Balancing Data and Theory in the Study of Phonological Variation and Change

Frans Hinskens
University of Nijmegen

Roeland van Hout
University of Tilburg

W. Leo Wetzels
Free University of Amsterdam

1. Introduction

During the second half of this century, linguistics has shown a steady evolution towards fragmentation, resulting in a multitude of different disciplines, with language variation and change becoming the object of historical linguistics, dialectology, sociolinguistics and — to some extent — creole linguistics. As a terminological side effect, the word 'linguistics' narrowed its semantic scope, referring now most of the time to the study of the 'core areas' of grammar: phonetics, phonology, morphology, syntax, and semantics (witness recent introductory textbooks such as, e.g. O'Grady, Dobrovolsky & Aronoff 1997: 10). Whereas the core areas of linguistics are generally concerned with aspects of language systems, oriented towards synchrony, and primarily theory-driven in their methodology, the 'peripheral' disciplines, such as dialectology and sociolinguistics, are more concerned with aspects of language use, with an inherently diachronic orientation and, above all, a data-driven methodology. Across the board, though especially in the 'peripheral' areas, there is a growing diversification in the types of data studied; apart from production also data from
perception and comprehension are investigated; in addition to ‘spontaneous’ data, experimental data more and more play a role (cf. Stemberger 1992).

No doubt, the methodological and theoretical differentiation that has taken place during the last decades, and that has led to the emancipation of the emerging disciplines, has yielded insights that otherwise might not have been achieved. Also, representatives of the study of language variation have tried to implement their analyses in mainstream theoretical linguistics, as is most visible in the work of William Labov and David Sankoff, who developed the rule formalism introduced in Chomsky & Halle’s (1968) *The Sound Pattern of English* — henceforth SPE — into a format capable of expressing the relative weight of internal as well as external factors in phonological variation (cf. Sankoff & Labov 1979; Sankoff 1987). In a comparable manner, at present attempts are being made to link statistical analyses to the formalism of Optimality Theory, as shown in some works of an upcoming generation of linguists (see, e.g., Anttila). From the viewpoint of an integrated theory of language, which should embody both the grammar and the various aspects of its actual use, the direct implementation of statistical data into the grammar seems beneficial.

Despite the obvious progress in the study of language variation that has been achieved over the last decades, much of the actual research still concentrates on isolated facts, and tends to lose sight of the structural relations among language phenomena. On the other hand, not all of the more theoretically oriented linguists seem to be sufficiently aware of the fact that variation is an essential characteristic of language, as well as a prerequisite to linguistic evolution. As was pointed out by Kiparsky: “heterogeneity and variation are not abnormalities but part of the normal condition of language” (1988: 370). The few theoretical linguists that do show sensitivity to the problem of variation rarely have clear ideas regarding the place of language variation in their models.

The question of the relative importance of empirical data for constructing theories is an old one, and, surely, part of the proliferation of linguistic disciplines originated as a reaction against the invariance assumption, an important methodological tool of mainstream theoretical linguistics. The danger of the alienation of the more data-oriented areas from linguistic theory is not only apparent, but, as Goldsmith puts it, there is a real danger of “the ‘Balkanization’ of linguistics — the unfortunate lack of communication across frameworks or paradigms” (1992: 161-62). It is the purpose of this book to contribute to the discovery of areas where theoretical linguistics and variation linguistics can profit from each other’s findings. It seems to us that the most efficient way of bridging the gap between different theories and methodologies is to discuss the
issue of variation on the basis of case studies that explicitly address the question of the tension between variation and theoretical abstraction.

2. Phonological background

As we pointed out in the previous section, modern linguistics is characterized by a growing diversification in the types of data considered. At the same time, the post-SPE development of several phonological (sub)theories, some of which represent specific modules in the overall organization of the phonological grammar, has made available a range of new ways of analysis. Almost three decades ago, DeCamp wrote: “Despite the apparent initial hostility, the generative-transformational model is the most hospitable to sociolinguistics of any theory yet known” (1970: 162). Today, DeCamp would definitely be more than happy about the developments that have taken place since then. According to Auer, these developments have made phonological theory more and more variation-friendly, partly at least, because phonological models are now better equipped to deal with low level phonological rules (Auer 1997: 80).

2.1 Non-linear generative phonology

Since the first comprehensive presentation of generative phonology by Noam Chomsky and Morris Halle (1968), phonological theory has gone through a number of spectacular modifications. Over a little less than 30 years, a well-articulated framework has been developed which, rather than a monolithic theory, represents a number of different subtheories, globally referred to as ‘non-linear’ phonology, which together define the descriptive and explanatory outlines of the phonological grammars of human language. Although it is still a ‘generative’ theory insofar as it supposes the existence of different levels of representation connected through rules, non-linear phonology is in many respects fundamentally different from the classical framework set out in The Sound Pattern of English. The major differences are the following.

• The authors of SPE did not recognize the syllable as a necessary descriptive unit. This decision was soon shown to embody a regrettable break with the phonological tradition. It was argued convincingly, most forcefully by Vennekamn (1972) and Hooper (1976), that an explicative theory of phonology could not do without the concept of the syllable, which has become completely rehabilitated in non-linear phonology. 1 Syllable Theory also deals with the internal structure of the syllable, more in particular with the question of how
much internal structure is needed to express phonologically significant generalizations.

- In linear generative phonology stress was considered to represent a segmental feature. By doing so, phonological theory was forced into a formal inconsistency by allowing this feature to be multivalued ([1stress], [2stress], [3stress], etc.), whereas all other features were claimed to be binary. In Metrical Stress Theory, stress is accounted for in a way that differs fundamentally from the one proposed in SPE. Stress is no longer considered to represent an inherent property of vowels, but a relative property of the syllable rhyme (nucleus and coda). The degree of stress for a given rhyme is derived from the position of that rhyme in a hierarchical prosodic structure, involving the foot and the word domains. In turn, the study of the phonologically relevant prosodic categories, which also involve domains larger than the phonological word, such as the clitic group, the phonological phrase, the intonational phrase, etc. is the object of Prosodic Phonology.

- In linear phonology speech sounds are represented as unordered sets of features, or feature matrices. Every segment is (positively or negatively) specified for all features, and phonological rules replace fully specified matrices by other fully specified matrices, and delete or epenthesize fully specified matrices. In doing so, the theory is unable to express in a principled way the fact that a feature may extend over domains larger than a single segment. Neither can it explain why some features consistently pattern together in phonological processes, while other features never do. Autosegmental Phonology (and, likewise, the derived Feature Geometry) express the autonomy of the individual feature as well as the relative solidarity among features. This approach has established a hierarchically structured segment that allows phonological rules to directly manipulate individual features or groups of features gathered under structural nodes. Assimilation is treated as spreading of features or feature bundles and is allowed to create structures in which segments share the same (set of) feature(s).

- In SPE, the pregenerative distinction between rules of phonology (allophony) and rules of morphophonology was abandoned. As a consequence, the fact that languages have rules with different clusters of characteristics was considered to be an accidental fact, or at any rate, irrelevant for the way the native speakers organize their knowledge about the phonological patterns of their language. In Lexical Phonology the importance of this dichotomy is reestablished in a slightly different way by the recognition of two different classes of rules, lexical rules, which interact with morphology, and postlexical rules, which are insensitive to the internal structure of the word.
Several of the contributions to this volume are couched within the framework of non-linear phonology or involve issues that relate to one of its modules. In Labov's study, the process of cross-word (re)syllabification is addressed. The author investigates the question whether the deletion of word-final /t,d/ is influenced by the possibility for the coronal stops to form an acceptable onset with the following word-initial consonant. The higher order prosodic domains of the phonological word and the phonological phrase play an important role in Auer's study of co-occurrence relations between linguistic variables. In her analysis of sonorant disimilation in the Romance languages, Lloret argues that the featural properties of sonorants are best captured by Avery & Rice's (1993) model of feature geometry.

