize the syllable shape of the emergent grammar. In addition, the acquisition of codas is a variable process, one that that reflects not only the phonology of the two languages involved, but also universal linguistic principles. In the following sections, we will show that variability in coda acquisition results from the interaction of both linguistic and extralinguistic factors as well as the influence of the two languages involved in the process.

3. Data Collection Procedures, VARBRUL Analysis and Results

The investigation adopts a sociolinguistic (variationist) methodology for data collection in order to obtain a reliable corpus of non-categorical data such as those encountered in developing languages. As is customary in variationist studies, our study includes a two-level distinction in a formality hierarchy. It is cross-sectional and involves the participation of 60 English learners, stratified into three proficiency groups. The methodology for data elicitation was inspired by Labov's (1966) seminal investigation of the pronunciation of (r) in coda position in three New York City department stores. Briefly, Labov attempted to elicit the phrase "fou<u>r</u>th floo<u>r</u>" from sales representatives by asking for the location of items found on the *fourth floor*. Whenever a response was provided, Labov would search for a more careful repetition of "fourth floor" by pretending not to hear the salesperson's response. The study examined the effects of style (casual versus careful speech), social class (via the three department stores, each representing a specific socioeconomic clientele), and the position of the variable within the word (word-internally preceding a consonant versus word-finally preceding a pause). The current study followed a similar procedure for data elicitation, except that the phrase "fourth floor" was replaced by a picture naming task consisting of 10 pictures that represented words ending in voiceless stops (e.g. shi[p], ca[t] and boo[k]). In the interview (to which we will occasionally refer as a picture-naming task), which was part of another larger and unrelated study on the development of codas (reported in Cardoso 2007), participants were sporadically asked the question "What's this?" Mimicking Labov's tactic, the interviewer pretended not to hear the participant's response by using body language or asking questions such as "Can you say that again?" For convenience and in the spirit of the variationist methodology adopted, the first and more spontaneous utterance was coded as *informal* while the second, more careful response was coded as *formal*. Because the developing grammars of L2 learners is characterized by monostylism (Cardoso 2007), this picture-naming interview was an opportune way to elicit data from two distinct levels of formality.

Motivated by the notion that certain prominent positions (e.g. stressed syllables) are more likely to maintain contrasts and thus less likely to undergo phonetic changes (e.g. Trubetzkoy 1939, Beckman 1998), the words selected for the picture-naming interview were all monosyllabic (and consequently stressed) to ensure the codas under investigation occurred exclusively in positions of prominence. In addition, the selection of the words for inclusion in the study followed two other criteria: (1) the words had to be highly frequent in the language in order to minimize word-frequency effects (Bybee 2001); one could argue, for instance, that learners are more likely to produce codas correctly in more frequently-occurring words; and (2) the words had to conform to a CVC syllable sequence (where V represents a lax vowel; e.g. [æ] [ɪ] [ɛ] [ʊ]). This was motivated by the hypothesis that final consonants in more complex structures such as CVV<u>C</u> and CVC<u>C</u> could in fact be syllabified as onsets of empty-headed syllables given that English rhymes abide by constituent binarity (i.e. bimoraicity); see Goad and Brannen (2003) for a comprehensive discussion of this proposal.

The 60 participants were post-pubescent male and female native BP speakers, with an age average of 25. The participants were divided into 3 groups of 20 participants each based on their proficiency in L2 English: (1) Beginner – students with an average of 20 hours of classroom experience learning the language; (2) Intermediate – participants with approximately 150 hours of English instruction; and (3) Advanced – students with more than 300 hours of exposure to the language. The data were collected by the author at a private language school in the city of Belém, Brazil, where English is rarely used outside of the learning environment. The interviews were audio recorded via a Marantz CDR300 CD/RW Recorder and an Audio-Technica AT831b lavaliere microphone, and later transcribed with the help of two research assistants and coded for VARBRUL 2 analysis. The 1,208 tokens collected were stratified among the following variables listed in Table 1.

(3) Table 1: Factor Groups for VARBRUL 2 Analysis

Factor Groups	**1**	**2**	**3**
Dependent Variables	Coda	Epenthesis	ONS
Place of articulation	Labial [p]	Coronal [t]	Dorsal [k]
Level of proficiency	Beginners (±20h)	Intermediate (±150h)	Advanced (±300h)
Style	Informal (first answer)	Formal (careful answer)	Interview
Participants	# 1	# 2	# 3, etc.

