22 Variation and Optionality

Arto Anttila
Stanford University

22.1 Preliminaries

The terms 'variation' and 'optionality' in phonology describe a situation where one phonological input has more than one output.[1] Consider the following examples from American English and South-Eastern Finnish. In both cases, variation results from the optional application of one or more phonological processes.

(1) (a) wɛst saɪd → wɛst saɪd ~ wɛs saɪd 'west side'

(b) lentä-i → lénti ~ lénsi ~ lént ~ léns 'fly-PAST'

In (1a), a coronal stop is variably deleted at the edge of a complex coda (*t/d*-deletion) (Labov 1997). In (1b), we have two variable processes working together: /t/ becomes [s] before [i] (Assibilation) and /i/ is deleted in an unstressed syllable (Apocope). This yields four logically possible outcomes, all attested (Laalo 1988).

Before embarking upon the discussion, it will be useful to draw some pre-theoretical distinctions. First, variation may occur *within* an individual (the same individual uses different forms at different times) or *across* individuals (different individuals use different forms). Second, variation may be *free* or contextually *conditioned*. Contextual conditions are usually divided into *internal* and *external* factors. Internal factors may be phonological, such as stress, syllable structure, foot structure, or segment quality, or they may be morphosyntactic, such as part of speech, morpheme type, or morphosyntactic domain. External factors include age, gender, style, register, identity, ethnicity, social class, and target audience. Third, contextual conditioning may be *categorical* or *quantitative*. In the first case, the occurrence of a variant is completely predictable from the context – i.e. we have a 'rule'. In the second case, the occurrence of a variant is not completely predictable, but there is a systematic quantitative pattern – i.e. we have a 'tendency'.

Variation and optionality are pervasive in the phonologies of natural languages and for this reason optional rules have always been part of the generative phonologist's descriptive toolbox (Chomsky and Halle 1968). It is much less clear how to go beyond pure optionality. One view holds that quantitative regularities have no place in the theory

[1] I thank Andries Coetzee, Paul de Lacy, and Kate Ketner for helpful comments. All errors are mine.

To appear in Paul de Lacy (ed.) *The Cambridge Handbook of Phonology*. Cambridge University Press.

of linguistic competence, but belong to the theory of performance (see e.g. Newmeyer 2003) and in practice quantitative aspects of phonological variation have been studied mainly by sociolinguists and phoneticians. There are two observations that are in a deep conflict with this view. First, categorical and quantitative regularities are often conditioned by the same grammatical factors. If the phonological grammar simply delivers the phonologically possible forms, it follows that any quantitative regularities must be explained by external factors, but in reality such regularities often refer to the grammar. Second, phonological variation may involve morphological and lexical conditioning and phonologically conditioned allomorph selection. This means that variation and quantitative regularities are potentially present at every level of phonology and cannot be reduced to 'low-level' phonology or phonetics. We conclude that a satisfactory theory of phonology must be able to provide an explicit theoretical interpretation for variable and quantitative regularities and show how such regularities relate to the more familiar invariant and categorical ones.

The main goal of this chapter is to give a brief overview of three current approaches to phonological variation. All these approaches assume some version of Optimality Theory (Prince and Smolensky 1993/2004, henceforth P&S 1993).[2] For other brief overviews that complement the picture given here, see especially Sankoff (1988) and Pierrehumbert (2003). The examples have been kept simple in the interest of conceptual clarity. For more detailed analyses, including several examples of quantitative modelling of phonological variation in large naturalistic corpora, the reader is referred to the work listed in the bibliography.

22.2 What should a phonological theory of variation explain?

Whenever we encounter a case of phonological variation, the following questions arise:

(2) (a) Why does variation occur in this environment as opposed to others?
(b) What determines the phonological shapes of the variants?
(c) What determines the quantitative preferences among the variants?
(d) What is universal and what is language-particular about this pattern?

Why does variation occur in certain environments, but not in others? English *t/d*-deletion is variable in complex codas (*lost* ~ *los*, *lift* ~ *lif*), but blocked in complex onsets (*train* ~ **rain, star* ~ **sar*). In Standard Finnish, Assibilation is variable in the past tenses of verbs if the preceding syllable is both stressed and bimoraic (*vúo.ti* ~ *vúo.si* 'leak-PAST'), but blocked if the preceding syllable is monomoraic (*vé.ti* / **vé.si* 'pull-PAST') and obligatory if the preceding syllable is trimoraic or unstressed (**káar.ti* / *káar.si* 'veer-PAST', **há.lu.ti* / *há.lu.si* 'want-PAST') (Anttila 2003). In both cases, the locus of variation can be defined in phonological terms. Thus, one would expect phonological theory to explain why variation arises in just these particular environments as opposed to others.

What determines the phonological shapes of the variants? Suppose English *t/d*-deletion occurs in coda clusters because both codas and clusters are universally marked.

[2] For general surveys of Optimality Theory, see de Lacy [ch.1], Kager (1999), and McCarthy (2002).

However, this does not yet explain why the repair is stop deletion. Why do we not have fricative deletion (*lost* → *lot*) or schwa-epenthesis (*lost* → *lost*[ə], *los*[ə]*t*) instead? In the case of Finnish Assibilation, why does /t/ become [s] (*vuoti* ~ *vuosi*) instead of getting deleted (**vuoi*), or geminated (**vuotti*). Again, one would expect phonological theory to explain why the variants take the particular shapes they do.

What determines the quantitative preferences among the variants? English *t/d*-deletion is systematically more common before consonants than vowels (Guy 1980, Labov 1997) and systematically more common after [s] than after [f] (Guy and Boberg 1997). In regional dialects of Finnish, Assibilation and Apocope are quantitatively related: Apocope is systematically more common if Assibilation has applied, e.g. *vuosi* → *vuos* (common) vs. *vuoti* → *vuot* (rare). This pattern holds true across dialects, although the absolute frequencies vary greatly. Such systematic quantitative asymmetries call for a phonological explanation.

Finally, some aspects of variation remain invariant across dialects and some are perhaps universal, whereas other aspects vary from dialect to dialect. For example, English *t/d*-deletion is preferred before a consonant (*cost me* ~ *cos me*) and dispreferred before a vowel (*cost again* ~ *cos again*). This quantitative generalization seems to hold true in all dialects for which sufficient data are available. In contrast, there are dialects where *t/d*-deletion is more common before pauses than before vowels, and there are dialects where we find the opposite pattern. A phonological theory of variation should explain why certain aspects of variation are invariant across dialects, but other aspects may vary from dialect to dialect.

22.3 Three theories of variation

22.3.1 Multiple Grammars

The Multiple Grammars Theory (Kroch 1989, Kiparsky 1993, Anttila 2002b) proposes that variation arises from competing invariant grammars within an individual. The simplest argument for multiple grammars is the phenomenon of multilingualism. It seems uncontroversial that we need two separate grammars to account for an individual's competence in two unrelated languages. The situation is less clear when we are faced with two dialects of the same language, variation among styles and registers, and free variation with little or no semiotic value. Do all these cases involve multiple grammars? The answer is not obvious and it is far from clear how we could even begin to answer this question at this level of generality. One way to make progress is to adopt a particular theory of grammar, combine it with the Multiple Grammars Theory, work out the predictions in special cases, and see whether they are empirically supported.

A multiple grammars analysis involves the following questions:

(3) Three questions involved in a multiple grammars analysis
- (a) What are the possible grammars?
- (b) What types of variation can be derived by combining possible grammars?
- (c) How well do the predicted types of variation match the observed types?

