

**ROBERTA QUINTANILHA AZEVEDO**

**A EPÊNTESE NO PORTUGUÊS BRASILEIRO (L2),  
EM SEGMENTOS PLOSIVOS EM CODAS MEDIAIS,  
POR FALANTES NATIVOS DO ESPANHOL COLOMBIANO (L1):  
UMA ANÁLISE VIA TEORIA DA OTIMIDADE ESTOCÁSTICA  
E GRAMÁTICA HARMÔNICA**

Dissertação de Mestrado em Letras, área de concentração em Linguística Aplicada, apresentada como requisito parcial para a obtenção do título de Mestre pelo Programa de Pós-Graduação em Letras da Universidade Católica de Pelotas

**Orientador: Profa. Dra. Carmen Lúcia Barreto Matzenauer**

**Co-Orientador: Prof. Dr. Ubiratã Kickhöfel Alves**

**Pelotas**

**2011**

## AGRADECIMENTOS

Considero este trabalho mais um desafio da minha vida pessoal e acadêmica. Enfrentei uma área de pesquisa que eu não conhecia e que hoje faz parte da minha formação acadêmica. Mas não teria conseguido sem o apoio de pessoas que foram fundamentais ao meu desenvolvimento nesta etapa.

À minha orientadora Carmen Matzenauer e ao meu co-orientador Ubiratã K. Alves, responsáveis por meus conhecimentos de fonologia e minha motivação e gosto pela Linguística, pelo acompanhamento paciente e participação integral no meu crescimento.

À querida professora Andréia Rauber, exemplo de profissional dedicado e atencioso, pelo apoio e segurança dados, em todos os momentos em que solicitei ajuda.

Aos meus pais e irmãos incondicionalmente.

À professora Mabel Mascarenhas Wiegand, Diretora do Departamento de Intercâmbio e Programas Internacionais da UFPel, pelo incentivo e cooperação durante estes dois anos de curso.

Aos estudantes colombianos, participantes da pesquisa.

E aos amigos, em especial ao Joe Edmer, e aos que conheci neste programa.

## RESUMO

Este trabalho tem por objetivo descrever e analisar, à luz da Gramática Harmônica - HG, (LEGENDRE, MIYATA & SMOLENSKY, 1990; SMOLENSKY & LEGENDRE, 2006) e da Teoria da Otimidade Estocástica - OT (BOERSMA & HAYES, 2001), o emprego de segmentos plosivos em codas mediais de palavras, passíveis de motivação de epêntese por parte de aprendizes colombianos de Português Brasileiro (PB), falantes nativos de Espanhol. A realização da pesquisa foi motivada pela licença que a L1 dos informantes oferece para a presença de segmentos plosivos em final de sílaba em oposição à evitação de tal estrutura silábica por falantes do PB. Para a constituição do *corpus*, foi proposto um instrumento composto de palavras cognatas do Português e Espanhol para a leitura em frases veículo, gravadas no Programa AUDACITY 1.3.5 (ANSI). Foram escolhidos, para este trabalho, quatro estudantes colombianos que permaneceram no Brasil por um período médio de seis meses, em mobilidade acadêmica na cidade de Pelotas-RS, com o cuidado de serem homogeneizados fatores extralinguísticos, como sexo, idade e nível de escolaridade. Os dados foram submetidos a uma análise acústica com o uso do *software* PRAAT, desenvolvido por Paul Boersma e David Weenink (versão 5.1.12). Foi confirmada a hipótese de que, mesmo em palavras cognatas nas duas línguas analisadas, há diferente tratamento às plosivas em coda, dependendo do sistema linguístico que está sendo utilizado pelo falante de Espanhol aprendiz de PB. Tal fato foi captado, à luz da HG e da OT Estocástica, via hierarquia de restrições universais.

## ABSTRACT

This study aims to describe and analyze, according to the frameworks of Harmonic Grammar – HG (LEGENDRE, MIYATA & SMOLENSKY, 1990; SMOLENSKY & LEGENDRE, 2006) and Stochastic Optimality Theory – OT, (BOERSMA & HAYES, 2001), the production, by Colombian Learners of Brazilian Portuguese (BP), of stop segments in word-mid codas, liable for the production of epenthetical segments. This investigation was motivated by the possibility of occurrence of stop codas in the learners' L1, but not in the L2, as Brazilian Portuguese tends to avoid this syllabic structure. In order to build the *corpus*, an instrument composed of Portuguese and Spanish cognate words was proposed, so that vehicle phrases could be recorded in AUDACITY 1.3.5 (ANSI) software. Four Colombian students, who remained in Brazil for a period of six months, in an academic mobility program in the city of Pelotas-RS, were chosen to take part in the investigation. In the selection of the participants, extralinguistic factors such as sex, age and scholarity were homogenized. The data were submitted to an acoustic analysis using PRAAT (5.1.12), developed by Paul Boersma and David Weenick. We confirmed the hypothesis that even in cognate words, there is a different treatment to plosive in coda position in both analyzed languages, depending on the language system which is being used by the speaker. Such fact was collected, according to the Harmonic Grammar and Stochastic Optimality Theory, through a hierarchy of universal constraints.

## LISTA DE TABELAS

Tabela 1 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /bd/ nas palavras “abdutor” e “abdome” .....	58
Tabela 2 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /bt/ nas palavras “obtusos” e “obtención” .....	59
Tabela 3 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /pt/ nas palavras “rapto” e “corrupto” .....	60
Tabela 4 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /pn/ para as palavras “apneia” e “hipnose” .....	60
Tabela 5 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /gm/ para as palavras “enigma” e “dogma” .....	61
Tabela 6 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /gn/ para as palavras “magno” e “signo” .....	62
Tabela 7 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /kt/ para as palavras “cacto” e “convicto” .....	63
Tabela 8 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /kn/ para as palavras “estricnina” e “tecnicismo” .....	64
Tabela 9 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /dm/ para as palavras “admisión” e “admirador” .....	65
Tabela 10 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /tm/ para as palavras “ritmo” e “algoritmo” .....	65
Tabela 11 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /tn/ para as palavras “etnia” e “vietnam” .....	66
Tabela 12 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /bd/ nas palavras “abdutor” e “abdome” .....	68
Tabela 13 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /bt/ nas palavras “obtenção” e “obtusos” .....	68
Tabela 14 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /pt/ nas palavras “rapto” e “corrupto” .....	69
Tabela 15 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /pn/ nas palavras “apneia” e “hipnose” .....	69

Tabela 16 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /gm/ nas palavras “enigma” e “dogma” .....	71
Tabela 17 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /gn/ nas palavras “magno” e “signo” .....	71
Tabela 18 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /kt/ nas palavras “cacto” e “convicto” .....	72
Tabela 19 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /kn/ nas palavras “tecnicismo” e “estricnina” .....	73
Tabela 20 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /dm/ nas palavras “admissão” e “admirador” .....	74
Tabela 21 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /tm/ nas palavras “algoritmo” e “ritmo” .....	74
Tabela 22 - <i>Outputs</i> das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /tn/ nas palavras “vietnam” e “etnia” .....	75
Tabela 23 - Produções das sequências [p.n, k.n, p.t, k.t] no Espanhol Colombiano.....	81
Tabela 24 - Produções das sequências [p.n, k.n, p.t, k.t] no Português Brasileiro.....	81
Tabela 25 – Produção das sequências [p.n, k.n, t.n, p.t, k.t] no EC .....	104
Tabela 26 - Produção das sequências [p.n, k.n, t.n, p.t, k.t] no PB .....	104

## LISTA DE TABLEAUX

<i>Tableau 1</i> - Exemplo de <i>tableau</i> no Modelo da OT Estocástica.....	35
<i>Tableau 2</i> - Exemplo de <i>tableau</i> no modelo da Gramática Harmônica.....	38
<i>Tableau 3</i> - Exemplo de <i>tableau</i> .....	39
<i>Tableau 4</i> - O problema <i>harmonic bounding</i> exemplificado em <i>tableau</i> .....	41
<i>Tableau 5</i> - Ilustrativo - Exemplo da formalização de Restrições de Conjunção Local .....	86
Conjunto de <i>Tableaux 6</i> - Simulação 1 OT / Aquisição do Espanhol Colombiano (0% de epêntese) .....	88
Conjunto de <i>Tableaux 7</i> - Simulação 1 HG / Aquisição do Espanhol Colombiano (0 de epêntese) .....	90
Conjunto de <i>Tableaux 8</i> - Simulação 2 OT / Aquisição do Português Brasileiro (100% de epêntese) .....	93
Conjunto de <i>Tableaux 9</i> - Simulação 2 HG / Aquisição do Português Brasileiro (100% de epêntese) .....	95
Conjunto de <i>Tableaux 10</i> - Simulação 3 OT / Interlândia ( <i>outputs</i> variáveis) .....	98
Conjunto de <i>Tableaux 11</i> - Simulação 3 OT / Interlândia –Segunda rodada.....	99
Conjunto de <i>Tableaux 12</i> - Simulação 3 HG / Interlândia ( <i>outputs</i> variáveis).....	101
Conjunto de <i>Tableaux 13</i> - Simulação 3 HG / Interlândia – Segunda rodada .....	102
Conjunto de <i>Tableaux 14</i> - Simulação 1 OT / Aquisição do Espanhol Colombiano (0% de epêntese) .....	108
Conjunto de <i>Tableaux 15</i> - Simulação 1 HG / Aquisição do Espanhol Colombiano (0% de epêntese) .....	110
Conjunto de <i>Tableaux 16</i> - Simulação 2 OT / Aquisição do Português Brasileiro (100% de epêntese) .....	114
Conjunto de <i>Tableaux 17</i> - Simulação 2 HG / Aquisição do Português Brasileiro (100% de epêntese) .....	117
Conjunto de <i>Tableaux 18</i> - Simulação 3 OT / Interlândia ( <i>outputs</i> variáveis) .....	121
Conjunto de <i>Tableaux 19</i> - Simulação 3 OT / Interlândia – Segunda rodada.....	122
Conjunto de <i>Tableaux 20</i> - Simulação 3 HG / Interlândia ( <i>outputs</i> variáveis).....	125
Conjunto de <i>Tableaux 21</i> - Simulação 3 HG / Interlândia – Segunda rodada .....	126
Conjunto de <i>Tableaux 22</i> - Simulação 1 OT / Aquisição do Espanhol Colombiano (0% de epêntese) .....	132

Conjunto de <i>Tableaux</i> 23 - Simulação 1 HG / Aquisição do Espanhol Colombiano (0% de epêntese) .....	136
Conjunto de <i>Tableaux</i> 24 - Simulação 2 OT / Aquisição do Português Brasileiro (100% de epêntese) .....	141
Conjunto de <i>Tableaux</i> 25 - Simulação 2 HG / Aquisição do Português Brasileiro (100% de epêntese) .....	145
Conjunto de <i>Tableaux</i> 26 - Simulação 3 OT / Interlíngua ( <i>outputs</i> variáveis) .....	151
Conjunto de <i>Tableaux</i> 27- Simulação 3 OT / Interlíngua – Segunda Rodada .....	153
Conjunto de <i>Tableaux</i> 28 - Simulação 3 HG / Interlíngua ( <i>outputs</i> variáveis) .....	157
Conjunto de <i>Tableaux</i> 29 - Simulação 3 HG / Interlíngua – Segunda Rodada .....	159

## LISTA DE FIGURAS

Figura 1 - Diagrama proposto por Bisol (1999) para a hierarquia prosódica de Nespor e Vogel (1986).....	19
Figura 2 - Representação da sílaba na Teoria Autossegmental .....	20
Figura 3 - Modelo Silábico de Selkirk (1982).....	20
Figura 4 - Escala de sonoridade .....	23
Figura 5 - Espraimento do traço coronal.....	27
Figura 6 - Exemplo da palavra “signo” na estrutura silábica proposta por Selkirk (1982).....	28
Figura 7 - Processo de ressilabação da palavra “signo” na estrutura silábica proposta por Selkirk (1982).....	28
Figura 8 - Exemplo de resultado da gramática na OT fornecida pelo PRAAT .....	37
Figura 9 - Produção da palavra “admirador” pelo sujeito “S” na frase “Diga admirador” .....	50
Figura 10 - Produção da palavra “rapto” pelo sujeito “D” nas frases “Diga rapto novamente” e “Puedes decir rapto de nuevo” .....	52
Figura 11 - Produção da palavra “rapto” na frase em Português “Diga rapto novamente” pelo sujeito “S” .....	56
Figura 12 - Produção da palavra “magno” na frase em Espanhol “Puedes decir magno de nuevo” pelo sujeito “A” .....	62
Figura 13 – <i>Output Distributions</i> da aquisição do Espanhol Colombiano – OT .....	90
Figura 14 – <i>Output Distributions</i> da aquisição do Espanhol Colombiano – HG .....	92
Figura 15 – <i>Output Distributions</i> da aquisição do Português Brasileiro – OT .....	94
Figura 16 – <i>Output Distributions</i> da aquisição do Português Brasileiro – HG .....	97
Figura 17 - <i>Output Distributions</i> da Interlíngua – OT .....	100
Figura 18 – <i>Output Distributions</i> da Interlíngua – HG .....	103
Figura 19 – <i>Output Distributions</i> da aquisição do Espanhol Colombiano – OT .....	109
Figura 20 - <i>Output Distributions</i> da aquisição do Espanhol Colombiano – HG.....	112
Figura 21 – <i>Output Distributions</i> da aquisição do Português Brasileiro – OT .....	116
Figura 22 – <i>Output Distributions</i> da aquisição do Português Brasileiro – HG .....	119
Figura 23 – <i>Output Distributions</i> da Interlíngua – OT .....	123
Figura 24 – <i>Output Distributions</i> da Interlíngua – HG .....	128
Figura 25 - <i>Output Distributions</i> da aquisição do Espanhol Colombiano – OT .....	135
Figura 26 - <i>Output Distributions</i> da aquisição do Espanhol Colombiano – HG.....	139

Figura 27 - <i>Output Distributions</i> da aquisição do Português Brasileiro – OT .....	144
Figura 28 - <i>Output Distributions</i> da aquisição do Português Brasileiro – HG.....	149
Figura 29 - <i>Output Distributions</i> da Interlíngua – OT .....	156
Figura 30 - <i>Output Distributions</i> da Interlíngua – HG .....	163

## LISTA DE QUADROS

Quadro 1 - Sequências com plosiva em coda escolhidas para a pesquisa .....	46
Quadro 2 - Sequências com plosiva em coda excluídas da pesquisa.....	46

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>14</b>
<b>2 REFERENCIAL TEÓRICO</b> .....	<b>18</b>
2.1 A TEORIA DA SÍLABA .....	18
2.1.1 A sílaba .....	18
2.1.2 Princípios universais na estruturação silábica: a sequência de sonoridade e o contato silábico .....	22
2.2 A SÍLABA NO PORTUGUÊS BRASILEIRO E NO ESPANHOL .....	24
2.2.1 Padrões silábicos no Português Brasileiro .....	24
2.2.2 Padrões Silábicos no Espanhol .....	31
2.3 A TEORIA DA OTIMIDADE ESTOCÁSTICA E A GRAMÁTICA HARMÔNICA .....	33
2.3.1 Pressupostos da Teoria da Otimidade Estocástica .....	33
2.3.1.1 A Teoria da Otimidade Estocástica.....	33
2.3.1.2 O Algoritmo de Aprendizagem Gradual (OT-GLA) (Boersma e Hayes, 2001).....	36
2.3.2 Pressupostos da Gramática Harmônica .....	38
2.3.2.1 A HG e o Algoritmo de Aprendizagem Gradual (HG-GLA) .....	38
<b>3 METODOLOGIA</b> .....	<b>42</b>
3.1 CARACTERIZAÇÃO DOS INFORMANTES .....	42
3.1.1 Os Informantes .....	42
3.1.2 O teste de proficiência em Português Brasileiro .....	43
3.2 O INSTRUMENTO DE COLETA DE DADOS .....	44
3.3 PROCEDIMENTOS DE APLICAÇÃO DO INSTRUMENTO AOS INFORMANTES, APRENDIZES DO PB .....	47
3.4 MÉTODO DE DESCRIÇÃO E ANÁLISE DOS DADOS .....	49
<b>4 DESCRIÇÃO DOS DADOS</b> .....	<b>54</b>
4.1 CONSIDERAÇÕES FUNDAMENTADORAS PARA AS DISCUSSÕES DOS DADOS NO PORTUGUÊS BRASILEIRO E NO ESPANHOL .....	55
4.1.1 A soltura e não soltura da plosiva em coda .....	55
4.1.2 A Sequência plosiva + nasal .....	57
4.2 DESCRIÇÃO DOS DADOS DE PRODUÇÃO NO ESPANHOL COLOMBIANO (L1) .....	58
4.2.1 Plosivas bilabiais em coda: [p] e [b] .....	58
4.2.2 Plosivas velares em coda: [k] e [g].....	61
4.2.3 Plosivas alveolares em coda: [t] e [d] .....	65
4.3 DESCRIÇÃO DOS DADOS DE PRODUÇÃO NO PORTUGUÊS BRASILEIRO (L2) .....	67

4.3.1 Plosivas bilabiais em coda: [p] e [b] .....	67
4.3.2 Plosivas velares em coda: [k] e [g].....	70
4.3.3 Plosivas alveolares em coda: [t] e [d] .....	74
4.4 COMPARAÇÃO DOS DADOS DE PRODUÇÕES DOS SUJEITOS NA L1 (ESPAÑHOL COLOMBIANO) E NA L2 (PORTUGUÊS BRASILEIRO) – GENERALIZAÇÕES .....	77
<b>5 AQUISIÇÃO DO PORTUGUÊS BRASILEIRO VIA TEORIA DA OTIMIDADE ESTOCÁSTICA E GRAMÁTICA HARMÔNICA.....</b>	<b>79</b>
5.1 A FORMALIZAÇÃO DAS RESTRIÇÕES PARA AS SEQUÊNCIAS [p.n, k.n, p.t, k.t]: UM ESTUDO-PILOTO.....	79
5.1.1 Simulação 1 - Aquisição do Espanhol Colombiano (L1) .....	87
5.1.2 Simulação 2 - Aquisição do Português Brasileiro (L2) .....	92
5.1.3 Simulação 3 – Interlândia ( <i>outputs</i> variáveis) .....	97
5.2 A FORMALIZAÇÃO DA HIERARQUIA DE RESTRIÇÕES PARA AS SEQUÊNCIAS [p.n, k.n, t.n, p.t, k.t]: A ANÁLISE FINAL .....	103
5.2.1 Análise dos dados via Teoria da Otimidade Estocástica e Gramática Harmônica: Conjunto Restrito de Restrições .....	105
5.2.1.1 Simulação 1 - Aquisição do Espanhol Colombiano (L1) .....	107
5.2.1.2 Simulação 2 - Aquisição do Português Brasileiro (L2) .....	113
5.2.1.3 Simulação 3 – Interlândia ( <i>outputs</i> variáveis) .....	119
5.2.2 Análise dos dados via Teoria da Otimidade Estocástica e Gramática Harmônica: Conjunto Completo de Restrições .....	128
5.2.2.1 Simulação 1 - Aquisição do Espanhol Colombiano (L1) .....	131
5.2.2.2 Simulação 2 - Aquisição do Português Brasileiro (L2) .....	140
5.2.2.3 Simulação 3 – Interlândia ( <i>outputs</i> variáveis) .....	150
5.2.3 Considerações finais das análises nos modelos da HG e OT Estocástica .....	163
<b>6 CONCLUSÕES .....</b>	<b>169</b>
<b>REFERÊNCIAS .....</b>	<b>174</b>
<b>ANEXOS.....</b>	<b>177</b>
<b>ANEXO A – FORMULÁRIO DE CONSENTIMENTO INFORMADO .....</b>	<b>178</b>
<b>ANEXO B – FICHA DE INFORMAÇÃO DO PARTICIPANTE .....</b>	<b>180</b>

## 1 INTRODUÇÃO

A ocorrência de epêntese nas várias línguas do mundo é fenômeno conhecido, mas ainda não esgotado nas discussões da área da fonologia. Para este trabalho, trataremos a epêntese como um processo fonológico caracterizado pela inserção de um som vocálico no *output* das palavras, som esse não presente no *input*. Tal fenômeno vai servir para tornar menos marcada a estrutura silábica de uma dada língua.

Em Collischonn (1997), é proposto que, na silabação do Português Brasileiro (PB), uma consoante não apta em posição silábica de ataque ou coda – consoante perdida – não seria ligada a nenhum nó silábico, o que desencadearia a criação de uma outra sílaba, desprovida temporariamente de núcleo vocálico. Esse seria o contexto perfeito para a inserção de um núcleo vocálico, como podemos perceber em palavras como *et[i]nia*<sup>1</sup> e *ac[i]ne*. Segundo Collischonn (2000, 2002), o processo de epêntese medial é variável no PB, mas tem emprego prevalente entre os falantes da língua. Pode afirmar-se que tal fenômeno é observado em falantes adultos de PB, bem como no processo de aquisição da língua, seja como primeira língua (L1) ou segunda língua (L2).

O fenômeno da epêntese na aquisição de uma segunda língua (L2)<sup>2</sup> tem dimensão relevante, porque pode demonstrar a forte relação do aprendiz com sua língua materna, durante o processo de aquisição de uma nova língua. Em se tratando de falante nativo de Português, por exemplo, diante de palavra com uma construção estranha às nossas estruturas silábicas, provavelmente tal aprendiz utilizaria alguma estratégia para facilitar a produção do padrão silábico que não ocorre na L1. É o caso, por exemplo, do que ocorre com a palavra *big[i]* (do inglês ‘big’, grande), quando pronunciada por falantes de PB, já que o Português não licencia plosivas em posição de coda silábica – a epêntese motiva uma ressilabação, o que simplifica a sílaba original CVC, tornando-a uma palavra dissilábica, com a estrutura CV.CV.

Um questionamento pertinente poderia ser proposto na situação inversa à exemplificada acima, ou seja, sobre o que ocorre com um aprendiz de uma língua materna de padrão mais marcado (como é o caso do Espanhol), em processo de aquisição de uma língua com um padrão menos marcado (como é o caso do Português Brasileiro). Qual seria o comportamento desse sujeito, diante de uma palavra em Português - cognata no Espanhol -

---

<sup>1</sup> Durante todo o trabalho utilizaremos colchetes [] para delimitarmos a transcrição fonética e barras oblíquas //, ao nos referirmos à transcrição fonológica.

<sup>2</sup> Este trabalho não fará distinção entre os termos Língua Estrangeira (LE) e Segunda Língua (L2).

como “*corrupto*”, que apresenta um segmento plosivo em coda, permitido no Espanhol, mas é realizada com epêntese por falantes do PB?

Neste trabalho, portanto, procuramos estabelecer qual seria a tendência de um falante de Espanhol Colombiano adquirindo Português, diante de palavras cognatas nas duas línguas. A hipótese é a de que os sujeitos aplicarão o processo de epêntese, assim como nós, brasileiros, por mostrar-se processo prevalente na gramática em aquisição, ainda que possa ser variável, e por tratar-se da aquisição de um padrão menos complexo, presente na L2.

Para verificar a hipótese suscitada de que os estudantes recorrerão à epêntese na produção das palavras selecionadas, foram gravadas produções de quatro sujeitos colombianos que estiveram na cidade de Pelotas (RS/Brasil) realizando intercâmbio por um período médio de seis meses. Eles foram submetidos a uma tarefa de leitura de seqüências recorrentes que continham o contexto fonológico, para o aparecimento do fenômeno da epêntese no Português Brasileiro.

Por fim, após a verificação acústica, os dados gravados foram dispostos para análise através dos modelos da Gramática Harmônica (LEGENDRE, MIYATA & SMOLENSKY, 1990; SMOLENSKY & LEGENDRE, 2006) e da Teoria da Otimidade Estocástica (BOERSMA & HAYES, 2001), que consideramos instrumentos teóricos interessantes para formalizar o fenômeno levantado. A Gramática Harmônica (*Harmonic Grammar* – HG) diferencia-se da Teoria da Otimidade (OT) no que diz respeito à avaliação do candidato ótimo; na avaliação do melhor *output*, a OT Estocástica converte os pesos numéricos em uma hierarquia em que a restrição em posição mais alta se mostra, por ela mesma, capaz de eliminar candidatos. Na HG, é abandonada a noção de dominância estrita, visto que a organização das restrições universais se dá por pesos numéricos que, na escolha do candidato ótimo, apresentam caráter cumulativo.

A contribuição desta pesquisa, portanto, diz respeito ao ineditismo da proposta. Desconhecemos, em nível de Mestrado, estudos em aquisição de L2 que trabalhem com o modelo teórico da Gramática Harmônica, bem como pesquisas sobre a aquisição do PB como L2 com foco no comportamento de seqüências consonantais em relação ao processo de epêntese. Além do mais, reconhecemos como relevante que os dados levantados das produções dos aprendizes colombianos do Português Brasileiro poderão servir como banco de dados a outras pesquisas, e até mesmo para trazer evidências sobre o comportamento de aprendizes estrangeiros, e suas possíveis dificuldades na aquisição do PB.

A fim de atender ao nosso objetivo geral de descrever e analisar, à luz da Gramática Harmônica e da Teoria da Otimidade Estocástica, o emprego de segmentos

plosivos em codas mediais de palavras, passíveis de motivação de epêntese na língua alvo, por parte de aprendizes colombianos de PB, falantes nativos de Espanhol, os quais produzem segmentos plosivos em final de sílaba em sua L1, apresentamos os objetivos específicos a seguir:

1 - Descrever os índices de ocorrência de produção de segmentos plosivos, em coda silábica medial, por aprendizes de Português Brasileiro, falantes nativos de Espanhol (colombianos), em palavras cognatas presentes em ambos os sistemas, observando a presença ou não do processo de epêntese;

2 – Identificar fatores linguísticos que contribuem para o emprego da epêntese por parte de falantes nativos de Espanhol (colombianos) no uso do PB como L2;

3 – Verificar, com base na Gramática Harmônica e na Teoria da Otimidade Estocástica, quais restrições são relevantes para o processo de aquisição do Português Brasileiro, por parte de aprendizes colombianos, no que se refere à produção das plosivas em posição medial de palavra;

4 - Identificar a(s) hierarquia(s) de restrições capaz(es) de caracterizar as produções linguísticas dos sujeitos da pesquisa, no processo de aquisição do Português Brasileiro como L2;

5 – Discutir o poder preditivo do modelo da Gramática Harmônica frente a estudos de aquisição de L2;

6 – Contribuir, em termos empíricos e teóricos, para o entendimento do processo de aquisição do PB por falantes nativos de Espanhol.

A partir dos objetivos supracitados, as questões de investigação que nortearão o nosso estudo são seguintes:

1 – A epêntese vocálica que ocorre no PB, após segmentos plosivos, em codas medias de palavras, será produzida por aprendizes colombianos de PB, cuja L1 admite segmentos plosivos em coda?

2 – Quais fatores linguísticos contribuem para a ocorrência da epêntese por parte de falantes nativos do Espanhol Colombiano?

3 – No que diz respeito à análise, à luz da Gramática Harmônica e da Teoria da Otimidade Estocástica, quais restrições serão relevantes para o processo de aquisição dos padrões silábicos do PB como L2, por parte de falantes do Espanhol Colombiano?

4 – Qual(is) ordenamento(s) de restrição(ões) dá(ão) conta das formas de *output* produzidas pelos aprendizes, seguindo os pressupostos da HG e da OT?

5 – O modelo da Gramática Harmônica consegue formalizar com adequação o processo de aquisição de L2? Há vantagens frente à Teoria da Otimidade Estocástica?

6 – Considerando tanto o plano empírico quanto o teórico, de que forma o presente estudo pode contribuir para o entendimento do processo de aquisição do PB por falantes do Espanhol colombiano?

Para a organização do estudo, o trabalho foi dividido em sete capítulos. Seguindo a Introdução, o segundo capítulo se compromete a apresentar o Referencial Teórico, que vai tratar desde a teoria da sílaba e padrões silábicos nas línguas de interesse (Português e Espanhol), passando por questões relevantes para o desenvolvimento do estudo como a ocorrência da epêntese no sul do Brasil, até chegar aos pressupostos dos modelos teóricos utilizados para a formalização e análise dos dados.

No terceiro capítulo, apresentamos a metodologia utilizada para o desenvolvimento do trabalho, que vai lidar com a caracterização dos informantes selecionados para a pesquisa, a elaboração do instrumento para a coleta de dados e procedimentos de aplicação desse instrumento, assim como com o método de descrição e análise dos dados e a escolha das restrições que irão compor as simulações futuras nos modelos teóricos da Gramática Harmônica e da Teoria da Otimidade Estocástica, para que possamos passar à descrição e à comparação dos dados de produção nas frases em Português e Espanhol, realizadas no capítulo 4.

Enfim, o capítulo 5 vai apresentar a análise dos dados na Teoria da Otimidade e na Gramática Harmônica para que, dessa forma, possamos, até o capítulo 4 - feita a descrição dos dados - responder às Questões Norteadoras 1 e 2 e com o capítulo 5, esgotar as demais questões, para apresentar no capítulo 6 as Conclusões do trabalho e, a seguir, as Referências Bibliográficas utilizadas.

## **2 REFERENCIAL TEÓRICO**

Este capítulo aborda os pressupostos teóricos que fornecem subsídios e suporte ao desenvolvimento do trabalho. Os assuntos relevantes ao presente estudo vinculam-se a dois focos de atenção: a verificação da possibilidade da ocorrência da epêntese no Português Brasileiro, por parte de falantes nativos de Espanhol Colombiano, em segmentos plosivos que ocupariam a posição de coda medial, e a análise desse fenômeno com base nos modelos teóricos da OT Estocástica e da Gramática Harmônica.

A História da Fonologia vem percorrendo um caminho importante desde o século passado, com início em autores como Jakobson e Trubetzkoy que, inspirados em Sausurre, lançam obras importantes à Fonologia Estrutural. Esse caminho vai ter continuidade com - dentre outras propostas - a Fonologia Gerativa (Chomsky e Halle, 1968), Fonologia Métrica (Selkirk, 1982), Fonologia Autossegmental (Clements e Hume, 1995), Fonologia Lexical (Kiparsky, 1982), a Teoria da Sílabas (Selkirk, 1982), a Teoria Prosódica (Booij, 1983; Nespor e Vogel, 1986) até chegar à Gramática Harmônica (Legendre; Myata; Smolensky, 1990) e à Teoria da Otimidade (Prince e Smolensky, 1993).

De início, acreditamos ser importante, primeiramente, dedicar um espaço para informações a respeito do nível prosódico com que iremos trabalhar – o nível da sílaba - e justificar a perspectiva teórica utilizada nesse caso, para, então, fazermos uma explanação da ocorrência da epêntese tanto na língua brasileira quanto na língua materna dos informantes, a fim de tratarmos o fenômeno na aquisição de língua estrangeira com base no funcionamento da Gramática Harmônica e da Teoria da Otimidade Estocástica, como modelos teóricos interessantes para a verificação da aquisição de língua estrangeira e das variações que se observam nesse processo.

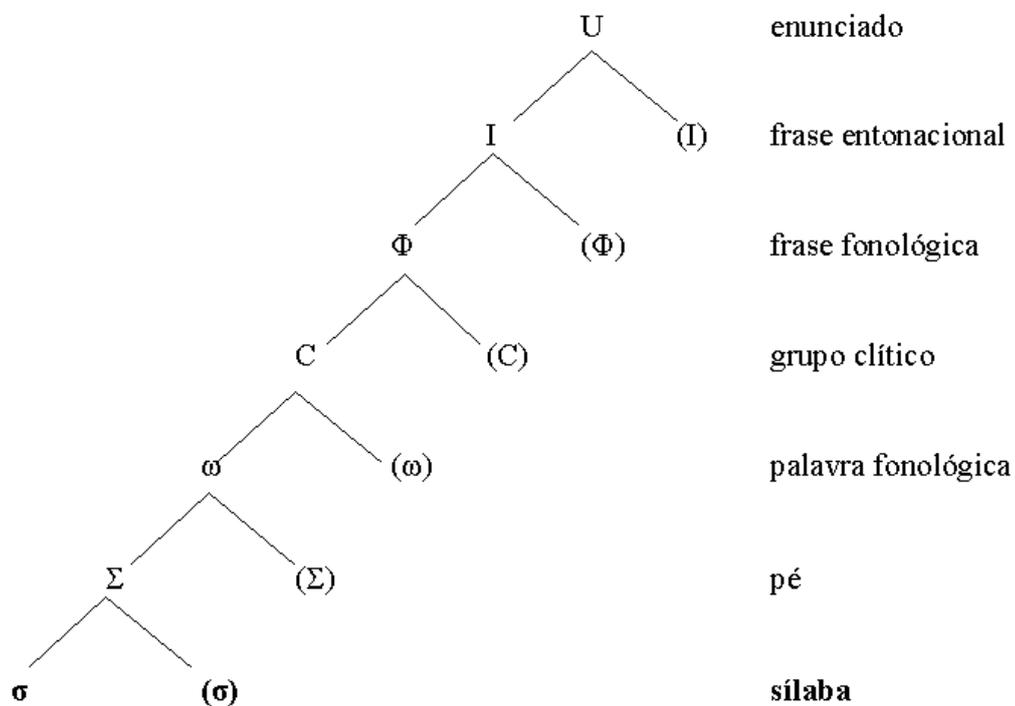
### **2.1 A TEORIA DA SÍLABA**

#### **2.1.1 A sílaba**

Cada língua possui um conjunto de sons com valor distintivo que compõe seu sistema fonológico. Porém, além desses sons, o sistema também é composto por unidades

prosódicas nas quais tais segmentos se organizam e cujo menor componente é a sílaba<sup>3</sup>. As análises fonológicas têm demonstrado que a sílaba é uma importante unidade de análise, cuja ocorrência, na Teoria da Otimidade e na Gramática Harmônica, será controlada por uma interação de restrições, conforme veremos.

De acordo com Bisol (1999), a sílaba é um constituinte que tem um espaço importante na teoria fonológica, e é fundamental dentro da hierarquia prosódica por se constituir no domínio de regras e processos. Abaixo, mostramos o diagrama proposto por Bisol (1999) para a hierarquia de constituintes prosódicos, segundo Nespor e Vogel (1986):



**Figura 1 - Diagrama proposto por Bisol (1999) para a hierarquia prosódica de Nespor e Vogel (1986)**

Com a estrutura prosódica que expusemos em (1), fica facilitada a visualização do nível prosódico com que vamos trabalhar na presente pesquisa.

A respeito da estrutura interna da sílaba, segundo Collischonn (2005a, p.101) podemos dizer que há basicamente duas teorias importantes: a Teoria Autossegmental e a Teoria Métrica da Sílaba. Dos autores que defendem a primeira, podemos citar Kahn (1976), Clements e Keyser (1983), Ito (1986) e Nespor e Vogel (1986). Já alguns autores que trabalham com a segunda perspectiva são Pike (1967), Kyparsky (1979), Selkirk (1982) e Harris (1983). É importante dizer, nesse sentido, que o que vai diferenciar uma concepção da

<sup>3</sup> Para Nespor e Vogel (1986), a sílaba é o menor constituinte prosódico.

outra com relação à sílaba é a visão sobre a aplicação de regras fonológicas, se será realizada no âmbito da sílaba ou em constituintes menores, como *onset* e rima.

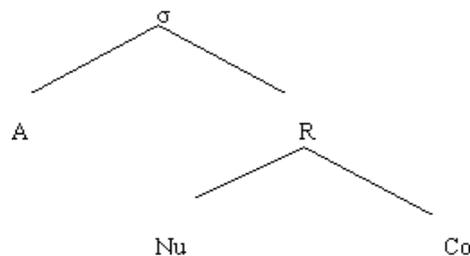
A Teoria Autossegmental vai representar a sílaba em camadas e vai formalizar os elementos que formam *onset* (ou ataque), núcleo e coda como constituintes independentes, conforme é mostrado na Figura 2.



**Figura 2 - Representação da sílaba na Teoria Autossegmental**

Fonte: Collischonn (1997, p.101)

Diferente da Teoria Autossegmental, a Teoria Métrica da Sílaba, proposta por Selkirk (1982), vai representar a sílaba em constituintes basicamente binários: ataque e rima, de modo que a rima se subdivida em núcleo e coda. Dessa forma, vai prever uma relação mais próxima entre o Núcleo e a Coda do que entre o Núcleo e o Ataque, permitindo a proposição de generalizações que se refiram a um destes constituintes e não necessariamente a toda a sílaba.



**Figura 3 - Modelo Silábico de Selkirk (1982)**

Os constituintes podem se apresentar preenchidos ou não, podendo, ainda, ser ramificados ou não ramificados.

Em se tratando da representação de Selkirk, a rima é o único constituinte obrigatório. Esse modelo teórico trata da sílaba como organização hierarquizada, que será constituída por *onset* (ou ataque), formado por uma consoante ou por um encontro consonantal, e rima, que se divide em duas partes: o núcleo, que é, no mais das vezes, constituído de uma vogal, e a coda, preenchida por consoante(s).

Porém, nem todas as sílabas das línguas, inclusive as do Português, possuem todos esses constituintes representados na Fig. 3. Há sílabas que possuem somente o ataque e o núcleo e não a coda, como, por exemplo, em “pé”; outras apresentam apenas o núcleo, não contendo o ataque nem a coda, como, por exemplo, em “é”; no entanto, é condição que todas as sílabas contenham um núcleo.

Selkirk (1982, p.337) vai considerar três fatores para o estudo da sílaba como uma unidade da fonologia: as restrições fonotáticas, a aplicação de regras fonológicas e o tratamento do fenômeno suprasegmental do acento.

As restrições fonotáticas dizem respeito aos constituintes silábicos licenciados pela fonologia da língua e aos segmentos que podem integrá-los. Um exemplo de restrição fonotática é a proibição de, no Português Brasileiro, segmentos plosivos ocuparem coda silábica. Em aparecendo plosiva em coda, de forma a evitar uma construção não ideal na nossa língua, acrescemos a epêntese como alternativa.

Exemplo:            acne - a[k]\_ne ~ a[ki].ne

Na aplicação de regras fonológicas, temos a unidade “sílaba” como determinante de domínio de aplicação de grande parte das regras. Temos um exemplo na regra de neutralização da sibilante antes de consoante e em final de palavra do Português:

Exemplos:            Pasma - /paSmo/            [ˈpazmu]  
                              Susto - /suSto/            [ˈsustu]

Podemos perceber, nos exemplos acima, uma perda de oposição entre os fonemas /s/ e /z/, o que indica que, em contextos como os dos exemplos dados, /paSmo/ e /suSto/, em que a sibilante aparece em coda de sílaba (final ou medial seguida de outra consoante), teremos perda de distinção de sonoridade entre os segmentos /s/ e o /z/. É, portanto, a unidade “sílaba” e/ou seu constituinte “coda” o contexto para a aplicação da regra.

No que diz respeito ao tratamento da propriedade suprasegmental do acento, é requerido que os segmentos sejam agrupados em unidades que tenham o tamanho da sílaba. A partir da sílaba e da unidade “pé métrico”, podemos descrever o padrão de acento nas várias línguas do mundo, conforme abaixo, adaptado de Bisol (1999), a partir de Halle e Vergnaud (1987) – segundo Bisol (1999) e Halle e Vergnaud (1987), para a atribuição do acento, as línguas consideram a organização das sílabas em pés métricos<sup>4</sup>.

<sup>4</sup> Para Halle e Vergnaud (1987), a demarcação dos pés métricos é feita por parênteses, sendo o cabeça do pé identificado por um asterisco. Para Bisol (1992), basta a construção de um pé métrico, na borda direita da palavra, para a atribuição do acento primário, no PB.

Exemplo.:        bor. bo. lé. ta  
                       (\* .) (\* .)

A contar da direita para a esquerda, teremos o acento primário na penúltima sílaba. Essa é a tendência geral em línguas como Português e Espanhol, que constroem pés binários de cabeça à esquerda, a partir da borda direita da palavra. Como veremos mais adiante, Collischonn (2002) relaciona o fenômeno da epêntese ao acento: explica a ocorrência da epêntese em palavras que contenham a posição da plosiva em coda que está passível de correção, estabelecendo relação com o acento primário da palavra, ou seja, a epêntese ocorre em posição silábica à esquerda do acento primário (ex.: admissão).

### **2.1.2 Princípios universais na estruturação silábica: a sequência de sonoridade e o contato silábico**

Na arquitetura da Teoria da Otimidade Estocástica e da Gramática Harmônica, o foco central está nas restrições, que são universais e violáveis, ou seja, para esses modelos teóricos, a relação entre o *input* (representação fonológica) e o *output* dá-se por meio de restrições, ao invés de se dar por meio de regras, como ocorria na Fonologia Gerativa anterior à OT. A formalização dessas restrições, que substituem as regras específicas de língua, ocorre no componente CON<sup>5</sup> do modelo. Esta seção, acompanhada das demais neste capítulo teórico, será fundamental para a explicitação nesse sentido.

Seguindo a suposição da existência de um conjunto de restrições comum a todas as línguas, teremos propriedades invariantes ligadas a princípios que serão universais. Através de estudos comparativos entre as línguas do mundo, tentam-se estabelecer, por exemplo, os princípios da Gramática Universal que estão relacionados à seleção de fonemas e à constituição dos inventários das línguas.

Da mesma forma, procuram-se estabelecer modelos silábicos universais e aceita-se o molde CV – formado por uma consoante no ataque e uma vogal no núcleo – como universal. A fonologia da sílaba, seguindo essa linha, vai operar com um conjunto de princípios, como, por exemplo: o princípio da Sequência da Sonoridade (Selkirk, 1984); o princípio do Licenciamento Prosódico/PLP (Itô, 1986); o princípio da Soância/PS (Clements, 1990); a Lei do Contato Silábico (Murray e Vennemann, 1983).

---

<sup>5</sup> A HG e OT são formadas dos componentes CON, GEN, EVAL. Nesse momento importa estarmos cientes de que CON é o conjunto de restrições universais e violáveis.

Sob a ótica da OT Estocástica e da HG, no presente estudo propomos uma análise das produções de palavras constituídas de sequências heterossilábicas, formadas por plosivas em codas mediais e por consoantes no *onset* da sílaba seguinte. Sendo assim, a sonoridade terá seu lugar de destaque na construção das restrições. Defendemos, neste trabalho, que a Lei do Contato Silábico possibilita explicar as ocorrências de epêntese nas sílabas que destacamos para a análise dos dados, durante as produções dos informantes, bem como nos permite, com argumentos plausíveis, estipular este princípio como base para o estabelecimento das restrições universais utilizadas na análise proposta neste trabalho, considerando que a sonoridade faz parte da Gramática Universal e atua na organização dos segmentos em sílaba.

Murray & Vennemann (1983) apontam o fato de que há um contato ideal, entre duas sílabas adjacentes, que ocorre quando o segmento final da primeira sílaba tem sonoridade maior do que o segmento inicial da próxima sílaba. Como exemplo, temos a palavra “carta” /kar.ta/, em que a sequência /rt/ forma um bom contato silábico, já que /r/ (segmento da coda da primeira sílaba) tem maior grau de sonoridade do que /t/ (segmento do *onset* da sílaba subsequente).

Baseado na escala de hierarquia de sonoridade de Clements (1990), a seguir, fica mais fácil visualizarmos que a líquida /r/ vai ter sonoridade bem maior do que a da obstruente /t/, no exemplo dado com a palavra “carta”:

O	<	N	<	L	<	G	<	V	
-	-	-	-	+	+	+	+	+	syllabic (silábico)
-	-	-	+	+	+	+	+	+	vocoid (vocoide)
-	-	+	+	+	+	+	+	+	approximant (aproximante)
-	+	+	+	+	+	+	+	+	sonorant (soante)
0	1	2	3	4	5	6	7	8	rank (escala de sonoridade)

**Figura 4 - Escala de sonoridade**

Fonte: Clements (1990, p. 292)

Nesse sentido, temos que uma palavra como “rap.to”, com um encontro heterossilábico entre duas obstruintes [p.t] apresenta caráter menos marcado, por formarem um *plateau* de sonoridade, do que a sequência heterossilábica de uma palavra como “ap.neia” que, de acordo com a escala de sonoridade apresentada na Figura 4, apresenta um acréscimo de sonoridade de zero (obstruente em coda) para 1 (nasal em *onset* da sílaba seguinte).

Dessa forma, tentamos justificar a ocorrência maior de epêntese em encontros heterossilábicos tais como [p.n] ou [k.n] (com aumento de sonoridade do segmento da coda para o segmento do *onset*), do que em [p.t] e [k.t] (sequência com *plateau* de sonoridade) e

uma menor possibilidade de ocorrência deste fenômeno em encontros como uma líquida (sonoridade = 2) e uma obstruinte (sonoridade = 0), como em “car.ta” (com diminuição de sonoridade do segmento da coda para o segmento do *onset*).

A sonoridade é relevante para a silabação não apenas em se tratando do contato silábico: também é fundamental para a forma como as línguas estruturam as sílabas. Segundo explica Selkirk (1984), pelo Princípio de Sequência de Sonoridade, em qualquer sílaba, há um segmento que constitui um pico de sonoridade, que é precedido e/ou seguido por uma sequência de segmentos que progressivamente decrescem em valores de sonoridade. Para Clements (1990), é basilar, na estruturação de sílabas, o Princípio de Soância, segundo o qual o perfil de sonoridade do tipo silábico preferido nas línguas apresenta um crescimento máximo de sonoridade do ataque para o núcleo e um decréscimo mínimo do núcleo para a coda. Daí explica-se por que uma palavra como “ata” é silabada como *a.ta* em lugar de ser silabada como *\*at.a*.

## 2.2 A SÍLABA NO PORTUGUÊS BRASILEIRO E NO ESPANHOL

### 2.2.1 Padrões silábicos no Português Brasileiro

#### a) Moldes silábicos no Português Brasileiro

Diferentes modelos teóricos, como o de Kahn (1976), Clements & Keyser (1983), Levin (1985), Itô (1986), Milliken (1988) ou o de Noske (1992) levam ao entendimento da existência de moldes silábicos diversos. No presente estudo, utilizamos a representação da sílaba de Selkirk (1982), que já foi empregada em diversos trabalhos sobre silabação no PB, sob distintos modelos teóricos, que operavam tanto por regras quanto por restrições. Nesse modelo, teremos um Ataque (A) e uma Rima (R), e a Rima, por sua vez, é constituída de Núcleo (Nu) e Coda (Co), conforme mostramos anteriormente na figura 3.

De um modo geral, podemos dizer que o molde diz respeito à estrutura possível de uma sílaba numa dada língua, de modo a determinar o número máximo e mínimo de elementos permitidos.

Apresentamos para o Português Brasileiro os padrões abaixo, conforme Collischonn (2005, p.117):

V	<u>é</u>	Núcleo
VC	<u>ar</u>	Núcleo e Coda
VCC	<u>instante</u>	Núcleo e Coda Complexa
CV	<u>cá</u>	Ataque e Núcleo
CVC	<u>lar</u>	Ataque, Núcleo e Coda
CVCC	<u>monstro</u>	Ataque, Núcleo e Coda Complexa
CCV	<u>tri</u>	Ataque Complexo e Núcleo
CCVC	<u>três</u>	Ataque Complexo, Núcleo e Coda
CCVCC	<u>transporte</u>	Ataque Complexo, Núcleo e Coda Complexa
VV	<u>aula</u>	Núcleo e Coda
CVV	<u>lei</u>	Ataque, Núcleo e Coda
CCVV	<u>grau</u>	Ataque Complexo, Núcleo e Coda
CCVVC	<u>claustro</u>	Ataque Complexo, Núcleo e Coda Complexa

Considerando-se que cada língua apresenta moldes silábicos específicos e levando-se em conta os moldes silábicos do PB acima descritos, temos a hipótese de que a representação do falante, em relação à estrutura silábica de sua língua, será fundamental no processo de aquisição de uma LE, já que deve unir as informações fonológicas da língua materna às adquiridas na língua estrangeira.

Também é fundamental atentarmos para o fato de que é específica de cada sistema a determinação do tipo de segmento que pode ocupar as diferentes posições nos moldes silábicos – a seguir são explicitadas tais particularidades com referência à fonologia do PB.

#### b) Segmentos licenciados nas posições silábicas do Português Brasileiro

Especificamente para o Português, Collischonn (1997) resume as condições silábicas da seguinte forma:

1) Se tivermos apenas uma Consoante no Ataque, esse pode ser preenchido por qualquer Consoante do inventário da Língua Portuguesa, como, por exemplo, em [ba.ra.ta]<sup>6</sup>;

2) Para o caso de um Ataque complexo, ou seja, o Português licencia até duas consoantes no Ataque: a primeira consoante só poderá ser uma Obstruinte (plosiva ou fricativa labial) e a segunda, uma Líquida (/r/ ou /l/), como em [graus], ou conforme ocorre

<sup>6</sup> Três segmentos não estão licenciados para ocupar o *onset* absoluto de palavras do PB: o r fraco - /r/, a lateral palatal - /ʎ/ e a nasal palatal - /ɲ/.

em ambas as sílabas da palavra [claus.tro]; estão excluídas as sequências que iniciem por Coronais Contínuas, que são as Fricativas [s, z, ʃ, ʒ];

3) Na Rima, teremos no máximo três elementos, dos quais um é obrigatório, que é o Núcleo, que se restringe às Vogais, e os outros dois não são obrigatórios, que constituem a coda; se a coda for simples, será formada pelas Soantes /N, l, r/ ou por /S/, ou, se for complexa, pela sequência “Soante + /S/”.