For the statistical analysis of the BPE corpus, we adopted the VARBRUL 2 program for DOS (Pintzuk 1988). This program has been extensively used in variationist studies in linguistics because, along with GoldVarb (Rand and Sankoff 1990, for Macintosh computers) and GoldVarb X (Sankoff et al 2005, for both PC and Macintosh computers), it is the only one explicitly designed to handle the types of data derived from studies of language variation; other probabilistic tests such as ANOVA were designed to handle data collected from controlled experimentation that result in rather balanced data. In Young and Bayley's (1996:258) terms, VARBRUL is able to manage "the distributional imbalances of linguistic features in sociolinguistic data." One critical difference between the different incarnations of these applications for running variable analyses is that VARBRUL is the only one able to handle multinomial analyses that contain more than two dependent variables or application values (Tagliamonte 2006:217). Because the development of codas in BPE involves three phenomena (and consequently three dependent variables), VARBRUL 2 was selected to analyze our data.

The results of a VARBRUL study should be interpreted as holding over the whole of the data corpus that is being investigated and, to the extent that this is a representative sample, to all similar speakers and linguistic and extralinguistic contexts. The output of a typical VARBRUL analysis contains the following information: (1) The raw number (N) and the percentage of rule application involving each factor. These results, however, do not provide enough information since they do not express the influence of each factor independently of the others. (2) The factor weight measures the influence that each factor has in the process under investigation, based on the corpus analyzed. It provides the most accurate view of the likelihood of variant occurrence. It consists of a list of values associated with each factor independently of others in the same factor group. The value indicates the degree to which a factor promotes the occurrence of each variant for the process being investigated. Because the development of coda production in BPE consists of three variants, the factor weight of .33 was established as the watershed between the weights that enhance the likelihood of a certain variant's occurrence (above .33) and those that inhibit its appearance (below .33) (see also Major 1996 and Preston 1996). Finally, (3) the input probability (also more descriptively called Overall Tendency) is the likelihood that each variant has of occurring in general, regardless of the specific contribution of particular factors. In other words, it represents the general propensity of the process to apply on its own, without the interference of the other factors included in the investigation.

The VARBRUL results in Table 2 (significant factors are shaded) indicate that the independent variables adopted in the study (i.e. the external variables *proficiency* and *style* and the internal variable *place of articulation*) have significant conditioning effects on the variable development of English codas by BP speakers. The factor group *participants* was excluded from the final VARBRUL analyses to prevent interference with the significant *proficiency* factor: every participant inherently belongs to a proficiency group. The results indicate that while i-epenthesis is favored in the speech of beginners, in less formal stylistic environments and in the context of labial or dorsal segments, potential codas are more likely to surface as such in more advanced proficiency levels, when more careful attention is paid to speech (Formal), and when the consonant in question is a coronal. ONS, on the other hand, is favored in intermediate stages of L2 acquisition (Intermediate), in more formal stylistic environments, and when the potential coda is a more marked labial or dorsal segment. As is typical in developmental phenomena, ONS is not an idiosyncrasy of developing grammars such as that of BPE: the lengthening and release properties (i.e. aspiration) that characterize ONS have also been observed in adult languages (e.g. in Yapese – Jensen 1977), in first language acquisition (e.g. in English – Leopold 1939; in Quebec French – Rose 2000), and in the development of second languages (e.g. in Korean English – Goad and Kang 2003; Mandarin English – Heyer 1986; Mandarin French – Steele 2002).

(4) Table 2: VARBRUL probabilistic results

		Epenthesis	ONS	Coda
Proficiency	Beginners (L1)	.81	.10	.09
	Intermediate (L2)	.21	.51	.28
	Advanced (L3)	.10	.31	.59
Style	Informal	.53	.26	.28
	Formal	.18	.37	.44
Place of Articulation	Labial	.41	.37	.22
	Coronal	.29	.30	.41
	Dorsal	.37	.39	.24
Input Probability		.49	.29	.22

The results clearly demonstrate that while the likelihood of the transfer process of i-epenthesis decreases as a function of increased proficiency, the reverse is observed for coda production. In contrast, the phenomenon of ONS increases and decreases over the course of L2 English acquisition. This is exactly what predicts Major's (2001) Ontogeny Phylogeny Model (OPM), which posits that learners progress from an initial stage with a prevalence of L1 features, to a final stage with a prevalence of L2 features, via medial stages, during which L1 features (i.e. i-epenthesis) gradually decline, L2 features (i.e. codas) gradually rise, and development phenomena (i.e. ONS) steadily rise, peak and then fall again over the course of L2 acquisition, as represented graphically in (5).

