Part II
Linguistic Structure

Linguistic Structure

Of all the subfields of sociolinguistics, the study of linguistic variation is perhaps the one with the strongest emphasis on the "linguistic" side of "sociolinguistics." While variationists are indeed concerned with understanding social structures and forces, as well as with helping to effect social change, they are also vitally interested in furthering the scientific understanding of language. Unlike theoretical linguists, who typically rely on idealized versions of homogeneous languages in their search for underlying structure, variationists maintain that any valid linguistic theory must give central place to the variation and change that pervade all human languages. Unfortunately, it has proven difficult to incorporate variation into linguistic theory or theoretical linguistic models into variation study. However, as the chapters in this section amply demonstrate, the effort has not been abandoned, and important inroads are being made.

In "Variation and phonological theory," Arto Anttila demonstrates that, despite the failure of early attempts to modify phonological rules to incorporate variation, current efforts at reconciling phonological theory and variable data have proven more successful. In large part, this is because of the dramatic transformation of phonological theory, as it has moved from early rule-based approaches to Optimality Theory, which is based on constraints on the well-formedness of output strings rather than on inviolable rules. Constraints are universal but violable, and differences between different languages are the result of different orderings of constraints among different groups of speakers. It is a small step to extend this notion to dialect differentiation. Using data from English and Finnish, Anttila takes us through step-by-step analyses that demonstrate that Optimality Theory can account quite well for observed patterns of intra-language variation, especially if the theory is modified slightly to allow for multiple constraint orderings (multiple grammars) within the individual speaker. Anttila further demonstrates that certain types of multiple-grammar models can account not only for the qualitative fact of variation but for observed quantitative patterns. Hence, Optimality Theory is a promising direction indeed as researchers work to bridge the gap between phonological theory and variation study.

Although it has been difficult to incorporate variation into theories of synchronic linguistic structure, variationists have been more successful in informing the study of language change. In "Investigating Chain Shifts and Mergers," Matthew J. Gordon discusses variationist investigations of two important types of language change and demonstrates how these studies have led us to rethink traditional notions regarding how and even why such changes take place. Investigations of vowel mergers have called into question the traditional functionalist notion that speakers make distinctions between sounds in order to make meaning distinctions, since researchers have uncovered cases in which speakers produce distinctions between vowel sounds but cannot perceive them – that is, cannot use them to distinguish meanings. Similarly, the variationist investigation of chain-shifting, the related movement of a series of vowels through phonetic space, also raises questions for long-established views on language change. Data on chain shifts currently in progress (especially the Northern Cities Shift in the USA) provide evidence against the functionalist notion that the motivation underlying chain shifting is the need to preserve phonological distinctions – that is, to avoid merger when one vowel encroaches on the phonetic space of another. It is only through investigating vowel changes in progress rather than "after the fact" that we gain understanding of not only the linguistic but also the social forces that drive language change.

Variation analysis has had little impact on syntactic theory over the decades. In "Variation and Syntactic Theory," Alison Henry discusses reasons for this, as well as reasons why syntactic theory should incorporate variable data – and why variation study should take better account of theoretical considerations. Although syntacticians have traditionally abstracted away from the variation they find, the fact remains that variation is pervasive and linguistic theories rest on shaky foundations indeed when we attempt to base them on invariant data. A central tenet of the Minimalist Program is that movement only occurs when it is forced; hence, variable movement is impossible. However, the observed facts of syntactic change seem difficult to reconcile with the notion that each speaker's individual grammar is, at heart, invariable. It may be possible to account for syntactic change in general by positing that speakers have several competing grammars, each of which is invariant, and that change results when one competitor wins out, but this view does not seem to allow for stable variation between old and new forms over long periods of time, a phenomenon that is by no means rare. Just as variable data can inform syntactic theory, so too can theory inform the study of language variation and change. For example, theoretical syntactic considerations can lead to a fuller understanding of the types of conditioning factors that are likely to affect variation, as well as the types of features (e.g. marked vs. unmarked) that are likely to be affected by change. Despite the longstanding gap between variation study and theoretical syntax, Henry notes that valuable contributions are currently being made toward bridging this chasm, to the mutual benefit of both fields.

In the final chapter in this section, "Discourse Variation," Ronald Macaulay investigates how variation study relates to discourse structure. Macaulay notes that there are many problems associated with conducting quantitative analyses in this area. It is difficult to establish the "envelope of variation" for discourse-level features (for example, discourse markers, tag questions), since it is nearly impossible to identify all and only the locations where such features could occur. In addition, it is difficult to establish variants of a particular variable (that is, different ways of saying the same thing), since the meanings associated with discourse features tend to be multifaceted and quite context-specific. However, variationists have conducted both quantitative and qualitative analyses of discourse-level features with fruitful results. Macaulay outlines a number of these studies, ranging from Labov's classic studies of narrative structure to investigations of the discourse patterns of women vs. men, to studies of discourse-level differences across social class groups, including Macaulay's own work in Scotland. Such studies are not without their limitations, and Macaulay urges caution as variationists proceed with their work in discourse. For example, sociolinguistic interviews may be quite different from naturally occurring conversations at the discourse level, since the interviewee usually does most of the talking. However, Macaulay stresses that even in the most one-sided interactions, talk is never simply an individual effort but always a joint production; hence variationists must pay attention to the contributions of *all* participants, no matter how seemingly unimportant.

Natalie Schilling-Estes

8 Variation and Phonological Theory

ARTO ANTTILA

In recent years, there have been several attempts to understand language variation from the perspective of Optimality Theory (Prince and Smolensky 1993).[1] This chapter has three main goals: (1) to point out why variation matters to the theoretical phonologist; (2) to list some central questions that any phonological theory of variation should address; (3) to give a brief introduction to the use of Optimality Theory in language variation, including concrete examples. Since the discussion will necessarily be selective, the reader will find it useful to follow up on the cited work for a fuller picture.

1 Internal Factors in Variation

Phonological variation is often studied from a sociolinguistic point of view, i.e. by examining the use of variants as a function of EXTERNAL FACTORS, such as sex, age, style, register, and social class. The fact that variation is also conditioned by INTERNAL FACTORS, such as phonology, morphology, syntax, and the lexicon, is what makes it interesting for the theoretical phonologist. We start with two concrete examples of internal conditioning. The purpose of these examples is to illustrate the sorts of phonological, morphological, and lexical facts that any grammatical theory of variation should explain.

In many dialects of English, word-final consonant clusters are variably simplified by deleting a coronal stop, e.g. *cost me~cos' me*. This is a well-studied phenomenon with interesting phonological and morphological structure (see e.g. Labov et al. 1968, Wolfram 1969, Fasold 1972, Guy 1980, Neu 1980, Guy and Boyd 1990, Guy 1991a, 1991b, Kiparsky 1993b, Reynolds 1994, Guy 1994, Guy and Boberg 1997, Labov 1997, among others). The examples in (1) illustrate two well-known generalizations: *t,d*-deletion rate depends on the quality of the following segment and the morphological status of the segment subject to deletion.

(1) a.
t,d-deletion rate: the following segment effect. Celeste S., Philadelphia, word-final clusters only. Figures from Labov 1997.

Following segment		Following segment	
stop	78%	/l/	40%
/w/	68%	pause	17%
fricative	65%	vowel	6%
nasal	57%	/r/	7%
/h/	45%	/y/	5%

b.
t,d-deletion rate: the morphological effect. Figures from Guy 1994.

Morphology	Guy 1991b	Santa Ana 1992
Monomorphemes (*cost*)	38.1%	57.9%
Irregular past (*lost*)	33.9%	40.7%
Regular past (*tossed*)	16.0%	25.7%

Various explanations for these effects have been offered. The following segment effect has been commonly attributed to syllable structure. Specifically, it has been proposed that the final coronal consonant tends to be retained if it can be resyllabified as a part of the following onset. This is the case before vowels, e.g. in *lost.Anna~los.tAnna*, but not before /l/, e.g. *lost.Larry* (**los.tLarry*) because *tl* is not a possible onset in English. The resyllabification hypothesis thus predicts that there should be more deletion before /l/ than before /r/ because *tr* is a possible onset in English. This is confirmed by the data (Guy 1994: 143–4, Labov 1997: 165).

As for the morphological generalization, at least two different explanations have been put forward. Kiparsky (1982) proposed that deletion is less frequent in the regular past tense (*tossed*) because it would make past and present tenses identical. The explanation would thus be homonymy avoidance, a principle with an obvious functional appeal (but see Labov 1994: ch. 19). A very different structural explanation was put forward by Guy (1991b) who posited a variable rule of *t,d*-deletion which applies inside out in the morphological structure: first to roots (*cost*), then to stems (*cost, los+t*) and finally to words (*cost, los+t, toss#ed*), subjecting roots to deletion three times, stems twice and words only once. This hypothesis makes the specific quantitative prediction that the retention rates in the three groups should be related exponentially, in the ratios of x, x^2 and x^3. This appears to be confirmed by the data (Guy 1991b).

As a second example, consider variable vowel coalescence in Colloquial Helsinki Finnish (Kiparsky 1993a, Paunonen 1995, Anttila to appear a). Vowel coalescence turns heterosyllabic vowel sequences into long vowels as follows:

(2) Colloquial Helsinki Finnish vowel coalescence
a. má.ke.a~má.kee "sweet"
b. lá.si.-a~lá.sii "glass-PAR(TITIVE)"

Rule: $V_1.V_2 \rightarrow V_1$: where V_1,V_2 are unstressed and V_2 is [+low]

Vowel coalescence is variable within an individual. The following dialogue fragment comes from an electronic corpus of spoken Helsinki Finnish (Paunonen 1995).[2] The same speaker sometimes does, sometimes does not coalesce, apparently unpredictably. The last line shows that both variants may occur within the same noun phrase (for more examples, see Paunonen 1995: 106–7).

(3) OH: Millas**ii** ihmis**ii** siel käy jud**oo**massa?
/millas-i-a/, /ihmis-i-ä/, /judoa-ma-ssa/
What sort of people practise judo there?
JS: Siel käy iha, nuor**ia** ja vanhojaki.
/nuor-i-a/
Some are really young, but there are old people too.
OH: Mieh**iä** nais**ia**?
/mieh-i-ä/, /nais-i-a/
Men? Women?
JS: Joo mieh**ii** ja nais**ia**.
/mieh-i-ä/, /nais-i-a/
Yes, men and women.

While the variation may initially seem random, vowel coalescence turns out to be subject to the following phonological and morphological conditions:

(4) a. Coalescence is favored in mid-low sequences (/ea/, /oa/, /öä/) and disfavored in high-low sequences (/ia/, /ua/, /yä/) (Paunonen 1995).
b. Coalescence is favored in derived environments and disfavored in roots (Anttila to appear a).
c. Coalescence is favored in adjectives and disfavored in nouns (Anttila to appear a).

Example (5) shows how these regularities emerge in Paunonen's corpus based on the phonological and morphological tagging carried out by the present author. The sample is based on 126 speakers and contains approximately 13,000 coalescence environments. The factors favoring coalescence are boldfaced. There were no examples of derived /ea/-final adjectives in this corpus.