2.2 Optimality Theory

A much more radical departure from standard generative phonology is represented by the recently proposed Optimality Theory, or OT (Prince & Smolensky 1993; McCarthy & Prince 1994). Here extrinsic ordering of language specific rules no longer exists. Instead, a set of universal constraints is assumed which determines the way in which surface structure is allowed to deviate from lexical representations. The only generative capacity of the model resides in a function called GEN, for 'generator,' provided by Universal Grammar. GEN projects an unlimited set of possible output candidates from a single lexical input form. It is the task of the language learner to discover what the lexical representations of the morphemes of their language are and what the relative importance is that their language attributes to the (potentially conflicting) universal constraints. All candidate analyses of the lexicalized form are rated according to their success in complying with the ordered constraint set. The candidate that best satisfies the constraints is selected as the optimal one. Thus, other than in previous theories of phonology, the output candidate is not the result of an operation executed on the input. It is part of the set projected by GEN and selected by parallel evaluation of all possible candidates. To illustrate constraint ranking and candidate evaluation in Optimality Theory, we will look at the well-known process of Syllable-Final Devoicing, based on Mascaró & Wetzels (forthcoming).

Devoicing can be analysed as the effect of three general unmarkedness properties of universal grammar, which are adequately formulated as constraints in (1) below. The first constraint requires that segmental properties, [œvoice] in this case, be preserved in onsets, but not necessarily in codas. Another, (1c), captures the unmarked character of [-voice] in obstruents by stating that they
must be voiceless. Finally, constraint (1b) expresses the general requirement that \([\alpha_{\text{voice}}]\) in the output be faithfull to \([\alpha_{\text{voice}}]\) in the input.

\[
\begin{align*}
(1) \quad \text{a. } & \text{IDENT (VOICE, ONSET)} \\
& \text{Correspondent (i.e., lexical and surface) elements have identical} \\
& \text{values for voice in the onset.} \\
\text{b. } & \text{IDENT (VOICE)} \\
& \text{Correspondent (i.e. lexical and surface) elements have identical} \\
& \text{values for voice.} \\
\text{c. } & \text{*VOICE/OBSTRUENT} \\
& [-\text{sonorant}] \text{ segments cannot be voiced.}
\end{align*}
\]

The constraints formulated in (1) above are of two kinds. The constraints (1a,1b) are part of the set of so-called Faithfulness Constraints, which impose a faithfull correspondence between lexical representation and surface form. The constraint (c) is a member of the set of universal Wellformedness Constraints. IDENT (VOICE, ONSET) is in a subset relation with IDENT (VOICE), and therefore the ordering is universally established as IDENT (VOICE, ONSET) » IDENT (VOICE).\(^5\) There are thus three possible orderings for the constraints, namely:

\[
\begin{align*}
(2) \quad \text{a. } & \text{IDENT (VOICE, ONSET) » IDENT (VOICE) » *VOICE/OBSTRUENT} \\
\text{b. } & \text{*VOICE/OBSTRUENT » IDENT (VOICE, ONSET) » IDENT (VOICE)} \\
\text{c. } & \text{IDENT (VOICE, ONSET) » *VOICE/OBSTRUENT » IDENT (VOICE).}
\end{align*}
\]

Since, ideally, any given order of the constraints in (2) should represent a natural language, three possible voicing effects are predicted: no voicing contrast, voicing contrast, but no coda devoicing, and coda devoicing. In the first case, faithfulness to the underlying voicing values will be guaranteed by the superordinate position of the IDENT constraints, and the result will be no devoicing as is illustrated for English in (3) below. If *VOICE/OBSTRUENT dominates the other two constraints, voiced obstruents will be completely disallowed. This accounts for languages like Hawaiian, which lacks voicing distinctions in its lexical inventory and on the surface. In the third case *VOICE/OBSTRUENT is ordered between the two faithfulness constraints. Since IDENT (VOICE, ONSET) dominates *VOICE/OBSTRUENT, onsets will remain faithful to underlying voicing values; codas will not, however, since at the same time *VOICE/OBSTRUENT outranks IDENT(VOICE). Obstruent codas will be able to violate faithfulness to the input in order to satisfy *VOICE/OBSTRUENT, and
they will change their underlying voicing value to 'voiceless.' This is illustrated in (4) for German.

In OT, evaluation is represented in the form of a constraint tableau, which must be interpreted as follows. Left to right order indicates constraint ranking, from higher to lower. Constraint violations are marked with an asterisk. The violation of a particular constraint does not automatically eliminate a candidate from further consideration. Constraint violation is fatal for a given candidate when there is at least one other candidate that does not violate the same constraint, or violates it less often. Fatal violations are marked by *!. In cases of fatal violation, cells that correspond to lower ranked constraints are shaded to mark the fact that evaluation for satisfaction of lower-ranked constraints is unnecessary. The symbol ♦ identifies the optimal candidate.

(3) English: no coda devoicing in *houseboy*

<table>
<thead>
<tr>
<th></th>
<th>IDENT(VO,ONS)</th>
<th>IDENT(VO)</th>
<th>*VOICE/OBSTR</th>
</tr>
</thead>
<tbody>
<tr>
<td>hou[z b] oy</td>
<td></td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>hou[s b] oy</td>
<td>*!</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>hou[z p] oy</td>
<td>*!</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

(4) German: coda devoicing *Hau[s d]iener* 'servant'

<table>
<thead>
<tr>
<th></th>
<th>IDENT(VO,ONS)</th>
<th>*VOICE/OBSTR</th>
<th>IDENT(VO)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hau[z d]iener</td>
<td></td>
<td>**!</td>
<td></td>
</tr>
<tr>
<td>Hau[s d]iener</td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Hau[z t]iener</td>
<td>*!</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

As the tableaux in (3) and (4) clearly show, the different typologies of devoicing are the result of a different ranking of the relevant constraints.

The majority of the papers contained in this volume argue for the greater explanatory potential of OT over traditional paradigms, specifically when it comes to explaining patterns of variation. OT is successfully applied to account for dialect differences by Rose and by Smith. Borowsky & Horvath propose that l-vocalisation in Australian English is shaped by constraints that relate to optimal syllable shape and sonority. Indeed, OT appears to offer a theoretical framework that is well suited to handle historical and geographical variation. Furthermore, different attempts are made to account for quantitative and stylistic variation in OT. The proposals to handle the quantitative patterns can be subsumed under two headings:
1. **Complete ordering of constraints in competing grammars** (see also Kiparsky 1993).

The view underlying this approach is reminiscent of the 'lectal' analysis of quantitative variation, based on implicational scaling techniques (Bailey 1973; Bickerton 1973), in which every single 'lect' is approached as a categorical grammar of its own. Quantitative patterns are the consequence of strategies of choice between grammars. Although Van Oostendorp passes over the quantitative aspect, his contribution is illustrative of a grammar competition analysis. The author starts from the assumptions that (a) style levels (or 'registers') all constitute subtly different grammars of their own, and that (b) the more elevated the style level, the higher the faithfulness constraints are ranked. He analyzes French liaison in *conversation familière*, *conversation soignée* and *discours et lecture*, Dutch vowel reduction in formal, semi-formal and informal style levels, and Turkish vowel epenthesis in loan words in careful, less careful and colloquial speech. In each case, the differences between style levels are accounted for through minimal differences in constraint ranking. Interestingly, the author claims that an OT type analysis allows one to “take any two style levels in a language system and predict which of the two is the more formal”.