(5) The OPM predictions across three proficiency levels

a. i-epenthesis by proficiency b. ONS by proficiency c. Coda by proficiency

With regards to level of formality, the VARBRUL results show the robust effect of this factor group, in which the more faithful coda and ONS variants are more likely to occur in more formal environments. In contrast, the least faithful of the three variants, i.e. i-epenthesis, is favored only when less attention is paid to speech (Informal). This pattern is compatible with a variety of studies in L2 acquisition (e.g. Gatbonton 1978, Major 1986; cf. Beebe 1980), and with the standard view that more formal styles require more "prestigious" and faithful forms (e.g. Labov 1966, van Oostendorp 1997, Cardoso 2001, 2003, and most of the sociolinguistic literature).

Finally, the results related to the effect of the place of articulation of the word-final stops attest that variant selection is also determined by the markedness of the potential coda. The probabilities assigned by the variable analysis indicate that the least marked coronal [t] is more likely to surface as a coda in BPE, more evidently in intermediate and advanced proficiency levels. The labial [p] and dorsal

[k], on the other hand, are less likely to syllabify as codas. Because consonant deletion is not an option in the BPE phonology, these consonants undergo one of the two strategies that the learner utilized in interlanguage: i-epenthesis and ONS. In Optimality Theory, this observation can be captured by the markedness sub-hierarchy of place of articulation (Prince and Smolensky 1993): $\text{NoCoda}_{[\text{LABIAL}]}$, $\text{NoCoda}_{[\text{DORSAL}]} >> \text{NoCoda}_{[\text{CORONAL}]}$. Being less marked than labials and dorsals, coronals are more likely to syllabify as codas in BP-based Interlanguage.

The discussions that follow are based on the cross-tabulations between level of proficiency and style and therefore the results are presented in percentages. We assume here the following standard views on the nature of grammar: (1) that proficiency levels correspond to different interlanguages and, by definition, to different grammars characterized by a series of transitional and permeable systems (Selinker 1972, Adjémian 1976, Adamson 1988, Preston 1996); and (2) that formality levels constitute separate grammars (Selkirk 1972, Chomsky 1988, van Oostendorp 1997, Cardoso 2001, 2003). Note that there are certain similarities between the assumptions made here and those inherent to the Multiple Grammar approach proposed by Kroch (1994) for whom variation is the result of distinct grammars in operation and in competition with each other. The view adopted here, however, differs from that of Kroch (1994) in a crucial way: in his Multiple Grammar approach, competing grammars are invoked to explain variable phenomena that are intrinsically *internal* to the grammar. In this study, however, we limit the assignment of grammar status to *external* factors such as style and proficiency, as is customary in the generative literature. By these assumptions, we may then conclude that the data investigated comprise the six distinct variable grammars illustrated in Table 3.

(6) Table 3: The development of word-final codas (%)

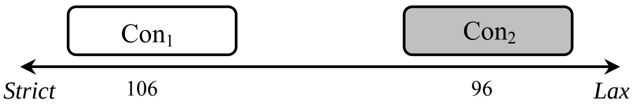
Grammars by Proficiency/Style	i-epenthesis	ONS	Coda
Level 1, Informal	89	7	5
Level 1, Formal	84	9	7
Level 2, Informal	51	38	10
Level 2, Formal	18	54	28
Level 3, Informal	39	25	36
Level 3, Formal	5	38	58

To conclude, we have shown in this section that the variable acquisition of codas in BPE is mediated by the transfer process of i-epenthesis and, in later stages, by the developmental phenomenon of ONS. It was also shown that the occurrence of each of the variants observed is strictly conditioned by a combination of extralinguistic (i.e. *proficiency* and *style*) and linguistic factors (i.e. *place of articulation of the potential coda*). In the following section, we will demonstrate how the quantitative, variable results obtained in the investigation can be modeled within a stochastic version of the framework of Optimality Theory.