Contudo, seguindo a linha de pensamento de que a condição de boa formação silábica vai determinar a estrutura silábica ideal das línguas, devemos imaginar que, nos casos de palavras que contenham sequências não dispostas conforme o molde silábico, teremos alternativas para contornar essa situação, como ocorre na palavra “*etnia*”, que na primeira sílaba apresenta a plosiva /t/ em coda, “et.nia”, ou seja, uma coda não licenciada pela língua. Nesse caso, tendemos a inserir uma vogal na palavra para contornar a situação não permitida, ficando a estrutura resultante com mais uma sílaba: [e.ti.ni.a] ~ [e.tʃi.ni.a]; todas as sílabas são, agora, licenciadas pela fonologia da língua.

c) A palatalização do /t/ e do /d/ antes do [i]

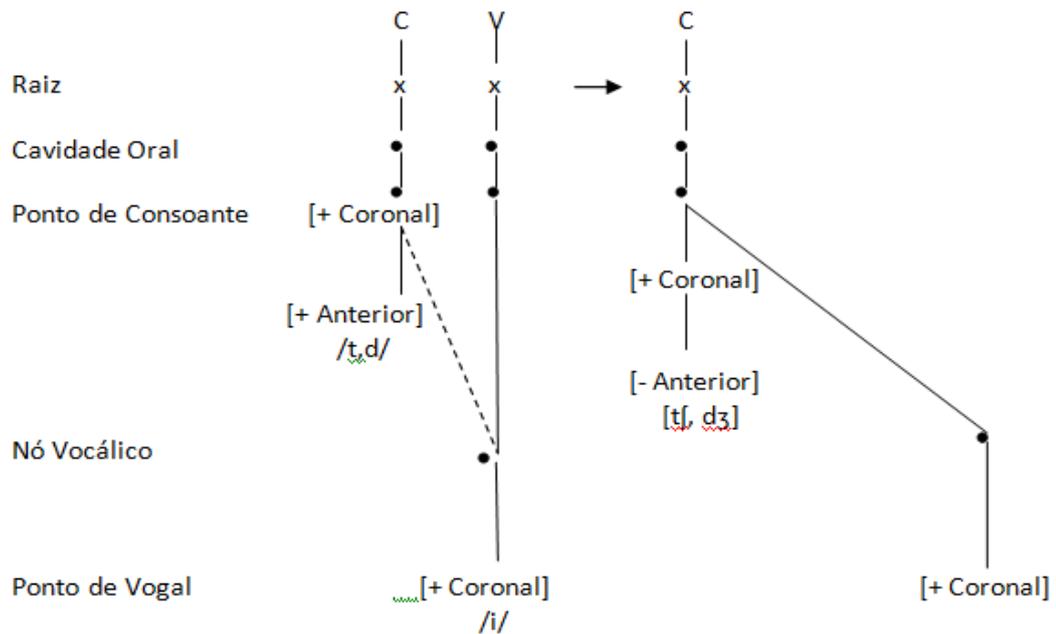
Reservamos este espaço para falar sobre o fenômeno da palatalização do /t/ e /d/ antes do [i] no PB e, particularmente, na cidade de Pelotas, para que, ao chegarmos nos dados de produção dos nossos informantes nas frases em Português Brasileiro, possamos estabelecer a coerência dos *outputs* realizados em palavras como “vietnam” ou “admissão”, presentes no instrumento de coleta do *corpus* do presente estudo.

A ocorrência desse processo nos dados coletados na fala dos nossos informantes é interessante porque, por um lado, evidencia uma possibilidade de *input* recebido pelos sujeitos da pesquisa na aquisição de PB que tiveram sua instrução no Português de forma não-formal e, por outro lado, revela a percepção dos estudantes diante de uma diferença que há entre os dialetos do PB.

Da literatura sobre o Espanhol, sabe-se que os segmentos /t/ e /d/ são dentais ([t] e [d]) diante da vogal [i], ou seja, sem palatalização. Não encontramos estudos variacionistas sobre a palatalização, específicos para a cidade de Bogotá na Colômbia, porém, pudemos observar nas produções coletadas na língua materna dos aprendizes que esses fonemas /t/ e /d/ diante do [i] são realizados como dentais, corroborando a literatura para esse contexto feita para o Espanhol Europeu.

Em alguns dialetos do Brasil, ocorre o fenômeno de palatalização de plosivas coronais. Autores como Clements & Hume (1995) caracterizam esse processo como de

espraiamento do traço coronal, que vai converter /t/ e /d/ [+anterior] em [tʃ] e [dʒ] [- anterior] antes do som [i]. A figura abaixo apresenta o caso, através da representação arbórea<sup>7</sup>.



**Figura 5 - Espraiamento do traço coronal**

Fonte: Matzenauer e Miranda (2010, p. 20)

Clements (1999), ainda sobre o mesmo processo, vai nos oferecer outra resposta. No trabalho em questão, o autor vai dizer que a palatalização ocorre não por assimilação, conforme identificado na estrutura arbórea acima, mas por uma regra de inserção do traço [+estridente]. Seguindo essa lógica, as africadas são plosivas simples, criadas por um desencontro fonético-fonológico.

Apesar de não termos estudos específicos para a cidade de Pelotas que possam sustentar nossa suspeita de que a palatalização dentre moradores nascidos seja categórica, em Matzenauer (2002) temos evidências do fenômeno em estudo sobre a produção do [t] e [d] antes de [i] por crianças de Pelotas e Porto Alegre. A autora vai discutir que, dentre as crianças pesquisadas, o alvo é a plosiva palato-alveolar, ou seja, o sistema lingüístico da criança vai dirigir-se à produção de [tʃ] e [dʒ] antes da vogal [i], indicando a tendência geral ao emprego do processo.

As considerações para o dialeto pelotense podem servir para explicar a ocorrência ou não da palatalização de [t] e [d] antes de [i], nas produções dos informantes da presente

<sup>7</sup> Segundo Clements (1999), as africadas são plosivas simples, criadas por um desencontro fonético-fonológico – em razão desse fato, diante da vogal [i], há a inserção do traço [+estridente] nas plosivas coronais, originando as africadas nesse contexto.

pesquisa nas frases em PB. Veremos posteriormente, na seção sobre os padrões silábicos do Espanhol, que não há motivação para que os colombianos palatalizem essas consoantes no contexto dessa língua, o que foi, inclusive, confirmado nos dados coletados. Logo, as situações de palatalização do [t] e [d] antes do [i] podem estar refletindo a interlíngua dos sujeitos da pesquisa.

d) A epêntese

À inserção de um elemento em uma palavra, nesse caso uma vogal, com a função de desfazer uma sequência indesejada, denominamos epêntese. Para melhor visualização do emprego do processo de epêntese no PB, utilizaremos o exemplo da palavra “signo”, cuja silabação é formalizada a seguir:

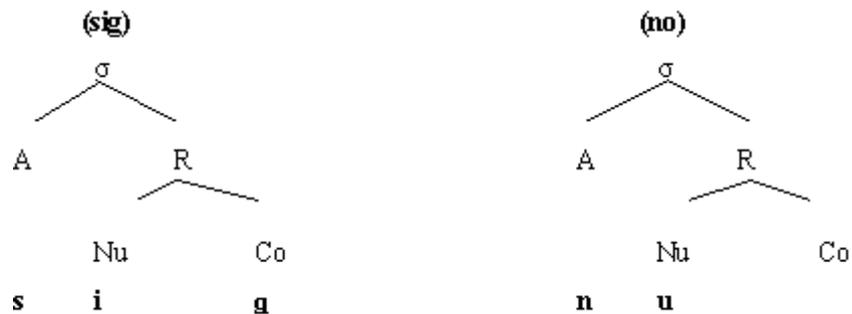


Figura 6 - Exemplo da palavra “signo” na estrutura silábica proposta por Selkirk (1982)

O problema que temos, para falantes de PB, na estrutura da palavra “signo”, conforme os moldes de sílaba discutidos na alínea (a) da seção 2.2.1, está na primeira sílaba, com a presença de uma plosiva [g] em posição não permitida, em coda silábica. Logo, para adequar esta consoante a um constituinte silábico em que seja licenciado, inserimos a vogal [i], após a plosiva [g], e criamos a situação para a nova silabação, em que o segmento plosivo dorsal, que ocupava uma posição de Coda não própria no PB, passa a ser Ataque de sílaba, conforme o molde ideal.

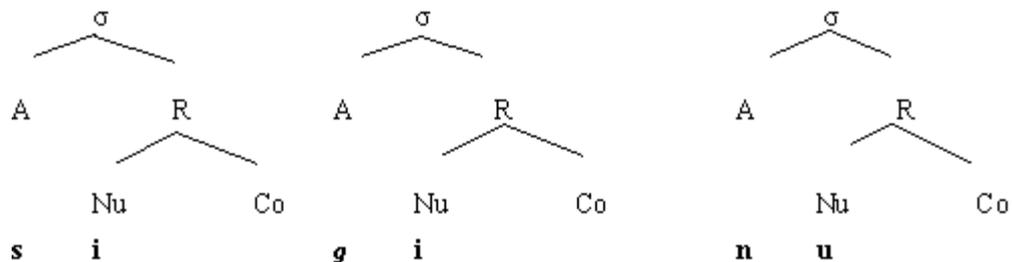


Figura 7 - Processo de ressilabação da palavra “signo” na estrutura silábica proposta por Selkirk (1982)

A figura 7, portanto, mostra o funcionamento da epêntese como mecanismo de correção, quando nos deparamos com sílabas não-licenciadas no Português Brasileiro. Retomando o que foi anteriormente explicitado, a coda silábica no PB pode ser constituída apenas pelas Soantes /N, l, r/ ou por /S/, ou pela sequência consecutiva de ambas, Soantes e /S/. Logo, o segmento [g] em Coda é indesejável, ou melhor, proibido pela fonologia da língua.

Não encontramos registro, na literatura, de estudos voltados à investigação do fenômeno da epêntese exclusivamente para falantes da cidade de Pelotas/RS, que venham a considerar todas as sequências-alvo contempladas na presente investigação. São limitados os estudos específicos a uma determinada cidade ou até mesmo a uma determinada região, que tratem de fenômenos variáveis como o da ocorrência da epêntese. Sabemos que numa mesma região encontraremos dialetos com diferenças significativas, em se tratando de um país como o Brasil de proporções territoriais e culturais enormes.

Dessa forma, devemos considerar algumas pesquisas que de alguma forma tragam evidências sobre o fenômeno variável da epêntese na região sul do Brasil, como o estudo de Alves e Lucena (2010), por exemplo, sobre a epêntese no dialeto gaúcho e paraibano em contatos silábicos de duas plosivas, produzidas por aprendizes de Inglês como L2.

O objetivo da pesquisa foi o de verificar as implicações, nos dialetos pelotenses e paraibano, do fenômeno variável do Afrouxamento da Condição de Coda/ACC, estudado por Bisol (1999), sobre as produções em Inglês por aprendizes falantes de tais dialetos do PB. Segundo os autores, a fricativa /f/ e as oclusivas /p/, /t/ e /k/ podem ocorrer de forma variável em codas não finais em palavras do Português, como a palavra /rapto/ = [‘Rap.tu] ou [‘Ra.pi.tu], nos dialetos pelotense e paraibano e a relação da produção variável das obstruintes em coda na L1 (Português) – com epêntese ou sem epêntese - com a taxa de produção de epêntese na aquisição de coda das obstruintes na L2 (Inglês).

Para tanto, os autores realizaram uma verificação nos dados de produção na L1 de doze gaúchos naturais de Pelotas/RS, do fenômeno de ACC, com vistas a identificar – dentre outros objetivos – se havia diferença entre os índices de produção de epêntese em dois estados brasileiros em que são usadas diferentes variantes do PB – Rio Grande do Sul e Paraíba -, para os segmentos [p], [k] e [f]. A conclusão foi a de que havia diferença entre os índices de epêntese encontrados nas localidades eleitas para o estudo e que é possível falarmos numa tendência mais forte, para o dialeto falado em Pelotas/RS, de Afrouxamento de Condição de Coda.

Em Collischonn (2002) também encontramos referências sobre o fenômeno fonológico de inserção vocálica após plosivas em coda - que estamos investigando -, embora o *corpus* estudado seja das capitais da Região Sul: Porto Alegre, Florianópolis e Curitiba. Com base em dados do projeto VARSUL, a autora conclui que a epêntese é variável, sendo dependente de fatores linguísticos que se mostram mais favoráveis à inserção da vogal, tais como:

1) A plosiva em Coda, motivadora da epêntese, estar em posição pré-tônica (ex.: vietnam) – explicada através das restrições do acento no PB, que favorecem a colocação do acento na penúltima sílaba da palavra, sendo evitada a inserção de sílaba à direita para impedir o deslocamento do acento;

2) O contexto seguinte à plosiva em Coda ser nasal (ex.: apneia) – explicada conforme Clements (1990), em que as sequências heterossilábicas oclusiva-nasal sofrem pressão considerável para serem modificadas em virtude de a primeira consoante da sequência ter grau de sonoridade menor que a segunda;

3) O contexto precedente ser uma consoante alveolar (ex.: admissão).

Porém, apesar de trazermos evidências com o estudo de Alves e Lucena (2010) sobre a ocorrência de um Afrouxamento de Condição de Coda variável na cidade de Pelotas e da existência de fatores mais favoráveis à inserção da vogal nos estudos de Collischonn (2002), considera-se que seria significativo, para o presente trabalho, um estudo sobre as possibilidades efetivas de *input* recebidas pelos sujeitos da pesquisa. Seria interessante um estudo voltado para o fenômeno na cidade de Pelotas, com a sequência de segmentos a que nos propusemos a analisar, para que pudéssemos estabelecer comparações com a nossa análise final junto aos modelos da Gramática Harmônica e Teoria da Otimidade.

Apesar dos poucos estudos sobre o tema, especialmente em se considerando a cidade de Pelotas/RS, local em que os estudantes receberam o *input* de PB, os resultados das pesquisas realizadas – que registraram presença predominante do uso da epêntese – permitiram que, para a presente investigação, tomássemos o processo como categórico; tal decisão foi relevante para caracterizarmos o estágio final de aquisição da L2 (PB), em um dos três tipos de simulação que realizamos à luz dos modelos teóricos da HG e da OT Estocástica.

Assim, tentamos com esta seção, justificar nossa opção por considerar, nas simulações computacionais a serem realizadas, a epêntese como alvo categórico a ser adquirido no PB como L2, ou seja, com 100% de produção de epêntese em coda medial de palavra na análise final.

Contudo, estamos cientes de que esta é uma limitação importante para o presente trabalho, embora não tenha sido nosso foco determinar os índices percentuais de epêntese que correspondem aos padrões-alvo a serem adquiridos por aprendizes de L2 residentes na cidade de Pelotas, o que deixaremos como questão para exploração futura.

## 2.2.2 Padrões Silábicos no Espanhol

### a) Moldes silábicos no Espanhol

Da mesma forma como ocorre no Português, no Espanhol também há um molde que vai permitir e vai restringir determinadas sequências, de forma a caracterizar a estrutura das sílabas nessa língua. Alguns padrões silábicos ocorrem no Português e no Espanhol da mesma forma, mas alguns padrões são específicos para cada uma das duas línguas.

Comparando-se os padrões silábicos do Português com os moldes silábicos do Espanhol, seguindo de padrões silábicos mais simples até os mais complexos, veremos que são possíveis, ao Espanhol, sílabas somente compostas de Vogais (V) e sílabas que admitem até cinco segmentos, assim como ocorre no Português, como, por exemplo, na palavra “claus.tro” (sílabas CCVVC). O Espanhol também terá sempre como Núcleo de sílaba uma Vogal, bem como no PB, conforme a seguir se mostra explicitado:

V	<u>al</u> amo	Núcleo
VC	<u>á</u> rbol	Núcleo e Coda
VCC	<u>ab</u> stracto	Núcleo e Coda Complexa
CV	<u>cal</u> or	Ataque e Núcleo
CVC	<u>par</u> te	Ataque, Núcleo e Coda
CVCC	<u>sans</u> crita	Ataque, Núcleo e Coda Complexa
CCV	<u>tri</u>	Ataque Complexo e Núcleo
CCVC	<u>tres</u>	Ataque Complexo, Núcleo e Coda
CCVCC	<u>trans</u> misor	Ataque Complexo, Núcleo e Coda Complexa
VV	<u>aula</u>	Núcleo e Coda
CVV	<u>rei</u> no	Ataque, Núcleo e Coda
CCVV	<u>trau</u> ma	Ataque Complexo, Núcleo e Coda
CCVVC	<u>claus</u> tro	Ataque Complexo, Núcleo e Coda Complexa

Mantemos, para a ilustração da estrutura silábica do Espanhol, o modelo que utilizamos para a sílaba do Português, de Selkirk (1982) – figura 3. Para a fundamentação teórica específica para a sílaba no Espanhol, consideramos as predições de Harris (1983), no que se refere aos componentes possíveis nas sequências silábicas. Vemos, assim, que o Espanhol também se assemelha ao Português no que se refere à obrigatoriedade de a sílaba conter um Núcleo, que seja Vogal, com os demais componentes da sílaba – Ataque e Coda – opcionais, porém, difere nos segmentos licenciados em posição de coda, conforme vemos a seguir.

b) Segmentos licenciados nas posições silábicas do Espanhol

Quanto às condições silábicas do Espanhol, temos que:

1) O Ataque é um constituinte opcional da sílaba no Espanhol; logo, uma sílaba nesse idioma pode existir sem Ataque, com Ataque simples (composto de uma consoante) ou Complexo, composto de duas Consoantes. No caso de Ataque simples, é permitida qualquer consoante do inventário fonológico do Espanhol, como em.: /zapato/, /hago/, /caja/;

2) Já em um Ataque Complexo, pode haver uma sequência de Obstruinte seguida de uma das líquidas /l/ ou /r/, como em.: /preso/, /blando/, /flan/;

3) Para a Rima, deve haver obrigatoriamente um Núcleo (Nu), que tem de ser Vogal. A Vogal, juntamente com o segmento que ocupa a Coda, formam as sequências listadas a seguir – nestas sequências que constituem “rimas silábicas”, pode-se ver que a Coda, que é um elemento opcional, pode ser ocupada por diferentes segmentos:

	Posição Medial	Posição Final
Vogal seguida de Glide	<u>au</u> .tor	lei
Vogal seguida de /s/	<u>pas</u> .ta	res
Vogal seguida de Líquida ou Líquida e /s/	<u>sal</u> .ta	mar/vals
Vogal seguida de Nasal ou Nasal e /s/	<u>com</u> .pra	sar.tén /Mayans
Vogal seguida de Obstruinte ou Obstruinte e /s/	<u>seg</u> .men.to	red / Felix [feliks]

A Rima no Espanhol, além disso, seguirá uma escala de sonoridade que respeitará o seguinte ordenamento: Vogal (V) > Glide (G) > Líquida (L) > Nasal (N) > Obstruinte (O). As Rimas que não foram especificadas como, por exemplo, as sequências abaixo, apesar de seguirem o ordenamento de sonoridade, surgem como sílabas impróprias ou mal formadas no Espanhol, segundo Harris (1983):

- Vogal Líquida Nasal (VLN);
- Vogal Líquida Obstruinte (VLO) – Com a Obstruinte diferente de /s/;
- Vogal Nasal Obstruinte (VNO) – Com a Obstruinte diferente de /s/.

Por outro lado, têm-se, também, as seguintes sequências, que representam exceções de construção silábica no Espanhol:

- Vogal Glide Nasal (VGN) – Encontrada em somente três palavras, em posição medial de palavra.: /veinte, treinta, aunque/.

- Vogal Glide Obstruinte (VGO) - para Obstruinte diferente de /s/ – Encontrada em um exemplo.: /auksilio/.

- Glide Vogal Glide (GVG) – Encontrada em posição final, em um pequeno número de palavras como.: /paraguai, buei, bioi/ e em posição interna, em palavras como nome próprio: Cuau.tla, ou palavras de origem indígena /gualaguaichu/.

O que se pode ver, contudo, é que a sequência que iremos analisar neste trabalho, com segmentos oclusivos em coda medial de palavras, passíveis de motivar epêntese no Português Brasileiro (já que não é uma sequência possível nos moldes do PB), é perfeitamente viável na estrutura da sílaba no Espanhol. Em outras palavras, nesta língua, oclusivas em codas mediais não constituem, com o *onset* seguinte, uma sequência passível de ressilabações para correções por má formação. Portanto, nas palavras cognatas em Português Brasileiro e Espanhol, que selecionamos para constituir o instrumento de frases-alvo, não devemos encontrar epêntese nas produções em Espanhol por parte dos aprendizes de PB, falantes nativos de Espanhol (Colombianos).

## 2.3 A TEORIA DA OTIMIDADE ESTOCÁSTICA E A GRAMÁTICA HARMÔNICA

### 2.3.1 Pressupostos da Teoria da Otimidade Estocástica

#### 2.3.1.1 A Teoria da Otimidade Estocástica

A Teoria da Otimidade ou OT (*Optimality Theory*), modelo proposto por Alan Prince e Paul Smolensky em 1993, preconiza que a gramática de cada língua seja determinada não mais por regras, que seriam incompatíveis às proposições do modelo, mas por hierarquias de restrições universais que podem ser violadas, diferenciando-se, nesse ponto, da Teoria

Gerativa Clássica. A OT abre mão das regras, porque opera com uma concepção de processamento linguístico inconciliável com elas, trabalhando com a ideia básica de que as línguas compartilham um conjunto de restrições universais ranqueáveis, porém contando com o pressuposto da violabilidade de tais restrições.

As restrições podem ser entendidas como exigências para que se alterem ou preservem estruturas que podem ou não ser violadas. De acordo com McCarthy (2002, p.13), as restrições são de dois tipos: (a) Marcação – que devem impor exigências na formação estrutural de *output*, ou seja, levam a *outputs* menos marcados. Exs.: As sílabas têm de ser abertas (No-Coda), oposição à consoante plosiva em coda (\*Stop-Coda) e (b) Fidelidade - que devem garantir que as formas ótimas de *output* sejam as mais semelhantes possível ao *input* recebido, conspirando contra qualquer tipo de mudança. Ex.: oposição à adição de segmento no *output* (DEP), oposição ao apagamento de um segmento no *output* (MAX).

Assim, as restrições que ocupam a posição mais alta na hierarquia de dado sistema, têm menor probabilidade de serem violadas pelo *output* ótimo, ao passo que as restrições ranqueadas em posição mais baixa estarão mais sujeitas à violação.

De acordo com McCarthy e Prince (1993), como propriedades fundamentais do modelo da Teoria da Otimidade, temos: **violabilidade**, **ranqueamento**, **inclusividade** e **paralelismo**. Quanto à violabilidade, as violações são permitidas, indicando que, na OT, as restrições universais são violáveis, fazendo com que a teoria seja mais explicativa; o ranqueamento vai ser responsável por ordenar as restrições de acordo com cada língua em particular; a inclusividade vai atuar de forma que a hierarquia de restrições possa avaliar o conjunto de candidatos que serão admitidos pelas considerações gerais de boa formação de estrutura; e o paralelismo trata de trabalhar com a melhor satisfação da hierarquia de restrições, considerando-se tanto a própria hierarquia, quanto o conjunto de candidatos ao mesmo tempo.

Ainda, a gramática nesse modelo de análise é formada pelos componentes:

- GEN (*generator*/gerador) - módulo da gramática capaz de tomar um *input* e criar infinitas possibilidades de candidatos a *output*;

- CON (*constraints*/conjunto de restrições) - o conjunto de restrições universais e violáveis; e

- EVAL (*evaluator*/avaliador) – módulo da gramática que realiza a avaliação dos candidatos, e determina o candidato ótimo a *output*.

Após a versão *Standard* da OT, temos a Teoria da Otimidade Estocástica (BOERSMA & HAYES, 2001), que, assim como na OT de Prince & Smolensky (1993), terá

restrições distribuídas em uma escala hierarquizada. Na avaliação dos concorrentes a *output*, o candidato ótimo será aquele que obedecer à restrição mais alta, independente do seu histórico frente às restrições mais baixas. Na OT Estocástica, porém, a hierarquia vai se dar através de índices numéricos que são atribuídos às restrições. Nesse sentido, a OT procura explicar a gramática de cada língua, atribuindo peso diferenciado às restrições, as quais, a partir de tais pesos, formarão uma hierarquia que explicará a gramática e particularmente cada fenômeno linguístico em análise para qualquer língua através do ordenamento das restrições.

O *tableau* a seguir, que apresenta restrições hierarquizadas por relações de dominância, e os *outputs* possíveis a partir da forma subjacente, representam esse modelo:

**Tableau 1 - Exemplo de *tableau* no Modelo da OT Estocástica**

Output/ Restrições	(40) Restrição A	(30) Restrição B	(15) Restrição C
[output 1]	*!		*
☞ [output 2]		***	

Como podemos ver, na primeira linha, da esquerda para direita, temos as restrições em ordem de dominância e, na primeira coluna, os candidatos possíveis, a partir de um *input* dado. O asterisco (\*) vai representar a violação de uma restrição, e o ponto de exclamação (!) vai determinar que houve uma violação fatal e que este candidato está fora da competição pelo *output* ótimo e a figura da mão apontando (☞), indica o candidato ótimo.

Podemos ver também, acima de cada Restrição (A, B, C), os índices numéricos responsáveis pelo ordenamento destas Restrições. A Restrição A, com um peso igual a 40, encontra-se mais altamente ranqueada do que as Restrições B e C, que, respectivamente apresentam pesos iguais a 30 e 15.

No exemplo 1, o candidato “1” foi eliminado, pois violou a restrição mais altamente ranqueada, deixando para o candidato “2” o *status* de candidato ótimo, já que viola uma restrição mais baixa no *ranking* e obedece à restrição mais altamente ranqueada, mesmo que tenha em seu histórico três violações à Restrição B. Isso quer dizer também que a Restrição A domina a Restrição B e C (RA>>RB>>RC).

Na análise via OT Estocástica podemos ainda, com o mesmo sistema de análise, ou seja, com as mesmas restrições, demonstrar a trajetória percorrida na aquisição da língua materna ou estrangeira. A relação entre os estágios desenvolvimentais é, nesse modelo, representada pelo deslocamento contínuo das restrições. Tal movimentação é tarefa de um algoritmo de aprendizagem, que tem a responsabilidade de guiar o sistema linguístico do

aprendiz, demovendo ou promovendo as restrições. O algoritmo proposto por Boersma e Hayes (2001), chamado Algoritmo de Aprendizagem Gradual (OT-GLA), será o algoritmo utilizado em nossa análise.

### **2.3.1.2 O Algoritmo de Aprendizagem Gradual (OT-GLA) (Boersma e Hayes, 2001)**

A proposta de Boersma e Hayes (2001) serve para dar conta da variação apresentada pelos aprendizes, bem como da gradualidade apresentada no processo de aquisição da língua à luz da OT Estocástica. Conforme explica Alves (2009), de acordo com os fundamentos do algoritmo em questão, o aprendiz vai receber um mapeamento *input-output* de cada vez e o estado corrente da gramática é que vai determinar o *output* ótimo.

A hierarquia de restrições será estabelecida em função de valores a serem assumidos pelas restrições em uma escala numérica, sob uma noção de ordenamento estocástico e não mais em relação a uma posição hierárquica horizontal. De acordo com a aprendizagem estocástica, a cada momento de avaliação dos candidatos, o ranqueamento é afetado por um valor de ruído estatístico (*noise*), que é sugerido em 2,0 na simulação computacional.

O algoritmo GLA é sensível aos erros do aprendiz, sendo, portanto, do tipo *error-driven*. Dessa forma, vai alterar o valor numérico das restrições quando o *output* da sua gramática se mostrar diferente do padrão encontrado na linguagem ambiente (erro). Esse índice de alteração se dá em função de um valor de plasticidade<sup>8</sup> que é adotado pelo algoritmo. Assim, o algoritmo vai diminuir (demover) o valor numérico das restrições que são violadas pelas formas encontradas de produção e vai aumentar (promover) o valor das restrições violadas pelo *output* “errado”, na hierarquia de restrições vigente.

Logo, cada restrição apresentará, na verdade, dois valores numéricos: o valor que corresponde ao ponto central da faixa de valores - valor central ou valor de ranqueamento (*ranking value*) – e, o ponto de seleção, que é um valor assumido dentro da faixa de valores, que corresponde ao valor exibido, no momento de avaliação dos candidatos.

O valor central das restrições, a ser aumentado e diminuído pelo algoritmo, vai ser atribuído em função do ruído determinado no sistema computacional. Então, em cada momento de produção podemos ter valores distintos, que Boersma e Hayes (2001) vão

---

<sup>8</sup> Plasticidade é um valor numérico através do qual o algoritmo vai ajustar o *ranking* das restrições. Esse valor é fornecido durante a simulação computacional, através da função *Learn* do PRAAT.

chamar de ponto de seleção, que estará contido em uma determinada faixa de valores.

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
Max	61.730	65.690	1.000000
*{k.n}	53.372	51.733	1.000000
*{p.n}	34.072	36.279	1.000000
*{k.t}	33.757	36.141	1.000000
*Dist {1}	34.072	35.240	1.000000
*{k}coda	33.757	30.266	1.000000
*{p.t}	23.134	26.555	1.000000
*Dist {1,0}	23.134	26.328	1.000000
*{stop}coda	23.134	22.678	1.000000
*{p}coda	23.134	20.816	1.000000
Dep	15.135	12.424	1.000000

Figura 8 - Exemplo de resultado da gramática na OT fornecida pelo PRAAT

O software é alimentado por *scripts*, com as restrições e com os índices percentuais de cada padrão de *output* produzido pelos sujeitos da pesquisa, a fim de poder simular o processo desenvolvimental por que passaram os falantes nativos de Espanhol na aquisição do PB. No resultado de uma simulação exemplificado na Figura 8 – que corresponde a uma etapa do processo de aquisição da língua estrangeira -, temos, em azul, as Restrições que foram utilizadas nessa análise. A primeira coluna após a lista das Restrições apresenta os valores que chamamos de ponto central, ao passo que a segunda coluna corresponde ao ponto de seleção do momento de avaliação. Por sua vez, a última coluna apresenta o valor de plasticidade que está sendo utilizado.

É, ainda, através dos valores centrais das Restrições, que este algoritmo OT-GLA vai representar *outputs* variáveis ou categóricas. Valores centrais bem afastados, ou seja, valores com distância superior a 10 pontos - como ocorre na Figura 8, entre os valores da Restrição DEP e as demais Restrições dadas – representam o resultado categórico, pois diferentes momentos de produção linguística, não causarão cruzamento na faixa de valores destas Restrições. Em contrapartida, um resultado da gramática fornecida pelo PRAAT, com valores centrais das Restrições com distância inferior a 10 pontos, estaria sujeito a cruzamento da faixa de valores, o que poderia alterar o ordenamento das Restrições e, conseqüentemente, o candidato ótimo, caracterizando a variação nos *outputs*.

Dessa forma, tentamos resumir a capacidade desse algoritmo de demonstrar os *outputs* variáveis na língua dos nossos aprendizes. Pretendemos, com isso, verificar se o GLA consegue dar conta dessa variação no processo de aquisição de uma L2, desde o estágio inicial, ainda correspondente à gramática da língua materna, até chegar à gramática da L2.

### 2.3.2 Pressupostos da Gramática Harmônica

Nas próximas seções, respectivamente, objetivamos descrever os pressupostos da Gramática Harmônica, além de fornecermos razões para a existência de uma interlíngua entre o Português Brasileiro e o Espanhol, ao descrevermos o funcionamento da aquisição da língua estrangeira sob o enfoque da Gramática Harmônica.

#### 2.3.2.1 A HG e o Algoritmo de Aprendizagem Gradual (HG-GLA)

O modelo da Gramática Harmônica, assim como a OT Estocástica, também vai trabalhar com as restrições com um caráter numérico. Além disso, como ocorre desde a OT considerada clássica - conforme explicam Jesney & Tessier (2007) - vai apresentar os módulos: GEN (*Generator*), CON (*Constraint*) e EVAL (*Evaluator*).

Entretanto, tal modelo diferencia-se, fundamentalmente, na avaliação do melhor *output* feita por EVAL, já que vai considerar, na escolha do candidato ótimo, o caráter cumulativo das violações de cada candidato considerando os pesos numéricos atribuídos às Restrições Universais, ao contrário do que ocorre na OT Estocástica, a qual não levava em conta valores cumulativos. Será escolhido como ótimo o candidato cujo valor de Harmonia for mais alto.

Basicamente, o princípio da HG é determinar para cada *input* um *output*, através da resolução de um conflito entre restrições, com a avaliação cumulativa dos valores atribuídos a essas restrições. Se utilizarmos o mesmo *tableau* já apresentado anteriormente em (1), formulado de acordo com o modelo da OT, tal *tableau* passaria a ser assim representado, de acordo com a HG, como em (2):

**Tableau 2 - Exemplo de tableau no modelo da Gramática Harmônica**

Output/Restrições	(40) Restrição A	(30) Restrição B	(15) Restrição C	H
☞ [Output 1]	*		*	-55
[Output 2]		***		-90

O *tableau* (2) tenta mostrar, então, que o módulo da gramática, EVAL, presente em ambos os modelos, OT e HG, vai apresentar um funcionamento diferenciado em cada uma das teorias. O valor numérico da Harmonia (H) é obtido ao multiplicarmos cada violação do

candidato (\*), que apresenta valor negativo, pelo valor de Ponto de Seleção (que corresponde ao valor da Restrição (R) que vem a ser assumido a cada momento de produção lingüística). Após tal operação, devem ser somados todos os resultados dessas multiplicações, referentes ao candidato.

Exemplo 1.:

$H \text{ do [output 1]} = R(A) \times (\text{n}^\circ \text{ violação}) + R(B) \times (\text{n}^\circ \text{ violação}) + R(C) \times (\text{n}^\circ \text{ violação})$

$H \text{ do [output 1]} = (40 \times -1) + (30 \times 0) + (15 \times -1)$

$H \text{ do [output 1]} = (-40) + (0) + (-15)$

$H \text{ do [output 1]} = -55$

Exemplo 2.:

$H \text{ do [output 2]} = R(A) \times (\text{n}^\circ \text{ violação}) + R(B) \times (\text{n}^\circ \text{ violação}) + R(C) \times (\text{n}^\circ \text{ violação})$

$H \text{ do [output 2]} = (40 \times 0) + (30 \times -3) + (15 \times 0)$

$H \text{ do [output 2]} = (0) + (-90) + (0)$

$H \text{ do [output 2]} = -90$

No modelo da OT Estocástica (ver *tableau 1*) observamos que, apesar de apresentar três violações, o *output 2* era o candidato ótimo, já que não violava a restrição mais alta. Porém, na HG, a situação será diferente (ver *tableau 2*), pois o número das violações de todas as Restrições, independentemente de seus valores hierárquicos, é relevante para o somatório final. Isso faz com que o *output 1*, seja o candidato ótimo e o *output 2*, eliminado, com um valor de Harmonia inferior ao do candidato 1.

Em outras palavras, a noção de cumulatividade das restrições violadas por um dado candidato, para a avaliação do candidato ótimo, no modelo da HG, vai distanciar tal teoria do modelo que opera sob uma noção de dominância estrita, conforme ocorre na OT Estocástica. Na Gramática Harmônica, o candidato ótimo é aquele que apresentar valor de Harmonia mais próximo de zero, conforme já foi referido.

**Tableau 3 - Exemplo de tableau**

/C1V2C3C4/	NoCoda (1.5)	Max (1.0)
HG: $\varphi$ [C1V2C3C4]	-1	
OT: $\varphi$ [C1V2]		-2

Fonte: Boersma e Pater (2008, p. 27)

No exemplo anterior (*Tableau 3*), à luz da HG, o candidato ótimo é o *output1*, ou [C1V2C3C4], já que seu valor de Harmonia (H) é o maior, ou o mais próximo a zero, em relação ao outro *output*. Conforme já explicado, o cálculo de H é feito multiplicando-se o número de violações na restrição, com o valor numérico da restrição:

$$H_{(\text{output})} = \text{número de violações} \times \text{valor da restrição}$$

$$H_{(\text{output1})} = (-1) \times (1.5) = -1.5$$

$$H_{(\text{output2})} = (-2) \times (1.0) = -2.0$$

O exemplo do *tableau 3* é interessante porque além de formalizar os dados no modelo da Gramática Harmônica, oferece possibilidade de comparação entre os modelos da OT e da HG. De acordo com o *tableau 3*, o candidato ótimo na HG é o candidato 1, porém, se com os mesmos dados utilizássemos o modelo teórico da OT, nosso candidato ótimo seria o *output2*, de acordo com o *tableau 1*.

Neste ponto, chamamos a atenção para a importância da escolha do modelo teórico que irá fundamentar a análise dos dados e, portanto, nossa preocupação em formalizar as produções encontradas dos sujeitos desta pesquisa nos dois modelos teóricos de análise (OT e HG): embora as restrições utilizadas sejam as mesmas, a avaliação dos candidatos e a escolha do candidato ótimo poderão ser diferentes.

Quanto ao algoritmo de aprendizagem nesse modelo, denominado de HG-GLA, assim como na OT (OT-GLA), tal algoritmo vai apresentar um valor de ruído, que é somado ao peso das Restrições. Esse algoritmo é disponibilizado no *software* PRAAT, e serve para expressar o processo de aquisição e as possíveis variações do *output*.

$$\text{Ex.: } \begin{array}{ccc} \overset{49}{50} & & \overset{52}{55} \\ 45 \quad \boxed{\text{R1}} \quad 55 & & 50 \quad \boxed{\text{R2}} \quad 60 \end{array}$$

No exemplo acima, procuramos mostrar duas Restrições com valores centrais diferentes, mas com pontos de seleção próximos. O valor de ruído produzido pelo algoritmo deve dar conta de tal fato, garantindo que, numa faixa estabelecida de 10 pontos (exemplos acima Restrição 1 = 45 a 55 e Restrição 2 = 50 a 60), o ponto de seleção varie, permitindo que ora R1 esteja acima, ora R2 esteja acima na hierarquia, o que vai garantir que tenhamos variação.

Cabe ressaltarmos, ainda, que tal algoritmo também é do tipo *error-driven* (guiado pelo erro), o que permite que as restrições se movimentem até o sujeito chegar ao seu alvo. E do tipo *Noisy* HG – especificação, Linear OT. Já que os valores de Harmonia podem ter pesos negativos, caso não utilizássemos o recurso Linear OT, poderíamos incorrer no

problema denominado de “*harmonic bounding*”<sup>9</sup>. De fato, a soma para que se chegue ao valor de Harmonia, após multiplicarmos as violações incorridas com os pesos das restrições, pode oferecer um sério engano, conforme podemos verificar no *tableau* 4.

**Tableau 4 - O problema *harmonic bounding* exemplificado em *tableau***

	R1(1.0)	R2 (-2.0)	
☞Output1	-1	-2	+3
Output2		-1	+2

O *tableau* 4 mostra, que uma restrição com peso negativo, acabou por erroneamente dar ao *output* 1 o status de candidato ótimo, quando, na verdade, deveria ser o *output* 2. Contudo, graças à seleção da função Linear OT no PRAAT, pode-se impedir que pesos negativos exerçam papel no cálculo do valor da Harmonia.

Encerrada a Revisão Teórica, passamos, a seguir, ao capítulo 3, que trata da Metodologia utilizada no desenvolvimento do presente trabalho.

---

<sup>9</sup> Lembramos que *harmonic bounding* é um problema da OT, não sendo exclusivo da versão Estocástica desse modelo teórico. Tal fenômeno é aqui referido pelo fato de ser, na utilização do algoritmo, o que motiva a escolha da opção “*Linear OT*” em lugar da opção “HG”.

### **3 METODOLOGIA**

#### **3.1 CARACTERIZAÇÃO DOS INFORMANTES**

##### **3.1.1 Os Informantes**

Fazem parte desta pesquisa quatro estudantes colombianos, que estiveram por um período médio de seis meses no Brasil, na cidade de Pelotas/RS, realizando um intercâmbio na Universidade Federal de Pelotas (UFPel). Três desses estudantes já haviam retornado ao país de origem antes do início desta pesquisa. Os quatro aprendizes cujos dados são analisados foram estudantes da Universidad de Ciências Aplicadas y Ambientales - UDCA, na cidade de Bogotá.

A justificativa para a escolha por informantes colombianos decorreu da busca de fatores que garantissem a consistência metodológica da pesquisa, como a disposição para a colaboração com a pesquisa, por parte dos sujeitos entrevistados, o número de sujeitos, o acesso a estes informantes durante a pesquisa e a homogeneização de fatores extralinguísticos, como sexo (todos são do sexo masculino), idade (entre 24 e 26 anos de idade), nível de escolaridade (último ano do ensino superior – 2 estudantes da Medicina Veterinária e 2 estudantes da Educação Física), o tempo de contato com o Português Brasileiro (6 meses), o uso não apenas da mesma L1, mas também da mesma variedade do Espanhol e o contato com o mesmo dialeto do PB.

Segundo as normas para a mobilidade internacional parcial da UFPel, não é necessário que estudantes de países que tenham como língua materna o Espanhol, e que almejam o intercâmbio na UFPel, apresentem qualquer tipo de comprovante de conhecimento ou proficiência no Português. Logo, grande parte dos estudantes de língua materna espanhola, que fazem a mobilidade internacional na UFPel, vêm a aprender o Português, durante o semestre acadêmico, no convívio diário, o que foi o caso dos aprendizes sujeitos desta pesquisa.

A ideia de selecionarmos estudantes que permaneceram o mesmo período em contato com a língua portuguesa em Pelotas parece ser crucial para o levantamento de dados como medida de controle extralinguístico, pois supõe-se que o tempo de contato com a língua estrangeira é fator relevante para a proximidade com a L2. A tentativa de padronizar o tempo

que os informantes permaneceram em imersão na L2 poderá nos resguardar da possibilidade de estarmos operando com sujeitos que apresentem níveis muito diferentes de proficiência, mesmo que as diferenças individuais, como extroversão, motivação, aptidão, tenham influência na aquisição da nova língua.

Durante a coleta dos dados, três dos quatro sujeitos da pesquisa já se encontravam de volta à Colômbia, o que implicou a realização da coleta de dados, por parte da pesquisadora, na cidade de Bogotá, no início do mês de abril de 2010, sendo que o último foi entrevistado em Pelotas. Os instrumentos para a coleta de dados foram aplicados pela própria pesquisadora, fixando-se dias diferentes para o instrumento referente a cada língua, para preservar a especificidade dos dois sistemas linguísticos analisados: o Português e o Espanhol.

### **3.1.2 O teste de proficiência em Português Brasileiro**

O período de contato dos informantes com a língua portuguesa tornou-se uma variável de controle, pois não encontramos no Brasil um teste de proficiência objetivo que pudesse ser aplicado aos estudantes a título de nivelamento. O Celpe-Bras – único certificado brasileiro de proficiência em Português como língua estrangeira, reconhecido oficialmente – é considerado um exame de desempenho e não de conhecimento sobre a língua. Em contato com profissionais ligados à aplicação e desenvolvimento do teste, deparamo-nos com uma discussão oportuna quando se objetiva nivelar estudantes adquirindo LE: o que queremos dizer com nível de proficiência?

[...] É um exame de uso de linguagem e não de conhecimento sobre a língua, testado em questões de múltipla escolha. As tarefas envolvem compreensão e produção de texto de forma integrada – parte coletiva – e uma entrevista – parte individual. [...] os critérios de avaliação são de uso da língua portuguesa e orientados, portanto, pela interlocução configurada no texto, seu propósito e no uso de recursos linguísticos coerentes e adequados tendo em vista as demandas da tarefa (de Margarete Schlatter, coordenadora da aplicação do CELPE-BRAS na UFRGS, por email, no dia 20/09/2010).

A definição da professora Margarete Schlatter para o exame de proficiência na língua portuguesa, o Celpe-Bras, confirma que o teste preza pela verificação do nivelamento através do uso da língua portuguesa e não de conhecimento estrutural da língua. O Celpe-Bras

é utilizado basicamente pelas Universidades brasileiras e algumas empresas no Brasil e no exterior e, portanto, é voltado a atender a essas necessidades.

Segundo o Manual do Celpe-Bras<sup>10</sup>, as habilidades verificadas incluem: comunicar-se em situações diárias como ler e escrever, conversar ou participar de atividades do contexto escolar, esclarecer dúvidas com o professor, fazer provas, apresentar seminários etc. e também incluem atividades externas ao contexto escolar como, fazer compras, obter informações, reclamar, ir ao médico etc. Não é objetivo desse exame analisar a aprendizagem em um determinado curso, mas examinar seu aprendizado na língua brasileira, independente de quando e como o aprendiz adquiriu a língua.

Diferente de outros testes de proficiência internacionais como o *TOEFL (Test of English as Foreign Language)* para o inglês e o *DELE (Diploma de Español como Lengua Extranjera)* para o Espanhol, que avaliam em separado a leitura, escrita, fala e compreensão auditiva, o Celpe-Bras é de natureza comunicativa, não integrando questões explícitas de gramática ou vocabulário, de modo que as habilidades sejam consideradas quando da elaboração de um texto oral ou escrito na avaliação do examinando.

Contudo, sendo o Celpe-Bras um teste de tarefas, ou seja, não apresentando perguntas objetivas, mas sim atividades constituídas, por exemplo, por convites a que o candidato passe um e-mail a um amigo ou atenda o telefone e conte uma reportagem a que assistiu sobre um tema atual do Brasil, a sua aplicação aos informantes da presente pesquisa exigiria, além de uma comissão de avaliação, um treinamento por parte da pesquisadora, o que seria inviável e desnecessário diante do tempo para concluir essa pesquisa, além de dispensável, diante dos objetivos do estudo.

Apesar de não pretendermos, nesta seção, ter assumido e esgotado uma discussão sobre testes de língua estrangeira como nivelamento, consideramos importante e necessária essa reflexão sobre avaliação de proficiência, até mesmo para que possamos justificar a não utilização de um teste de proficiência em PB aos estudantes colombianos, que nos levou a apenas controlar o tempo de imersão na língua portuguesa no Brasil.

### 3.2 O INSTRUMENTO DE COLETA DE DADOS

Para esta pesquisa, optou-se pelo uso de instrumentos de leitura de frases, um em Português e outro em Espanhol. Trata-se de testes de leitura em voz alta, que apresentam

---

<sup>10</sup> Disponível em: <<http://www.inep.gov.br/celpebras/default.asp>>.

palavras com plosiva em coda medial, as quais pertencem ao léxico das duas línguas: *enigma, dogma, cacto, convicto, estricnina, tecnicismo, admirador, admissão, rapto, corrupto, apneia, hipnose, obtuso, obtenção, abductor, abdome, etnia, Vietnam, ritmo, algoritmo, magno, signo*.

Tanto na versão em Português como na versão em Espanhol, as palavras foram apresentadas em frases-veículo. A escolha de palavras cognatas para comporem os instrumentos nas duas línguas teve a preocupação de preservar não apenas as mesmas consoantes plosivas na posição de coda medial, mas também os mesmos segmentos na posição de *onset* da sílaba seguinte, obtendo-se, assim, sequências idênticas nos dois sistemas linguísticos.

Um primeiro recorte para que chegássemos às sequências eleitas em palavras cognatas, no Português e no Espanhol, foi a escolha de palavras de no máximo quatro sílabas, de apenas duas classes morfológicas: substantivos e adjetivos, e com estrutura morfológica sem prefixo, dando-se preferência a palavras primitivas. Outra preocupação inicial foi a de se evitarem sequências de plosiva em coda medial em palavras não usuais ou pouco conhecidas, o que foi feito considerando-se o nível e o meio acadêmico de convivência dos estudantes participantes da pesquisa. Procurou-se, ainda, manter, para os exemplos nas duas línguas, a mesma relação da posição da sequência de consoantes com referência ao acento primário do item lexical.

Determinadas as variáveis linguísticas referentes ao número de sílabas das palavras a serem testadas, à classe gramatical, à estrutura morfológica, bem como à posição da sequência de consoantes em relação ao acento primário do item lexical, o próximo passo foi a seleção das sequências heterossilábicas que seriam utilizadas no instrumento de coleta de dados, levando-se em conta todas as observações anteriores, o que levamos a efeito com a utilização do Dicionário Aurélio Eletrônico da Língua Portuguesa (AE) Século XXI, versão 3.0, utilizando-se o mecanismo de busca por ele disponibilizado (a busca manual para esse processo tornaria o processo excessivamente demorado e complexo).

Dentre todas as possibilidades, as sequências escolhidas como objeto da pesquisa são as arroladas a seguir em sua forma ortográfica e em sua forma fonológica:

	<b>Sequência</b>		<b>Nº palavras encontradas na busca eletrônica pelo AE</b>	<b>Contexto</b>
1)	“gm”	/gm/	579 palavras	Plosiva + Nasal
2)	“ct”	/kt/	1954 palavras	Plosiva - Nasal
3)	“cn”	/kn/	219 palavras	Plosiva + Nasal
4)	“dm”	/dm/	73 palavras	Plosiva + Nasal
5)	“pt”	/pt/	978 palavras	Plosiva - Nasal
6)	“pn”	/pn/	195 palavras	Plosiva + Nasal
7)	“bt”	/bt/	64 palavras	Plosiva - Nasal
8)	“bd”	/bd/	143 palavras	Plosiva - Nasal
9)	“tn”	/tn/	91 palavras	Plosiva + Nasal
10)	“tm”	/tm/	105 palavras	Plosiva + Nasal
11)	“gn”	/gn/	579 palavras	Plosiva + Nasal

**Quadro 1 - Sequências com plosiva em coda escolhidas para a pesquisa**

Foram excluídas as sequências a seguir apresentadas no Quadro 2, em virtude de sua baixa frequência ou inexistência no Português:

	<b>Sequência</b>		<b>Nº palavras encontradas na busca eletrônica pelo AE</b>	<b>Contexto</b>
1)	“bn”	/bn/	31 palavras	Plosiva + Nasal
2)	“pm”	/pm/	0 palavras	Plosiva + Nasal
3)	“cm”	/km/	32 palavras	Plosiva + Nasal
4)	“bm”	/bm/	0 palavras	Plosiva + Nasal
5)	“dn”	/dn/	0 palavras	Plosiva + Nasal
6)	“pd”	/pd/	0 palavras	Plosiva - Nasal
7)	“cd”	/kd/	0 palavras	Plosiva - Nasal
8)	“gd”	/gd/	0 palavras	Plosiva - Nasal
9)	“gt”	/gt/	0 palavras	Plosiva - Nasal

**Quadro 2 - Sequências com plosiva em coda excluídas da pesquisa**

Também foram desconsideradas, conforme já foi referido, as palavras que contivessem prefixos como /sub/, como em *submarino*, considerando-se apenas palavras sem afixos, levando em conta a suspeita de Collischonn (2004) de que alguns prefixos possam estar atuando como palavra fonológica independente. Nesse caso, dependendo do prefixo,

estariamos sujeitos a índices de epêntese completamente diferentes. Do mesmo modo, também foram desconsideradas sequências de plosivas mais sibilantes, pela possibilidade, ainda referida em Collischonn (2004), da formação de uma africada na realização fonética ([ps], [ts], [ks]).