Anttila argues for the possibility of having partial ranking of constraints, producing multiple winners as output. In the case of the Finnish genitive plural allomorphy, the grammar underdetermines the output. Complete ranking has its price, as adding ranks complicates the grammar. The variation patterns are produced within one and the same grammar. Nagy & Reynolds obtain the same effect by permitting constraints to 'float'. This property of floating is defined as follows: “We propose floating constraints, whereby some particular constraint within a single grammar may be represented as falling anywhere within a designated range in the ranking hierarchy” (1997: 37). Borowsky & Horvath acknowledges the two options of non-ranking and floating as relevant to their data, but, since they are not concerned with a precise prediction of the quantitative patterns, they leave their preference open.

Optimality Theory appears to be a source of inspiration for the study of variation and change. The theory is still young, and a lot remains to be researched. As always, not every proposal will turn out to be satisfactory in the long run. A critical discussion and an empirical test of the otherwise peaceful OT practice of 'exploding constraints' into 'families of constraints' to account for
quantitative variation can for instance be found in Guy (see also Pierrehumbert 1994: 245 and Auer 1997: 69 for critical discussion of other aspects).

3. Accounting for variation and change in historical linguistics, dialectology and sociolinguistics

According to Bloomfield (1933: Ch. 1 passim), the systematic, scientific study of language started with the attempts by the Neogrammarians, a group of nineteenth-century German philologists, and their immediate predecessors to establish the historical relationships between the Indo-European languages as well as their individual evolution. Initially they focused on the regularity of the correspondences between the sounds of the different languages and on the sound structure of the hypothetical proto-language. To this end, the Neogrammarians further elaborated and refined the methods of external comparison and internal reconstruction, still used in essentially identical ways in contemporary historical linguistics.

The discovery of significant amounts of apparent or real counter-evidence ('residual forms') against the Neogrammariand assumption that sound change was regular and exceptionless played an important role in the development of dialectology. A major difference between dialectology and the Neogrammariand tradition is the fact that extralinguistic matters (such as geography, natural and man-made borders, the cultural effects administrative and economic centres can have on the 'periphery' of a language area etc.) play an explicit role in dialectology. In this respect, sociolinguistics can be seen as a continuation of the anti-Neogrammariand approach to the study of sound change. However, in several other respects, the development from dialectology to sociolinguistics is marked by some important disruptions, both conceptually and methodologically.

The Neogrammarians distinguished sound change from analogy and borrowing. Language change in the strict sense takes the form of either sound change, which is phonetically motivated, or analogical change, which is of a morphological nature. At the core of the Neogrammariand position is the claim that sound change is a merely mechanical, physiologically induced, phenomenon. Therefore it is supposed to operate blindly, allowing no exceptions. For this reason the notion of 'sound law' (German Lautgesetz) was introduced. However, the Neogrammarians always left the door open for types of linguistic change other than sound change. One of these is borrowing from co-existing linguistic systems, either dialects or other languages. Residual forms could also be explained as the result of either analogy or competing sound changes. An impor-
tant difference between the processes of analogy and (mechanical) sound change is that the former is more clearly motivated by the grammar, correcting the lack of regularity in morphological paradigms caused by sound change. An interesting case of the working of analogy is discussed in Rose’s study of the evolution of the way second person singular feminine subjecthood is expressed in Muher, a member of the South Ethio-Semitic dialect cluster known as Gurage. In Smith’s contribution, an elaborate attempt is made to recapitulate the stages of evolution that account for the differences in sound structure between a group of indigenous languages of Australia and their reconstructed protolanguage. Although Smith’s reconstruction is based upon the premisses of Optimality Theory, the claim that the changes that occurred can be described in terms of the reordering of universal constraints implies the regularity of the relevant evolutionary processes. As in the time of the Neogrammarians, the effects of analogy and borrowing are considered secondary or peripheral.

Contrary to the Neogrammian position, only few processes of linguistic change that have been completed appear to be entirely exceptionless. Within a speech area one can usually find dialects that did not undergo a certain historical change. Not every linguistic change necessarily reaches full geographical spread throughout the entire speech community. Moreover, within the grammar and lexicon of a single dialect, regular and exceptional forms can often be found to exist side by side. This fact has led some scholars to subscribe to the idea that especially sound change tends to be ‘sporadic.’ In short, both the extensive and the intensive diffusion of linguistic change may be incomplete. As regards the intensive diffusion of linguistic change, the distinction between the Neogrammian type of sound change (which is phonetically gradual and lexically abrupt) and lexically diffuse sound change (which is phonetically abrupt and lexically gradual — Scheutz 1987: 1608), the so-called ‘Neogrammian controversy’, has been the subject of a particularly fruitful exchange of ideas between Labov and Kiparsky. In section 5.1 below, we will briefly return to this issue.

Like historical linguistics, dialectology is concerned with the results of past processes of change. Yet, the object of dialectology is not a diachronic description or the comparison of several historical phases of a language, but a description of synchronic dialectal diversity. Variation across dialects typically results from extensive numbers of exceptions to some process of linguistic change. Dialect geography concentrates on the interdialectal, spatial diffusion of historical changes. Three contributions to this volume can be mentioned in this connection. Rose compares the morpheme expressing the second person singular feminine subject in Muher and Chaha (members of the Gurage dialect cluster) and reconstructs the diachrony. Lloret systematically analyzes crossdialectal dif-
ferences in the behavior of the infinitive marker after cliticization in Catalan. Finally, Guy devotes a few paragraphs to the well-known difference in the proportion of word-final t/d-deletion before vowels and pauses between the New York City and the Philadelphia dialects of American English.

Like much work in historical linguistics, traditional dialectology implicitly assumes linguistic systems to be homogeneous, which is clearly manifested in the fact that most research is based on the production of one or a few informants for every variety. Auer’s study is based on data from introspection, specifically judgements of very few informants on the wellformedness of forms which are intermediate between the dialect and the standard language in three different speech communities. Most sociolinguists would probably approach this problem on the basis of data gathered in a sample or samples of informants. On the other hand, Auer should definitely be credited for presenting an in-depth study of a problem that most sociolinguists are aware of, but that as yet hardly anyone has studied empirically.

One of the great merits of dialectology has been to provide a robust body of empirical evidence showing that the archetypical Neogrammarian conceptualization of linguistic change is rather one-sided and, in particular, that the alleged exceptionlessness of sound change usually does not imply ‘uniformity,’ although it does not in principle exclude regularity. The first of the two case studies presented by Lloret is concerned with the dissimilation of sonorants in a range of Romance languages, mainly in several dialects of Catalan. The author shows that, although far from being exceptionless from the point of view of the lexicon (hence ‘sporadic’), this dissimilation is regular from the point of view of its phonological conditioning and outcome.

Findings of historical linguistics and dialectology led to the insight that, regardless of its origin, a process of language change that is not completed in some respect results in variation between and/or within varieties.