As a concrete example, we consider Kiparsky's (1993) analysis of *t/d*-deletion in American English. The analysis is based on Optimality Theory (see Reynolds 1994, Côté 2000, and Coetzee 2004 for alternative analyses). We will adopt the following terminological convention: the term 'grammar' (in regular font) will be used to refer to a total ranking of constraints, and the term 'GRAMMAR' (in small caps) will refer to an individual's collection of grammars. Under the Multiple Grammars Theory, an individual's GRAMMAR may contain several different grammars.

To keep things simple, we only consider three possible contexts: prevocalic (_V), preconsonantal (_C) and prepausal (_##):

(4)

cost me ~ cos me	*t, d* → ∅ / _ C
cost ~ cos	*t, d* → ∅ / _ ##
cost again ~ cos again	*t, d* → ∅ / _ V

Kiparsky proposes that *t/d*-deletion is driven by syllable structure: *t*/*d* is deleted if it ends up being extrasyllabic.[3] The analysis assumes the following five constraints:

(5)

*COMPLEX	No tautosyllabic clusters (P&S 1993§6.2.1)
ONSET	Onsets required (P&S 1993§2.2)
PARSE	Segments belong to syllables. (P&S 1993§6.1)
ALIGN-LEFT-WORD	No resyllabification across word boundaries. (McCarthy and Prince 1993§5)
ALIGN-RIGHT-PHRASE	Phrase-final consonants are also syllable-final. (McCarthy and Prince 1993§3)

The resulting constraint violations are shown below. Brackets ([,]) mark syllable and phrase boundaries. We consider three candidates: (a) the faithful candidate; (b) the deletion candidate; (c) the resyllabification candidate. No rankings are assumed.

(6) *The violation pattern*

		*COM-PLEX	ONSET	ALIGN-L-W	ALIGN-R-P	PARSE
1. /kast əgɛn/	(a) kast][əgɛn	*	*			
	(b) kas]t[əgɛn		*			*
	(c) kas][təgɛn			*		
2. /kast mi/	(a) kast][mi	*				
	(b) kas]t[mi					*
	(c) kas][tmi	*		*		
3. /kast/	(a) kast]]	*				
	(b) kas]]t				*	*

[3] Kiparsky's analysis dates from 1993 and assumes the then current Containment Theory of Faithfulness (P&S 1993): *t/d* is not literally deleted, but simply left unparsed. While Containment Theory is no longer generally accepted, we have retained Kiparsky's original formulation here.

We can now begin to answer the questions in (3). First, what are the possible grammars? The answer: the possible grammars are the 5! = 120 possible ways of ranking the five constraints. Second, what kinds of *t/d*-deletion patterns do these 120 grammars yield? The answer is easy to work out using the OTSoft software package (Hayes, Tesar, and Zuraw 2003). We only get the six distinct patterns shown in (7).

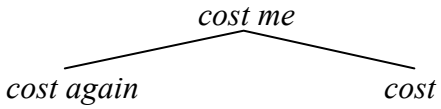
(7) *The factorial typology*

	cost again	*cost*	*cost me*
(a)	faithful	faithful	faithful
(b)	**deletion**	faithful	**deletion**
(c)	**deletion**	**deletion**	**deletion**
(d)	resyllab	faithful	faithful
(e)	resyllab	faithful	**deletion**
(f)	resyllab	**deletion**	**deletion**

Table (7) reveals an important general prediction: if *t/d*-deletion occurs before a pause ((c), (f)) or before a vowel ((b), (c)) it also occurs before a consonant. In contrast, if *t/d*-deletion occurs before a vowel ((b), (c)), it may or may not occur before a pause, and if *t/d*-deletion occurs before a pause ((c), (f)), it may or may not occur before a vowel.

More generally, the constraints induce a partial ordering on the inputs in terms of their deletion potential. We call this the AISSEN ORDERING (Aissen 2003). An Aissen ordering can be represented as a Hasse diagram: if *t/d*-deletion is possible for a lower input (*cost again, cost*), it will be possible for a higher input (*cost me*).

(8) *The Aissen Ordering for t/d-deletion*



In its simplest form, the Multiple Grammars Theory says that any combination of possible grammars is a possible GRAMMAR. While this may sound unrestrictive and raise fears that anything will be possible, recall that the factorial typology limits us to the six possible patterns in (7). Since all the possible grammars conform to the ordering in (8), so will all the possible GRAMMARS. For example, assume a GRAMMAR with three rankings: one that derives the output pattern (a), another that derives (b), and yet another that derives (e). The GRAMMAR {a, b, e} predicts that there can be a speaker with variable deletion before vowels and consonants, but no deletion before pauses. This is consistent with the Aissen ordering.

(9) The GRAMMAR {a, b, e}

	cost again	*cost*	*cost me*
(a)	faithful	faithful	faithful
(b)	**deletion**	faithful	**deletion**
(e)	resyllab	faithful	**deletion**

In contrast, there can be no speaker with variable deletion before vowels, but no deletion before consonants. In such a dialect, *t/d*-deletion would be possible in *cost again*, but not in *cost me*, contradicting the Aissen ordering.

The Multiple Grammars Theory has a straightforward quantitative interpretation:

(10) *A quantitative interpretation of multiple grammars* (Anttila 1997)
(i) A candidate is predicted if it wins by some grammar;
(ii) If a candidate wins by *n* grammars and *t* is the total number of grammars, then the candidate's probability of occurrence is *n/t*.

If we assume the quantitative interpretation in (10), it follows that *t/d*-deletion rate should decrease from top to bottom in the Aissen ordering: any higher input should undergo deletion at least at the same rate as any lower input. For example, the GRAMMAR {a, b, e} predicts the following quantitative pattern:

(11) *GRAMMAR {a, b, e} and the probability of deletion for each input*

	cost again	*cost*	*cost me*
(a)	faithful	faithful	faithful
(b)	**deletion**	faithful	**deletion**
(e)	resyllab	faithful	**deletion**
	1/3	0	2/3

Different deletion probabilities can be derived by varying the number of grammars that predict each output pattern. This is possible because there are several distinct rankings that predict the same output pattern: we have 120 grammars, but only 6 distinct output patterns. Thus, the GRAMMAR {a, b, b, e, e} would predict the deletion probabilities 2/5, 0, and 4/5 for *cost again*, *cost*, and *cost me*, respectively. Crucially, there is no way to construct a grammar that would subvert the Aissen ordering which is thus universal and independent of rankings.[4]

We can now address the last question in (3): how well do the predictions match the observations? The summary in (12) comes from Coetzee (2004:218). The observed quantitative patterns are consistent with the Aissen ordering: in all dialects, *t/d*-deletion is most frequent before consonants; in five dialects, *t/d*-deletion is more common before pauses than vowels; in one dialect, *t/d*-deletion is more common before vowels than pauses.

[4] If we further allow for the possibility that a speaker may have multiple copies of the *same* grammar, we can model arbitrarily fine quantitative distinctions, but crucially, still only within the limits of the Aissen ordering.

(12) *The influence of following context on* t/d-*deletion* (Coetzee 2004: 218)

		_C	_V	_##
Chicano English (Los Angeles)	*n*	3,693	1,574	1,024
(Santa Ana 1991:76, 1996:66)	% deleted	62	45	37
Tejano English (San Antonio)	*n*	1,738	974	564
(Bayley 1995:310)	% deleted	62	25	46
African American English (Washington, DC)	*n*	143	202	37
(Fasold 1972:76)	% deleted	76	29	73
Jamaican mesolect (Kingston)	*n*	1,252	793	252
(Patrick 1991:181)	% deleted	85	63	71
Trinidadian acrolect	*n*	22	43	16
(Kang 1994:157)	% deleted	81	21	31
Neu data	*n*	814	495	–
(Neu 1980:45)	% deleted	36	16	–

The Multiple Grammars Theory is a very simple theory of variation: it only assumes that an individual may possess multiple grammars, something that is independently necessary. It is in no way wedded to Optimality Theory or to any particular quantitative interpretation: it can be easily combined with different grammatical theories as well as different quantitative interpretations. However, despite its simplicity and versatility, the Multiple Grammars Theory does not trivialize the study of variation: it makes falsifiable empirical predictions when combined with particular theories and analyses. In this section, we have used Optimality Theory and the Multiple Grammars Theory to draw a distinction between two kinds of quantitative variation patterns: those that are independent of constraint rankings and therefore universal (Aissen orderings), and those that depend on constraint rankings and can be expected to vary from language to language.

