

# VARIATION AND OPTIMALITY THEORY: REGRESSIVE ASSIMILATION IN VIMEU PICARD

Walcir CARDOSO

Concordia University (Montreal, Canada), Centre for the Study of Learning  
and Performance (CSLP)

## RESUMO

*Este artigo apresenta uma avaliação das propostas pelas quais a variação tem sido analisada dentro da Teoria da Optimidade (TO). Através da análise do processo fonológico de assimilação regressiva entre palavras em picardo conclui-se que uma versão estocástica da TO é mais adequada para a investigação de fenómenos variáveis.*

## ABSTRACT

*This article presents an assessment of how variation has been analyzed within the framework of Optimality Theory (OT). Analyzing the phonological process of acrossword regressive assimilation in Picard, we conclude that a stochastic version of OT is better suited for the investigation of variable phenomena.*

## PALAVRAS-CHAVE

*Assimilação Regressiva. Picardo de Vimeu. Variação. Teoria da Optimidade.*

## KEY WORDS

*Optimality Theory. Regressive Assimilation. Variation. Vimeu Picard.*

## Introduction

With the emergence of Optimality Theory (OT – Prince and Smolensky 1993) and the consequent demise of variable rules (Labov 1972, Cedergren 1973, Cedergren and Sankoff 1974, Guy 1975) in favor of constraint interaction for the analysis of variability (e.g.Reynolds

1994, Anttila 1997; see also Fasold 1996 and Bergen 2000 for a critique of variable rules), it has been argued that intra-language variation can be satisfactorily accounted for via constraint interaction (e.g. Reynolds 1994, Anttila 1997, Taler 1997, Cardoso 2001). In a constraint-based approach like OT, variability can be expressed without resorting to a separate grammar for each variant or, in the case of a process with more than two variants, without the postulation of more than one rule for a single phenomenon: the framework allows for variation to be encoded in (and therefore predicted by) a single constraint hierarchy. On the other hand, the theory also allows for the assignment of separate grammars for cases in which variation truly involves different grammars (e.g. different dialects, different registers).

Several approaches have been proposed for the analysis of variation in OT. In this paper, we will discuss and evaluate four of these proposals: (1) Kiparsky's (1993) grammars in competition, (2) Reynolds' (1994) floating constraints, (3) Anttila's (1997) partial grammars, and (4) Boersma's (1998) and Boersma and Hayes' (2001) stochastic OT. To level the playing field and to provide empirical evidence for my claims and arguments, the present investigation focuses on one single phenomenon: Across-Word Regressive Assimilation (AWRA) in Vimeu Picard (spoken in northern France), a phonological process that operates variably depending, among other factors, on the geographical distribution of its speakers. The data come from a database consisting of an oral fieldwork corpus of tape-recorded interviews with Picard speakers, and written documents (e.g. private letters, unpublished stories, etc.).

This study offers variationist OT accounts for AWRA within these four approaches. The general goals are: (1) to provide an overview of how variation has been analyzed within the framework of Optimality Theory, and (2) to assess four different approaches that have been proposed for the analysis of variable phenomena in OT. More specifically, the study addresses the following research question: Which of these approaches is better capable of accounting for sociolinguistically-grounded variation,

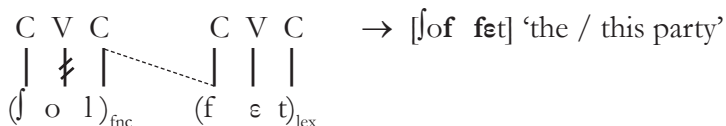
under the traditional variationist assumption that the grammar must include quantitative information, and that the manipulation of frequency is part of a speaker's linguistic competence (e.g. Guy 1975, 1997, Labov 1969, Cedergren and Sankoff 1974)?

This article is composed of four main sections. In section 1, I provide the data that illustrate AWRA, its domain of application, and the variation patterns that characterize the phenomenon. Section 2 presents the data collection procedures and a discussion of the relevant quantitative VARBRUL results for one of the variables considered in the study: the geographic distribution of the participants (originally reported in Cardoso 2001, 2003). In section 3, four different approaches to variationist OT are presented and evaluated vis-à-vis their abilities to satisfactorily account for the variable patterns that characterize AWRA in Picard. It is also shown how a stochastic, GLA-based approach to the analysis of variation is preferable to standard, ordinal OT to analyze the type of variation described in this study: It presupposes a simpler (and more precise) grammar, governed by the same constraints and principles that govern categorical phenomena. Finally, section 4 concludes the study.

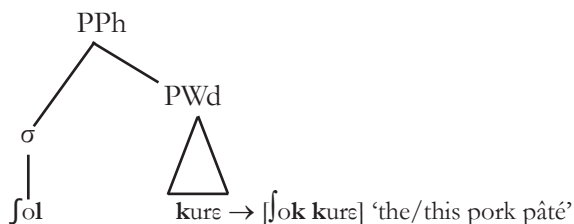
## 1 The data: Across-Word Regressive Assimilation

Across-Word Regressive Assimilation (AWRA henceforth) is a phonological process of Vimeu Picard that operates exclusively at the domain juncture of a (CV)l shape clitic (*inc* in (1)), followed by a consonant-initial lexical word (*lex*) (e.g. (dol)<sub>inc</sub> bibin)<sub>lex</sub> → [dob**b**ibin] ‘some brandy’; (al)<sub>inc</sub> (pek)<sub>lex</sub> → [ap**p**ek] ‘at the fishing’). When both phonological and morphosyntactic contexts are met, the root node of the lexical word's initial consonant associates to the timing slot of the preceding clitic-final /l/, resulting in a geminate across the two words (CV tier is used for illustrative purposes only):

## (1) The AWRA process



As implied in the discussion and representation above, AWRA is a domain-sensitive phenomenon that applies exclusively at the domain juncture of an /l/-final syllable and the following consonant-initial Prosodic Word, within the Phonological Phrase domain (it does not operate within words – e.g. /kalfa/ → \*[kaf.fa], ✓[kal.fa] ‘cauker’ – and across words in higher prosodic domains – e.g. /bel de bel/ → \*[bed de bel], ✓[bel de bel] ‘the very last match’). Following Nespor and Vogel’s (1986) Prosodic Phonology approach to domain-sensitive phenomena, I will refer to this domain juncture as  $\Phi$  (see Cardoso 1999, 2003 for a comprehensive analysis of the prosodic domain of AWRA).

 (2) The Prosodic Domain of AWRA in Picard:  $\Phi$ 


Contrary to what is illustrated in (1) and implied in the representation in (2), AWRA does not apply categorically. In this prosodic context, three distinct patterns can be observed: (a) faithfulness of input /l/ (/l/-preservation); (b) Across-Word Regressive Assimilation (AWRA); and (c) /l/-deletion. For convenience, I will use the acronym AWRA as both a general term for the phenomenon (to which I will occasionally refer as the “AWRA phenomenon”), and as a term for one of its variants, the one illustrated in (3b) below.

## (3) Variants of AWRA

a. /l/-preservation				
/ʃol kure/	→	[ʃol kure]		‘the/this pork pâté’
/dol tart/	→	[dol tart]		‘some pie’
/al kât/	→	[al kât]		‘she sings’
b. AWRA				
/ʃol kure/	→	[ʃok kure]		
/dol tart/	→	[dot tart]		
/al kât/	→	[ak kât]		
c. /l/-deletion				
/ʃol kure/	→	[ʃo kure]		
/dol tart/	→	[do tart]		
/al kât/	→	[a kât]		

In the following section, I will show that one of the factors that determine the patterns observed above is the geographic distribution of the speakers. These results will then serve to demonstrate how the four approaches proposed in OT can be used for the analysis of sociolinguistically-grounded variation.

## 2 AWRA and geographical distribution: a variationist investigation

In this section, I present a discussion of the data collection procedures that were adopted in order to obtain samples of non-categorical data such as those illustrated in (3) above. To illustrate the variable phenomenon in (3) across four different approaches to variation that have been proposed in OT, the discussion will focus exclusively on the geographic distribution of the speakers (geographic location factor). For a comprehensive discussion of the variationist study, see Cardoso (2001, 2003).