With respect to the observability of processes of language change, Bloomfield (1933), like most of his contemporaries, was pessimistic. He was convinced that only analogical change and borrowing can be observed to some extent. This pessimism is not shared by most sociolinguists. Since Labov (1966) demonstrated that linguistic variation can be a synchronic slice of a process of change in progress, it is clear that the study of language variation is of inherent interest to historical linguistics. So, whereas the efforts of historical linguistics and dialectology made clear that processes of language change that are not completed in some respect result in variation, sociolinguistic investigations of the Labovian type show that synchronic variation is typically a phase in a process that may eventually result in a categorical change. Quantitative variation is a necessary,
but, as such, not a sufficient condition for linguistic change, witness the existence of instances of stable variation. Examples are the variation in the place of articulation of the nasal in the suffix -ing ~ -in' (i.e. [ɪŋ] ~ [ɪn]) in certain varieties of English, the word-final deletion of coronal stops in several varieties of English and Dutch, and the age-old variation in the allomorphy of the Finnish genitive plural, studied by Anttila. One of the main questions Anttila attempts to answer is why this variation could linger on for centuries without very much change. After all, the 'free' variation, in which recent loans and foreign names are involved, shows that the phenomenon is still productive. Anttila’s answer is straightforward: the elimination of variation would complicate the grammar, because it requires the ranking of constraints that are now only partially ranked.

Sociolinguistic attention is concentrated on heterogeneity within linguistic systems. The limitation of research to intra-systemic variation in sociolinguistics should be explained first and foremost as a reaction against the explicit idealization that was typical of the classic generative approach to language as a homogeneous system. The merit of sociolinguistics is not so much that it stresses the heterogeneity of linguistic systems, but rather that it has developed the — essentially quantitative — techniques capable of revealing the order that exists in the heterogeneity. Indeed, many alleged instances of ‘free variation’ have indeed turned out to be cases of ‘orderly heterogeneity.’ The application of these techniques and methods also led to more general insights, such as the fact that today’s heterogeneity may in the future turn out to be a change in progress. A good example is the vocalization of postvocalic /l/ in coda position in the Adelaide dialect of Australian English which Borowsky & Horvath show to be both linguistically and socially ‘regular.’ Especially in the latter respect, the variation displays patterns that are usually associated with processes of language change. The authors point out that this change, which occurs in several speech communities without the likelihood of borrowing, strongly suggest “that we must look for systemic [i.e. internal] causes of change.”

A considerable amount of sociolinguistic literature on phonological variation and change has accumulated over the past 25 years. Most of the models that deal with the the interaction between internal and external factors in the emergence and spread of phonological change are not, in essence, very different from those presented in Labov (1972: Ch. 7) and Kroch (1978).

An entirely different, though equally productive area of sociolinguistic research concerns the structural consequences of language contact. Traditionally, most work on language contact focussed on morphosyntactic issues; in the last decade more attention has been paid to the provinces of phonology and morphophonology (Campbell 1997). Auer’s present study of the co-occurrence re-
restrictions between linguistic variables which manifest themselves in the (un)acceptability of intermediate forms is located at the crossroads between intra-systemic (hence 'sociolinguistic') variation, inter-systemic (hence cross-dialectal) variation and contact linguistics.

In accordance with the general practice, we use the notion of variation to refer to both intrasystemic, quantitative variation and intersystemic variation, i.e., variation across related dialects. The first case involves non-categorical phenomena, the second differences between related varieties in the nature or distribution of (either categorical or variable) elements or structures. In this connection, the question immediately comes up of whether sociostylistic variation, the topic of Van Oostendorp's contribution, is to be considered as intra- or intersystemic variation. In the light of the Optimality Theoretical view on language variation as the competition between subtly different grammars (see section 2.2 above), it does not come as a surprise that Van Oostendorp analyzes stylistic variation as intersystemic variation. 'Temporal' variation, in the sense of variation across successive historical stages of a linguistic system, is usually referred to as language change.

Whereas attention in historical linguistics is focused on (aspects of) the language system, either in its diachronic development or in its historical relation to other systems, sociolinguistics analyzes various products of language use, typically recorded speech. Hence, historical linguistics, like dialectology, investigates the products of past change in the 'langue.' Sociolinguistics, on the other hand, investigates the process of change in 'parole.' Paramount in the difference is which aspect of the change is examined: whereas historical linguists most often limit their study to the results of intensive generality, in dialectology attention is focused on the areal reflection of extensive generality. Sociolinguists generally study both the extensive and intensive generalization of change in progress in a given speech community, usually a city or some village.

Apart from fundamental differences in orientation that exist between historical linguistics and sociolinguistics, there are also basic differences in methodology. These differences concern aspects like

• the nature of the data: written (historical linguistics) or oral, elicited or spontaneous;
• the ways in which the material is collected: 'armchair-method' (historical linguistics) or 'tape recorder-method';
• the types of analyses: quantitative, statistical (rarely used, sometimes not feasible, in historical linguistics).

Also, the fact that historical linguists often try to follow large numbers of changes in outline over a long period of time, whereas sociolinguists usually
investigate a comparatively small amount of changes in great detail affects not only the research questions, but also the type of conclusions and inferences that are possible.

Despite the many conceptual and methodological differences, there is one insight which is shared by *historical linguistics*, *dialectology* and *sociolinguistics*, namely the insight that the closer two linguistic systems are (in time, space or in the relevant social dimensions), the greater the degree of similarity. As far as geographical and social space are concerned, this also results from processes such as borrowing and convergence. Conversely, structural dissimilarity tends to grow with distance (cf. Chambers 1995: 58-66).

### 4. The relevance of linguistic structure for the study of variation and change

In its tendency to concentrate on linguistic forms rather than structures, and in its inclination towards studying linguistic forms in isolation, the sociolinguistic approach to language variation and change unmistakably inherited some of the features of dialectology (see also Chambers & Trudgill 1980: 38). In many a sociolinguistic study, the implicit methodological position seems to be that the less a description is embedded in the grammar, the more reliable and useful it is. This, however, is but one manifestation of the widespread conviction that a thorough empirical study of variation has little to gain from theoretical linguistics in general.

Undoubtedly, those who expect from linguistic theory a ‘hard’ answer to the question of what is a possible or impossible type of variation, and, in a second instance, what is a possible grammatical change, will be disappointed. More in particular: the present state of linguistic theory hardly ever allows for deductive or causal predictions/explanations of language change. The structure of a given language and/or a well-founded theory do in principle sometimes allow for probabilistic predictions/explanations, but in our actual state of knowledge, there is very little that linguistic theory can predict with more than a reasonable probability. Even if the great majority of linguists is by now convinced of the existence of an innate language acquisition device, our understanding of what exactly its properties are is still very poor. At least as poor is our understanding of how it relates to an integrated theory of language competence and performance. We only have vague ideas about how linguistic competence of a given language is acquired or what components a model of performance must contain, or what exactly the characteristics of these components are.
Most linguists do indeed believe that the form of a specific grammar is a dynamic mechanism globally kept in shape by our innate capacity to create language but, beyond that, under constant pressure of a number of rivalling forces. Language must be learnable, which probably explains part of its regularity: whether syntactic case is expressed by a fixed word order or by morphological case markers, whether morphological categories are expressed by suffixes or by modifications of the stem, the mechanisms used are strikingly regular in every language. Regularity is indeed typical for all areas of grammar, including phonology and phonetics. As Labov points out, phonological variation usually leads to a state of regular complementary distribution between the original value of a sound and a new contextually determined variant. Even context-free phonetic properties of segments can be amazingly homogeneous among speakers of a language. Dutch, English and Venezuelan Spanish systematically produce [t] with dental, alveolar and interdental contact respectively. Typological universals, often presented in the form of context-free universal implicational laws, show a clear cross-linguistic preference for articulatorily simple segments over marked or complex ones. Surely, human perception is an important force in organizing linguistic systems. As Clements and Herz suggest “articulatory organization is oriented towards the goal of achieving relatively stable acoustic outputs with optimal perceptual properties” (1995: 7). We can also expect that the properties of a specific grammar, once they are mastered, co-determine the type of variation that is likely to occur. Variation is an act of speech, but “speech is a physical and behavioral manifestation of cognitively-represented linguistic structure, and, as such, cannot be fully understood without reference to the linguistic structure that underlies it” (Clements and Herz 1995: 2). Even if all of the foregoing is correct, our knowledge of how exactly the different factors mentioned contribute to define the possible range of variation -and of possible grammars, of course- is at this point not detailed enough to make any precise predictions (for more detailed discussion of this problem, see Labov 1994). However, this does not mean that nothing can be said. The following is an illustration of a type of variation that seems to occur typically in languages with a very specific type of underlying sound system.