22.3.2 Partially Ordered Grammars

We now turn to a theory that derives variation from a single grammar: the theory of Partially Ordered Grammars (Anttila 1997, 2002a, Anttila and Cho 1998; see Reynolds 1994 and Nagy and Reynolds 1997 for a closely related approach). This theory is empirically more restrictive than the Multiple Grammars Theory: we will see that it excludes certain quantitative patterns predicted to be possible under the Multiple Grammars Theory. Based on evidence from Finnish, we will show how the two theories differ and how Partially Ordered Grammars are empirically superior to Multiple Grammars, at least in this particular case.

The empirical data come from a variable process of Vowel Coalescence found in many dialects of Finnish (Paunonen 1995, Anttila in press). Vowel Coalescence applies to sequences of unstressed short vowels, both derived and nonderived, where the second vowel is [+low].

(13) (a) /suome-a/ súomea ~ súomee 'Finnish-PAR(TITIVE)'
(b) /ruotsi-a/ rúotsia ~ rúotsii 'Swedish-PAR(TITIVE)'

The process is variable within an individual. In a corpus of Colloquial Helsinki Finnish (Paunonen 1995), we find several examples of variation within a single noun phrase and even within a single word. The input /usea-mp-i-a/ 'many-COMPARATIVE-PL-PAR' has four logically possible surface variants, all attested.

(14) úseampia ~ úseempia ~ úseampii ~ úseempii

The central phonological generalization is that coalescence is more common in sequences of mid and low vowels, e.g. *ea ~ ee*, than in sequences of high and low vowels, e.g. *ia ~ ii* (Paunonen 1995:110). This asymmetry is found across regional dialects: coalescence in *ea* is much more widespread than coalescence in *ia* (Paunonen 1995:106-114).

(15) Vowel Coalescence patterns in some Finnish dialects (Paunonen 1995:109-111)

/ea/	/ia/	
ea	ia	Literary Finnish
ee	ia	General Häme
ea ~ ee	ia	Töölö (Helsinki), old upper middle class females
ee	ia ~ ii	Western Uusimaa
ea ~ ee	ia ~ ii	Colloquial Helsinki Finnish

To get the analysis off the ground, we posit the optimality-theoretic constraints in (16). The constraint violations are illustrated in (17).

(16) *Constraints*

*EA	Avoid /ea, oa, öä/ hiatus.
*IA	Avoid /ia, ua, yä/ hiatus.
FAITH	No coalescence

(17) The constraint violation pattern

1. /suome-a/ 'Finnish-PAR'	FAITH	*EA	*IA
(a) suomea		*	
(b) suomee	*		
2. /ruotsi-a/ 'Swedish-PAR'	FAITH	*EA	*IA
(c) ruotsia			*
(d) ruotsii	*		

The vowel height asymmetry can be captured by positing the fixed ranking *EA >> *IA that holds across all dialects of Finnish. This ranking produces the desired typological pattern:

(18) The factorial typology

		suome-a	*ruotsi-a*
(a)	FAITH » *EA » *IA	faithful	faithful
(b)	*EA » FAITH » *IA	**coalesce**	faithful
(c)	*EA » *IA » FAITH	**coalesce**	**coalesce**

How about the variable dialects? Instead of adopting Multiple Grammars, we adopt a slightly more relaxed notion of grammar: we define a grammar as a partial order in a set of constraints. A partial order is a binary relation (i.e. a set of ordered pairs) that is irreflexive, asymmetric, and transitive.[5] By this new definition, {*EA » *IA} qualifies as a grammar, so does {*EA » *IA, *EA » FAITH}. We call such generalized optimality-theoretic grammars Partially Ordered Grammars. A classical optimality-theoretic grammar is a Partially Ordered Grammar where all the pairs are ordered, e.g. {*EA » *IA, *EA » FAITH, *FAITH » *IA}. Any Partially Ordered Grammar can be translated into a set of totally ranked grammars; examples are provided in (19). The reverse does not hold, a fact that will become empirically relevant in a moment.

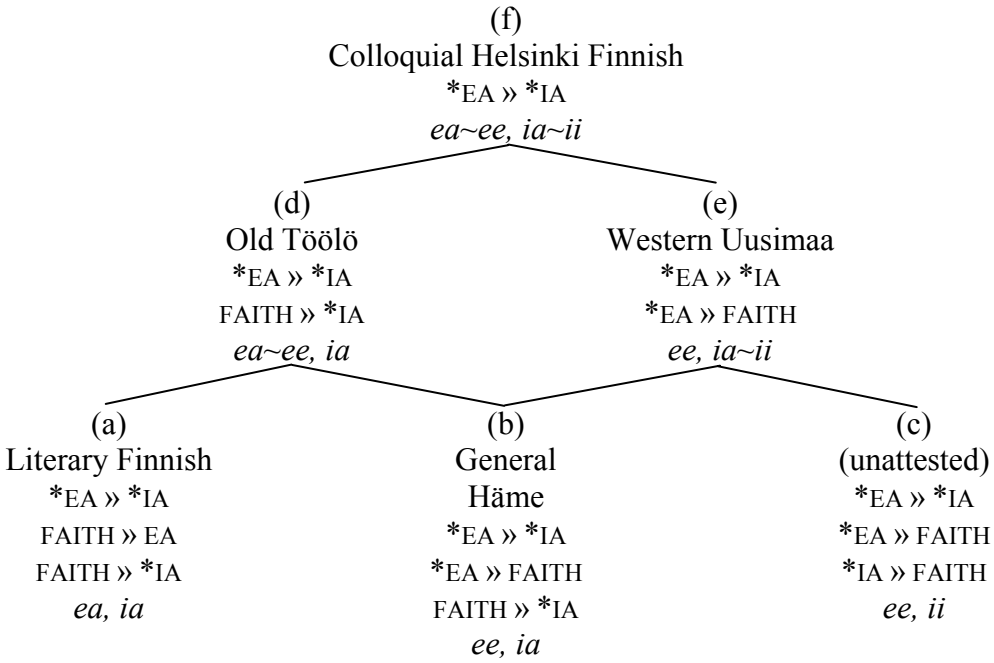
[5] Let *C* be the set of constraints and *R* a binary ranking relation in *C*. *R* is
(i) irreflexive iff for every *x* in *C*, *R* contains no ordered pair <*x, x*> with identical first and second members.
(ii) asymmetric iff for any ordered pair <*x, y*> in *R*, the pair <*y, x*> is not in *R*.
(iii) transitive iff for all ordered pairs <*x, y*> and <*y, z*> in *R*, the pair <*x, z*> is also in *R*.

(19) Partially Ordered Grammars translated into total rankings

	GRAMMAR	TOTAL RANKINGS (TABLEAUX)		
(a)	*EA » *IA	*EA	*IA	FAITH
		*EA	FAITH	*IA
		FAITH	*EA	*IA
(b)	*EA » *IA	*EA	*IA	FAITH
	*EA » FAITH	*EA	FAITH	*IA
(c)	*EA » *IA	*EA	FAITH	*IA
	*EA » FAITH			
	FAITH » *IA			

Since grammars are sets of ordered pairs, a grammar may include other grammars. Diagram (20) spells out the Finnish system in terms of the subset relation. Each node is a grammar and each mother grammar is the intersection of its daughter grammars. We call the resulting structure a 'grammar lattice'. Each grammar is annotated with the predicted output pattern.