In brief, the study consisted of 2,783 tokens of variants of AWRA collected in the field by Julie Auger for the Picard project during the summers of 1996 and 1997, which were further transcribed by four research assistants. The data collected were stratified among six independent variables and later analyzed by the VARBRUL 2 program (Pintzuk 1988): Three extralinguistic factor groups: (1) level of formality, (2) speaker, and (3) geographic location; and three linguistic factor groups: (1) grammatical status of the *l*-clitic, (2) place of articulation of the following consonant, and (3) manner of articulation of the following consonant. As indicated above, I will only report and discuss the results for the geographic location factor in this paper.

The participants (Speakers 1-8) were eight male adult native speakers of Picard, with an average age of more than 70 years old; they inhabited five villages in the Picardie region in northern France: Feuquières, Fressenneville, Bienfay, Bouillancourt and Nibas. Women and younger speakers were not included in the investigation because the vast majority of native speakers of Picard who still use the language routinely are older men.

In order to collect tokens from a wide range of stylistic levels, the data collection methodology used in this study provides a three-level distinction in a formality hierarchy: (1) informal interview, (2) formal interview, and (3) collection of written documents. (1) The informal interview consisted of tape-recorded conversations between the field worker and the interviewee or between the interviewee and other native speakers of Picard. (2) The formal interview consisted of an audio-recorded translation task (designed for the purpose of this study) in which the participants were asked to orally translate French sentences into Picard.<sup>1</sup> (3) The collection of written documents consisted of the selection of such documents from at least one speaker from each region investigated. These documents were extracted from articles from the Picard magazine *Ch'Lanchron*, and unpublished material (including short stories, articles and a few private letters).

<sup>1</sup> The only way to elicit more formal oral data was through the translation task because Picard, as a dying language, is characterized by monostylism (e.g. Dressler 1972, Dorian 1977, Dressler 1988).

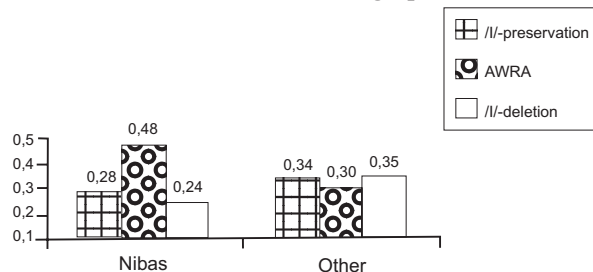
From all the linguistic and extralinguistic factors that were initially included in the investigation, VARBRUL's probabilistic results indicate that the external variables level of formality and geographic location and the internal variable status of the l-clitic have significant conditioning effects on determining the output of the AWRA phenomenon. Focusing solely on the results for geographic location, two general patterns can be observed with regards to the AWRA phenomenon: One in which AWRA is favored (.48) as opposed to /l/-preservation (.28) and /l/-deletion (.24) (2 participants), and another pattern in which all three variants are relatively equally distributed (average around .33) (6 participants). The VARBRUL results (in probabilities) for this factor group are illustrated below (from Cardoso 2001, 2003).

(4) TABLE 1: Probabilities in five geographic locations

	/l/-preservation	AWRA	/l/-deletion
Nibas	.28	.48	.24
Feuquières	.31	.30	.39
Fressenneville	.38	.29	.32
Bienfay	.30	.32	.38
Bouillancourt	.38	.30	.32

For ease of exposition (and because exactly two patterns can be observed), this factor group is regrouped into two major categories: Nibas and Other (which includes all the remaining villages). The results (in probability) are illustrated in Figure 1.

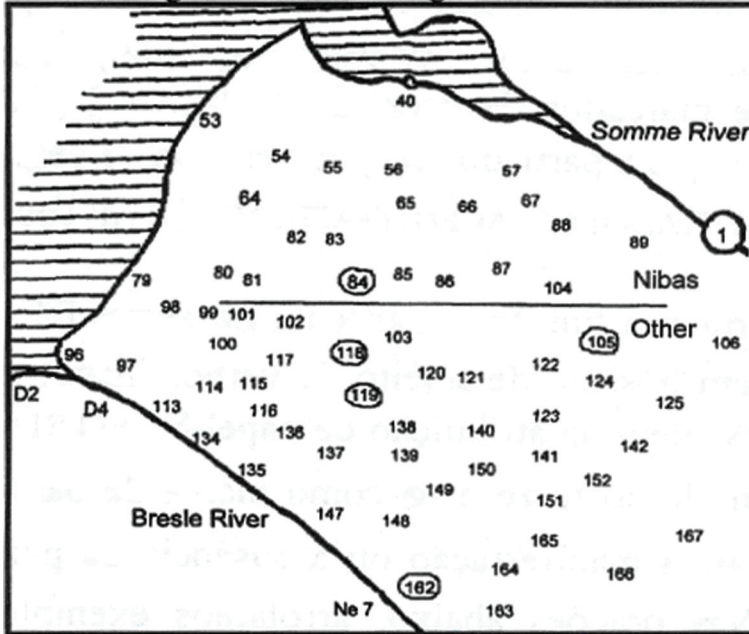
FIGURE 1: AWRA and Geographic Location



Some readers could argue that the variation pattern illustrated in Figure 1 merely demonstrates intraspeaker variation, especially in the context of a limited number of tokens and participants. A brief look at previous studies on regressive assimilation in Picard leads me to conclude that, even though intraspeaker variation is a logical alternative for describing the AWRA phenomenon in the language, more needs to be said about the effect of geographic variation as a determining factor in the results observed above. In the introduction to his *Dictionnaire des parlers picards du Vimeu* and in his grammar *Grammaire des parlers picards du Vimeu (Somme)*, Vasseur (1963, 8 and 1996, 7-8 respectively) refers to the region in which regressive assimilation applies “always and without exception” as the “region of Nibas”. He also acknowledges that assimilation occurs in other regions, but to a lesser extent and sometimes only involving the determiner /*fol*/. Likewise, Debie (1981, 455) observes that regressive assimilation “is concentrated to its maximum intensity in Nibas and in other neighboring villages: 65 [Arrest], 66 [Mons], 86 [Franleu], 87 [Quesnoy] and 122 [Toeuffles] (see Figure 2). It is no coincidence that our results display a relatively similar pattern for the AWRA phenomenon: Of the five villages included in this investigation, it is in Nibas that the AWRA variant is more likely to appear. In Figure 2 below (adapted from Dubois 1957), I show the geographic location of the five villages in the region of Vimeu. The numbers that relate to the villages investigated are circled: 84 = Nibas; 105 = Bienfay (which belongs to the commune of Moyenneville); 118 = Fressenneville; 119 = Feuquières; and 162 = Bouillancourt. The straight line on the map indicates an isogloss-like geographical boundary between two probable dialects: One in which AWRA is highly favored (represented by the region of Nibas – 84), and one in which the three variants are equally likely to appear (represented by the other villages).



FIGURE 2: Five villages in Vimeu



As has been proposed for the analysis of distinct dialectal varieties (e.g. Selkirk 1997, Alber 2001, Boersma 2001), I argue that these two sets of villages define separate dialects, which are formally represented by two grammars: One for the village of Nibas, in which the AWRA variant is favored as opposed to the other two variants, and one for the other villages (Other) in which the three variants are equally predicted. The establishment of these two distinct variable grammars will allow us to analyze and evaluate the variable AWRA phenomenon within four OT approaches for the analysis of variation. This will be the topic of the following section.

### 3 Optimality Theory and variation: AWRA

In this section, I demonstrate how the variable AWRA phenomenon can be analyzed within four approaches that have been proposed for

the analysis of variation in OT: (1) Grammars in competition (section 3.1), (2) Crucial nonranking: Floating constraints (section 3.2), (3) Crucial nonranking: Partial grammars (section 3.3), and (4) Stochastic OT (section 3.4).<sup>2</sup> After an evaluation of these four approaches in the context of the AWRA quantitative results, I will argue that a stochastic version of OT, the one proposed by Boersma (1998) and Boersma and Hayes (2001), is better suited for the analysis of variable phenomena.

To account for the variable aspects of AWRA, I adopt the following well-established OT constraints in (5). I assume that the reader is well versed in OT and knows the basic mechanisms involved in constraint-based analyses.