Após o levantamento das palavras que satisfizessem as condições requeridas, foram confrontadas as palavras selecionadas em Português com vocábulos registrados no Dicionário da Real Academia Espanhola (RAE), (22<sup>a</sup>. edição, 2001), para a confirmação de que os itens escolhidos para o Português realmente apresentavam palavras cognatas no Espanhol. Terminada essa fase, adaptaram-se frases simples (frases veículo) no Português e no Espanhol, para a constituição de dois instrumentos – um para cada língua – para a posterior etapa de leitura pelos sujeitos. As frases escolhidas foram as seguintes:

a) Frases em Português:

1 - Diga \_\_\_\_\_;

2 - Diga \_\_\_\_\_ novamente;

b) Frases em Espanhol:

1 - ¿Puedes decir \_\_\_\_\_?

2 - ¿Puedes decir \_\_\_\_\_ de nuevo?

Em trabalhos sobre aquisição de L2, é muito usual a coleta de dados feita a partir de situações espontâneas do uso da língua, inclusive porque é uma situação ideal, pois representa o uso “real” da língua. Entretanto, esse tipo de coleta consome mais tempo e pode se apresentar muito dificultosa já que os informantes podem não utilizar os elementos de interesse da pesquisa. Além do mais, através de instrumento constituído de frase com palavras-alvo, é possível controlar variáveis linguísticas tais como o ambiente fonológico – tal possibilidade, na presente pesquisa, permitiu que fossem estabelecidas generalizações acerca do fenômeno objeto de estudo.

### 3.3 PROCEDIMENTOS DE APLICAÇÃO DO INSTRUMENTO AOS INFORMANTES, APRENDIZES DO PB

Para a aplicação do Instrumento, as frases-veículo foram apresentadas em *slides*, no programa Microsoft Office PowerPoint 2003, com as vinte e duas palavras-alvo tanto em

posição medial, quanto em posição final de frase, randomizadas eletronicamente uma única vez.<sup>11</sup>

As vinte e duas palavras selecionadas foram repetidas três vezes, para que, em caso de problemas durante a gravação, a palavra não fosse perdida. As frases foram divididas em três arquivos com quarenta e quatro palavras, o que ocasionou uma duração média de três minutos por arquivo, para que fosse feita uma pausa proposital, que tornasse a aplicação do instrumento menos desgastante aos sujeitos.

A gravação foi feita no editor de áudio digital *Audacity* 1.3 beta, com a utilização do *headset* Multilaser PH53102 (microfone + fone de ouvido), para facilitar a análise acústica que foi realizada posteriormente, já que as gravações foram imediatamente salvas em arquivos de extensão *wave*. Enquanto o sujeito lia a apresentação dos slides em um computador, em outro eram feitas as gravações, para que não houvesse qualquer tipo de problema que pudesse representar um empecilho à qualidade da gravação, como atraso na voz ou peso de arquivos abertos ao mesmo tempo.

Durante a aplicação do Instrumento, o procedimento foi o de realizar o teste em dois dias diferentes – em um dia foi aplicado o teste em Espanhol e, no dia seguinte, o teste em Português –, para que não houvesse o uso dos dois sistemas em uma única entrevista, na tentativa de evitação de interferência. Para os três sujeitos que já haviam retornado à Colômbia, as gravações foram feitas no início de abril de 2010, na cidade de Bogotá: no primeiro dia, foi realizada a coleta de dados em língua Espanhola. Para o sujeito que havia retornado ao Brasil, a coleta foi realizada no mês seguinte, seguindo-se o mesmo encaminhamento.

Na aplicação do instrumento com as frases-veículo, cada um dos sujeitos da pesquisa foi rotulado pela primeira letra do nome, para que a identificação durante a fase de tratamento dos dados fosse facilitada. As apresentações do instrumento foram organizadas em três arquivos, contendo quarenta e quatro frases-veículo cada um – uma frase por *slide* - no Programa *Microsoft Office PowerPoint* 2003, em Português, e três arquivos contendo quarenta e quatro frases em Espanhol – também, uma frase por *slide* - separados, para que o total de cento e trinta e duas frases do instrumento não se tornasse exaustivo aos sujeitos durante a gravação no editor de áudio *Audacity*.

Antes do início das gravações, foi combinado que não seriam fornecidas maiores informações acerca do foco da pesquisa, até que os participantes houvessem terminado de

---

<sup>11</sup> Disponível em: <<http://www.random.org/lists/>>

gravar suas produções. Os informantes foram convidados a entrar em uma sala um a um, sem que pudessem escutar o outro. No dia seguinte, os mesmos procedimentos para a gravação em Português foram tomados, porém as instruções foram dadas em Português.

Após a aplicação da pesquisa, foram entregues dois formulários que complementam o estudo:

a) Formulário de Consentimento Informado, com as informações sobre a pesquisa, inclusive sobre a confidencialidade dos dados coletados;

b) Ficha de Informação do Participante, para confirmar informações importantes sobre os sujeitos da pesquisa, como tempo de permanência no Brasil e nível de escolaridade.

### 3.4 MÉTODO DE DESCRIÇÃO E ANÁLISE DOS DADOS

O *corpus*, depois de coletado, foi submetido a uma descrição detalhada. Descrevemos aqui os passos por que passamos durante o processo da leitura acústica, importante para a organização final dos dados.

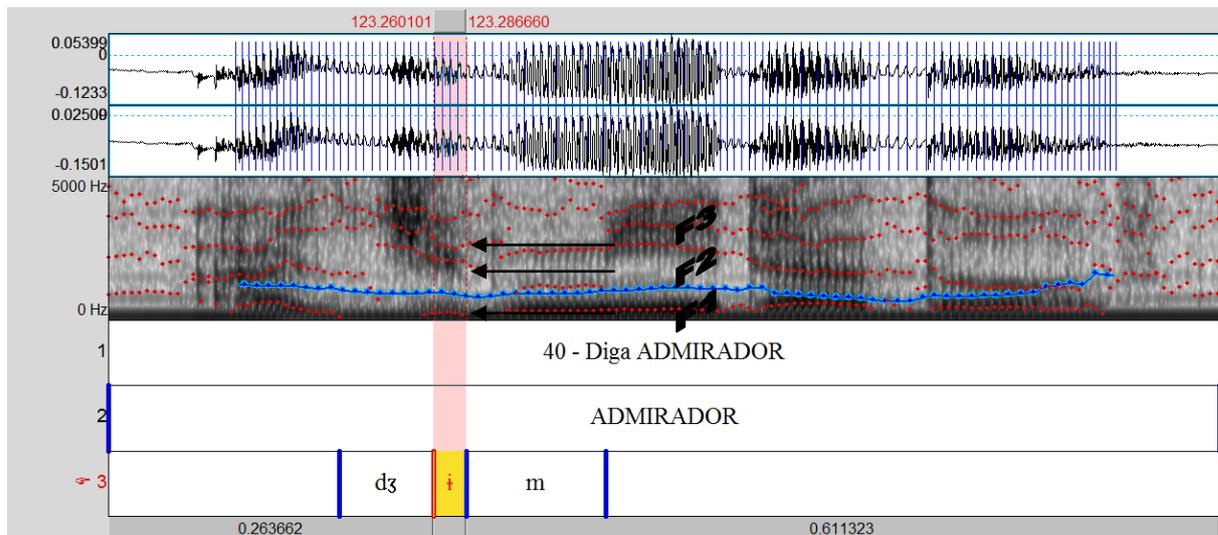
A análise acústica dos dados coletados foi realizada com o auxílio do programa PRAAT, que é um software comumente utilizado na análise de dados de fala. A partir dos arquivos de áudio, gravados em extensão *wave*, com as frases repetidas pelos sujeitos da pesquisa, o PRAAT vai gerar formas de onda e espectrogramas dos sons, capazes de fazer com que possamos identificar cada segmento de uma palavra.

Segundo Ladefoged (1996), podemos verificar, no espectrograma, os picos das ondas de uma vogal, que correspondem à frequência básica de vibração do ar no trato vocal. A esse modo natural de vibração denominamos de formantes, que estão intimamente ligados ao tamanho do trato vocal e à posição dos articuladores do trato vocal. São infinitos os números de formantes de um som, porém, para a análise de vogais, os três primeiros são suficientes, mas, principalmente, os dois primeiros são fundamentais.

Especificamente para a análise da vogal alta [i] - foco do nosso estudo, já que se tem mostrado como a vogal epentética por excelência no PB - devemos encontrar uma frequência do Formante 1 (F1) relativamente baixa, já que o primeiro formante varia inversamente à altura do dorso da língua quando da sua produção. Já o Formante 2 (F2), que está relacionado com a anterioridade/posterioridade da língua, deve aparecer alto para a vogal epentética [i].

As vogais podem também ser descritas através da sua disposição no espectro, localização dos picos e vales das ondas e medida da duração; trabalhos atuais têm considerado os formantes para a caracterização de vogais, pois, entre outros aspectos, é fácil sua identificação na análise acústica de dados, bem como a identificação de suas trajetórias relacionadas aos segmentos adjacentes.

A figura abaixo é uma reprodução da imagem do PRAAT onde estão assinalados os formantes da vogal epentética, em trecho selecionado de uma das sentenças produzidas por um dos sujeitos da pesquisa:



**Figura 9 - Produção da palavra “admirador” pelo sujeito “S” na frase “Diga admirador”**

O espectrograma identificado pela figura 09, com a produção da palavra “admirador”, mostra como se deu a visualização da vogal epentética na produção do sujeito “S”, no instrumento em Português Brasileiro. Os formantes identificados mostram um F1 baixo, já que se relaciona ao inverso da altura do dorso da língua, e F2 e F3 altos, conforme se esperaria para a vogal [i].

O lugar utilizado para as gravações das produções dos estudantes, sujeitos da pesquisa, foi uma sala silenciosa – já que não tivemos a possibilidade da utilização de um local isolado acusticamente, o que seria ideal. Apesar da facilidade da visualização da vogal epentética no espectrograma do programa PRAAT, através da observação dos formantes, sabemos que qualquer barulho captado no local das gravações pode escurecer as imagens no espectrograma, o que pode ser fator de distração na hora da definição do que é ruído e o que é vibração de cordas vocais.

O fator levantado acima é especialmente relevante pelo fato de a análise deste estudo focalizar uma vogal (a vogal epentética). Isso ocorre porque, quando o foco são

consoantes, o tratamento segue rumo diverso, já que, diferentemente das vogais, as consoantes, de maneira geral, são descritas de outra maneira, inclusive porque são acusticamente diversas entre si: enquanto algumas são produzidas com estreitamento do trato vocal, outras envolvem a cavidade nasal e, ainda, no caso das plosivas, há a produção com o bloqueio total à passagem do ar.

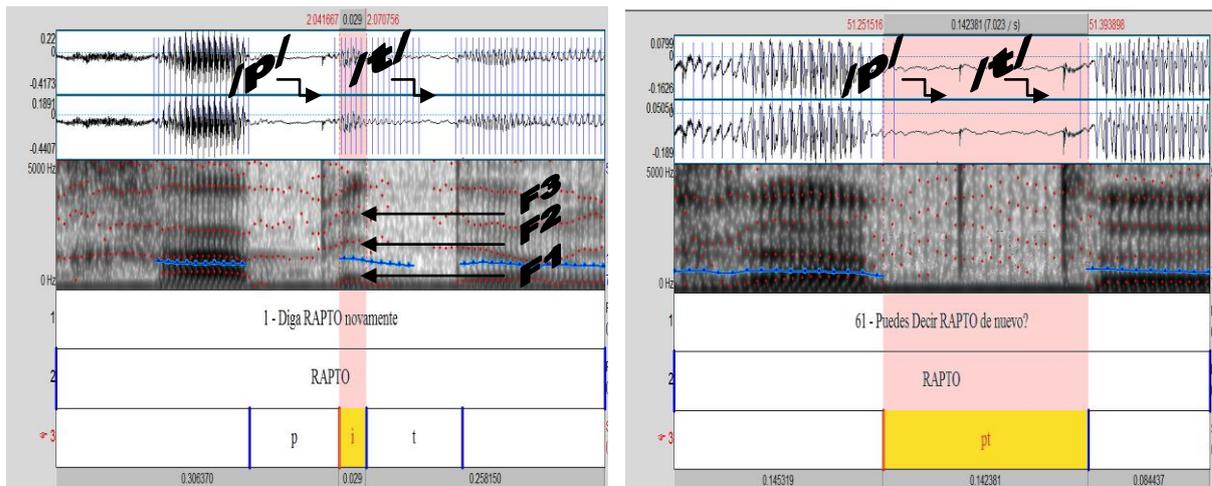
Segundo Kent & Read (1992), as características acústicas das plosivas dependem das suas posições na sílaba. Por exemplo, a oclusiva em início de sílaba deverá, primeiro, realizar o fechamento do trato vocal (*stop gap* ou período de clusura), depois a soltura (*burst* ou explosão) e, por fim, a transição dos formantes. Porém, quando em fim de sílaba, primeiro haverá a transição dos formantes, para ser feito o fechamento e a soltura ou não-soltura.

As consoantes plosivas que vamos analisar em coda medial de palavras cognatas do Português e Espanhol são: [p, b] - bilabiais, [t, d] - alveolares, [k, g] - velares. Porém, como a sílaba é a categoria mais baixa na hierarquia prosódica e está sob o domínio da palavra fonológica (BISOL, 2005), procuramos atentar, também, ao contexto que envolve a consoante plosiva em foco, como forma de tentar explicar os fenômenos que podem estar interferindo na presença ou não da epêntese.

Os recursos do programa PRAAT permitem notar no traçado de cada sinal os correlatos acústicos dos gestos articulatórios que interessam ao exame dos nossos dados. A presença da barra de sonoridade, no espectro acústico, por exemplo, vai indicar a vibração de cordas vocais; por sua vez, a ocorrência da barra de explosão vai se apresentar como uma das características básicas de oclusivas e africadas.

Uma evidência importante para as plosivas em posição de coda é a de que, nesta posição da sílaba, esses segmentos parecem sofrer um enfraquecimento. Logo, procuramos tratá-las de forma diversa à das plosivas em início de sílaba. Como, para este trabalho, focamos as plosivas em coda de sílabas CVC como em “**rap**.to” e VC, “vi.**et**.nam”, observamos a soltura ou não dos articuladores, ou seja, a presença ou não da explosão da oclusiva, como forma de cuidar a sua ocorrência.

Conforme justificamos no início do trabalho, para o Espanhol, esperava-se a produção das plosivas em coda e nas produções, para o Português, a ocorrência de epêntese mesmo que de forma variável. Procuramos, com os espectrogramas seguintes, demonstrar essa situação, demarcando as explosões das consoantes [p] e [t] da palavra “rpto”, sendo pronunciada nas duas línguas.



**Figura 10 - Produção da palavra “rapto” pelo sujeito “D” nas frases “Diga rapto novamente” e “Puedes decir rapto de nuevo”**

Na Figura 10, o espectrograma da esquerda, representa a produção da palavra “rapto” em uma frase do Português e o espectrograma da direita, representa a mesma palavra, porém, sendo produzida em uma frase no Espanhol. Procuramos demonstrar, que na produção da palavra “rapto” na frase em Português, temos a ocorrência da epêntese evidenciada com a posição dos formantes (F1, F2 e F3) e com o aparecimento de uma onda sonora após a explosão do /p/. Na frase em Espanhol, diferentemente do espectrograma da produção no PB, não vemos epêntese, mas uma closura, ou espaço em silêncio, após a explosão da plosiva /p/ e antecedendo a plosiva /t/.

Na análise das produções, procuramos extrair algumas características que consideramos relevantes descrever, no que se refere aos segmentos eleitos para a pesquisa. Ainda que nosso foco seja a presença ou não da epêntese após a plosiva em coda, observamos, também se houve apagamento dessa plosiva, se ela apresenta explosão ou soltura do ar e se tal soltura se mostra curta ou longa.

Durante o processo de análise, tivemos situações de dúvidas na leitura de alguns dados. Para esses casos, houve a participação de um segundo juiz, por vezes de até três juízes, para que pudéssemos fazer a melhor escolha. Superadas as etapas listadas acima, com a classificação acústica dos dados pronta, os padrões empíricos foram organizados e tratados estatisticamente, para posterior submissão à análise à luz da Gramática Harmônica e da Teoria da Otimidade Estocástica, através de simulação computacional, ainda no *software* PRAAT.

Nas próximas seções, portanto, apresentamos uma descrição dos *outputs* encontrados nas tentativas de produção das seqüências estudadas, para que possamos promover uma análise, via Teoria da Otimidade Estocástica e Gramática Harmônica, de

padrões de saída que acarretam a modificação do padrão silábico da L2 em função do sistema hierárquico interlinguístico do aprendiz.

## 4 DESCRIÇÃO DOS DADOS

Com os dados de *output* de aquisição do Português Brasileiro (L2) obtidos junto aos quatro informantes colombianos que participaram da pesquisa, pretendemos, neste momento, demonstrar os padrões de produções encontrados nas sequências heterossilábicas de segmentos plosivos selecionadas [g.m, k.t, k.n, d.m, p.t, p.n, b.t, b.d, t.n, t.m e g.n].

O intuito deste capítulo é atender ao objetivo de verificar se a epêntese vocálica que ocorre no uso do PB por falantes brasileiros, após segmentos plosivos em codas mediais de palavras, vai ocorrer também na produção dos sujeitos da pesquisa, estudantes colombianos, enquanto adquirindo português. Também discorreremos sobre as possíveis variáveis linguísticas que podem estar contribuindo para os tipos de resposta encontrados nas produções dos estudantes.

Descrevemos, para tanto, as realizações fonéticas nas tentativas de produção das onze sequências formadas de segmentos plosivos em coda medial de palavra, eleitas para o estudo, com vistas a apontar os *outputs* encontrados nas produções dos colombianos enquanto produziam as palavras cognatas em frases, no Português Brasileiro.

As realizações fonéticas dessas consoantes, na situação em análise, se apresentam, nas diferentes línguas, de forma diversa, conforme vimos no capítulo 3. No Espanhol, teremos as consoantes oclusivas articuladas em posição de coda silábica, já que essa constitui uma situação possível na fonologia do Espanhol; assim, teoricamente, não haveria motivação para que falantes nativos dessa língua se utilizem de alternativas como epêntese ou apagamento, em casos de presença de plosiva em coda.

Por outro lado, no PB, é utilizada a alternativa da epêntese, já que plosiva em coda não constitui estrutura silábica licenciada. Assim, uma palavra como “rap.to”, com oclusiva labial em coda, seria transformada em “ra.p[i].to”, na qual uma sílaba complexa, com coda não licenciada pelo sistema, seria transformada em duas sílabas abertas CV. Teóricos como Jakobson (1972) elegeram a sílaba CV como o padrão silábico ótimo, universal; assim, seguindo esse raciocínio e reconhecendo a relevância de tendências universais no tocante à marcação, nossa hipótese inicial era a de que um falante nativo de Espanhol, com uma situação de coda marcada - como em “sig.no” - na LE faria epêntese, assim como os falantes nativos de Português.

Para fins de organização e facilitação da visualização dos dados, dividimos este capítulo em duas partes, uma de dados dos *outputs* no Espanhol e outra no Português. Por sua

vez, cada parte está subdividida de acordo com o ponto de articulação dos segmentos plosivos em coda analisados: plosivas bilabiais [p, b], plosivas alveolares [t, d] e velares [k, g], além de subdivisões de acordo com a sonoridade.

Para a disposição dos dados nas tabelas, não foi feita a separação das produções das sequências em função da frase-veículo, pois não encontramos diferenças, quanto a essa variável, nem de padrões acústicos, nem de número de epêntese. Sendo assim, foi feita apenas uma tabela por sequência, para representar as ocorrências encontradas.

Para todas as sequências analisadas, elencamos algumas possibilidades de *output*, como (a) a ocorrência de epêntese após a consoante plosiva em coda, (b) o apagamento da plosiva em coda e (c) a soltura ou não soltura da consoante. Assim, enumeramos em “**Outros**” os casos diversos aos sugeridos para as possibilidades dadas de *output* e em “**Erro**”, qualquer tipo de problema durante a gravação que inutilizasse aquela sequência.

#### 4.1 CONSIDERAÇÕES FUNDAMENTADORAS PARA AS DISCUSSÕES DOS DADOS NO PORTUGUÊS BRASILEIRO E NO ESPANHOL

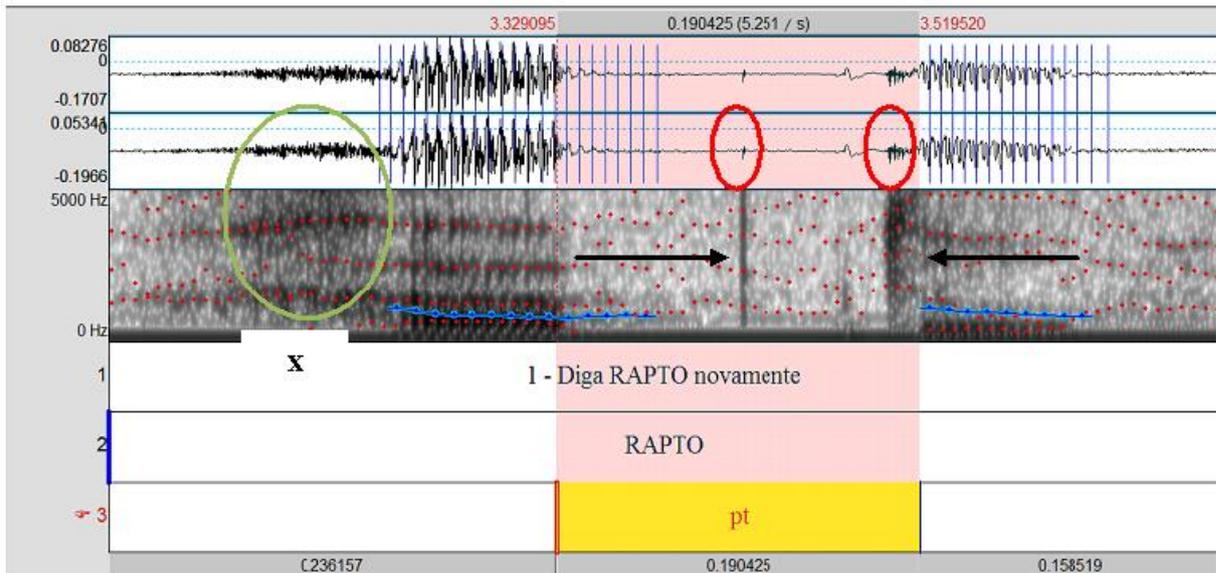
##### 4.1.1 A soltura e não soltura da plosiva em coda

Conforme veremos nos dados a seguir, tanto nas produções em Português quanto nas produções em Espanhol, temos uma lenição em grande parte das produções da plosiva em coda. Consideramos relevante demonstrar o reconhecimento da possibilidade de ocorrência de não soltura da plosiva em coda, o que aconteceu, por exemplo, com a palavra [ab<sup>h</sup>dutor].

Consideramos como ocorrência de soltura ou explosão o caso das plosivas que apresentaram dois parâmetros acústicos: a presença da barra de sonoridade e o *burst*<sup>12</sup> -, sem a intenção de qualificar tais características. Com o espectrograma seguinte, pretendemos demonstrar essa ocorrência:

---

<sup>12</sup> O “*burst*” é a explosão de ar obtida com a soltura dos articuladores, durante a realização da oclusiva.



**Figura 11 - Produção da palavra “rapto” na frase em Português “Diga rapto novamente” pelo sujeito “S”**

As flechas indicadas no espectrograma da sequência [p.t] estão apontando as barras de sonoridade do [p] e do [t], da esquerda para a direita, respectivamente, e o círculo em vermelho marca o *burst*. A situação demonstrada na figura 11 capta inclusive a velarização do [x] realizada na produção do estudante, possível de perceber pela duração da consoante, pela forma de onda - que se assemelha a uma representação de ruído característico das fricativas -, e pela presença de barras verticais, que se devem à vibração da parte posterior da língua. Este fato da velarização da líquida, sinalizado com o círculo verde, é significativo, já que estamos tratando de um informante cuja língua materna é o Espanhol, língua em que a realização fonética do *r-forte* é alveolar.

Nas produções na língua materna dos estudantes, assim como na língua em aquisição, a soltura da plosiva em coda se mostrou variável em todas as sequências, com exceção da sequência [d.m], que revelou um valor de cem por cento de produção sem soltura audível. Por outro lado, tal sequência seguiu o padrão mostrado pelas demais consoantes plosivas em coda seguidas de nasal, em se considerando o vozeamento, conforme discussão apresentada na seção 4.1.2. Por sua vez, as plosivas não vozeadas seguidas de plosiva também não vozeada, como nas sequências [p.t] e [k.t], ao contrário, vão apresentar mais situações de soltura audível e menos casos de soltura não audível.

Como forma de aproveitar a leitura fonética realizada para todas as sequências selecionadas, optamos por descrever, além do fenômeno da epêntese, outros fenômenos que chamaram a atenção do pesquisador; não temos, no entanto, pretensão de formalizar através

da OT e da HG todos estes padrões fonéticos, mas apenas relatar os fatos para que, de alguma forma, possam servir a trabalhos futuros.

#### 4.1.2 A Sequência plosiva + nasal

Antes de relatarmos os dados relativos à sequência de plosiva em coda e nasal em *onset* da próxima sílaba (“g.m”, “c.n”, “d.m”, “p.n”, “t.n”, “t.m” e “g.n”), devemos discutir uma situação importante verificada nas ocorrências nesse contexto. Para todos os casos tivemos porcentagem considerável de vozeamento da plosiva surda que está em posição de coda silábica. Esse fato não é surpreendente, inclusive faz sentido quando falamos de contexto adjacente com consoante nasal.

Segundo Clements (1990), as sequências heterossilábicas oclusiva-nasal sofrem uma pressão considerável para serem modificadas, já que a coda é menos sonora do que a consoante do ataque seguinte, violando a estrutura silábica considerada ideal. Conforme foi apresentado no Referencial Teórico, na seção 2.1.2, a diferença entre a oclusiva e a nasal na escala de sonoridade é de um ponto. Entretanto, se considerarmos ainda subdividir a oclusiva em surda e sonora, menor será a diferença entre a nasal e a plosiva sonora, com relação à surda. Essa pode ser a resposta para o vozeamento das oclusivas surdas em coda, assim como é a resposta de Clements (1990) para a inserção da epêntese nesse caso, ou seja, a resposta está vinculada à escala de sonoridade e à lei do contato silábico e sua influência na constituição de estruturas silábicas ótimas.

Foneticamente, esse vozeamento da plosiva surda em coda remeteu a uma situação relevante na análise acústica dos dados. A forma de onda apresentada no espectrograma nas palavras selecionadas para compor este estudo, constituídas por uma sequência de duas plosivas vozeadas, como em “**ab.du.tor**”, fez com que a análise fosse dificultosa, pois fez parecer ao pesquisador que há uma situação de apagamento da plosiva em coda, já que a forma de onda das consoantes que formam essa sequência é muito semelhante.

Para tanto, a análise do contexto adjacente foi importante, assim como a observação dos formantes, da duração da nasal e da oclusão da plosiva. Nesse sentido, foi possível nos acercarmos de excluir a hipótese de apagamento, e consideramos que houve uma não soltura da plosiva em coda, com vozeamento da consoante em coda, no caso das plosivas surdas.

## 4.2 DESCRIÇÃO DOS DADOS DE PRODUÇÃO NO ESPANHOL COLOMBIANO (L1)

### 4.2.1 Plosivas bilabiais em coda: [p] e [b]

Para a sequência /bd/, exposta na tabela 1, apresentamos quatro possibilidades de *outputs* para as produções no Espanhol: [b.d] (*output* fiel ao *input*), [bi.d] (*output* com epêntese), [ḅ.d] (*output* sem a explosão da plosiva em coda) e [ɸ.d] (*output* com apagamento do /b/). Para a sequência /bt/, exposta na tabela 2, oferecemos, também, quatro possibilidades: [b.t] (*output* fiel ao *input*), [bi.t] (*output* com epêntese), [ḅ.t] (*output* sem a explosão da plosiva em coda) e [ɸ.t] (*output* com apagamento do [b]). Nossa previsão para essas sequências é a de que tenhamos cem por cento de produção da plosiva em coda, sem dificuldades aos aprendizes nas produções em Espanhol.

**Tabela 1 - *Outputs* das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /bd/ nas palavras “abdutor” e “abdome”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
b.d	83% (10/12)	50% (6/12)	0% (0/12)	25% (3/12)	40% (19/48)
bi.d	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
ḅ.d	17% (2/12)	25% (3/12)	100% (12/12)	67% (8/12)	52% (25/48)
ɸ.d	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Erro	0 (0/12)	25% (3/12)	0 (0/12)	8% (1/12)	8% (4/48)

**Tabela 2 - Outputs das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /bt/ nas palavras “obtuso” e “obtención”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
b.t	75% (9/12)	83% (10/12)	92% (11/12)	83% (10/12)	83% (40/48)
bi.t	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
ḅ.t	25% (3/12)	17% (2/12)	8% (1/12)	17% (2/12)	17% (8/48)
ɸ.t	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Erro	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)

Conforme esperávamos, zero foi o número de epêntese apresentado nas tabelas referentes à produção dos sujeitos colombianos nas sequências /bd/ e /bt/, já que o Espanhol - língua materna dos sujeitos - permite oclusivas em posição final de sílaba. Porém, no que diz respeito às características fonéticas da plosiva [b] em coda, outras observações podem ser feitas.

O [b] em coda, de maneira variável, se apresenta ora com explosão audível, ora sem explosão para os quatro informantes e, mesmo quando apresenta explosão, é fraca, em relação a sua presença em posição de ataque. Situação interessante também foi o fato de, nessas sequências em análise, as plosivas sonoras se apresentarem com pouco vozeamento, mesmo quando sofriam explosão.

Encontramos ainda, nos dados referentes a /bt/ e /bd/, com um número expressivo de ocorrências, um pequeno desvozeamento da bilabial, nos casos em que tivemos a explosão da plosiva em coda e nos momentos em que não apresentou soltura de ar audível.

A seguir, ainda com as consoantes bilabiais em coda, apresentamos as produções referentes às sequências /pt/ e /pn/. Para a sequência /pt/, apresentamos quatro possibilidades de *outputs* para as produções no Espanhol: [p.t] (*output* fiel ao *input*), [pi.t] (*output* com epêntese), [p̣.t] (*output* sem a explosão da plosiva em coda) e [ɸ.t] (*output* com apagamento do [p]). Por sua vez, para a sequência /pn/, também, previmos quatro possibilidades: [p.n] (*output* fiel ao *input*), [pi.n] (*output* com epêntese), [p̣.n] (*output* sem a explosão da plosiva em coda) e [ɸ.n] (*output* com apagamento do [p])<sup>13</sup>.

<sup>13</sup> Outras possibilidades, como a soltura longa da consoante plosiva em coda/aspiração, foram observadas, porém, identificamos nas tabelas das sequências heterossilábicas que estamos estudando [b.d, b.t, p.t, p.n, g.m, g.n, k.t, k.n, d.m, t.m, t.n], apenas os fenômenos verificados nos dados de produção ou teoricamente possíveis.

**Tabela 3 - Outputs das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /pt/ nas palavras “rapto” e “corrupto”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
p.t	75% (9/12)	83% (10/12)	92% (11/12)	92% (11/12)	85% (41/48)
pi.t	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
p̣.t	25% (3/12)	17% (2/12)	8% (1/12)	8% (1/12)	15% (7/48)
ɸ.t	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Erro	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)

**Tabela 4 - Outputs das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /pn/ para as palavras “apneia” e “hipnose”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
p.n	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
pi.n	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
p̣.n	75% (9/12)	0 (0/12)	0% (0/12)	0 (0/12)	19% (9/48)
ɸ.n	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	25% (3/12)	100% (12/12)	100% (12/12)	100% (12/12)	81% (39/48)
Erro	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)

Verificamos para a sequência /pt/ - com o encontro de consoantes plosivas surdas - assim como ocorreu na sequência /bt/, índices maiores de soltura de ar audível da consoante em coda, em relação às sequências /bd/ e /pn/. Porém, em nenhuma dessas sequências encontramos uma soltura exagerada ou qualquer outro tipo de correção de estrutura, confirmando a literatura sobre a possibilidade do aparecimento da consoante oclusiva em coda no Espanhol.

Além disso, ao contrário das sequências /bd/ e /bt/, encontramos para a sequência /pn/ um vozeamento da plosiva em coda, que enumeramos em ”Outros”, visto o número significativo de sua ocorrência. Tanto na palavra “apneia” quanto em “hipnose”, a produção realizada por todos os estudantes colombianos resultou em [am<sup>h</sup>neja] ou [ab<sup>h</sup>neja] e [ib<sup>h</sup>nozi] ou [im<sup>h</sup>nozi], ou seja, com o vozeamento da plosiva em coda diante de nasal no *onset* seguinte

ou com a assimilação do traço [+ nasal], pela plosiva da coda, com a preservação do ponto de articulação original dessa plosiva.

Conforme já exposto em seção anterior, esse tipo de encontro de oclusiva em coda seguida de nasal em *onset* de próxima sílaba está sujeito à alteração, se considerarmos a escala de sonoridade. No Português Brasileiro, encontramos a epêntese como forma de correção ao molde silábico, como já vimos no trabalho de Collischonn (2002). No Espanhol, estamos verificando a ocorrência de vozeamento da plosiva surda diante da força da nasal, podendo haver também assimilação do traço [+ nasal].

#### 4.2.2 Plosivas velares em coda: [k] e [g]

No seguimento da descrição dos dados, tanto para [g.m] quanto para [g.n], próximas sequências apresentadas, temos quatro possibilidades de *outputs* para as produções no Espanhol: [g.m] e [g.n] (*output* fiel ao *input*), [gi.m] e [gi.n] (*output* com epêntese), [g̃.m] e [g̃.n] (*output* sem a explosão da plosiva em coda) e [Ɂ.m] e [Ɂ.n] (*outputs* com apagamento do [g]).

Conforme já verificamos nos dados anteriores, as sequências constituídas de nasal parecem apresentar, todas, uma lenição da plosiva em coda. Até o momento, podemos ver que a maior parte das produções se apresenta sem a soltura de ar da plosiva em coda. Esta situação pode ser caracterizada como interessante, visto que será observada até o fim da análise.

**Tabela 5 - *Outputs* das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /gm/ para as palavras “enigma” e “dogma”**

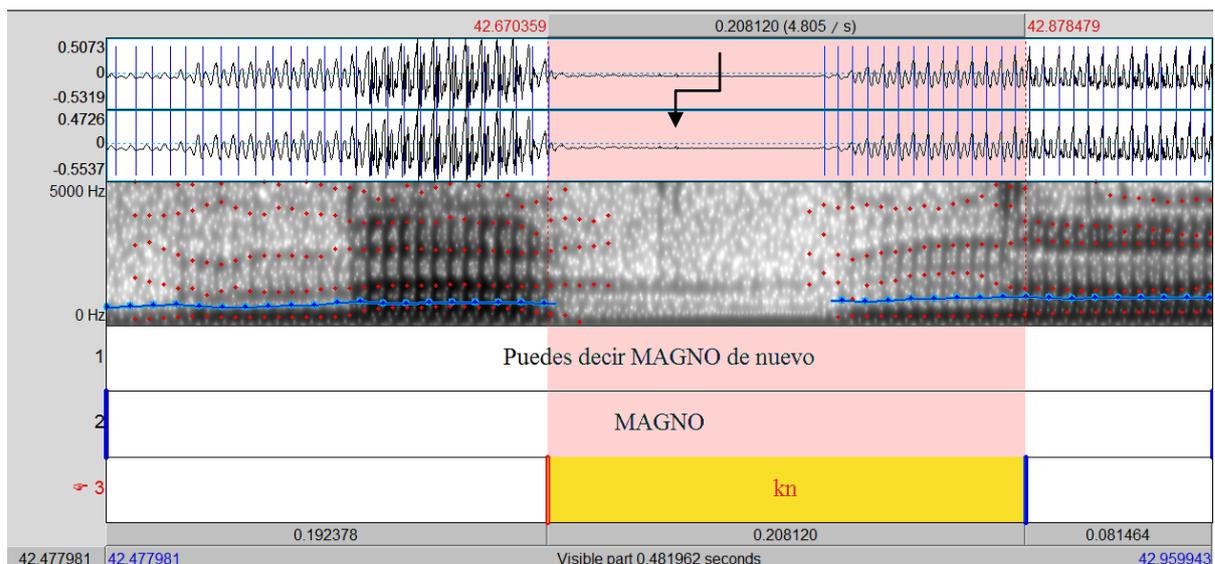
	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
g.m	0% (0/12)	17% (2/12)	8% (1/12)	8% (1/12)	8% (4/48)
gi.m	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
g̃.m	8% (1/12)	75% (9/12)	75% (9/12)	58% (7/12)	55% (26/48)
Ɂ.m	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	92% (11/12)	0 (0/12)	17% (2/12)	34% (4/12)	35% (17/48)
Erro	0 (0/12)	8% (1/12)	0 (0/12)	0 (0/12)	2% (1/48)

**Tabela 6 - Outputs das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /gn/ para as palavras “magno” e “signo”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
g.n	8% (1/12)	0 (0/12)	42% (5/12)	0 (0/12)	13% (6/48)
gi.n	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
g̃.n	17% (2/12)	83% (10/12)	58% (7/12)	75% (9/12)	58% (28/48)
g̃.n	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	75% (9/12)	17% (2/12)	0 (0/12)	25% (3/12)	29% (14/48)
Erro	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)

Além da confirmação da expectativa de encontrarmos a maior parte da produção sem soltura de ar, outro aspecto deve ser relatado, já que apresenta uma ocorrência considerável nas duas tabelas acima. Para os dois casos /gn/ e /gm/ tivemos um desvozeamento da plosiva em coda, que está representado em “Outros”.

O espectrograma apresentado a seguir, tem a finalidade de demonstrar a nitidez do desvozeamento do /g/ na palavra “magno”. A seta indica exatamente a posição da explosão da plosiva, onde não temos forma de onda, mas um período de closura característico da oclusiva surda, com baixíssima energia dissipada.



**Figura 12 - Produção da palavra “magno” na frase em Espanhol “Puedes decir magno de nuevo” pelo sujeito “A”**

É clara a visualização do ensurdecimento da plosiva sonora pela realização de seu correspondente surdo [k]. Tal padrão foi apresentado não apenas por um sujeito, mas, de

forma variável, por todos eles, chegando a 25% em se tratando da sequência [g.m] e a uma porcentagem de 39% em se referindo à sequência /gn/.

Dando continuidade à descrição das sequências com plosivas velares, apresentamos agora os resultados relativos a /kt/ e /kn/, com quatro possibilidades de *outputs*, para cada uma das sequências para, ainda, produções no Espanhol: [k.t] (*output* fiel ao *input*), [ki.t] (*output* com epêntese), [ḳ.t] (*output* sem explosão da plosiva em coda) e [ɕ.t] (*output* com apagamento do [k]). E, para /kn/: [k.n] (*output* fiel ao *input*), [ki.n] (*output* com epêntese), [ḳ.n] (*output* sem a explosão da plosiva em coda) e [ɕ.n] (*output* com apagamento do [k]).

**Tabela 7 - *Outputs* das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /kt/ para as palavras “cacto” e “convicto”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
k.t	50% (6/12)	25% (3/12)	100% (12/12)	25% (3/12)	50% (24/48)
ki.t	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
ḳ.t	50% (6/12)	75% (9/12)	0 (0/12)	75% (9/12)	50% (24/48)
ɕ.t	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Erro	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)

**Tabela 8 - Outputs das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /kn/ para as palavras “estricnina” e “tecnicismo”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
k.n	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
ki.n	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
ḳ.n	25% (3/12)	33% (4/12)	0 (0/12)	17% (2/12)	19% (9/48)
œ.n	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	75% (9/12)	67% (8/12)	100% (12/12)	83% (10/12)	81% (39/48)
Erro	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)

Quanto à sequência [k.t], os resultados mostram 50% do total das produções da plosiva em coda com explosão e 50% sem explosão e, quanto à sequência [k.n], não foi surpresa, mais uma vez, a ocorrência de um vozeamento da plosiva em coda (substituição do [k] por um [g]) num contexto de nasal. Porém, nessas produções, mais de 80% das ocorrências apresentaram plosiva vozeada (representada em “Outros”), ou seja, os resultados registraram nesse caso apenas 19% de [k] em coda, e sem explosão.

De qualquer forma, não encontramos nenhum sinal de inserção de vogal, mas apenas a produção da plosiva em coda, mesmo que com a substituição da plosiva surda por sua correspondente sonora.

### 4.2.3 Plosivas alveolares em coda: [t] e [d]

**Tabela 9 - Outputs** das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /dm/ para as palavras “admisión” e “admirador”

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
d.m	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
dʒi.m	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
dʒ.m	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
d̃.m	100% (12/12)	100% (12/12)	100% (12/12)	100% (12/12)	100% (48/48)
ɬ.m	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Erro	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)

**Tabela 10 - Outputs** das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /tm/ para as palavras “ritmo” e “algoritmo”

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
t.m	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
tʃi.m	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
tʃ.m	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
t̃.m	8% (1/12)	0 (0/12)	0 (0/12)	0 (0/12)	2% (1/48)
ɬ.m	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	92% (11/12)	75% (9/12)	100% (12/12)	100% (12/12)	92% (44/48)
Erro	0 (0/12)	25% (3/12)	0 (0/12)	0 (0/12)	6% (3/48)

**Tabela 11 - Outputs das produções das palavras cognatas Português/Espanhol em frases do Espanhol para a sequência /tn/ para as palavras “etnia” e “vietnam”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
t.n	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
tʃi.n	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
tʃ.n	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
t̃.n	67% (8/12)	0 (0/12)	0 (0/12)	0 (0/12)	17% (8/48)
s.n	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	25% (3/12)	92% (11/12)	100% (12/12)	100% (12/12)	79% (38/48)
Erro	8% (1/12)	8% (1/12)	0 (0/12)	0 (0/12)	4% (2/48)

Encerramos esta seção discutindo as produções referentes às sequências /tn, tm, dm/. Começando por /dm/: zero foi o número de epêntese na realização fonética dessa sequência, e ainda podemos observar que 100% das produções foram realizadas sem soltura de ar, bem como todas as demais sequências que têm uma nasal em *onset* da próxima sílaba. Nem todas tiveram tal porcentagem sem soltura, mas os resultados apresentaram a maior parte das produções sem explosão, incluindo as sequências que vamos verificar a seguir, /tn/ e /tm/.

Para o Espanhol, vimos que as alveolares não são palatalizadas, mas produzidas como dentais mesmo diante da vogal /i/, exatamente o que observamos nos *outputs* nessas três últimas sequências. O interessante será verificar nas produções nas frases em Português, na próxima seção, a ocorrência de palatalização, o que demonstrará a percepção dos estudantes das diferenças entre as duas línguas.

Consideramos oportuno separar em “Outros”, nas produções de /tn/ e /tm/, as situações diversas às previstas nas sequências elencadas nas tabelas, principalmente porque renderam uma porcentagem bastante grande. Essas produções referem-se a *outputs* com vozeamento do /t/. Em todas as palavras, as produções dos aprendizes se apresentavam com um processo de substituição do [d] em lugar do [t], assim como já havíamos visto em se tratando das sequências anteriores de plosiva mais nasal, como [k.n] e [p.n].

A seguir, realizamos a análise das produções nas frases do PB, fazendo uma comparação com as ocorrências na língua materna dos aprendizes colombianos.

### 4.3 DESCRIÇÃO DOS DADOS DE PRODUÇÃO NO PORTUGUÊS BRASILEIRO (L2)

Passamos agora às produções dos aprendizes no Português Brasileiro. A análise realizada nesta seção é fundamental na verificação de vários questionamentos desta pesquisa, como, por exemplo, se os estudantes colombianos realizam epêntese nas palavras cognatas do Espanhol em frases no Português.

Considerando que, na aquisição de uma nova língua, a marcação é um fator preponderante nas dificuldades do aprendiz, esperávamos encontrar, nas produções dos sujeitos descritas nesta seção, padrão semelhante ao *input* por eles recebido, já que é menos marcada uma produção com epêntese, igual à forma o.[bi.t]u.so, do que o.[b.t]u.so, sem epêntese. Conforme os dados mostrados a seguir, vemos que, de fato, há a inserção de vogal nas produções, porém baixos serão os números nas realizações.

A organização das tabelas de produções respeitou a mesma ordem seguida na descrição das produções no Espanhol, começando pelas plosivas bilabiais, passando pelas velares e terminando nas alveolares em coda. Da mesma forma, utilizamos a notação “**Erro**” para separar os problemas de gravação e a categoria “**Outros**” para identificar respostas diferentes das elencadas na tabela.

#### 4.3.1 Plosivas bilabiais em coda: [p] e [b]

Para a sequência /bd/, exposta na tabela abaixo, apresentamos quatro possibilidades de *outputs* para as produções no Português: [b.d] (*output* sem epêntese), [bi.d] (*output* com epêntese), [ḅ.d] (*output* sem a explosão da plosiva em coda) e [ɸ.d] (*output* com apagamento do [b]). Para a sequência /bt/, oferecemos, também, quatro possibilidades: [b.t] (*output* sem epêntese), [bi.t] (*output* com epêntese), [ḅ.t] (*output* sem a explosão da plosiva em coda) e [ɸ.t] (*output* com apagamento do [b]). A previsão para essas sequências neste momento é a de que tenhamos alguma ocorrência de epêntese, assim como acontece no dialeto pelotense.

**Tabela 12 - Outputs das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /bd/ nas palavras “abductor” e “abdome”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
b.d	42% (5/12)	25% (3/12)	0 (0/12)	34% (4/12)	25% (12/48)
bi.d	0 (0/12)	0 (0/12)	8% (1/12)	8% (1/12)	4% (2/48)
ḅ.d	58% (7/12)	50% (6/12)	92% (11/12)	58% (7/12)	65% (31/48)
ɐ.d	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Erro	0 (0/12)	25% (3/12)	0 (0/12)	0 (0/12)	6% (3/48)

**Tabela 13 - Outputs das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /bt/ nas palavras “obtenção” e “obtusó”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
b.t	50% (6/12)	75% (9/12)	92% (11/12)	100% (12/12)	79% (38/48)
bi.t	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
ḅ.t	50% (6/12)	25% (3/12)	8% (1/12)	0 (0/12)	21% (10/48)
ɐ.t	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Erro	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)

Os dados mostraram, para a sequência /bd/, 4% de epêntese, que corresponde a duas inserções de vogal num universo de 48 produções. Tal índice constitui um número pequeno, mas muito importante no contexto em que estamos trabalhando, considerando os objetivos e as questões de pesquisa propostos. Podemos dizer, portanto, que os estudantes que produziram epêntese perceberam que há diferença entre a estrutura silábica do Português e Espanhol. Porém, a sequência seguinte /bt/ apresentou 100% de plosiva em coda, com a característica diferencial de que apresenta muito mais soltura da plosiva /b/ em coda, em se comparando com a sequência /bd/.

Ladefoged (1996) fala sobre a limitação das plosivas em coda, contando com essa possibilidade de não soltura do ar. Porém, o que observamos nas sequências /bd/ e /bt/ foi que, tanto para o Português como para o Espanhol, a não soltura é bem mais frequente quando

o segmento seguinte à plosiva em coda é sonoro, fato que procuramos destacar nos demais contextos analisados.

Ainda falando em plosivas bilabiais em coda, apresentamos, no seguimento, as sequências /pn/ e /pt/, que são utilizadas no nosso ensaio final nos modelos da Gramática Harmônica e Teoria da Otimidade.

**Tabela 14 - Outputs das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /pt/ nas palavras “rapto” e “corrupto”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
p.t	92% (11/12)	75% (9/12)	100% (12/12)	100% (12/12)	92% (44/48)
pi.t	0 (0/12)	8% (1/12)	0 (0/12)	0 (0/12)	2% (1/48)
p̣.t	8% (1/12)	17% (2/12)	0 (0/12)	0 (0/12)	6% (3/48)
ɤ.t	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Erro	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)

**Tabela 15 - Outputs das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /pn/ nas palavras “apneia” e “hipnose”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
p.n	8% (1/12)	17% (2/12)	0 (0/12)	8% (1/12)	8% (4/48)
pi.n	0 (0/12)	0 (0/12)	25% (3/12)	59% (7/12)	21% (10/48)
p̣.n	8% (1/12)	0 (0/12)	0 (0/12)	0 (0/12)	2% (1/48)
ɤ.n	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	84% (10/12)	75% (9/12)	75% (9/12)	33% (4/12)	67% (32/48)
Erro	0 (0/12)	8% (1/12)	0 (0/12)	0 (0/12)	2% (1/48)

Para a sequência /pt/, evidenciamos quatro possibilidades de *outputs* para as produções no Português: [p.t] (*output* sem epêntese), [pi.t] (*output* com epêntese), [p̣.t] (*output* sem a explosão da plosiva em coda) e [ɤ.t] (*output* com apagamento do [p]).