(5) Vowel coalescence rate by environment

		/**ea**/	/ia/	*Examples*
a.	Noun & Root	14.8%	0%	hopea, rasia
	Noun & **Derived**	41.0%	20.0%	suome-a, lasi-a
	Adjective & Root	72.4%	0%	makea, kauhia
	Adjective & **Derived**	–	30.2%	–, uus-i-a

		Derived	Root	
b.	Noun & /ia/	20.0%	0%	lasi-a, rasia
	Noun & **/ea/**	41.0%	14.8%	suome-a, hopea
	Adjective & /ia/	30.2%	0%	uus-i-a, kauhia
	Adjective & **/ea/**	–	72.4%	–, makea

		Adjective	Noun	
c.	Root & /ia/	0%	0%	kauhia, rasia
	Root & **/ea/**	72.4%	14.8%	makea, hopea
	Derived & /ia/	30.2%	20.0%	uus-i-a, lasi-a
	Derived & **/ea/**	–	41.0%	–, suome-a

Glosses: *hopea* "silver", *rasia* "box", *makea* "sweet", *kauhia* "terrible", *suome-a* "Finnish-PAR", *lasi-a* "glass-PAR", *uus-i-a* "new-PL-PAR".

The regularities in (4) sometimes yield categorical, sometimes quantitative effects. For example, as (5a) shows, changing /ea/ to /ia/ in nouns results in categorical blocking in roots, but only in a quantitative dispreference in derived environments. If we divide the speakers into groups by sex (2 groups), age (3 groups), social class (3 groups), and neighborhood (2 groups), the same regularities continue to hold in each subgroup.

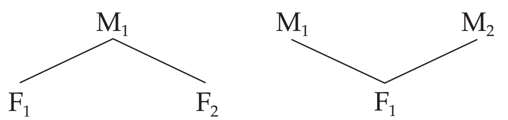
In addition to phonological and morphological conditions, both cases of variation show lexical conditions as well. Myers and Guy (1997) report that *t,d*-deletion frequency is higher in high-frequency words than low-frequency words, but intriguingly, only in monomorphemes. Thus, the deletion rate is higher in *past* (high-frequency) than *priest* (low-frequency), whereas regularly inflected past tense forms, such as *passed* (high-frequency) and *kissed* (low-frequency) are not significantly different. As for Finnish, native noun roots like *hopea* "silver" coalesce variably, whereas otherwise identical recent borrowings like *idea* "idea" never coalesce.

These two examples illustrate internal conditioning: the distribution of variant pronunciations is conditioned by phonology, morphology, and the lexicon. It is evident that phonological, morphological, and lexical regularities sometimes yield obligatory and categorical effects (i.e. "rules"), sometimes variable and quantitative effects (i.e. "tendencies"). Note that in neither case can the variability and quantitative effects be reduced to phonetics: both *t,d*-deletion and vowel coalescence are morphophonemic alternations, witness the presence of morphological and lexical conditions.

One possibility would be to assume that phonological theory should only explain categorical, but not quantitative regularities. However, while it is perfectly acceptable to simplify the object of study, it is hard to see why one would want to seriously commit oneself to such a position. It would be like declaring a priori that phonological theory should only explain alternations, but not static phonotactic regularities, or that it should only explain allophonic, but not morphophonemic alternations. There seems to be no reason to limit the scope of phonological theory in such ways. The question is how exactly phonological theory should address facts of this sort; for a general discussion of various options, see Pierrehumbert (1994). The work reviewed in this chapter takes the view that both categorical and quantitative regularities derive from the same grammatical principles, which need to be identified, and that the differences in realization strength (invariant vs. variable, categorical vs. quantitative) can be explained by referring to the ways these principles interact.

Finally, a brief note on the definition of variation is due. One possibility is to regard variation as a form-meaning relation of a special kind. It has been suggested that the form-meaning relation in natural languages is ideally one-to-one and that this is a principle that languages strive to satisfy (Anttila 1989). The two possible types of deviations from this ideal state are illustrated in (6): one meaning (M_1) corresponds to several forms (F_1, F_2) i.e. we have VARIATION; one form (F_1) corresponds to several meanings (M_1, M_2), i.e. we have AMBIGUITY.

(6) Variation and ambiguity



While the existence of ambiguity is usually accepted as a mundane fact, the existence of intra-individual variation is sometimes questioned. For example, it has been suggested that apparent variation results from lumping together individuals with different, but invariant dialects (Bickerton 1971). However, well-documented cases of intra-individual variation clearly exist, among them our two examples. Another common observation is that putative free variants sometimes have subtly different meanings, suggesting that genuine variation may be hard to find. We will return to this point in section 3.3. In any case, the issue is less important than it might first seem. What makes phenomena like *t,d*-deletion and vowel coalescence interesting for the phonologist is that they reflect the workings of phonology, morphology, and the lexicon, sometimes categorically, sometimes quantitatively, and these facts call for an explanation, no matter whether the phenomena are instances of "genuine" variation or not.

2 The Goals of a Phonological Theory of Variation

The following is an open-ended list of questions that any phonological theory of variation should address. The discussion draws upon Liberman (1994) where many of the questions were originally stated.

2.1 The locus of variation

Why does variation occur in some environments, but not in others? While it is of some interest to study those isolated cases of variation that we happen to notice, it would be more interesting to know why just these cases of variation exist instead of the many other conceivable ones that do not. For example, *t,d*-deletion in e.g. *cos(t) me* has been attributed to cluster simplification due to ease of articulation, itself presumably a functional universal (Chambers 1995: 235–7). But why is the deleted segment /t/, instead of /s/ or /m/, even though deleting any of the three would yield a simpler cluster? Why are only coronal stops deleted instead of all obstruents or all consonants? Why only final clusters, but not initial as well (*stress* → **sress*)? Why consonant deletion instead of vowel epenthesis (**cost*[ə]*me*) which would also serve to break up the cluster? At this point, the functional explanation is no longer very helpful. Moreover, in some languages we find the reverse of the English pattern. In Lardil, only coronals fail to be deleted in the coda (Wilkinson 1988, Kiparsky 1993b). In Finnish, we find variable initial cluster simplification in loanword nativization, e.g. *stress* becomes *stressi~tressi~ressi*, depending on the speaker. While articulatory economy may be a driving force behind *t,d*-deletion, it leaves us puzzled about the details. The conclusion is inevitable: if we want to explain what varies and how in each language, we need a good understanding of the language-specific phonologies. One commonsense intuition is that optionality/variation arises in environments where the regularities of the language are somehow "relaxed" or where they "conflict." What exactly this means in any given language is a question that can only be answered through phonological analysis.

2.2 The degrees of variation

Why are phonological alternations sometimes obligatory, but sometimes only optional? Within the latter group, how can we explain frequency effects that reflect internal factors such as syllable structure, segment quality, vowel height, morphological constituency and part of speech? If the same grammatical principles are responsible for both categorical and quantitative effects, what determines the realization strength in a particular case? One grammatical constraint that has been studied cross-linguistically from this point of view is the Obligatory Contour Principle (Leben 1973, Goldsmith 1976): "Adjacent identical elements are prohibited." For example, Guy and Boberg 1997 argue that *t,d*-deletion is more likely the more similar the preceding segment is to *t,d*. Thus, *wrist* shows a higher deletion rate than *rift* because /s/ and /t/ share two features, [+cor] and [–son], whereas /f/ and /t/ share only one, [–son]. For other quantitative OCP-effects, see McCarthy (1986, 1988, 1994), Frisch et al. (1997, Arabic), Berkley (1994, English), Anttila (to appear b, Finnish), Kang (1996, Korean) and Liberman (1994, Latin).

2.3 Markedness

How are the loci and degrees of variation related to the cross-linguistic issues of naturalness or markedness? For example, are new variants which are initially optional/variable always easier to articulate, either easier universally (cf. cluster simplification as a result of articulatory economy) or easier in some language-particular relativized sense? Does variation arise in situations where there are two equally marked alternatives available and obligatory patterns in situations where there is one clearly least marked alternative available? Does the quantitative distribution of variants reflect markedness, i.e. are the common variants less marked than the less common variants in some independently definable sense?

2.4 Interfaces

How do phonological, morphological, and lexical factors interact in variation? How should phonological theory explain cases where variation shows a phonological pattern, perhaps even a universal tendency, yet at the same time different word classes and individual lexical items show their own characteristic patterns of behavior, perhaps by being totally or partially exempt from variation or by showing a different quantitative pattern? The interaction between phonology, morphology, and the lexicon is a central problem in synchronic phonological theory; see e.g. Hargus (1993) and Mohanan (1995) for recent overviews. On the diachronic side, the problem is related to the long-standing controversy between neogrammarians and lexical diffusionists (Wang 1969, Labov 1994, Kiparsky 1995): is sound change mechanical and phonetically conditioned or does it proceed word by word? Variation is a unique source of evidence as it often reveals subtle implicational and quantitative relations among phonological, morphological, and lexical factors.

2.5 External factors

How do internal factors interface with external factors in variation? While it is not the business of grammatical theory to explain the effects of sex, age, style, register and social class, one would like to have at least a plausible scenario of where such facts fit. There would seem to be two possibilities: (1) The modular view: internal and external factors are of a fundamentally different nature: grammars are structural objects built out of (innate) universal principles; external factors reflect the ways in which these structural objects are used. This implies that external factors can be reduced to choices among grammars; (2) The anti-modular view: there is no important theoretical difference between internal and external factors, which interact with each other fairly directly. This seems to be the position assumed in a good part of the Variable Rule literature where internal and external constraints are freely mixed in the context conditions of variable rules.

2.6 *Language change*

Change presupposes a period of variation although variation need not produce change. Why do some cases of variation linger for centuries without very much change, while others move, quickly or slowly, towards a categorical resolution? (Liberman 1994: 1)

In the current state of theoretical phonology, the immediate problem facing anyone who wants to work on variation is to find a formal framework that is capable of at least describing variable phenomena, including their quantitative aspects. As we will see, several proposals have been put forward. These proposals can be evaluated in terms of descriptive adequacy, e.g. how accurate the resulting quantitative predictions are. The next step is to evaluate them in terms of their restrictiveness, i.e. whether they exclude anything, and if they do, whether they exclude the linguistically bizarre and hence systematically unattested types of variation and allow the linguistically plausible and, in particular, the actually attested types of variation. Finally, one can evaluate the proposals in terms of whether they generalize in the right way beyond the variation problem. For example, does a particular approach shed light on the interaction of phonology and morphology more generally? Does a theory that is developed for intra-individual variation lend itself to explaining cross-linguistic variation as well? Are the proposed classes of grammars learnable?

3 Standard Optimality Theory

Classical generative phonology outlined in Chomsky and Halle's (1968) *The Sound Pattern of English* (SPE) recognized the existence of optionality in phonology and provided two devices for describing it: optional rules and allomorphy. Going beyond optionality, Labov (1969) and Cedergren and Sankoff (1974) proposed a statistical technique for estimating the relative weight of different internal and external factors based on corpus frequencies. They further proposed to associate the resulting weights with the structural descriptions of SPE-style phonological rules. This well-known Variable Rule model has been highly influential in variationist phonology for over three decades.

Meanwhile, phonological theory has moved on. In particular, there has been a gradual shift away from rule-based theories to constraint-based theories which leave little if any role for rules and derivations. Consequently, the notion "variable rule" is not easy to interpret in the context of current phonological theory. At the same time, work on phonological variation has continued largely independently of phonological theory, often consciously emphasizing its empirical character. While variationist phonological analyses make heavy use of the quantitative analysis tools developed in connection with the Variable Rule model, this is often done with no particular commitment to the theoretical notion "variable rule" itself (Fasold 1991).

One constraint-based framework that emerged in the early 1990s and soon became very influential, especially in phonology, is Optimality Theory (OT) (Prince and Smolensky 1993). Among other things, Optimality Theory opened up new ways of thinking about variation. In the rest of this chapter, we will examine some of these recent approaches, emphasizing their similarities and differences, and, where possible, evaluate their merits and shortcomings. Some of the papers discussed here are available at the Rutgers Optimality Archive (http://roa.rutgers.edu).