(5) Constraint definitions

---

<b>FAITH-Lex</b>	The outputs of lexical words are faithful to their inputs.
<b>MAX-IO</b>	Every segment of the input has a correspondent in the output.
<b>Linearity</b>	The input reflects the precedence structure of the output, and vice versa.
<b>NoCoda-Rt</b>	A Coda cannot license a Root node.

---

The constraint FAITH-Lex, which should be interpreted as a cover term for a set of constraints on correspondent elements (e.g. MAX-Lex, DEP-Lex, Linearity-Lex), expresses the cross-linguistic tendency for preservation of information contained in lexical words ather than in function words. It was proposed by Casali (1997) and Pulleyblank (1997) (under ‘Faith-Stem’ – see also Trubetzkoy 1939, Steriade 1995, Casali 1996 and Beckman 1997), although implicit in McCarthy and Prince (1995)

---

<sup>2</sup> A fifth approach for the analysis of variation in OT was proposed by Coetzee (2006): A rank-ordering model of EVAL. The model, however, was left out from the discussion and analyses because it is unable to make precise quantitative predictions: the approach is restricted to predicting the *relative* frequencies of variants. Due to space limitations, I will not provide details of Coetzee’s rank-ordering model; suffice it to say that in the context of the AWRA phenomenon presented here, this model does not differ considerably from Kiparsky’s (1993) grammars in competition approach, which will be discussed in forthcoming section 3.1.

under the Root-Affix Faithfulness Metaconstraint: Root-Faith  $\gg$  Affix-Faith. In the context of AWRA, FAITH-Lex predicts the directionality of AWRA, and thus prevents cases of *progressive* assimilation (e.g. / $[\text{ʃol kure}]$ /  $\rightarrow$  / $[\text{ʃol lure}]$  / ‘the pork pâté’). Due to space limitations, I will not illustrate unattested cases of progressive assimilation in the forthcoming rankings and discussions. It must be assumed, however, that FAITH-Lex is ranked at the higher, undominated end in the constraint hierarchy that characterizes the grammar of Picard.

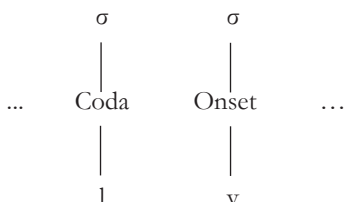
MAX-IO is a constraint that militates against deletion and is violated in cases in which the clitic-final /l/ is deleted from the output; e.g. / $[\text{ʃo kure}]$ / violates MAX-IO because the input /l/ does not have a correspondent in the output.

The Linearity constraint rules out candidates in which the sequence of input segments is reversed or otherwise not obeyed in the surface representation. In cases of regressive assimilation, the precedence relation of  $S_1$  /l-k/ is not reflected in  $S_2$  [k-k]: /l/ precedes /k/ in  $S_1$  but the correspondent of /l/ does not precede the correspondent of /k/ in the output.

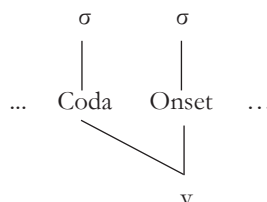
Finally, NoCoda-Rt is member of a family of constraints that captures the crosslinguistic observation on syllabic well-formedness that coda segments are marked. As originally proposed by Prince and Smolensky (1993) (i.e. Syllables do not have Codas), the general version of the constraint is inadequate to account for the range of behavior that coda consonants display cross-linguistically, since languages impose different types of restrictions on codas (in OT, see McCarthy and Prince 1993, Benua 1995, Lombardi 1995, Kawasaki 1998, among others). Observe that NoCoda-Rt is formulated in terms of licensing; consequently, a syllable final consonant can only surface without incurring a violation of this constraint if all of its features are linked to and therefore licensed by a following onset (cf. Piggott 2003). In languages in which NoCoda-Rt is highly ranked, the only codas permitted will be geminates. This is exactly the behavior observed in some dialects of Inuit (e.g. Kalaallisut

and Labrador – Bobaljik 1996). In Picard, NoCoda-Rt rules out forms in which the clitic-final coda /l/ bears and therefore licenses its own Root node, e.g. [ʃol vak] – see (6a). In cases of assimilation, however, NoCoda-Rt is not violated because the assimilated coda’s segmental content (i.e. Root node) is licensed by the onset of the following word, e.g. [ʃov vak] – see (6b). This is shown in the representations below, using standard Onset-Rhyme theory (segments stand for Root nodes).

(6) a. Violation of NoCoda-Rt



b. Satisfaction of NoCoda-Rt



Recall from section 1 that AWRA is a domain-sensitive phenomenon that operates exclusively at the domain juncture  $\Phi$  (see (3) above). As such, we cannot assume that the constraints discussed above have the same weight in the grammar of Picard. For instance, while the constraints MAX-IO, Linearity and NoCoda-Rt can be assumed to be operative at the domain juncture  $\Phi$ , it is clear that this assumption does not hold at other domains, since variable AWRA is unattested in domains lower or higher than  $\Phi$  (see discussion under (1) above). For the analysis of domain-driven phenomena such as AWRA, I adopt Cardoso’s (2003, 2005) domain-specific constraint approach, in which each constraint is decomposable into their domain-specific (e.g. Phonological Phrase – PPh, Prosodic Word – PWd) and locusspecific (i.e. juncture, limit, and span) counterparts, each of which may be ranked independently within a single grammar to yield the alternations observed across domains. Influenced by insights from Prosodic Phonology (Selkirk 1972, 1997, and Nespor and Vogel 1986), the decomposition of constraints within this approach is restricted to those established by this theory (e.g. there

are no direct references to morphosyntactic domains or words). In the forthcoming discussions, the constraints in (5) will appear specified for  $\Phi$  whenever relevant (e.g.  $\text{MAX-IO}_{\Phi}$ ,  $\text{Linearity}_{\Phi}$ ). Constraints that have an Utterance span effect (i.e. those that operate within the entire span of the Utterance domain), on the other hand, will not be specified for a domain (e.g.  $\text{MAX-IO}$ ,  $\text{Linearity}$ ), implying that they operate across all domains of the grammar.

Let us now address how the different approaches proposed for the analysis of variation in OT are able to account for AWRA and its variation patterns.

### 3.1 Grammars in competition

The first account of variation within the OT framework was that of Kiparsky (1993). Within Kiparsky's approach, which follows a stricter view of constraint domination (i.e. a view in which constraints are crucially ranked), variation is seen as a result of competing grammars (or distinct constraint rankings). For instance, in order to account for t/d deletion in English, he assigns a separate constraint ranking for each set of environments favoring the application of the phenomenon. Adapting Kiparsky's approach to the AWRA context, his view would require the assignment of three separate grammars to account for the variation patterns encountered in Nibas and in the other villages (Other):

(7) Kiparsky's approach to variation and AWRA in Other and Nibas

Grammars (constraint rankings)	Output
(a) $\text{MAX-IO}_{\Phi}$ , $\text{Linearity}_{\Phi}$ >> $\text{NoCoda-Rt}_{\Phi}$	/l/-preservation
(b) $\text{MAX-IO}_{\Phi}$ , $\text{NoCoda-Rt}_{\Phi}$ >> $\text{Linearity}_{\Phi}$	AWRA
(c) $\text{Linearity}_{\Phi}$ , $\text{NoCoda-Rt}_{\Phi}$ >> $\text{MAX-IO}_{\Phi}$	/l/-deletion

Kiparsky's approach to variation resembles the cophonology approach (e.g. Itô and Mester 1995ab, Orgun 1996, Inkelas et al. 1997, Inkelas and Zoll 2000), which also appeals to separate grammars to account for different types of variation (e.g. those triggered by different morphological or prosodic constituents, by a class of specific morphemes). Consequently, his approach inherits one of the shortcomings of the use of cophonologies, namely, the proliferation of grammars. Furthermore, Kiparsky's approach to variation is unable to predict the likelihood of occurrence of each variant involved in AWRA. Based on the rankings shown in (7), each variant of AWRA in Nibas is equally likely to appear, which is inconsistent with the results illustrated in Figure 1 in section 2: In this village, AWRA is more likely to apply (.48) than the other two variants (.28 and .24 for /l/-preservation and /l/-deletion respectively).