Para a sequência /pn/, oferecemos, também, quatro possibilidades: [p.n] (*output* sem epêntese), [pi.n] (*output* com epêntese), [p̣.n] (*output* sem a explosão da plosiva em coda), e [ɤ.n] (*output* com apagamento do [p]).

As sequências que vemos agora implicam um contato de dois segmentos surdos /pt/ e de um segmento surdo e outro sonoro /pn/. Na descrição anterior, vimos que a sequência que contava com uma consoante sonora no *onset* da próxima sílaba (/bd/) apresentou epêntese, enquanto que, na sequência de consoantes surdas, não foi aplicado esse processo. Entre /pn/ e /pt/ veremos caso semelhante. Enquanto para /pn/ houve 21% de epêntese, para /pt/ ocorreu apenas 2% de epêntese. Tal resultado pode estar evidenciando a influência do segmento sonoro na inserção do segmento epentético, diferenciando esse tipo de sequência daquelas em que há contato entre dois segmentos surdos.

Além disso, outra observação com relação à diferença entre /pn/ e /pt/ ocorre em se considerando a explosão do /p/: enquanto em /pt/ houve quase 100% da plosiva com explosão no PB e no EC, no /pn/ esses valores chegam a zero no Espanhol. Os casos relatados em “Outros” na sequência /pn/ referem-se ao vozeamento do /p/ como, por exemplo, [abneja]. Não tivemos epêntese nessa produção, mas continuamos tendo essa influência do segmento sonoro no contato silábico. Nesses casos, não encontramos explosão, seguindo uma lógica de maior enfraquecimento da plosiva nos contatos com sequências sonoras.

#### **4.3.2 Plosivas velares em coda: [k] e [g]**

Em se considerando as sequências com consoantes velares em coda, essas foram trabalhadas em separado, primeiramente, com as sequências /gm/ e /gn/, para depois entrarmos nas sequências /kt/ e /kn/, destacando que essas duas últimas, juntamente com /pt/ e /pn/, serão consideradas na análise final na OT e HG, a ser realizada no capítulo 5.

**Tabela 16 - Outputs das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /gm/ nas palavras “enigma” e “dogma”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
g.m	17% (2/12)	0 (0/12)	17% (2/12)	67% (8/12)	25% (12/48)
gi.m	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
g̃.m	75% (9/12)	25% (3/12)	75% (9/12)	33% (4/12)	52% (25/48)
ɶ.m	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	8% (1/12)	25% (3/12)	8% (1/12)	0 (0/12)	10% (5/48)
Erro	0 (0/12)	50% (6/12)	0 (0/12)	0 (0/12)	13% (6/48)

**Tabela 17 - Outputs das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /gn/ nas palavras “magno” e “signo”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
g.n	50% (6/12)	34% (4/12)	34% (4/12)	42% (5/12)	39% (19/48)
gi.n	0 (0/12)	0 (0/12)	50% (6/12)	0 (0/12)	13% (6/48)
g̃.n	42% (5/12)	58% (7/12)	16% (2/12)	50% (6/12)	42% (20/48)
ɶ.n	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	8% (1/12)	8% (1/12)	0 (0/12)	0 (0/12)	4% (2/48)
Erro	0 (0/12)	0 (0/12)	0 (0/12)	8% (1/12)	2% (1/48)

Para a sequência /gm/, quatro são as opções de *outputs* para as produções no Português: [g.m] (*output* sem epêntese), [gi.m] (*output* com epêntese), [g̃.m] (*output* sem a explosão da plosiva em coda) e [ɶ.m] (*output* com apagamento do [g]). Para a sequência /g.n/, as mesmas caracterizações são dadas: [g.n] (*output* sem epêntese), [gi.n] (*output* com epêntese), [g̃.n] (*output* sem a explosão da plosiva em coda), e [ɶ.n] (*output* com apagamento do [g]). A previsão para essas sequências era a de que tivéssemos alguma ocorrência de epêntese, seguindo a lógica da pressão da consoante sonora sobre a sequência.

Conforme os dados nas tabelas, podemos ver que /gm/ não segue a tendência de apresentar inserção de vogal. Por outro lado, devemos considerar que a presença de epêntese que vimos na sequência /bd/ foi muito baixa, o que nos faz repensar a hipótese de considerar

que uma consoante sonora qualquer possa estar incentivando a inserção da epêntese após a plosiva em coda.

Para /gn/, os dados mostram 13% de epêntese. Acreditando que o fenômeno da epêntese tenha de seguir um padrão, ou seja, que os dados não devem ocorrer de forma aleatória, devemos captar essas generalizações de forma a tentar dar conta do sistema de interlíngua dos estudantes colombianos, aprendizes de PB. Esse encaminhamento é fundamental para podermos formalizar as restrições que serão utilizadas na aplicação dos modelos teóricos na análise dos dados mostrada a seguir.

Neste momento da observação do *corpus*, também podemos descartar a ideia de que a escala de ponto de articulação – dorsal > labial > coronal – esteja agindo sobre os dados, de forma que os segmentos mais marcados da escala (dorsais), que poderiam ser os responsáveis pelo emprego da epêntese como forma de facilitar a produção, não são os que mais apresentam o fenômeno da inserção vocálica.

O único fato que podemos considerar categórico até o momento é o da explosão da plosiva em coda. Até agora, as plosivas sonoras mostraram que sofrem lenição maior em coda, com relação às plosivas surdas. Tanto nas produções em Português Brasileiro, quanto em Espanhol Colombiano, as plosivas surdas têm mais soltura do que o inverso, situação diversa das plosivas sonoras.

**Tabela 18 - Outputs das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /kt/ nas palavras “cacto” e “convicto”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
k.t	33% (4/12)	42% (5/12)	100% (12/12)	50% (6/12)	56% (27/48)
ki.t	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
k'.t	67% (8/12)	58% (7/12)	0 (0/12)	50% (6/12)	44% (21/48)
ɬ.t	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Erro	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)

**Tabela 19 - Outputs das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /kn/ nas palavras “tecnicismo” e “estricnina”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
k.n	8% (1/12)	0 (0/12)	0 (0/12)	59% (7/12)	17% (8/48)
ki.n	0 (0/12)	0 (0/12)	33% (4/12)	8% (1/12)	10% (5/48)
ḳ.n	0 (0/12)	0 (0/12)	17% (2/12)	0 (0/12)	4% (2/48)
∅.n	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	92% (11/12)	92% (11/12)	42% (5/12)	33% (4/12)	65% (31/48)
Erro	0 (0/12)	8% (1/12)	8% (1/12)	0 (0/12)	4% (2/48)

Tratando das plosivas dorsais em coda /kt/ e /kn/, apresentamos, para ambas as sequências, quatro possibilidades de *outputs* para as produções no Português: [k.t] (*output* sem epêntese), [ki.t] (*output* com epêntese), [ḳ.t] (*output* sem a explosão da plosiva em coda) e [∅.t] (*output* com apagamento do [k]).

E, para /kn/, as possibilidades foram: [k.n] (*output* sem epêntese), [ki.n] (*output* com epêntese), [ḳ.n] (*output* sem a explosão da plosiva em coda) e [∅.n] (*output* com apagamento do [k]).

As porcentagens de epêntese foram: [k.t] igual a zero e [k.n] igual a 10%. Mais uma vez, temos mais explosões na plosiva surda e temos a sequência com sonora apresentando epêntese. Os casos relatados em “Outros” tratam do vozeamento do [k]. Em uma palavra como “tecnicismo”, encontramos uma produção que substitui o [k] pelo [g], de modo que os casos não tenham sido raros, 81% no Espanhol e 65% no Português.

Resta, ainda, a análise da sequência /tn/, para podermos concluir uma generalização específica às sequências que são consideradas para a formalização das restrições. Porém, acreditamos que, dadas as circunstâncias apresentadas até o presente momento, podemos já antecipar que o contato silábico está sendo significativo no aparecimento da epêntese.

### 4.3.3 Plosivas alveolares em coda: [t] e [d]

**Tabela 20 - Outputs das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /dm/ nas palavras “admissão” e “admirador”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
d.m	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
dʒi.m	0 (0/12)	0 (0/12)	0 (0/12)	75% (9/12)	19% (9/48)
dʒ.m	0 (0/12)	0 (0/12)	0 (0/12)	8% (1/12)	2% (1/48)
d̃.m	100% (12/12)	100% (12/12)	100% (12/12)	17% (2/12)	79% (38/48)
ɐ.m	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Erro	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)

**Tabela 21 - Outputs das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /tm/ nas palavras “algoritmo” e “ritmo”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
t.m	58% (7/12)	0 (0/12)	25% (3/12)	0 (0/12)	21% (10/48)
tʃi.m	0 (0/12)	0 (0/12)	0 (0/12)	8% (1/12)	2% (1/48)
tʃ.m	0 (0/12)	0 (0/12)	0 (0/12)	92% (11/12)	23% (11/48)
t̃.m	8% (1/12)	17% (2/12)	0 (0/12)	0 (0/12)	6% (3/48)
ɐ.m	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	34% (4/12)	83% (10/12)	75% (9/12)	0 (0/12)	48% (23/48)
Erro	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)

**Tabela 22 - *Outputs* das produções das palavras cognatas Português/Espanhol em frases do Português para a sequência /tn/ nas palavras “vietnam” e “etnia”**

	Sujeito A	Sujeito D	Sujeito J	Sujeito S	Total
t.n	42% (5/12)	0 (0/12)	0 (0/12)	0 (0/12)	10% (5/48)
tʃi.n	0 (0/12)	0 (0/12)	0 (0/12)	33% (4/12)	8% (4/48)
tʃ.n	0 (0/12)	8% (1/12)	0 (0/12)	67% (8/12)	19% (9/48)
t̃.n	42% (5/12)	8% (1/12)	17% (2/12)	0 (0/12)	17% (8/48)
ɛ.n	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)
Outros	16% (2/12)	83% (10/12)	83% (10/12)	0 (0/12)	46% (22/48)
Erro	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/12)	0 (0/48)

Para a sequência /dm/, são cinco as possibilidades de *outputs* para as produções no Português: [d.m] (*output* sem epêntese), [dʒi.m] (*output* com epêntese), [d̃.m] (*output* sem a explosão da plosiva em coda), [ɛ.m] (*output* com apagamento do [d]) e [dʒ] (*output* com a palatalização da plosiva em coda sem a inserção de epêntese).

Para a sequência /tm/, oferecemos cinco possibilidades também: [t.m] (*output* sem epêntese), [tʃi.m] (*output* com epêntese), [t̃.m] (*output* sem a explosão da plosiva em coda), [ɛ.m] (*output* com apagamento do [t]) e [tʃ] (*output* com a palatalização da plosiva em coda, ou seja, palatalização sem epêntese). Finalmente, para /tn/, encontramos: [t.n] (*output* sem epêntese), [tʃi.n] (*output* com epêntese), [t̃.n] (*output* sem a explosão da plosiva em coda), [ɛ.n] (*output* com apagamento do [t]) e [tʃ] (*output* com a palatalização da plosiva em coda, ou seja, palatalização sem epêntese).

Na seção 2.2 do Referencial Teórico, que trata dos padrões de coda no PB e questões pertinentes de análise, procuramos mostrar que, para o Português falado na cidade de Pelotas, temos o [t] e o [tʃ], bem como o [d] e o [dʒ], como alofones diante da vogal [i]. Palavras como “atividade” ou mesmo as palavras em que acrescentamos epêntese, como “ad[i]missão”, são produzidas com a plosiva coronal palatalizada. Com a leitura dos dados de produção nas tabelas acima, foi possível confirmar esse fenômeno, já que tivemos *outputs* palatalizados pelos estudantes colombianos, mesmo sem a inserção da vogal. Quanto à ocorrência de epêntese, encontramos, respectivamente, os seguintes índices para /dm/, /tm/ e /tn/: 19%, 2% e 8%.

Para os casos em que a plosiva em coda é /t/, assim como encontramos no Espanhol, também nos dados do Português houve vozeamento em coda. Tanto em /tm/ quanto em /tn/, metade das produções – que estão relatadas em “Outros” – sofreu assimilação da sonoridade da nasal com conseqüente substituição da correspondente surda /t/ pela sonora /d/.

A análise das plosivas alveolares em coda revelou algumas situações importantes, como a confirmação de que as nasais ou as consoantes sonoras em *onset* da sílaba seguinte à plosiva em coda exercem, de alguma forma, influência sobre a coda precedente. Também foi possível confirmar que, diante de nasais, o enfraquecimento da plosiva se dá de forma diferente do que ocorre nos casos de sequências com *plateaux*. Por fim, foi possível comprovar, pelas produções realizadas pelos aprendizes, que eles estiveram expostos a *inputs* de alveolares palatalizadas, já que foram reproduzidas mesmo sem inserção de epêntese.

A partir deste momento, já estamos aptos a responder a duas das seis questões norteadoras postas:

1) A epêntese vocálica que ocorre no PB, após segmentos plosivos, em codas mediais de palavras, será produzida por aprendizes colombianos de PB, cuja L1 admite segmentos plosivos em coda?

Podemos dizer que sim, que os aprendizes de PB que são colombianos produziram epêntese após segmentos plosivos em codas mediais de palavras, porém, em baixos índices, o que demonstra que os aprendizes ainda aplicam, mesmo para uma LE, padrões da LM, sendo que já apresentam um sistema de interlíngua diferente dessa língua materna. Entretanto, em se considerando que, no sul do Brasil, temos o registro da presença predominante da epêntese após a plosiva em posição de coda medial, caracterizamos que os aprendizes se encontram distantes da aquisição completa de tal padrão da L2.

2) Qual(is) fator(es) linguístico(s) contribui(em) para a ocorrência da epêntese por parte de falantes nativos do Espanhol Colombiano?

Com a análise acústica dos dados finalizada, verificamos que o contato silábico teve papel significativo na inserção de epêntese por parte dos aprendizes do Português Brasileiro, nas sequências de plosiva em coda que selecionamos para esta pesquisa. Observamos que as sequências que apresentavam a consoante nasal seguinte à plosiva em coda apresentaram maior número de epêntese em comparação às sequências com *plateau* de duas plosivas no contato silábico.

Este capítulo de descrição dos dados mostra-se fundamental para formalização das restrições nas análises via Teoria da Otimidade e Gramática Harmônica, que nos propusemos.

#### 4.4 COMPARAÇÃO DOS DADOS DE PRODUÇÕES DOS SUJEITOS NA L1 (ESPAÑHOL COLOMBIANO) E NA L2 (PORTUGUÊS BRASILEIRO) – GENERALIZAÇÕES

Os *outputs* encontrados na análise acústica foram muito significativos para a continuação da pesquisa ora proposta porque, por um lado, confirmam nossa hipótese sobre a realização da epêntese pelos aprendizes colombianos do Português Brasileiro no uso da L2 e, por outro, apresentam situações que corroboram os padrões de coda plosiva no Espanhol.

Primeiramente, podemos informar que, em se tratando de todas as sequências heterossilábicas no Espanhol Colombiano e no Português Brasileiro, não encontramos apagamento e soltura longa. No PB isso não é surpreendente, porque o sistema de correção silábica em situação de uma sequência não-ideal é a inserção da vogal, a epêntese. Para o Espanhol, não raramente foi também visto que a plosiva em coda caracteriza uma situação perfeitamente possível na sua estrutura e, conforme Harris (1983), ainda para as sequências não tautossilábicas propostas, a aspiração não seria o usual.

Outra observação decorrente da análise acústica foi relativa ao enfraquecimento da plosiva em posição de coda. Essa situação foi uma constante nas produções nas duas línguas. Relatamos este fato na representação fonética que utilizamos para caracterizar que não houve explosão da plosiva (ex.: p̃). Porém, nas formas de onda das sequências que propusemos, nos espectrogramas, vemos que, diante de outra plosiva surda, este fato é mais raro. Nas sequências [p.t, k.t, b.t], mais da metade das produções se deu com soltura de ar audível; em contrapartida, em todas as sequências diferentes do contato de plosiva qualquer mais plosiva surda, ou seja, nas sequências [b.d, k.n, p.n, d.m, g.m, t.m, g.n, t.n], mais da metade das produções apresentam lenição da plosiva em coda.

Com exceção das sequências [p.t, k.t, d.m, b.t, b.d], as demais sequências [k.n, p.n, t.m, t.n, g.n, g.m], todas seguidas de consoante nasal, sofreram fenômeno da sonorização, nos casos de plosiva em coda surda, como, por exemplo, [ab̃neja] para a palavra “apneia”, e dessonorização, nos casos de a plosiva em coda ser sonora, como, por exemplo, [mak̃no] para a palavra “magno”. Todos os casos foram mais frequentes nas produções realizadas em Espanhol, mas aconteceram também nas produções em Português Brasileiro. Em Harris (1969), já temos descrita esta mesma ocorrência para o espanhol mexicano, tanto para o vozeamento, quanto para o desvozeamento das plosivas em posição de coda. Entretanto, Harris aponta que, no Espanhol mexicano, este fenômeno ocorre de forma parcial, ou seja, os

segmentos não se sonorizam completamente. Em nossa análise, não entramos nesta discussão, mas acreditamos ser uma constatação interessante no que se refere ao espanhol latino.

Destacamos que não foi realizado teste de reconhecimento ou de frequência de palavra, para verificar se estaria havendo algum tipo de influência na inserção de epêntese nesse sentido. Isso se deve, principalmente, ao recorte proposto na metodologia, que não permitiu muitas opções de palavras, apesar de termos feito a escolha baseada nas atividades cotidianas dos estudantes, como palavras que tivessem a ver com atividades acadêmicas e com os cursos que estavam frequentando. Sabemos que a realização de um tipo de teste de frequência de ocorrência de palavra poderia ter enriquecido a análise, mas acreditamos que esse fato não alteraria o resultado final, já que acreditamos ser a estrutura silábica o fator motivador de ocorrência da epêntese.

Do conjunto dos dados que aqui se encerra, podemos dizer então que há ocorrência do fenômeno da epêntese, mesmo que de forma tímida, no uso de Português por falantes de Espanhol, e que o contato silábico com um segmento sonoro em *onset*, especificamente, uma consoante nasal, faz diferença na inserção da vogal. Encontramos, nas produções em PB, as seguintes porcentagens de epêntese:

[b.t] = 0, [b.d] = 4%, [k.t] = 0, [p.t] = 2%, [g.m] = 0, [t.m] = 2%, [t.n] = 8%, [k.n] = 10%, [g.n] = 13%, [d.m] = 19%, [p.n] = 21%

Nos casos das sequências /dm/, /tm/ e /tn/, tivemos, ainda, casos de palatalização da plosiva, mesmo sem epêntese, demonstrando uma aproximação da produção dos aprendizes aos padrões realizados no Português Brasileiro diante da vogal [i]. Podemos acrescentar que, se seguirmos uma escala que classifica como menos marcados os casos em que a coda seja mais sonora do que o *onset* da sílaba seguinte, temos que as sequências de plosiva mais nasal serão as mais marcadas, sendo exatamente essas as que estão demonstrando mais epêntese.

No próximo capítulo, vamos apresentar a análise dos dados através dos modelos da Teoria da Otimidade e Gramática Harmônica. Vamos, assim, determinar as restrições que serão utilizadas, bem como discutir o comportamento dos dados nos dois modelos, para que possamos responder aos demais questionamentos propostos.

## 5 AQUISIÇÃO DO PORTUGUÊS BRASILEIRO VIA TEORIA DA OTIMIDADE ESTOCÁSTICA E GRAMÁTICA HARMÔNICA

### 5.1 A FORMALIZAÇÃO DAS RESTRIÇÕES PARA AS SEQUÊNCIAS [p.n, k.n, p.t, k.t]: UM ESTUDO-PILOTO

Conforme já vimos no capítulo do Referencial Teórico, no que diz respeito às seções referentes à Teoria da Otimidade e à Gramática Harmônica, ambos os modelos têm, como base de funcionamento, a interação entre restrições que são violáveis e universais. Logo, definir as restrições é uma etapa fundamental e cabe ao pesquisador refletir sobre o tipo de restrições capaz de dar conta das gramáticas das línguas em análise.

Segundo Alves (2008), com foco na aquisição de L1 há uma série de trabalhos (DEMUTH 1995, LEVELT 1995, PATER & PARADIS 1996, SMOLENSKY 1996, GNANADESIKAN 2004, LEVELT & VAN DE VIJVER 2004, DAVIDSON ET AL. 2004) que apontam para um estágio inicial em que as restrições de marcação dominam as restrições de fidelidade. Ou seja, no primeiro estágio de aquisição de L1, a criança vai ter produções de *outputs* menos marcadas e vai modificando a hierarquia das restrições de forma que as relações de ordenamento entre fidelidade e marcação se modifiquem gradualmente, para que as estruturas mais marcadas passem a emergir.

Alves (2008) ainda afirma que, para o caso de aquisição de L2, uma série de outros trabalhos (BROSELOW ET AL., 1998, PATER 1997, HANCIN-BHATT & BHATT 1998, DAVIDSON ET AL. 2004) argumentam que a hierarquia inicial em direção à L2 corresponde ao ranking da L1. Logo, podemos apontar quais *outputs* indicam um estágio mais ou menos avançado na aquisição da segunda língua, utilizando-se as mesmas restrições, porém, com ordenamentos diferentes para tais restrições. Esse será um passo importante à construção dos sistemas de restrições que iremos utilizar para identificar os estágios dos aprendizes do PB nas aquisições da língua materna, da língua estrangeira e a interlíngua, que será o estágio intermediário entre as duas outras hierarquias.

Em outubro de 2010, realizamos uma primeira análise sob os modelos da Gramática Harmônica e da Teoria da Otimidade, para verificarmos a viabilidade de utilização de tais modelos para os dados já levantados no presente trabalho. Do total de palavras selecionadas para o estudo naquele momento (22 palavras), o teste foi feito com apenas oito

palavras, que contemplavam as sequências [p.n, k.n, p.t, k.t], (duas palavras para cada sequência). Naquela etapa da pesquisa, ainda não tínhamos a análise acústica completa.

No estudo piloto em questão, utilizamos somente as produções realizadas nas frases:

- Puedes Decir \_\_\_\_\_.

- Diga\_\_\_\_\_.

Assim, o Estudo-Piloto contou com menos da metade das produções totais levantadas para a pesquisa em sua integralidade.

Esse estudo foi fundamental, pois revelou que não seria simples encontrar uma regularidade motivada por um fenômeno linguístico, que dê conta da formalização da epêntese nas restrições nos dois sistemas que estamos estudando (Português Brasileiro e Espanhol), no que se refere à aquisição do segmento plosivo em coda e à presença ou não de epêntese no uso do PB como L2. O número total de sequências que havíamos escolhido para compor o nosso trabalho – 11 sequências – pareceu exigir muito mais atenção para que se chegasse a uma definição sobre quais restrições seriam capazes de explicitar o nosso alvo.

Dessa forma, mostramos aqui como chegamos às restrições para o Estudo-Piloto e para a análise final dos dados da pesquisa. Demonstramos, também, o comportamento das mesmas quatro sequências desse Estudo-Piloto, com o acréscimo da sequência [t.n] para diferenciar a amostra daquele estudo preliminar e caracterizar a análise final, tentando também justificar a não-utilização das 11 sequências propostas inicialmente para esta pesquisa.

Já vimos que, no PB, a epêntese é a estratégia de reparo utilizada de modo a corrigir uma estrutura silábica como a de uma plosiva em coda. Além disso, vimos também que, no Espanhol, não é necessária uma alternativa como a epêntese, já que é perfeitamente possível a presença de plosiva em coda. Logo, as restrições tiveram que ser pensadas de forma a refletir as gramáticas responsáveis por essas formas de saída.

Os resultados das produções no Espanhol Colombiano e no Português Brasileiro para parte dos dados – dados utilizados no Estudo-Piloto, que constituíam o *corpus* de Quintanilha & Alves (2010) - podem ser verificados nas tabelas 23 e 24:

**Tabela 23 - Produções das seqüências [p.n, k.n, p.t, k.t] no Espanhol Colombiano**

Consoante em Coda	Epêntese	Apagamento
Consoante em Coda	Epêntese	Apagamento total
[k.n]	[ki.n]	[ɔ.n]
100%	0	0
Consoante em Coda	Epêntese	Apagamento total
[k.t]	[ki.t]	[ɔ.t]
100%	0	0
Consoante em Coda	Epêntese	Apagamento total
[p.n]	[pi.n]	[ɔ.n]
100%	0	0
Consoante em Coda	Epêntese	Apagamento total
[p.t]	[pi.t]	[ɔ.t]
100%	0	0

**Tabela 24 - Produções das seqüências [p.n, k.n, p.t, k.t] no Português Brasileiro**

Consoante em Coda	Epêntese	Apagamento
Consoante em Coda	Epêntese	Apagamento total
[k.n]	[ki.n]	[ɔ.n]
77,27%	22,73%	0
Consoante em Coda	Epêntese	Apagamento total
[k.t]	[ki.t]	[ɔ.t]
100%	0	0
Consoante em Coda	Epêntese	Apagamento total
[p.n]	[pi.n]	[ɔ.n]
83%	17%	0
Consoante em Coda	Epêntese	Apagamento total
[p.t]	[pi.t]	[ɔ.t]
100%	0	0

Com os dados de produção levantados nesse momento, nossa análise nos levou a crer que tanto o contato silábico quanto o ponto de articulação estariam influenciando nas produções com epêntese dos informantes. Se pensarmos em uma escala de ponto tal como dorsal > labial > coronal, aliada à lei de contato silábico, teríamos que: [k.n] > [p.n] > [k.t] >

[p.t]. Isso foi exatamente o que os nossos dados estavam refletindo, conforme é possível visualizar na tabela 24, das produções em Português.

Assim, as sequências mais marcadas, ([k.n] - 22,73% e [p.n] - 17%) apresentaram, nas produções em Português, maior número de epêntese com relação às sequências listadas mais abaixo nessa escala ([k.t] - zero e [p.t] - zero).

Isso fez com que formalizássemos restrições que refletissem essa situação, conforme abaixo:

### RESTRICÇÕES DE FIDELIDADE

- Dep – Oposição à ocorrência de epêntese
- Max – Oposição à ocorrência de apagamento

### RESTRICÇÕES DE MARCAÇÃO

#### **RESTRICÇÕES DE CONTATO SILÁBICO**

- \*Dist {1,0} (cf. Gouskova, 2004) – Oposição à distância de sonoridade
- \*Dist {1} (cf. Gouskova, 2004)

#### **RESTRICÇÕES REFERENTES A PONTO DE ARTICULAÇÃO**

- \*{stop}<sub>coda</sub> - Oposição à produção de segmentos plosivos em coda silábica.
- \*{k,p}<sub>coda</sub> – Oposição à produção de segmentos dorsais e labiais em coda silábica.
- \*{k}<sub>coda</sub> – Oposição à produção de segmentos dorsais em coda silábica

#### **RESTRICÇÕES CONJUNTAS<sup>14</sup>**

- \*{p.t} [\*Dist {1,0} & \*{k,p}<sub>coda</sub>]
- \*{p.n} [\*Dist {1} & \*{k,p}<sub>coda</sub>]
- \*{k.t} [\*Dist {1,0} & \*{k}<sub>coda</sub>]
- \*{k.n} [\*Dist {1} & \*{k}<sub>coda</sub>]

Vimos que as restrições podem pertencer a duas famílias: Fidelidade e Marcação. Na primeira, devemos estabelecer restrições que busquem a preservação do *input*. Para o caso ora posto, de plosiva em posição de coda medial de palavra, acreditamos que as restrições de fidelidade DEP e MAX (McCARTHY E PRINCE, 1995) podem dar conta, em substituição às restrições PARSE e FILL<sup>15</sup> de Prince e Smolensky (1993).

Para as restrições de marcação referentes à coda – nosso alvo –, achamos necessário agregar restrições de contato silábico, restrições de caráter estrigente e restrições

<sup>14</sup> A formalização de restrições conjuntas de caráter estrigentes baseia-se na proposta de Alves (2008).

<sup>15</sup> Não entraremos nessa discussão, mas consideramos interessante observar que em Prince e Smolensky (1993), PARSE era a restrição semelhante a MAX e FILL a DEP.

conjuntas. Ao longo desta seção, pretendemos exatamente explicar e justificar o uso de tais restrições no Estudo-Piloto.

A primeira decisão tomada foi a da utilização do mecanismo de Alinhamento Harmônico (HA) de Prince & Smolensky (1993), pois esse mecanismo propiciará um conjunto de restrições que pode mostrar quais segmentos se apresentam mais aptos a ocorrer em coda silábica a partir de uma escala linguística. Porém, para a utilização desse mecanismo, foi necessário ainda decidir se as restrições seriam formalizadas através de uma hierarquia universalmente fixa ou através de restrições em relação de estringência<sup>16</sup>.

Numa relação de ranking fixo, temos que algo será sempre mais marcado em relação a outro. Ex.: “A” é mais marcado do que “B” em qualquer situação. Na noção de estringência, teremos uma consideração um pouco diferente da anterior. Partindo da nossa análise, podemos dizer que segmentos labiais ([p], por exemplo) nunca serão mais marcados do que dorsais ([k], por exemplo), mas que podem ser igualmente marcados, na interação das restrições no ranking.

Como no presente estudo só estamos lidando com plosivas em coda, o que nos interessa é o ponto de articulação dessa plosiva (dorsal > labial > coronal) e o seu valor com relação à consoante seguinte, conforme a Lei do Contato Silábico. De acordo com a noção de estringência, precisávamos fazer com que a restrição que se opõe ao menos marcado fosse a mais violada, considerando-se [kn] > [kt] e [pn] > [pt].

Logo, enquanto a relação de ranking fixo vai sempre afirmar que A é mais marcado do que B, a relação de estringência não vai admitir que as restrições tenham esse tipo de relação de dominância, visto que, sob tal relação, uma restrição mais específica vai implicar demissão da restrição mais geral.

Dessa forma, ficamos com as restrições  $\ast\{\text{stop}\}_{\text{coda}}$  – que equivale a [k], [p], [t], em função do Alinhamento Harmônico, que vai prever a existência desta restrição -,  $\ast\{k, p\}_{\text{coda}}$ ,  $\ast\{k\}_{\text{coda}}$  e, num estágio inicial de aquisição, em que as restrições de marcação receberão valor inicial igual a 100 e as de fidelidade 0, a presença de um candidato [k] - ponto mais marcado na escala de articulação - ocasionaria a demissão de todas essas restrições em estringência. Assim, a restrição  $\ast\{k, p\}_{\text{coda}}$  seria demovida não só pela presença de um [k], mas

---

<sup>16</sup> A relação de estringência mostrou-se plenamente pertinente na análise de Alves (2008). Para a definição de estringência, segue-se Matzenauer & Miranda (2010), que explicam ser esse um tipo de relação de subconjunto com referência a violações de restrições, sendo formalmente assim expressa, segundo McCarthy (2008: 65-66): “A restrição R1 é mais estrigente do que a restrição R2 se toda a violação de R2 implicar também a violação de R1, sendo que algumas violações de R1 não implicam violação de R2”. Para maiores explicações sobre sistemas de estringência, veja-se Alves & Keller (2010).



(\*DISTANCE), onde \*Dist (0) >> \*Dist (-1) >> \*Dist (-2), já que, nesse caso, quanto maior a queda de sonoridade, mais harmônica é a distância. Em outras palavras, é melhor termos um encontro [k.t] do que um [k.n], ou um encontro [p.t] a um [p.n], conforme já discutimos.

Já que trabalhamos, no Estudo-Piloto, apenas com as sequências [p.n, k.n, p.t, k.t], abordamos apenas dois diferentes graus de distância de sonoridade entre a coda e o *onset* da sílaba seguinte: Dist {0} e Dist {+1}. Julgamos interessante mencionar que o Mecanismo de Alinhamento Relacional, conforme proposto por Gouskova (2004), não sugere relações de estringência, mas de ranking fixo entre tais restrições. Em nossa análise, pensamos, para esta e as demais análises realizadas, numa organização que trabalhe esta relação em estringência, para que possamos demonstrar a emergência da distância menos marcada (Dist{0}) sobre uma distância mais marcada (Dist{+1}), tal como fizemos com as demais restrições utilizadas.

As duas restrições obtidas do Alinhamento Relacional, entretanto, não conseguiam dar conta das diferenças de marcação que existiam entre [p.n] e [k.n] ou [p.t] e [k.t], diferenças essas que eram efeito, também, dos pontos de articulação dos segmentos de coda. Logo, era necessário combinar as possibilidades de ponto de articulação em coda (já que todas as codas são plosivas) e a escala de sonoridade em *onset* da sílaba seguinte (já que contamos com plosivas e nasais em *onset* da sílaba seguinte), de forma a construirmos restrições que pudessem satisfazer tais condições.

a) Coda – **dorsal** > **labial** > coronal = [k] > [p]

b) *Onset* - líquida > **nasal** > fricativa > **plosiva** = [n] > [t]

Nossas restrições nesse sentido foram:

- \*Dist {1,0} (cf. GOUSKOVA, 2004)
- \*Dist {1} (cf. GOUSKOVA, 2004)

Para esse fim, fizemos uso do mecanismo de Conjunção Local (SMOLENSKY, 1995). Julgávamos necessário estabelecer restrições que, através de uma ação conjunta, se responsabilizassem pela distinção entre ponto de articulação em coda e contato silábico. A ideia é trabalharmos com domínios onde um domínio “D” será violado sempre que ambas as restrições são violadas.

Abaixo, são listadas as Restrições de Conjunção Local propostas:

- \*{p.t} restrição conjunta – [\*Dist {1,0} & \*{k,p}coda]
- \*{p.n} restrição conjunta – [\*Dist {1} & \*{k,p}coda]
- \*{k.t} restrição conjunta – [\*Dist {1,0} & \*{k}coda]
- \*{k.n} restrição conjunta – [\*Dist {1} & \*{k}coda]

Com o *tableau* ilustrativo a seguir, pretendemos demonstrar que a Restrição Conjunta vai permitir que as restrições que a compõem, mesmo mais baixas no *ranking*, tenham um efeito que as considere mais marcadas em conjunto, através da formalização [ $*\text{Dist } \{+1\} \& * \{k\}_{\text{coda}}$ ].

**Tableau 5 - Ilustrativo - Exemplo da formalização de Restrições de Conjunção Local**

	{k.n}restrição conjunta - [ $*\text{Dist } \{+1\} \& * \{k\}_{\text{coda}}$ ]	Dist {+1}	$*k_{\text{coda}}$
k			*
k.n	*	*	*
p.n		*	

Juntas, estas duas Restrições representadas no *tableau* ilustrativo terão condições de fazer efeito, como se estivessem ranqueadas em posição mais alta. Mas, para tal, a restrição conjunta deve se encontrar em posição mais alta no ranking e, também, devemos cuidar para que as violações ocorram apenas num mesmo domínio, sob pena de poder exercer efeitos em ambientes sob os quais não deveria atuar.

Se, por exemplo, tivéssemos uma Restrição DEP, entre a Restrição conjunta e a Restrição Dist {+1}, [p.n] seria possível, [k] seria possível, mas [k.n] não. DEP ganharia a disputa com cada uma das Restrições sozinhas, mas perderia delas em conjunto.

A ideia de utilizarmos as restrições conjuntas surge com a intenção de preservarmos as restrições de marcação por contato silábico (Dist) e ponto de articulação, no sentido de não permitirmos que nenhuma delas pudesse agir sobre a outra. A presença de um [k.t], por exemplo – que, por ser menos marcada, deve ter as restrições que a ela se referem demovidas primeiramente, com relação às restrições referentes à sequência [k.n] -, faria com que as restrições referentes a [p.t] e [p.n] também fossem violadas, de forma a manter-se um ordenamento das restrições por marcação de ponto.

Um aspecto interessante que veremos no Estudo-Piloto, ainda em relação às Restrições de Conjunção Local, diz respeito ao fato de que, na HG, as restrições já agem de forma cumulativa; logo, nosso empenho em resolver um problema da OT Estocástica poderia ter algum efeito indesejado sobre a Gramática Harmônica, como o fato de as restrições conjuntas aparecerem como redundantes. Porém, conforme veremos a seguir, não tivemos nenhum problema em utilizar restrições conjuntas em ambos os modelos de análise. Com o conjunto de restrições que formalizamos, foi possível fazer com que os dados nos dessem a resposta que esperávamos. No que segue, veremos os resultados organizados em *tableaux*.

### 5.1.1 Simulação 1 - Aquisição do Espanhol Colombiano (L1)

Feitos os esclarecimentos a respeito do papel de cada restrição no processo de aquisição da língua materna e estrangeira dos aprendizes, vamos agora tratar de cada caso em separado. Serão realizadas três simulações em cada modelo teórico, respectivamente: aquisição da L1 (Espanhol); aquisição plena da L2 (Português Brasileiro) e aquisição parcial (interlíngua), essa última com os índices de *outputs* variáveis encontrados em nosso *corpus*. Essa proposta de realizarmos três simulações em cada modelo teórico, tentando representar a hierarquia de restrições para cada língua e ainda para a interlíngua, foi realizada em todas as análises que fizemos até o fim do trabalho.

Em todos os conjuntos de *tableaux* a serem discutidos, fornecidos pelo *software PRAAT*, a coluna denominada *ranking value* fornece o valor central das restrições. Conforme demonstramos no Referencial Teórico, este é o valor numérico que será promovido ou demovido pelo algoritmo de aprendizagem. Na coluna denominada *disharmony*, teremos os valores de ponto de seleção, que determinam um momento de produção linguística.

a) A aquisição do Espanhol Colombiano (L1) via Teoria da Otimidade Estocástica

Apresentamos, neste item, a aquisição dos segmentos plosivos [k, p], em coda medial de palavra, com uma ocorrência de 0% de epêntese no Espanhol Colombiano, via Teoria da Otimidade Estocástica. A simulação 1, realizada no *software PRAAT*, é dada no conjunto de *tableaux* 6, a seguir:

Conjunto de Tableaux 6 - Simulação 1 OT / Aquisição do Espanhol Colombiano (0% de epêntese)

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
<b>Dep</b>	60.270	61.773	1.000000
<b>Max</b>	61.730	59.912	1.000000
<b>*{k.n}</b>	48.417	49.167	1.000000
<b>*{p.n}</b>	24.193	25.456	1.000000
<b>*{k}coda</b>	20.365	21.615	1.000000
<b>*Dist {1}</b>	24.193	20.030	1.000000
<b>*{k.t}</b>	20.365	19.301	1.000000
<b>*{stop}coda</b>	-22.001	-18.226	1.000000
<b>*{p.t}</b>	-22.001	-18.284	1.000000
<b>*{p}coda</b>	-22.001	-21.108	1.000000
<b>*Dist {1,0}</b>	-22.001	-22.325	1.000000

pt	Dep	Max	*{k.n}	*{p.n}	*{k}coda	*Dist {1}	*{k.t}	*{stop}coda	*{p.t}	*{p}coda	*Dist {1,0}
☞ p.t								*	*	*	*
pi.t	*!										
t		*!									

kt	Dep	Max	*{k.n}	*{p.n}	*{k}coda	*Dist {1}	*{k.t}	*{stop}coda	*{p.t}	*{p}coda	*Dist {1,0}
☞ k.t					*		*	*	*	*	*
ki.t	*!										
t		*!									

pn	Dep	Max	*{k.n}	*{p.n}	*{k}coda	*Dist {1}	*{k.t}	*{stop}coda	*{p.t}	*{p}coda	*Dist {1,0}
☞ p.n				*		*		*	*	*	*
pi.n	*!										
n		*!									

kn	Dep	Max	*{k.n}	*{p.n}	*{k}coda	*Dist {1}	*{k.t}	*{stop}coda	*{p.t}	*{p}coda	*Dist {1,0}
☞ k.n			*	*	*	*	*	*	*	*	*
ki.n	*!										
n		*!									

Esta primeira simulação trata da aquisição do Espanhol Colombiano, língua materna dos aprendizes; logo, os *outputs* devem responder a este estímulo de forma a evidenciar um resultado de 100% de produção da consoante plosiva em coda, sem epêntese ou apagamento, assim como podemos verificar nas ocorrências de todas as sequências dadas pelas produções dos aprendizes de PB [p.t, p.n, k.t, k.n].

Nesta primeira simulação na Teoria da Otimidade Estocástica, contudo, o sistema foi programado a responder a um estágio inicial de uma criança adquirindo uma língua, de modo que o estágio final seja a primeira língua desta criança - no caso desse estudo - o Espanhol Colombiano. Esta situação sugere lidarmos com marcação com peso 100% e Fidelidade com peso igual a 0 no estágio inicial. O *ranking* das restrições, apresentado no conjunto de *tableaux* 6, deve dar conta desse aprendizado, ao revelar a hierarquia das restrições que vai se dar pelos pesos numéricos.

A hierarquia de restrições, formalizada para representar a aquisição das codas medias no Espanhol Colombiano, mostra que é necessário que as restrições de fidelidade DEP

e MAX sejam promovidas e que, além disso, alcancem um distanciamento do valor<sup>17</sup> das restrições de marcação, de forma a demonstrar que é categórica a produção da plosiva em posição de coda medial no Espanhol Colombiano.

Ainda que a cada avaliação tenhamos diferentes pontos de seleção, nosso *output* ótimo será sempre o mesmo, pois o ranqueamento de DEP (60.270) é suficientemente superior ao das restrições de marcação. Por exemplo,  $\{k.n\}$ , que é a restrição mais próxima de DEP, apresenta um peso igual a 48.417. Disso resultam os *outputs* categóricos, sem epêntese sob um *input* encerrado por segmentos plosivos.

Nesse estágio final da gramática da língua materna dos nossos aprendizes, verificamos, ainda, que os valores de ranqueamento das restrições de marcação respeitam um ordenamento de marcação, que atribuem índices mais baixos às restrições que se referem a sequências menos marcadas e mais gerais, em termos de estringência. Conseguimos demonstrar, dessa forma, uma aquisição em ordem de marcação, onde [p.t] emerge anteriormente a [k.t] e a [p.n] e [k.n], o que nos é garantido pela estringência entre as restrições. De fato, uma restrição mais específica como  $\{Dist\ 1\}$  – que se refere ao contato de uma obstruente e uma consoante nasal, quando violada, vai demover também  $\{Dist\ 1,0\}$ .

Logo, o algoritmo consegue convergir para uma gramática que se apresenta categórica, em que os segmentos plosivos em codas mediais são produzidos sem epêntese (estágio final da aquisição da L1), conforme indicado no conjunto de *tableaux* 6 das sequências [k.t, k.n, p.t, p.n], na simulação 1 da OT Estocástica.

A figura 13, a seguir, trata da distribuição probabilística dos *outputs* (*output distributions*). No comando “*to output distributions*” do *software* PRAAT, temos condições de submeter os nossos candidatos a 100.000 avaliações, para comprovar se a gramática em questão está resultando nos índices percentuais desejados.

---

<sup>17</sup> De acordo com o sistema computacional, o distanciamento abaixo de 10 pontos entre os valores das restrições (*ranking value*) corresponde à variação, ou seja, à possibilidade de termos *outputs* variáveis.



No que se refere à escolha do candidato ótimo, na HG devemos considerar, além dos pesos das restrições, as violações sofridas por cada restrição. Assim, o candidato ótimo será dado pelo valor de Harmonia - que é o somatório do produto entre as violações e o peso das restrições - que se encontra ao lado do *tableau* e não mais diretamente através do *ranking value* como na OT Estocástica, que tem poder eliminativo na escolha do candidato ótimo. Novamente, o algoritmo vai convergir, ou seja, vai atingir o estágio final da gramática, conforme podemos ver na indicação do candidato ótimo em cada um dos *tableaux*.

Os valores de Harmonia dos candidatos ótimos - [p.t] = 0, [k.t] = -10.337, [p.n] = -4.553, [k.n] = -54.986 -, são bastante superiores (e, portanto, mais próximos de zero) aos valores de Harmonia dos candidatos perdedores com epêntese - [pi.t, ki.t, pi.n, ki.n] = -78.546 - de forma a caracterizar que não está havendo variação nos *outputs*. Isso quer dizer que é categórica a produção da plosiva em coda.

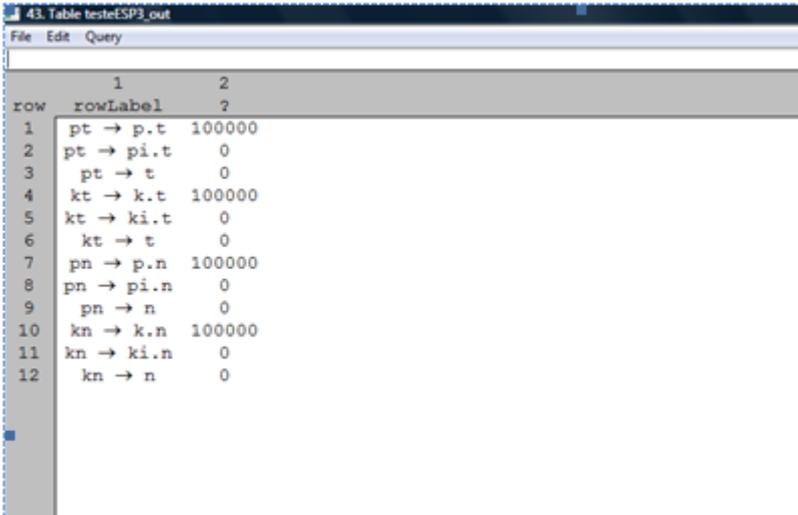
Porém, não podemos desconsiderar as diferenças nos valores numéricos dos pesos das restrições dados em cada modelo teórico e o comportamento das restrições conjuntas nesse modelo da HG. Os pesos das restrições promovidas nesse modelo - DEP (78.063) e MAX (78.110) - são superiores em comparação aos da OT Estocástica - DEP (60.270) e MAX (61.730), e os pesos das restrições demovidas -  $\ast\{k.n\}$ (42.384),  $\ast\{p.n\}$ (-0.138),  $\ast\{stop\}_{coda}$  (-56.173),  $\ast\{k.t\}$ (5.780),  $\ast\{p.t\}$ (-56.173),  $\ast\{k\}coda$ (5.780),  $Dist\{+1\}$ (-0.138),  $Dist\{+1,0\}$ (-56.173) e  $\ast\{p\}coda$ (-56.173) - são inferiores aos verificados no modelo anterior -  $\ast\{k.n\}$ (48.417),  $\ast\{p.n\}$ (24.193),  $\ast\{stop\}_{coda}$  (-22.001),  $\ast\{k.t\}$ (20.365),  $\ast\{p.t\}$ (-22.001),  $\ast\{k\}coda$ (20.365),  $Dist\{+1\}$ (24.193),  $Dist\{+1,0\}$ (-22.001) e  $\ast\{p\}coda$ (-22.001) - , para que se possa garantir o maior valor de Harmonia ao candidato ótimo.

Conforme já explicitado, o valor de Harmonia de cada candidato a *output* é resultado da soma do produto entre o número de violações e o peso de cada restrição. Logo, por exemplo, para [k.n], que tem um valor de Harmonia igual a -54.986, basta considerarmos todos os valores de *disharmony* dados para as restrições violadas por tal candidato, conforme podemos ver no *tableau 7*, referente a esta sequência. Ou seja, basta somarmos os valores das restrições:  $\ast\{k\}coda$ ,  $\ast\{k.n\}$ ,  $\ast\{k.t\}$ ,  $\ast\{Dist\{1\}\}$ ,  $\ast\{p.n\}$ ,  $\ast\{p\}coda$ ,  $\ast\{stop\}coda$ ,  $\ast\{p.t\}$ ,  $\ast\{Dist\{1,0\}\}$ . Devemos lembrar, ainda, que os valores negativos das restrições não contam no cálculo, conforme a opção que fizemos por utilizar o *linear OT*<sup>18</sup>.

<sup>18</sup> O *software PRAAT* oferece algumas possibilidades de versão de análise à luz da HG. Entretanto, conforme vemos no conjunto de *tableaux 7*, caso considerássemos os valores de ranqueamento negativo das restrições sagraríamos como ótimo candidatos que são limitados harmonicamente, ou seja, que nunca seriam ótimos na OT. Para resolver esse problema, a opção *linear OT* não considera, no cálculo de H, as restrições com valores centrais negativos.

Dessa forma, fica mais claro perceber por que é preciso promover DEP e MAX e demover o valor das restrições de marcação de forma mais acentuada do que acontece na OT Estocástica, para que possamos alcançar o candidato ótimo. Nesse modelo, DEP deve atingir um valor superior à soma de todas as restrições de marcação para que seja o candidato perdedor. A seguir, na simulação 2, que se refere à aquisição plena do Português, com 100% de epêntese, vamos ter uma situação que se apresentará oposta à aquisição do Espanhol no que se refere à hierarquia das restrições.

Na figura 14, a seguir, mostramos também para a HG o resultado do comando *output distributions* fornecido pelo programa, representando as probabilidades de ocorrência de cada uma das sequências dadas [p.t, k.t, p.n, k.n] nesse exercício teórico.



row	rowLabel	?
1	pt → p.t	100000
2	pt → pi.t	0
3	pt → t	0
4	kt → k.t	100000
5	kt → ki.t	0
6	kt → t	0
7	pn → p.n	100000
8	pn → pi.n	0
9	pn → n	0
10	kn → k.n	100000
11	kn → ki.n	0
12	kn → n	0

Figura 14 – *Output Distributions* da aquisição do Espanhol Colombiano – HG

Mais uma vez, podemos ver a probabilidade de 100% para a produção das plosivas em coda para a aquisição do Espanhol Colombiano, o que demonstra, a partir de tal confirmação, a acuidade dos valores fornecidos pelo algoritmo.