3.1 A synopsis of Optimality Theory

In this section, we give a brief synopsis of Optimality Theory. (For thorough textbook discussions, see e.g. Archangeli and Langendoen 1997 and Kager 1999.) An optimality-theoretic grammar consists of the following components:

(7) a. Lexicon: A set of inputs (underlying forms).
   b. Generator (Gen): A function that takes an input and returns a set of output candidates. Gen is maximally permissive: any output candidate is permitted for any input within very general limits of structural well-formedness. Thus, for the input /kæt/ "cat" Gen would yield {[kʰæt], [kæt], [kæ], [ba], . . . }.
   c. Evaluator (Eval) which consists of two parts: (1) A universal constraint set Con which contains two principal types of constraints: faithfulness constraints that prefer the output that most closely resembles the input and markedness constraints that prefer the phonologically least marked output; (2) A language-particular ranking of the universal constraints. Ranking selects the least offensive candidate from among the output candidates and calls it optimal. This is the actual output.

We now turn to a concrete example. Assume the following three constraints. For more discussion and exemplification, see Kager (1999: ch. 3).

(8) Onset Onsets are required. (markedness)
   *CxCod No coda clusters. (markedness)
   Max(C) No consonant deletion. (faithfulness)

One of the central hypotheses of Optimality Theory is that constraints are universal. This means that variation among languages should be reducible to differences in ranking. Let us assume that the three constraints in (8) are ranked as in (9) in a particular dialect of English. Ranking is indicated by ">>".

(9) Onset >> *CxCod >> Max(C)

We now work out the pronunciation this grammar assigns to the input /mɪst/ using the tableau in (10). For the purposes of this illustration, we only consider two output candidates: [mɪst] and [mɪs].

(10)

Input: /mɪst/	Onset	*CxCod	Max(C)
1a. [mɪst]		*!	
1b. ⇒ [mɪs]			*

Both candidates have exactly one syllable which has an onset ([m]). Thus, both satisfy Onset. The candidates differ with respect to the last two constraints. Candidate (1a) [mɪst] is perfectly faithful: it realizes all the input segments and only those. However, it contains a coda cluster [st] in violation of *CxCod. A constraint violation is marked by "*". Candidate (1b) [mɪs] does not violate *CxCod because [t] has been deleted, but the deletion amounts to a violation of Max(C). Thus, both (1a) and (1b) incur exactly one constraint violation. The decisive factor is that in this dialect *CxCod is ranked higher than Max(C). This makes the *CxCod violation fatal for (1a). A fatal violation is marked by "!". As a consequence, (1a) loses and (1b) wins. Since we now have a winner, the remaining constraint columns are irrelevant, which is indicated by shading. More generally, given two candidates, we can find the winner by checking the highest constraint on which the two candidates differ (here *CxCod): the winner is the candidate that better satisfies this constraint. The candidate that beats all its competitors is called the optimal candidate and marked by the arrow. Note that the optimal candidate need not be perfect: it is just better than any of its competitors.

The dialect described here is a *t,d*-deletion dialect of a slightly odd type: the final *t* is invariably deleted under the pressure to avoid coda clusters created by the ranking *CxCod >> Max(C). The reverse ranking Max(C) >> *CxCod yields a dialect where *t* is invariably retained. Since most dialects are variable, the question arises how variation can be accommodated in standard Optimality Theory. It turns out that there are at least two ways of getting at variation without modifying any of the assumptions laid out so far. We will now examine these two approaches in turn.

3.2 *Tied violations*

In *tied violations*, two or more candidates incur exactly the same violations with respect to all the constraints in the grammar. As an example, consider variable stress in Walmatjari, spoken by fewer than a thousand people in the

Kimberley division of Western Australia (Hudson and Richards 1969). The analysis is based on Hammond (1994); the data come from Hudson and Richards (1969: 183–5). The basic generalization goes as follows: (1) in all two-syllable words, main stress falls on the first syllable; (2) in words with three syllables, main stress may fall on the first or second syllable, with most words showing variable pronunciation; (3) in words with four syllables, main stress may fall on the first or second syllable, with some words showing variable pronunciation.[3]

(11)	a.	yápa	"child"
		pálma	"creek"
	b.	máṇalu~maṇálu	"we (pl. excl.)-him"
		yútanti~yutánti	"sit"
	c.	páljmanàna~paljmánana	"touching"
		ʈúṇmanàna~ʈuṇmánana	"burying"

The question is what determines the locus of variation: why does stress vary between the first and second syllable, instead of, say, the first and the last syllable? And why only in words longer than two syllables? Let us assume that stress in Walmatjari is assigned by grouping syllables into binary feet where the strong syllable is on the left. The resulting trochaic stress pattern is cross-linguistically very common. Variability now follows from the fact that it is sometimes possible to group syllables into feet in more than one way. Following the ideas in Hammond (1994), we propose the constraints in (12). For a textbook discussion of stress in Optimality Theory, see Kager (1999: ch. 4).

(12)	a.	Troch	Feet are left-headed (trochaic).
	b.	FtBin	Feet are binary.
	c.	*Lapse	No sequences of two unfooted syllables.

We now display the tableaux. Example (13) shows that stress is correctly predicted to fall on the first syllable in disyllables, with no variation. Note also that ranking does not matter. The same winners would emerge under any ranking.

(13) Disyllables: No variation

/yapa/	Troch	FtBin	*Lapse
a. yapa			*!
b. ⇒ (yápa)			
c. ya(pá)		*!	
d. (yá)(pá)		*!*	
e. (yapá)	*!		

Leaving the word completely unfooted (13a) fatally violates *LAPSE; forming monosyllabic feet (13c–d) fatally violates FTBIN; forming a right-headed (iambic) foot (13e) fatally violates TROCH.

We now come to the interesting part: the variable cases.

(14) Trisyllables: Variation

/yutanti/	TROCH	FTBIN	*LAPSE
a. yutanti			*!*
b. ⇒ (yútan)ti			
c. ⇒ yu(tánti)			
d. yutan(tí)		*!	*
e. (yú)(tánti)		*!	
f. (yú)tan(tí)		*!*	
g. yu(tán)(tí)		*!*	
h. (yú)(tán)(tí)		*!**	
i. yu(tantí)	*!		

As (14) shows, the grammar correctly predicts that stress varies between the first and second syllables in three-syllable words. This is because there are two (and only two) acceptable ways of grouping syllables into feet. The constraints cannot tell the two candidates apart and both slip through, hence variation. As the reader can easily verify, the same variation pattern is also predicted in four-syllable words. The upshot is that the grammar correctly predicts the locus of variation.

Despite this success, the tied violations approach has two inherent limitations. First, truly identical violation profiles seem extremely difficult to achieve given two common assumptions: (1) all constraints are universal and thus present in all grammars; (2) constraints are sufficiently fine-grained to distinguish any two non-identical candidates. Thus, there will always be some low-ranking constraint that will distinguish between the purported optional variants, blocking variation. The above analysis works only because we have failed to include all the members of the universal constraint set in our tableau. Second, the approach provides no way of modeling degrees of variation, but reduces variation to optionality. Nevertheless, there is an interesting insight: the approach claims that variation involves candidates that are equally good in terms of some subset of high-ranking constraints. This idea reappears in subsequent work, as we will see shortly.

3.3 Pseudo-optionality

Another approach to variation is to attribute it to free choice between alternative inputs. Following Müller (1999), we call this approach PSEUDO-OPTIONALITY. Thus, in the case of *t,d*-deletion, we could posit two underlying forms: /cost/ and /cos/ which both mean "cost". The fact that a variant sometimes conveys information about the context of utterance, linguistic or otherwise, might be taken as an argument for pseudo-optionality. For example, in Colloquial Helsinki Finnish, we might posit two distinct lexical items with slightly different meanings:

(15) a. /makea/ "sweet"
b. /makee/ "sweet, colloquial, most typically uttered by a young working-class female"

The "uncoalesced" (15a) and the "coalesced" (15b) would thus be distinct inputs. We could now set up the grammar in such a way that the two variants would be the unique winners in their respective competitions, e.g. by ranking faithfulness reasonably high. The two variants would thus not compete in phonology at all, but at the point where lexical items are introduced into syntax. Interestingly, this seems to be the favored solution to optionality in optimality-theoretic syntax. Müller (1999) notes that proponents of pseudo-optionality often try to argue for subtle differences in meaning (i.e. different inputs) in cases like *A review of this article came out yesterday* and *A review came out yesterday of this article* (for criticism, see Newmeyer 2000). The reasoning is analogous.

The pseudo-optionality approach to variation is attractive because of its austerity: it proposes to derive variation from the unavoidable assumption that the speaker can select from among different inputs. However, as a general model of phonological variation it has a number of weaknesses. Essentially, pseudo-optionality proposes to reduce all apparent variation to linguistic free will, manifested for example in lexical alternations like *thirsty* vs. *hungry* in the frame *I'm* __ where the choice is clearly unpredictable from grammar-internal factors and for all intents and purposes also from grammar-external factors. As we have seen, this is not generally true of phonological variation: phonological (morphological, lexical, . . .) factors do emerge in variation, sometimes categorically, sometimes quantitatively. But if variation is independent of phonology and reducible to optionality at the point of lexical insertion, then it is not clear how such phonological effects can be modeled, or even how they are possible in the first place.

The argument from putative meaning differences is not convincing either. While it is true that in Colloquial Helsinki Finnish *makea* and *makee* have different connotations, so do hundreds of other word pairs, and in exactly the same way: the coalesced variant is always "more colloquial." Thus, the additional shade of meaning is clearly not a property of a specific lexical item, but rather, coalescence itself. In other words, the young working-class females have a "colloquial" phonology rather than a "colloquial" vocabulary.

We have now reviewed two optimality-theoretic approaches to variation. These approaches are attractive because they do not modify the standard assumptions of Optimality Theory in any way. At the same time, both have conceptual and empirical weaknesses. We will now turn to approaches that have given up one or more of the basic assumptions in order to account for variation, in particular its quantitative aspects.

4 The Multiple Grammars Model

The multiple grammars model proposes that variation arises from the competition of distinct grammatical systems within an individual. This view has been defended by e.g. Kiparsky (1993b) for phonology and Kroch (1989, 1994, 2001) for morphosyntax. That an individual can simultaneously possess several grammars is uncontroversial in the case of multilingualism. The question is whether multilingualism, multidialectalism, an individual's ability to switch among styles and registers, and ultimately, an individual's ability to involve in free variation are all fundamentally similar phenomena.

In optimality-theoretic terms, the multiple grammars model implies that a single individual commands a set of distinct total rankings of constraints, or tableaux for short. For example, an individual grammar might consist of the three tableaux in (16). As the reader can easily verify, grammar (16a) deletes the final *t* in /mɪst/, whereas both (16b) and (16c) retain it.

(16)	Input	Tableau	Output
	a. /mɪst/	Onset >> *CxCod >> Max(C)	[mɪs]
	b. /mɪst/	Onset >> Max(C) >> *CxCod	[mɪst]
	c. /mɪst/	Max(C) >> Onset >> *CxCod	[mɪst]

The simplest interpretation of the multiple grammars model is that any set of tableaux is a possible system for an individual. On each occasion of use, defined as e.g. word, utterance, etc., the speaker would reach into the grammar pool and select a tableau. In the absence of good reasons to think otherwise, it would be natural to assume that each tableau has the same probability of being selected. This does not imply that variation is free: the grammar pool available to the speaker may be biased in various ways. For example, in (16) the number of *t*-retention grammars is greater than the number of *t*-deletion grammars. Specifically, one might expect this speaker to delete *t* in *mist* approximately 33 percent of the time. We might even propose an explicit QUANTITATIVE INTERPRETATION of the multiple grammars model along these lines:

(17) *A quantitative interpretation of multiple grammars*. The number of grammars that generate a particular output is proportional to the relative frequency of this output.