Assuming the traditional variationist view that the grammar must include quantitative information and that the manipulation of frequency is part of a speaker's linguistic competence (e.g. Guy 1975, 1997, Labov 1969, Cedergren and Sankoff 1974), if the quantitative aspect of a variable grammar is ignored, variation is reduced to random selection, similar to the notion of "free variation". In standard, non-variationist OT, this is commonly claimed to result from the interaction of "freely-ranked" constraints (e.g. Clements 1997, 315), the antithesis of variationist linguistics. This is exactly what Kiparsky's approach implies: That variation such as that found in Nibas is merely the result of the random selection of grammars.

### 3.2 Crucial non-ranking: floating constraints

In an effort to account for variation by assuming the existence of a single grammar that allows the inclusion of both variable outputs and quantitative information, Reynolds (1994) and Anttila (1997), previously published as a manuscript in 1995) pursued an idea hinted at by Prince and Smolensky (1993) in a footnote, about the possibility of crucial nonranking of constraints. In the early stages of OT, it was not evident

why the crucial nonranking of constraints, an essential assumption for the concept that variation can be encoded within a single grammar, should be tolerated in a framework that advocates a *strict dominance hierarchy* (i.e. that each constraint must have absolute priority over all the constraints lower in the hierarchy). In the context of constraint ranking in OT, there could exist a situation in which a constraint set imposes crucial non-dominance (i.e. nonranking) of its components. When a given grammar is unable to categorically yield one of two or more rankings allowed by a set of constraints, the result is the possibility of two or more acceptable forms or outputs in that grammar, i.e. variation per se.

Based on the notion of crucial nonranking, two different proposals have been made in the OT literature: (1) Reynolds' (1994) floating constraint approach (discussed in this section); and (2) Anttila's (1997) partial ranking of constraints approach (discussed in section 3.3). In this section, the focus is on Reynolds' "floating constraints" approach. In Reynolds' view, a variation grammar consists of variably ranked constraints (or *floating* constraints, using the author's terminology). In this approach, the grammar is defined by a single constraint hierarchy, in which one or more constraints may float with respect to another constraint or set of constraints. For example, in a constraint set (call it S), some subset S' may float with respect to some other subset S". Within each subset, constraints may float with respect to each other, as is the case in subset S" below.

(8) Reynolds' floating constraints

$$\{A \gg \boxed{\{\{B\}S' \{C D\}S''\}} \gg E\}S$$

From the number of rankings allowed by a set of variably ranked constraints, distinct outputs can be predicted. For instance, from the variable ranking of S' and S" above, four different rankings and therefore potentially different outputs are expected:

## (9) Rankings allowed by a set of floating constraints

A >>	<b>B</b> >> C >> D	>> E
A >>	<b>B</b> >> D >> C	>> E
A >>	C >> D >> <b>B</b>	>> E
A >>	D >> C >> <b>B</b>	>> E

Anttila (1997) demonstrates that the probability of each variant's occurrence is the result of the number of rankings for which each variant wins, divided by the total number of rankings (or tableaux) generated by the variably ranked constraints. This is formalized in (10):

## (10) Variant probabilistic prediction (Anttila 1997)

- (a) A candidate is predicted by the grammar *if and only if* it wins in some tableaux.
- (b) If a candidate wins in  $n$  tableaux and  $t$  is the total number of tableaux, then the candidate's probability of occurrence is  $n/t$ .

To illustrate, suppose that in a given grammar, GRAM, two constraints B and C float with respect to each other (11a). This is indicated by the semi-colon (to distinguish crucial nonranking from cases of indeterminate ranking, indicated by a comma) between the two constraints involved, with the curly brackets delimiting the set of floating constraints. As a result, two different rankings are possible as illustrated in (11b):

## (11) A variably ranked grammar

- (a) Constraint ranking: A >> {B; C} >> D
- (b) Ranking possibilities: A >> **B** >> **C** >> D  
A >> **C** >> **B** >> D



Imagine that two optimal forms are possible in GRAM, i.e. Cand1 and Cand2. Cand1 is selected when B is ranked higher than C, while Cand2 is selected in the reverse situation. This is illustrated in the two tableaux in (12).

(12) Tableau 1: Variation in GRAM

Tableau (a) = A >> B >> C >> D

Tableau (b) = A >> C >> B >> D

	A	B	C	D			A	C	B	D
☞ Cand <sub>1</sub>			*				Cand <sub>1</sub>		*!	
Cand <sub>2</sub>		*!					☞ Cand <sub>2</sub>		*	

Following Anttila's (1997) variant probabilistic prediction, the variable ranking of constraints B and C results in a pattern in which two outputs are possible, and the probability of each output occurrence can be predicted by (10). For example, candidates 1 and 2 in (12) win in exactly one tableau each ( $n=1$ ), and two is the total number of tableaux ( $t=2$ ).  $n/t = 1/2 = 0.5$  or 50%. Each candidate's probability of occurrence is thus 0.5 and each variant is likely to occur 50% of the time in the same grammar.

The constraint ranking in (11a) and the tableaux in (12) emphasize a crucial distinction in the context of variation in OT, i.e. the distinction between grammars and tableaux. While in (11a) one ranking or grammar yields two tableaux and consequently two outputs (see Tableau 1 in (12)), a categorical grammar yields only one tableau and consequently only one output (i.e. no variation).

Reynolds' approach to variation can be straightforwardly applied to the investigation of AWRA.<sup>3</sup> For instance, to account for the variation patterns observed in the village of Nibas (where AWRA is more likely to apply (.48) than /l/-preservation or /l/- deletion – .28 and .24 respectively), two subsets of domain-specific floating constraints would be required:

<sup>3</sup> In fact, Reynolds' floating constraints approach was adopted in the analysis of variation in AWRA in an earlier version of this investigation, as reported in Cardoso (2001).

## (13) Floating constraints and AWRA

 Nibas:  $\{ \{ \text{MAX-IO}_\phi; \text{NoCoda-Rt}_\phi \}; \text{Linearity}_\phi \}$ 

 Other:  $\{ \text{MAX-IO}_\phi; \text{NoCoda-Rt}_\phi; \text{Linearity}_\phi \}$ 

Consider the village of Nibas, where the hierarchy in (13) yields four rankings, two of which select AWRA as the optimal candidate, while the other two rankings select either /l/- preservation or /l/-deletion as the output. The application of Anttila's (1997) variant probability prediction ( $n/t$ ) in (10) yields probabilistic results that tightly match the ones observed in the village of Nibas. The same applies to the other villages (Other) in (15).

(14) TABLE 2: Floating constraints – output selection for Nibas

Possible Rankings	Output	Observed	Predicted
$\text{MAX-IO}_\phi \gg \text{NoCoda-Rt}_\phi$ $\gg \text{Linearity}_\phi$	AWRA		
$\text{NoCoda-Rt}_\phi \gg \text{MAX-IO}_\phi$ $\gg \text{Linearity}_\phi$	AWRA	.48	$2/4 = .50$
$\text{Linearity}_\phi \gg \text{MAX-IO}_\phi \gg$ $\text{NoCoda-Rt}_\phi$	/l/- preservation	.28	$1/4 = .25$
$\text{Linearity}_\phi \gg \text{NoCoda-Rt}_\phi \gg$ $\text{MAX-IO}_\phi$	/l/-deletion	.24	$1/4 = .25$

(15) TABLE 3: Floating constraints – output selection for Other

Possible Rankings	Output	Observed	Predicted
$\text{MAX-IO}_\phi \gg \text{NoCoda-Rt}_\phi$ $\gg \text{Linearity}_\phi$	AWRA	.30	$1/3 = .33$
$\text{Linearity}_\phi \gg \text{MAX-IO}_\phi \gg$ $\text{NoCoda-Rt}_\phi$	/l/- preservation	.34	$1/3 = .33$
$\text{Linearity}_\phi \gg \text{NoCoda-}$ $\text{Rt}_\phi \gg \text{MAX-IO}_\phi$	/l/-deletion	.35	$1/3 = .33$

In sum, from an empirical perspective, Reynolds' approach can satisfactorily account for the AWRA phenomenon in Picard, as indicated above. The approach, however, is flawed from a conceptual perspective. Most importantly, the model is too permissive in the possibilities of rankings allowed within the grammar. For instance, to account for a variation pattern in which four constraints (A, B, C, D) interact to yield two distinct variants (X, Y), several possibilities of rankings (from which I include only five) are possible within Reynolds' approach (assume that X violates A and B, while Y violates C and D) (adapted from Taler 1997). In addition, the frequencies predicted by the variably-ranked hierarchy are always in small integer fractions (e.g. 1/2, 1/3, 2/3). While I have shown that this is exactly what is found in variable AWRA (e.g. in Other, the three variants are predicted 1/3 of the time), other studies have shown that this is not always the case in variationist linguistics (e.g. Cardoso 2007; see also the critique of the /kelyia/ results in Nagy and Reynolds' 1994 analysis of word-final deletion in Faetar, discussed in Guy 1997, 136-8).