(20) A grammar lattice



Grammars (a), (b), and (c) are the totally ranked grammars that describe invariant dialects: (a) is Literary Finnish (*ea, ia*); (b) is General Häme (*ee, ia*); (c) is a hypothetical dialect that shows categorical coalescence everywhere. Grammars (d) and (e) contain two ranked pairs each: (d) is the dialect of old upper middle class female residents of Töölö, a traditional upper-class neighborhood of Helsinki, who allow coalescence in *ea*-sequences, but never in *ia*-sequences (*ea~ee, ia*) (Paunonen 1995:111); (e) is the dialect of Western Uusimaa (*ee, ia~ii*). Grammar (f) contains only one ranked pair. This

grammar translates into three total rankings and predicts variation in both vowel sequences (*ea~ee, ia~ii*), a pattern typical of Colloquial Helsinki Finnish.

If we continue to assume the quantitative interpretation of Multiple Grammars in (10), we derive the predictions in (21). The most important prediction is that *ea* should coalesce at a higher rate than *ia* in all variable dialects. In other words, the analysis derives Paunonen's quantitative generalization.

(21) *A quantitative typology of Vowel Coalescence*

	ea → *ee*	*ia* → *ii*
(a)	0	0
(b)	1	0
(c)	1	1
(d)	1/2	0
(e)	1	1/2
(f)	2/3	1/3

Interestingly, the Multiple Grammars Theory fails to derive Paunonen's generalization. Under this theory, any combination of grammars is a possible GRAMMAR. Now, consider the GRAMMAR {a, c}.

(22) *GRAMMAR {a, c} and the probability of deletion for each input*

		suome-a	*ruotsi-a*
(a)	FAITH » *EA » *IA	faithful	faithful
(c)	*EA » *IA » FAITH	**coalesce**	**coalesce**
		1/2	1/2

This combination of two total rankings predicts a variable dialect where both *ea* and *ia* coalesce at exactly the same rate, contradicting Paunonen's generalization. This dialect is correctly excluded by Partially Ordered Grammars because it is not a partial order: there is no set of ordered pairs that would pick out exactly these two total rankings. The most specific grammar that is a subset of both (a) and (c) is *EA >> *IA, but this grammar also contains (b) which predicts the familiar vowel height asymmetry. The systematic absence of variation patterns like (22) constitutes empirical evidence for Partially Ordered Grammars against Multiple Grammars. The fact that arguments like this are possible in principle underlines the importance of quantitative evidence for phonological theory.

If we delve deeper into the data, more quantitative generalizations emerge. Two of them are morphological:

(23) *Morphological conditions on Vowel Coalescence* (Anttila in press)

(a) THE ROOT FAITHFULNESS EFFECT: Vowel Coalescence is more common across morphemes than within roots.

(b) THE PART-OF-SPEECH EFFECT: Vowel Coalescence is more common in adjectives than in nouns.

The root faithfulness effect reflects the cross-linguistic generalization that roots are more resilient under markedness pressure than affixes (McCarthy and Prince 1995b). In Finnish, this generalization emerges quantitatively in *ea*-sequences and categorically in *ia*-sequences: *ea*-roots coalesce at a lower rate than derived *ea*-sequences (e.g. *hopea* ~ *hopee* 'silver' vs. *suome-a* ~ *suome-e* 'Finnish-PAR') whereas *ia*-roots never coalesce in any dialect (*lattia* / **lattii* 'floor' vs. *ruotsi-a* ~ *ruotsi-i* 'Swedish-PAR'). In order to derive this asymmetry, we adopt the constraint IDENT(Root) which strives to preserve the identity of root segments (McCarthy and Prince 1995b, Beckman 1998, Alderete 2001). The special pattern of *ia*-roots can be captured by the fixed ranking $\text{FAITH}_{\text{root}}$ >> *IA which holds in all dialects of Finnish. The updated typology is given in (25).

(24) *An analysis of the root faithfulness effect*

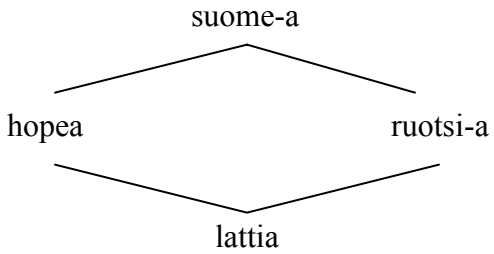
(a)	$\text{FAITH}_{\text{root}}$	Avoid coalescence in roots.
(b)	$\text{FAITH}_{\text{root}}$ » *IA	No coalescence in *ia*-roots.

(25) *The updated typology*

		suome-a	*hopea*	*ruotsi-a*	*lattia*
(a)	FAITH » *EA » $\text{FAITH}_{\text{root}}$ » *IA	faithful	faithful	faithful	faithful
(b)	FAITH » $\text{FAITH}_{\text{root}}$ » *EA » *IA	faithful	faithful	faithful	faithful
(c)	$\text{FAITH}_{\text{root}}$ » FAITH » *EA » *IA	faithful	faithful	faithful	faithful
(d)	$\text{FAITH}_{\text{root}}$ » *EA » FAITH » *IA	**coalesce**	faithful	faithful	faithful
(e)	$\text{FAITH}_{\text{root}}$ » *EA » *IA » FATH	**coalesce**	faithful	**coalesce**	faithful
(f)	*EA » FAITH » $\text{FAITH}_{\text{root}}$ » *IA	**coalesce**	**coalesce**	faithful	faithful
(g)	*EA » $\text{FAITH}_{\text{root}}$ » FAITH » *IA	**coalesce**	**coalesce**	faithful	faithful
(h)	*EA » $\text{FAITH}_{\text{root}}$ » *IA » FAITH	**coalesce**	**coalesce**	**coalesce**	faithful

We now draw the Aissen ordering for Vowel Coalescence based on the typology in (25). The prediction is that a higher input should undergo Vowel Coalescence at least at the same rate as any lower input.

(26) *The Aissen ordering for Vowel Coalescence*



The part-of-speech effect is more problematic. There are at least two kinds of nouns: recently borrowed nouns that never coalesce, e.g. *idea* / **idee* 'idea' and a handful of native nouns that coalesce to some extent, e.g. *hopea* ~ *hopee* 'silver'. This means that adding a constraint like IDENT(Noun) (Smith 1997) will not suffice because it cannot distinguish between the two noun classes. An alternative solution is to assume that recently borrowed nouns and adjectives subscribe to slightly different COPHONOLOGIES

(Orgun 1996, Inkelas 1998, Anttila 2002). The cophonologies needed to capture the part-of-speech effect are given in (27).

(27) *A cophonology analysis of the part-of-speech effect*

(a) FAITH$_{root}$ » *EA (*idea*-nouns only)	No coalescence in *idea*-nouns
(b) *EA » FAITH$_{root}$ (adjectives only)	Coalescence in *ea*-adjectives

We have used Finnish Vowel Coalescence as an example to illustrate two common phenomena: (i) phonological variation involves both phonological and morphological conditions; (ii) both kinds of conditions yield both categorical and quantitative surface patterns. How well does the present analysis succeed in capturing these patterns? Table (28) summarizes the observations and predictions. The numbers are based on Paunonen's corpus of Colloquial Helsinki Finnish that represents the output of 126 individual speakers (Paunonen 1995, Anttila in press). The counts include all stems ending in the vowel sequences *ea*, *eä, ia*, *iä*.