There are two common objections to the multiple grammars model. First, it has been suggested that the number of grammars per individual becomes implausibly large (Reynolds 1994, Guy 1997b). Three grammars may not sound too bad, but how about an individual with 120 or 40,320 grammars? While initially persuasive, this objection is difficult to evaluate unless we have some way of determining how many grammars are too many. Second, the model seems unconstrained. If any combination of grammars is possible for an individual speaker, we run the risk that any kind of variation, including any kind of frequencies, can be modeled. This would imply that the theory does not exclude anything and is thus vacuous (Liberman 1994). For example, the *t*-retention bias in (16) follows from nothing in particular. In fact, we could have chosen any tableaux, with consequently different results. However, biases like (16) may have a principled basis. As we will see shortly, phonological and morphological constraints themselves introduce inherent biases into the system, setting firm limits on how many and what kinds of grammars can be constructed in the first place. In addition, external factors like sex, age, style, register and social class may be responsible for systematic differences among individuals. For example, old and young speakers may have systematically different grammar pools, although they may be overlapping.

We now exemplify the multiple grammars model by outlining the beginnings of an optimality-theoretic analysis of English *t,d*-deletion based on the fuller analysis in Kiparsky (1993b). The model predicts that certain types of variable dialects are possible, others impossible, and that certain types of statistical distributions of variants are possible, others impossible. Our goal is to show that the multiple grammars model is a serious hypothesis concerning variation, although not without problems.

Following the approach in Liberman (1994), let us first define a grammar as a set of input/output pairs where for every input there is some fixed output. We start by assuming a particularly simple grammar that does not impose any constraints on the possible input/output mappings, but accepts them all. While such a grammar is unlikely to do anything linguistically interesting, it will give us a baseline: any analysis that we may come up with should do better, and certainly not worse, than this linguistically naive grammar.

Next, we define the possible inputs and possible outputs in the domain of interest, which is *t,d*-deletion. Here we will only consider two inputs and three outputs. On the input side, *t* can occur either before a vowel or before a consonant. On the output side, we consider three options: *t* in the syllable coda, *t* resyllabified as part of the following onset and *t*-deletion.

(18) Inputs: /cost##V/, /cost##C/
Outputs: [cost.X], [cos.tX], [cos.X]

Next, we ask what are the possible input/output mappings, i.e. grammars. Since there are two inputs and three outputs, we have $3^2 = 9$ distinct grammars, shown in (19).

(19) The nine possible grammars (the naive model)

	/cost##V/	/cost##C/	
1	cost.V	cost.C	complex coda, complex coda
2	cost.V	cos.tC	complex coda, resyllabification
3	cost.V	cos.C	complex coda, *t*-deletion
4	cos.tV	cost.C	resyllabification, complex coda
5	cos.tV	cos.tC	resyllabification, resyllabification
6	cos.tV	cos.C	resyllabification, *t*-deletion
7	cos.V	cost.C	*t*-deletion, complex coda
8	cos.V	cos.tC	*t*-deletion, resyllabification
9	cos.V	cos.C	*t*-deletion, *t*-deletion

It is immediately obvious that some of these grammars are linguistically natural, others are not. For example, grammar 6 yields resyllabification pre-vocalically (*cos.tAnna*) and deletion pre-consonantally (*cos.me*), which seems plausible. On the other hand, grammar 8 yields the reverse pattern (*cos.Anna*, *cos.tme*) which is unheard of. In addition, given the multiple grammars model, we predict a speaker who combines grammars 4 and 6. Such an individual shows invariant resyllabification pre-vocalically (*cos.tAnna*) and optional deletion pre-consonantally (*cost.me~cos.me*), which is again plausible. On the other hand, a combination of 8 and 9 gives a dialect with invariant deletion pre-vocalically (*cos.Anna*) and variable deletion pre-consonantally (*cos.tme~cos.me*), which seems extremely odd. What we clearly need is a phonological theory that rules out the unnatural grammars and keeps the natural ones.

Following Kiparsky (1993b), let us now enrich our linguistically naive model by adopting some basic optimality-theoretic constraints on syllable structure. Three of these constraints have already been introduced. To simplify discussion, we substitute Max(t) for the more general Max(C).

(20)
*CxCod No coda clusters.
*CxOns No onset clusters.
Onset Onsets are required.
Max(t) No *t*-deletion.
Align Morpheme and syllable boundaries coincide.

We now ask the following question: given these five constraints, Optimality Theory and the multiple grammars model, what patterns are predicted? The answer could be one of the following:

(21) 1 We continue to predict all the nine logically possible patterns. This would show that our constraints and Optimality Theory serve no useful purpose at all.
2 Some of the nine logically possible patterns are excluded. Now the question is: which patterns? The best scenario is that our model rules out all the linguistically unnatural and hence unattested

patterns, and keeps all the natural patterns, including the attested ones. A worse scenario is that the resulting pattern is more or less random in which case we would not be doing any better than the naive model. The worst scenario is that our system rules out all the right patterns and keeps all the wrong ones.

In order to see what is being predicted, we must take all the possible tableaux and apply them to all the possible inputs. Five constraints can be ranked in 5! = 5 × 4 × 3 × 2 × 1 = 120 possible ways. This so-called FACTORIAL TYPOLOGY exhausts the space of grammatical possibilities. Recall that there are two inputs: /cost##V/ (the pre-vocalic environment) and /cost##C/ (the pre-consonantal environment). This means we will have to check 240 tableaux in all. To get started, we select one tableau at random and apply it to both inputs.

(22) A sample tableau. . = syllable boundary, | = morpheme boundary

/cost##V/	*CxOns	*CxCod	Onset	Align	Max(t)
1a. cost|.V		*!	*		
1b. cos|.V			*!		*
1c. ⇒ cos.t|V				*	
/cost##C/					
2a. cost|.C		*!			
2b. ⇒ cos|.C					*
2c. cos.t|C	*!			*	

This grammar resyllabifies *t* before a vowel (*cos.tAnna*) and deletes it before a consonant (*cos.me*). Looking back to (19), we see that this is a type 6 grammar, one of the linguistically plausible types. This is a good start, but since we are testing the limits of the model, we still need to check what happens under the remaining 119 rankings. This can be done either by brute force, i.e. by working through the tableaux one by one, either by hand or by computer (as has been done here), or by ingenious reasoning. The results are shown in (23).

(23) The five possible grammars (Optimality Theory)

	/cost##V/	/cost##C/	
1	cost.V	cost.C	complex coda, complex coda
4	cos.tV	cost.C	resyllabification, complex coda
5	cos.tV	cos.tC	resyllabification, resyllabification
6	cos.tV	cos.C	resyllabification, *t*-deletion
9	cos.V	cos.C	*t*-deletion, *t*-deletion

Instead of the logically possible nine output patterns, we only get five. This means that there is no way of ranking the five constraints such that we would get patterns 2, 3, 7 and 8. These patterns are thus excluded as impossible.[4] Taking a closer look at (23), we can see that all three outputs (complex coda, resyllabification, *t*-deletion) are predicted to be possible both pre-vocalically and pre-consonantally. However, the environments show clear preferences. Resyllabification is favored pre-vocalically, whereas complex coda and *t*-deletion are favored pre-consonantally. This is an example of an inherent bias introduced by the constraints. A more accurate picture is shown in (24) which sums up the number of tableaux predicting each output in each environment, both in actual numbers and percentages (out of 120 tableaux).

(24) The number of tableaux predicting each output

	V	C	V%	C%
resyllabification	70	20	58%	17%
complex coda	25	50	21%	42%
t-deletion	25	50	21%	42%

Assuming the quantitative interpretation in (17), we can take the percentages in (24) to represent the quantitative output of a speaker who possesses all the 120 grammars. This hypothetical speaker exhibits the well-documented vowel/consonant asymmetry, deleting *t,d* approximately 21 percent of the time before vowels and about 42 percent of the time before consonants. Different percentages can be obtained by throwing out grammars from the pool. For example, by throwing out all grammars except types 6 and 9 we get a dialect where *t*-deletion is categorical in the pre-consonantal environment, but optional in the pre-vocalic environment (50 percent resyllabification, 50 percent *t*-deletion), another dialect with a vowel/consonant asymmetry of the right kind.

A very important question is what kinds of dialects are excluded. Recall the bizarre dialects predicted by the naive model, for example dialect 8 with pre-vocalic deletion (*cos.Anna*) and pre-consonantal resyllabification (*cos.tme*). This dialect is now excluded, a good result. Also excluded is the dialect with invariant pre-vocalic deletion (*cos.Anna*) and variable pre-consonantal deletion (*cos.me~cost.me* or *cos.me~cos.tme*). To see why this should be, note that to get invariant pre-vocalic deletion we are limited to tableaux of type 9. But this means that we will also get invariant pre-consonantal deletion. Yet another excluded dialect is one where deletion is more frequent pre-vocalically than pre-consonantally. In conclusion, a very simple analysis that assumes five phonological constraints, Optimality Theory, and the multiple grammars model excludes several unattested *t,d*-deletion patterns, among them unattested quantitative patterns. This is a promising initial result.

However, some odd predictions persist. For example, the model predicts variable dialects without a consonant/vowel asymmetry. Such a dialect can be obtained by combining tableaux of types 1, 5 and 9, and those only. Such a speaker will vary between resyllabification, complex coda and *t*-deletion, but show no quantitative differences between the vowel and consonant environments. Dialects of this kind do not seem to exist. There are at least three possible diagnoses: (1) not all relevant constraints have been included in the analysis; (2) the phonology of English contains rankings that are responsible for the robust vowel/consonant asymmetry; (3) the multiple grammars model is false. Here it suffices to note that there are clearly many constraints that matter to *t*-deletion and that have not been included, for example constraints referring to segment quality. In addition, the phonology of English is likely to have at least some partial rankings (see sections 5.2 and 5.3) that restrict the pool of available grammars. Thus, concluding (3) is clearly premature.

Finally, the question arises how to incorporate morphological effects in the analysis. Kiparsky's (1993b) solution is to posit morphologically specialized phonological constraints that check syllable well-formedness separately for roots, stems and words. This puts *cost*, *los+t* and *toss#ed* in unequal positions as the schematic tableau (25) shows.

(25)

	*CxCod$_{root}$	*CxCod$_{stem}$	*CxCod$_{word}$
1a. cost	*	*	*
1b. cos			
2a. los+t		*	*
2b. los			
3a. toss#t			*
3b. toss			

The regularly inflected *tossed* only violates the word-level constraint, the semi-regular *lost* violates both stem-level and word-level constraints, and the monomorphemic *cost* violates all three. Interspersed among the five constraints discussed so far, these new constraints will bias the output pattern in a particular way. The overall effect is easy to see even without working out the tableaux: all else being equal, there will be more tableaux that delete *t* in *cost* than in *lost*, and more tableaux that delete *t* in *lost* than in *tossed*, the desired result.

In sum, we have presented a multiple grammars approach to variation within Optimality Theory and exemplified it by outlining a possible analysis of *t,d*-deletion along the lines of Kiparsky (1993b). We have seen that the model yields falsifiable, and mostly reasonable, predictions both in the categorical and quantitative domains. Certain types of variable dialects are predicted to be possible, others impossible; certain types of statistical distributions of variants are predicted to be possible, others impossible. The limits of the model can be explored further by applying it to specific *t,d*-deletion dialects. Empirically the model needs to be augmented to cover the effects of segment quality, the

Obligatory Contour Principle, and the pause environment (for the treatment of pause, see Kiparsky 1993b), both in their categorical and quantitative manifestations. All this seems straightforward in principle, but detailed analyses may prove otherwise.