(16) TABLE 4: Reynolds' floating constraint approach: a permissive model

Possible Constraint Rankings: <i>several</i>	Predictability ( <i>n/t</i> )
a. { A; B; C; D }	X = .5; Y = .5
b. { { A; B }; { C; D } }	X = .5; Y = .5
c. { { A; C }; { B; D } }	X = .5; Y = .5
d. { { A; B >> C }; D }	X = .5; Y = .5
e. { { B; A >> C }; D }	X = .5; Y = .5

### 3.3 Crucial nonranking: partial grammars

The third approach to variation in OT was proposed by Anttila (1997). Instead of the use of sets of floating constraints, each of which may contain one or more constraints (see (13) and (16b-e)), Anttila's model accounts for variation by means of a more restricted version of

crucial nonranking. In his approach, the only partial rankings allowed are those composed of single constraints. For instance, to account for the variation pattern illustrated in (16), only the crucial nonranking of all of the constraints A, B, C and D (i.e. (16a)) is permitted in an Anttila-like approach: {A; B; C; D}.

To account for the disparity of results observed involving the factor *geographic location* within Anttila's approach, consider the two distinct variable grammars below (from Cardoso 2003), composed of domain-specific constraints (those not specified for a domain should be assumed to be operative over the span of the Utterance domain, as indicated earlier): (1) one grammar for the village of Nibas, in which the crucial nonranking of five constraints yields 120 tableaux, and (2) one grammar for Other, in which the nonranking of three constraints yields 6 tableaux.

(17) Geographic location and AWRA

**a. Nibas Grammar:**

{MAX-IO<sub>Φ</sub>; MAX-IO; NoCoda-Rt<sub>Φ</sub>; NoCoda-Rt; Linearity<sub>Φ</sub>} >>  
 Linearity

**b. Other Grammar:**

{MAX-IO<sub>Φ</sub>; NoCoda-Rt<sub>Φ</sub>; Linearity<sub>Φ</sub>} >> MAX-IO, NoCoda-Rt,  
 Linearity

The application of Anttila's variant probability prediction in (10) yields the results illustrated in Table 5. Observe that under each variant, the left column (under Pred) indicates the predicted probability of each variant's occurrence, calculated by  $n/t$ , and the parenthesized numbers illustrate the number of rankings (or tableaux) in which that candidate is the winner for each subset of villages (i.e. Nibas or Other). The values in the right column (Obs), on the other hand, indicate the actual VARBRUL weight established for each variant (values from Figure 1).

(18) TABLE 5: Predicted &amp; observed probability of variant occurrence by geographic location

Geographic location	Total # of tableaux	/l/- preservation		AWRA		/l/-deletion	
		Pred	Obs	Pred	Obs	Pred	Obs
Nibas	120	.23 (28)	.28	.53 (64)	.48	.23 (28)	.24
Other	6	.33 (2)	.34	.33 (2)	.30	.33 (2)	.35

In sum, the crucial nonranking of the constraints  $\text{MAX-IO}_\phi$ ,  $\text{MAX-IO}$ ,  $\text{NoCoda-Rt}_\phi$ ,  $\text{NoCoda-Rt}$ , and  $\text{Linearity}_\phi$  in Nibas yields a pattern in which the AWRA variant is more often favored (.53) in relation to the other variants (.23 for both /l/-preservation and /l/- deletion). In the grammar assigned for Other, on the other hand, the crucial nonranking of  $\text{MAX-IO}_\phi$ ,  $\text{NoCoda-Rt}_\phi$  and  $\text{Linearity}_\phi$  results in a pattern in which each of the three variants of the AWRA phenomenon is equally expected to surface (probability .33). As shown in (18) under “Obs”, the predictions made here closely correspond to the VARBRUL results obtained.

Comparing the options that are possible in the two approaches that appeal to the crucial nonranking of constraints, Anttila’s is more advantageous for the analysis of variation. Firstly, Anttila’s model is more constrained because it is less permissive on the possibilities of rankings allowed by the grammar. In fact, Anttila’s approach constitutes a subset of Reynolds’, as implied at the outset of this section. Secondly, Reynolds’ model presents problems from a learnability perspective because the range of options that the language learner will entertain when confronted with the numerous ranking possibilities that his model predicts is too vast. In other words, the hypothesis space in Reynolds’ approach is too large in comparison to Anttila’s. A desirable effect of Anttila’s approach in comparison to Reynolds’ is that it reinforces the notion that different rankings produce different results. More importantly, his model determines the shape of a variable grammar – a partial order composed exclusively of unranked constraints.

Even though Anttila's approach constitutes an improvement over Kiparsky's and Reynolds' approaches to variation in OT, his proposal is exceedingly restrictive in the ranking possibilities allowed in the grammar: Only partial grammars are allowed, similar to what was illustrated in (17). It has been shown that this is not always possible without resorting to (sometimes) random constraints for the mere sake of matching probabilities (see Cardoso 2007). In other words, the partial grammar approach encourages the proliferation of constraints (see Guy 1997, 139 for a critical assessment of the consequences if "OT allows unlimited decomposition of its putative universal constraints"). As was the case with Reynolds' floating constraints, another serious shortcoming of this approach to variation is that it predicts that frequencies should be in small integer fractions (e.g.  $2/3$ ,  $1/2$ ,  $1/3$ ). While this is certainly the case in the present study (e.g. each variant is likely to occur  $1/3$  of the time in Other), other studies have shown that this is not always the case (Cardoso 2007; see also Pater and Werle 2001 for a similar view).

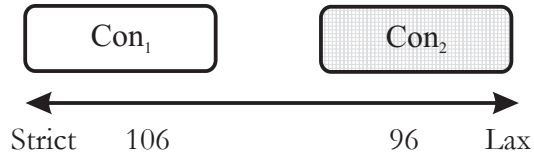
In the following section, I present another approach to variation in OT that addresses the limitations of the three proposals discussed thus far.

### 3.4 Crucial ranking: Stochastic Optimality Theory

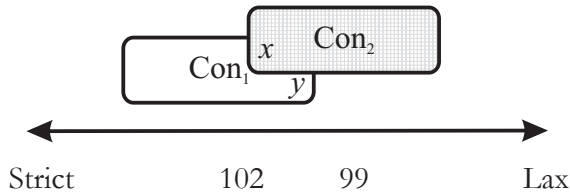
The fourth approach proposed for investigating and representing variability in the framework of Optimality Theory is that of Boersma's (1998) and Boersma and Hayes' (2001): Stochastic OT (StOT). StOT employs an associated learning algorithm: the Gradual Learning Algorithm (GLA). Within the StOT approach, variation and gradient wellformedness are accounted for by a probabilistically determined reranking of constraints at certain intervals during evaluation time (i.e. during the process of speaking). Briefly, StOT postulates a continuous scale of constraint strictness in which constraints (e.g.  $Con_1$  and  $Con_2$  in (19)) are annotated with arbitrary numerical strictness values established by a GLA (e.g. Boersma and Weenink's (2000) Praat program, Hayes et al's (2003) OTSoft software). The probability of reranking (i.e. variation)

is determined by the distance between  $\text{Con}_1$  and  $\text{Con}_2$  on the strictness scale and by the amount of evaluation noise (i.e. standard deviation, typically 2.0) added to the strictness values. A standard deviation of 2.0, the amount of noise used in this study, ensures that the distance of 10 or more values between two constraints will yield categorical rankings, as the hypothetical hierarchy in (19) illustrates. Under StOT, constraints not only dominate other constraints (as is the case in standard OT), but they are also specific distances apart. The two figures in (19) and (20) below illustrate the distinction between a categorical grammar (in which crucially ranked constraints are distant on the strictness scale) and a variable one (in which crucially ranked constraints overlap in their distribution). Note that the superposition of  $\text{Con}_2$  above  $\text{Con}_1$  in (20) is merely to illustrate the boundaries that separate the overlapping constraints, whose distances are computed in the horizontal dimension only.