### 5.1.2 Simulação 2 - Aquisição do Português Brasileiro (L2)

a) A aquisição do Português Brasileiro (L2) via Teoria da Otimidade Estocástica

Apresentamos, neste item, a aquisição dos segmentos plosivos [k, p], em coda medial de palavra, com uma ocorrência de 100% de epêntese no Português Brasileiro, via Teoria da Otimidade Estocástica – modelo que trabalha com as restrições em ordem estrita. A simulação 2, realizada no *software* PRAAT, é dada no conjunto de *tableaux* 8, a seguir:

Conjunto de Tableaux 8 - Simulação 2 OT / Aquisição do Português Brasileiro (100% de epêntese)

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
Max	61.730	65.690	1.000000
*{k.n}	53.372	51.733	1.000000
*{p.n}	34.072	36.279	1.000000
*{k.t}	33.757	36.141	1.000000
*Dist {1}	34.072	35.240	1.000000
*{k}coda	33.757	30.266	1.000000
*{p.t}	23.134	26.555	1.000000
*Dist {1,0}	23.134	26.328	1.000000
*{stop}coda	23.134	22.678	1.000000
*{p}coda	23.134	20.816	1.000000
Dep	15.135	12.424	1.000000

pt	Max	*{k.n}	*{p.n}	*{k.t}	*Dist {1}	*{k}coda	*{p.t}	*Dist {1,0}	*{stop}coda	*{p}coda	Dep
p.t							!	*	*	*	
☞ pi.t											*
t	!										

kt	Max	*{k.n}	*{p.n}	*{k.t}	*Dist {1}	*{k}coda	*{p.t}	*Dist {1,0}	*{stop}coda	*{p}coda	Dep
k.t				!		*	*	*	*	*	
☞ ki.t											*
t	!										

pn	Max	*{k.n}	*{p.n}	*{k.t}	*Dist {1}	*{k}coda	*{p.t}	*Dist {1,0}	*{stop}coda	*{p}coda	Dep
p.n			!		*		*	*	*	*	
☞ pi.n											*
n	!										

kn	Max	*{k.n}	*{p.n}	*{k.t}	*Dist {1}	*{k}coda	*{p.t}	*Dist {1,0}	*{stop}coda	*{p}coda	Dep
k.n		!	*	*	*	*	*	*	*	*	
☞ ki.n											*
n	!										

Ao considerarmos que o primeiro estágio de aquisição de L2 é a gramática da L1, tanto nesta simulação com o algoritmo OT GLA, como com o algoritmo de aprendizagem da HG, teremos de alimentar o sistema com os valores de ranqueamento obtidos na simulação anterior. No caso da OT Estocástica, os valores obtidos na simulação 1, conforme pôde ser visto no conjunto de *tableaux* 6, foram os seguintes:

- MAX = 78.110; DEP = 78.063

- \*{k.n} = 42.384; \*{k}coda = 5.780; \*{k.t} = 5.780; \*Dist {1} = -0.138; \*{p.n} = -0.138; \*{p}coda = -56.173; \*{stop}coda = -56.173; \*{p.t} = -56.173 e \*Dist {1,0} = -56.173.

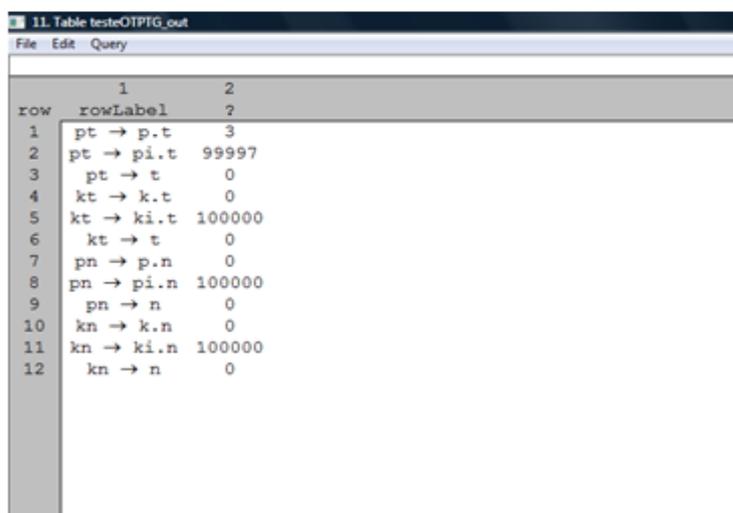
Informamos ao sistema que, partindo da língua materna dos aprendizes – sistema-alvo da simulação 1 – devemos obter uma gramática a partir da qual o fenômeno da epêntese ocorre em 100% das produções.

Para que o candidato ótimo com epêntese possa emergir a partir da noção de ranking estrito, nesta simulação 2, vamos verificar a demção de DEP. Esta restrição de Fidelidade vai ter seu valor diminuído pelo algoritmo de aprendizagem a cada exemplar de *input* que receber. Como o sistema está alimentado com dados que requerem um *output* categórico, com 100% de epêntese, DEP deve sofrer uma queda no seu peso, fazendo com que ocupe uma posição baixa na hierarquia.

Em contraposição à queda de DEP, vemos as restrições de marcação assumindo valores mais altos com relação à simulação anterior. Na simulação 1, vimos as restrições de marcação com valores tão baixos, que alguns chegavam a ser negativos, demonstrando com nitidez um processo de aquisição que viola com mais frequência as restrições que dizem respeito às sequências menos marcadas da hierarquia, graças às restrições estringentes e conjuntas que acreditamos ser fundamentais nesse processo de aquisição.

Nesta segunda simulação, essa hierarquia deve ser modificada, porque neste momento nosso candidato ótimo deve violar apenas DEP, como podemos ver nos conjuntos de *tableaux*. Por sua vez, as restrições conjuntas e estringentes devem manter a relação de marcação dos componentes primitivos a partir da escala que as originou. As restrições que se referem a sequências menos marcadas devem, também aqui, apresentar uma ordem que considere que as restrições referentes à sequência menos marcada [p.t] - \*Dist{1,0}, \*{Stop}<sub>coda</sub>, \*{p}<sub>coda</sub>, \*{p.t}/[\*Dist{1,0} & \*{k,p}<sub>coda</sub>] -, com relação às demais sequências dadas, sejam as mais violadas. Logo, tais restrições devem vir em seguida de DEP no ranking. Ainda que a cada momento de avaliação dos candidatos (ou seja, momentos de produção linguística) as restrições possam assumir diferentes pesos – pontos de seleção determinados por um ruído estatístico do algoritmo computacional utilizado –, os nossos *outputs* ótimos devem ser sempre os mesmos, uma vez que DEP é sempre suficientemente inferior às restrições de marcação.

A figura 15, apresentada a seguir, tem o intuito de demonstrar para esta simulação da aquisição do PB, por parte dos estudantes colombianos, na OT, se a probabilidade de ocorrências está condizente com as porcentagens com as quais o algoritmo convergiu.



row	rowLabel	1	2
1	pt → p.t	3	
2	pt → pi.t	99997	
3	pt → t	0	
4	kt → k.t	0	
5	kt → ki.t	100000	
6	kt → t	0	
7	pn → p.n	0	
8	pn → pi.n	100000	
9	pn → n	0	
10	kn → k.n	0	
11	kn → ki.n	100000	
12	kn → n	0	

Figura 15 – *Output Distributions* da aquisição do Português Brasileiro – OT

Mais uma vez, os índices de porcentagem de ocorrência dos candidatos a *output* é condizente com os valores dados pelo algoritmo. Contudo, observamos que, apesar de o *output distributions* ter apresentado valor igual a 99.997 de probabilidade de ocorrência para a sequência [pi.t], estatisticamente, podemos considerar essa diferença para 100.000 sem significância, ou seja, podemos considerar que, também para essa sequência, o percentual de ocorrência de epêntese foi igual a 100%.

b) A aquisição do Português Brasileiro (L2) via Gramática Harmônica

Seguindo a organização do estudo, o conjunto de *tableaux* 9, a seguir, continuará tratando da aquisição do Português Brasileiro, porém, no modelo da Gramática Harmônica, demonstrando um candidato ótimo que vai emergir através da soma dos pesos das restrições.

Conjunto de Tableaux 9 - Simulação 2 HG / Aquisição do Português Brasileiro (100% de epêntese)

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
Max	78.110	77.599	1.000000
*{k.n}	43.248	42.829	1.000000
*{k}coda	20.889	21.993	1.000000
*{k.t}	20.889	19.830	1.000000
*{p.n}	17.783	18.128	1.000000
*Dist {1}	17.783	16.766	1.000000
Dep	13.237	11.903	1.000000
*{stop}coda	8.653	9.067	1.000000
*{p}coda	8.653	8.359	1.000000
*{p.t}	8.653	7.558	1.000000
*Dist {1,0}	8.653	7.188	1.000000

pt	Max	*{k.n}	*{k}coda	*{k.t}	*{p.n}	*Dist {1}	Dep	*{stop}coda	*{p}coda	*{p.t}	*Dist {1,0}	
p.t								*	*	*	*	-32.172
☞ pi.t							*					-11.903
t	*											-77.599

kt	Max	*{k.n}	*{k}coda	*{k.t}	*{p.n}	*Dist {1}	Dep	*{stop}coda	*{p}coda	*{p.t}	*Dist {1,0}	
k.t			*	*				*	*	*	*	-73.995
☞ ki.t							*					-11.903
t	*											-77.599

pn	Max	*{k.n}	*{k}coda	*{k.t}	*{p.n}	*Dist {1}	Dep	*{stop}coda	*{p}coda	*{p.t}	*Dist {1,0}	
p.n					*	*		*	*	*	*	-67.067
☞ pi.n							*					-11.903
n	*											-77.599

kn	Max	*{k.n}	*{k}coda	*{k.t}	*{p.n}	*Dist {1}	Dep	*{stop}coda	*{p}coda	*{p.t}	*Dist {1,0}	
k.n		*	*	*	*	*		*	*	*	*	-151.719
☞ ki.n							*					-11.903
n	*											-77.599

Para esta simulação do sistema do Português à luz da HG, alimentamos o algoritmo com as mesmas informações fornecidas ao algoritmo OT-GLA. O algoritmo foi informado de que, no sistema alvo, *inputs* com plosivas finais devem ser epentetizados e que o estágio inicial da aquisição são os valores do sistema-alvo da simulação 1 – aquisição do Espanhol Colombiano. Devemos considerar, nesse sentido, que os pesos das restrições resultantes da simulação 1 são diferentes para os dois modelos teóricos. Na simulação 1 da

HG – Aquisição da língua materna/EC (conjunto de *tableaux* 7), encontramos os seguintes valores:

DEP – 78.063, MAX – 78.110

\*{k.n} – 42.384, \*{k}<sub>coda</sub> – 5.780, \*{k.t} – 5.780, \*Dist{1} – - 0.138, \*{p.n} – - 0.138, \*{p}<sub>coda</sub> – -56.173, \*{stop}<sub>coda</sub> – -56.173, \*{p.t} – -56.173, \*Dist{1,0} – -56.173

No que diz respeito à gramática resultante do processo de aquisição, é interessante observar no *ranking value* que DEP não apresenta mais o menor valor de todas as restrições, como aconteceu na OT Estocástica. Entretanto, tal restrição apresenta o menor valor de Harmonia com relação aos candidatos perdedores, que é o que nos importa para a seleção do candidato ótimo, neste modelo. Além disso, os candidatos perdedores, com a produção da plosiva em coda, também nos evidenciam outra situação, a de que mantêm valores de Harmonia que preservam a relação de marcação entre as restrições. A situação é evidente porque a própria formalização das restrições por estringência garante que essas relações sejam preservadas.

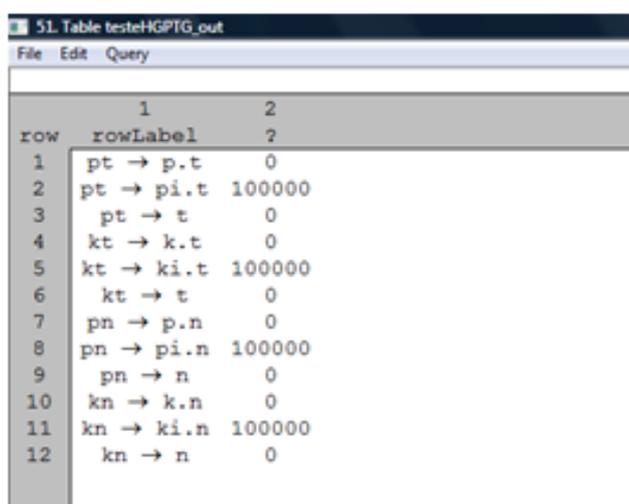
Conforme discutido, a escolha do candidato ótimo, neste modelo, vai se dar a partir do valor de Harmonia das restrições. Assim, no *tableau* referente à sequência [p.t], por exemplo, o candidato [pi.t] sagra-se ótimo pelo fato de seu valor de H (-11.903) ser o mais próximo de zero. O candidato perdedor [p.t], por sua vez, apresenta um valor de H (-32.172), equivalente à soma dos valores de ponto de seleção assumidos, naquele momento de avaliação específico, por cada uma das restrições violadas pelo candidato ((-9.067) + (-8.359) + (-7.558) + (-7.188)). Finalmente, o candidato [t] é o que apresenta valor de Harmonia mais baixo ou mais negativo (-77.599), valor obtido através da multiplicação do valor do ponto de seleção da restrição MAX, por -1 violação.

Uma outra comparação interessante entre os dois modelos diz respeito aos valores das restrições apresentadas por cada um dos algoritmos. Assim como ocorreu na Simulação 1, novamente as promoções e demissões dos valores das restrições ocorrem de forma mais acentuada na HG. A razão para tal situação advém do caráter cumulativo que rege a análise na HG. Não basta que DEP seja superior a qualquer restrição de marcação que proíba plosiva em coda. É preciso que a soma dos valores dos pontos de seleção de todas as restrições de marcação estridentes seja inferior ao valor do ponto de seleção a ser assumido por DEP.

Em suma, tanto o algoritmo de aprendizagem vinculado à OT-GLA quanto o algoritmo da HG-GLA conseguem convergir em gramáticas categóricas, que expressam o sistema-alvo a ser atingido pelos aprendizes. Dessa forma, acreditamos ter explicado a diferença entre os índices numéricos assumidos por DEP à luz dos dois algoritmos. Sob a OT-

GLA, na Simulação 2 (conjunto de *tableaux* 8), a restrição de Fidelidade em questão apresentou valor igual a 15.135, valor esse que já era bem afastado e inferior ao das restrições de marcação. À luz da HG, foi necessário que este valor de Harmonia fosse superior ao mais alto valor da soma dos pontos de seleção a serem assumidos pelas restrições de marcação violadas pelo candidato perdedor.

A seguir, através da figura 15, apresentamos os *output distributions* para a aquisição do PB na HG.



row	1	2
1	pt → p.t	0
2	pt → pi.t	100000
3	pt → t	0
4	kt → k.t	0
5	kt → ki.t	100000
6	kt → t	0
7	pn → p.n	0
8	pn → pi.n	100000
9	pn → n	0
10	kn → k.n	0
11	kn → ki.n	100000
12	kn → n	0

Figura 16 – *Output Distributions* da aquisição do Português Brasileiro – HG

Fazendo a prova real, com a submissão dos candidatos a 100.000 avaliações, vemos que os índices de porcentagem de 100%, para a produção de epêntese no PB, foram também confirmados pelo comando *output distributions*.

A última simulação, a seguir, para o Estudo-Piloto, vai tratar dos *outputs* variáveis apresentados pelos estudantes colombianos no processo de aquisição do Português Brasileiro.

### 5.1.3 Simulação 3 – Interlândia (*outputs* variáveis)

a) A aquisição dos *outputs* variáveis (Interlândia) via Teoria da Otimidade Estocástica

Apresentamos, neste item, a aquisição dos segmentos plosivos [k, p], em coda medial de palavra, com uma ocorrência variável de epêntese no Português Brasileiro, via Teoria da Otimidade Estocástica. A simulação 3, realizada no *software* PRAAT, é dada no conjunto de *tableaux* 10, a seguir:

Conjunto de Tableaux 10 - Simulação 3 OT / Interlúngua (*outputs* variáveis)

			<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>							
	<b>Max</b>		61.730	63.113	1.000000							
	<b>Dep</b>		44.045	45.690	1.000000							
	<b>*Dist {1}</b>		40.418	42.521	1.000000							
	<b>*{k.n}</b>		40.592	39.719	1.000000							
	<b>*{p.n}</b>		40.418	37.146	1.000000							
	<b>*{k.t}</b>		12.540	16.245	1.000000							
	<b>*{k}coda</b>		12.540	16.154	1.000000							
	<b>*Dist {1,0}</b>		-5.776	-3.954	1.000000							
	<b>*{p}coda</b>		-5.776	-4.254	1.000000							
	<b>*{stop}coda</b>		-5.776	-4.383	1.000000							
	<b>*{p.t}</b>		-5.776	-6.410	1.000000							

pt	Max	Dep	*Dist {1}	*{k.n}	*{p.n}	*{k.t}	*{k}coda	*Dist {1,0}	*{p}coda	*{stop}coda	*{p.t}
☞ p.t								*	*	*	*
pi.t		*!									
t	*!										

kt	Max	Dep	*Dist {1}	*{k.n}	*{p.n}	*{k.t}	*{k}coda	*Dist {1,0}	*{p}coda	*{stop}coda	*{p.t}
☞ k.t						*	*	*	*	*	*
ki.t		*!									
t	*!										

pn	Max	Dep	*Dist {1}	*{k.n}	*{p.n}	*{k.t}	*{k}coda	*Dist {1,0}	*{p}coda	*{stop}coda	*{p.t}
☞ p.n			*		*			*	*	*	*
pi.n		*!									
n	*!										

kn	Max	Dep	*Dist {1}	*{k.n}	*{p.n}	*{k.t}	*{k}coda	*Dist {1,0}	*{p}coda	*{stop}coda	*{p.t}
☞ k.n			*	*	*	*	*	*	*	*	*
ki.n		*!									
n	*!										

Nesta simulação, lidamos com *outputs* variáveis, que refletirão a gramática em desenvolvimento dos nossos aprendizes de PB. Diferentemente das Simulações 1 e 2, os valores centrais das restrições, neste momento, são muito próximos. Assim, as alterações nos valores numéricos em momentos de produção linguística constituirão uma fonte de motivação para que tenhamos mudança no ranking das restrições.

Voltando os olhos à tabela 24, podemos ver que, na sequência [p.n], tivemos 83% de produção sem epêntese, contra 17% com epêntese. Então, a probabilidade de termos candidatos sem epêntese é muito maior do que a possibilidade de candidatos com epêntese. Se ficarmos realizando diversas avaliações sucessivas, em algum momento devemos testemunhar a presença de um candidato ótimo com epêntese, conforme podemos ver no conjunto de *tableaux* 11, a seguir.

Conjunto de Tableaux 11 - Simulação 3 OT / Interlúngua –Segunda rodada

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
Max	61.730	63.023	1.000000
*{p.n}	40.418	45.048	1.000000
Dep	44.045	44.377	1.000000
*Dist {1}	40.418	41.167	1.000000
*{k.n}	40.592	37.662	1.000000
*{k}coda	12.540	13.542	1.000000
*{k.t}	12.540	13.448	1.000000
*{p.t}	-5.776	-3.423	1.000000
*Dist {1,0}	-5.776	-4.098	1.000000
*{p}coda	-5.776	-7.136	1.000000
*{stop}coda	-5.776	-9.215	1.000000

pt	Max	*{p.n}	Dep	*Dist {1}	*{k.n}	*{k}coda	*{k.t}	*{p.t}	*Dist {1,0}	*{p}coda	*{stop}coda
☞ p.t								*	*	*	*
pi.t			*!								
t	*!										

kt	Max	*{p.n}	Dep	*Dist {1}	*{k.n}	*{k}coda	*{k.t}	*{p.t}	*Dist {1,0}	*{p}coda	*{stop}coda
☞ k.t						*	*	*	*	*	*
ki.t			*!								
t	*!										

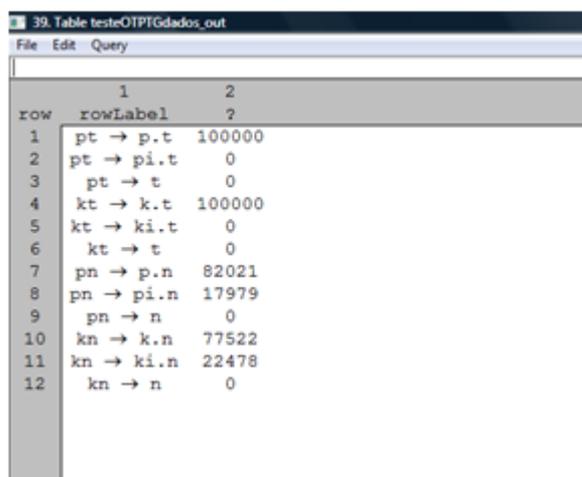
pn	Max	*{p.n}	Dep	*Dist {1}	*{k.n}	*{k}coda	*{k.t}	*{p.t}	*Dist {1,0}	*{p}coda	*{stop}coda
p.n		*!		*				*	*	*	*
☞ pi.n			*								
n	*!										

kn	Max	*{p.n}	Dep	*Dist {1}	*{k.n}	*{k}coda	*{k.t}	*{p.t}	*Dist {1,0}	*{p}coda	*{stop}coda
k.n		*!		*	*	*	*	*	*	*	*
☞ ki.n			*								
n	*!										

Na primeira rodada (conjunto de *tableaux* 10), tínhamos uma hierarquia que apresentava DEP acima de todas as restrições de marcação, mostrando que a plosiva em coda estava emergindo sob os demais candidatos a *output*, em um modelo que seleciona o candidato ótimo via *ranking* estrito. Na segunda rodada, tentamos demonstrar como se dá esta alteração da configuração hierárquica. DEP está abaixo de *\*{p.n}* no conjunto de *tableaux* 11, disso resultando a emergência dos *outputs* das sequências [p.n] e [k.n] com epêntese.

Assim como foi esperado, a possibilidade de *outputs* variáveis ocorreu somente com as sequências [p.n] e [k.n]. No momento em que tivemos 0% de epêntese nas sequências [p.t] e [k.t], não poderíamos ter valores de ranqueamento próximos entre DEP (44.045) e as restrições relacionadas às sequências com *plateau* *\*{k}coda* (12.540), *\*{k.t}* (12.540), *\*Dist{1,0}* (-5.776), *\*{p.t}* (-5.776), *\*{stop}coda* (-5.776), *\*{p}coda* (-5.776).

A figura 17 mostra a distribuição probabilística dos dados, para a interlúngua dos estudantes colombianos na OT, para que possamos ver se as probabilidades dos candidatos a *output* estão condizentes com os valores dados pelo algoritmo OT-GLA a partir dos conjuntos de *tableaux* desta simulação.



row	rowLabel	?
1	pt → p.t	100000
2	pt → pi.t	0
3	pt → t	0
4	kt → k.t	100000
5	kt → ki.t	0
6	kt → t	0
7	pn → p.n	82021
8	pn → pi.n	17979
9	pn → n	0
10	kn → k.n	77522
11	kn → ki.n	22478
12	kn → n	0

**Figura 17 - Output Distributions da Interlíngua – OT**

Mais uma vez, podemos confirmar que, após a submissão dos candidatos a 100.000 avaliações, a probabilidade de ocorrência de cada um deles é perfeitamente compatível com as porcentagens para as quais o algoritmo deveria convergir. De acordo com os dados da tabela 24, que utilizamos para alimentar o sistema, tivemos:

$$[p.t] = 100\%$$

$$[k.t] = 100\%$$

$$[p.n] = 83\% \qquad [pi.n] = 17,979\%$$

$$[k.n] = 77,27\% \qquad [ki.n] = 22,73\%$$

Podemos afirmar, assim, que os valores do *output distributions* estão condizentes com os valores fornecidos pelo algoritmo.

b) A aquisição dos *outputs* variáveis (Interlíngua) via Gramática Harmônica

Conforme os dados levantados na análise acústica, a probabilidade de ocorrência de produção de epêntese foi menor que a ocorrência das produções da consoante plosiva em condição de coda medial de palavra. Logo, o algoritmo deve refletir esta porcentagem, apresentando as variações condizentes com os dados de produção que extraímos da análise acústica das produções dos aprendizes. A seguir, ainda na Simulação 3 da aquisição dos *outputs* variáveis, vamos descrever o comportamento do algoritmo vinculado ao modelo da HG, com os índices percentuais de produção, que servirão como alvo, e verificar se, assim como ocorreu na OT Estocástica, o algoritmo HG vai convergir.

Conjunto de Tableaux 12 - Simulação 3 HG / Interlúngua (*outputs* variáveis)

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
<b>Max</b>	78.110	78.313	1.000000
<b>Dep</b>	53.034	53.344	1.000000
<b>*Dist {1}</b>	24.891	23.910	1.000000
<b>*{p.n}</b>	24.891	23.441	1.000000
<b>*{k.n}</b>	-0.375	2.813	1.000000
<b>*Dist {1,0}</b>	-31.144	-29.816	1.000000
<b>*{stop}coda</b>	-31.144	-30.492	1.000000
<b>*{p}coda</b>	-31.144	-30.575	1.000000
<b>*{p.t}</b>	-31.144	-31.009	1.000000
<b>*{k}coda</b>	-36.979	-33.871	1.000000
<b>*{k.t}</b>	-36.979	-37.149	1.000000

pt	Max	Dep	*Dist {1}	*{p.n}	*{k.n}	*Dist {1,0}	*{stop}coda	*{p}coda	*{p.t}	*{k}coda	*{k.t}	
☞ p.t						*	*	*	*			0
pit		*										-53.344
t	*											-78.313

kt	Max	Dep	*Dist {1}	*{p.n}	*{k.n}	*Dist {1,0}	*{stop}coda	*{p}coda	*{p.t}	*{k}coda	*{k.t}	
☞ kt						*	*	*	*	*	*	0
kit		*										-53.344
t	*											-78.313

pn	Max	Dep	*Dist {1}	*{p.n}	*{k.n}	*Dist {1,0}	*{stop}coda	*{p}coda	*{p.t}	*{k}coda	*{k.t}	
☞ p.n			*	*		*	*	*	*			-47.351
pi.n		*										-53.344
n	*											-78.313

kn	Max	Dep	*Dist {1}	*{p.n}	*{k.n}	*Dist {1,0}	*{stop}coda	*{p}coda	*{p.t}	*{k}coda	*{k.t}	
☞ k.n			*	*	*	*	*	*	*	*	*	-50.164
ki.n		*										-53.344
n	*											-78.313

Dados os valores de Harmonia no conjunto de *tableaux* 12, vemos que o algoritmo consegue convergir para uma gramática que expressa os índices percentuais de 83% e 17% para a sequência [p.n] sem epêntese e com epêntese, respectivamente, e 78% e 22% para a sequência [k.n], sem epêntese e com epêntese, respectivamente. Ou seja, DEP assume valores de Harmonia muito próximos aos candidatos vencedores [p.n] e [k.n], demonstrando o caráter variável da gramática no que diz respeito a essas duas sequências, conforme vemos, também, no conjunto de *tableaux* 13, a seguir.

Conjunto de Tableaux 13 - Simulação 3 HG / Interlúngua – Segunda rodada

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>	
Max	78.110	79.632	1.000000	
Dep	53.034	53.084	1.000000	
*Dist {1}	24.891	28.376	1.000000	
*{p.n}	24.891	27.057	1.000000	
*{k.n}	-0.375	-0.819	1.000000	
*{stop}coda	-31.144	-30.282	1.000000	
*{p.t}	-31.144	-31.158	1.000000	
*{p}coda	-31.144	-32.273	1.000000	
*{k}coda	-36.979	-35.145	1.000000	
*Dist {1,0}	-31.144	-35.251	1.000000	
*{k.t}	-36.979	-38.194	1.000000	

pt	Max	Dep	*Dist {1}	*{p.n}	*{k.n}	*{stop}coda	*{p.t}	*{p}coda	*{k}coda	*Dist {1,0}	*{k.t}	
☞ p.t						*	*	*		*		0
pi.t		*										-53.084
t	*											-79.632

kt	Max	Dep	*Dist {1}	*{p.n}	*{k.n}	*{stop}coda	*{p.t}	*{p}coda	*{k}coda	*Dist {1,0}	*{k.t}	
☞ k.t						*	*	*	*	*	*	0
ki.t		*										-53.084
t	*											-79.632

pn	Max	Dep	*Dist {1}	*{p.n}	*{k.n}	*{stop}coda	*{p.t}	*{p}coda	*{k}coda	*Dist {1,0}	*{k.t}	
p.n			*	*		*	*	*		*		-55.433
☞ pi.n		*										-53.084
n	*											-79.632

kn	Max	Dep	*Dist {1}	*{p.n}	*{k.n}	*{stop}coda	*{p.t}	*{p}coda	*{k}coda	*Dist {1,0}	*{k.t}	
k.n			*	*	*	*	*	*	*	*	*	-55.433
☞ ki.n		*										-53.084
n	*											-79.632

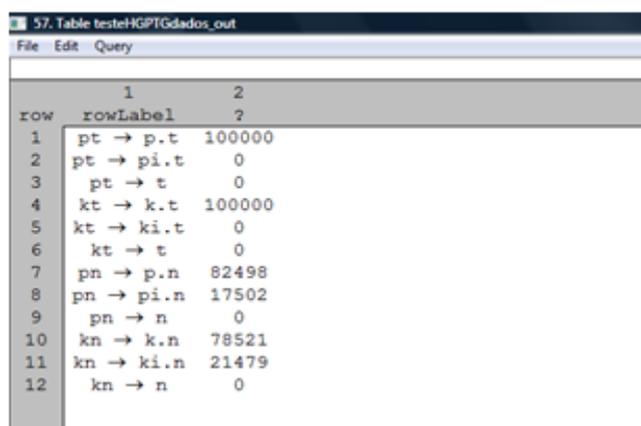
Ao compararmos o conjunto de *tableaux* 12 com estes valores da segunda rodada (conjunto de *tableaux* 13), teremos DEP, responsável pelo valor de Harmonia das sequências [pi.n] e [ki.n], com valor igual a -53.084 (-1 violação x valor *disharmony* de DEP). Por sua vez, os valores de [p.n] (-55.433) e [k.n] (-55.433) – resultados da soma do produto dos valores *disharmony* de cada restrição violada, pelo número de violações com valor - são, neste momento, maiores do que o de DEP, ou mais próximos a zero, permitindo que a epêntese ocorra. Devemos lembrar, mais uma vez, que as restrições com valor negativo são eliminadas do cálculo.

Observando os índices numéricos do conjunto de *tableaux* 13 ou do conjunto de *tableaux* 12, ressaltamos que, ainda que o valor central de DEP (53.034) se mostre distante e seja superior ao valor de todas as restrições de marcação, ou seja, com uma distância superior a 10 pontos, verificada através da soma dos pontos de seleção das restrições violadas, teremos valores de Harmonia próximos aos candidatos com possibilidade de variação no *output*. Vimos, anteriormente, que essa condição referente aos valores das restrições implicaria de imediato, à luz da OT-GLA, ausência de epêntese.

Além disso, é importante considerar que, para [p.t] e [k.t], que não apresentaram epêntese nas produções dos aprendizes, não tivemos proximidade nos valores de Harmonia entre tais candidatos vencedores ([p.t] e [k.t]) e os candidatos perdedores ([pi.t], [ki.t], [t]), demonstrando *outputs* categóricos para estas sequências. Dessa forma, visualizamos a capacidade do modelo em lidar com estas duas situações.

Portanto, ambos os algoritmos convergiram com essas restrições ao estágio final de gramática considerado nas três simulações realizadas. Porém, o resultado final das gramáticas de cada algoritmo apresenta características próprias ao modo como cada um deles seleciona o *output* ótimo. Esse exercício teórico foi fundamental à etapa final deste trabalho, que vai se utilizar das sequências [p.n, p.t, k.n, k.t, t.n], conforme estabelecido nos objetivos descritos na Introdução.

Trazemos a figura 18, a seguir, para comprovar que a gramática em questão resulta nos percentuais desejados de [p.n]=83%, [pi.n]=17%, [k.n]=77,27%, [ki.n]=22,73% e [k.t]/[p.t]=100%.



row	rowLabel	?
1	pt → p.t	100000
2	pt → pi.t	0
3	pt → t	0
4	kt → k.t	100000
5	kt → ki.t	0
6	kt → t	0
7	pn → p.n	82498
8	pn → pi.n	17502
9	pn → n	0
10	kn → k.n	78521
11	kn → ki.n	21479
12	kn → n	0

**Figura 18 – Output Distributions da Interlíngua – HG**

Feitas essas considerações sobre o Estudo-Piloto, partimos para a análise final, que se dará em duas partes.

## 5.2 A FORMALIZAÇÃO DA HIERARQUIA DE RESTRIÇÕES PARA AS SEQUÊNCIAS [p.n, k.n, t.n, p.t, k.t]: A ANÁLISE FINAL

Conforme justificamos na seção anterior, apesar das onze sequências de segmentos plosivos em coda levantadas por esta pesquisa, apenas cinco delas serão tomadas

para a análise final [p.n, k.n, t.n, p.t, k.t] à luz da Teoria da Otimidade Estocástica e da Gramática Harmônica. No que segue, refletiremos sobre a escolha das restrições que servirão para compor a análise e a resposta de cada modelo ao processo de aquisição do PB pelos informantes colombianos.

Antes de iniciarmos a reflexão sobre as restrições, adiantamos que, para este grupo de sequência de segmentos [p.n, k.n, t.n, p.t, k.t], trabalhamos com a hipótese de que o contato silábico está influenciando na epêntese pelos aprendizes de Português Brasileiro. Logo, a tabela dos dados abaixo, bem como a discussão das restrições a seguir, pretendem refletir essa ideia.

No que segue, apresentamos um resumo das tabelas 3, 4, 7, 8 e 11 (Espanhol Colombiano) e 14, 15, 18, 19 e 22 (Português Brasileiro), referentes, respectivamente, às sequências [p.t, p.n, k.t, k.n, t.n].

**Tabela 25 – Produção das sequências [p.n, k.n, t.n, p.t, k.t] no EC**

Consoante em coda	Epêntese	Apagamento
Consoante em coda	Epêntese	Apagamento total
[k.n, p.n, t.n]	[ki.n, pi.n, ti.n]	[∅.n]
100%	0	0
Consoante em coda	Epêntese	Apagamento total
[k.t, p.t]	[ki.t, pi.t]	[∅.t]
100%	0	0

**Tabela 26 - Produção das sequências [p.n, k.n, t.n, p.t, k.t] no PB**

Consoante em coda	Epêntese	Apagamento
Consoante em coda	Epêntese	Apagamento total
[k.n, p.n, t.n]	[ki.n, pi.n, ti.n]	[∅.n]
86,52%	13,48%	0
Consoante em coda	Epêntese	Apagamento total
[k.t, p.t]	[ki.t, pi.t]	[∅.t]
98,96%	1,04%	0

Os dados das tabelas 25 e 26, assim, servirão para as análises dos dados via Teoria da Otimidade Estocástica e Gramática Harmônica, que serão realizadas nas seções 5.2.1 – Conjunto Restrito de Restrições e 5.2.2 – Conjunto Completo de Restrições, a seguir.

### 5.2.1 Análise dos dados via Teoria da Otimidade Estocástica e Gramática Harmônica: Conjunto Restrito de Restrições

Conforme justificado na metodologia, trabalhamos com DEP e MAX, como restrições de fidelidade, para que atuem na busca de *outputs* mais fieis.

Já dissemos que, para a formalização da marcação da plosiva em coda, faremos uso de restrições de marcação que reflitam os efeitos do contato silábico. Ao afirmarmos que o contato silábico é o responsável pela ocorrência de epêntese nas sequências formadas pela consoante plosiva em coda, devemos perceber que o ranking das restrições de coda não é o mesmo de *onset*. A coda vai privilegiar segmentos com valores mais altos na escala de sonoridade, enquanto que o *onset* vai preferir os mais baixos na escala em questão.

Prince & Smolensky (1993) propõem o Alinhamento Harmônico, como forma de obtermos um ranking de restrições que podem expressar quais segmentos que, a partir da escala de sonoridade, vão se mostrar mais aptos em dada posição silábica, para trabalharmos a emergência de segmentos plosivos em posição de coda.

Além disso, optamos por utilizar restrições da família “Dist”, propostas em Gouskova (2004), que, através da escala de Clements (1990), expressa a distância de sonoridade entre as duas consoantes que compõem as sequências heterossilábicas que vamos utilizar nesta análise, como forma de contemplar as ocorrências de epêntese por consequência de contato silábico.

No Estudo-Piloto que realizamos, quando utilizamos a escala de sonoridade de Clements (1990): obstruinte – nasal – líquida – glide – vogal, conforme estabelecido no Mecanismo de Alinhamento Relacional de Gouskova (2004), não estávamos fazendo uma distinção importante no que se refere à posição da obstruinte em coda no Português Brasileiro. Dizer que no contato silábico do PB é proibida a distância entre duas obstruintes (dist 0) ou entre uma obstruinte e uma nasal seguinte (dist +1) é afirmar que toda fricativa em coda, seguida de plosiva ou de nasal em *onset*, também é proibida. Dessa forma, estaríamos excluindo a possibilidade de termos palavras como “cas.ta” ou “as.no” no PB, sem a ocorrência de correção por epêntese.

De forma a refletirmos esta situação nas restrições que comporão esta análise com um conjunto restrito de restrições, acrescentamos aos nossos candidatos a *output* as sequências [s.n] e [s.t], que, no PB, não são passíveis de epêntese, e procuramos diferenciá-las das sequências heterossilábicas [p.t, k.t] e [p.n, k.n, t.n], através da ação de restrições conjuntas e

de restrições de proibição a plosivas em coda, que construímos para se oporem apenas às sequências com plosiva em coda. Por exemplo, utilizamos a restrição  $\ast\{\text{stop}\}_{\text{coda}}$ , que vai ser responsável pela demorção apenas das sequências [p.t, k.t, p.n, k.n, t.n], em detrimento de [s.n, s.t], para estabelecer as diferenças dessas sequências para determinada hierarquia de restrições.

Dessa forma, as restrições que utilizamos nesta análise final – Conjunto Restrito de Restrições, foram:

#### **RESTRICÇÕES DE FIDELIDADE**

DEP

MAX

#### **RESTRICÇÕES DE MARCAÇÃO**

$\ast\text{Dist}\{1\}$

$\ast\text{Dist}\{1,0\}$

$\ast\{\text{stop}\}_{\text{coda}}$ <sup>19</sup>

$\ast\text{Dist}\{1,0\} \& \{\text{stop}\}_{\text{coda}}$

$\ast\text{Dist}\{1\} \& \{\text{stop}\}_{\text{coda}}$

Assim, para dar conta da aquisição de segmentos plosivos em posição de coda medial de palavras no EC e no PB, contamos com duas restrições de fidelidade, que trabalharão para que possamos garantir *outputs* mais fiéis: DEP – oposição à ocorrência de epêntese - e MAX – oposição à ocorrência de apagamento. Contamos também com cinco restrições de marcação, que trabalharão de forma a se oporem a *outputs* não permitidos em dada língua:

$\ast\text{Dist}\{1\}$  e  $\ast\text{Dist}\{1,0\}$  – Oposição ao acréscimo de sonoridade do segmento em coda em relação ao segmento em posição de *onset* da sílaba seguinte e ao *plateau* silábico; tais restrições são utilizadas para diferenciarmos as distâncias silábicas entre as sequências utilizadas na análise, já que estamos considerando esse um fator importante à ocorrência da epêntese nas produções dos aprendizes colombianos de Português Brasileiro.

$\ast\{\text{stop}\}_{\text{coda}}$  – oposição à plosiva em coda; tal restrição é necessária para, além de representar oposição à ocorrência de um segmento indesejado em posição de coda no PB, permitir que façamos a distinção entre a ocorrência de tal segmento e a produção da fricativa [s] em coda, que é possível na língua.

<sup>19</sup> Na próxima seção (5.2.2) de análise – Conjunto Completo de Restrições, todos os membros da escala serão utilizados, mesmo que, a princípio, tais restrições pareçam desnecessárias para a análise.

\*Dist {1,0} & \*{stop}<sub>coda</sub> e \*Dist {1} & \*{stop}<sub>coda</sub> – restrições conjuntas, de caráter estrigente; frente à ideia de restrições de marcação de impor exigências na formação do nosso *output*, propiciando que as formas de saída respeitem uma hierarquia de marcação, utilizamos as restrições conjuntas para diferenciar contato silábico (\*Dist{1,0}) e modo de articulação (\*{stop}<sub>coda</sub>) e definir a hierarquia de marcação (“[plosiva].[nasal]” [p.n, k.n, t.n]) > (“[plosiva].[plosiva]” [p.t, k.t]) > [s.n] > [s.t].

Sabemos que a restrição que proíbe plosivas em coda (\*{stop}<sub>coda</sub>), da forma como sugerimos as restrições, exercerá papel redundante junto à restrição \*Dist{1,0} & \*{stop}<sub>coda</sub>, pois ambas serão responsáveis pela violação das mesmas sequências “t.t”(sequência de plosivas) e “t.n” (sequência de plosivas mais nasais). Porém, a restrição \*{stop}<sub>coda</sub> é importante para que possamos formar as restrições conjuntas \*Dist {1} & \*{stop}<sub>coda</sub> e \*Dist {1,0} & \*{stop}<sub>coda</sub>. Em termos de simulação computacional, isto deve implicar apenas valores iguais entre as restrições redundantes, conforme veremos a seguir nas 3 simulações realizadas para cada um dos sistemas.

### **5.2.1.1 Simulação 1 - Aquisição do Espanhol Colombiano (L1)**

a) A aquisição do Espanhol Colombiano (L1) via Teoria da Otimidade Estocástica

A simulação 1 constitui-se numa etapa fundamental para o processo de aquisição da L2 – que será realizado na simulação posterior –, já que é preciso mapear também o estágio inicial do processo de aquisição de língua materna. Assim como realizamos no Estudo-Piloto, esta etapa deve ocorrer nas duas análises finais (Conjunto Restrito de Restrições e Conjunto Completo de Restrições) que propusemos neste capítulo, nas simulações realizadas sob ambos os algoritmos de aprendizagem.

O sistema, neste momento, foi programado a responder a um estágio inicial de um bebê adquirindo o Espanhol Colombiano, conforme o conjunto de *tableaux* 14, a seguir.

Conjunto de *Tableaux* 14 - Simulação 1 OT / Aquisição do Espanhol Colombiano (0% de epêntese)

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
<b>Dep</b>	60.908	62.204	1.000000
<b>Max</b>	58.945	59.287	1.000000
<b>*Dist {1} &amp; *{stop}coda</b>	46.969	47.083	1.000000
<b>*Dist {1}</b>	21.091	22.269	1.000000
<b>Dist {1,0} &amp; *{stop}coda</b>	21.209	19.811	1.000000
<b>*{stop}coda</b>	21.209	18.952	1.000000
<b>*Dist {1,0}</b>	-19.852	-16.269	1.000000

tt	Dep	Max	*Dist {1} & *{stop}coda	*Dist {1}	*Dist {1,0} & *{stop}coda	*{stop}coda	*Dist {1,0}
☞ p.t/k.t					*	*	*
pi.t/ki.t	*!						
t		*!					

tn	Dep	Max	*Dist {1} & *{stop}coda	*Dist {1}	*Dist {1,0} & *{stop}coda	*{stop}coda	*Dist {1,0}
☞ p.n/k.n/t.n			*	*	*	*	*
pi.n/ki.n/ti.n	*!						
n		*!					

st	Dep	Max	*Dist {1} & *{stop}coda	*Dist {1}	*Dist {1,0} & *{stop}coda	*{stop}coda	*Dist {1,0}
☞ s.t							*
si.t	*!						
t		*!					

sn	Dep	Max	*Dist {1} & *{stop}coda	*Dist {1}	*Dist {1,0} & *{stop}coda	*{stop}coda	*Dist {1,0}
☞ s.n				*			*
si.n	*!						
n		*!					

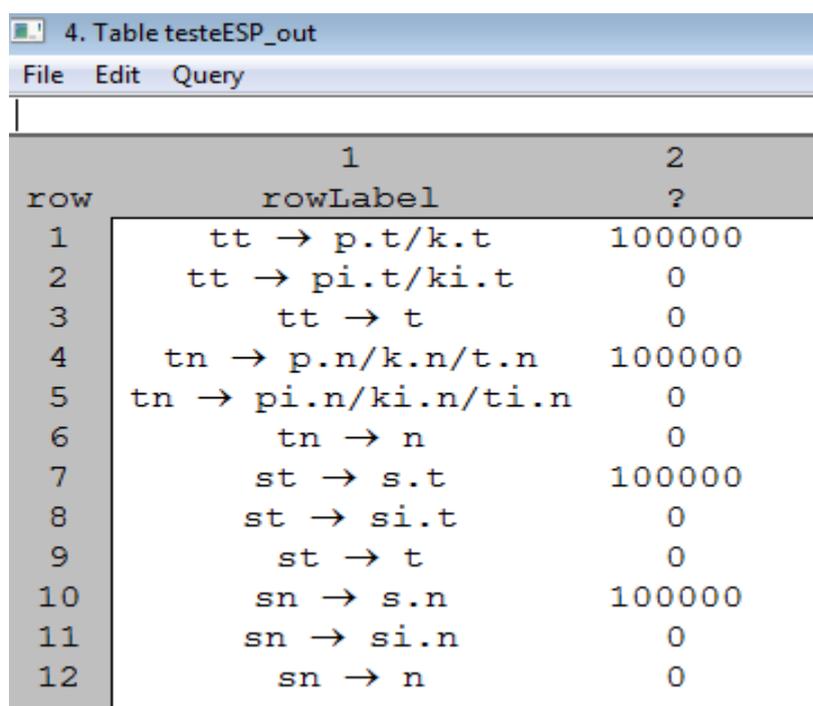
Programamos os algoritmos para que eles atingissem um sistema-alvo a partir do qual plosivas finais são permitidas. Dessa forma, as restrições de marcação devem ser afetadas pelo algoritmo, de forma a serem demovidas e alcançarem as posições mais baixas da hierarquia no estágio final, invertendo o ranking do estágio inicial ( $M \gg F$ ).

Como estamos trabalhando com a Teoria da Otimidade Estocástica neste momento, para fazermos com que venha a emergir um candidato ótimo sem epêntese ou apagamento, é importante mantermos as restrições que se referem à proibição da produção da plosiva em coda mais abaixo na hierarquia, para que os candidatos que apresentem tal configuração não recebam uma violação fatal - que é dada aos candidatos que violarem as restrições mais altamente ranqueadas -, vindo, conseqüentemente, a ser eliminados da disputa a *output* ótimo.

Logo, com relação aos resultados numéricos apresentados pelas restrições de fidelidade DEP (60.908) e MAX (58.945), verificamos que, como neste caso de produções de

coda do Espanhol não temos apagamento nem epêntese, estas restrições devem apresentar valores superiores aos das restrições de marcação, fazendo com que ocupem os lugares mais altos na hierarquia. Além disso, estas duas restrições de fidelidade podem apresentar valores de ranqueamento próximos entre elas, já que não afetam a escolha do candidato ótimo, pois, na gramática em questão, o importante é que as restrições de marcação assumam valores centrais bem inferiores aos valores das restrições de fidelidade. Dessa forma, ainda que tais restrições de marcação apresentem faixas de valores probabilísticos que mantenham contato entre si, tal fato não afeta a escolha do candidato ótimo, que independe da relação  $MAX \gg DEP$  ou  $DEP \gg MAX$ , visto os altos valores numéricos dessas duas restrições

Assim, em se tratando da aquisição da língua materna dos aprendizes, o algoritmo conseguiu convergir na OT Estocástica para uma simulação que sugere *outputs* categóricos. Os valores de DEP e MAX encontram-se bem superiores aos das restrições de marcação (acima de 10 pontos), não permitindo, portanto, que haja variação no candidato ótimo. A figura 19, a seguir, vai corroborar esta informação.



row	1 rowLabel	2 ?
1	tt → p.t/k.t	100000
2	tt → pi.t/ki.t	0
3	tt → t	0
4	tn → p.n/k.n/t.n	100000
5	tn → pi.n/ki.n/ti.n	0
6	tn → n	0
7	st → s.t	100000
8	st → si.t	0
9	st → t	0
10	sn → s.n	100000
11	sn → si.n	0
12	sn → n	0

**Figura 19 – Output Distributions da aquisição do Espanhol Colombiano – OT**

Conforme podemos ver na distribuição probabilística dos dados (figura 19), realizada no modelo da Teoria da Otimidade Estocástica, tivemos, como resposta, 100% de produção da consoante plosiva em coda para todas as sequências consideradas

“[plosiva].[plosiva]” [p.t, k.t], “[plosiva].[nasal]” [p.n, k.n, t.n], [s.t], [s.n], tal como ocorre no Espanhol Colombiano.

b) A aquisição do Espanhol Colombiano (L1) via Gramática Harmônica

O estágio inicial de aquisição no modelo da HG é o mesmo definido para a OT. As restrições de marcação receberam um valor numérico igual a 100, e as restrições de fidelidade DEP e MAX, um valor numérico igual a 0. A diferença entre os dois algoritmos deve se dar na resposta numérica para o estágio final, pois, enquanto na OT temos os pesos das restrições definindo o ranking entre elas, na HG, a partir dos pesos das restrições, devemos calcular o valor de Harmonia. Dessa forma, as restrições de marcação também devem ser afetadas pelo algoritmo, de forma a terem seus pesos diminuídos, porém, numa proporção diferenciada com relação à OT, para que mantenham os *outputs* ótimos após o cálculo de H, conforme verificamos no conjunto de *tableaux* 15.