An interesting general aspect of the model is that it explicitly identifies cross-linguistic variation and intra-individual variation: both reduce to differences in constraint ranking. In this particular sense, intra-individual variation is cross-linguistic variation within an individual. One would thus expect to find the same regularities in both domains. Thus, for example, cross-linguistic variation in syllable structure parameters should emerge intra-individually in the relevant types of variation, perhaps in a statistical form. This is indeed what seems to happen in the case of *t,d*-deletion.

To refute the multiple grammars model, one can proceed in two ways:

(26) a. Show that the model is too powerful. This would involve a demonstration that the model predicts systematically unattested types of variation, i.e. the model overgenerates and a more restrictive theory is needed. Alternatively, one could adopt a less powerful model (see e.g. section 5.3) and show that it is sufficient for all the attested cases of variation.

b. Show that the model is too weak. This would involve documenting cases of variation that cannot be captured in the multiple grammars model under reasonable linguistic assumptions, i.e. the model undergenerates and something more powerful is needed.

Since most optimality-theoretic work on variation has so far concentrated on getting the descriptive facts right, one way or the other, very little attention has been paid to general theory comparison issues such as (26a) and (26b). However, both kinds of arguments are already on record. In the rest of this chapter, we will review two analyses of the same corpus of data: 28,000 tokens of variable genitive plurals in Finnish. The first analysis (Anttila 1997a) adopts multiple tableaux, but places a particular restriction on their possible combinations. The second analysis (Boersma and Hayes 2001) introduces numerical ranking in order to better capture quantitative distinctions in the data.

5 Alternatives to Multiple Grammars

5.1 The data

The Finnish genitive plural can be formed in two principal ways which we will here call the WEAK ENDING /-ien/ (also /-jen/) and the STRONG ENDING /-iden/. We will now examine the distribution of the two types of endings in vowel-final stems. (Consonant-final stems are systematically different and will not be discussed.)

(27) Genitive plural inflection in Finnish

	Stem	Variants	Gloss
a.	/puu/	pu-iden	"tree-PL-GEN"
	/maa/	ma-iden	"land-PL-GEN"
b.	/lasi/	las-ien	"glass-PL-GEN"
	/margariini/	margariin-ien	"margarine-PL-GEN"
	/sosialisti/	sosialist-ien	"socialist-PL-GEN"
c.	/naapuri/	naapure-iden~naapur-ien	"neighbor-PL-GEN"
	/ministeri/	ministere-iden~minister-ien	"minister-PL-GEN"
	/aleksanteri/	aleksantere-iden~aleksanter-ien	"Alexander-PL-GEN"

The stems in (27a) choose the strong variant, the stems in (27b) the weak variant, and the stems in (27c) are variable. This raises the locus of variation problem: why does variation only occur in the last group? An important observation is that the first group contains all the monosyllabic stems, the second group all the disyllabic stems (as well as certain longer stems, e.g. /margariini/ "margarine" and /sosialisti/ "socialist") and the variable stems may have any number of syllables greater than two. This shows that word length in syllables is somehow involved, reminding us of the Walmatjari pattern (section 3.2).

In addition, variable stems are variable to different degrees. Native speakers often report that one variant sounds better than the other, although both are possible. These subtle differences in judgments are reflected in corpus frequencies: the better-sounding variant tends to be more common. The following numbers are based on the 1987 issues of the *Suomen Kuvalehti* weekly (approximately 28,000 genitive plurals).[5]

(28)

	Stem type	Strong	Weak
a.	/ka.me.ra/	ka.me.roi.den (99.4%)	?ka.me.ro.jen (0.6%)
b.	/sai.raa.la/	sai.raa.loi.den (50.5%)	sai.raa.lo.jen (49.5%)
c.	/naa.pu.ri/	naa.pu.rei.den (37.2%)	naa.pu.ri.en (62.8%)
d.	/po.lii.si/	?po.lii.sei.den (1.4%)	po.lii.si.en (98.6%)

Glosses: *kamera* "camera", *sairaala* "hospital", *naapuri* "neighbor", *poliisi* "police"

Two phonological generalizations emerge: (1) stems ending in an underlyingly [+low] vowel /a/ prefer the strong variant; stems ending in an underlyingly [+high] vowel /i/ prefer the weak variant (Anttila 1997a); (2) Stems with a light penultimate syllable (CV) prefer the strong variant; stems with a heavy penultimate syllable (CVV) prefer the weak variant (Itkonen 1957).[6] These tendencies have a cumulative effect: /ka.me.ra/ (final /a/, light penult) favors the strong variant almost categorically, and /po.lii.si/ (final /i/, heavy penult) favors the weak variant almost categorically. The mixed cases /sai.raa.la/ and /naa.pu.ri/ fall in between.

To summarize, the locus of variation depends on the number of syllables in the stem; the degree of variation depends on the height of the final vowel and the weight of the penultimate syllable of the stem. The question is: why would such things matter? We suggest that they do because the shapes of the variants are different: the strong variant /-iden/ makes the preceding syllable heavy, e.g. *naa.pu.***rei***.den* whereas the weak variant /-ien/ makes the preceding syllable light, e.g. *naa.pu.***ri***.en*. While this may seem to be yet another puzzling observation, this is the crucial phonological fact that makes everything fall into place.

5.2 The constraints

We now introduce the constraints. For a more detailed discussion, see Anttila (1997a). First, heavy syllables tend to be stressed and stressed syllables tend to be heavy universally (see e.g. Kager 1999: 155, 268). This is expressed as the constraint hierarchy in (29). The constraints are stated negatively, e.g. *′L means "Avoid a stressed light syllable."

(29) The Weight-Stress Principle
{*H, *′L} >> {*′H, *L}
"Unstressed heavies and stressed lights are worse than stressed heavies and unstressed lights."

Second, we propose that stress, syllable weight and vowel height are preferably combined as stated in (30)–(31):

(30) The Weight-Sonority Principles
a. *H/I >> *H/O >> *H/A
b. *L/A >> *L/O >> *L/I
"A heavy syllable with a high nucleus is worse than a heavy syllable with a mid nucleus, which is worse than a heavy syllable with a low nucleus. The reverse holds for light syllables."

(31) The Stress-Sonority Principles
a. *′I >> *′O >> *′A
b. *A >> *O >> *I
"A stressed syllable with a high nucleus is worse than a stressed syllable with a mid nucleus, which is worse than a stressed syllable with a low nucleus. The reverse holds for unstressed syllables."

Finally, we propose the following rhythmic principles:

(32) The Rhythmic Principles

a.	*′X.′X	"No adjacent stressed syllables."
	*X.X	"No adjacent unstressed syllables."
b.	*H.H	"No adjacent heavy syllables."
	*L.L	"No adjacent light syllables."

5.3 *Analysis 1: Stratified grammar*

First, we address the locus of variation problem. There are two separate questions here: (1) Why do monosyllabic and disyllabic stems choose different variants? (2) Why does variation require a minimum of three syllables? We start by noting that in Finnish words the first syllable is always stressed, the second always unstressed. This follows from two undominated constraints: InitialStress (not mentioned in the tableaux) and *ʹX.ʹX "No adjacent stressed syllables". The correct distribution of variants now follows by ranking the Weight-Stress constraints *H and *ʹL as in (33). Note that these two constraints need not be ranked with respect to each other: both rankings (Tableau 1, Tableau 2) yield the same winner.

(33) /maa/ "land"

Tableau 1		*X́.X́	*Ĺ	*H	. . .
a.	⇒ mái.den			*	
b.	*má.jen		*!	*	

Tableau 2		*X́.X́	*H	*Ĺ	. . .
a.	⇒ mái.den		*		
b.	*má.jen		*	*!	

Since the first syllable is always stressed, and stressed syllables are preferably heavy, the strong variant with the heavy penult invariably wins. In disyllables the reverse situation obtains, as shown in (34): since the second syllable is always unstressed and unstressed syllables are preferably light, the weak variant with the light penult invariably wins.

(34) /lasi/ "glass"

Tableau 1		*X́.X́	*Ĺ	*H	. . .
a.	*lá.sei.den		*	**!	
b.	⇒ lá.si.en		*	*	

Tableau 2		*X́.X́	*H	*Ĺ	. . .
a.	*lá.sei.den		**!	*	
b.	⇒ lá.si.en		*	*	

We are here making the idealizing assumption that the final syllable is never stressed, implying an undominated constraint *FinalStress. In fact, final syllables may be optionally stressed if heavy (cf. Hanson and Kiparsky 1996). Here we abstract away from this additional variable pattern.

We now understand why monosyllabic and disyllabic stems show no variation: this is an effect of the invariant initial stress. However, nothing has been said about stress beyond the second syllable. This has the consequence in (35):

(35) /naapuri/ "neighbor"

Tableau 1		*X́.X́	*Ĺ	*H	...
a.	⇒ náa.pu.rèi.den			*	
b.	⇒ náa.pu.ri.en			*	
Tableau 2		*X́.X́	*H	*Ĺ	...
a.	⇒ náa.pu.rèi.den		*		
b.	⇒ náa.pu.ri.en		*		

As (35) shows, the third syllable may be stressed, in which case we get the strong variant, or it may be unstressed, in which case we get the weak variant.[7] This explains why variation requires a minimum of three syllables: after the second syllable stress is essentially optional. However, not all long words show variation. As the reader can easily verify, the analysis predicts variation in /mi.nis.te.ri/ "minister" (*minister-ien~ministere-iden*) and /a.lek.san.te.ri/ "Alexander" (*aleksanter-ien~aleksantere-iden*), but not in the very similar /mar.ga.rii.ni/ "margarine" (*margariin-ien/*margariine-iden*) and /so.si.a.lis.ti/ "socialist" (*sosialist-ien/*sosialiste-iden*), which is correct.

The remaining question is how to predict degrees of variation in cases like /naapuri/ "neighbor". For this purpose, we place another set of constraints below the three constraints discussed so far. This is shown in (36). Note that the universal rankings proposed in section 5.2 are respected; we are simply adding rankings.

(36) A grammar for Finnish: augmented version
*′X. ′X >> {*H, *′L} >> {*L.L, *H/I, *′I}

The new set contains three mutually unranked constraints. They can be spelled out as the six tableaux in (37). However, unlike (33)–(35), where different rankings produce the same winner, here different rankings produce different winners:

(37) /naapuri/ "neighbor"

Tableau 1		...	*H/I	*Í	*L.L
a.	náa.pu.rèi.den		*!	*	
b.	⇒ náa.pu.ri.en				*
Tableau 2		...	*H/I	*L.L	*Í
a.	náa.pu.rèi.den		*!		*
b.	⇒ náa.pu.ri.en			*	
Tableau 3		...	*Í	*L.L	*H/I
a.	náa.pu.rèi.den		*!		*
b.	⇒ náa.pu.ri.en			*	
Tableau 4		...	*Í	*H/I	*L.L
a.	náa.pu.rèi.den		*!	*	
b.	⇒ náa.pu.ri.en				*
Tableau 5		...	*L.L	*H/I	*Í
a.	⇒ náa.pu.rèi.den			*	*
b.	náa.pu.ri.en		*!		
Tableau 6		...	*L.L	*Í	*H/I
a.	⇒ náa.pu.rèi.den			*	*
b.	náa.pu.ri.en		*!		