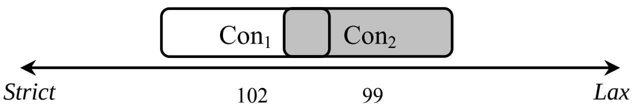
4. A Stochastic OT Analysis

For the analysis of the variable data discussed in the previous section, we adopt Boersma's (1998, 2000) and Boersma and Hayes' (2001) methodology for investigating and representing variability in the framework of Optimality Theory (Prince and Smolensky 1993): Stochastic OT (StOT), an approach that includes an associated learning algorithm: the Gradual Learning Algorithm (GLA). Within StOT, variation and gradient well-formedness are accounted for by a probabilistically determined reranking of constraints at certain intervals during evaluation time (i.e. during the process of speaking). Briefly, StOT postulates a continuous scale of constraint strictness in which constraints (e.g. Con_1 and Con_2) are annotated with arbitrary numerical strictness values established by a GLA (e.g. Boersma and Weenink's 2000 Praat program, Hayes et al's 2003 OTSoft software). The probability of reranking (i.e. variation) is determined by the distance between Con_1 and Con_2 on the strictness scale and by the amount of evaluation noise (i.e. standard deviation) added to the strictness values. This way, constraints not only dominate other constraints (as is the case in standard OT), but they are also specific distances apart. The two figures in (7) and (8) below illustrate the distinction between a categorical grammar (in which crucially ranked constraints are distant, similar to Con_1 >> Con_2 in standard OT) and a variable one (in which crucially ranked constraints overlap).

(7) A categorical ranking



(8) A variable ranking



In the context of a variable ranking, as shown in (8), the grammar might select any point within the overlap of Con1 and Con2. Most often, the grammar will select the ranking Con_1 >> Con_2 because of the higher ranking of Con_1 over Con_2. However, it is also possible for the grammar to select a point within the leftmost (higher ranked) area of Con_2 and the rightmost (lower ranked) area of Con_1. In this case, Con_2 is ranked higher than Con_1 and a different candidate is selected.

For the analysis of variation, the main advantages of this approach are: (1) we are able to obtain a simpler grammar, with fewer constraints (cf. Reynolds 1994, Anttila 1997, Cardoso 2001, 2003), and still predict the same frequency data: variation is governed by the same linguistic constraints and principles that govern categorical phenomena (e.g. the crucial ranking of constraints in standard OT); (2) other approaches to variation (i.e. Reynolds' 1994 "floating constraints" and Anttila's 1997 "crucial non-ranking of constraints") predict that quantities of variation should be in small integer fractions (e.g. 2/3, 1/2, 1/3). The data in this study do not always show this type of variation (see also Pater and Werle 2001 for a similar observation).

Consider the relevant OT constraints that we adopt in this analysis:

(9) OT constraints

MAX-IO	Every element of the input has a correspondent in the output (i.e. No deletion) (McCarthy and Prince 1995)
DEP-IO	Every element of the output has a correspondent in the input (i.e. No epenthesis) (McCarthy and Prince 1995)
NoCoda(Obs)	Syllables do not have obstruent codas (e.g. Broselow et al 1998), which comprise the following place-specific constraints: - $NoCoda(Obs)_{[coronal]}$: No obstruent coronals in coda - $NoCoda(Obs)_{[labial]}$: No obstruent labials in coda - $NoCoda(Obs)_{[dorsal]}$: No obstruent dorsals in coda
Nuc/V	Nuclei contain vowels (e.g. Prince and Smolensky 1993, Oostendorp 1995, Tranel 1996)

For the stochastic analyses, a series of computer simulations was performed using the OTSoft 2.1 software package (Hayes et al, 2003). The GLA component of the package was fed with a set of inputs, relevant surface forms and their respective quantitative values (established by VARBRUL in the previous section), incorrect rival candidates, and constraint violations (just like in a standard OT analysis). In brief, the simulations proceeded as follows: as indicated in Table 3 in section 3, the data set investigated represents six distinct grammars. Each of these grammars was individually "learned" by OTSoft, which was supplied with the following information (all numbers are arbitrary) for the learning simulation to take place: (1) Number of times to go through forms (or total of learning trials): 1,000,000. This number indicates how many learning trials the GLA will perform. The higher the number, the more likely the observed and predicted probabilities will match. (2) The initial state: 100