**Conjunto de Tableaux 15 - Simulação 1 HG / Aquisição do Espanhol Colombiano (0% de epêntese)**

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
<b>Dep</b>	70.258	69.027	1.000000
<b>Max</b>	71.834	67.222	1.000000
<b>*Dist {1} &amp; *(stop)coda</b>	38.531	36.622	1.000000
<b>*Dist {1}</b>	11.633	13.261	1.000000
<b>*Dist {1,0} &amp; *(stop)coda</b>	0.725	2.428	1.000000
<b>*(stop)coda</b>	0.725	-1.010	1.000000
<b>*Dist {1,0}</b>	-42.092	-44.157	1.000000

tt	Dep	Max	*Dist {1} & *(stop)coda	*Dist {1}	*Dist {1,0} & *(stop)coda	*(stop)coda	*Dist {1,0}	
☞ p.t/k.t					*	*	*	-2.428
pi.t/ki.t	*							-69.027
t		*						-67.222

tn	Dep	Max	*Dist {1} & *(stop)coda	*Dist {1}	*Dist {1,0} & *(stop)coda	*(stop)coda	*Dist {1,0}	
☞ p.n/k.n/t.n			*	*	*	*	*	-52.310
pi.n/ki.n/ti.n	*							-69.027
n		*						-67.222

st	Dep	Max	*Dist {1} & *(stop)coda	*Dist {1}	*Dist {1,0} & *(stop)coda	*(stop)coda	*Dist {1,0}	
☞ s.t							*	0
si.t	*							-69.027
t		*						-67.222

sn	Dep	Max	*Dist {1} & *(stop)coda	*Dist {1}	*Dist {1,0} & *(stop)coda	*(stop)coda	*Dist {1,0}	
☞ s.n				*			*	-13.261
si.n	*							-69.027
n		*						-67.222

Os valores à direita do Conjunto de *Tableaux*, definidores do candidato ótimo na HG, são os valores de Harmonia. O programa PRAAT, que utilizamos para realizar as

simulações, faz o cálculo do valor de Harmonia para cada candidato, que é a soma da multiplicação do valor de cada violação pelo peso da restrição. A simples conferência deste cálculo de H pode ser realizada com base no conjunto de *tableaux* 15, que nos dá o peso das restrições e as violações – que apresentam valores negativos para o cálculo de Harmonia, neste modelo.

Realizando uma demonstração do cálculo de Harmonia para o candidato [s.n], serão três as somas que devemos fazer, uma para cada candidato a *output*. A primeira refere-se ao candidato [s.n] - com a produção da plosiva em coda; outra soma deve ser feita com relação ao candidato [si.n] - com epêntese. Finalmente, a última soma deve ser feita para o candidato [n], que representa o apagamento da plosiva em coda:

Antes de iniciarmos a demonstração dos cálculos de Harmonia, duas informações importantes devem ser retomadas: os valores numéricos das restrições utilizados no cálculo de H correspondem aos valores apresentados na coluna *disharmony*, ou seja, aos valores de ponto de seleção. Além disso, as restrições com valores negativos não contam no cálculo, porque selecionamos a opção Linear OT, como ocorre com \*Dist{1,0} (-44.157), automaticamente eliminado do cálculo de H.

Candidato com a produção da plosiva em coda

$$/sn/ = (-1 \text{ violação} \times *Dist\{1\}) + \cancel{(-1 \text{ violação} \times *Dist\{1,0\})}$$

$$/sn/ = (-1 \times 13.261) = -13.261$$

Candidato com epêntese

$$/sin/ = (-1 \times 69.027)$$

$$/sin/ = -69.027$$

Candidato com apagamento

$$/n/ = (-1 \times 67.222)$$

$$/n/ = -67.222$$

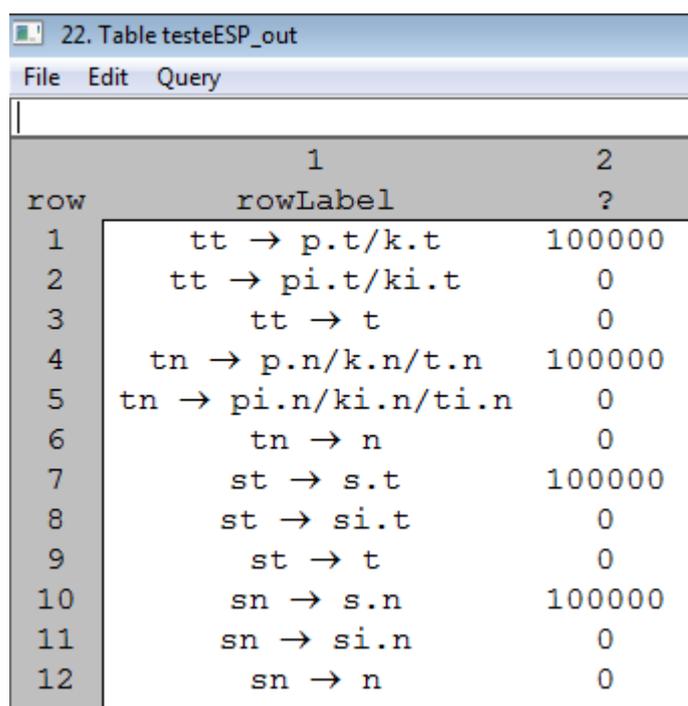
Para este candidato [s.n], que apresentou valor de Harmonia igual a -13.261, bem superior ou mais próximo a zero, com relação aos candidatos com epêntese [si.n] e apagamento [n], já podemos adiantar que tal forma de *output* é a vencedora categoricamente, ou seja, a grande distância no valor de Harmonia que assume frente aos outros candidatos não permite que haja variação neste resultado. Por mais que a cada avaliação, ou a cada produção linguística realizada, tenhamos valores diferentes para as restrições, os dois candidatos perdedores nunca chegarão a causar variação entre os candidatos ótimos, dada a distância no valor de Harmonia.

Assim como tivemos para o candidato [s.n], os outros candidatos ótimos [s.t] = 0, [p.t, k.t] = -2.428 e [p.n, k.n, t.n] = -52.310 também apresentaram um valor de H que supera o valor dos demais candidatos a output: candidatos com epêntese [pi.t, ki.t], [si.t], [pi.n, ki.n, ti.n], [si.n] = -69.027 e candidatos com apagamento [st] e [sn] = -67.222. Neste modelo, essa observação é fundamental para que se considere o caráter categórico de ocorrência da obstruente em coda, na aquisição do Espanhol Colombiano.

Porém, em se comparando os valores de ranqueamento das restrições, ou valor central, na HG, com os valores que obtivemos na OT, vamos ver que as restrições de fidelidade, que no estágio final de aquisição do EC foram promovidas na hierarquia, tiveram, neste algoritmo da HG, seus valores superiores aos obtidos na OT. Isso também ocorreu com as restrições de marcação, que tiveram seus pesos inferiores aos dados no ranqueamento da OT. De fato, se estamos falando que na HG trabalhamos com a cumulatividade das restrições, faz sentido que DEP e MAX tenham que assumir valores tão altos que sejam superiores à soma dos pesos de todas as restrições de marcação, para a preparação do cálculo da Harmonia.

Assim, em se tratando da aquisição da língua materna dos aprendizes, o algoritmo conseguiu convergir também na HG para uma simulação que sugere *outputs* categóricos, ainda que com valores distintos aos dados na OT.

O resultado do *output distributions*, dado pela figura 20, a seguir, representa as probabilidades de ocorrência de cada uma das sequências dadas nesta análise final.



row	rowLabel	?
1	tt → p.t/k.t	100000
2	tt → pi.t/ki.t	0
3	tt → t	0
4	tn → p.n/k.n/t.n	100000
5	tn → pi.n/ki.n/ti.n	0
6	tn → n	0
7	st → s.t	100000
8	st → si.t	0
9	st → t	0
10	sn → s.n	100000
11	sn → si.n	0
12	sn → n	0

Figura 20 - *Output Distributions* da aquisição do Espanhol Colombiano – HG

A probabilidade de 100%, apresentada no *output distributions*, para a produção dos segmentos plosivos em coda que estamos trabalhando [plosiva].[plosiva], [plosiva].[nasal], e para a fricativa em coda [s.t] e [s.n], serve como prova real nesta verificação da aquisição do Espanhol Colombiano no modelo teórico da Gramática Harmônica.

### 5.2.1.2 Simulação 2 - Aquisição do Português Brasileiro (L2)

a) A aquisição do Português Brasileiro (L2) via Teoria da Otimidade Estocástica

Esta simulação 2, referente à aquisição do PB, que também estará presente na análise final com o conjunto completo de restrições, vai permitir que possamos verificar a capacidade dos algoritmos OT-GLA e HG-GLA de convergirem em uma gramática que reflita um padrão categórico de saída, agora com 100% da produção de epêntese após a plosiva em posição de coda na L2, mas não após a fricativa [s], perfeitamente possível, no PB, em posição de coda, tal como em “**cas.pa**”.

Ao considerarmos que o estágio inicial de aquisição de L2 é o resultado final, ou alvo da L1, tanto nesta simulação com o algoritmo OT GLA, quanto com o algoritmo de aprendizagem da HG, tivemos que alimentar o sistema com os valores de ranqueamento obtidos na simulação anterior, dados para cada modelo. No caso do algoritmo vinculado à OT, tivemos:

- MAX = 58.945; DEP = 60.908

- \*Dist {1} = 21.091; \*Dist {1,0} = -19.852; \*{stop}coda = 21.209; \*Dist {1} & \*{stop}coda = 46.969; \*Dist {1,0} & \*{stop}coda = 21.209.

Com isso, podemos dizer que o estágio inicial da L2 vai contar com as restrições de fidelidade (MAX e DEP) dominando as de marcação (\*Dist {1}, \*Dist {1,0}, \*{stop}coda, \*Dist {1} & \*{stop}coda, \*Dist {1,0} & \*{stop}coda), na hierarquia das restrições propostas para lidar com a aquisição da consoante plosiva em coda medial de palavra. O estágio final é dado no conjunto de *tableaux* 16.

Conjunto de *Tableaux* 16 - Simulação 2 OT / Aquisição do Português Brasileiro (100% de epêntese)

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
Max	58.945	56.495	1.000000
*Dist {1} & *{stop}coda	51.700	50.804	1.000000
*Dist {1,0} & *{stop}coda	46.529	47.218	1.000000
*{stop}coda	46.529	46.851	1.000000
Dep	36.632	39.686	1.000000
*Dist {1}	24.778	26.577	1.000000
*Dist {1,0}	4.424	5.843	1.000000

tt	Max	*Dist {1} & *{stop}coda	*Dist {1,0} & *{stop}coda	*{stop}coda	Dep	*Dist {1}	*Dist {1,0}
p.t/k.t			*!	*			*
☞ pit/ki.t					*		
t	*!						

tn	Max	*Dist {1} & *{stop}coda	*Dist {1,0} & *{stop}coda	*{stop}coda	Dep	*Dist {1}	*Dist {1,0}
p.n/k.n/t.n		*!	*	*		*	*
☞ pi.n/ki.n/ti.n					*		
n	*!						

st	Max	*Dist {1} & *{stop}coda	*Dist {1,0} & *{stop}coda	*{stop}coda	Dep	*Dist {1}	*Dist {1,0}
☞ s.t							*
si.t					*!		
t	*!						

sn	Max	*Dist {1} & *{stop}coda	*Dist {1,0} & *{stop}coda	*{stop}coda	Dep	*Dist {1}	*Dist {1,0}
☞ s.n						*	*
si.n					*!		
n	*!						

Nesta segunda simulação, o algoritmo da OT Estocástica foi programado para atingir um sistema-alvo a partir do qual plosivas finais não são permitidas, mas a fricativa [s] em coda sim. Sendo assim, neste momento que tratamos da aquisição do PB, as restrições devem ser afetadas pelo algoritmo, de forma a atenderem esta situação.

Dessa forma, as restrições de marcação (\*Dist{1} e \*Dist{1,0}), violadas pelas sequências [s.n, s.t], devem continuar sendo demovidas e ainda serem inferiores às restrições de fidelidade (DEP e MAX), para que possam continuar fazendo com que tenhamos os candidatos ótimos sem epêntese e sem apagamento, [s.t] e [s.n], assim como vimos na simulação anterior no Espanhol Colombiano (L1).

Por outro lado, as restrições de marcação (\*{stop}coda, \*Dist {1} & \*{stop}coda, \*Dist {1,0} & \*{stop}coda), que se referem especificamente aos *outputs* das sequências de plosiva + plosiva (“t.t”), representadas pelas sequências [p.t, k.t] e plosiva + nasal (“t.n”), representadas por [p.n, k.n, t.n]), devem ser promovidas e superarem DEP, de forma a

receberem uma violação fatal e deixarem emergir candidatos ótimos com epêntese, como, por exemplo: “ra.p[i].to” e “te.c[i].ni.cis.mo”.

Logo, a restrição de fidelidade DEP (36.632) assume um valor central que possibilita que supere \*Dist {1} (24.778) e \*Dist{1,0} (4.424), fazendo emergir [st] e [sn]. Ao mesmo tempo, o valor da restrição em questão encontra-se abaixo dos de \*{stop}coda (46.529), \*Dist {1} & \*{stop}coda (51.700), \*Dist {1,0} & \*{stop}coda (46.529) no ranking, para que se obtenham, na simulação do PB, [pi.t, ki.t] (ex.: “ca.c[i].to”) e [pi.n, ki.n, ti.n] (ex.: “hi. p[i]. no.se”).

Com relação à restrição de fidelidade MAX, como não temos nenhum candidato ótimo com apagamento, esta restrição deve aparecer sempre com valor superior ao das demais, do que resulta em candidatos com apagamento que sempre incorrerão em uma violação fatal. Esta restrição deve ainda manter uma distância (acima de 10 pontos) sobre as outras restrições utilizadas na análise, para que tenhamos um candidato ótimo categórico. Assim, é fundamental que seu valor assuma posição superior e bastante afastada do de DEP nos casos dos candidatos ótimos [pi.n, ki.n, ti.n] e [pi.t, ki.t], bem como superior aos de \*Dist{1} e \*Dist{1,0}, no caso dos *outputs* [s.t] e [s.n].

A restrição de fidelidade DEP deve apresentar um peso de valor superior ao das restrições \*Dist{1} e \*Dist{1,0}, para que os candidatos com a fricativa em coda ([s.t] e [s.n]) venham a emergir e sejam categóricos. Por sua vez, para garantir os *outputs* [pi.n, ki.n, ti.n] e [pi.t, ki.t], tal restrição deve apresentar valor inferior ao das restrições conjuntas e da restrição \*{stop}coda, para que os candidatos com plosiva em coda, passíveis de epêntese, não sofram violação fatal e sejam eliminados da disputa pelo candidato ótimo.

Assim, em se tratando da aquisição plena da língua estrangeira – Português Brasileiro - pelos aprendizes colombianos, o algoritmo conseguiu convergir, na OT Estocástica, para uma simulação que sugere, assim como na simulação 1, *outputs* categóricos. Porém, enquanto no Espanhol Colombiano era importante que DEP assumisse valor superior ao das restrições de marcação, no PB, tal restrição de fidelidade teve não somente de ser demovida até encontrar um peso que superasse as restrições \*Dist{1} e \*Dist{1,0}, referentes aos *outputs* ótimos sem epêntese [s.n, s.t], tendo sido necessário, também, que DEP estivesse abaixo das restrições \*{stop}coda, \*Dist{1}&{stop}coda e \*Dist{1,0}&{stop}coda, referentes aos *outputs* com epêntese “ti.t”, “ti.n”.

A figura 21, dada pelo PRAAT, mostra a distribuição probabilística dos dados na aquisição do PB como língua estrangeira, para esta segunda simulação, corroborando a ideia de que a aquisição do Português Brasileiro admite epêntese em sequências heterossilábicas

formadas de consoante plosiva em coda, como [plosiva].[plosiva] – rap[i]to e [plosiva].[nasal] – et[i]nia, mas não em sequências como [s.t] – “astuto” e [s.n] – “asma”, de consoante fricativa em posição de coda medial.

row	1 rowLabel	2 ?
1	tt → p.t/k.t	0
2	tt → pi.t/ki.t	100000
3	tt → t	0
4	tn → p.n/k.n/t.n	0
5	tn → pi.n/ki.n/ti.n	100000
6	tn → n	0
7	st → s.t	100000
8	st → si.t	0
9	st → t	0
10	sn → s.n	100000
11	sn → si.n	0
12	sn → n	0

**Figura 21 – Output Distributions da aquisição do Português Brasileiro – OT**

Assim, para a aquisição de codas mediais no Português Brasileiro (L2), com 100% de epêntese após a consoante plosiva, demonstramos, através do recurso *output distributions* (figura 21), que, ao submetermos os candidatos a 100.000 avaliações, a gramática em questão resultou nos índices percentuais desejados, expressando a acuidade dos valores fornecidos pelo algoritmo às restrições.

b) A aquisição do Português Brasileiro (L2) via Gramática Harmônica

Nesta simulação do sistema do Português à luz da HG, alimentamos o algoritmo com as mesmas informações fornecidas ao algoritmo OT-GLA. O algoritmo foi informado de que, no sistema-alvo, *inputs* com plosivas finais devem ser epentetizados, além de que *inputs* com a fricativa [s] devem manter-se sem epêntese. O algoritmo também foi alimentado com os valores do sistema-alvo da simulação 1 – aquisição do Espanhol Colombiano na HG, para a determinação do estágio inicial da aquisição:

- MAX = 71.834; DEP = 70.258

- \*Dist {1} = 11.633; \*Dist {1,0} = -42.092; \*{stop}coda = 0.725; \*Dist {1} & {stop}coda = 38.531; \*Dist {1,0} & {stop}coda = 0.725

A diferença entre os dois algoritmos nesta simulação referente à aquisição do Português Brasileiro (L2), contudo, se deu tanto na resposta numérica para o estágio inicial da L2, quanto com relação aos dados do estágio final, conforme podemos verificar no conjunto de *tableaux* 17.

**Conjunto de *Tableaux* 17 - Simulação 2 HG / Aquisição do Português Brasileiro (100% de epêntese)**

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>	
<b>Max</b>	71.834	73.680	1.000000	
<b>*Dist {1} &amp; *{stop}coda</b>	40.740	41.390	1.000000	
<b>Dep</b>	41.929	38.741	1.000000	
<b>*{stop}coda</b>	29.054	28.181	1.000000	
<b>Dist {1,0} &amp; *{stop}coda</b>	29.054	27.142	1.000000	
<b>*Dist {1}</b>	13.842	12.401	1.000000	
<b>*Dist {1,0}</b>	-13.763	-12.764	1.000000	

tt	Max	*Dist {1} & *{stop}coda	Dep	*{stop}coda	*Dist {1,0} & *{stop}coda	*Dist {1}	*Dist {1,0}	
p.t/k.t				*	*		*	-55.323
☞ pi.t/ki.t			*					-38.741
t	*							-73.680

tn	Max	*Dist {1} & *{stop}coda	Dep	*{stop}coda	*Dist {1,0} & *{stop}coda	*Dist {1}	*Dist {1,0}	
p.n/k.n/t.n		*		*	*	*	*	-109.114
☞ pi.n/ki.n/ti.n			*					-38.741
n	*							-73.680

st	Max	*Dist {1} & *{stop}coda	Dep	*{stop}coda	*Dist {1,0} & *{stop}coda	*Dist {1}	*Dist {1,0}	
☞ s.t							*	0
si.t			*					-38.741
t	*							-73.680

sn	Max	*Dist {1} & *{stop}coda	Dep	*{stop}coda	*Dist {1,0} & *{stop}coda	*Dist {1}	*Dist {1,0}	
☞ s.n						*	*	-12.401
si.n			*					-38.741
n	*							-73.680

Conforme já observamos anteriormente, os valores à direita do *tableau* correspondem aos valores de Harmonia, definidores do candidato ótimo na HG. Bem como ocorreu na simulação 2 da OT Estocástica, os candidatos selecionados a *output* ótimo devem apresentar valor de Harmonia superior, ou mais próximo a zero, frente aos demais candidatos. Além disso, devem garantir um valor tal (10 pontos, conforme padrão computacional), que os distancie dos valores de Harmonia dos candidatos perdedores.

Olhando para os valores de Harmonia dos candidatos ótimos no Conjunto de *Tableaux* 17, vimos que esta meta foi alcançada. É importante que tenhamos esta distância entre os valores de Harmonia para que, conforme já discutimos, em momentos de produção linguística, não tenhamos variação no candidato ótimo. A distância superior aos 10 pontos

entre os valores de H na HG - garantida na OT, por sua vez, pelos valores de ranqueamento – nos assegura de que estamos obtendo como resposta um candidato ótimo categórico.

Porém, quando comparados os valores dos pesos das restrições nos dois diferentes algoritmos vinculados à OT e à HG, podemos verificar comportamentos numéricos bem diferentes. É interessante mencionar que a restrição de fidelidade MAX, por exemplo, apesar de apresentar peso maior na HG (71.834), com relação à OT (58.945), em ambos os modelos conquista uma posição muito alta, superando as demais restrições. Estes índices são perfeitamente explicáveis, pois, já que não temos candidato ótimo com apagamento, esta restrição deve ser sempre promovida no *ranking*, com a diferença de que, na HG, seu peso deve superar a soma dos pesos das demais restrições responsáveis por cada candidato ótimo.

No que se refere à restrição DEP, vimos que, na OT, esta restrição assumiu valor tal que a fizesse obter um peso que superasse ao das restrições de marcação referentes às fricativas em coda (\*Dist{1} e \*Dist{1,0}), deixando emergir os candidatos [s.t, s.n]. Por sua vez, o valor das restrições em questão ficou abaixo do das restrições de marcação referentes às plosivas epentetizadas (\*Dist {1,0} & {stop}coda, \*Dist {1} & {stop}coda e \*{stop}coda), fazendo com que os candidatos “t[i]t” e “t[i]n” fossem ótimos. Na HG, a restrição DEP vai assumir outra posição no *ranking*, se estabelecendo logo abaixo de MAX, com a segunda posição do *ranking*.

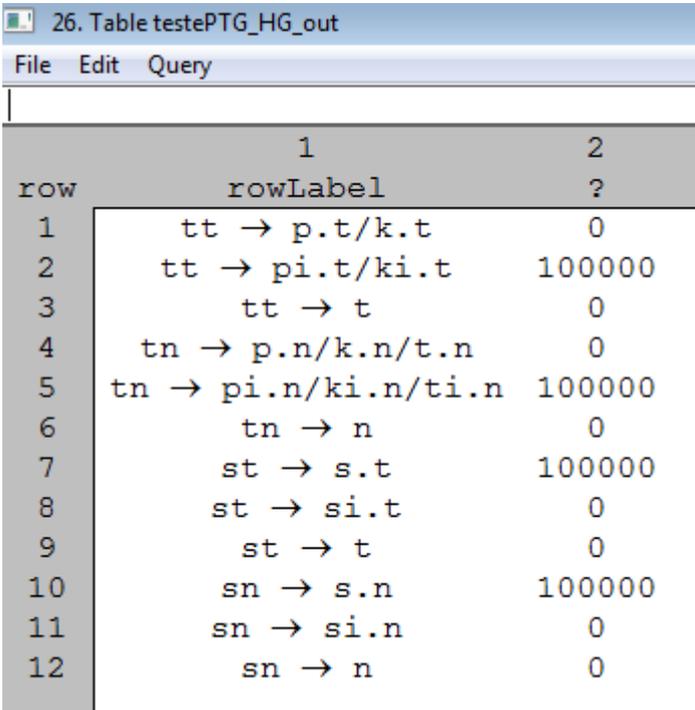
Esta posição de DEP faz-se, também, totalmente justificável, no momento em que estamos trabalhando com um modelo que sugere a cumulatividade do peso das restrições. No caso dos candidatos “t.t” [p.t, k.t] e “t.n”[p.n, k.n, t.n], o peso de DEP poderia ser alto na hierarquia desde que não ultrapassasse a soma dos pesos das restrições violadas por estes candidatos perdedores. Como estes candidatos violam mais de uma restrição de marcação, e tais restrições apresentam pesos significativos na hierarquia, o peso de DEP sozinho não ultrapassaria este valor de Harmonia, alcançando o menor valor de H, ou mais próximo a zero, de modo a garantir os candidatos com plosiva em coda seguido de epêntese, a escolha ótima.

Com relação aos *outputs* ótimos [s.t] e [s.n], ao levarmos em conta que a restrição de marcação \*Dist{1,0} – que está relacionada à sequência [s.t] - apresentou valor negativo propositadamente, para que atingisse um índice inferior a zero, que a fizesse sair fora da definição dos pesos das restrições, automaticamente recebendo valor zero para o cálculo de H, DEP poderia expressar qualquer valor positivo como peso. Já com relação à sequência [s.n], que vai violar \*Dist{1,0} e \*Dist{1}, só teremos o valor do ponto de seleção de \*Dist {1} (12.401) no cálculo da Harmonia, já que o valor do ponto de seleção de \*Dist{1,0} (-12.764)

é negativo, estipulando que DEP poderia apresentar, no mínimo, valor acima de 12.401, para que se mantivesse o candidato ótimo categórico.

Assim, em se tratando da aquisição da língua estrangeira dos aprendizes (Português Brasileiro), o algoritmo conseguiu convergir também na HG para uma simulação que sugere *outputs* categóricos, ainda que com valores distintos aos dados na OT. Em outras palavras, o HG-GLA conseguiu atingir o resultado esperado.

No algoritmo vinculado à HG, demonstramos, a seguir, o resultado do recurso *output distributions*, que representa a simulação das probabilidades de ocorrência de cada uma das sequências dadas nesta análise final.



row	1 rowLabel	2 ?
1	tt → p.t/k.t	0
2	tt → pi.t/ki.t	100000
3	tt → t	0
4	tn → p.n/k.n/t.n	0
5	tn → pi.n/ki.n/ti.n	100000
6	tn → n	0
7	st → s.t	100000
8	st → si.t	0
9	st → t	0
10	sn → s.n	100000
11	sn → si.n	0
12	sn → n	0

Figura 22 – *Output Distributions* da aquisição do Português Brasileiro – HG

Como ainda estamos tratando da aquisição do Português Brasileiro, tivemos a probabilidade esperada de 100% para a produção das consoantes plosivas em coda com epêntese, e 100% da produção da fricativa [s] em coda para as sequências [s.t, s.n], o que demonstra coerência com os valores numéricos fornecidos pelo algoritmo.

### 5.2.1.3 Simulação 3 – Interlíngua (*outputs* variáveis)

a) A aquisição dos *outputs* variáveis (Interlíngua) via Teoria da Otimidade Estocástica

Através da Simulação 3, observamos o tratamento de ambos os algoritmos de aprendizagem frente a gramáticas com *outputs* variáveis, de acordo com os índices percentuais encontrados no *corpus* do presente trabalho. Esta possibilidade de variação do candidato ótimo dada pelos algoritmos de aprendizagem que estamos utilizando nesta pesquisa é um grande diferencial destes modelos, diferencial esse que será verificado nesta última simulação.

Diferentemente das simulações 1 e 2, os valores centrais das restrições, neste momento, precisam ser muito próximos. Assim, as alterações nos valores dos pontos de seleção, em momentos de produção linguística, não representarão apenas uma mera alteração numérica dos pesos das restrições, mas motivação para que tenhamos mudança no ranking das restrições, no caso da OT, e valores de Harmonia próximos, no caso da HG.

Na terceira e última simulação, consideramos que o estágio inicial da interlíngua é o resultado final, ou estágio-alvo, da L1. Assim, temos que alimentar o sistema com os valores de ranqueamento obtidos na simulação 1, ou seja, nosso estágio inicial para o desenvolvimento da interlíngua é o mesmo que utilizamos para a aquisição plena do Português Brasileiro (L2), na simulação anterior.

- MAX = 58.945; DEP = 60.908

- \*Dist {1} = 21.091; \*Dist {1,0} = -19.852; \*{stop}<sub>coda</sub> = 21.209; \*Dist {1} & \*{stop}<sub>coda</sub> = 46.969; \*Dist {1,0} & \*{stop}<sub>coda</sub> = 21.209.

Com isso, podemos repetir que o estágio inicial da interlíngua vai contar com as restrições de fidelidade (MAX e DEP) dominando as restrições de marcação (\*Dist {1}, \*Dist {1,0}, \*{stop}<sub>coda</sub>, \*Dist {1} & \*{stop}<sub>coda</sub>, \*Dist {1,0} & \*{stop}<sub>coda</sub>), na hierarquia das restrições propostas para lidar neste momento.

Os *outputs* variáveis, que representam a gramática em desenvolvimento dos aprendizes, podem ser vistos no Conjunto de *Tableaux* 18.

Conjunto de Tableaux 18 - Simulação 3 OT / Interlíngua (outputs variáveis)

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
Max	60.081	61.701	1.000000
Dep	44.083	47.568	1.000000
*Dist {1,0} & *{stop}coda	36.899	37.350	1.000000
*Dist {1} & *{stop}coda	40.937	37.280	1.000000
*{stop}coda	36.899	36.197	1.000000
*Dist {1}	15.059	14.551	1.000000
*Dist {1,0}	-4.162	-2.678	1.000000

tt	Max	Dep	*Dist {1,0} & *{stop}coda	*Dist {1} & *{stop}coda	*{stop}coda	*Dist {1}	*Dist {1,0}
☞ p.t/k.t			*		*		*
pi.t/ki.t		*!					
t	*!						

tn	Max	Dep	*Dist {1,0} & *{stop}coda	*Dist {1} & *{stop}coda	*{stop}coda	*Dist {1}	*Dist {1,0}
☞ p.n/k.n/t.n			*	*	*	*	*
pi.n/ki.n/ti.n		*!					
n	*!						

st	Max	Dep	*Dist {1,0} & *{stop}coda	*Dist {1} & *{stop}coda	*{stop}coda	*Dist {1}	*Dist {1,0}
☞ s.t							*
si.t		*!					
t	*!						

sn	Max	Dep	*Dist {1,0} & *{stop}coda	*Dist {1} & *{stop}coda	*{stop}coda	*Dist {1}	*Dist {1,0}
☞ s.n						*	*
si.n		*!					
n	*!						

Diferentemente das simulações 1 e 2, os pontos de seleção da restrição de fidelidade DEP (47.568) e das restrições de marcação, que se referem exclusivamente às sequências com plosiva em coda \*Dist{1,0} & \*{stop}coda (37.350), \*Dist{1} & \*{stop}coda (37.280), \*{stop}coda (36.197), são próximos, representando que está havendo variação na seleção do *output* ótimo, no que diz respeito às sequências [p.t, k.t] e [p.n, k.n, t.n]. Dessa forma, alterações nos pontos de seleção, dados por avaliações sucessivas, ao contrário do que ocorreu nas simulações anteriores (1 e 2), representam uma fonte de motivação para mudança na *ranking* das restrições.

Os dados expostos na tabela 26 mostraram que, em 98,8% das produções de [p.t, k.t], tivemos a produção da plosiva em coda. Entretanto, há 1,2% de ocorrência de produção da plosiva com inserção de epêntese. Logo, o algoritmo deve refletir esta situação, possibilitando que, em 1,2% das avaliações, tenhamos os candidatos ótimos [pi.t, ki.t], com epêntese. Da mesma forma, vamos ter que a sequência [p.n, k.n, t.n] aparece 86,6% das vezes

nas produções dos aprendizes colombianos e em 13,4% dos casos como [pi.n, ki.n, ti.n], com epêntese. Assim, também nesta sequência, temos de formalizar a emergência de ora um candidato ótimo com epêntese, ora sem epêntese, de acordo com a porcentagem de ocorrência de cada um.

O conjunto de *tableaux* 19 vai nos fornecer uma amostra desta variação, que pode ser percebida, se compararmos esta hierarquia da segunda rodada, com a do conjunto de *tableaux* 18, que mostrava, para um outro momento de produção linguística, a mesma simulação 3, na Teoria da Otimidade Estocástica.

**Conjunto de Tableaux 19 - Simulação 3 OT / Interlândia – Segunda rodada**

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>				
Max	60.081	57.482	1.000000				
*Dist {1,0} & *{stop}coda	36.899	43.439	1.000000				
Dep	44.083	43.388	1.000000				
*Dist {1} & *{stop}coda	40.937	39.858	1.000000				
*{stop}coda	36.899	35.091	1.000000				
*Dist {1}	15.059	14.895	1.000000				
*Dist {1,0}	-4.162	-5.264	1.000000				

tt	Max	*Dist {1,0} & *{stop}coda	Dep	*Dist {1} & *{stop}coda	*{stop}coda	*Dist {1}	*Dist {1,0}
p.t/k.t		*!			*		*
☞ pit/ki.t			*				
t	*!						

tn	Max	*Dist {1,0} & *{stop}coda	Dep	*Dist {1} & *{stop}coda	*{stop}coda	*Dist {1}	*Dist {1,0}
p.n/k.n/t.n		*!		*	*	*	*
☞ pi.n/ki.n/ti.n			*				
n	*!						

Enquanto que, no conjunto de *tableaux* 18, tínhamos DEP com um valor de ponto de seleção igual a 47.568 – permitido pela margem de valores assumidos pelo *ranking value* desta restrição (39.083 a 49083) -, que superava todas as restrições de marcação, apesar da proximidade nos valores, nesta segunda rodada do conjunto de *tableaux* 19, DEP assume um valor de seleção menor, igual a 43.388, sendo ultrapassada pela restrição \*Dist{1,0} & {stop}coda, que fará emergir novos candidatos ótimos para as sequências formadas de plosiva mais nasal e plosiva mais plosiva: candidatos ótimos com epêntese.

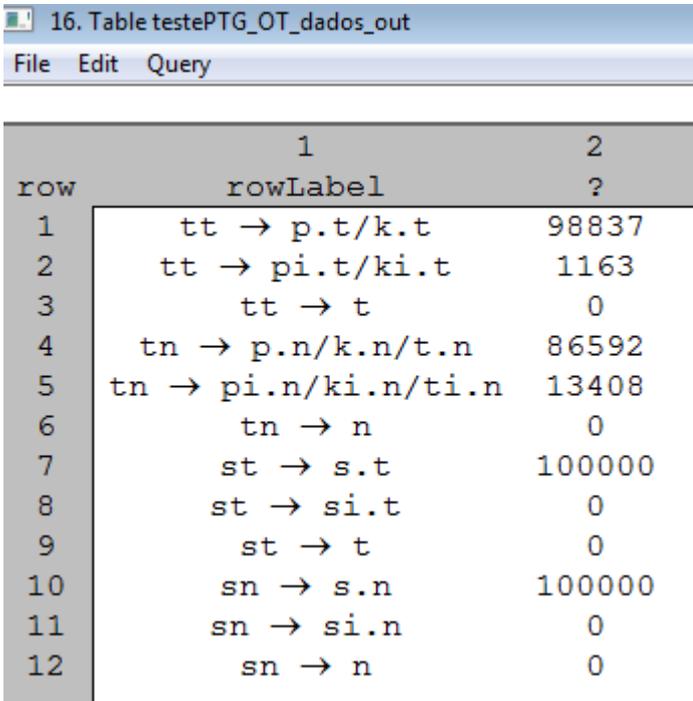
Além de esta hierarquia de restrições mostrar que restrições com valores centrais próximos podem ocasionar variação no candidato ótimo, ela está evidenciando uma organização das restrições que vai garantir que as restrições referentes aos candidatos menos marcados ocupem posições mais baixas na hierarquia com relação às restrições que se referem a estruturas mais marcadas. Assim, não por acaso, vamos ter maior frequência da ocorrência

de epêntese na sequência com plosiva + nasal, que diante das demais sequências é a mais marcada.

Quanto às sequências [s.t] e [s.n], menos marcadas se comparadas a [p.n, k.n, t.n] e [p.t, k.t], encontramos 100% da produção desta fricativa em coda, já que o /S/ no PB é perfeitamente possível nesta situação. O reflexo disto pode ser verificado no posicionamento que assumem as restrições de marcação \*Dist{1} e \*Dist{1,0}, que se referem a esta fricativa, na hierarquia das restrições dada na OT Estocástica. Se observarmos o conjunto de *tableaux* 18, veremos que estas restrições, além de assumirem as posições mais baixas no *ranking*, apresentam valores afastados das demais restrições.

Assim, em se tratando de uma gramática em desenvolvimento, com a presença de *outputs* variáveis, o algoritmo de aprendizagem vai convergir na OT Estocástica.

A figura 23, a seguir, apresenta uma prova real da simulação realizada no PRAAT, para os dados de interlândia, com *outputs* variáveis, dos estudantes colombianos aprendendo Português Brasileiro. Após a submissão dos candidatos a 100.000 avaliações, a distribuição probabilística dos dados apontou para uma probabilidade de ocorrência de porcentagens esperada.



row	rowLabel	2
1	tt → p.t/k.t	98837
2	tt → pi.t/ki.t	1163
3	tt → t	0
4	tn → p.n/k.n/t.n	86592
5	tn → pi.n/ki.n/ti.n	13408
6	tn → n	0
7	st → s.t	100000
8	st → si.t	0
9	st → t	0
10	sn → s.n	100000
11	sn → si.n	0
12	sn → n	0

Figura 23 – *Output Distributions* da Interlândia – OT

Conforme podemos ver no *output distributions*, para a Interlândia dos colombianos aprendendo PB, na Teoria da Otimidade Estocástica, tivemos 100% da produção

da fricativa em coda, e outputs variáveis para as sequências heterossilábicas formadas por plosiva em coda.

A seguir, continuamos com os dados de interlíngua do processo de aquisição das codas mediais pelos estudantes colombianos, no Português Brasileiro, porém, no modelo teórico da Gramática Harmônica.

b) A aquisição dos *outputs* variáveis (Interlíngua) via Gramática Harmônica

Para esta simulação do sistema da gramática em desenvolvimento à luz da HG, seguimos a mesma lógica que utilizamos no algoritmo OT-GLA. O algoritmo de aprendizagem da HG foi informado de que o estágio inicial da interlíngua é o resultado final, ou estágio-alvo da L1. Assim, alimentamos o sistema com os valores de ranqueamento obtidos na simulação 1, ou seja, nosso estágio inicial para o desenvolvimento da interlíngua é o mesmo que utilizamos para o estágio inicial da aquisição plena do Português Brasileiro (L2), na simulação 2 da HG:

- MAX = 71.834; DEP = 70.258

- \*Dist {1} = 11.633; \*Dist {1,0} = -42.092; \*{stop}<sub>coda</sub> = 0.725; \*Dist {1} & {stop}<sub>coda</sub> = 38.531; \*Dist {1,0} & {stop}<sub>coda</sub> = 0.725

A diferença entre os dois algoritmos de aprendizagem nesta terceira simulação para a gramática em desenvolvimento dos aprendizes colombianos de Português Brasileiro, contudo, se deu tanto na resposta numérica para o estágio inicial, quanto na caracterização do estágio final, conforme verificamos no conjunto de *tableaux* 20.

Conjunto de *Tableaux* 20 - Simulação 3 HG / Interlúngua (*outputs* variáveis)

	<i>ranking value</i>	<i>disharmory</i>	<i>plasticity</i>	
Max	72.819	73.028	1.000000	
Dep	49.345	53.496	1.000000	
*Dist {1,0} & *{stop}coda	20.652	21.693	1.000000	
*{stop}coda	20.652	20.066	1.000000	
*Dist {1} & *{stop}coda	3.481	3.686	1.000000	
*Dist {1,0}	-22.165	-21.580	1.000000	
*Dist {1}	-23.417	-22.563	1.000000	

tt	Max	Dep	*Dist {1,0} & *{stop}coda	*{stop}coda	*Dist {1} & *{stop}coda	*Dist {1,0}	*Dist {1}	
☞ p.t/k.t			*	*		*		-41.760
pi.t/ki.t		*						-53.496
t	*							-73.028

tn	Max	Dep	*Dist {1,0} & *{stop}coda	*{stop}coda	*Dist {1} & *{stop}coda	*Dist {1,0}	*Dist {1}	
☞ p.n/k.n/t.n			*	*	*	*	*	-45.445
pi.n/ki.n/ti.n		*						-53.496
n	*							-73.028

st	Max	Dep	*Dist {1,0} & *{stop}coda	*{stop}coda	*Dist {1} & *{stop}coda	*Dist {1,0}	*Dist {1}	
☞ s.t						*		0
si.t		*						-53.496
t	*							-73.028

sn	Max	Dep	*Dist {1,0} & *{stop}coda	*{stop}coda	*Dist {1} & *{stop}coda	*Dist {1,0}	*Dist {1}	
☞ s.n						*	*	0
si.n		*						-53.496
n	*							-73.028

Olhando para os valores de Harmonia dos candidatos ótimos no conjunto de *tableaux* 20, temos que os candidatos ótimos são [s.n], [s.t], [p.n, k.n, t.n] e [p.t, k.t]. Porém, se observarmos além da indicação do candidato ótimo (☞), mas o valor de Harmonia de cada candidato ótimo com relação aos candidatos perdedores, percebemos que [p.n, k.n, t.n] e [p.t, k.t] não são categóricos, ao contrário do que ocorre com [s.n] e [s.t]. Os candidatos ótimos com a plosiva em coda (“t.t”[p.t, k.t] = -41.760 e “tn” [p.n, k.n, t.n] = -45.445) expressam um valor de H próximo ao dos candidatos com epêntese (“t[i].t” e “t[i].n” = -53.496), indicando que, em algum momento de produção, os candidatos ótimos irão mudar.

Esses valores que determinam que tais candidatos ótimos devem variar não são aleatórios, visto que foram dados pela análise acústica que realizamos a partir das produções dos aprendizes de PB nestas sequências. Logo, as restrições responsáveis pelo valor de H de [p.n, k.n, t.n] (\*{stop}coda, \*Dist{1,0}, \*Dist{1}, \*Dist {1,0} & \*{stop}coda, \*Dist {1} & \*{stop}coda), devem se organizar numericamente, de forma a responder a esta situação. Assim, o fato de as restrições \*Dist{1,0} e \*Dist{1} receberem valor negativo, retirando-as do cálculo do valor de H, vai fazer com que as restrições \*{stop}coda, \*Dist {1,0} & \*{stop}coda e

\*Dist {1} & \*{stop}coda, somadas levem a um valor de Harmonia próximo ao do candidato que viola unicamente DEP.

O conjunto de *tableaux* 21, a seguir, tem o propósito de demonstrar essa variação entre os candidatos com a produção da plosiva em coda e com epêntese.

**Conjunto de Tableaux 21 - Simulação 3 HG / Interlúngua – Segunda rodada**

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>	
Max	72.819	75.745	1.000000	
Dep	49.345	47.479	1.000000	
*Dist {1,0} & *{stop}coda	20.652	21.416	1.000000	
*{stop}coda	20.652	20.279	1.000000	
*Dist {1} & *{stop}coda	3.481	7.303	1.000000	
*Dist {1,0}	-22.165	-20.562	1.000000	
*Dist {1}	-23.417	-21.399	1.000000	

tt	Max	Dep	*Dist {1,0} & *{stop}coda	*{stop}coda	*Dist {1} & *{stop}coda	*Dist {1,0}	*Dist {1}	
☞ p.t/kt			*	*		*		-41.695
pi.t/kit		*						-47.479
t	*							-75.745

tn	Max	Dep	*Dist {1,0} & *{stop}coda	*{stop}coda	*Dist {1} & *{stop}coda	*Dist {1,0}	*Dist {1}	
p.n/k.n/t.n			*	*	*	*	*	-48.998
☞ pi.n/ki.n/ti.n		*						-47.479
n	*							-75.745

Se compararmos os valores de *disharmony* do conjunto de *tableaux* 20 (MAX – 73.028, DEP – 53.496, \*Dist {1,0}&{\*stop}coda – 21.693, \*{stop}coda – 20.066, \*Dist {1}&{\*stop}coda – 3.686, \*Dist {1} – -22.563, \*Dist {1,0} – -21.580) com os do conjunto de *tableaux* 21, veremos que as mudanças numéricas dos pesos das restrições fizeram com que o candidato com epêntese (“t[i].n”), emergisse, atingindo valor de H superior ou mais próximo de zero, com relação ao candidato com a plosiva em coda (“t.n”). É dessa forma que o algoritmo de aprendizagem vai nos mostrar candidatos ótimos não-categóricos na HG.

É interessante notar que as hierarquias dadas pelo conjunto das restrições tanto no conjunto de *tableaux* 20, quanto no conjunto de *tableaux* 21, caso estivessem operando à luz da OT-GLA, implicariam, de imediato, um candidato ótimo categórico e sem epêntese, já que o valor central de DEP se encontra alto na hierarquia com relação aos valores das demais restrições, indicando a diferença na concepção de gramática para cada um dos modelos.

Se compararmos os valores de ranqueamento das restrições nos dois diferentes algoritmos vinculados à OT e HG, vamos ter, no geral, comportamentos numéricos bem diferentes. A restrição de fidelidade MAX, por exemplo, que em ambos os modelos conquista uma posição alta na hierarquia, superando as demais restrições, na HG terá seu alto valor explicado pela ação cumulativa que o modelo sugere. Conforme já salientamos, sob tal

modelo teórico, não basta que uma restrição esteja no topo da hierarquia como na OT, pois tal restrição, além disso, deve assumir valor maior que o valor de H dos demais candidatos, para que não deixe que o candidato com apagamento possa emergir.

Com relação aos *outputs* ótimos [s.t] e [s.n], levando em conta que as restrições  $*Dist\{1,0\}$  e  $*Dist\{1\}$  – que são violadas pelas sequências [s.t] e [s.n] - apresentam valor negativo, DEP poderia expressar qualquer valor positivo como peso, já que o valor negativo da restrição está indicando um valor de H igual a zero para ambas as sequências de fricativa em coda.

No que se refere à restrição DEP, vimos que, na OT, esta restrição assumiu valor tal que a fizesse obter um peso que superasse as restrições de marcação referentes às fricativas em coda ( $*Dist\{1\}$  e  $*Dist\{1,0\}$ ), deixando emergir os candidatos [s.t e s.n], e, ao mesmo tempo, que ficasse com valor próximo ao das restrições de marcação referentes às sequências com plosivas ( $*Dist\{1,0\} \& *\{stop\}coda$ ,  $*Dist\{1\} \& *\{stop\}coda$  e  $*\{stop\}coda$ ), que têm produção variável com a realização de epêntese, fazendo com que os candidatos “t[i].t” e “t[i].n”, em algum momento, fossem ótimos. Na HG, a restrição DEP vai assumir outra posição no ordenamento de restrições. Tal restrição vai assumir valor maior, mas próximo à soma dos valores das restrições de marcação, para que, assim como na OT, também expresse candidatos com e sem a plosiva em coda, variavelmente.

Nesta simulação, sempre vamos ter as restrições referentes à sequência “t.n” (obstruinte + nasal) mais próximas a DEP, o que indica que esta sequência, além de mais marcada, está assumindo maior ocorrência de epêntese (13%), com relação à “t.t” (obstruinte + obstruinte) (1%) e, conseqüentemente, com relação às sequências [s.t] e [s.n], que devem ter suas restrições sempre abaixo para que se mantenham categóricas.

Assim, em se tratando da aquisição da língua estrangeira dos aprendizes (Português Brasileiro), o algoritmo conseguiu convergir também na HG para uma simulação que sugere *outputs* categóricos e variáveis, ainda que com valores distintos aos dados na OT.

Em suma, ainda que ambos os algoritmos tenham convergido com essas restrições ao estágio final de gramática considerado nas três simulações realizadas, o resultado final das gramáticas de cada algoritmo apresenta características próprias ao modo como cada um deles seleciona o *output* ótimo.

Demonstramos, na figura 24, o resultado do recurso *output distributions*, representando as probabilidades de ocorrência simuladas em cada uma das sequências dadas nesta análise final.

row	rowLabel	?
1	tt → p.t/k.t	98880
2	tt → pi.t/ki.t	1120
3	tt → t	0
4	tn → p.n/k.n/t.n	86822
5	tn → pi.n/ki.n/ti.n	13178
6	tn → n	0
7	st → s.t	100000
8	st → si.t	0
9	st → t	0
10	sn → s.n	100000
11	sn → si.n	0
12	sn → n	0

**Figura 24 – Output Distributions da Interlíngua – HG**

Ao submetermos o conjunto de candidatos a 100.000 avaliações, a Gramática em questão resultou em índices iguais a 87% sem epêntese e 13% com a inserção da vogal, para a produção da plosiva em coda seguida de consoante nasal. Inputs com a sequência de obstruintes, apresentando 98% da produção da plosiva em coda e 2% com epêntese, são facilmente visíveis. Além disso, os *inputs* com a fricativa [s] mantêm-se sem epêntese, já que são possíveis não só no PB como na língua materna dos colombianos. Tais evidências demonstram a acuidade dos valores que o algoritmo nos forneceu.

### **5.2.2 Análise dos dados via Teoria da Otimidade Estocástica e Gramática Harmônica: Conjunto Completo de Restrições**

Antes de respondermos às Questões Norteadoras finais (3 a 6) propostas para esta pesquisa, que questionam, fundamentalmente, sobre quais restrições serão relevantes para o processo de aquisição dos padrões silábicos do PB como L2, bem como quais hierarquias dão conta das formas de *output* produzidas pelos aprendizes, trabalharemos com mais uma proposta de restrições, tentando dar conta de mais uma possibilidade silábica no PB, não

abrangida nas análises anteriores (Estudo-Piloto e análise final – Conjunto Restrito de Restrições).