The candidate *naapur-ien* wins in four tableaux, the candidate *naapure-iden* in two tableaux. Assuming the quantitative interpretation in (17), the predicted frequencies are 2/3 and 1/3, respectively. This is reasonably close to the actually observed frequencies 62.8 percent and 37.2 percent.

The grammar in (36) consists of internally unranked strata of constraints which are strictly ranked with respect to each other. Following the terminology of Tesar and Smolensky (1995) and Boersma (2001), we call such grammars STRATIFIED GRAMMARS. The complete stratified grammar needed to account for the Finnish facts is given in (38). The quantitative results will be displayed in section 5.5.

(38) A grammar for Finnish: final version

{*'X. 'X} >>	Stratum 1
{*H, *'L} >>	Stratum 2
{*L.L, *H/I, *'I} >>	Stratum 3
{*L/A, *H/O, *'O, *H.H, *X.X, *'H}	Stratum 4

A stratified grammar like (38) can be viewed from two angles. If we look at (38) as a relation, i.e. as a set of ordered pairs of constraints, we have a single grammar, not multiple grammars. Alternatively, if we spell out all the possible linearizations of the constraints in (38), we have multiple tableaux. From the second perspective it is clear that a stratified grammar is a special case of the multiple grammars model. Thus, all stratified grammars are instances of the multiple grammars model, but the converse does not hold: there are instances of the multiple grammars model that are not stratified grammars. For example, the tableaux in (39) constitute a perfectly well-formed grammar in the multiple grammars model, but not a stratified grammar.

(39) *'X.'X >> *H >> *'L
*'L >> *H >> *'X.'X

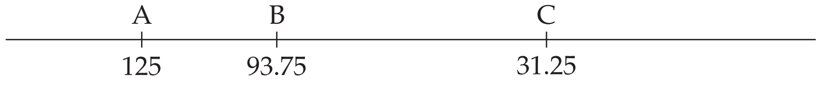
Other special cases of the multiple grammars model we have not discussed for reasons of space include FLOATING CONSTRAINTS (Reynolds 1994, Nagy and Reynolds 1997) and PARTIALLY ORDERED GRAMMARS (Anttila and Cho 1998, Anttila to appear a, Anttila to appear b). The latter subsume stratified grammars as a special case, as pointed out by Boersma (2001).

It is not clear which of these approaches is correct, if any. It may turn out that some restricted version of the multiple grammars model (e.g. stratified grammars) is sufficient. It may also be that the full power of multiple grammars is necessary. Finally, it is possible that something quite different is needed. This claim has been made by proponents of CONTINUOUS RANKING which generalizes Optimality Theory itself by introducing numerical ranking.

5.4 Analysis 2: Continuous ranking

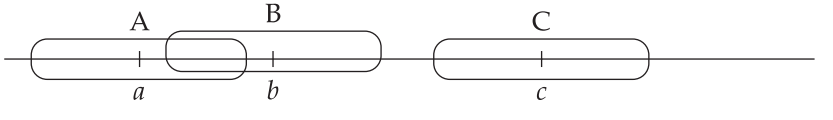
In standard Optimality Theory, the ranking A >> B >> C means that A dominates B which dominates C. There is no sense in which A and B could lie "closer together" than B and C. In the continuous ranking model (Zubritskaya 1997, Hayes 2000, Hayes and MacEachern 1998, Boersma 1998 and especially Boersma and Hayes 2001 which we will follow here) such a statement is meaningful. This model generalizes standard Optimality Theory by adopting a CONTINUOUS RANKING SCALE. Each constraint has a fixed RANKING VALUE along a real-number scale where higher values correspond to higher-ranked constraints. For example, the ranking values of A, B, and C could be 125, 93.75, and 31.25, respectively, as shown in (40). How ranking values are determined will be discussed shortly.

(40) Ranking values

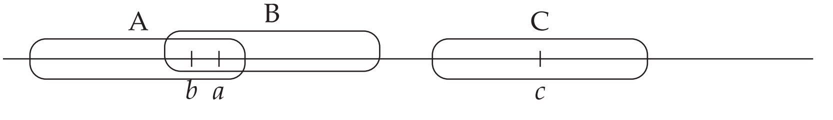


In order to describe variation, the model introduces STOCHASTIC CANDIDATE EVALUATION. Boersma and Hayes (2001) propose that, at evaluation time (i.e. the moment of speaking), a random positive or negative value is temporarily added to the ranking value of each constraint. The resulting actual ranking value is called the SELECTION POINT. As the selection points will vary around the fixed ranking value from evaluation to evaluation, the constraints begin to act as if they were associated with ranges of values, instead of points. This means that constraint ranges may overlap to different degrees. Consider the hypothetical examples in (41) and (42). A, B, and C are constraints with ranking values 125, 93.75, and 31.25, respectively, and *a*, *b*, and *c* are their respective selection points in a particular evaluation.

(41) Evaluation 1: common result



(42) Evaluation 2: rare result



In evaluation (41), the selection points happen to fall roughly in the middle of each constraint which results in the ranking *a* >> *b* >> *c*. In evaluation (42), *a* happens to be chosen very near the bottom of A and *b* very near the top of B which results in the ranking *b* >> *a* >> *c*. In the continuous ranking model, such differences between evaluations are responsible for variation. Boersma and Hayes (2001) explicitly assume that selection points are distributed normally around the ranking value, which means that they tend to cluster near the center of the constraint, but will occasionally fall near the edges. They also assume that every constraint has the same standard deviation, or equal "breadth." The degree of variation will depend on how close the fixed ranking values are to each other. Thus, the ranking *a* >> *b* will be common, *b* >> *a* rare, and C is too far away for *c* to ever rise above either *a* or *b*, i.e. the ranking is categorical.[8]

An advantage that continuous ranking currently enjoys over various other models is that there exists an algorithm, the GRADUAL LEARNING ALGORITHM, that is able to learn continuously ranked grammars, including variation.[9] The algorithm has been implemented as a computer program and it seems to perform well in empirical tests. This is also good news for stratified grammars: as Boersma (2001) notes, stratified grammars are a special case of continuously ranking grammars. The algorithm works roughly as follows (for a more detailed description, see Boersma and Hayes 2001). The algorithm starts out with a set of constraints with arbitrary (e.g. identical) ranking values. It reads a learning datum that consists of a surface form and its corresponding underlying form and checks whether the grammar is able to generate this surface form from the given underlying form. If not, the ranking must be incorrect, and the algorithm responds by promoting and demoting constraints by small amounts along the continuous ranking scale in a way that is likely to improve the situation. Given the adjusted grammar and further learning data, the algorithm will keep repeating these steps. If the algorithm behaves as desired, the end result is a ranking that correctly describes the data, including the quantitative preferences observed during learning.

To test the performance of continuous ranking and the Gradual Learning Algorithm, Boersma and Hayes replicated Anttila's (1997a) study of the Finnish genitive plural based on the same 28,000-token corpus. In this experiment, each variant was presented to the algorithm in the relative frequency in which it occurs in the corpus. In a representative run, the following ranking values emerged:

(43)

*H	288.000	*O	196.754
*ʹI	207.892	*X.X	188.726
*L.L	206.428	*ʹO	3.246
*A	199.864	*ʹA	0.136
*H.H	199.274	*I	–7.892

5.5 A comparison of results

The tables in (44)–(47), reproduced from Boersma and Hayes (2001), contrast the performance of the two models. SG stands for stratified grammar, CRG for continuously ranking grammar.

(44) 2-syllable stems

Example	Candidates	OBS.%	SG	CRG	Tokens
kala	ká.lo.jen	100	100	100	500
'fish'	ká.loi.den	0	0	0	0
lasi	lá.si.en	100	100	100	500
'glass'	lá.sei.den	0	0	0	0

(45) 3-syllable stems

Example	Candidates	Obs.%	SG	CRG	Tokens
kamera	ká.me.ro.jen	0	0	0.52	0
'camera'	ká.me.ròi.den	100	100	99.48	720
hetero	hé.te.ro.jen	0.5	0	0.57	2
'hetero'	hé.te.ròi.den	99.5	100	99.43	89
naapuri	náa.pu.ri.en	63.1	67	69.51	368
'neighbor'	náa.pu.rèi.den	36.9	33	30.49	215
maailma	máa.il.mo.jen	49.5	50	42.03	45
'world'	máa.il.mòi.den	50.5	50	57.97	46
korjaamo	kór.jaa.mo.jen	82.2	80	81.61	350
'repair shop'	kór.jaa.mòi.den	17.8	20	18.39	76
poliisi	pó.lii.si.en	98.4	100	100	806
'police'	pó.lii.sèi.den	1.6	0	0	13

(46) 4-syllable stems

Example	Candidates	Obs.%	SG	CRG	Tokens
taiteilija	tái.tei.li.jo.jen	0	0	0.52	0
'artist'	tái.tei.li.jòi.den	100	100	99.48	276
luettelo	lú.et.te.lo.jen	0	0	0.56	0
'catalogue'	lú.et.te.lòi.den	100	100	99.44	25
ministeri	mí.nis.te.ri.en	85.7	67	69.49	234
'minister'	mí.nis.te.rèi.den	14.3	33	30.51	39
luon.neh.din.ta	lúon.neh.dìn.to.jen	100	100	100	1
'charaterization'	lúon.neh.dìn.noi.den	0	0	0	0
edustusto	é.dus.tùs.to.jen	100	100	100	84
'representation'	é.dus.tùs.toi.den	0	0	0	0
margariini	már.ga.rìi.ni.en	100	100	100	736
'margarine'	már.ga.rìi.nei.den	0	0	0	0

(47) 5-syllable stems

Example	Candidates	Obs.%	SG	CRG	Tokens
ajattelija	á.jat.te.li.jo.jen	0	0	0.52	0
'thinker'	á.jat.te.li.jòi.den	100	100	99.48	101
televisio	té.le.vi.si.o.jen	0	0	0.57	0
'television'	té.le.vi.si.òi.den	100	100	99.43	41
Aleksanteri	á.lek.sàn.te.ri.en	88.2	67	69.51	15
'Alexander'	á.lek.sàn.te.rèi.den	11.8	33	30.49	2
evankelista	é.van.ke.lìs.to.jen	100	100	100	2
'evangelist'	é.van.ke.lìs.toi.den	0	0	0	0
italiaano	í.ta.li.àa.no.jen	100	100	100	1
'Italian'	í.ta.li.àa.noi.den	0	0	0	0
sosialisti	só.si.a.lìs.ti.en	100	100	100	99
'socialist'	só.si.a.lìs.tei.den	0	0	0	0

Example	Candidates	Obs.%	SG	CRG	Tokens
koordinaatisto	kóor.di.nàa.tis.to.jen	80	80	81.61	8
'coordinate grid'	kóor.di.nàa.tis.tòi.den	20	20	18.39	2
avantgardisti	á.vant.gàr.dis.ti.en	100	100	100	2
'avant-gardist'	á.vant.gàr.dis.tèi.den	0	0	0	0

Boersma and Hayes conclude that both models predict the empirical frequencies fairly well. The mean absolute error for the percentage predictions is 2.2 percent for the stratified grammar and 2.53 percent for the machine-learned continuously ranking grammar, averaged over 100 runs. Since stratified grammars are less powerful than continuously ranking grammars, the result is very encouraging for stratified grammars. However, more case studies are needed to evaluate the empirical adequacy of the two models in different domains. The question whether standard Optimality Theory needs to be replaced by numerical ranking still remains open.