for both markedness and faithfulness (the concrete value is inessential). By default, the initial state is set with the arbitrary number of 100 for markedness and faithfulness constraints, a value that ensures that the ranking values will always be positive. This value can be manipulated by the researcher depending on his/her views regarding language learning. For instance, s/he might decide that the learning process starts with a grammar in which markedness (e.g. 100) is ranked above faithfulness (e.g. 50) – a standard hypothesis for first language acquisition (e.g. Smolensky 1996, Davidson et al 2004, Hayes 2004, Prince and Tesar 2004). Because the acquisition of codas involve violations of both markedness (NoCoda(Obs), Nuc/V) and faithfulness (DEP-IO), we adopt an equal value of 100 for the two types of constraints. (3) Initial/final plasticity: 2/.002 respectively, which are the default values in OTSoft 2.1. They serve to adjust the GLA results by comparing the outcome of the learning algorithm with the results entered for each pair of input-output. In Hayes' (2004:21) own words, "[p]lasticity is the size of the change in the grammar that the GLA makes every time its own guess [does not] match the learning datum it encounters." Note that the algorithm will only make adjustments to the simulated grammar if it detects discrepancies between what is observed and what it predicts – it is *error-driven*. The plasticity values are usually set high in the beginning to accelerate learning, and low late in the learning process, for more refined results. Finally, (4) the number of times to test grammar: 2,000 cycles (default). This number indicates the number of times the GLA will repeat the process of stochastic evaluation and compare the results to the relative frequencies that were observed in the data (the GoldVarb results in our particular case). As will be shown later, the predictions established by the algorithm closely match the frequencies observed in the corpus analyzed. At the end of the simulations, the algorithm arrived at a final grammar that attempted to mimic the relative frequency of variants in the data, by assigning a ranking value for each of the constraints included in the analysis. The procedure was repeated for each one of the six grammars established in the investigation.

To illustrate how the GLA works in the context variable phenomena, consider the results obtained in Level 2 Formal, a grammar characterized by a higher incidence of ONS (54%) in comparison with i-epenthesis (18%) and coda production (28%); results compiled from Table 3. These frequencies obtained were then learned by the GLA (OTSoft), which located an empirically appropriate ranking value for each constraint in the grammar, as illustrated in (10); shaded constraints indicate an overlap in their distribution (similar to the hypothetical hierarchy depicted in (8)), whose outcome is variation.

(10) Level 2 Formal grammar: ranking values[1]

Constraint	Ranking value
MAX-IO	106.000
NoCoda(Obs)	98.688
Nuc/V	98.242
DEP-IO	97.070

[1] To ensure a higher likelihood of coda production for the set of word-final coronal consonants (recall from (4) above that, in contrast to labials and dorsals, coronals favor the pronunciation of codas), it is assumed, for convenience sake, that the following place-specific constraints and their relative ranking values are encoded into a single NoCoda(Obs). The values assigned by the GLA in this case are unimportant; what is crucial is the predictability of these values with respect to the distribution of coronals over labials and dorsals to syllabify as codas or via ONS in interlanguage. For instance, observe below that the GLA-generated frequencies predict that coronals will syllabify as codas approximately 50% of the time, while labials and dorsals are equally likely to trigger other interlanguage phenomena such as i-epenthesis and ONS. NoCoda(Obs) should then be understood as a family of constraints on place of articulation markedness for codas, according to the sub-hierarchy proposed by Prince and Smolensky (1993) and, more importantly, in line with the data and analyses presented in the study.

(i) Ranking values for place-specific versions of NoCoda(Obs) and GLA predictions

Constraint	Ranking value	GLA predictions
NoCoda(Obs)$_{[\text{LABIAL}]}$	98.416	24%
NoCoda(Obs)$_{[\text{DORSAL}]}$	98.372	25%
NoCoda(Obs)$_{[\text{CORONAL}]}$	97.212	51%

Because MAX-IO is ranked considerably higher than the other three constraints, Level 2 Formal grammar will always rank MAX-IO at the higher end of the hierarchy and coda deletion will never take place. The constraints NoCoda(Obs), Nuc/V and DEP-IO, however, overlap in their distribution resulting in a pattern in which they can provisionally outrank each other resulting in variable patterns that resemble the ones observed in the investigation. According to the ranking values assigned to the overlapping constraints in (10), DEP-IO will be outranked by the other two constraints 18% of the time and consequently the candidate depicting i-epenthesis will be selected in 18% of the tableaux evaluations in this grammar. In contrast, Nuc/V will be dominated 54% of the time resulting in ONS at the same percentage level. Finally, the candidate representing the target-like coda variant will be selected 28% of the time because this is the number of times in which the constraint NoCoda(Obs) will be dominated by its overlapping neighbors. In Table 4, compare the results predicted by the GLA (under *GLA*) with the ones encountered in the data (under O*bs*): the grammar learned by the GLA generates output frequencies that are virtually identical to those observed in the investigation.