Com as restrições que propusemos no Estudo-Piloto, tivemos resultados positivos junto aos algoritmos de aprendizagem HG e OT. Ou seja, os algoritmos convergiram às gramáticas-alvo estabelecidas, tanto no momento em que eram categóricas – caso da aquisição do Espanhol Colombiano e Português Brasileiro –, quanto na representação da fase desenvolvimental, enquanto relacionadas a *outputs* variáveis. Porém, como estávamos trabalhando com uma quantidade menor de dados, com relação à análise final, tivemos uma percepção de estarmos lidando com contato silábico e escala de ponto de articulação, o que nos fez trabalhar com restrições voltadas a esta situação.

Também no Estudo-Piloto, utilizando a escala de Clements (1990) - obstruintes (t), nasais (n), líquidas (l) glides (w) vogais (v) -, não consideramos que, dentre as obstruintes, temos as sequências com fricativa em coda, [s.n] e [s.t], como, por exemplo, na palavra “**as.no**” ou “**ás.pe.ro**” - palavras cognatas nas duas línguas em estudo - que nunca sofrerão epêntese no PB, diferentemente do que deve ocorrer com as demais obstruintes representadas por “t.t” e “t.n”.

Com o aumento de dados na análise final (tabelas 25 e 26), com relação aos dados utilizados no Estudo-Piloto (tabelas 23 e 24), percebemos que a escala de ponto não mais exerceria efeito em nossa análise, uma vez que apenas o contato silábico estaria sendo relevante para o fenômeno em questão. Dessa forma, a análise final – Conjunto Restrito de Restrições - se prestou a corrigir estas duas observações, acrescentando as sequências [s.t] e [s.n], que não podem ser tratadas da mesma forma que as demais obstruintes utilizadas no PB. Também na simulação em questão, foram desconsideradas as restrições estridentes  $\ast\{\text{stop}\}_{\text{coda}}$ ,  $\ast\{\text{k,p}\}_{\text{coda}}$ ,  $\ast\{\text{k}\}_{\text{coda}}$ , que tinham o propósito de lidar com a escala de ponto.

Dessa forma, para uma última análise, propomos utilizar restrições que contemplam toda a escala de contato silábico, considerando as famílias de restrições Dist de Gouskova (2004), para que possamos lidar com todas as possibilidades de distância entre contato silábico no PB e ainda caracterizarmos as restrições numa escala de marcação estridente por modo de articulação em coda. Assim, vamos ver emergirem *outputs* ótimos como “**car.ta**”, **cau.da**, **man.ta**, sem efeito da epêntese, já que são formados por uma estrutura silábica perfeitamente possível no PB.

Na análise final – Conjunto Restrito de Restrições, utilizamos apenas a restrição de marcação  $\ast\{\text{stop}\}_{\text{coda}}$ , desconsiderando os outros membros da escala. Se, por um lado, não utilizar todas as restrições não teria efeito na gramática da OT, na HG é possível que tal

tomada de decisão resulte em diferenças no cômputo final, uma vez que tal modelo considera o peso de todas as restrições utilizadas. Essa indagação motivou o aumento do número de restrições, inclusive porque desconhecemos discussão a esse respeito. Em suma, queremos perseguir o seguinte questionamento: devem todos os membros de uma escala ser formalizados através de restrições na HG, ainda que, sob a OT Estocástica, muitos desses membros já se mostrem irrelevantes?

No que segue, demonstramos as restrições de fidelidade e marcação formalizadas para esta última análise.

### **RESTRICÇÕES DE FIDELIDADE**

MAX

DEP

### **RESTRICÇÕES DE MARCAÇÃO**

\*Dist {+3} – Pior contato. O mais marcado.

\*Dist {+3, +2}

\*Dist {+3, +2, +1}

\*Dist {+3, +2, +1, 0}

\*Dist {+3, +2, +1, 0, -1}

\*Dist {+3, +2, +1, 0, -1, -2}

\*Dist {+3, +2, +1, 0, -1, -2, -3}

\*{stop}<sub>coda</sub>

\*{stop, fric}<sub>coda</sub>

\*{stop, fric, nas}<sub>coda</sub>

\*{stop, fric, nas, liq}<sub>coda</sub>

\*{stop, fric, nas, liq, glide}<sub>coda</sub>

\*Dist {1, 0} & \*{stop}<sub>coda</sub>

\*Dist {1} & \*{stop}<sub>coda</sub>

No que se refere às restrições de marcação, temos as restrições “\*Dist”, apresentadas em caráter estrigente, que servirão para representar oposição às distâncias de sonoridade entre a consoante em coda e a consoante da sílaba seguinte - nesse caso, plosivas ou nasais.

As restrições referentes a modo de articulação “\*{stop...glide}<sub>coda</sub>” dos segmentos em coda, também dadas em caráter estrigente, se oporão à ocorrência de determinado segmento em coda e as restrições conjuntas se oporão às consoantes plosivas em coda [p.t, k.t] e [p.n, k.n, t.n].

Em comparação à análise anterior, com o Conjunto Restrito de Restrições, contamos agora com um número superior de Restrições. Na outra análise estávamos trabalhando com as Restrições:

- DEP e MAX

- \*Dist {1, 0}, \*Dist {1}, \*{stop}<sub>coda</sub>, \*Dist {1, 0} & {stop}<sub>coda</sub> e \*Dist {1} & {stop}<sub>coda</sub>

Agora, estamos lidando com a escala de contato silábico completa das Restrições Dist de Gouskova (2004). Logo, nesta análise com o Conjunto Completo de Restrições, teremos, além das Restrições anteriores, as Restrições a seguir:

- \*Dist {+3}, \*Dist {+3, +2}, \*Dist {+3, +2, +1}, \*Dist {+3, +2, +1, 0}, \*Dist {+3, +2, +1, 0, -1}, \*Dist {+3, +2, +1, 0, -1, -2}, \*Dist {+3, +2, +1, 0, -1, -2, -3}, \*{stop, fric}<sub>coda</sub>, \*{stop, fric, nas}<sub>coda</sub>, \*{stop, fric, nas, liq}<sub>coda</sub>, \*{stop, fric, nas, liq, glide}<sub>coda</sub>.

### 5.2.2.1 Simulação 1 - Aquisição do Espanhol Colombiano (L1)

a) A aquisição do Espanhol Colombiano (L1) via Teoria da Otimidade Estocástica

Nesta etapa da análise que utiliza o conjunto completo de restrições, estamos trabalhando com 7 sequências de segmentos consonantais (“t.t” – obstruente + obstruente, “t.n” – obstruente + nasal, “n.t” – nasal + obstruente, “l.t” – líquida + obstruente, “w.t” – glide + obstruente, [s.t], [s.n]) que representam as possibilidades de contato silábico entre os segmentos que compõem a escala de sonoridade de Clements (1990) – glides (w), nasais (n), líquidas (l), obstruintes (t), com exceção da vogal.

Cada sequência que utilizamos representará uma distância de sonoridade dada pelo mecanismo de alinhamento relacional de Gouskova (2004), com o cuidado de termos apenas sequências compostas de um segmento consonantal em coda e uma obstruente ou nasal em onset da sílaba seguinte. Somadas a estas sequências da escala “Dist”, ainda temos [s.t] e [s.n], que, dentre as sequências de obstruintes utilizadas (“t.t” – obstruente + obstruente, “t.n” – obstruente + nasal), têm comportamento diferente dos demais segmentos, pois, em posição de coda medial no PB, não são passíveis de epêntese.

Dando início à primeira simulação, alimentamos o sistema com dados que tratam da aquisição do Espanhol Colombiano, ou seja, que expressem, para todas as sequências de segmentos utilizadas (“t.t” – obstruente + obstruente, “t.n” – obstruente + nasal, “n.t” – nasal + obstruente, “l.t” – líquida + obstruente, “w.t” – glide + obstruente, [s.t], [s.n]), 100% da

produção dos segmentos em coda, sem epêntese ou apagamento, tal como deve ocorrer no Espanhol.

Como o Espanhol Colombiano é a língua materna dos nossos aprendizes de Português Brasileiro, o sistema é programado a responder a um estágio inicial de um bebê adquirindo as consoantes em coda no Espanhol Colombiano.

A seguir, demonstramos os valores apresentados pelo algoritmo OT-GLA para a aquisição do Espanhol Colombiano.

Conjunto de *Tableaux* 22 - Simulação 1 OT / Aquisição do Espanhol Colombiano (0% de epêntese)

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
<b>*Dist {+3}</b>	100.000	100.830	1.000000
<b>*Dist {+3, +2}</b>	100.000	98.947	1.000000
<b>Max</b>	64.120	67.952	1.000000
<b>Dep</b>	64.947	60.265	1.000000
<b>*Dist {1} &amp; {stop}coda</b>	50.193	48.349	1.000000
<b>*Dist {+3, ...+1}</b>	32.870	33.138	1.000000
<b>*Dist {1,0} &amp; {stop}coda</b>	30.027	31.651	1.000000
<b>*{stop}coda</b>	30.027	29.004	1.000000
<b>*Dist {+3, ... 0}</b>	-4.003	-4.045	1.000000
<b>*{stop, fric}coda</b>	-4.003	-7.097	1.000000
<b>*Dist {+3, ...-1}</b>	-16.793	-15.529	1.000000
<b>*{stop, ... nas}coda</b>	-16.793	-16.823	1.000000
<b>*Dist {+3, ...-2}</b>	-22.087	-23.439	1.000000
<b>*{stop, ... liq}coda</b>	-22.087	-25.167	1.000000
<b>*{stop, ... glide}coda</b>	-29.067	-28.050	1.000000
<b>*Dist {+3, ...-3}</b>	-29.067	-31.327	1.000000



Nesta primeira simulação (conjunto de *tableaux* 22), conforme aconteceu no Estudo-Piloto e na Análise Final – Conjunto Restrito de Restrições, o algoritmo aprendeu o Espanhol Colombiano, que em seu estágio de fluência deve articular as consoantes plosivas em coda, sem a utilização de epêntese, que tende a ocorrer após plosivas no PB. Sendo assim, é necessário que DEP seja promovida, assumindo uma posição na hierarquia que supere as restrições vinculadas às consoantes em coda, de modo a fazer com que o candidato com epêntese receba uma violação fatal, por violar uma restrição mais altamente ranqueada.

Juntamente com a promoção da restrição de fidelidade DEP, devemos ter as restrições MAX e \*Dist {+3}, \*Dist{+3, +2}, que se referem, respectivamente, a apagamento e contatos inexistentes no EC, acima das demais restrições de marcação, independentemente dos pesos que assumam, para que, neste modelo de dominância estrita, recebam posições mais altas no ordenamento das restrições e excluam candidatos com apagamento e epêntese da disputa pelo *output* ótimo.

Ainda que a cada avaliação tenhamos diferentes pontos de seleção para MAX, DEP, \*Dist {+3}, \*Dist {+3, +2}, o importante é que tais restrições recebam valores que os mantenham distanciadas de todas as demais restrições, para que o *output* ótimo seja sempre o mesmo. Ou seja, estas quatro restrições podem receber valores próximos que as façam variar entre si, em termos de *status* hierárquico, em diversos momentos de produção linguística, mas não podem estar próximas das outras restrições que se referem à oposição das sequências de consoante em coda que estamos utilizando, pois isso alteraria o candidato ótimo que deve ser categórico.

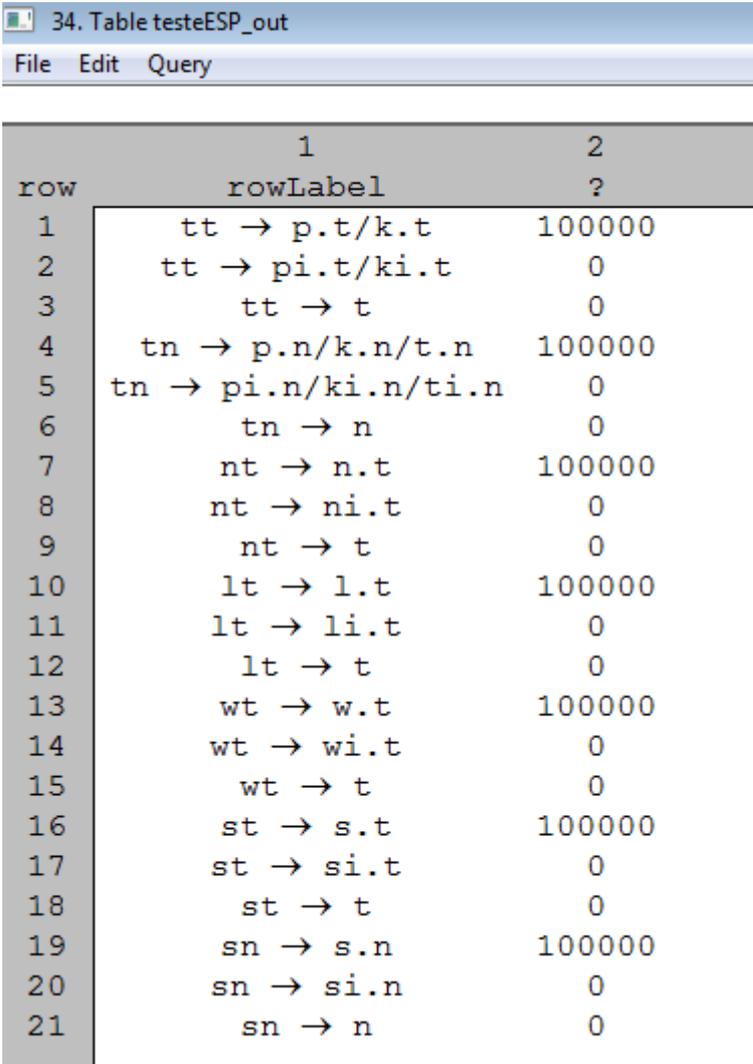
Quanto às restrições de marcação, temos valores de ranqueamento que respeitam um ordenamento de marcação, que atribuem índices mais baixos às restrições que no conjunto de estringência fazem oposição a candidatos menos marcados e mais gerais. Conseguimos demonstrar, contudo, uma aquisição em ordem de marcação, onde “t.n” (Dist {+1}) é mais marcada do que “t.t” (Dist {0}), que, por sua vez, é mais marcada do que “n.t” (Dist {-1}) e “l.t” (Dist {-2}) e “w.t” (Dist {-3}), por estringência. As restrições que fazem oposição aos candidatos menos marcados, dessa forma, devem ser mais violadas e, portanto, mais demovidas, assumindo posições mais baixas no *ranking*.

Assim, as restrições \*Dist {+3,...-3} e \*{stop, ...glide}<sub>coda</sub>, que abrangem todas as proibições de distância entre os segmentos em coda e *onset* e todas as proibições de modo de articulação em coda, são as restrições mais baixas na hierarquia, demovidas por estringência. Desse modo, a escala segue sucessivamente até chegarmos às restrições mais marcadas e específicas, que, juntamente com a restrição de fidelidade DEP, se encontram mais altamente

ranqueadas: \*Dist{1} & {stop}coda e \*{+3,...+1}, que apresentam oposição a [k.n, p.n, t.n], sequências mais marcadas entre as analisadas.

Logo, o algoritmo consegue convergir para uma gramática que se apresenta categórica na simulação 1 da OT Estocástica, a partir da qual os segmentos consonantais em codas mediais são produzidos sem epêntese (estágio final da aquisição), seguindo uma lógica de marcação que é determinada por estringência.

A distribuição probabilística dos dados, realizada no modelo da Teoria da Otimidade Estocástica, pode ser visualizada na figura 25.



row	rowLabel	2
1	tt → p.t/k.t	100000
2	tt → pi.t/ki.t	0
3	tt → t	0
4	tn → p.n/k.n/t.n	100000
5	tn → pi.n/ki.n/ti.n	0
6	tn → n	0
7	nt → n.t	100000
8	nt → ni.t	0
9	nt → t	0
10	lt → l.t	100000
11	lt → li.t	0
12	lt → t	0
13	wt → w.t	100000
14	wt → wi.t	0
15	wt → t	0
16	st → s.t	100000
17	st → si.t	0
18	st → t	0
19	sn → s.n	100000
20	sn → si.n	0
21	sn → n	0

Figura 25 - *Output Distributions* da aquisição do Espanhol Colombiano – OT

Conforme é possível ver no *output distributions*, se compararmos tais valores com os utilizados na simulação (tabela 25), veremos que os índices percentuais, após a submissão dos candidatos às 100.000 avaliações, são os percentuais desejados.

b) A aquisição do Espanhol Colombiano (L1) via Gramática Harmônica

Relembrando que, nesta simulação, as restrições de marcação receberam um valor numérico inicial igual a 100 e as restrições de fidelidade DEP e MAX um valor numérico igual a 0 no estágio inicial, temos que a diferença entre os dois algoritmos vai se dar na resposta numérica para o estágio final. Isso se explica porque, enquanto na OT Estocástica temos os pesos das restrições classificados em um *ranking* entre elas, na HG, o candidato ótimo vai se dar a partir do cálculo do valor de Harmonia. Dessa forma, o valor das restrições de marcação também deve ser afetado pelo algoritmo, de maneira a terem seus pesos diminuídos, numa proporção diferenciada com relação à OT, para que mantenham os *outputs* ótimos após o cálculo de H, conforme verificamos no conjunto de *tableaux* 23.

**Conjunto de Tableaux 23 - Simulação 1 HG / Aquisição do Espanhol Colombiano (0% de epêntese)**

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
<b>*Dist {+3, +2}</b>	100.000	102.090	1.000000
<b>*Dist {+3}</b>	100.000	95.380	1.000000
<b>Dep</b>	88.626	92.441	1.000000
<b>Max</b>	89.792	88.833	1.000000
<b>*Dist {1} &amp; {stop}coda</b>	39.061	40.793	1.000000
<b>*{stop}coda</b>	11.140	13.855	1.000000
<b>*Dist {1,0} &amp; {stop}coda</b>	11.140	12.086	1.000000
<b>*Dist {+3, ...+1}</b>	9.423	9.048	1.000000
<b>*{stop, fric}coda</b>	-31.453	-29.687	1.000000
<b>*Dist {+3, ... 0}</b>	-31.453	-31.262	1.000000
<b>*Dist {+3, ...-1}</b>	-53.405	-51.027	1.000000
<b>*{stop, ... nas}coda</b>	-53.405	-55.718	1.000000
<b>*{stop, ... liq}coda</b>	-66.469	-63.629	1.000000
<b>*Dist {+3, ...-2}</b>	-66.469	-68.799	1.000000
<b>*Dist {+3, ...-3}</b>	-78.418	-76.188	1.000000
<b>*{stop, ... glide}coda</b>	-78.418	-77.640	1.000000



Os candidatos ótimos são aqueles que apresentam a articulação do segmento consonantal em coda. Já que estamos ainda trabalhando com *outputs* categóricos, além de apresentarem valores superiores aos dos demais candidatos, os vencedores mantiveram a distância necessária entre os valores de Harmonia dos demais candidatos a *output*. Por mais que, a cada avaliação, tenhamos valores diferentes de ponto de seleção para as restrições, tais diferenças não chegarão a causar variação entre os candidatos perdedores e os candidatos ótimos, dada a distância nos valores de Harmonia entre os candidatos perdedores e vencedores.

No que diz respeito ao ordenamento das restrições, assim como na OT, tivemos as restrições de fidelidade e as restrições de marcação  $*\text{Dist}\{+3\}$  e  $*\text{Dist}\{+3, +2\}$  com os valores mais altos de ranqueamento. Porém, em se comparando os índices numéricos obtidos na presente simulação aos valores de ranqueamento que obtivemos na OT, vamos ver que tais restrições - mais altas no *ranking*, no estágio final de aquisição do Espanhol Colombiano - neste algoritmo da HG, tiveram seus valores superiores aos obtidos na OT. Isso também ocorreu com as restrições de marcação, que tiveram seus pesos mais demovidos aos dados no ranqueamento da OT. Enquanto DEP apresentou, na OT, um valor de ranqueamento igual a 64.947, na HG, DEP tem seu valor igual a 88.626, para poder desbancar, por exemplo, a sequência mais marcada [p.n, k.n, t.n], cujo candidato tem seu valor de H igual a -69.612.

É interessante observar que o algoritmo da HG manteve as restrições de marcação referentes às sequências mais marcadas ( $*\text{Dist}\{+3, \dots, +1\}$ ,  $*\text{Dist}\{1\}$  &  $*\{\text{stop}\}_{\text{coda}}$ ,  $*\text{Dist}\{1,0\}$  &  $*\{\text{stop}\}_{\text{coda}}$  e  $*\{\text{stop}\}_{\text{coda}}$ ) com valor positivo, em detrimento das demais restrições de marcação que ficaram com valores negativos, sendo excluídas do cálculo do valor de H. Para esta simulação, bastava termos DEP e MAX assumindo posição superior à cumulatividade do valor das demais restrições, referente a cada candidato ótimo. O algoritmo deu conta desta demanda excluindo do cálculo as restrições menos marcadas, mais baixas na hierarquia, facilitando a DEP e MAX a tarefa de virem a ter de assumir valores tão altos para que, no cálculo da H, os mantivesse vencedores diante dos candidatos produzidos com a consoante em coda.

Dessa forma, em se tratando da aquisição da língua materna dos aprendizes, em um sistema estridente como o aqui apresentado, o algoritmo conseguiu convergir também na HG para uma simulação que sugere *outputs* categóricos, ainda que com valores distintos aos dados na OT. DEP e MAX apresentaram valores de ponto de seleção que permitiram que os candidatos que os violam atingissem valores de Harmonia superiores aos daqueles que violam

as restrições de marcação (acima de 10 pontos), não permitindo, portanto, que houvesse variação nos candidatos ótimos.

A figura 26, a seguir, visa a demonstrar a distribuição probabilística dos dados na Gramática Harmônica, que deve ser igual à verificada na OT, já que os dados de entrada são os mesmos para os dois modelos.

row	rowLabel	2
1	tt → p.t/k.t	100000
2	tt → pi.t/ki.t	0
3	tt → t	0
4	tn → p.n/k.n/t.n	100000
5	tn → pi.n/ki.n/ti.n	0
6	tn → n	0
7	nt → n.t	100000
8	nt → ni.t	0
9	nt → t	0
10	lt → l.t	100000
11	lt → li.t	0
12	lt → t	0
13	wt → w.t	100000
14	wt → wi.t	0
15	wt → t	0
16	st → s.t	100000
17	st → si.t	0
18	st → t	0
19	sn → s.n	100000
20	sn → si.n	0
21	sn → n	0

**Figura 26 - Output Distributions da aquisição do Espanhol Colombiano – HG**

Dessa forma, podemos comprovar a acuidade dos valores fornecidos pelo algoritmo, pois a gramática em questão, após avaliação dos candidatos, está apresentando os índices percentuais desejados, de 100% da produção das consoantes em coda, para todas as sequências heterossilábicas utilizadas nesta análise Parte 2.

### 5.2.2.2 Simulação 2 - Aquisição do Português Brasileiro (L2)

a) A aquisição do Português Brasileiro (L2) via Teoria da Otimidade Estocástica

A aquisição plena do PB vai ter no estágio final de aquisição a produção de 100% de epêntese após as consoantes plosivas em posição de coda. Como estamos trabalhando com a escala de Clements (1990), estamos tratando as plosivas em conjunto com as fricativas e africadas, sob a categoria das obstruintes (t). No PB, esta classificação, em princípio, não parece ser desejável, porque dentre as fricativas temos o segmento [s], perfeitamente possível em posição de coda, ex.: “cos.ta” em comparação aos demais segmentos que ensejam o acréscimo da epêntese como correção a uma situação indesejada como em “a. f[i]. ta” ou “ca.[ki].to”.

No entanto, a partir da presente análise – Conjunto Completo das Restrições -, é fundamental que tenhamos a diferenciação das demais sequências formadas de obstruinte em posição de coda, com as sequências [s.t] e [s.n], tal como viemos observando desde a análise com o conjunto restrito de restrições.

Ao alimentarmos o sistema com os valores das restrições obtidos na simulação anterior (estágio final do Espanhol Colombiano e inicial para o Português Brasileiro), temos que o estágio inicial da L2 vai contar com as restrições de fidelidade (MAX e DEP) dominando as restrições de marcação, na hierarquia das restrições propostas para lidar com a aquisição de sequências com consoantes em coda medial de palavra. O estágio final é dado no conjunto de *tableaux* 24.

Conjunto de *Tableaux* 24 - Simulação 2 OT / Aquisição do Português Brasileiro (100% de epêntese)

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
<b>*Dist {+3, +2}</b>	100.000	99.949	1.000000
<b>*Dist {+3}</b>	100.000	98.415	1.000000
<b>Max</b>	64.120	61.648	1.000000
<b>*{stop}coda</b>	56.644	55.124	1.000000
<b>*Dist {1} &amp; {stop}coda</b>	55.997	54.909	1.000000
<b>*Dist {1,0} &amp; {stop}coda</b>	56.644	54.779	1.000000
<b>Dep</b>	44.951	49.044	1.000000
<b>*Dist {+3, ...+1}</b>	32.053	35.545	1.000000
<b>*Dist {+3, ... 0}</b>	15.993	15.302	1.000000
<b>*{stop, fric}coda</b>	15.993	13.117	1.000000
<b>*{stop, ... nas}coda</b>	3.203	5.804	1.000000
<b>*Dist {+3, ...-1}</b>	3.203	2.758	1.000000
<b>*{stop, ... liq}coda</b>	-2.091	2.306	1.000000
<b>*Dist {+3, ...-2}</b>	-2.091	-2.040	1.000000
<b>*Dist {+3, ...-3}</b>	-9.071	-8.477	1.000000
<b>*{stop, ... glide}coda</b>	-9.071	-11.179	1.000000



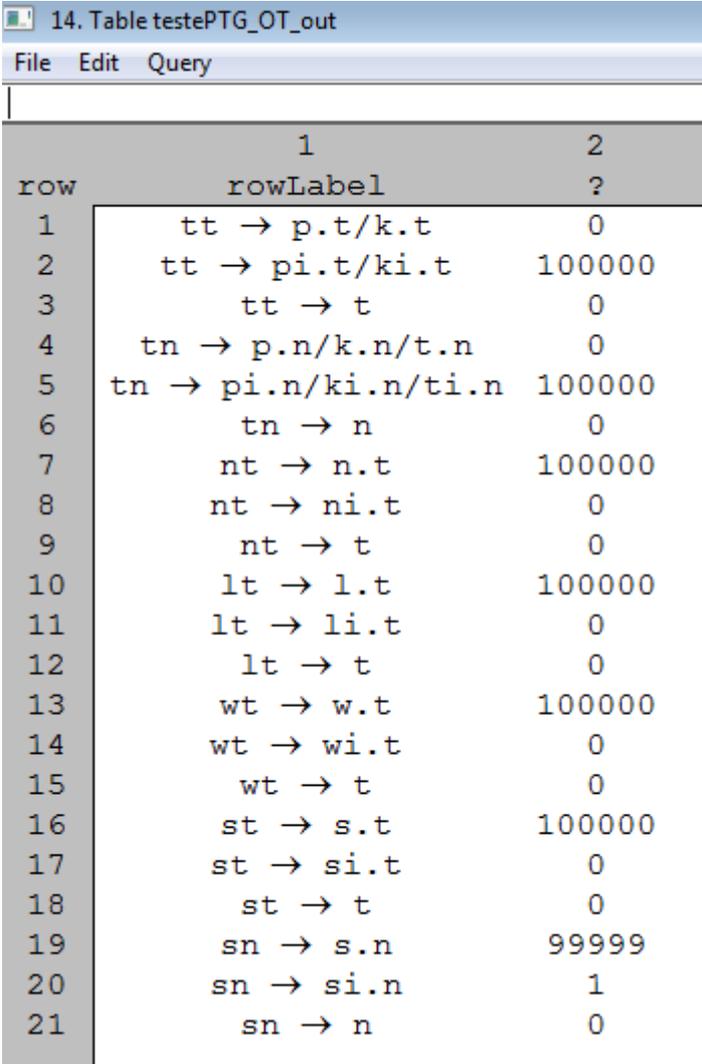
Como nesta simulação – Conjunto Completo das Restrições, o algoritmo de aprendizagem deve atingir um sistema-alvo a partir do qual plosivas finais são epentetizadas, mas a fricativa [s] em coda e demais sequências sem plosiva em coda são permitidas, DEP, que atingia as posições mais altas nas hierarquias de HG e OT na simulação anterior, neste momento, deve ser demovida na hierarquia. Por sua vez, a restrição de fidelidade MAX e as restrições de marcação  $*Dist\{+3\}$  e  $*Dist\{+3, +2\}$ , como nunca serão responsáveis pela emergência de candidatos ótimos, devem permanecer em todas as simulações nas posições mais altas na hierarquia.

Como não temos apagamento em nenhum dos sistemas (Português Brasileiro ou Espanhol Colombiano) em análise, a promoção de MAX indica que serão excluídos da disputa candidatos com apagamento. As restrições de marcação  $*Dist\{+3\}$  e  $*Dist\{+3, +2\}$  mantêm-se com valores altos, pois dentre os candidatos a *output* que estamos considerando na análise, nenhum deles é capaz de violar tais restrições, já que estas sequências de contato silábico entre obstruinte em coda e glide (“t.w”) e obstruinte e líquida (“t.l”) ou nasal e glide (“n.w”) não constituem estruturas admitidas nos moldes silábicos nem no PB, nem no EC.

O papel de DEP é fundamental em se tratando de segmentos plosivos em coda no PB, foco desta pesquisa, visto que é com a mobilidade desta restrição de fidelidade que vai ser possível a emergência dos candidatos que estamos almejando. Ao falarmos de Teoria da Otimidade, em que o candidato ótimo se dá a partir da hierarquia de restrições dada pelos pesos das restrições em dominância estrita, vemos que DEP (49.044), nesta simulação, só precisou encontrar um peso que fosse inferior e mantivesse um distanciamento (distância de 10 pontos) das restrições específicas à plosiva em coda ( $*Dist\{1\} \& *\{stop\}_{coda} = 54.909$ ,  $\{stop\}_{coda} = 55.124$  e  $*Dist\{1,0\} \& *\{stop\}_{coda} = 54.779$ ), ao mesmo tempo em que fosse superior e mantivesse um valor de distanciamento das demais restrições de marcação.

Assim, em se tratando da aquisição plena da língua estrangeira – Português Brasileiro - pelos aprendizes colombianos, o algoritmo conseguiu convergir na OT Estocástica para uma simulação com *outputs* categóricos. Porém, enquanto no Espanhol Colombiano era importante que DEP assumisse valor superior ao das restrições de marcação, no PB, tal restrição teve de ser demovida até encontrar um peso que superasse as restrições referentes aos *outputs* ótimos sem epêntese [s.n, s.t], “w.t”, “l.t”, “n.t” e, ao mesmo tempo, estivesse abaixo das restrições referentes aos *outputs* com epêntese “t[i].t” e “t[i].n” (obstruintes em coda).

Demonstramos, através do *output distributions* (figura 27), os índices percentuais obtidos, após o processo de avaliação a que submetemos os candidatos a *output* no comando “to *output distributions*” do software PRAAT.



row	rowLabel	?
1	tt → p.t/k.t	0
2	tt → pi.t/ki.t	100000
3	tt → t	0
4	tn → p.n/k.n/t.n	0
5	tn → pi.n/ki.n/ti.n	100000
6	tn → n	0
7	nt → n.t	100000
8	nt → ni.t	0
9	nt → t	0
10	lt → l.t	100000
11	lt → li.t	0
12	lt → t	0
13	wt → w.t	100000
14	wt → wi.t	0
15	wt → t	0
16	st → s.t	100000
17	st → si.t	0
18	st → t	0
19	sn → s.n	99999
20	sn → si.n	1
21	sn → n	0

**Figura 27 - Output Distributions da aquisição do Português Brasileiro – OT**

Podemos comprovar, assim, que a gramática em questão apresenta os índices percentuais desejados, de 100% de produção de epêntese para o caso das sequências formadas de obstruintes em coda [t.t, t.n], bem como de 100% de produção da consoante em coda, para as demais sequências [n.t, l.t, w.t, s.t, s.n].

#### b) A aquisição do Português Brasileiro (L2) via Gramática Harmônica

Nesta simulação do sistema do Português à luz da HG, alimentamos o algoritmo com as mesmas informações fornecidas ao algoritmo OT-GLA. O algoritmo foi informado de que, no sistema-alvo, *inputs* com plosivas seguidas de outra obstruinte ou nasal devem ser

epentetizados e que *inputs* com a fricativa [s], além de que as sequências compostas de “n.t”(nasal em coda e obstruente em *onset* da sílaba seguinte), “l.t” (líquida em coda e obstruente em *onset*), “w.t” (glide em coda e obstruente em *onset* da sílaba seguinte) devem manter-se sem epêntese.

O conjunto de *tableaux* 25 possibilita a verificação da resposta do algoritmo da HG para a aquisição plena do Português Brasileiro (L2) e, dessa forma, permite-nos comparar o presente resultado com aquele obtido da OT Estocástica.

**Conjunto de *Tableaux* 25 - Simulação 2 HG / Aquisição do Português Brasileiro (100% de epêntese)**

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
<b>*Dist {+3}</b>	100.000	101.427	1.000000
<b>*Dist {+3, +2}</b>	100.000	98.707	1.000000
<b>Max</b>	89.792	87.900	1.000000
<b>Dep</b>	60.725	59.613	1.000000
<b>*Dist {1} &amp; {stop}coda</b>	41.838	40.943	1.000000
<b>*{stop}coda</b>	39.041	40.482	1.000000
<b>*Dist {1,0} &amp; {stop}coda</b>	39.041	38.616	1.000000
<b>*Dist {+3, ...+1}</b>	12.200	8.197	1.000000
<b>*{stop, fric}coda</b>	-3.552	-2.626	1.000000
<b>*Dist {+3, ... 0}</b>	-3.552	-4.345	1.000000
<b>*Dist {+3, ...-1}</b>	-25.504	-24.888	1.000000
<b>*{stop, ... nas}coda</b>	-25.504	-26.827	1.000000
<b>*Dist {+3, ...-2}</b>	-38.568	-37.087	1.000000
<b>*{stop, ... liq}coda</b>	-38.568	-39.476	1.000000
<b>*{stop, ... glide}coda</b>	-50.517	-51.347	1.000000
<b>*Dist {+3, ...-3}</b>	-50.517	-53.283	1.000000



Ao verificar os dados de aquisição via HG, cabe-nos questionar como o alto número de restrições utilizadas – decorrente das escalas que embasaram os conjuntos estridentes de restrições – podem (ou não) exercer efeitos sobre o valor de Harmonia dos candidatos.

Diferentemente da OT Estocástica, neste modelo o que importa são os valores de H. Assim, não é surpresa termos DEP com um valor de ranqueamento que o faz aparecer, numericamente, acima das demais restrições de marcação, relacionadas às sequências formadas de segmentos consonantais em coda. Se o que importa é DEP assumir valor superior ao da soma das restrições que violam os candidatos com plosiva em coda, ao mesmo tempo em que exibe um valor inferior à Harmonia dos demais candidatos, não podemos nos ater unicamente aos valores do *ranking*.

Ao olharmos para os valores de Harmonia dos candidatos ótimos no conjunto de *tableaux* 25, vimos que esta meta foi alcançada e que, ainda, foi possível garantir a distância necessária entre o valor de Harmonia (H) dos candidatos ótimos e dos candidatos perdedores, distância essa que faz com que não tenhamos variação nos candidatos ótimos.

No que se refere à restrição de fidelidade MAX e às restrições de marcação \*Dist{+3} e \*Dist{+3, +2}, reafirmamos que não coincidentemente, em todas as simulações – aquisição de L1, aquisição de L2 e Interlíngua – e em ambos os algoritmos de aprendizagem – HG e OT – tais restrições sempre serão promovidas com relação às demais. Além disso, \*Dist{+3} e \*Dist{+3, +2}, que nunca serão violadas pelas sequências de segmentos propostos (pois só estamos trabalhando com sequências existentes no Português Brasileiro (PB) e Espanhol Colombiano (EC), e tais restrições contemplam contatos não existentes nem no PB, nem no EC (como, por exemplo, uma consoante plosiva em coda seguida de um glide ou uma consoante nasal em coda, também seguida de um glide), receberão sempre os valores mais altos de ranqueamento, de modo que seus valores não se modifiquem de simulação para simulação e de algoritmo para algoritmo, pelo fato de tais restrições não sofrerem promoção ou demissão em função da evidência positiva.

Em outras palavras, se observarmos todos os conjuntos de *tableaux* nesta Análise Final – Conjunto Completo de Restrições (conjunto de *tableaux* 22 ao 27) -, veremos que o valor de ranqueamento para as restrições \*Dist{+3} e \*Dist{+3, +2} será sempre o máximo, igual a 100.000. Isto quer dizer que sob nenhuma hipótese tais restrições serão demovidas, frente às sequências presentes na evidência positiva desta análise “t.t”, ”t.n”, ”n.t”, ”l.t”, ”w.t”, [s.t] e [s.n]. É importante mencionar que a restrição de fidelidade MAX, vai ter seu valor diferenciado de acordo com o algoritmo utilizado. Na OT, a restrição em questão vai

assumir valor de ranqueamento igual a 64.120 em todas as simulações; por sua vez, na HG, esta mesma restrição vai assumir valores mais altos, em torno de 90.000, pois não basta que ela esteja alta na hierarquia, mas, sim, que o valor de Harmonia do candidato que exhibe apagamento seja mais negativo do que o valor do candidato ótimo.

Vimos, na análise desenvolvida à luz da OT, que DEP teve de assumir um valor que o fizesse estar abaixo das restrições de marcação específicas às sequências com a consoante plosiva em coda “t.t” e “t.n”, para que fosse possível a emergência da epêntese nestas sequências. Na HG, para que possamos ver candidatos ótimos com epêntese apenas seguidos de obstruintes em coda (com exceção de [s.n] e [s.t]), só precisamos que DEP assumira valor maior, de modo que o candidato com epêntese apresente um valor de H superior ao das sequências com plosiva em coda. Isto quer dizer que o candidato que viola unicamente DEP deve estar acima de valores de H com relação às seguintes sequências:

$$[t.t] = (-1 \text{ violação } X * \text{Dist}\{1,0\} \& * \{\text{stop}\}_{\text{coda}}) + (-1 \text{ violação } X * \{\text{stop}\}_{\text{coda}})$$

$$[t.t] = (-38.616) + (-40.482) = -79.098$$

$$[t.n] = (-1 X * \text{Dist}\{+3, \dots +1\}) + (-1 X * \{\text{stop}\}_{\text{coda}}) + (-1 \text{ violação } X * \text{Dist}\{1,0\} \& * \{\text{stop}\}_{\text{coda}}) + (-1 \text{ violação } X * \text{Dist}\{1\} \& * \{\text{stop}\}_{\text{coda}})$$

$$[t.n] = (-8.197) + (-40.482) + (-38.616) + (-40.943) = -128.238$$

Estar acima quer dizer que DEP deve levar o candidato com epêntese a ter valor menos negativo, ou mais próximo a zero, com relação aos valores de Harmonia calculados. Por outro lado, ao mesmo tempo, o candidato que viola unicamente DEP deve ser mais negativo do que os candidatos que exibem as demais sequências [s.t], [s.n], n.t, l.t, w.t, não epentetizadas no PB. Como a HG trabalha com a cumulatividade dos pesos das restrições, seria difícil DEP alcançar valor mais negativo do que as restrições referentes à sequência [s.n], por exemplo, que requer a soma de  $*\text{Dist}\{+3, \dots +1\}$ ,  $*\{\text{stop, fric}\}_{\text{coda}}$ ,  $*\text{Dist}\{+3, \dots 0\}$ ,  $*\text{Dist}\{+3, \dots -1\}$ ,  $*\{\text{stop, \dots nas}\}_{\text{coda}}$ ,  $*\text{Dist}\{+3, \dots -2\}$ ,  $*\{\text{stop, \dots liq}\}_{\text{coda}}$ ,  $*\{\text{stop, \dots glide}\}_{\text{coda}}$ ,  $*\{+3, \dots -3\}_{\text{coda}}$ , não fosse a estratégia utilizada pelo algoritmo de fazer com que estas restrições atinjam valor negativo, sendo anulados do cálculo de H.

Utilizando na HG a opção “*Linear OT*”, no programa PRAAT, assumimos que restrições com pesos negativos devem ser excluídas do cálculo de H. Para esta situação de aquisição do PB, em que DEP teve de assumir duas posições diante das restrições de marcação postas em estringência – na qual a violação de uma restrição mais específica implica violação às restrições mais gerais - esta opção foi importante. Foi possível, dessa forma, que as consoantes em coda fossem adquiridas no PB, com a particularidade de a

aquisição das plosivas, em detrimento das demais consoantes em coda, exigirem a inserção da epêntese para a aquisição plena do Português Brasileiro.

Assim, em se tratando da aquisição da língua estrangeira dos aprendizes colombianos (Português Brasileiro), o algoritmo conseguiu convergir também na HG para uma simulação que sugere *outputs* categóricos, ainda que com valores distintos aos dados na OT. Em outras palavras, o algoritmo conseguiu atingir o resultado esperado, mesmo diante de restrições conjuntas e estridentes que pareciam ser redundantes em um modelo que trabalha com a cumulatividade do peso das restrições.

Demonstramos o *output distributions*, representando as probabilidades de ocorrência de cada uma das sequências dadas nesta análise final com o conjunto completo de restrições, após a submissão dos candidatos a 100.000 avaliações, conforme mostra a figura 28.

row	1 rowLabel	2 ?
1	tt → p.t/k.t	0
2	tt → pi.t/ki.t	100000
3	tt → t	0
4	tn → p.n/k.n/t.n	0
5	tn → pi.n/ki.n/ti.n	100000
6	tn → n	0
7	nt → n.t	100000
8	nt → ni.t	0
9	nt → t	0
10	lt → l.t	100000
11	lt → li.t	0
12	lt → t	0
13	wt → w.t	100000
14	wt → wi.t	0
15	wt → t	0
16	st → s.t	100000
17	st → si.t	0
18	st → t	0
19	sn → s.n	100000
20	sn → si.n	0
21	sn → n	0

Figura 28 - *Output Distributions* da aquisição do Português Brasileiro – HG

Dada, na figura 28, as probabilidades de ocorrências do conjunto de candidatos a *output* com que estamos lidando, podemos afirmar a acuidade dos valores fornecidos pelo algoritmo, levando em conta que estamos considerando 100% de produção de epêntese para as sequências formadas com consoante obstruinte em coda na aquisição plena do Português Brasileiro, conforme já discutido na seção 2.2.1 - dos padrões silábicos no PB -, no Referencial Teórico.

### 5.2.2.3 Simulação 3 – *Interlândia* (*outputs variáveis*)

a) A aquisição dos *outputs* variáveis (*Interlândia*) via Teoria da Otimidade Estocástica

Nesta simulação, devemos observar e analisar o tratamento de ambos os algoritmos de aprendizagem frente à gramática em desenvolvimento dos aprendizes, com *outputs* variáveis. Diferentemente das simulações 1 e 2, os valores centrais das restrições DEP e das restrições específicas às sequências com plosiva em coda, neste momento, devem ser muito próximos. As alterações nos valores dos pontos de seleção, diante de avaliações sucessivas, não representarão apenas uma mera variação numérica dos pesos das restrições, mas motivação para que tenhamos mudança no *ranking* das restrições, no caso da OT, e valores de H próximos, no caso da HG.

Partindo do pressuposto de que contamos com colombianos - cuja língua materna é o Espanhol – aprendendo Português, o estágio inicial nesta simulação deve ser a aquisição plena do Espanhol, tal como fizemos na simulação anterior, e o estágio final é esta etapa desenvolvimental da língua, em direção à aquisição plena do PB, conforme podemos ver no conjunto de *tableaux* 26.

Conjunto de *Tableaux* 26 - Simulação 3 OT / Interlíngua (*outputs* variáveis)

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
<b>*Dist {+3}</b>	100.000	100.659	1.000000
<b>*Dist {+3, +2}</b>	100.000	98.071	1.000000
<b>Max</b>	65.078	65.989	1.000000
<b>Dep</b>	50.650	49.310	1.000000
<b>*{stop}coda</b>	43.366	44.198	1.000000
<b>*Dist {1} &amp; {stop}coda</b>	47.595	44.164	1.000000
<b>*Dist {1,0} &amp; {stop}coda</b>	43.366	42.893	1.000000
<b>*Dist {+3, ...+1}</b>	30.272	32.248	1.000000
<b>*{stop, fric}coda</b>	9.336	10.670	1.000000
<b>*Dist {+3, ... 0}</b>	9.336	5.783	1.000000
<b>*Dist {+3, ...-1}</b>	-3.454	-1.435	1.000000
<b>*{stop, ... nas}coda</b>	-3.454	-3.391	1.000000
<b>*Dist {+3, ...-2}</b>	-8.748	-8.044	1.000000
<b>*{stop, ... liq}coda</b>	-8.748	-10.300	1.000000
<b>*{stop, ... glide}coda</b>	-15.728	-14.812	1.000000
<b>*Dist {+3, ...-3}</b>	-15.728	-17.790	1.000000



De fato, nesta simulação 3, na qual continuamos tendo tratamento diferencial às sequências com plosiva em coda, vamos ter próximos os valores centrais da restrição de fidelidade DEP (50.650) com relação às restrições de marcação, que se referem exclusivamente às sequências com plosiva em coda, fato esse que representa variação no candidato a *output* ótimo.

Não por acaso, a restrição *\*Dist{1} & \*{stop}coda* é a mais próxima de DEP. Isto quer dizer que “t.n” é a nossa sequência mais marcada dentre os candidatos a *output*. Conseqüentemente, tal sequência vai apresentar maiores índices de epêntese com relação à “t.t”, causando variações mais frequentes entre os candidatos a *output* com e sem epêntese.

O conjunto de *tableaux* 27 apresenta uma amostra dessa variação, que pode ser percebida se compararmos os valores das restrições nesta hierarquia da segunda rodada com os valores das restrições dados na 1ª rodada (conjunto de *tableaux* 26 – Simulação 3 OT / Interlíngua).

**Conjunto de *Tableaux* 27- Simulação 3 OT / Interlíngua – Segunda Rodada**

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
<b>*Dist {+3}</b>	100.000	101.327	1.000000
<b>*Dist {+3, +2}</b>	100.000	101.133	1.000000
<b>Max</b>	65.078	63.423	1.000000
<b>*Dist {1} &amp; {stop}coda</b>	47.595	51.028	1.000000
<b>Dep</b>	50.650	49.625	1.000000
<b>*Dist {1,0} &amp; {stop}coda</b>	43.366	46.476	1.000000
<b>*{stop}coda</b>	43.366	42.035	1.000000
<b>*Dist {+3, ...+1}</b>	30.272	29.754	1.000000
<b>*Dist {+3, ... 0}</b>	9.336	10.949	1.000000
<b>*{stop, fric}coda</b>	9.336	7.922	1.000000
<b>*{stop, ... nas}coda</b>	-3.454	0.353	1.000000
<b>*Dist {+3, ...-1}</b>	-3.454	-4.720	1.000000
<b>*{stop, ... liq}coda</b>	-8.748	-8.050	1.000000
<b>*Dist {+3, ...-2}</b>	-8.748	-9.427	1.000000
<b>*{stop, ... glide}coda</b>	-15.728	-15.449	1.000000
<b>*Dist {+3, ...-3}</b>	-15.728	-17.979	1.000000



O conjunto de *tableaux* 27 está nos demonstrando a emergência de “hi.p[i].no.se”, em um modelo linguístico que trabalha com a noção de dominância estrita. No momento em que a restrição  $*\text{Dist}\{1\} \& *{\text{stop}}_{\text{coda}}$ , específica da sequência “t.n” ([p.n, k.n, t.n]) é promovida na hierarquia das restrições, superando DEP no ordenamento, poderá haver a emergência do candidato com epêntese.

Conforme já comentamos, um dos aspectos diferenciais deste algoritmo de aprendizagem diante do modelo *standard* da OT é sua capacidade de trabalhar com *outputs* variáveis, dados pela alteração dos pontos de seleção (*disharmony*). Os valores centrais (*ranking values*) estão determinando de quanto a quanto poderá variar o peso da restrição, e o ponto de seleção demarca dentro desta escala em que ponto se encontra a restrição em um dado momento de produção linguística. Sendo assim, estando DEP e  $*\text{Dist}\{1\} \& *{\text{stop}}_{\text{coda}}$ , com valores de ranqueamento próximos, essas podem alterar sua posição no *ranking*, em qualquer momento de avaliação, respeitando as probabilidades de ocorrência de cada restrição, dadas no conjunto de *tableaux* 27.

Assim como DEP foi capaz de alterar de posição com a restrição específica da sequência “t.n”, vemos, no *ranking*, que a restrição de fidelidade DEP também está suscetível diante da sequência “t.t” ([p.t, k.t]). Se é necessária uma diferença de mais de 10 pontos entre as restrições para que haja variação, “t.t” também é capaz de emergir, graças à também proximidade de  $*\text{Dist}\{1,0\} \& *{\text{stop}}_{\text{coda}}$  (43.366) com DEP (50.650), de modo a respeitar a probabilidade de variação de 98,96% para o candidato “t.t” ser o vencedor e 1,04% para “t[i]t” emergir, conforme dados de produção dos estudantes colombianos (tabela 26).