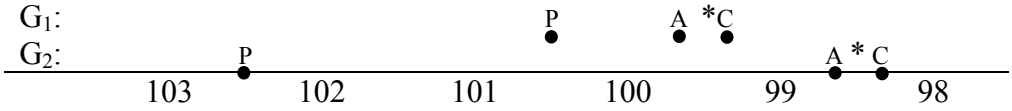
(28) *Observations and predictions*

	OBS%	PRED%	EXAMPLE	GLOSS	*n*
/e-a/, ADJ	--	75.0	--		--
/e-a/, NOUN	41.0	50.0	suome-a	'Finnish-PAR'	714
/ea/, ADJ	72.4	75.0	makea	'sweet'	1,745
/ea/, NOUN (native)	18.8	37.5	hopea	'silver'	48
/ea/, NOUN (recent)	0.0	0.0	idea	'idea'	12
/i-a/, ADJ	30.2	25.0	uus-i-a	'new-PL-PAR'	4,264
/i-a/, NOUN	20.0	25.0	ruotsi-a	'Swedish-PAR'	5,059
/ia/, ADJ	0.0	0.0	kauhia	'terrible'	261
/ia/, NOUN	0.0	0.0	lattia	'floor'	847

22.3.3 Stochastic Optimality Theory

In classical Optimality Theory (P&S 1993), ranking is an ordering relation. For example, the ranking PARSE » ALIGN-LEFT-WORD » *COMPLEX can be described as a set of ordered pairs of constraints. In Stochastic Optimality Theory (Boersma 1998, Boersma and Hayes 2001), we encounter a richer notion of ranking. Each constraint is associated with a fixed real-number value called the RANKING VALUE, marked as a • in the diagrams below. For example, assume two grammars, G_1 and G_2, with the constraints PARSE (= P), ALIGN-LEFT-WORD (= A), and *COMPLEX (*C). The exact real number values used are arbitrary; only the degree of ranking difference is important.

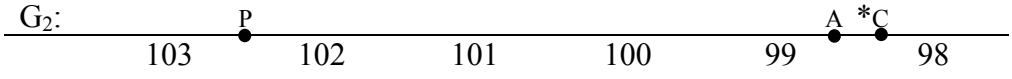
(29) *Two grammars* (G_1, G_2) *with different ranking values*



In both grammars, the constraints appear in the same order. However, in grammar G_1 the constraints PARSE and ALIGN-LEFT-WORD are relatively close to each other, whereas in grammar G_2 they are far apart. In both grammars, the constraints ALIGN-LEFT-WORD and *COMPLEX are ranked about equally close to each other. This notion of a *continuous ranking scale* forms the basis for an alternative approach to variation and quantitative patterns.

The central idea in Stochastic Optimality Theory is *stochastic candidate evaluation*. This means that a random positive or negative value ('noise') is temporarily added to the ranking value of each constraint at the moment of evaluation, i.e. a particular speaking event. The resulting value is called the 'selection point' and it is this selection point that determines the ranking in actual evaluation. Selection points are assumed to be normally distributed around the ranking value and each constraint has the same standard deviation ('breadth'). An example will make this more concrete. Assume ranking values as in Grammar G_2. If we test this grammar a large number of times, the selection points will oscillate around the fixed ranking values from evaluation to evaluation. The result might look like the following:

(30) *Testing grammar G_2*



(a) PARSE » ALIGN-L-W » *COMPLEX → kɑst][əgɛn (likely)

(b) PARSE » *COMPLEX » ALIGN-L-W → kɑs][təgɛn (less likely)

(c) ALIGN-L-W » *COMPLEX » PARSE → kɑs]t[əgɛn (least likely)

Ranking (a) arises when the selection points fall near the center of each constraint and the actual ranking is identical to the ranking in the grammar. Ranking (b) reverses the close neighbors ALIGN-L-W and *COMPLEX, which is not very surprising since their ranking values are so close to each other. Ranking (c) puts the highest-ranked PARSE at the bottom, a selection that is possible, but highly unlikely. It is now clear how variation arises from stochastic candidate evaluation: selection points vary from evaluation to evaluation, which results in different rankings, which in turn results in different outputs. However, the fixed ranking values guarantee that some rankings will have a higher probability than others and the output will quantitatively reflect this.

Stochastic Optimality Theory comes with a learnability algorithm called the 'Gradual Learning Algorithm' (GLA, Boersma and Hayes 2001, Tesar [ch.24]). Given a set of constraints and variable learning data, GLA attempts to find the ranking values responsible for the learning data. The input to GLA consists of a set of arbitrarily ranked constraints (e.g. all ranked at 100) and some </input/, [output]> pairs that constitute the learning data. The algorithm first checks whether the current ranking correctly generates the current learning datum. If the answer is no, some learning must take place. This is done by adjusting the ranking values as follows: for every constraint violated by the learning datum (= desired winner), its ranking value is decreased by a small step, and for every constraint violated by the current winner (= wrong winner), its ranking value is increased by a small step. GLA is able to cope with free variation because the individual adjustments to the ranking values are very small: two constraints can be pushed slightly

closer to each other while maintaining their relative ranking.[6] Empirical tests of GLA have shown that the algorithm can approximate the relative frequencies in the learning data very well, given a reasonable set of constraints.

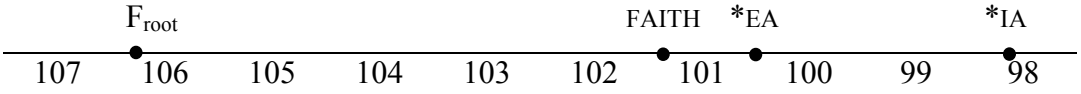
We now illustrate how GLA works using Finnish Vowel Coalescence data. To simplify things, we abstract away from the part-of-speech effect and only consider nouns, ignoring the four exceptional roots that coalesce to some extent (*häpeä* 'shame', *hopea* 'silver'*, aukea* 'opening'*, lipeä* 'lye'). The following tableau was given as input to the version of GLA implemented in OTSOFT.

(31) Vowel Coalescence in Colloquial Helsinki Finnish: regular nouns

			FREQUENCY	*EA	*IA	$\text{FAITH}_{\text{root}}$	FAITH
(a)	/suome-a/	suomea	421	*			
	'Finnish-PAR'	suomee	293				*
(b)	/idea/	idea	12	*			
	'idea'	idee	0			*	*
(c)	/ruotsi-a/	ruotsia	4045		*		
	'Swedish-PAR'	ruotsii	1014				*
(d)	/lattia/	lattia	847		*		
	'floor'	lattii	0			*	*

In a representative test run where the algorithm was allowed 50,000 learning trials and the grammar 2,000 test cycles, the following ranking values emerged:

(32) $\text{FAITH}_{\text{root}}$ = 106.240, FAITH = 101.306, *EA = 100.568, *IA = 98.126



(33) *Observations and predictions*

	obs%	**pred%**	**Example**	**Gloss**	***n***
/e-a/, NOUN	0.410	0.420	suome-a	'Finnish-PAR'	714
/ea/, NOUN (recent)	0.000	0.016	idea	'idea'	12
/i-a/, NOUN	0.200	0.130	ruotsi-a	'Swedish-PAR'	5,059
/ia/, NOUN	0.000	0.001	lattia	'floor'	847

As shown in (33), GLA performed quite well, with an average error of 2.414 percent per candidate. In general, one would expect Stochastic Optimality Theory to match quantitative patterns much better than its ordinal competitors because of the increased descriptive power drawn from real numbers. Note that the categorical patterns are not quite categorical: the grammar predicts marginal variation even in roots (**idee, *rasii*). If we increase the number of exposures to learning data, these ungrammatical forms will eventually disappear: in a run with 100,000,000 learning trials and 50,000 test cycles, $\text{FAITH}_{\text{root}}$ reached the ranking value 114.000 resulting in a virtually categorical pattern.