6 Conclusion

In this chapter, I have discussed the following optimality-theoretic approaches to variation:

(48)
1. Tied violations
2. Pseudo-optionality
3. Multiple grammars (the generic model)
4. Stratified grammars
5. Continuously ranking grammars

Tied violations and pseudo-optionality are theoretically the most conservative approaches: they do not assume anything beyond standard Optimality Theory. However, neither seems capable of accommodating the kinds of data variationist linguists are used to seeing in their daily work. The multiple grammars model goes further. In particular, there are obvious ways of extending this model to quantitative regularities. Here the problem takes on the opposite character: unless constrained in principled ways, the multiple grammars model may be able to accommodate any kind of variation, linguistically natural or unnatural, a theoretically bad result. Approaches like stratified grammars address this concern by proposing restrictions on possible ranking relations. Not enough work has been done to determine how well motivated the proposed restrictions are empirically and whether they make the right kinds of predictions beyond the particular analyses at hand. Finally, the continuous ranking hypothesis challenges one of the basic assumptions of Optimality Theory itself by adopting real-number weighting. While this added power is a descriptive advantage, it remains to be seen whether it is necessary and whether it has undesirable consequences in other domains.

All the approaches, except pseudo-optionality, take essentially the same view on locus of variation and the role of markedness: variation occurs in environments where the constraints are unable to distinguish between candidates that are almost equally good in terms of markedness and faithfulness. It is important to see that, even in the absence of quantitative predictions, an analysis that correctly predicts the locus of variation has made a major contribution. This step is usually taken for granted in quantitative analyses of variation.

The proposals differ in how they approach degrees of variation. Tied violations and pseudo-optionality do not address the problem at all. The multiple grammars model and its derivatives (e.g. stratified grammars) usually rely on some version of grammar-counting, although this is not a necessary feature of the approach, and there are many conceivable ways of linking abstract grammatical structure with concrete corpus frequencies and/or gradient well-formedness judgments. As always, it is not obvious how to cross the competence/performance divide. Finally, continuous ranking is clearly motivated by the desire to enhance the power of Optimality Theory in the quantitative direction. This is also a domain where the competing proposals seem fairly easy to compare in concrete terms.

Less work has been done on the problem of interfaces: how do phonological, morphological, and lexical information interact in variation? In addition to morphologically specialized phonological constraints, such as *CxCod$_{\text{ROOT}}$, *CxCod$_{\text{STEM}}$, and *CxCod$_{\text{WORD}}$, there is the possibility that different morphological and lexical categories subscribe to (slightly) different rankings; for general discussion, see e.g. Itô and Mester (1995a, 1995b, 1998), Orgun (1996), Inkelas (1998, 1999), Anttila (to appear a, to appear b), Kiparsky (to appear). The morphological and lexical aspects of variation are an interesting general challenge for all these models and are likely to be useful in distinguishing them empirically. Relatively little work has been done on the interface of grammar and external factors; however, see Oostendorp (1997) and Morris (1998) for analyses of stylistic variation. There is a steadily growing literature on Optimality Theory and language change, often explicitly related to variation. A useful starting point is the bibliography compiled by Gess (2000).

While the work reviewed here is still in its beginnings, it seems fair to conclude that Optimality Theory has opened up a new perspective for the study of phonological variation. A number of technical proposals have been put forward, some initial results have been obtained, and a concrete research agenda is taking shape. The proposed models remain to be evaluated both formally, i.e. by working out the genuine (as opposed to notational) similarities and differences between the models, and empirically, i.e. by investigating many more cases of variation in many more languages, with the specific goal of finding evidence that would decide between the models. The initial empirical success of Optimality Theory gives one hope that generative phonology is beginning to answer some of the empirical questions raised by variationist linguists. It should also encourage theoretical phonologists to pay close attention to variable (including quantitative) data since such data are likely to have implications beyond variation.

Appendix

OT analyses of phonological variation

Anttila (1997a), Anttila (1997b), Anttila (to appear a, to appear b), Anttila and Cho (1998), Anttila and Revithiadou (2000), Auger (2000), Boersma (1997, 1998), Boersma and Hayes (2001), Borowsky and Horvath (1997), Demuth (1997), Hammond (1994), Hayes (2000), Hayes and MacEachern (1998), Hinskens et al. (1997), Holt (1997), Itô and Mester (1997), Iverson and Lee (1994), Kager (1996), Kang (1996), Kiparsky (1993b, to appear), Liberman (1994), Morris (1998), Nagy and Reynolds (1997), Oostendorp (1997), Reynolds (1994), Ringen and Heinämäki (1999), Rose (1997), Zubritskaya (1997).

Of related interest

Bailey (1973), Berkley (1994), Cedergren and Sankoff (1974), Frisch (1996), Frisch et al. (1997), Guy (1991a, 1991b, 1997a, 1997b), Guy and Boberg (1997), Labov (1969, 1972, 1994), Müller (1999), Myers (1995), Myers and Guy (1997), Paolillo (2000), Pierrehumbert (1994), Vennemann (1988), Weinreich et al. (1968).

NOTES

1 This paper was written while I was a Faculty of Arts and Social Sciences Postdoctoral Fellow at the Department of English Language and Literature, National University of Singapore. I thank Jack Chambers, Vivienne Fong and Natalie Schilling-Estes for valuable comments.

2 The corpus is available on the University of Helsinki Language Corpus Server at http://www.ling.helsinki.fi/uhlcs/.

3 For reasons of space, we will not consider the following additional facts: (1) besides variable stress, four-syllable words show two fixed stress patterns: initial stress (*tjí.ni.njà.ṛa* "midday"), a pattern limited to monomorphemic words, and second syllable stress (*tja*: *ḷá.na.na* "dispersing"); (2) five-syllable and six-syllable words also show variation, but only in secondary stress placement. See Hammond (1994) for an analysis.

4 As always, we need to add the caveat "all else being equal." By adding a linguistically ill-motivated constraint into the system, this prediction might no longer hold.

5 The corpus is available on the University of Helsinki

Language Corpus Server at http://www.ling.helsinki.fi/uhlcs/.

6 In Finnish (C)VV and (C)VC count as heavy syllables for the purposes of stress assignment, whereas (C)V counts as light.

7 The system predicts that secondary stresses only occur on heavy syllables. It must be duly noted that secondary stresses on light syllables have been reported, see e.g. Hanson and Kiparsky (1996), Elenbaas (1999), Elenbaas and Kager (1999). Thus, the present analysis only captures a subset of actual secondary stresses. The fact that these light syllable secondary stresses do not play a role in allomorph selection may imply that they have a postlexical origin.

8 This statement is not exactly true, but it is close enough. See Boersma and Hayes (2001) for discussion.

9 See http://www.fon.hum.uva.nl/praat/.

REFERENCES

Anttila, Arto (1997a). Deriving variation from grammar. In Frans Hinskens, Roeland van Hout and Leo Wetzels (eds.), *Variation, Change and Phonological Theory*, Amsterdam/Philadelphia: John Benjamins Publishing Company. 35–68. Also: http://www.roa.rutgers.edu, ROA-63.

Anttila, Arto (1997b). *Variation in Finnish phonology and morphology*. Doctoral dissertation, Stanford University, Stanford, California.

Anttila, Arto (to appear a). Derived environment effects in Colloquial Helsinki Finnish. Stanford, California: CSLI Publications. Also http://www.roa.rutgers.edu, ROA-406-08100.

Anttila, Arto (to appear b). Morphologically conditioned phonological alternations. *Natural Language and Linguistic Theory*. Also http://www.roa.rutgers.edu, ROA-425-10100.

Anttila, Arto and Young-mee Cho. (1998). Variation and change in Optimality Theory. *Lingua* 104: 31–56. Special issue on Optimality Theory.

Anttila, Arto and Anthi Revithiadou (2000). Variation in Allomorph Selection. *Proceedings of NELS 30*.

Anttila, Raimo (1989) *Historical and Comparative Linguistics*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Archangeli, Diana and D. Terence Langendoen (eds.) (1997). *Optimality Theory: An Overview*. Oxford: Blackwell.

Auger, Julie (2000). Phonological Variation and Optimality Theory: Evidence from word-initial vowel epenthesis in Picard. In Julie Auger and Andrea Word-Allbritton (eds.), *The CVC of Sociolinguistics: Contact, Variation, and Culture*. Indiana University Working Papers in Linguistics, vol. 2, pp. 1–20.

Bailey, Charles-James N. (1973). *Variation and Linguistic Theory*. Arlington, Virginia: Center for Applied Linguistics.

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: CLS. 1–12.

Bickerton, Derek (1971). Inherent variability and variable rules. *Foundations of Language* 7: 457–92.

Boersma, Paul (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* 21. 43–58. Also http://roa.rutgers.edu, ROA-221-1097.

Boersma, Paul (1998). Functional phonology. Doctoral dissertation, University of Amsterdam. The Hague: Holland Academic Graphics. Also: http://uvafon.hum.uva.nl/paul/.

Boersma, Paul (2001). Review of Arto Anttila (1997): Variation in Finnish Phonology and Morphology. *Glot International*. 5(1): 33–40. Also http://uvafon.hum.uva.nl/paul/.

Boersma, Paul and Bruce Hayes (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32(1): 45–86. Also http://www.roa.rutgers.edu, ROA-348-1099.

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*. Amsterdam/Philadelphia: John Benjamins Publishing Company. 101–23.

Cedergren, Henrietta J. and David Sankoff (1974). Variable rules: Performance as a statistical reflection of competence. *Language* 50: 333–55.

Chambers, J. K. (1995). *Sociolinguistic Theory*. Oxford, UK, and Cambridge, USA: Blackwell.

Chomsky, Noam and Morris Halle (1968). *The Sound Pattern of English*. Cambridge, Massachusetts: MIT Press.

Demuth, Katherine (1997). Multiple optimal outputs in acquisition. In Viola Miglio and Bruce Morén (eds.), *University of Maryland Working Papers in Linguistics* 5, College Park, MD: Linguistics Department, University of Maryland. 53–71.

Elenbaas, Nine (1999). *Foot typology and ternary stress systems*. Doctoral dissertation, Utrecht Institute of Linguistics, OTS. The Hague: Holland Academic Graphics. Also: http://www.roa.rutgers.edu, ROA-397-06100.

Elenbaas, Nine and René Kager (1999). Ternary rhythm and the lapse constraint. *Phonology* 16(3): 273–329.

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

Fasold, Ralph (1991). The quiet demise of variable rules. *American Speech* 66: 3–21.

Frisch, Stefan (1996). *Similarity and frequency in phonology*. Doctoral dissertation, Northwestern University. http://www.roa.rutgers.edu, ROA 198-0597.

Frisch, Stefan, Michael Broe and Janet Pierrehumbert (1997). Similarity and phonotactics in Arabic. http://www.roa.rutgers.edu, ROA 223–1097.

Gess, Randall (2000). Bibliography: Optimality Theory and Language Change. http://www.roa.rutgers.edu, ROA-359-1199.

Goldsmith, John (1976). *Autosegmental phonology*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Guy, Gregory and Sally Boyd (1990). The development of a morphological class. *Language Variation and Change* 2: 1–18.

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*, New York: Academic Press. 1–36.