In many indigenous languages of the Americas, more specifically of South America, but also of Australia and elsewhere, nasal stops may show an intricate allophonic, illustrated by the examples in (5), which are taken from the Brazilian language Kaingang.
Probably the best-known feature of Kaingang phonology is the occurrence of double and triple contour segments, which occur as allophones of the nasal phonemes /m, n, ñ, ñ/. Post-oralized\textsuperscript{10} allophones occur syllable-initially before oral vowels, pre-oralized allophones occur syllable-finally after oral vowels, and circum-oralized allophones are surface manifestations of ambisyllabic nasal consonants between oral vowels. The different allophones of /m/, as well as their phonotactic distributions, are representative for all the nasal consonants in Kaingang (cf. Wetzels 1995). The range of allophonic variation found in Kaingang is in fact very common in South American indigenous languages, although circum-oralized nasals are a little less common than pre- or post-oralized variants.\textsuperscript{11}

Contour segments can be found in Kuyawi, Yuhup, Hupda, Nukak, and Kakua, all members of the Makú family (Martins 1995). They occur in Karo (Ramarama family — Gabas 1989), Maxacali (Macro Jê family — Gudschinsky, Popovich & Popovich 1970), Southern Barasano (Tucano — Smith & Smith 1971), Guarani (Adelaar 1986), and many other languages.

The variation under discussion, which occurs across language families, is unheard of in any of the Indo-European languages. For example, in French, which, like Kaingang, has a contrast between oral and nasal vowels, as in pot [po] ‘pot’, pont [pɔ̃] ‘bridge’, beau [bo] ‘beautiful’, bon [bɔ̃] ‘good’, mot [mo] ‘word’, mont [mɔ̃] ‘mountain’, contour segments do not occur. One might wonder therefore whether there is some structural reason why these sounds only occur in certain linguistic areas. One suggestion for a structural explanation comes from Steriade, who writes with regard to the circum-oralized segments of Kaingang: “The delay [in the onset of nasalization of the consonant] is obviously motivated by the fact that the preceding vowel is\textit{distinctively} oral: had nasalization started on ‘time,’ at the beginning of the stop closure, the possibility of anticipatory nasalization affecting the preceding vowel would have muddled the contrast between oral and nasal vowels” (1993: 448, with original emphasis). Steriade’s explanation is interesting, and it might very well be the case that part of the motivation for the relative stability of contour segments in these languages...
resides in their effect of preserving the oral/nasal contrast on vowels. We are reluctant to believe, however, that the preservation of vocalic contrast is the primary *motivation* for the existence of this type of allophony. First of all, it does not account for the areal distribution of the phenomenon; why are contours not found in French or Portuguese? Secondly, there are languages that have nasal/oral contour segments without having a nasal contrast on vowels. One such language is Wari (Chapakuara family), described in Everett & Kern (1998). Wari has [mb] and [nd] in free variation with [m] and [n] in syllable-initial position before vowels, which are always underlyingly oral. Finally, it seems to be less easy for languages to maintain an oral/nasal contrast in vowels before a nasal consonant than after a nasal consonant. One would consequently expect that contour segments are more common in codas that in onsets. This prediction appears to be doubtful, at the very least. To the extent that the relatively small sample of languages we had at our disposal permits any strong conclusions, the opposite appears to be true: almost all the languages that have pre-oralized nasal consonants ([b|m]) also have post-oralized ones ([mb]), whereas the opposite implication does not hold, as in Wari. One noticeable exception is Mundurukú (Tupí family — Crofts 1973), which has contours in the syllable coda, but not in the syllable onset. However, in Mundurukú, codas are in a different way exceptional as compared to onsets, and it is this difference that turns out to provide the structural cue for the presence of nasal contours in all the languages mentioned: whereas Mundurukú allows the triple contrast /p, b, m/ to be realized in the syllable onset, voiced stops are banned from the coda position. Indeed, in all the languages that have the kind of allophonic variation under discussion, the underlying system of consonants is one where the voiced series is completely lacking. This suggests that contour segments of the kind discussed have a perceptual rather than an articulatory motivation, on the assumption that a segment with a sonority contour is perceptually more salient than a segment with a ‘flat’ voiced or nasal articulation.12

The above example illustrates that phonological variation or, in this case, allophony may be determined by properties of the underlying system of segments. Language typology is another area where variation studies may profit from insights acquired by theoretical linguistics. For example, it is a well-known fact that intervocalic consonants almost exceptionlessly function as onsets of the syllable of which the immediately following vowel is the nucleus. This generalisation is almost categorical cross-linguistically in non-derived words. In derived words, specifically across prefix boundaries, as in German *ent=erben* or Dutch *ont=erven* ‘disinherit’, and more generally across word boundaries exceptions may be found. The situation is a little different as regards complex onsets. Many
languages do not allow specific sequences to form complex onsets in derived words of any kind, whereas the same consonants obligatorily cluster to form syllable onsets in non-derived words. For example, in Brazilian Portuguese, a prefix-final consonant functions as the onset of the base initial vowel in a word like *inegal ‘unequal’, unlike in Dutch or German. Equally, across word boundaries onsetless syllables are avoided at a normal rate of speech. In the same language however, a surface syllabification contrast can be observed between tautomorphemic /bl/ in *sublime ‘sublime’, and heteromorphemic /bl/ in *sub[^i]linhar ‘underline’, where the superscript [i] represents an (optional) epenthetic vowel, which is typically inserted after coda consonants in Brazilian Portuguese. In Kaingang the sequence /ŋr/ is a regular onset in non-derived words. In the word /ren̩re/ ‘two’, derived by reduplication and /ŋ/ infixation, the sequence /ŋr/ behaves clearly as a heterosyllabic cluster with /ŋ/ residing exclusively in the coda of the first syllable, as can be derived from its surface contour properties. Phonetically the word appears as [reŋ̩re] instead of the otherwise expected pronunciation with the triple contour *[reŋ̩f̩re] which is typical for non-derived /ŋr/ sequences, as in *[ŋ̩f̩i8]f̩re] from underlying /ŋ̩i̯re/ ‘to dance’, where intervocalic /ŋ/ is ambisyllabic at the surface. Furthermore, among the clusters of the type muta cum liquida, the sequences /tl/ and /dl/ are relatively rare crosslinguistically. For example, unlike /tr/, /dr/, /sl/, /pl/, etc., the clusters /tl/ and /dl/ are systematically avoided in the Indo-European languages (see Wetzels 1985).

The above-mentioned facts represent some well-known cross-linguistic tendencies, and could easily be translated into OT-like constraints: a) syllables have onsets: create CV syllables, b) align syllable boundaries with morpheme/word boundaries: do not create CLV syllables if some boundary intervenes between C and L, and c) the OCP: avoid sequences /tl/, /dl/ which are composed of segments which are nearly identical with respect to constriction and place. If transposed into a situation of variable word-final consonant deletion, which is a very popular process in variation studies, these cross-linguistic regularities enable us to make the following (probabilistic) prediction regarding the quantitative pattern:

(6) Least deletion before vowel-initial words;
    ↓
    Most deletion before liquid-initial words, except when final consonant is /t, d/ and the initial consonant is /l/;
    any other consonant
The above discussion concerns a simple and straightforward example of the way in which variation studies can contribute to falsifying claims derived from typological studies, which are usually based on invariant linguistic data. Exactly this type of study is undertaken by Labov. See also Guy's contribution, especially with regard to the predicted unwelformedness of the /tl/ and /dl/ sequences.