(19) Categorical ranking



(20) Variable ranki



In the context of a variable ranking, as shown in (20), the grammar might select for evaluation any point within the overlap of  $\text{Con}_1$  and  $\text{Con}_2$ . Most likely, the grammar will select the ranking  $\text{Con}_1 \gg \text{Con}_2$  because of the higher ranking of  $\text{Con}_1$  over  $\text{Con}_2$ . However, it is also possible for the grammar to select a point within the leftmost (higher ranked) area of  $\text{Con}_2$  (i.e.  $x$ ) and the rightmost (lower ranked) area of  $\text{Con}_1$  (i.e.  $y$ ). In this case,  $\text{Con}_2$  is ranked higher than  $\text{Con}_1$  (i.e.  $\text{Con}_2 \gg \text{Con}_1$ ) and a different candidate is selected. In sum, the distance between two or more constraints determines variability (e.g. outputs *a* and *b*) and, more importantly, it encodes predictability of variant occurrence into the grammar (e.g. output *a*'s likelihood of occurrence is 12% while output *b*'s is 88%).

I will now illustrate how StOT is able to account for the geographical distribution of AWRA across the two regions in Vimeu. With the same set of constraints used in the preceding analyses at hand (along with a set of inputs, surface forms and their respective quantitative values established by VARBRUL, incorrect rival candidates, constraints, and constraint violations – just like in a standard OT analysis), a series of computer simulations was performed using the OTSoft 2.1 software package (Hayes et al 2003). In brief, the simulations proceeded as follows: As indicated in section 2, the data set investigated represents two distinct grammars: Nibas and Other. Each of these grammars was individually “learned” by OTSoft, which was supplied with the following information (all numbers are arbitrary) for the learning simulation to take place:

(1) Number of times to go through forms (or total of learning trials): 1,000,000. This number indicates how many learning trials the GLA will perform. The higher the number, the more likely the observed and predicted probabilities will match.

(2) The initial state: 100 for both markedness and faithfulness (the actual value is not essential). By default, the initial state is set with the arbitrary number of 100 for markedness and faithfulness constraints,



a value that ensures that the ranking values will be always positive. This value can be manipulated by the researcher depending on his/her views regarding the initial state in language learning. For instance, s/he might decide that the learning process starts with a grammar in which markedness (e.g. 100) is ranked above faithfulness (e.g. 50) – a standard hypothesis for first language acquisition (e.g. Smolensky 1996, Davidson, Jusczyk, and Smolensky 2004, Hayes 2004).

(3) Initial/final plasticity: 2/.002 respectively, which are the default values in OTSoft 2.1. They serve to adjust the GLA results by comparing the outcome of the learning algorithm with the results entered for each pair of input-output. In Hayes' (2004, 21) own words, "[p]lasticity is the size of the change in the grammar that the GLA makes every time its own guess [does not] match the learning datum it encounters." Note that the algorithm will only make adjustments to the simulated grammar if it detects discrepancies between what is observed and what it predicts – it is *error-driven*.

(4) Number of times to test grammar: 2,000 cycles (default). This number indicates the number of times the GLA will repeat the process of stochastic evaluation and compare the results to the relative frequencies that were observed in the data (the VARBRUL results in our particular case). As will be shown later, the predictions established by the algorithm closely match the frequencies observed in the corpus analyzed.

At the end of the simulations, the algorithm arrived at a final grammar that attempted to mimic the relative frequency of variants in the data, by assigning a ranking value for each of the constraints included in the analysis. Note that there is a degree of randomness in the learning, suggesting that separate GLA simulations will never generate the exact same ranking values and match-up to input frequencies. The procedure described above was repeated for each one of the two grammars established in the investigation.

For the AWRA analysis within StOT, consider the results obtained in Nibas, in which the AWRA variant is more likely to occur (.48) than

l-preservation (.27) and l-deletion (.24). These frequencies in the data were learned by the GLA (OTSoft), which generated the following ranking values for Nibas (note that because the constraints overlap in their distribution – similar to what was illustrated in (20) above – the outcome is variation). The ranking values for these constraints are shown in (21). Note that FAITH-Lex (not illustrated in (21) due to space limitations and its irrelevance to the variable results) is assumed to be ranked at the higher end (approximately 10 values higher than MAX-IO) in the hierarchy of the language. Using standard OT, the hierarchy can be represented as: FAITH-Lex<sup>110</sup> >> MAX-IO<sub>ϕ</sub><sup>100.42</sup> >> NoCodaRt<sub>ϕ</sub><sup>100.24</sup> Linearity<sub>ϕ</sub><sup>99.34</sup> (where the superscripted numbers indicate the ranking value assigned by the GLA – OTSoft).

(21) Nibas: Ranking values

Constraint	Ranking value
MAX-IO <sub>ϕ</sub>	100.42
NoCodaRt <sub>ϕ</sub>	100.24
Linearity <sub>ϕ</sub>	99.34

Because the three equally ranked constraints overlap in their distribution, it is predicted by the values established by the GLA that the set comprised of MAX-IO<sub>ϕ</sub> and NoCoda-Rt<sub>ϕ</sub> will outrank Linearity<sub>ϕ</sub> most of the time (i.e. 48% of the time) since the latter has a lower ranking value; consequently, the outcome will be AWRA (see (22a) for the rankings that select AWRA as the optimal candidate). The GLA also predicts that NoCoda<sub>ϕ</sub> will be over-ranked by the two other constraints 28% of the time and l-preservation will result (22b), while MAX-IO<sub>ϕ</sub> will occasionally rank at the lower end of the hierarchy (24% of the time) whose consequence is l-deletion (22c). In (22), observe that the values established by the GLA simulations for each AWRA variant (under GLA) generate predictions that closely match those observed in the corpus analyzed (under Obs).

(22) TABLE 6: Predicted &amp; observed probability of variant occurrence in Nibas

Rankings	Output Selection			Results	
	[l]	AWRA	Ø	GLA	Obs
a. $\text{MAX-IO}_\Phi \gg \text{NoCoda-Rt}_\Phi$ $\gg \text{Linearity}_\Phi$ $\text{NoCoda-Rt}_\Phi \gg \text{MAX-IO}_\Phi$ $\gg \text{Linearity}_\Phi$		✓		.48	.48
b. $\text{Linearity}_\Phi \gg \text{MAX-IO}_\Phi$ $\gg \text{NoCoda-Rt}_\Phi$ $\text{MAX-IO}_\Phi \gg \text{Linearity}_\Phi$ $\gg \text{NoCoda-Rt}_\Phi$	✓			.28	.27
c. $\text{NoCoda-Rt}_\Phi \gg \text{Linearity}_\Phi$ $\gg \text{MAX-IO}_\Phi$ $\text{Linearity}_\Phi \gg \text{NoCoda-Rt}_\Phi$ $\gg \text{MAX-IO}_\Phi$			✓	.24	.24

The same procedures described above were applied to the results obtained for the other villages. After learning the frequencies entered into the simulation, the algorithm determined the following ranking values for the constraints responsible for the AWRA phenomenon in Other:

(23) Other: Ranking values

Constraint	Ranking value
$\text{Linearity}_\Phi$	100.19
$\text{NoCodaRt}_\Phi$	99.99
$\text{MAX-IO}_\Phi$	99.82

For convenience sake, I will not illustrate or discuss the rankings that select the three AWRA variants in Other. Instead, I summarize in (24) the stochastic analyses for the two grammars that characterize the phenomenon of Across-Word Regressive Assimilation in Picard.