(11) Table 4: Tableaux selection for Level 2 Formal – Predicted (GLA) and observed results (Obs)

	Output Selection			**Results**	
Rankings	**[i]**	**ONS**	**Coda**	**GLA**	**Obs**
Nuc/V >> NoCoda(Obs) >> DEP-IO	✓			18	18
NoCoda(Obs) >> Nuc/V >> DEP-IO	✓				
DEP-IO >> NoCoda(Obs) >> Nuc/V		✓		54	55
NoCoda(Obs) >> DEP-IO >> Nuc/V		✓			
DEP-IO >> Nuc/V >> NoCoda(Obs)			✓	28	28
Nuc/V >> DEP-IO >> NoCoda(Obs)			✓		

Due to space limitations, we will neither illustrate the tableaux evaluations nor discuss the rankings assigned for each grammar assumed in the study. Instead, we summarize below the stochastic analyses for the six distinct grammars that characterize the development of English codas by Brazilian Portuguese speakers (where the superscripted numbers indicate the ranking values assigned by the GLA). Observe under each variant that the grammar learned by the GLA generates output frequencies that are strikingly close or similar to those observed in the data. We thus conclude the StOT analysis of the variable phenomena that characterize the development of English codas by BP speakers.

(12) Summary of grammars by proficiency and style: Predictions (GLA) and observed results (Obs)

Grammars by Proficiency & Style	**Epenthesis** >> DEP-IO		**ONS** >> Nuc/V		**Coda** >> NoCoda	
	GLA	**Obs**	**GLA**	**Obs**	**GLA**	**Obs**
L1, Informal: MAX-IO^{106} >> NoCoda(Obs)$^{99.8}$ >> Nuc/V$^{99.1}$ >> DEP-IO$^{95.1}$	89	90	7	6	5	4
L1, Formal: MAX-IO^{106} >> NoCoda(Obs)$^{99.3}$ >> Nuc/V$^{99.1}$ >> DEP-IO$^{95.6}$	84	84	9	9	7	8
L2, Informal: MAX-IO^{106} >> NoCoda(Obs)$^{99.5}$ >> Nuc/V$^{97.5}$ >> DEP-IO^{97}	51	52	38	39	10	9
L2, Formal: MAX-IO^{106} >> DEP-IO$^{98.7}$ >> NoCoda(Obs)$^{98.2}$ >> Nuc/V^{97}	18	18	54	55	28	28
L3, Informal: MAX-IO^{106} >> Nuc/V$^{98.5}$ >> NoCoda(Obs)$^{97.8}$ >> DEP-IO$^{97.7}$	39	39	25	23	36	38
L3, Formal: MAX-IO^{106} >> DEP-IO$^{99.9}$ >> Nuc/V$^{97.2}$ >> NoCoda(Obs)$^{96.9}$	5	5	38	39	58	56

5. Conclusions

We have provided in this study evidence for Goad's (2002) proposal that the process of acquiring codas in first (Goad and Brennan 2003) and second languages (Goad and Kang 2003) may be mediated by a transitory stage which, in the case of Brazilian Portuguese English, is triggered by a combination of L1 and L2 features as well as universal principles of syllabification: Onset-Nucleus Sharing. Under this approach, word-final consonants are syllabified as onsets and, in order to be licensed, they spread certain inherent features to empty nucleus positions. This analysis was confirmed in our study via the lengthening and aspiration of word-final voiceless stops in intermediate stages of L2 acquisition.

More generally, within a sociolinguistic approach for data collection and analysis, this study investigated the variable development of word-final codas in BPE across three stages of proficiency and two formality levels. The VARBRUL statistical results have shown that the occurrence of the three variants that characterize L2 coda acquisition (namely i-epenthesis, ONS and coda production) is conditioned by both extralinguistic (i.e. proficiency and style) and linguistic factors (i.e. the markedness of labials and dorsals vis-à-vis coronals when they syllabify in coda position).

Finally, we have demonstrated how a stochastic version of OT can satisfactorily serve as a model for modeling variability: it not only allows for variation to be directly encoded in the grammar, but it also incorporates into the same grammar a mechanism that captures the quantitative aspect of variable phenomena, i.e. continuous constraint ranking and stochastic evaluation. In addition, StOT is able to account for variation and its predictability via the same linguistic constraints and principles that govern categorical phenomena. This results in a simpler grammar, with fewer constraints (cf. Reynolds 1994, Anttila 1997, Cardoso 2001, 2003). The claim that the probability of each variant's occurrence may be encoded in (and therefore predicted by) the grammar yields important consequences for the study of variation and linguistic theory in general, because it constitutes an attempt to narrow down the distinction between what is traditionally labeled as competence versus performance. By proposing an analysis in which variation as well as the predictability of each variable output is encoded in the grammar (and therefore into competence), we obtain a more accurate and comprehensive approach to the study of language. This analysis (among many others in the sociolinguistic literature) presupposes that variation is an inherent part of what is normally referred to as competence.