5. The relevance of variation and change for phonological theories

Although variation and heterogeneity are an inherent part of natural languages, all phonological theories posit the existence of a categorical core (cf. Pierrehumbert 1994: 239). Apparently, the need is felt for distinguishing between core data and quantitative peripheral data, between categorical facts and 'irrelevant' free variation. Nevertheless, the relevance of such a distinction is not self-evident. The area of tension between accepting and rejecting variations as relevant facts is illustrated by the following statement of Labov:

"The basic mode of operation of linguistics and its raison d'être follow from the need to resolve this contradiction: we find in principle and in fact that some differences don't make a difference" (1975: 7).

The use and exploration of data on language use may help to clarify the tension between quantitative variation and theoretical pursuits. This point is raised explicitly in Borowsky & Horvath, who refer to Bailey, whom they praise for his commitment to integrating variation and linguistic theory (Bailey 1973, 1982). Although, as Borowsky & Horvath observe, Bailey is scornful of quantitative data as such, he seeks to explain variation and change in terms of deterministic implicational relations. According to Bailey, quantitative or statistical data only show that specific phenomena are more frequent than other ones, but he goes on suggesting that "what is statistically LESS is lighter (less marked), slower, later and what is statistically MORE is heavier (more marked), faster, and earlier" (1973: 82, with original emphasis). He adds that it may well be the case that speakers use the fact that "a is more frequent than b is more frequent than c" to deduce implicational patterns between a, b, and c. Consequently, quantitative patterning appears to have acquired the status of an interesting, although peripheral phenomenon. From an optimistic perspective, one may interpret Bailey's view of the matter as a suggestion for bridging the gap between data-driven and theory-driven approaches. Less optimistically, one could conclude that quantitative patterns are uninteresting for theoretical interpretation unless they can be raised to the power of implicational scales, specifically for re-
searchers in the tradition of Bailey and Bickerton, among many others, who cling to the belief that all linguistic processes are basically deterministic.

In the preface to the book they edited on symbolic and statistical approaches to language, Klavans & Resnik (1996) conclude that the hostility to quantification has been a dominant property of modern linguistics. Work on natural language processing (henceforth NLP) has been dominated by the theoretical orientation of generative grammar, which is a rule-governed, symbolic system. In the late 1980s and early 1990s, the field of NLP underwent a shift by successfully adopting inductive statistical methods and techniques (including the use of learning algorithms, also popular in language acquisition research). Klavans and Resnik plead for combining symbolic and statistical approaches, for the benefit of both, but especially, of course, for the benefit of the power and quality of language research. Such an approach is not in conflict with Bailey's point of view, in that it supports the search for a proper balance between theory building and data collection.

We will briefly discuss the relevance of variation data for theoretical linguistics under the following headings:
1. the direct relevance of variation data for developing theories,
2. the empirical testing of theories and models,
3. the emergence and structure of language continua.

5.1 The direct relevance of variation data for developing theories

An old problem of historical linguistics and a problem that, except for major publications by Labov (1981, 1994: Chs. 15-18) and Kiparsky (1988, 1995), has received surprisingly little attention, concerns the so-called ‘Neogrammarian controversy’. It relates to the distinction between lexically exceptionless, Neogrammarian sound change on the one hand, and lexically diffuse sound change on the other. Part of the reason for the lack of attention given to this distinction resides undoubtedly in the complexity of the problem.

Labov (1981) presented an overview of the — opposite — characteristics of two types of phonological change. Kiparsky (1988: 399 ff.; 1995) suggested to analyse the distinction between the two types in terms of the rule typology proposed in Lexical Phonology. In Lexical Phonology a distinction is made between lexical and postlexical rules (cf. section 2.1 above). Neogrammarian sound change, which is exceptionless, not conditioned by morphology and typically not neutralizing, results according to Kiparsky from a postlexical rule. On the other hand, lexically diffuse sound change can be analyzed as the simultaneous analogical extension of a lexicalized (structure-building) rule through the
generalization of the rule’s context, accompanied by a markedness reversal of
the feature affected by the rule; the markedness reversal affects the relevant lexi-
con in an item-by-item fashion (cf. Kiparsky 1995). Lexical diffusion is what can
happen to a rule which has lost its productivity (cf. Hinskens 1996). In the light
of the life cycle of a phonological rule, which starts out as a rule of phonetic
implementation, becomes a postlexical phonological rule, subsequently acquires
morphological conditioning and eventually ceases to exist, sometimes getting
lexically ‘fossilized’, it is only to be expected that a dialect feature can be entirely
‘exceptionless’ in one dialect while being lexically diffuse in a sister dialect. Ex-
amples are

• the tensing and raising of /ae/ in dialects of English. As far as American English
is concerned, the ‘rule’ is lexically diffuse in, e.g., Philadelphia and New York
City (Labov 1994), but postlexical in the Northern Cities as well as in the
Midwest (for Columbus, Ohio, cf. Hartman Keiser et al. 1997);
• word-final t-deletion in several different dialects of Dutch. The process is
postlexical in e.g. Nijmegen, postlexical and lexical in Limburg (cf. Hinskens
1992; Hinskens & van Hout 1994), but it has been lexicalized in Afrikaans, just
like in the Black English Vernacular, at least in certain items (cf. Labov 1972:
216).

As far as the contributions to this volume are concerned, Kiparsky’s pro-
posal seems to be relevant to Borowsky & Horvath’s study of the vocalization
of /l/ in Adelaide, which appears to be a case of regular sound change (as pre-

The Neogrammarian controversy constitutes a clear example of an impor-
tant and immanent, but very thorny problem, for a deeper understanding of
which variation studies and phonological theory need each other’s insights. Ob-
viously, data regarding language use will play a decisive role.

There are other areas where variation data are directly relevant for theory
construction and testing. Identifying the role of external structure, the exploita-
tion of databases and the inclusion of probability mechanisms appear important
methodological instruments for the description and theoretical explanation of
language variation. We will briefly discuss these aspects here.

5.1.1. The role of external structure

All types of variation can be highly relevant to the study of grammar.
Sociolinguistic studies show time and again that within speech communities
gradient patterns of variation exist and sometimes even occur abundantly and
tenaciously. Similarly, geographical dialect continua repeatedly confirm the
naturalness of gradual variation. Sometimes, the differences between neighbo-
ring dialects can also give us detailed insights into the type of categorical differences that can exist within structural frameworks such as the diasystem (Weinreich 1954). The concepts of distance or ‘closeness’ are often decisive in discussions about the relationships between linguistic systems. As Smith puts it: “The closer the systems compared, the closer the grammars of these systems should be. Near-identical systems should have minimally differing grammars.” The plausibility of a proposed analysis of a given phenomenon in one dialect can be tested against the way it is implemented in the grammar of a related dialect. As Rose shows in her study of differences between two Gurage dialects, a specific analysis of a part of verb inflection (in Chaha) may look entirely uncontroversial when considered in isolation, but becomes highly suspicious when confronted with the corresponding phenomenon in a neighboring dialect (in this case Muher). In the second part of her study, Lloret presents a unified account of differences in the behavior of the infinitive marker after cliticization in various dialects of Catalan.