Dado que as sequências [s.t], [s.n], “w.t”, “l.t”, “n.t” são produzidas no Português Brasileiro e no Espanhol Colombiano, de modo a emergir sem a inserção de epêntese, é necessário que DEP, frente às restrições que representam estas sequências ( $*{\text{stop}, \dots \text{fric}}_{\text{coda}}$ ,  $*{\text{stop}, \dots \text{nas}}_{\text{coda}}$ ,  $*{\text{stop}, \dots \text{liq}}_{\text{coda}}$ ,  $*{\text{stop}, \dots \text{glide}}_{\text{coda}}$ ,  $*\text{Dist}\{+3, \dots +1\}$ ,  $*\text{Dist}\{+3, \dots 0\}$ ,  $*\text{Dist}\{+3, \dots -1\}$ ,  $*\text{Dist}\{+3, \dots -2\}$ ,  $*\text{Dist}\{+3, \dots -3\}$ ), apresente ranqueamento superior e com um distanciamento de mais de 10 pontos. De fato, este distanciamento deve ser maior entre as restrições de fidelidade e as restrições de caráter mais geral, que, por estringência, devem ser demovidas toda vez que tivermos o algoritmo submetido a segmentos mais marcados em coda. Por exemplo, a presença de uma líquida em coda (“l.t”) será capaz de demover  $*{\text{stop}, \dots \text{liq}}_{\text{coda}}$  e  $*{\text{stop}, \dots \text{glide}}_{\text{coda}}$ .

Assim, em se tratando de uma gramática em desenvolvimento, com a presença de *outputs* variáveis, o algoritmo de aprendizagem vai convergir na OT Estocástica.

A figura 29 demonstra a distribuição probabilística dos dados de produção dos aprendizes colombianos de PB em Português Brasileiro.

row	1 rowLabel	2 ?
1	tt → p.t/k.t	98882
2	tt → pi.t/ki.t	1118
3	tt → t	0
4	tn → p.n/k.n/t.n	86306
5	tn → pi.n/ki.n/ti.n	13694
6	tn → n	0
7	nt → n.t	100000
8	nt → ni.t	0
9	nt → t	0
10	lt → l.t	100000
11	lt → li.t	0
12	lt → t	0
13	wt → w.t	100000
14	wt → wi.t	0
15	wt → t	0
16	st → s.t	100000
17	st → si.t	0
18	st → t	0
19	sn → s.n	100000
20	sn → si.n	0
21	sn → n	0

Figura 29 - *Output Distributions da Interlândia – OT*

Retomando os valores utilizados para alimentar o sistema nesta simulação (tabela 26) [t.t] = 98,96%, [ti.t] = 1,04%, [t.n] = 86,52%, [ti.n] = 13,48%, e considerando que nem o Espanhol, nem o Português realizam epêntese para sequências heterossilábicas formadas por nasais, líquidas, glides e fricativas, podemos perceber que as probabilidades de ocorrências dadas aos candidatos estão condizentes com as porcentagens com as quais o algoritmo deveria convergir.

b) A aquisição dos *outputs* variáveis (Interlândia) via Gramática Harmônica

Para esta simulação do sistema, da gramática em desenvolvimento, à luz da HG, seguimos a mesma lógica que utilizamos no algoritmo OT-GLA. O algoritmo de aprendizagem da HG foi informado de que o estágio inicial da interlândia é o resultado final

da L1. As diferenças entre os dois algoritmos de aprendizagem nesta simulação 3 para a gramática em desenvolvimento dos aprendizes colombianos de Português Brasileiro, contudo, é identificada a seguir. O conjunto de *tableaux* 28 demonstra a gramática em desenvolvimento dos estudantes, a partir do algoritmo de aprendizagem vinculado à HG.

Conjunto de *Tableaux* 28 - Simulação 3 HG / Interlândia (outputs variáveis)

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
<b>*Dist {+3}</b>	100.000	98.003	1.000000
<b>*Dist {+3, +2}</b>	100.000	94.449	1.000000
<b>Max</b>	91.695	92.463	1.000000
<b>Dep</b>	67.922	71.107	1.000000
<b>*Dist {1,0} &amp; {stop}coda</b>	29.941	31.129	1.000000
<b>*{stop}coda</b>	29.941	29.750	1.000000
<b>*Dist {1} &amp; {stop}coda</b>	3.560	7.220	1.000000
<b>*Dist {+3, ... 0}</b>	-12.652	-11.276	1.000000
<b>*{stop, fric}coda</b>	-12.652	-17.022	1.000000
<b>*Dist {+3, ...+1}</b>	-26.078	-24.064	1.000000
<b>*Dist {+3, ...-1}</b>	-34.604	-33.458	1.000000
<b>*{stop, ... nas}coda</b>	-34.604	-36.958	1.000000
<b>*{stop, ... liq}coda</b>	-47.668	-47.225	1.000000
<b>*Dist {+3, ...-2}</b>	-47.668	-51.571	1.000000
<b>*{stop, ... glide}coda</b>	-59.617	-60.210	1.000000
<b>*Dist {+3, ...-3}</b>	-59.617	-61.978	1.000000



Olhando para os valores de Harmonia de cada candidato no conjunto de *tableaux* 28, temos que os candidatos ótimos são “t.n”, “t.t”, “l.t”, “w.t”, [s.t], [s.n], “n.t”. Porém, se observarmos além da indicação do candidato ótimo, mas seu valor de Harmonia com relação aos candidatos perdedores, percebemos que “t.n” [p.n, k.n, t.n] e “t.t” [p.t, k.t] não são categóricos, diferentemente do que ocorre com as demais sequências. Os candidatos ótimos com a plosiva em coda (“t.t” = -60.879 e “t.n” = -68.099) expressam um valor de H tão próximo aos candidatos com epêntese (t[i].t e t[i].n = -71.107) que indicam que, em algum momento de produção, a hierarquia das restrições dada no conjunto de *tableaux* 28 para estas sequências com plosiva em coda pode mudar.

Conforme já vimos, estes valores, que determinam que tais candidatos ótimos devem variar, não são aleatórios, uma vez que foram obtidos a partir da análise acústica que realizamos a partir das produções dos aprendizes de PB nestas sequências, demonstrados na tabela 26. Logo, as restrições responsáveis pelo valor de Harmonia de “t.n” e “t.t” devem se organizar numericamente de forma a responder a esta situação. O conjunto de *tableaux* 29, a seguir, tem o propósito de demonstrar esta variação entre os candidatos com a produção da plosiva em coda e com epêntese.

Conjunto de *Tableaux* 29 - Simulação 3 HG / Interlíngua – Segunda Rodada

	<i>ranking value</i>	<i>disharmony</i>	<i>plasticity</i>
<b>*Dist {+3}</b>	100.000	101.473	1.000000
<b>*Dist {+3, +2}</b>	100.000	99.538	1.000000
<b>Max</b>	91.695	91.336	1.000000
<b>Dep</b>	67.922	66.935	1.000000
<b>*Dist {1,0} &amp; {stop}coda</b>	29.941	32.449	1.000000
<b>*{stop}coda</b>	29.941	31.761	1.000000
<b>*Dist {1} &amp; {stop}coda</b>	3.560	3.068	1.000000
<b>*Dist {+3, ... 0}</b>	-12.652	-10.625	1.000000
<b>*{stop, fric}coda</b>	-12.652	-10.717	1.000000
<b>*Dist {+3, ...+1}</b>	-26.078	-27.142	1.000000
<b>*Dist {+3, ...-1}</b>	-34.604	-33.527	1.000000
<b>*{stop, ... nas}coda</b>	-34.604	-35.988	1.000000
<b>*Dist {+3, ...-2}</b>	-47.668	-47.249	1.000000
<b>*{stop, ... liq}coda</b>	-47.668	-50.064	1.000000
<b>*Dist {+3, ...-3}</b>	-59.617	-56.615	1.000000
<b>*{stop, ... glide}coda</b>	-59.617	-57.577	1.000000



Se compararmos os valores do ponto de seleção (*disharmony*) do conjunto de *tableaux* 29 com os valores obtidos na 1ª rodada desta simulação na HG que trata da interlíngua dos estudantes colombianos (conjunto de *tableaux* 28), veremos que as mudanças numéricas dos pesos das restrições fizeram com que o candidato com epêntese (“t[i].n”) emergisse, atingindo valor de H (-66.935) superior ou mais próximo de zero, com relação ao candidato com a plosiva em coda “t.n” (-67.277). É dessa forma que o algoritmo de aprendizagem mostrou que os candidatos ótimos relacionados à plosiva em coda não são categóricos na HG, em função da proximidade dos valores de H entre os candidatos.

É interessante notar que tais valores de restrições, apresentados tanto no conjunto de *tableaux* 29, quanto no conjunto de *tableaux* 28, à luz da OT-GLA implicaria, de imediato, um candidato ótimo categórico e sem epêntese, já que DEP se encontra alto na hierarquia com relação às demais restrições, indicando a diferença na concepção de gramática para cada um dos modelos teóricos que estamos adotando nesta análise.

Com isso, se compararmos os valores de ranqueamento das restrições nos dois diferentes algoritmos vinculados à OT e HG, teremos, no geral, comportamentos numéricos bem diferentes. A restrição de fidelidade MAX, por exemplo, que em ambos os modelos conquista uma posição alta na hierarquia, superando as demais restrições, na HG terá seu valor explicado pela ação cumulativa que o modelo sugere. Assim, não basta que tal restrição esteja no topo da hierarquia, como na OT, mas, sim, que a restrição em questão assumira valor maior que o valor de H dos demais candidatos que violam as restrições de marcação, para que não possibilite que o candidato com apagamento venha a emergir. Isso explica o fato de termos, na HG, valores de seleção em torno de 91.000, ao passo que, na OT, tenhamos obtido valores de 65.000.

Com relação aos *outputs* ótimos [s.t], [s.n], “n.t”, “w.t”, “l.t”, ao levarmos em conta que as restrições  $\{stop, fric\}_{coda}$ ,  $\{stop, \dots nas\}_{coda}$ ,  $\{stop, \dots liq\}_{coda}$ ,  $\{stop, \dots glide\}_{coda}$  e  $\{Dist\{+3, \dots +1\}$ ,  $\{Dist\{+3, \dots 0\}$ ,  $\{Dist\{+3, \dots -1\}$ ,  $\{Dist\{+3, \dots -2\}$ ,  $\{Dist\{+3, \dots -3\}$ , que estão relacionadas a estas sequências heterossilábicas com consoante diferente de plosiva em coda, apresentam valor negativo, concluímos que DEP poderia expressar qualquer valor positivo como peso, já que o valor negativo da restrição está indicando um valor de H igual a zero para as sequências formadas de consoante diferente de plosiva em coda. Essa situação pode também ser uma estratégia do modelo para dizer que tais restrições de marcação estão tendo uma função redundante, no momento em que se referem a sequências com o mesmo comportamento diante da sua exposição à coda medial de palavras.

Também é importante o fato de, em ambos os modelos, lidarmos com uma hierarquia que vai organizar as restrições em termos de relações de marcação, preservando as relações de estringência. Nesta simulação, que trabalha com os dados de produção dos estudantes colombianos no PB (interlíngua), sempre vamos ter as restrições referentes à sequência “t.n” mais próximas a DEP, o que indica que esta sequência, além de mais marcada, está assumindo maior ocorrência de epêntese (13%), com relação à “t.t” (1,9%) e, conseqüentemente, com relação às demais sequências, que mantêm suas restrições de marcação sempre abaixo na hierarquia para que se mantenham categóricas. Por sua vez, as restrições conjuntas  $*Dist\{1\} \& *\{stop\}_{coda}$  e  $*Dist\{1,0\} \& *\{stop\}_{coda}$ , que caracterizam e diferenciam as sequências heterossilábicas de plosiva em coda das demais sequências, foram fundamentais neste processo, indicando que, mesmo em um modelo que trabalha com a cumulatividade dos dados, restrições conjuntas, que expressam a ação conjunta de duas restrições, são importantes.

Assim, em se tratando da aquisição da língua estrangeira dos aprendizes (Português Brasileiro), o algoritmo conseguiu convergir também na HG para uma simulação que sugere *outputs* categóricos e variáveis, ainda que com valores distintos aos dados na OT. Dessa forma, visualizamos a capacidade do modelo em lidar com essas duas situações em paralelo.

Em suma, ambos os algoritmos convergiram, com as restrições propostas, ao estágio final de gramática considerado nas três simulações realizadas. Porém, o resultado final das gramáticas de cada algoritmo apresenta características próprias ao modo como cada um deles seleciona o *output* ótimo.

A figura 30 vai demonstrar a distribuição probabilística dos dados na HG, num processo de 100.000 avaliações dos candidatos a *output*, para que possamos verificar se a probabilidade de ocorrências condiz com as porcentagens para as quais o algoritmo convergiu.

row	rowLabel	?
1	tt → p.t/k.t	99143
2	tt → pi.t/ki.t	857
3	tt → t	0
4	tn → p.n/k.n/t.n	86452
5	tn → pi.n/ki.n/ti.n	13548
6	tn → n	0
7	nt → n.t	100000
8	nt → ni.t	0
9	nt → t	0
10	lt → l.t	100000
11	lt → li.t	0
12	lt → t	0
13	wt → w.t	100000
14	wt → wi.t	0
15	wt → t	0
16	st → s.t	100000
17	st → si.t	0
18	st → t	0
19	sn → s.n	100000
20	sn → si.n	0
21	sn → n	0

**Figura 30 - Output Distributions da Interlândia – HG**

A figura 30 comprova que a gramática em questão está apresentando os índices percentuais desejados, assim como expressa a acuidade dos valores fornecidos pelo algoritmo nesta simulação 3, da interlândia dos estudantes colombianos na Gramática Harmônica.

### 5.2.3 Considerações finais das análises nos modelos da HG e OT Estocástica

Ao considerarmos que o primeiro estágio de aquisição da L2 é o sistema da L1, as simulações em ambos os algoritmos de aprendizagem precisam expressar, primeiramente, a aquisição da língua materna. Assim, foi necessário verificar se os algoritmos seriam capazes de convergir na L1 dos aprendizes, que caracteriza um caso de produção categórica, de 100% de plosiva em coda (Espanhol Colombiano), e na língua estrangeira, que apresenta 100% de

produção da epêntese vocálica (Português Brasileiro). A partir daí, pudemos verificar o estágio desenvolvimental dos aprendizes colombianos em direção ao alvo (aquisição do PB), com as produções da vogal epentética ocorrendo variavelmente. Por essa razão é que trabalhamos sempre com três diferentes simulações para cada algoritmo:

- Simulação 1 – Aquisição do sistema do Espanhol Colombiano (L1);
- Simulação 2 – Aquisição plena do sistema do Português Brasileiro (L2);
- Simulação 3 – Aquisição incompleta do sistema do Português Brasileiro, de

forma a refletir uma gramática em desenvolvimento (interlíngua).

Os aprendizes colombianos, quando adquirindo Português Brasileiro, produziram epêntese vocálica, mesmo que de forma tímida. Nesse sentido, os dados sugeriram que a emergência do segmento vocálico seguiu uma tendência na direção de estruturas mais marcadas às menos marcadas, ou seja, vimos a emergência da epêntese em número maior nas sequências mais marcadas, representadas pela consoante plosiva em coda e, ainda, em maior número naquela sequência que mantinha contato silábico mais marcado (plosiva + nasal). Com as restrições em stringência propostas, tentamos expressar, ainda, que a aquisição tardia das restrições referentes às sequências marcadas é explicada pelo alto ranqueamento de tais restrições na hierarquia.

Vimos que ambos os algoritmos de aprendizagem convergiram, conseguindo refletir uma gramática para os dados de aquisição, com e sem variação. É importante, ainda, destacar os valores das restrições, que se mostraram diferentes nos algoritmos: na OT Estocástica, vai importar a distância dos valores das restrições; por sua vez, na HG, importarão as distâncias de Harmonia. Uma vez que o foco principal da análise foi analisar o emprego de segmentos plosivos em codas mediais de palavras, passíveis de motivação de epêntese no PB, podemos acrescentar, ainda, que o processo de aquisição da L2, seguindo o algoritmo vinculado à OT, forçou uma queda de DEP com relação à hierarquia das restrições na aquisição da língua materna e a promoção das restrições de marcação, de forma que alterasse a posição desta restrição de fidelidade no *ranking*, para que pudessemos ver emergirem as sequências “t[i].t” e “t[i].n” epentetizadas, diferentemente do algoritmo vinculado à HG, que teve de diminuir em índices numéricos muito menores o peso de DEP para que o valor de H do candidato com epêntese fosse mais negativo do que os valores de Harmonia referentes às sequências heterossilábicas formadas de plosiva em coda.

Neste momento, acreditamos ser possível termos as respostas às demais questões norteadoras elaboradas para esta pesquisa:

Questão 3 – No que diz respeito à análise, à luz da Gramática Harmônica e da Teoria da Otimidade Estocástica, quais restrições serão relevantes para o processo de aquisição dos padrões silábicos do PB como L2, por parte de falantes do Espanhol Colombiano?

Questão 4 – Qual(is) hierarquia(s) de restrição(ões) dá(ão) conta das formas de *output* produzidas pelos aprendizes, seguindo os pressupostos da HG e OT Estocástica?

Quanto às Questões Norteadoras 3 e 4, referentes aos dados obtidos da análise acústica dos padrões silábicos produzidos pelos aprendizes colombianos em Português Brasileiro via Gramática Harmônica, verificamos que foi utilizada a estratégia da epêntese nas produções em Português das palavras formadas por plosivas em coda medial, como em “ra.p[i].to”, apesar de esta estratégia não acontecer na língua materna destes aprendizes. Além disso, verificamos que as consoantes em coda seguidas de nasal foram muito mais alvo de epêntese do que as consoantes em coda seguidas de plosiva. Sendo assim, formalizamos restrições de marcação que dessem conta de contato silábico.

As restrições \*Dist, originadas a partir do mecanismo de alinhamento relacional de Gouskova (2004), em stringência, serviram, assim, para diferenciar e hierarquizar as sequências de consoante em coda, seguida de plosiva ou nasal, de forma que as sequências com nasal fossem mais marcadas. Aliado a isto, pensamos em combinar, com as restrições \*Dist, restrições de marcação que, também de forma stringente, pudessem dar conta das demais possibilidades de codas mediais no PB. Dessa forma, conseguiríamos ver emergirem palavras como: “car.ta” – “l.t”, “cau.da” – “w.t”, “san.ta” – “n.t”, “as.no” – “s.n”, “cas.ta” – “s.t”, que têm tratamento diferente das plosivas em coda no PB e nos *outputs* dos aprendizes colombianos.

Sendo assim, consideramos relevantes todas as possibilidades de restrições referentes a contato silábico (\*Dist{+3} até \*Dist{+3,...-3}) e a modo de articulação (\*{stop}<sub>coda</sub> até \*{stop,...glide}<sub>coda</sub>), na análise final com o conjunto completo de restrições. Além disso, expressamos restrições conjuntas (\*Dist{1} & \*{stop}<sub>coda</sub> e \*Dist{1} & \*{stop}<sub>coda</sub>), para que, mesmo em um modelo que trabalha com a cumulatividade dos pesos das restrições, pudessemos diferenciar ou caracterizar “t.n” e “t.t” em detrimento das demais sequências (l.t, w.t, n.t, s.n, s.t).

Porém, o que percebemos no modelo da Gramática Harmônica foi que, para a definição de uma plosiva em coda variando entre (“t.n”/“t.t”) (produção da plosiva em coda sem epêntese) e (“t[i].n”/“t[i].t”) (produção da plosiva em coda com epêntese), ou seja, “t.n” variando com “t[i].n” e “t.t” variando com “t[i].t”, e as demais sequências com consoante em

coda sendo produzidas de forma categórica, sem a produção de epêntese, tivemos a restrição DEP com seu valor de ranqueamento reduzido com relação ao peso que apresentou na aquisição da língua materna e um pouco maior ao valor obtido na aquisição do PB como LE.

Essa alteração do valor de DEP na Gramática Harmônica – interlíngua fez com que o valor de Harmonia do candidato que viola unicamente tal restrição ficasse próximo do valor de H dos candidatos com plosiva em coda, de forma que tivéssemos *outputs* ótimos variando, mas não tão proximamente, que nos trouxesse, na aquisição do PB, plosivas com epêntese de forma categórica na aquisição do PB.

A observação para as simulações 2 e 3, que representavam, respectivamente, aquisição categórica do PB e os *outputs* variáveis das produções dos aprendizes é a de que as restrições de marcação que representavam as demais sequências com as possibilidades de consoantes em coda no PB ( $\{stop, fric\}_{coda}$ ,  $\{stop, \dots nas\}_{coda}$ ,  $\{stop, \dots liq\}_{coda}$ ,  $\{stop, \dots glide\}_{coda}$  e  $\{Dist\{+3, \dots +1\}$ ,  $\{Dist\{+3, \dots 0\}$ ,  $\{Dist\{+3, \dots -1\}$ ,  $\{Dist\{+3, \dots -2\}$ ,  $\{Dist\{+3, \dots -3\}$ ) apresentaram valor negativo, o que quer dizer que possivelmente não exercem efeitos na análise, o que as tornava dispensáveis mesmo em um modelo tal como a HG, em que o peso de cada uma das restrições se mostra fundamental para a definição do candidato ótimo. Se a resposta para o valor de H das sequências [s.t], [s.n], “w.t”, “l.t” e “n.t” é zero, tais restrições definitivamente não estavam explicando a emergência destas sequências, sendo desnecessárias para demonstrar que tais sequências sem epêntese ou apagamento são os candidatos ótimos na análise. Tais resultados podem ser indicativos, portanto, de que o conjunto de restrições apresentadas na segunda simulação já se mostrava suficiente para a apropriada análise do fenômeno, sem que fossem considerados todos os membros do conjunto de estringências, derivados da escala completa utilizada para a formação de restrições via Alinhamento Harmônico.

Sendo assim, cremos ter conseguido dar conta da aquisição completa da L2 e do processo desenvolvimental dos aprendizes colombianos através do modelo teórico da HG, com restrições de fidelidade DEP e MAX, conforme McCarthy e Prince (1995), e restrições de marcação conjuntas e em estringência, representadas pela oposição ao contato silábico e à consoante plosiva em coda.

Retomemos, agora, a quinta e a sexta Questões Norteadoras.

Questão 5 – O modelo da Gramática Harmônica consegue formalizar com adequação o processo de aquisição de L2? Há vantagens frente à Teoria da Otimidade Estocástica?

Os padrões silábicos produzidos pelos estudantes colombianos em sua língua

materna demonstraram que não há utilização de estratégias como epêntese e apagamento para a produção das consoantes em coda que selecionamos para fazer parte da nossa análise. Em compensação, na aquisição da L2 destes aprendizes (Português Brasileiro), temos comportamento diferenciado entre sequências com plosiva em coda medial com relação às demais possibilidades de consoantes em coda e, ainda, entre plosiva em coda seguida, no *onset* da sílaba seguinte, de nasal ou plosiva. Sendo assim, foi necessária a formalização de restrições que considerassem todas estas situações do PB.

O modelo da Gramática Harmônica conseguiu, com acuidade, formalizar este processo de aquisição da L2, de modo a garantir que as restrições que não estavam relacionadas aos candidatos ótimos (MAX – referente a apagamento; \*Dist {+3}, \*Dist{+3, +2} – proibição às sequências heterossilábicas “t.l”, “n.w”, “t.w”, inexistentes no PB) mantivessem valores altos o suficiente para que nunca fossem possíveis, como candidatos ótimos, aqueles candidatos que violam unicamente cada uma destas restrições. O modelo também conseguiu dar conta do comportamento diferenciado das sequências “t.n” e “t.t” representadas por [p.n, k.n, t.n] e [p.t, k.t], que, no PB, são realizadas com a ocorrência de epêntese em detrimento das demais, inclusive, da também obstruinte [s] em coda.

Para tanto, o algoritmo HG-GLA manteve DEP em uma posição acima das restrições de marcação na hierarquia das restrições, com um peso que permitisse que o valor de H do candidato com epêntese fosse menos negativo do que o dos candidatos que violam unicamente as restrições referentes às sequências formadas de plosiva em posição de coda, para, assim, vermos a emergência de [pi.n/ki.n/ti.n] e [pi.t/ki.t]. Ao mesmo tempo, o valor atribuído à restrição DEP era mais negativo, ou menos harmônico, com relação às restrições referentes às demais sequências de consoantes em coda possíveis no PB, para termos, para essas, candidatos ótimos com a produção da consoante em coda, sem a ocorrência de epêntese ou apagamento. A emergência dessas sequências que não são passíveis de epêntese no PB foi facilitada com a exclusão das restrições referentes à “n.t”, “l.t”, “w.t”, [s.t] e [s.n] da análise, já que essas receberam pesos negativos, sugerindo uma explicação de que formam restrições redundantes ou desnecessárias na análise<sup>20</sup>.

Dessa maneira de o modelo da HG tratar da avaliação para a verificação do candidato a *output* ótimo, não temos um candidato vencedor baseado na dominância estrita (princípio fundamental da OT) - conforme confirmamos no posicionamento mantido por DEP

<sup>20</sup> O fato de uma restrição ter valor negativo não indica que ela seja necessariamente redundante. Restrições desnecessárias podem acabar tendo valor negativo, mas, muitas vezes, restrições ativas também podem exibir este comportamento – pode ser o caso das sequências [s.t] e [s.n].

na hierarquia - mas a cumulatividade dos pesos das restrições, atuando em conjunto na escolha das formas de *output*. Tal algoritmo também parece diferir do associado à OT no momento em que acusa restrições que são redundantes na análise, se mostrando um modelo formalmente mais simples, ou mais econômico, e acusando que, muito mais do que diferenças no módulo EVAL, há diferenças na formalização das restrições, uma vez que, através do recurso *Linear* OT associado ao algoritmo, é possível ter o indicativo de algumas restrições que se mostram dispensáveis na análise, uma vez que as restrições em questão poderão assumir caráter negativo.

Questão 6 – Considerando tanto o plano empírico quanto o teórico, de que forma o presente estudo pode contribuir para o entendimento do processo de aquisição do PB por falantes do Espanhol Colombiano?

Acreditamos que o ineditismo da proposta de trabalharmos com os modelos da HG e OT para aquisição do PB como L2, com foco no comportamento de sequências consonantais em relação ao processo de epêntese, pode ser considerada uma contribuição importante do estudo. Além disso, o levantamento dos dados das produções dos aprendizes colombianos do PB poderá servir como banco de dados a outras pesquisas e, até mesmo, trazer evidências sobre o comportamento dos aprendizes estrangeiros e suas possíveis dificuldades na aquisição do PB.

Com relação à contribuição dada pela presente análise aos estudiosos dos modelos da OT e da HG, destacamos que também uma contribuição foi dada no que se refere à discussão sobre a formalização das restrições de marcação utilizadas para a ocorrência do fenômeno da epêntese na aquisição do PB, como L2.

## 6 CONCLUSÕES

Nosso objetivo de descrever e analisar, à luz da OT Estocástica e da Gramática Harmônica, o emprego de segmentos plosivos em codas mediais, em sequências heterossilábicas, de palavras cognatas do Português Brasileiro e do Espanhol Colombiano, resultou na formulação de seis questões norteadoras, que nos fizeram refletir desde a possibilidade de ocorrência de epêntese nas produções dos aprendizes colombianos em Português Brasileiro e as restrições relevantes para tal processo de aquisição, até a adequação do modelo da HG nesse processo.

Acreditamos que foi possível dar conta das questões, na ordem numérica em que foram propostas, através dos diferentes capítulos que buscaram discutir assuntos que consideramos relevantes tanto para que se pudesse alcançar o objetivo final, quanto para que fossem aprofundadas reflexões teóricas e empíricas de questões que se apresentaram fundamentais no decorrer do trabalho, dentre as quais destacamos: (a) a palatalização de /t/ e /d/ antes da vogal [i], na cidade de Pelotas; (b) a ocorrência do fenômeno da epêntese na cidade de Pelotas, fundamental para que se estabelecesse o *input* recebido pelos aprendizes; (c) o teste de proficiência no Português Brasileiro; (d) a sonorização das plosivas diante de segmentos nasais e, finalmente, (e) a formalização das restrições na HG e OT, com as sequências [pt, kt, pn, kn, tn].

No que diz respeito à primeira questão norteadora, sobre a possibilidade da ocorrência da epêntese nos *outputs* dos aprendizes no Português Brasileiro após consoantes plosivas em codas mediais de palavras em sequências não-tautossilábicas, demonstramos que podem ser produzidas vogais de reparo, mesmo que de forma tímida. Foi através do Capítulo 4, da Discussão e Descrição dos Dados, com a descrição fonética das ocorrências nas produções de cada sequência heterossilábica eleita para este estudo, que pudemos fazer esta constatação, que julgamos compreender um valioso banco de dados, para estudos futuros.

A segunda questão norteadora tratava de questionar os fatores linguísticos que estariam contribuindo para a ocorrência do fenômeno da epêntese por parte dos falantes nativos do Espanhol Colombiano nas produções em PB. A primeira questão foi extremamente relevante para que se chegasse à conclusão de que o contato silábico estava se mostrando significativo na inserção da vogal. Foi após a verificação fonética dos dados que tivemos condições de estabelecer qual fator linguístico estaria influenciando no aparecimento da epêntese nas produções dos nossos aprendizes.

Para o Estudo-Piloto exposto no Capítulo 5, - estudo esse elaborado antes do término da leitura total dos dados -, com as sequências [p.t, k.t, p.n, k.n], parecia que contato silábico e ponto de articulação estariam motivando efeito nos *outputs* com epêntese. Os dados respeitavam uma escala de marcação tal que [k.n] apresentava mais epêntese que [p.n], que, por sua vez, apresentava mais epêntese do que a sequência [k.t], que exibia mais epêntese do que [p.t]. Entretanto, após a leitura total dos dados, verificamos que apenas o Contato Silábico estaria favorecendo a inserção vocálica nas sequências heterossilábicas.

A partir da resposta às duas questões iniciais, tivemos condições de trabalhar com uma teoria que lidasse com a nossa suposição de que o Contato Silábico estaria sendo significativo na inserção da vogal após a plosiva em coda, o que nos possibilitou responder ao terceiro questionamento proposto para esta investigação, questionamento esse que versava sobre as restrições que seriam relevantes para a aquisição dos padrões silábicos do PB como L2. Conforme evidenciamos no desenvolvimento do trabalho, tanto na OT Estocástica, quanto na Gramática Harmônica, foi necessária uma reflexão sobre o tipo de restrições capaz de dar conta das gramáticas das línguas em análise.

Para tanto, trabalhamos com teorias que envolviam cada uma das estruturas que abrangem cada língua, principalmente para que pudéssemos perceber que o estágio de interlíngua dos nossos aprendizes de PB estava mais próximo à língua materna dos estudantes. Consideramos cabíveis as restrições Dist, de Goukova (2004), como base para nossas restrições de marcação à análise final, com as sequências [k.t, p.t, k.n, p.n, t.n], como forma de oposição a dadas produções de estruturas de *outputs*.

Abordamos, na sequência, a quarta Questão Norteadora, que questionava sobre a hierarquia assumida pelas restrições na Gramática Harmônica e na Teoria da Otimidade Estocástica nas formas de *output* produzidas. A hierarquia das restrições utilizadas nas simulações se prestou a defender tanto as gramáticas categóricas da língua materna, com 0% de epêntese, e da língua estrangeira, com 100% de epêntese, até a gramática de *outputs* variáveis, da interlíngua. É importante ressaltarmos que, neste modelo da HG, os pesos numéricos de todas as restrições contribuem, cumulativamente, para a escolha da forma de saída, diferentemente do que ocorre na OT Estocástica.

Verificamos que, na L1 dos aprendizes, as restrições de fidelidade DEP e MAX assumem um peso que as mantém afastadas das restrições de marcação, de tal forma que, mesmo se fizessemos avaliações sucessivas, não veríamos variação entre as restrições, ou seja, diferentes momentos de produções linguísticas não modificariam seu ranqueamento. A demção das restrições de marcação em ambos os modelos teóricos – OT Estocástica e HG -

demonstrou que o Espanhol Colombiano sugere que as restrições de fidelidade assumam as posições mais altamente ranqueadas de forma a caracterizar o *ranking* no Espanhol.

Os valores apresentados pela Gramática Harmônica, ainda na simulação que tratava da aquisição da língua materna dos estudantes colombianos, tiveram de ser maiores nas restrições de fidelidade e menores nas restrições de marcação, para que os valores de Harmonia dos candidatos mantivessem a mesma lógica da OT Estocástica, não permitindo variações de ranqueamento, mas resultando apenas em um candidato ótimo, categórico, sem epêntese nem apagamento.

A hierarquia correspondente à aquisição plena da língua estrangeira – Português Brasileiro – já teve suas alterações marcantes para que se pudesse assumir outro candidato ótimo e, também, categórico. Na aquisição plena do PB, assumimos que o candidato ótimo deveria ser aquele com 100% de epêntese e sem variação; logo, DEP teve de ser demovido para que a epêntese pudesse emergir como a estratégia de reparo capaz de adaptar, ao padrão de PB, os *inputs* que resultariam em plosivas em codas mediais das palavras dadas. Logo, o algoritmo de aprendizagem teve de atualizar os pesos das restrições e novos *rankings* se estabeleceram, para que, à luz da HG, os valores de Harmonia do candidato ótimo com epêntese fosse superior aos candidatos perdedores.

A simulação 3, que pretendia demonstrar os *outputs* variáveis das produções dos aprendizes no PB, teve de restabelecer os pesos das restrições, que se reorganizaram em ranqueamento, demonstrando semelhança à hierarquia de restrições de língua materna. Esse fato sugeriu que a interlíngua dos aprendizes se encontra ainda próxima ao sistema da língua materna, no que se refere à aquisição das codas mediais de consoantes plosivas no PB. Destaquemos, no que diz respeito a esta simulação, a diferença fundamental de que tal simulação resultou em valores de Harmonia próximos para o candidato dado como ótimo e o candidato com epêntese, proximidade essa que indicava a probabilidade de variação entre tais candidatos, permitindo a produção variável da epêntese. O algoritmo consegue, assim, retratar os estágios desenvolvimentais por que passaram os aprendizes.

Para finalizar, ao respondermos às últimas Questões Norteadoras – cinco e seis – que, respectivamente, tratavam de verificar se o modelo da HG consegue formalizar com adequação o processo de aquisição de L2 e suas vantagens frente à OT Estocástica, bem como discutir de que forma o presente estudo pode contribuir para o entendimento do processo de aquisição do Português Brasileiro por falantes do Espanhol Colombiano, podemos afirmar que o modelo que trabalha com a cumulatividade das restrições conseguiu convergir em todas as gramáticas a que foi submetido, assim como aconteceu na Teoria da Otimidade Estocástica.

Durante o desenvolvimento do estudo, deixamos claras as diferenças não somente no que diz respeito à formalização da avaliação das formas de saída, mas, também, no que concerne à concepção de gramática que tentamos defender, dada através dos pesos das restrições para que assumam uma hierarquia que estabeleça o sistema gramatical desejado. A diferença fundamental dos modelos, no que diz respeito aos valores assumidos pelas restrições apresentados por cada um dos algoritmos, é que, na HG, as promoções e demissões desses valores ocorrem de maneira bem mais acentuada, de modo que tivemos momentos em que as restrições acabaram por assumir valor negativo.

Logo, o caráter cumulativo que rege o modelo da Gramática Harmônica é que faz com que tenhamos tais alterações acentuadas dos valores das restrições. Vimos, em qualquer das simulações da aquisição plena do PB à luz da HG, que era necessário, para que os candidatos alcançassem o PB, que DEP fosse bastante demovido, o que garantiria que o valor de Harmonia do candidato ótimo fosse sempre mais próximo de zero do que dos demais candidatos. Com isso, acreditamos que esses valores vão ter implicações não somente na determinação do estágio final da gramática do aprendiz, mas na concepção de aquisição de linguagem expressa por cada um dos dois algoritmos.

Essa reflexão que realizamos acerca das diferenças nos valores numéricos apresentados por cada um dos algoritmos (HG-GLA e OT-GLA), em cada uma das três simulações realizadas (1 – aquisição do sistema do Espanhol com 0% de epêntese; 2 – aquisição do sistema de Português Brasileiro, com 100% de epêntese; 3 – aquisição incompleta do Português Brasileiro, com *outputs* variáveis), serve para demonstrarmos, contudo, as particularidades e as implicações na escolha de uma ou outra proposta teórica.

Podemos dizer ainda que o modelo da HG, diante das restrições de marcação propostas, como  $\text{*}\{\text{stop}, \dots, \text{nas}\}_{\text{coda}}$ ,  $\text{*}\{\text{stop}, \dots, \text{liq}\}_{\text{coda}}$ ,  $\text{*}\{\text{stop}, \dots, \text{glide}\}_{\text{coda}}$  e  $\text{*Dist}\{+3, \dots, -1\}$ ,  $\text{*Dist}\{+3, \dots, -2\}$ ,  $\text{*Dist}\{+3, \dots, -3\}$ , referentes às sequências “w.t”, “l.t” e “n.t”, na última análise (Análise Final – Conjunto Completo de Restrições), demonstrou que, para a emergência de tais sequências, talvez não fossem necessárias estas restrições. Julgamos que, ao fornecer valores negativos às restrições em todas as simulações desta última análise, o modelo poderia estar acusando o caráter desnecessário de tais restrições para explicar o processo que estávamos analisando. Isso fez com que sugeríssemos outra diferença entre HG e OT que não se refere ao EVAL, mas que evidencia o caráter de necessidade ou não das restrições de CON.

Acreditamos que as descobertas empíricas e teóricas do presente estudo se mostram relevantes não só aos investigadores do processo de aquisição fonológica do Português como língua estrangeira, ou para os pesquisadores dos modelos da OT e HG, pois a

investigação proposta fornece insumos teóricos que tendem a facilitar o entendimento das sutis diferenças entre duas línguas semelhantes. Estudos sobre a aquisição de sons em Língua Estrangeira são importantes também porque podem fornecer subsídios para o ensino da pronúncia em LE. A compreensão de como se adquire, se utiliza e se articula a língua estrangeira pode dar ao professor melhores condições de avaliar o ensino e oferecer alternativas para a sua prática em sala de aula.

Feitas essas considerações, concluímos que, para analisarmos tanto os sistemas linguísticos da língua materna dos aprendizes colombianos, quanto da língua estrangeira (o Português Brasileiro), ou mesmo o sistema intermediário entre as duas gramáticas, foi necessária uma escolha fundamentada da teoria a ser utilizada. Demonstramos que o estudo de aquisição da língua estrangeira à luz dos modelos da Gramática Harmônica e da Teoria da Otimidade Estocástica, bem como a avaliação das diferenças entre os dois modelos, permitiu-nos uma explicação formal para os padrões de produção encontrados em nossos dados.

## REFERÊNCIAS

ALVES, U. K. *A Aquisição das Sequências Finais de Obstruintes do Inglês (L2) por Falantes do Sul do Brasil: Análise via Teoria da Otimidade*. Tese de doutorado inédita. Porto Alegre: PUCRS, 2008.

\_\_\_\_\_. *A Epêntese Vocálica na Aquisição das Plosivas Finais do Inglês (L2): Tratamento pela OT Estocástica e Pela Gramática Harmônica*. Trabalho apresentado no II SIS Vogais. Belo Horizonte, 2009.

BISOL, L. A Sílabas e seus Constituintes. In: NEVES, M. H. M. (Org.). *Gramática do Português Falado*, v. 7. Campinas: Ed. Da UNICAMP, 1999.

\_\_\_\_\_. (org). *Introdução a Estudos de Fonologia do Português Brasileiro*. 4. ed. Porto Alegre: EDIPUCRS, 2005.

\_\_\_\_\_; Schwindt, L.C (org). *Teoria da Otimidade: Fonologia*. Campinas, SP: Pontes Editores, 2010.

BOERSMA, P. HAYES, B. Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32, 45-86, 2001.

BOERSMA, P.; PATER, J. Convergence Properties of a Gradual Learning Algorithm for Harmonic Grammar. Amsterdam: University of Amsterdam, UMass Amherst, 2008. (Manuscript).

CLEMENTS, G. N. The Role of the Sonority Cycle in Core Syllabification. In: KINGSTON, J. & BECKMAN, M. (Orgs.). *Papers in Laboratory Phonology I*. Cambridge: CUP, p. 283-333, 1990.

\_\_\_\_\_.; HUME, E. V. The Internal Organization of Speech Sounds. In: GOLDSMITH, J. *The Handbook of Phonological Theory*. London: Blackwell, 1995.

\_\_\_\_\_. *Affricates as Noncontoured Stops*. Praga: Charles University in Prague, 1999.

COLLISCHONN, G. *Análise Prosódica da Sílabas em Português*. Tese de doutorado inédita. Porto Alegre: PUCRS, 1997.

\_\_\_\_\_. *Epêntese Vocálica e Restrições de Acento no Português do Sul do Brasil*. Signum: Estudos Linguísticos, Londrina (p.61-78). Junho, 2004.

\_\_\_\_\_. A Epêntese Vocálica no Português do Sul do Brasil. In BISOL, Leda; BRESCANCINI, C. *Fonologia e Variação. Recortes do Português Brasileiro*. Porto Alegre: EDIPUCRS, 2002, p.205 a 230.

\_\_\_\_\_. A Sílabas em Português. In: BISOL, Leda. (Org.). Introdução a Estudos de Fonologia do Português Brasileiro. 4. ed. Porto Alegre: EDIPUCRS, p. 101-133, 2005a.

Dicionários – eletrônico e convencional – *Dicionário Aurélio Eletrônico (AE) – Século XXI*, versão 3.0 (novembro de 1999), *Minidicionário Luft* (2000) e *Dicionário Houaiss da língua portuguesa* (2001).

Dicionários – eletrônico e convencional – Dicionário de la Lengua Española – Real Academia Española (RAE), vigésima segunda edición, 2001. Disponível em: <<http://www.rae.es>> Acesso em: 10 abr. 2010.

FERREIRA, A. B. de H. Míni Aurélio. O Dicionário da Língua Portuguesa. 6. ed., Curitiba: Positivo, 2004.

FERREIRA, A.P. Pet or Petty? Diferenças entre Palavras CVC e CVCV do Inglês por Aprendizes Brasileiros: uma Análise Acústica. Dissertação (Mestrado em Letras), Programa de Pós-Graduação em Letras. Curitiba: UFP, 2007.

GOUSKOVA, M. Relational Hierarchies in Optimality Theory: The case of Syllable Contact. *Phonology* 21:2, p. 201-250, 2004.

HARRIS, J. W. *Syllable Structure and Stress in Spanish: A Nonlinear Analysis*. Cambridge, Mass: MIT Press, 1983.

JESNEY, K.; TESSIER, A. Re-evaluating Learning Biases in Harmonic Grammar. In: BECKER, M. (Org.). University of Massachusetts Occasional Papers in Linguistics, 36: Papers in Theoretical and Computational Phonology. Amherst: GLSA, 2007. p.1-42.

KENT, R.D.; READ, C. *The Acoustic Analysis of Speech*. San Diego: Singular Publishing Group, 1992. 238 p.

LADEFOGED, P. *Elements of Acoustic Phonetic*. 2. Ed. Chicago: The University of Chicago Press, 1996.

LADO, R. 1957. *Linguistics Across Cultures, Applied Linguistics Language Teachers*. Ann Arbor, MI: University of Michigan Press. 141 p.

LEGENDRE, G.; MIYATA, Y.; SMOLENSKY, P. Can connectionism contribute to syntax? Harmonic Grammar, with an application. In: ZIOLKOWSKI, M.; NOSKE, M. DEATON, K. (eds.). *Proceedings of the 26th Regional Meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 1990.

LUCENA, R. M; ALVES, U. K. Implicações Dialetais (Dialeto Gaúcho vs. Paraibano) na Aquisição de Obstruintes em Coda por Aprendizes de Inglês: Uma análise Variacionista. *Letras de Hoje*, v. 45, p. 35-42, 2010.

\_\_\_\_\_; SORACE, A.; SMOLENSKY, P. The Optimality-Theory – Harmonic Grammar Connection. In: SMOLENSKY, P.; LEGENDRE, G. *The Harmonic Mind*. The MIT Press, 2006.

McCARTHY. A Thematic Guide to Optimality Theory. Cambridge University Press, 2002.

\_\_\_\_\_; PRINCE, A. Faithfulness and Reduplicative Identity. In: BECKMAN, J. et al (eds.). University of Massachusetts Occasional Papers in Linguistics 18. Amherst, MA: GLSA Publications, 1995. [Disponível: ROA 60].

MATZENAUER, C. L. B. Homogeneidade/heterogeneidade na aquisição e na variação fonológica. In: VANDRESEN, Paulino (org.). *Variação e mudança no português falado da região sul*. Pelotas: Educat, 2002, p. 95-114.

MATZENAUER, C.; MIRANDA, A. R. Traços Distintivos. In: BISOL, Leda; SCHWINDT, Luiz Carlos (org). *Teoria da Otimidade: Fonologia*. Campinas: Pontes Editores, 2010, p. 11-55.

MUNRO, M. J; BOHN, O. *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Amsterdam: John Benjamins, 2007.

MURRAY, R.; VENNEMANN, T. Sound Change and Syllable Structure [: problems] in Germanic Phonology. *Language* 59, 1983.

PRINCE, A; SMOLENSKY, P. Optimality Theory: Constraint Interaction in Generative Grammar. Technical Report, Rutgers University and University of Colorado at Boulder, 1993. Revised version published by Blackwell, 2004. [Disponível: ROA-573].

QUINTANILHA, R; ALVES, U. A Epêntese no Português Brasileiro (L2) em Segmentos Obstruintes em Codas Mediais de Palavras, por Falantes Nativos do Espanhol Colombiano (L1): Uma Análise via Gramática Harmônica. Palhoça-S.C: Unisul, 2010, p. 248.

SELKIRK, E. The syllable. HULST, H. V. D., SMITH. *The structure of phonological representations (part II)*. Foris, Dordrecht, p. 337-383, 1982.

SMOLENSKY, P. On the internal structure of the constraint component Con of UG. Handout da palestra apresentada na UCLA, Los Angeles, California. ROA 86. Disponível em: <http://www.roa.rutgers.edu>, 1995.

SMOLENSKY, P.; LEGENDRE, G. *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. Cambridge: MIT, 2006.

## **ANEXOS**

## **ANEXO A – FORMULÁRIO DE CONSENTIMENTO INFORMADO**

### **FORMULÁRIO DE CONSENTIMENTO INFORMADO**

Prezado informante:

Por favor, leia o texto a seguir. Ele apresenta informações importantes a respeito do estudo de que você fará parte. Após isso, assinie o documento, indicando que você entende a natureza desta pesquisa e que você consente participar dela.

#### **DADOS DA PESQUISA:**

Investigador – Roberta Quintanilha Azevedo, Mestranda.

Pós-Graduação em Letras da Universidade Católica de Pelotas – UCPel

Orientadora: Prof<sup>a</sup> Dr<sup>a</sup>. Carmen Lúcia Matzenauer

Co-Orientador: Prof. Dr. Ubiratã Kickhöfel Alves

#### **OBJETIVOS E PROCEDIMENTOS:**

Este estudo tem por objetivo investigar qual seria a tendência de um falante de Espanhol colombiano adquirindo Português, diante de palavras que pertençam ao léxico das duas línguas. A pesquisa visa a contribuir com os estudos referentes à aquisição de língua portuguesa como língua estrangeira.

A pesquisa em questão é requisito parcial para a obtenção do grau de Mestre em Letras, com área de concentração em Linguística Aplicada.

#### **PROCEDIMENTOS**

O informante realizará a tarefa de leitura em voz alta de sequências que contenham o contexto para o aparecimento do fenômeno da epêntese no Português Brasileiro, em dois dias. No primeiro dia, o informante fará a leitura de frases em Espanhol, expostas em 132 slides, divididos em 3 arquivos de 44 frases cada. No segundo dia, as frases serão em Português, também expostas em 132 slides, divididos em 3 arquivos de 44 frases cada. As frases encontram-se em ordem aleatória. O informante deverá ler em voz alta cada uma das frases. Tal tarefa será também gravada em áudio, para ser posteriormente analisada.

#### **OUTRAS INFORMAÇÕES**

A participação no estudo é de caráter voluntário. Todos os participantes têm a liberdade de cancelar a participação a qualquer momento. As gravações em áudio ficarão de posse do pesquisador, e os dados coletados poderão ser utilizados em estudos posteriores.

A identidade de todos os participantes permanecerá confidencial - ao participar da pesquisa, o informante receberá um Código de Identificação, de modo que o nome do participante não seja nunca divulgado.

**DECLARAÇÃO**

Declaro que li e compreendi a informação acima e que consinto participar desta pesquisa.

.....

Nome

.....

Assinatura

Data

**ANEXO B – FICHA DE INFORMAÇÃO DO PARTICIPANTE****FICHA DE INFORMAÇÃO DO PARTICIPANTE**

Por favor, preencha o formulário abaixo. Sua participação neste estudo é muito importante.

Obrigada!

1 – NOME:

2 – DATA DE NASCIMENTO:

3 – CIDADE NATAL:

4 – ESCOLARIDADE

( ) Superior Incompleto ( ) Superior Completo ( ) Pós-Graduação

5 – NOME E CIDADE DA LOCALIZAÇÃO DA INSTITUIÇÃO DE ENSINO SUPERIOR COLOMBIANA DE ORIGEM:

6 – TEMPO TOTAL DE PERMANÊNCIA NA CIDADE DE PELOTAS/BRASIL:

Data de Chegada –

Data de Partida –

7 – VOCÊ FREQUENTOU ALGUM CURSO DE LÍNGUA PORTUGUESA ANTES DA CHEGADA NA CIDADE DE PELOTAS/BRASIL? POR QUANTO TEMPO?

8 – VOCÊ ESTEVE EM OUTRA CIDADE DO BRASIL? QUAL FOI O PERÍODO DE PERMANÊNCIA NESSE LUGAR?

9 – VOCÊ JÁ VISITOU ALGUM PAÍS DE LÍNGUA PORTUGUESA ALÉM DO BRASIL? QUAL(IS)? DE QUANTO TEMPO FOI A VISITA?

10 – QUE OUTRAS LÍNGUAS VOCÊ PODE FALAR ALÉM DO ESPANHOL E DO PORTUGUÊS?