Note that what distinguishes these two grammars from each other is the distance between the same three constraints. For instance, while in Nibas  $\text{Linearity}_{\Phi}$  is ranked relatively lower in the hierarchy, in Other, the three constraints are somewhat equidistant (with  $\text{Linearity}_{\Phi}$  slightly over-ranking all other constraints). Furthermore, observe under each variant in (24) that each grammar learned by the GLA generates output frequencies that are strikingly close or identical to those observed in the data, as was illustrated in Figure 1.

(24) TABLE 7: Predicted & observed probability of variant occurrence by geographic location

Grammars of Vimeu Picard	/l/		AWRA		Ø	
	Pred	Obs	Pred	Obs	Pred	Obs
<b>Nibas:</b> $\text{MAX-IO}_{\Phi}^{100.42}$ $>> \text{NoCodaRt}_{\Phi}^{100.24} >>$ $\text{Linearity}_{\Phi}^{99.34}$	.28	.27	.48	.48	.24	.24
<b>Other:</b> $\text{Linearity}_{\Phi}^{100.19} >>$ $\text{NoCodaRt}_{\Phi}^{99.99} >> \text{MAX-IO}_{\Phi}^{99.82}$	.35	.34	.30	.30	.36	.35

In comparison with the three approaches that have been proposed for the analysis of variable phenomena in OT and, more specifically, in the context of variable AWRA, StOT has proved to be superior: the approach is able to account for the same variable phenomenon (AWRA) via a simpler grammar, with fewer and well-motivated constraints (cf. Reynolds 1994, Anttila 1997, Cardoso 2001, 2003, but see Cardoso 2007 for a similar approach), and predict even more accurate frequency data.<sup>4</sup>

<sup>4</sup> I am aware that StOT has its disadvantages. For instance, the approach is mathematically so powerful that it is possible to match any frequencies with a very small number of reasonable constraints. Whether this is a positive or negative aspect of the approach requires further studies based on both empirical and theoretical grounds.

## 4 Concluding remarks

The study's main goals were: (1) to provide an overview of how variation has been analyzed within the framework of Optimality Theory, and (2) to assess four different approaches that have been proposed for the analysis of variable phenomena in OT. To level the playing field and to provide empirical evidence for my claims and arguments, the overview and the assessment involved the investigation of a single phenomenon: Across-Word Regressive Assimilation in Vimeu Picard, a phonological process that operates variably depending, among other factors, on the geographical distribution of its speakers – while in the village of Nibas the variant AWRA is more likely to occur than l-deletion and l-preservation, in the other villages the three variants are relatively equally likely to occur. While three of the approaches (namely Kiparsky's grammars in competition, Reynolds' floating constraints, and Anttila's partial grammars) were able to adequately capture certain aspects of the phenomenon, they also presented serious limitations that have been minimized or eradicated with the advent of a new way of doing variation in OT: Via stochastic evaluation, in which constraints are not only crucially ranked (as is the case in standard OT), but they occupy specific ranking positions in the hierarchy. Within this approach, variation is the mere result of when two or more constraints overlap in their distribution within a given strictness scale.

For the analysis of variation, this paper argued in favor of a stochastic version of the framework of Optimality Theory: The Gradual Learning Algorithm proposed by Boersma (1998) et seq. and Boersma and Hayes (2001). I have argued in the context of variable data from AWRA that this approach is superior in comparison with its predecessors because it accounts for variation and its frequency effects via the same linguistic constraints and principles that govern categorical phenomena (e.g. the crucial ranking of constraints). The result is the stipulation of simpler and more accurate grammars, with fewer constraints (cf. Reynolds 1994, Anttila 1997, Cardoso 2001, 2003): “[a] grammar with fewer constraints

should in principle be preferred to a grammar with more constraints, providing they make identical predictions” (Asudeh 2001, 9).

## References

ALBER, Birgit. **Regional variation at edges**: Glottal stop epenthesis and dissimilation in Standard and Southern varieties of German. Ms., University of Padova. To appear in *Zeitschrift für Sprachwissenschaft*. 2001.

ANTTILA, Arto. **Deriving variation from grammar**: A study of Finnish genitives. In: HINSKENS, F.; HOUT R. van; WETZELS L. (eds.), **Variation, Change and Phonological Theory**. Amsterdam: John Benjamins, 1997. p. 35-68. (First appeared as a manuscript in 1995, Stanford University).

ASUDEH, Ash. **Linking, optionality, and ambiguity in Marathi**. In: SELLS, Peter (ed.), **Formal and empirical issues in optimality-theoretic syntax**. Stanford, CA: CSLI Publications. 2001. p. 257-312.

BERGEN, Benjamin K. **Probability in phonological generalizations**: Modeling French optional final consonants. Ms., University of California, Berkeley and International Computer Science Institute. 2000.

BENUA, Laura. **Identity effects in morphological truncation**. In: BECKMAN, J.; DICKEY, L. Walsh-; URBANCZYK, S. (eds.), **University of Massachusetts Occasional Papers in Linguistics 18**: Papers in Optimality Theory. Amherst, MA: GLSA, 1995. p. 77-136.

BECKMAN, Jill. **Positional faithfulness, positional neutralization and Shona vowel harmony**. *Phonology* 14, 1997. p. 1-46.

BOBALJIK, Jonathan David. **Assimilation in the Inuit languages and the place of the uvular nasal**. *International Journal of American Linguistics* 62, 1996. p. 323-350.

BOERSMA, Paul. **Functional phonology**. Formalizing the interactions between articulatory and perceptual drives. LOT 11. The Hague, Netherlands: Holland Academic Graphics, 1998.

BOERSMA, Paul. **Review of Arto Anttila: Variation in Finnish phonology and morphology**. *GLOT International* 5, 2001. p. 31-40.

BOERSMA, Paul; HAYES, Bruce. **Empirical tests of the gradual learning algorithm**. *Linguistic Inquiry* 32, 2001. p. 45-86.

BOERSMA, Paul; WEENINK, David. **Praat, a system for doing phonetics by computer**. 2000. Disponível em: <<http://www.praat.org>>

CARDOSO, Walcir. **The domain of across-word regressive assimilation in Picard – an optimality theoretic account**. *Southwest Journal of Linguistics* 17(2), 1999. p. 1-22.

\_\_\_\_\_. **Variation patterns in regressive assimilation in Picard**. *Language Variation and Change* 13, 2001. p. 305-341.

\_\_\_\_\_. **Topics in the phonology of Picard**. PhD thesis, McGill University. McGill Working Papers in Linguistics, 2003.

\_\_\_\_\_. **An integrated approach to variation in Optimality Theory: Evidence from Brazilian Portuguese and Picard**. In: GEERTS; GINNEKEN van; JACOBS (Eds.), **Romance Languages and Linguistic Theory**. Amsterdam and Philadelphia: John Benjamins, 2005. p. 1-15.

\_\_\_\_\_. **The variable development of English word-final stops by Brazilian Portuguese speakers: A stochastic optimality theoretic account**. *Language Variation and Change* 19 (3). 2007.

CASALI, Roderic. **Resolving hiatus**. PhD dissertation, UCLA. 1996.  
\_\_\_\_\_. **Vowel elision in hiatus contexts**: Which vowel goes? *Language* 73, 1997. p. 493-533.

CLEMENTS, George N. **Berber syllabification**: Derivations or constraints? In: ROCA, I. (ed.), **Derivations and Constraints in Phonology**. Oxford: Oxford University Press, 1997. p. 289-330.

CEDERGREN, Henrietta. **The interplay of social and linguistic factors in Panama**. PhD dissertation, Cornell University, 1973.

CEDERGREN, Henrietta; SANKOFF, David. **Variable rules**: Performance as a statistical reflection of competence. *Language* 50, 1974. p. 333-355.

COET'ZEE, Andries W. **Variation as accessing 'non-optimal' candidates**. *Phonology* 23, 2006. p. 337-385.

DAVIDSON, Lisa; JUSCZYK, Peter; SMOLENSKY, Paul. **The Initial and Final States**: Theoretical Implications and Experimental Explorations of Richness of the Base. In: KAGER, R.; PATER, J.; ZONNEVELD, W. (Eds.), **Constraints in phonological acquisition**. Cambridge: Cambridge University Press, 2004. p. 321-368.