As we pointed out in section 3 above, the principle of the positive correlation between linguistic and extra-linguistic ‘closeness’ between language systems is an insight shared by historical linguistics, dialectology and sociolinguistics. The external dimensions of geography, time and social coherence define the extra-linguistic domains within which linguists analyze linguistic structure. The dependency of language variation and change on inherent properties of these main external dimensions can be summarized under the heading of the proximity principle. Inversely, the in-depth study of social structures and communication situations in speech communities could considerably improve our understanding of how and to which extent linguistic differences can be indicative of extra-linguistic structuring.

It can be difficult to disentangle the internal and external forces at work in processes of language change. Analyzing a specific process of change in two closely related languages or language varieties which — and this is crucial — are structurally identical in the relevant respect, is one way to assess the role of extra-linguistic factors. The analysis of the embedding of the change in internal structure on the one hand and the careful manipulation and control of external factors on the other are indispensable when it comes to answering questions such as to which extent similarities in a process of language change in related linguistic systems are motivated either by common — or even universal — internal tendencies or rather by common external factors.
5.1.2 The exploitation of databases

For several studies presented in this volume, existing data collections or databases are used for data (re)analysis. Anttila uses an on-line corpus of approximately 1.3 million words of written standard Finnish to gain insight into the use of the Finnish plural genitive. Labov re-analyzes t/d-deletion data from Philadelphia. Auer presents re-analyses of data concerning dialect-standard continua. More and more, extensive databases are being used in language research. The same practice can be observed in the study of first language acquisition, where the CHILDES databank catalyzed the comparative use, collection, and exchange of databases on spontaneous, spoken first language data (MacWhinney 1995). The use of computer resources almost inevitably leads to the exploitation of more and larger databases. Simultaneously, the quality of the data improves when one incorporates flexible, incremental coding schemes and systematically adds information in several layers (annotations). Data exploration and the comparative use of databases may well become part of the standard research equipment, for theoretical linguists as well.

Another feature worth mentioning is the growing importance of the diversity of the database, i.e. the exploitation of different and multiple datasources. An example of the potential usefulness of different datasources is provided in Labov’s study of syllable structure and late (re)syllabification. Labov considers the hypothesis that “the retention of a final consonant is favored when it can form part of a following onset.” Hence one would expect that the nature of the right-hand environment, and particularly the sonority level of the following segment, affects consonant cluster simplification, as suggested in section 4 above. In production data from the Philadelphia dialect, this hypothesis is tested by looking for phonetic indications of the non-occurrence of word-final t/d-deletion. According to Labov, “most of the evidence is negative. Yet it would be a serious error to argue that resyllabification does not take place: it is not a myth, but a reality.” Perception, specifically the role of the perceptual prominence of allophonic processes in clusters, is considered as an alternative explanation (at least for the unexpected patterning of word-final t/d-deletion in glides). This idea is tested by the use of natural misunderstandings, a new and innovative datasource in the study of language variation.

5.1.3 The inclusion of probability mechanisms

The most popular tool for the handling of quantitative variation is the variable rule, a type of stochastic analysis, which is, for example, being used in the GOLDVARB approach, very popular in contemporary variation studies (see, for
instance, Borowsky & Horvath). One could consider the stochastic analysis as the ideal strategy to formalize linguistic reality, because, as Abney puts it, "properties such as the gradualness of rule learning, the gradualness of language change, dialect continua, and statistical universals, make a great deal more sense if we assume weighted or stochastic grammars" (1996: 24).

The question of integrating variation into the grammar is taken very seriously by Guy, who addresses the following classic question of variation studies: if the output of grammar is categorical, how is orderly heterogeneity, i.e. structured variation, generated? Among other things, he discusses the type of models that posit a 'separate performance grammar.' On the basis of the outcomes of his analyses of the variable process of word-final t/d-deletion, he argues that it is suspect that an alleged performance grammar has effects which are essentially identical to those of the competence grammar — an unjustified case of "theoretical apartheid." Anttila equally considers quantitative variation as an internal part of the grammar, as does Labov in most of his work. A different position is taken by Van Oostendorp and Smith, as well as by Kiparsky (cf. 1988: 386-88 for his discussion of patterns in word-final t/d-deletion in English).

We will refrain here from giving a technical exposé of the statistical aspects involved in a GOLDVARB type of analysis. Instead, we will limit ourselves to pointing out one serious misunderstanding that exists with respect to this method of statistical analysis, especially because it is related to probability mechanisms in phonological theories other than SPE, including OT. Variable rule analysis is in fact a statistical technique for modelling discrete choices — i.e. nominal variables — and the way these choices are conditioned (cf. Sankoff & Labov 1979). As such, the statistical technique is not dependent upon the rule concept at all. On the contrary, the technique of 'logistic regression,' as the GOLDVARB type of method is called in the statistical jargon, can be applied equally well to the analysis of choice mechanisms or strategies for competing grammars. It also means that variable rule analysis does not solve the problem of how to incorporate probability mechanisms in grammars. Linguistics is still in need of productive and creative proposals which have both a linguistic and a mathematical embedding. The availability of all sorts of sociolinguistic data and the variationist expertise with GOLDVARB offer a reasonable starting point for working out more detailed theoretical proposals about choices and conditioning factors.

5.2 The empirical testing of theories and models

Variation data can shed light on the usefulness of concepts which are essential to a given model or theory. The outcomes of some of the studies in this volume
pertain to prosodic constituents. Lloret finds that the categorical differences in the behavior of the infinitive marker after cliticization between several dialects of Catalan are brought about by differences in the domains in which the infinitive marker ‘r’ can be dropped. The data demonstrate that there are good reasons not to dismiss the clitic group, the existence of which has been put to question by many phonologists. In his study of syllable structure and late resyllabification, Labov mainly employs quantitative techniques.

In general, quantitative data can play an important role in evaluating a specific hypothesis or in choosing among different theories. Hinskens (1995) compares two theories of phonological weight on the basis of data regarding the apparent time change in a lexicalized deletion rule in a group of Limburg dialects of Dutch. Anttila claims that his data cannot be explained with a variable rule analysis of the classical type, and shows that the OT analysis he proposes fares much better. More concretely, he shows that the frequencies predicted by his non-ranking hypothesis closely match the frequencies observed in the data.

It is interesting to note that there hardly exists a tradition in linguistics of testing theories on the basis of behavioral data (with the obvious exception of language acquisition research). Perhaps, the best known example in sociolinguistics is Guy’s 1991 study, where he puts his exponential model of t/d-deletion to a test. In his contribution to this volume, Anttila cites Guy on his effort to achieve maximal precision in the explanation of variation: “The development of models that have explanatory value in this sense — models from which one can derive precise quantitative predictions — is one of the fundamental challenges facing our discipline” (Guy 1991: 1-2). If Guy is right, quantitative data could play a vital role in testing theories. A solid tradition of testing still needs to be established in linguistics. How this may be done can be learned from the social sciences, which offer numerous examples of how models with probability components can be tested. To give an example, Lave & March (1993) describe four primary models of social research. They discuss models for the phenomena of choice, exchange, adaptation and diffusion, which, interestingly enough, all relate to communication and language behavior. All models they present incorporate probability mechanisms and for each of them the problem of how they can be empirically tested is discussed. More and better empirical testing in the study of language variation and change clearly would raise the status of the variation data.
5.3 The emergence and structure of language continua

Research on language continua has a tradition that is different from quantitative sociolinguistics. It is a data-oriented tradition linked to code-switching research. This is demonstrated by Auer, who investigates co-occurrence restrictions on emerging and existing continua between dialect and standard language. The term co-occurrence restriction refers to the fact that linguistic variables do not always co-occur freely within a given prosodic domain such as the phonological word, the phonological phrase, or any other relevant linguistic unit. Auer attempts to find out which restrictions hold in a language continuum when two structurally closely related language varieties combine (as in the case of standard and dialect varieties) and what the nature of these restrictions is. Auer confines his study to the linguistic properties of the morphological and phonological continua. In the overwhelming majority of cases two structural principles appear to account for the unidirectional, implicational patterning of the data: (1) the precedence of morphology over phonology, (2) the degree of lexicalization of phonological processes. As Auer shows, in a few cases external forces may overrule internal ones. This observation is reminiscent of Fasold's view of syntactic variation that exists between related languages. In his discussion of pragmatically motivated violations of the Subjacency Condition in Swedish and Norwegian, he reaches the conclusion that "the community of speakers of these two languages has tacitly agreed to allow a sociolinguistic principle to overrule a grammatical principle" (1992: 355). Auer shows that structural consequences of the contact between related varieties can yield essential information about linguistic properties of the systems involved and about the strength of these properties.