References

Adamson, H. Douglas (1988). *Variation theory and second language acquisition.* Washington, D.C.: Georgetown University Press.

Anttila, Arto (1997). Deriving variation from grammar: A study of Finnish genitives. In F. Hinskens, R. van Hout and L. Wetzels (eds.), *Variation, Change and Phonological Theory*. Amsterdam: John Benjamins, 35-68.

Archangeli, Diana (1997). Optimality Theory: An introduction to linguistics in the 1990s. In D. Archangeli and D. T. Langedoen (eds.), *Optimality Theory – An Overview.* Oxford: Blackwell, 1-32.

Archibald, John. (1993). *Language Learnability and L2 Phonology*. Dordrecht: Kluwer Academic Publishers.

Beckman, Jill (1998). *Positional faithfulness*. PhD dissertation, University of Massachusetts, Amherst.

Bernhardt, Barbara and Joseph Stemberger (1998). *Handbook of Phonological Development: From the Perspective of Constraint-Based Nonlinear Phonology*. San Diego CA: Academic Press.

Boersma, Paul (1998). *Functional phonology: formalizing the interactions between articulatory and perceptual drives.* The Hague, Netherlands: Holland Academic Graphics.

Boersma, Paul (2000). Learning a grammar in functional phonology. In J. Dekkers, V. van der Leeuw, and J. van der Weijer (eds.), *Optimality Theory: Phonology, Syntax and Acquisition*. Oxford: Oxford University Press, 465-523.

Boersma, Paul and Bruce Hayes (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry 3*2, 45-86.

Broselow, Ellen and Daniel Finer (1991). Parameter Setting in Second Language Phonology and Syntax. *Second Language Research* 7 (1), 35-59.

Broselow, Ellen, Su-I. Chen and Chilin Wang (1998). The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition* 20, 261-280.

Bybee, Joan (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Cardoso, Walcir (1999). A quantitative analysis of word-final /r/-deletion in Brazilian Portuguese. *Linguistica Atlantica* 21, 13-52.

Cardoso, Walcir (2001). Variation patterns in regressive assimilation in Picard. *Language Variation and Change* 13, 305-341.

Cardoso, Walcir. (2003). *Topics in the phonology of Picard*. PhD thesis, McGill University. Published by the McGill Working Papers in Linguistics.

Cardoso, Walcir (2007). The variable development of English word-final stops by Brazilian Portuguese speakers: A stochastic optimality theoretic account. *Language Variation and Change* 19, 1-30.

Davidson, Lisa, Peter Jusczyk and Paul Smolensky (2004). The Initial and Final States: Theoretical Implications and Experimental Explorations of Richness of the Base. In R. Kager, J. Pater, & W. Zonneveld (Eds.), *Constraints in phonological acquisition*. Cambridge: Cambridge University Press. 321.368.

Chomsky, Noam (1988). *Language and problems of knowledge: the Managua lectures.* Cambridge: MIT Press.

Gatbonton, Elizabeth (1975). *Systematic variation in second language speech: a sociolinguistic study.* PhD thesis, McGill University.

Goad, Heather (2002). Markedness in right-edge syllabification: Parallels across populations. *CJL 47*, 151-186.

Goad, Heather and Kathleen Brannen (2003). Phonetic evidence for phonological structure in syllabification. In J. van de Weijer, V. van Heuven, and H. van der Hulst (eds.), *The Phonological Spectrum,* Vol. 2. Amsterdam: John Benjamins, 3-30.

Goad, Heather and Hyun-Sook Kang (2003). Word-final syllabification in L2 acquisition with emphasis on Korean learners of English. In J. Liceras et al. (eds.), *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA)*, 122-129.

Hayes, Bruce, Bruce Tesar, and Kie Zuraw (2003). OTSoft 2.1 software package. URL: http://www.linguistics.ucla.edu/people/hayes/otsoft

Heyer, Sarah (1986). *English final consonants and the Chinese learner*. MA Thesis, Southern Illinois University.

Jensen, John Thayer (1977). *Yapese reference grammar*. Honolulu: University of Hawaii Press.