DEBRIE, René. **Problèmes posés par la présence de l'assimilation régressive dans le sud-ouest du domaine picard**. *Revue de Linguistique Romane* 45, 1981. p. 422-464.

DORIAN, Nancy C. **The problem of the semi-speaker in language death**. *International Journal of the Sociology of Language* 12, 1977. p. 23-32.



DRESSLER, Wolfgang. **On the phonology of language death.** In: PERANTEAU, P. (ed.), **Papers from the Eighth Regional Meeting of the Chicago Linguistic Society.** Chicago: Chicago Linguistic Society, 1972. p. 448-457.

\_\_\_\_\_. **Language death.** In: NEWMAYER, F. (ed.), **Linguistics: The Cambridge Survey**, v. IV. Cambridge: Cambridge University Press, 1988. p. 184-92.

FASOLD, Ralph. **The quiet demise of variable rules.** In: SINGH, R. (ed.), **Towards a critical sociolinguistics.** Amsterdam: Benjamins, 1996. p. 79-98.

GUY, Gregory. **Use and application of the Cedergren-Sankoff variable rule program.** In: FASOLD, R.; SHUY, R. (eds.), **Analyzing Variation in Language.** Papers from the Second Colloquium on New Ways of Analyzing Variation. Washington, D.C.: Georgetown University Press, 1975. p. 59-69.

\_\_\_\_\_. **Competence, performance and the generative grammar of variation.** In: HINSKENS, F.; HOUT, R. van; WETZELS, W. (eds.), **Variation, Change and Phonological Theory.** Amsterdam: John Benjamins, 1997. p. 125-143.

HAYES, Bruce. **OTSoft: Constraint Ranking Software – Manual.** University of California, Los Angeles (UCLA), 2004/2007.

HAYES, Bruce; TESAR, Bruce; ZURAW, Kie. **OTSoft 2.1** software package. 2003. Disponível em: <<http://www.linguistics.ucla.edu/people/hayes/otsoft>>

INKELAS, Sharon. **Phonotactic blocking through structural immunity.** Ms., University of California, Berkeley, 1997.

INKELAS, Sharon; ORGUN, Orhan; ZOLL, Cheryl. **The implications of lexical exceptions for the nature of grammar**. In: ROCA, I. (ed.), **Derivations and Constraints in Phonology**. Oxford: Oxford University Press, 1997. p. 393-418.

INKELAS, Sharon; ZOLL, Cheryl. **Reduplication as morphological doubling**. Ms., University of California, Berkeley and Massachusetts Institute of Technology, 2000.

ITÔ, Junko; MESTER, Ralf-Armin. **The core-periphery structure of the lexicon and constraints on reranking**. In: BECKMAN, J.; DICKEY, L. Walsh-; URBANCZYK, S. (eds.), **University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory**. Amherst, MA: GLSA. 1995a. p. 181-209.

ITÔ, Junko; MESTER, Ralf-Armin. **Japanese phonology**. In: GOLDSMITH, J. (ed.), **The Handbook of Phonological Theory**. Oxford: Blackwell, 1995b. p. 817-838.

KAWASAKI, Takako. **Coda constraints: Optimizing representations**. PhD dissertation, McGill University, 1998.

KIPARSKY, Paul. **Variable rules**. Paper presented at the Rutgers Optimality Workshop, New Brunswick, N.J. 1993.

KIRCHNER, Robert. **Contrastiveness is an epiphenomenon of constraint ranking**. Proceedings of the Berkeley Linguistics Society 21, 1995. 198-208. [ROA-51].

KROCH, Anthony. **Morphosyntactic variation**. In: BEALS, K. et al. (eds.), **Papers from the 30th Regional Meeting of the Chicago Linguistic Society** v. 2: The Parasession on Variation in Linguistic Theory. 1994. p. 180-201.

LABOV, William. **Contraction, deletion, and inherent variability of the English copula.** *Language* 45, 1969. p. 715-762.

\_\_\_\_\_. **Sociolinguistic patterns.** Philadelphia: University of Pennsylvania Press, 1972.

LOMBARDI, Linda. **Laryngeal neutralization, alignment, and markedness.** In: BECKMAN, J.; WALSH-DICKEY, L.; URBANCZYK, S. (eds.), **University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory.** Amherst, MA: GLSA, 1995. p. 225-248.

MCCARTHY, John; PRINCE, Alan. **Faithfulness and reduplicative identity.** In: BECKMAN, J.; WALSH-DICKEY, L.; URBANCZYK, S. (eds.), **University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory.** Amherst, MA: GLSA, 1995. p. 49-384.

MCCARTHY, John; PRINCE, Alan. **Generalized alignment.** In: BOOIJ, G.; MARLE, J. van (eds.), **Yearbook of Morphology.** Dordrecht: Kluwer Academic, 1993. p. 79-153.

NESPOR, Marina; VOGEL, Irene. **Prosodic phonology.** Dordrecht: Foris, 20, 1986.

ORGUN, Cemil Orhan. **Sign-based morphology and phonology with special attention to Optimality Theory.** PhD dissertation, University of California, Berkeley, 1996.

PATER, Joseph; WERLE, Adam. **Typology and variation in child consonant harmony.** In: FÉRY, C.; GREEN, A. Dubach; VIJVER, R. van de (eds.). **Proceedings of HILP5.** University of Potsdam, 2001.

PIGGOTT, Glyne. The phonotactics of a ‘Prince’ language: a case study. In: **Living on the Edge**: 28 Papers in honour of Jonathan Kaye. Mouton de Gruyter: Berlin, 2003. p. 401-425.

PINTZUK, Susan. **VARBRUL programs [computer program]**. **Philadelphia**: University of Pennsylvania Department of Linguistics, 1988.

PRINCE, Alan; SMOLENSKY, Paul. **Optimality Theory**: Constraint Interaction in Generative Grammar. Oxford/Cambridge: Blackwell Publishers, 1993/2004.

PULLEYBLANK, Douglas. **Optimality Theory and features**. In: ARCHANGELI, D.; LANGEDOEN, D. T. (eds.), **Optimality Theory – An Overview**. Oxford: Blackwell, 1997. p. 59-101.

REYNOLDS, William. **Variation and phonological theory**. PhD thesis, University of Pennsylvania. Selkirk, Elisabeth 1972. The phrase phonology of English and French. PhD thesis, MIT. Distributed in 1981 by the Indiana University Linguistics Club, Bloomington, Indiana, 1994.

SELKIRK, Elisabeth. **Prosodic domains in phonology: Sanskrit revisited**. In: ARONOFF, M.; KEAN, M.-L. (eds.), **Juncture**. Saratoga, California: Anma Libri, 1980. p. 107-129.

\_\_\_\_\_. **The prosodic structure of function words**. In: MORGAN J.; DEMUTH, K. (eds.), **Signal to Syntax**: Bootstrapping from speech to grammar in early acquisition. Mahwah, NJ: Lawrence Erlbaum, 1997. p. 187-213.

SMOLENSKY, Paul. **The initial state and ‘richness of the base’ in Optimality Theory**. Technical Report JHU-CogSci-96-4, Cognitive Science Department, Johns Hopkins University. Rutgers Optimality Archive: #154, 1996.

STERIADE, Donca. **Underspecification and markedness**. In: GOLDSMITH, J. (ed.), **The Handbook of Phonological Theory**. Cambridge, Massachusetts: Blackwell, 1995. p. 114-174.

TALER, Vanessa. **S-Weakening in the Spanish of San Miguel, El Salvador**. MA thesis, McGill University. 1997.

TRUBETZKOY, Nikolai **Grundzuge der phonologie**. Göttingen: Vandenhoeck und Ruprecht. 1939. Appears in English translation as *Principles of Phonology*, translated by C. Baltaxe. Berkeley and Los Angeles: University of California Press, 1969.

VASSEUR, Gaston. **Dictionnaire des parlers picards du Vimeu (Somme)**. Collection de la Société de Linguistique Picarde. Amiens, France: Musée de Picardie, 1963.

\_\_\_\_\_. **Grammaire des parlers picards du Vimeu** – Avec consideration spéciale du dialecte de Nibas. Abbeville, France: Le Conseil Régional de Picardie, F. Paillart, 1996.