6. Conclusion

Of the many problems and research fields that are worthwhile topics for further research, a considerable number were discussed in the previous sections. Nevertheless, we would like to mention two areas that in our modest opinion are of particular interest; both have not received the scholarly attention they deserve. One has been discussed at several places in this introductory chapter, the 'Neogrammarians controversy', in which, as we believe, some of the essentials of phonological variation and change come together. More attention should be given to this research area in the future than has been done until now. Secondly, we wish to emphasize the fruitfulness of the research in the field of language contact. Auer investigates the contact between different language varieties. In
connection with research into the contact between languages, more specifically between phonologies which greatly differ in several dimensions, we mention Singh (1996) who points to the vital place of Universal Grammar in loan phonology. Inspiring research results are also obtained in language studies concerning second generation immigrants, especially in circumstances in which two non-related languages are involved, as shown in El Aisatti (1996), who found indications of language loss in the production, but not in the perception, of Moroccan Arabic of second generation Moroccans living in the Netherlands.

It is obvious that formal theory and the study of language variation and change can enrich each other. Cf. Benincà's (1992: 32, 34) considerations regarding the possible symbiosis of dialect geography and syntactic theory: "Collecting data in function of a theory is advantageous for both elements at play: of course it is indispensable for the theory, which can be checked for being confirmed, improved, corrected or falsified on the basis of the data. But contrary to what one might think, it is also useful for the data, for if one has a theory to check, one is urged to look for a type of data that otherwise might have gone unnoticed or that might have remained completely unknown" (p. 34. Our translation —Eds.).

In-depth structural analysis and linguistic theory can also be indispensable when deciding whether something does or does not constitute a case of quantitative variation. Singh & Ford (1989) reanalyzed five "so-called" variable processes. They found that only one out of these five processes is indeed variable (Singh & Ford 1989: 377), the others result from a lack of descriptive adequacy. On the other hand, variation data can be an excellent means "to cross-check the relevant postulates about a specific language and the relevant claims of a linguistic theory", as Lloret puts it; a similar position is taken by Smith (in section 2 of his contribution). To that end, according to Lloret, the methodology of theoretical linguistics will need to be accommodated such that it no longer solely relies on "systematic and recurrent facts, i.e. the regular data", but also takes "non-systematic but recurrent facts, i.e. the partial regularity of the marginal data" seriously.

However, apart from such general methodological observations and recommendations, it is probably not feasible to reveal in a few statements how an appropriate balance between theory and data can be achieved in the study of phonological variation and change. This is not to be interpreted as a subterfuge. Rather, it is the consequence of the complexity of this specific shared field of research. We feel that the answer to the question of how to obtain a proper balance depends on many different factors, not the least important of which is the specific linguistic object under study.
Each study presented in this volume is marked by an obvious determination to reach a deeper understanding of the facts of language variation and change. Despite differences in both orientation and methodology, all studies show promising results of the scholarly interchange between the insights of theoretical grammar and research in language variation and change. We hope that this volume will contribute to the cooperation between two areas of linguistic research which, while coming of age, grew more apart than might be desirable for a genuine understanding of processes of language change.

Notes

1. For an introduction to the different components of nonlinear phonology mentioned here, as well as for a discussion of their historical roots, the reader is referred to Goldsmith (1990) or Kenstowicz (1994).

2. The way to account for stress as described here is typical of Liberman and Prince (1977). Later developments include Prince (1983), who argued for a ‘grid-only’ theory, Halle & Vergnaud (1987) defending a ‘bracketed grid’ theory, and Hayes (1991), who proposed that languages shape their stress systems by selecting a foot type from a very limited universally determined set.

3. For a discussion of the different characteristics of strictly phonological rules and rules of morphophonology, see Pulleyblank (1986).

4. Whenever we use [α voice], we refer to the two privative features ‘voiceless’ and ‘voiced’ or, in Halle and Stevens’ (1971) terminology ‘stiff vocal cords’ and ‘slack vocal cords.’ For motivation of the idea that [-voice] and [+voice] represent independent monovalent features, see Wetzels (1994) and Mascaro & Wetzels (forthcoming).

5. See McCarthy (1997) for justification of universal ordering based on proper subset relations between constraints.

6. A second possibility to relieve the strict order hierarchy of constraints is to quantify their rank order. Zubritskaya (1997), who introduces a separate production module to capture probabilistic properties of speech, proposes this option. It has to be sorted out whether it is really necessary to stipulate an independent production module in the grammar to account for the patterning of quantitative variation. Another point that remains unclear is how to reconcile weighting with complete ranking.

7. Important exceptions being e.g. the studies of phonological ‘Sprachbünde’ by Jakobson (1930) and Lehiste (1978).


9. Guy argues for an integrated model of language, a model in which competence and performance are not relegated to different modules of the grammar.
10. Instead of pre-nasalized, post-nasalized, and medio-nasalized, we will use the less common terms post-oralized, pre-oralized, and circum-oralized, respectively. This is to distinguish the contour sounds like [m'b], ['m], [b'mb] of Kaingang, which are not derived by nasal spread, from similar ones, which are so derived.

11. Other examples of languages that have triple contours are Karitiana (Arikém family) and, at some shallow level of abstraction, Uruewauwau (Tupi-Guarani family — see Angenot & Sampaio 1996).

12. From the foregoing it should not be concluded that the absence of the triple contrast /p, b, m/ in some language necessarily implies the existence of surface nasal contours. For example Umutina (Bororo family — Telles 1995), among many other languages, has no voiced series and yet no contour segments. Also, the terminology used to define the different contour segments suggests that we attribute underlying (phonemic) status to nasal consonants in Kaingang and elsewhere. This question is, however, far from clear, as argued in Wetzels (1995), but independent from the important fact at issue here, which is that languages which have contour segments of the type discussed very rarely, if ever, have a three-way phonemic opposition between /p, b, m/, etc.

13. Notice that this formulation implies that a CL sequence can be syllabified as a complex onset across morpheme or word boundary if the following syllable is onsetless. This seems indeed to be the case, witness French [k.trom] from /katr onV quatre hommes ‘four men’. Furthermore, languages differ with regard to the prosodic domain within which the avoidance of onsetless syllables is applicable.

14. At present usually referred to as African American Vernacular English or Ebonics.

15. Cf. Auer & Hinskens (1996: 20-21). Compare also the discussion regarding mentalistic (‘language bioprogram’) versus socio-historical explanations of structural similarities between otherwise in many cases radically different creole languages (Bickerton 1984 and following ‘open peer commentary’).

References


— 1981. “Resolving the Neogrammarian Controversy”. Language 57.267-308