Guy, Gregory R. (1991a). Contextual conditioning in variable lexical phonology. *Language Variation and Change* 3: 223–39.
Guy, Gregory R. (1991b). Explanation in variable phonology. *Language Variation and Change* 3: 1–22.
Guy, Gregory R. (1994). The phonology of variation. In Beals et al. (eds.), *CLS 30, Volume 2: The Parasession on Variation in Linguistic Theory*, Chicago: CLS. 133–49.
Guy, Gregory (1997a). Competence, performance, and the generative grammar of variation. In Frans Hinskens, Roeland van Hout and Leo Wetzels (eds.), *Variation, Change and Phonological Theory*, Amsterdam/Philadelphia: John Benjamins Publishing Company. 125–43.
Guy, Gregory (1997b). Violable is variable: Optimality theory and linguistic variation. *Language Variation and Change* 9: 333–47.
Guy, Gregory and Charles Boberg (1997). Inherent variability and the Obligatory Contour Principle. *Language Variation and Change* 9: 149–64.
Hammond, Michael (1994). An OT account of variability in Walmatjari stress. http://www.roa.rutgers.edu, ROA-20-0794.
Hanson, Kristin and Paul Kiparsky (1996). A parametric theory of poetic meter. *Language* 72: 287–335.
Hargus, Sharon (1993). Modeling the phonology–morphology interface. In Sharon Hargus and Ellen Kaisse (eds.), *Phonetics and Phonology, Volume 4: Studies in Lexical Phonology*, San Diego, California: Academic Press. 45–74.
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: Oxford University Press. 88–120. Also: http://www.linguistics.ucla.edu/people/hayes/gradient.htm.
Hayes, Bruce and Margaret MacEachern (1998). Quatrain form in English folk verse. *Language* 74: 473–507. Also: http://www.linguistics.ucla.edu/people/hayes/metrics.htm.
Hinskens, Frans, Roeland van Hout and Leo Wetzels (eds.) (1997). *Variation, Change and Phonological Theory*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
Holt, David Eric (1997). The role of the listener in the historical phonology of Spanish and Portuguese: An optimality-theoretic account. Doctoral dissertation, Georgetown University. Also: http://www.roa.rutgers.edu, ROA-278-0898.
Hudson, J. and E. Richards (1969). The Phonology of Walmatjari. *Oceanic Linguistics* 8: 171–88.
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*, Dordrecht: Kluwer Academic Publishers. 121–55.
Inkelas, Sharon (1999). Exceptional stress-attracting suffixes in Turkish: representations versus the grammar. In René Kager, Harry van der Hulst and Wim Zonneveld (eds.), *The Prosody–Morphology Interface*, Cambridge: Cambridge University Press. 134–87.
Itkonen, Terho (1957). Mellakoihin vai mellakkoihin? Yleiskielemme eräiden taivutushorjuvuuksien taustaa. *Virittäjä*. 259–86. (Abstract: Des raisons de certaines hésitations dans les déclinaisons du finnois littéraire).
Itô, Junko and Armin Mester (1995a). The core-periphery structure

in the lexicon and constraints on re-ranking. In Jill N. Beckman, Laura Walsh Dickey and Suzanne Urbanczyk (eds.), *Papers in Optimality Theory*, University of Massachusetts, Amherst: GLSA. 181–210.

Itô, Junko and Armin Mester (1995b). Japanese phonology. In John A. Goldsmith (ed.), *The Handbook of Phonological Theory*, Oxford: Blackwell. 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*, Oxford University Press. 419–62. Also: http://www.roa.rutgers.edu, ROA-145-0996.

Itô, Junko and Armin Mester (1998). The phonological lexicon. In Natsuko Tsujimura (ed.), *A Handbook of Japanese Linguistics*. Oxford: Blackwell. 62–100.

Iverson, Gregory K. and Shinsook Lee (1994). Variation as optimality in Korean cluster reduction. Paper presented at ESCOL 94, University of South Carolina.

Kager, René (1996). On affix allomorphy and syllable counting. In U. Kleinhenz (ed.), *Interfaces in Phonology*, *Studia Grammatica* 41, Berlin: Akademie Verlag. 155–71. Also: http://www.roa.rutgers.edu, ROA-88.

Kager, René (1999). *Optimality Theory*. Cambridge: Cambridge University Press.

Kang, Hyeon-Seok (1996). The deletion of *w* in Seoul Korean and its implications. In David Dowty, Rebecca Herman, Elizabeth Hume and Panayiotis A. Pappas (eds.), *Working Papers in Linguistics No. 48*, The Ohio State University, Department of Linguistics. 56–76.

Kiparsky, Paul (1982). *Explanation in Phonology*. Dordrecht, Holland: Foris Publications.

Kiparsky, Paul (1989). Phonological change. In Frederick Newmeyer (ed.), *Linguistics: The Cambridge Survey*, Cambridge: Cambridge University Press. 363–415.

Kiparsky, Paul (1993a). Blocking in nonderived environments. In Sharon Hargus and Ellen Kaisse (eds.), *Phonetics and Phonology, Volume 4: Studies in Lexical Phonology*, San Diego, California: Academic Press. 277–313.

Kiparsky, Paul (1993b).Variable rules. Handout distributed at the Rutgers Optimality Workshop (ROW1).

Kiparsky, Paul (1995). The phonological basis of sound change. In John A. Goldsmith (ed.), *The Handbook of Phonological Theory*, Oxford: Blackwell. 640–70.

Kiparsky, Paul (to appear). *Paradigm Effects and Opacity*. Stanford, CA: CSLI Publications.

Kroch, Anthony (1994). Morphosyntactic variation. In Beals et al. (eds.), *CLS 30, Volume 2: The Parasession on Variation in Linguistic Theory*, Chicago: CLS. 180–201.

Kroch, Anthony S. (1989). Reflexes of grammar in patterns of language change. *Journal of Language Variation and Change* 1(3): 199–244.

Kroch, Anthony (2001). Syntactic change. In Mark Baltin and Chris Collins (eds.) *The Handbook of Contemporary Syntactic Theory*. Oxford: Blackwell. 699–729.

Labov, William (1969). Contraction, deletion and inherent variability of the English copula. *Language* 45: 715–62.

Labov, William (ed.) (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William (1994). *Principles of Linguistic Change: Internal Factors*.

Oxford, UK, and Cambridge, USA: Blackwell.

Labov, William (1997). Resyllabification. In Frans Hinskens, Roeland van Hout and Leo Wetzels (eds.), *Variation, Change and Phonological Theory*, Amsterdam/Philadelphia: John Benjamins Publishing Company. 145–79.

Labov, William, Paul Cohen, Clarence Robins and John Lewis (1968). *A Study of the Nonstandard English of Negro and Puerto Rican Speakers of New York City*. Cooperative Research Report 3288, Vols I and II, Philadelphia: U.S. Regional Survey.

Leben, William R. (1973). *Suprasegmental phonology*. Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Liberman, Mark (1994). Optionality and optimality. Fragment of a draft, Department of Linguistics, University of Pennsylvania.

McCarthy, John (1986). OCP Effects: Gemination and Antigemination. *Linguistic Inquiry* 17(2): 207–63.

McCarthy, John (1988). Feature geometry and dependency: A review. *Phonetica* 43: 84–108.

McCarthy, John (1994). The phonetics and phonology of Semitic pharyngeals. In Patricia Keating (ed.), *Papers in Laboratory Phonology III*, Cambridge: Cambridge University Press. 191–283.

Mohanan, K. P. (1995). The organization of the grammar. In John A. Goldsmith (ed.), *The Handbook of Phonological Theory*, Oxford: Blackwell. 24–69.

Morris, Richard E. (1998). *Stylistic variation in Spanish phonology*. Doctoral dissertation, Ohio State University. http://www.roa.rutgers.edu, ROA-292-0199.

Müller, Gereon (1999). Optionality in optimality-theoretic syntax. *Glot International* 4(5).

Myers, James (1995). The categorical and gradient phonology of variable *t*-deletion in English. MS., York University.

Myers, James and Gregory R. Guy (1997). Frequency effects in variable lexical phonology. In Charles Boberg, Miriam Meyerhoff, Stephanie Strassel, and the PWPL series editors (eds.), *University of Pennsylvania Working Papers in Linguistics,* 4(1): 215–27.

Nagy, Naomi and William Reynolds (1997). Optimality theory and variable word-final deletion in Faetar. *Language Variation and Change* 9(1): 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*, New York: Academic Press. 37–54.

Newmeyer, Frederick J. (2000). Optimality and functionality: Some critical remarks on OT syntax. http://www.roa.rutgers.edu, ROA-402-08100.

Oostendorp, Marc van (1997). Style registers in conflict resolution. In Frans Hinskens, Roeland van Hout and Leo Wetzels (eds.), *Variation, Change and Phonological Theory*. Amsterdam/Philadelphia: John Benjamins Publishing Company. 207–29.

Orgun, Cemil Orhan (1996). Sign-based morphology and phonology with special attention to optimality theory. Doctoral dissertation, Department of Linguistics, University of California, Berkeley. http://www.roa.rutgers.edu, ROA-171.

Paolillo, John (2000). A probabilistic model for Optimality Theory. In

Julie Auger and Andrea Word-Allbritton (eds.), *The CVC of Sociolinguistics: Contact, Variation, and Culture*. Indiana Working Papers in Linguistics, vol. 2, pp. 89–107.

Paolillo, John (forthcoming). *Analyzing Linguistic Variation: Statistical Models and Methods*. Stanford, CA: CSLI Publications.

Paunonen, Heikki (1995). *Suomen kieli Helsingissä. Huomioita Helsingin puhekielen historiallisesta taustasta ja nykyvariaatiosta* (The Finnish language in Helsinki. Observations on the historical background of and current variation in spoken Helsinki Finnish). Helsinki: Helsingin yliopiston suomen kielen laitos.

Pierrehumbert, Janet (1994). Knowledge of variation. In Beals et al. (eds.), *CLS 30, Volume 2: Parasession on Variation in Linguistic Theory*, Chicago: CLS. 232–56.

Prince, Alan and Paul Smolensky (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers University, New Brunswick, and University of Colorado, Boulder.

Reynolds, William Thomas (1994). *Variation and phonological theory*. Doctoral dissertation, University of Pennsylvania.

Ringen, Catherine O. and Orvokki Heinämäki (1999). Variation in Finnish vowel harmony. *Natural Language and Linguistic Theory* 17: 303–37.

Rose, Sharon (1997). Featural morphology and dialect variation. In Frans Hinskens, Roeland van Hout and Leo Wetzels (eds.), *Variation, Change and Phonological Theory*, Amsterdam/Philadelphia: John Benjamins Publishing Company. 231–66.

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*, Chicago: CLS. 277–87.

Tesar, Bruce and Paul Smolensky (1995). The learnability of Optimality Theory. In Raul Aranovich, William Byrne, Susanne Preuss and Martha Senturia (eds.), *The Proceedings of the Thirteenth West Coast Conference on Formal Linguistics (WCCFL XIII)*, Stanford, California: Stanford Linguistics Association/CSLI. 122–37.

Vennemann, Theo (1988). *Preference Laws for Syllable Structure and the Explanation of Sound Change*. Amsterdam: Mouton de Gruyter.

Wang, William S.-Y. (1969). Competing changes as a cause of residue. *Language* 45: 9–25. Reprinted in Philip Baldi and Ronald N. Werth (eds.), (1978). *Readings in Historical Phonology*. University Park and London: The Pennsylvania State University Press.

Weinreich, Uriel, William Labov and Marvin I. Herzog (1968). Empirical foundations for a theory of language change. In W. P. Lehmann and Yakov Malkiel (eds.), *Directions for Historical Linguistics: A Symposium*, Austin: University of Texas Press. 95–195.

Wilkinson, Karina (1988). Prosodic Structure and Lardil Phonology. *Linguistic Inquiry* 19: 325–34.

Wolfram, Walter A. (1969). *A Sociolinguistic Description of Detroit Negro Speech*. Washington, DC: Center for Applied Linguistics.

Zubritskaya, Katya (1997). Mechanism of sound change in OT. *Language Variation and Change* 9(1): 121–48.