Kroch, Anthony (1994). Morphosyntactic variation. In K. Beals, J. Denton, R. Knippen, L. Melnar, H. Suzuki and E. Zeinfeld (eds.), *Papers from the 30th Regional Meeting of the Chicago Linguistic Society*. Volume 2: The Parasession on Variation in Linguistic Theory, 180-201.

Labov, William (1966). *The social stratification of English in New York City.* Washington, DC: Center for Applied Linguistics.

Leopold, Werner (1939). *Speech development of a bilingual child: a linguist's record.* New York: AMS Press.

Major, Roy (1986). Paragoge and degree of foreign accent in Brazilian English. *Second Language Research 2*, 53-71.

Major, Roy (1996). Markedness in second language acquisition of consonant clusters. In R. Bayley and D. Preston (eds.), *Second Language Acquisition and Linguistic Variation*. Amsterdam: John Benjamins, 75-96.

Major, Roy (2001). *Foreign Accent. The Ontogeny and Phylogeny of Second Language Phonology.* New Jersey: Lawrence Erlbaum Associates.

Oostendorp, Marc van (1995). *Vowel quality and phonological projection*. PhD thesis, Tilburg University.

Oostendorp, Marc van (1997). Style levels in conflict resolution. In F. Hinskens, R. van Hout and W. Wetzels (eds.), *Variation, Change and Phonological Theory.* Amsterdam: John Benjamins, 207-229.

Pater, Joseph (1996). *Consequences of constraint ranking.* PhD dissertation, McGill University.

Pater, Joe and A. Werle (2001). Typology and variation in child consonant harmony. In C. Féry, A. Dubach Green and R. van de Vijver (eds.). *Proceedings of HILP5*, University of Potsdam.

Pintzuk, Susan (1988). *VARBRUL programs*. Philadelphia: University of Pennsylvania.

Preston, Dennis (1996). Variationist perspectives on second language acquisition. In R. Bayley and D. Preston (eds.), Second *Language Acquisition and Linguistic Variation.* Amsterdam: Benjamins, 1-45.

Prince, Alan and Paul Smolensky (1993). *Optimality Theory: Constraint interaction in generative grammar*. Cambridge, Massachusetts: MIT Press.

Reynolds, William (1994). *Variation and phonological theory*. PhD thesis, University of Pennsylvania.

Rose, Yvan (2000). *Headedness and prosodic licensing in the L1 acquisition of phonology*. PhD thesis, McGill University.

Selinker, Larry (1972). Interlanguage. *International Review of Applied Linguistics 10*, 209-31.

Selkirk, Elisabeth (1972). *The phrase phonology of English and French.* PhD thesis, MIT.

Steele, Jeffrey (2002). L2 learner's modification of target language syllable structure: Prosodic licensing effects in interlanguage phonology. In A. James and J. Leather (eds.), *New sounds 2000.* University of Klagenfurt, 315-324.

Tagliamonte, Sali A. (2006). *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Tranel, Bernard (1996). French liaison and elision revisited: A unified account within Optimality Theory. In C. Parodi, C. Quicoli, M. Saltarelli, and M. Zubizarreta (eds.), *Aspects of Romance Linguistics*. Washington, DC: Georgetown University Press, 433-455.

Trubetzkoy, Nikolai. (1939). Grundzuge der phonologie. Göttingen: Vandenhoeck und Ruprecht. (It also appears in English translation as Principles of Phonology, translated by C. Baltaxe. Berkeley and Los Angeles: University of California Press, 1969).

Wilson, John and Alison Henry (1998). Parameter setting within a socially realistic linguistics. *Language in Society* 27, 1-21.

Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition North America (GALANA)

edited by Alyona Belikova,
Luisa Meroni, and Mari Umeda

Cascadilla Proceedings Project Somerville, MA 2007

Copyright information

ISBN 978-1-57473-419-5 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Cardoso, Walcir and Denis Liakin. 2007. Onset-Nucleus Sharing in Interlanguage: Evidence from Brazilian Portuguese English. In *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition North America (GALANA)*, ed. Alyona Belikova, Luisa Meroni, and Mari Umeda, 61-72. Somerville, MA: Cascadilla Proceedings Project.

or:

Cardoso, Walcir and Denis Liakin. 2007. Onset-Nucleus Sharing in Interlanguage: Evidence from Brazilian Portuguese English. In *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition North America (GALANA)*, ed. Alyona Belikova, Luisa Meroni, and Mari Umeda, 61-72. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1547.