[6] Boersma and Hayes (2001) speculate that the size of these adjustment steps (called the PLASTICITY value) may decrease in the course of learning. The hypothesis is that a mature learner will be more reluctant to change her grammar in the face of unexpected data than an immature learner.

Somewhat surprisingly, the average error per candidate simultaneously increased to 4.050 percent.

Despite its great numerical power, there are variation patterns that are hard for Stochastic Optimality Theory, but easy for Multiple Grammars and Partially Ordered Grammars. For example, consider a pattern that requires a single constraint to range over a set of fixed constraints. In the following hypothetical example, the ranking A » B is fixed whereas the constraint C is unranked. In Reynolds' (1994) terminology, C is a "floating constraint".

(34) C » A » B
A » C » B
A » B » C

A grammar of this type is not possible under Stochastic Optimality Theory if we want to maintain the assumption that each constraint has the same standard deviation (Boersma and Hayes 2001). Faced with cases like this, one would either have to reanalyze the data using different constraints or complicate the theory by allowing different constraints to have different standard deviations.

22.4 External factors

So far, we have only discussed cases where variation is conditioned by internal (phonological, morphological) factors. There remains the important question how external factors such as age, gender, style, register, identity, ethnicity, social class, and target audience fit into the picture. Vowel Coalescence in Colloquial Helsinki Finnish provides an example of external conditioning that is just as systematic as the cases of internal conditioning discussed above. Of the several external factors discussed by Paunonen, we only consider age: the younger the speaker, the higher the coalescence rate (Paunonen 1995, Anttila in press).

(35) *The Age Effect*

	OLD (65-)		MIDDLE-AGED (40-45)		YOUNG (15-20)	
/ea/, NOUN	3.3%	(1/30)	18.2%	(4/22)	50%	(4/8)
/ia/, NOUN	0%	(0/337)	0%	(0/263)	0%	(0/247)
/e-a/, NOUN	24.3%	(65/267)	30.5%	(60/197)	67.2%	(168/250)
/i-a/, NOUN	7.8%	(147/1,886)	11.4%	(182/1,597)	43.5%	(685/1,576)
/ea/, ADJ	49.3%	(242/491)	64.9%	(334/515)	93.0%	(687/739)
/ia/, ADJ	0%	(0/80)	0%	(0/100)	0%	(0/81)
/e-a/, ADJ	–		–		–	
/i-a/, ADJ	9.6%	(146/1,519)	16.8%	(220/1,308)	64.2%	(923/1,437)

The age effect has a natural synchronic interpretation in all three models: young speakers, middle-aged speakers, and old speakers have internalized different GRAMMARS (Multiple Grammars), different partial orders (Partially Ordered Grammars), or different ranking

values (Stochastic Optimality Theory).[7] More generally, it is tempting to identify internal factors with grammatical constraints and external factors with choices among grammars, rankings, or ranking values. This view has the advantage of being consistent with a modular view of language: internal factors are about grammars, external factors are about how the grammars are used. This has the perhaps unintuitive consequence that the part-of-speech effect in Colloquial Helsinki Finnish becomes an external factor since it is described in terms of distinct rankings (cophonologies) as opposed to being hard-wired in the constraints. Such analytical ambiguities are to be expected since the distinction between grammar and usage is not given in nature, but remains a central topic of foundational debate in linguistics (Newmeyer 2003, Bybee 2005).

22.5 A methodological note

We conclude with a methodological remark. The typical way of obtaining data in linguistics is to use well-formedness judgments elicited from native speakers or reported in dictionaries and descriptive grammars. The study of variation usually requires more effort. In particular, the quantitative preferences that accompany variation are not easily accessible to intuition, neither are they commonly reported in descriptive grammars. In order to study phonological variation, especially its quantitative aspects, one will typically need large amounts of usage data, such as naturally occurring text or speech corpora, databases compiled by field linguists, dialectologists, and sociolinguists, or acoustic and articulatory databases. While general-purpose data resources are usually not designed to answer one's specific research questions, they can often be converted into a more useful form with a reasonable amount of work. For example, raw text and speech corpora may not be very useful in themselves, but their usefulness can be significantly enhanced by annotating them phonetically, phonologically, morphologically, and syntactically. This can be done either manually or (semi-)automatically with the help of computational parsers. Annotated corpora are extremely useful for hypothesis testing and they often reveal unobvious typological and quantitative generalizations that would be difficult or impossible to establish based on other types of data. Developing such resources for the purposes of phonological research is an important task for the future.

[7] Andries Coetzee (p.c.) points out that there is another way in which stochastic grammars can account for the influence of extragrammatical factors: the standard deviation of the normal distribution of the ranking values can be increased or decreased. Increasing the standard deviation will lead to more variation, decreasing it will lead to less variation. Such changes could be correlated with the formality of the speech situation: in more relaxed speech situations a speaker might be more lenient with her grammar in the sense of increasing the standard deviation of constraint rankings.

Bibliography

Aissen, Judith. 2003. 'Differential Object Marking: Iconicity vs. Economy', *Natural Language and Linguistic Theory* 21, 435-483.

Alderete, John. 2001. 'Root-Controlled Accent in Cupeño', *Natural Language and Linguistic Theory* 19, 455 - 502.

Anttila, Arto. in press. 'Derived environment effects in Colloquial Helsinki Finnish', in Kristin Hanson and Sharon Inkelas, (eds.), *The Nature of the Word: Essays in Honor of Paul Kiparsky*, MIT Press, Cambridge, Massachusetts.

Anttila, Arto. 2003. 'Finnish Assibilation', in Makoto Kadowaki and Shigeto Kawahara, (eds.), *Proceedings of NELS 33*, GLSA, Amherst, Massachusetts, pp.13-24.

Anttila, Arto. 2002a. 'Morphologically Conditioned Phonological Alternations', *Natural Language and Linguistic Theory* 20, 1-42.

Anttila, Arto. 2002b. 'Variation and phonological theory', in Jack Chambers, Peter Trudgill, and Natalie Schilling-Estes, (eds.), *Handbook of Language Variation and Change*, Blackwell, Malden, Massachusetts, and Oxford, U.K., pp. 206-243.

Anttila, Arto. 1997. 'Deriving variation from grammar', in Hinskens et al., (eds.), *Variation, Change and Phonological Theory*, John Benjamins, Amsterdam / Philadelphia, pp. 35-68.

Anttila, Arto, and Young-mee Yu Cho. 1998. 'Variation and change in Optimality Theory', *Lingua* 104, 31-56, Special issue on Conflicting Constraints. [See also John J. McCarthy, (ed.), 2003, pp. 569-580.]

Anttila, Arto, and Anthi Revithiadou. 2000. 'Variation in allomorph selection', in Masako Hirotani, Andries Coetzee, Nancy Hall and Ji-yung Kim, (eds.), *NELS 30, Proceedings of the North East Linguistic Society*, Vol. 1, pp. 29-42.

Auger, Julie. 2001. 'Phonological variation and Optimality Theory: Evidence from word-initial vowel epenthesis in Picard', *Language Variation and Change* 13, 253-303.

Bayley, Robert. 1995. 'Consonant cluster reduction in Tejano English', *Language Variation and Change* 6, 303-326.

Beckman, Jill N. 1998. *Positional Faithfulness*, Ph.D. dissertation, Department of Linguistics, University of Massachusetts, Amherst, ROA-234.

Benua, Laura. 1995. 'Identity effects in morphological truncation', in Jill N. Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, (eds.), *Papers in Optimality Theory*, University of Massachusetts Occasional Papers 18, GLSA, University of Masssachusetts, Amherst, pp. 77-136.

Berkley, Deborah Milam. 1994. 'Variability in Obligatory Contour Principle Effects', in Beals et al., (eds.), *CLS 30, Volume 2: Parasession on Variation in Linguistic Theory*, Chicago Linguistic Society, Chicago, pp. 1-12.

Boersma, Paul. 1998. *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. Ph.D. dissertation, University of Amsterdam, Holland Academic Graphics, The Hague.

Boersma, Paul, and Bruce Hayes. 2001. 'Empirical tests of the Gradual Learning Algorithm', *Linguistic Inquiry* 32, 45-86.

Boersma, Paul. 2001. 'Review of Arto Anttila (1997): Variation in Finnish Phonology and Morphology', *Glot International* 5, 33-40.

Borowsky, Toni, and Barbara Horvath. 1997. 'L-vocalization in Australian English', in Frans Hinskens, Roeland van Hout and Leo Wetzels, (eds.), *Variation, Change and Phonological Theory*, John Benjamins, Amsterdam / Philadelphia, pp. 101-123.

Bybee, Joan L. 2005. 'Grammar is usage and usage is grammar', Presidential Address, the Linguistic Society of America Annual Meeting, Oakland, CA, January 8, 2005. http://www.unm.edu/~jbybee/Bybee%20plenary.pdf.

Bybee, Joan L. 2001. *Phonology and Language Use*, Cambridge University Press, Cambridge, UK.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change, Language Variation and Change 14, pp. 261-290.

Cardoso, Walcir. 2001. 'Variation patterns in regressive assimilation in Picard', *Language Variation and Change* 13, 305-342. [In the same volume as Auger's paper, but different year. Something's wrong.]

Casali, Roderic F. 1996. *Resolving hiatus*. Ph.D. dissertation, Department of Linguistics, University of California, Los Angeles, ROA-215.

Cedergren, Henrietta J. and David Sankoff. 1974. 'Variable rules: performance as a statistical reflection of competence', *Language* 50, 333-355.

Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*, Harper and Row, New York.

Coetzee, Andries. 2004. *What it Means to be a Loser: Non-optimal Candidates in Optimality Theory*, Ph.D. dissertation, University of Massachusetts, Amherst.

Côté, Marie-Hélène. 2000. *Consonant Cluster Phonotactics: A Perceptual Approach.* Ph.D. dissertation, Department of Linguistics and Philosophy, Massachusetts Institute of Technology.

Demuth, Katherine. 1997. 'Multiple Optimal Outputs in Acquisition', in Viola Miglio and Bruce Morén, (eds.), University of Maryland Working Papers in Linguistics 5, Linguistics Department, University of Maryland, College Park, MD, pp. 53-71.

Fasold, Ralph. 1972. *Tense Marking in Black English*. Center for Applied Linguistics, Arlington.

Frisch, Stefan A., Janet B. Pierrehumbert, and Michael B. Broe. 2004. 'Similarity Avoidance and the OCP', *Natural Language and Linguistic Theory* 22, 179-228.

Guy, Gregory R. 1980. 'Variation in the group and the individual: The case of final stop deletion', in William Labov, (ed.), *Locating Language in Time and Space*, Academic Press, New York, pp. 1-36.

Guy, Gregory. 1991a. 'Explanation in variable phonology', *Language Variation and Change*, 3, 1-22.

Guy, Gregory. 1991b. Contextual conditioning in variable lexical phonology, *Language Variation and Change*, 3, 223-239.

Guy, Gregory R. 1994. 'The Phonology of Variation', in Beals et al., (eds.), *CLS 30, Volume 2: Parasession on Variation in Linguistic Theory*, The Chicago Linguistic Society, Chicago, pp. 133-149.

Guy, Gregory R. 1997a. 'Competence, Performance, and the Generative Grammar of Variation', in Frans Hinskens, Roeland van Hout and Leo Wetzels, (eds.),

Variation, Change and Phonological Theory, John Benjamins, Amsterdam / Philadelphia, pp. 125-143.

Guy, Gregory, and Charles Boberg. 1997a. 'Inherent Variability and the Obligatory Contour Principle', *Language Variation and Change* 9, 149-164.

Guy, Gregory R. 1997b. 'Violable is variable: Optimality Theory and linguistic variation', *Language Variation and Change* 9, 333-347.

Hammond, Michael. 1994. 'An OT account of variability in Walmatjari stress', ROA-20.

Hayes, Bruce. 2004. 'Stochastic Phonological Knowledge: The Case of Hungarian Vowel Harmony', Ms., UCLA. http://www.humnet.ucla.edu/humnet/linguistics/people/hayes/HayesNYUHandout.pdf.

Hayes, Bruce, and Margaret McEachern. 1998. 'Quatrain Form in English Folk Verse', *Language* 64, 473-507.

Hayes, Bruce, Bruce Tesar, and Kie Zuraw (2003) 'OTSoft: Optimality Theory Software, version 2.1', http://www.linguistics.ucla.edu/people/hayes/otsoft (13 Jan. 2004).

Hayes, Bruce. 2000. 'Gradient Well-formedness in Optimality Theory', in Joost Dekkers, Frank van der Leeuw and Jeroen van de Weijer, (eds.), *Optimality Theory: Phonology, Syntax, and Acquisition*, Oxford University Press, pp. 88-120.

Hinskens, Frans, Roeland van Hout, and Leo Wetzels, (eds.), *Variation, Change and Phonological Theory*, John Benjamins, Amsterdam / Philadelphia.

Holt, D. Eric (ed.), 2003. *Optimality Theory and Language Change*. Dordrecht: Kluwer Academic Publishers. *Studies in Natural Language and Linguistic Theory* 56.

Inkelas, Sharon, Orhan Orgun, and Cheryl Zoll. 1997. 'The implications of lexical exceptions for the nature of grammar', in Iggy Roca, (ed.), *Derivations and Constraints in Phonology*, Clarendon Press, Oxford, pp. 393-418.

Inkelas, Sharon. 1998. 'The theoretical status of morphologically conditioned phonology: a case study of dominance effects', in Geert Booij and Jaap van Marle, (eds.), *Yearbook of Morphology 1997*, Kluwer, Dordrecht, pp. 121-155.

Inkelas, Sharon. and Cheryl Zoll. 2003. 'Is Grammar Dependence Real?' Ms., University of California at Berkeley and Massachusetts Institute of Technology, ROA-587.

Itô, Junko, and Armin Mester. 1999. 'The phonological lexicon', in Natsuko Tsujimura, (ed.), *The Handbook of Japanese Linguistics,* Blackwell, Oxford, pp. 62-100.

Itô, Junko, and Armin Mester. 1995. 'Japanese Phonology', in John A. Goldsmith, (ed.), *The Handbook of Phonological Theory*, Blackwell, Oxford. pp. 817-38.

Itô, Junko, and Armin Mester. 1997. 'Correspondence and Compositionality: The Ga-gyō Variation in Japanese Phonology', in Iggy Roca, (ed.), *Derivations and Constraints in Phonology*, Clarendon Press, Oxford, pp. 419-462.

Iverson, Gregory K. and Shinsook Lee, 1994. 'Variation as Optimality in Korean Cluster Reduction', in Janet Fuller, Ho Han, and David Parkinson, (eds.), *ESCOL '94*, *Proceedings of the Eastern States Conference on Linguistics.*

Kager, René. 1999. *Optimality Theory*. Cambridge University Press, Cambridge.

Kang, Hyeon-Seok.1994. 'Variation in past-marking and the question of the system in Trinidadian English', in Beals et al., (eds.), *CLS 30, Volume 2: Parasession on*

Variation in Linguistic Theory, The Chicago Linguistic Society, Chicago, pp. 150-164.

Kang, Yoonjung. 2002. 'Frequency of use and analogical change in Korean nouns: a case against abstract representation', Poster presented at NELS 33, MIT, Cambridge, MA.

Kiparsky, Paul. 1993. 'Variable rules', Handout from Rutgers Optimality Workshop 1.

Kiparsky, Paul. 1995. 'The phonological basis of sound change', in John A. Goldsmith, (ed.), *The Handbook of Phonological Theory*, Blackwell, Oxford. pp. 640-670.

Kroch, Anthony. 1989. 'Reflexes of grammar in patterns of language change', *Language Variation and Change* 1, 199-244.

Labov, William, Paul Cohen, Clarence Robins, John Lewis. 1968. 'A Study of the Nonstandard English of Negro and Puerto Rican Speakers of New York City', U.S. Regional Survey. Philadelphia.

Labov, William. 1969. 'Contraction, deletion and inherent variability of the English copula.' *Language* 45, 715-62.

Labov, William. 1994. *Principles of Linguistic Change: Internal Factors*, Blackwell, Oxford,U.K. and Cambridge, U.S.A.

Labov, William. 1997. 'Resyllabification', in Frans Hinskens, Roeland van Hout and Leo Wetzels, (eds.), *Variation, Change and Phonological Theory*, John Benjamins, Amsterdam / Philadelphia, pp. 145-179.

Laalo, Klaus. 1988. *Imperfektimuotojen ti~si -vaihtelu suomen kielessä* [*The Past Tense ti~si Alternation in Finnish*], Suomalaisen Kirjallisuuden Seura, Helsinki.

Liberman, Mark Y. 1993. 'Optionality and Optimality', fragment of a draft, Department of Linguistics, University of Pennsylvania.

McCarthy, John J. and Alan S. Prince. 1993. 'Generalized Alignment', in Geert Booij and Jaap van Marle, (eds.), *Yearbook of Morphology,* Kluwer, Dordrecht, pp. 79-153.

McCarthy, John J. and Alan S. Prince. 1995 'Faithfulness and Reduplicative Identity', in Jill Beckman, Suzanne Urbanczyk, and Laura Walsh Dickey, (eds.), *University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory*, pp. 249-384.

McCarthy, John J. (ed.). 2003. *Optimality Theory in Phonology: A Reader*, Blackwell Publishing, Malden, Mass.

McCarthy, John J. 2002. *A Thematic Guide to Optimality Theory*, Cambridge University Press, Cambridge.

Morris, Richard E. 1998. *Stylistic Variation in Spanish Phonology*, Ph. D. dissertation, Ohio State University.

Myers, James. 2003. 'Frequency effects in Optimality Theory', Ms., National Chung Cheng University.

Myers, James, and Gregory Guy. 1997. 'Frequency effects in Variable Lexical Phonology', *University of Pennsylvania Working Papers in Linguistics* 4, 215-227.

Nagy, Naomi and William Reynolds. 1997. 'Optimality Theory and variable word-final deletion in Faetar', *Language Variation and Change* 9, 37-55.

Neu, Helene. 1980. 'Ranking of Constraints on /t,d/ Deletion in American English: A Statistical Analysis', in William Labov, (ed.), *Locating Language in Time and Space*, Academic Press, New York, pp. 37-54.

Newmeyer, Frederick. 2003. 'Grammar is grammar and usage is usage', *Language* 79, 682-707.

Oostendorp, Mac van.1997. 'Style registers in conflict resolution', in Frans Hinskens, Roeland van Hout and Leo Wetzels, (eds.), *Variation, Change and Phonological Theory*, John Benjamins, Amsterdam / Philadelphia, pp. 207-229.

Orgun, Cemil Orhan. 1996. *Sign-Based Morphology and Phonology with special attention to Optimality Theory*, Ph.D. dissertation, Department of Linguistics, University of California, Berkeley. ROA-171.

Paolillo, John C. 2002. *Analyzing Linguistic Variation: Statistical Models and Methods*. CSLI Publications, Stanford, California.

Pater, Joe. 2000. 'Nonuniformity in English stress: the role of ranked and lexically specific constraints', *Phonology* 17, 237-274.

Patrick, Peter. 1991. 'Creoles at the intersection of variable processes: *-t, d* deletion and past-marking in the Jamaican mesolect', *Language Variation and Change* 3, 171-189.

Paunonen, Heikki. 1995. *Suomen kieli Helsingissä* [*The Finnish Language in Helsinki*], Helsingin yliopiston suomen kielen laitos, Helsinki.

Phillips, B.S. 1984. 'Word frequency and the actuation of sound change', *Language* 45, 9-25.

Phillips, B.S. 2001. 'Lexical diffusion, lexical frequency, and lexical analysis', in Joan. L. Bybee and Paul Hopper, (eds.), *Frequency and the Emergence of Linguistic Structure*, Amsterdam, John Benjamins, pp. 123-136.

Pierrehumbert, Janet. 1994. 'Knowledge of Variation', in Beals et al., (eds.), *CLS 30, Volume 2: Parasession on Variation in Linguistic Theory*, The Chicago Linguistic Society, Chicago, pp. 232-256.

Pierrehumbert, Janet. 2003. 'Probabilistic Phonology: Discrimination and Robustness', in Rens Bod, Jennifer Hay, and Stefanie Jannedy, (eds.), *Probabilistic Linguistics*. The MIT Press, Cambridge, Massachusetts, and London, England, pp. 177-228.

Prince, Alan, and Paul Smolensky 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell Publishing, Malden, MA.

P&S 1993. See Prince and Smolensky 1993.

Prince, Alan, and Paul Smolensky. 1993 [P&S 1993]. *Optimality Theory: Constraint Interaction in Generative Grammar*, Rutgers University, New Brunswick, and University of Colorado, Boulder. [See also Prince and Smolensky 2004]

Reynolds, William T. 1994. *Variation and Phonological Theory*. Ph.D. dissertation, Department of Linguistics, University of Pennsylvania.

Ringen, Catherine, and Orvokki Heinämäki. 1999. 'Variation in Finnish vowel harmony: An OT account', *Natural Language and Linguistic Theory* 17, 303-337.

Sankoff, David. 1988. 'Variable rules', in U. Ammon, N. Dittmar, and K. J. Mattheier, (eds.), *Sociolinguistics: An International Handbook of the Science of Language and Society*, Vol. 2, Walter de Gruyter, Berlin, pp. 984-997.

Santa Ana, Otto. 1996. 'Sonority and syllable structure in Chicano English', *Language Variation and Change* 8, 63-89.

Santa Ana, Otto. 1992. 'Locating the linguistic cycle in vernacular speech: Chicano English and the Exponential Hypothesis', in J. M. Denton, G. P. Chan and C. P. Canakis, (eds.), *CLS 28: Papers from the 28th Regional Meeting of the Chicago Linguistic Society, Vol. 2: The cycle in linguistic theory*, pp. 277-287.

Santa Ana, Otto. 1991. *Phonetic Simplification Processes in the English of the Barrio: A Cross-Generational Sociolinguistic Study of the Chicanos of Los Angeles.* Ph.D. dissertation, University of Pennsylvania.

Smith, Jennifer L. 1997. 'Noun faithfulness: On the privileged status of nouns in phonology', ROA-242.

Tesar, Bruce. (2005). 'Learnability' in Paul de Lacy (ed.) *The Cambridge Handbook of Phonology*. Cambridge: Cambridge University Press, ch.24.

Zubritskaya, Katya. 1997. 'Mechanism of sound change in Optimality Theory', *Language Variation and Change* 9, 121-48.

Zuraw, Kie. 2000. *Patterned Exceptions in Phonology*, Ph.D. dissertation, Department of Linguistics, University of California, Los Angeles.

To appear in Paul de Lacy (ed.) *The Cambridge Handbook of Phonology*. Cambridge University Press.