5 Dialectology and Formal Linguistic Theory: The Blind Man and the Lame

FRANS HINSKENS

"The men of experiment are like the ant, they only collect and use; the reasoners resemble spiders, who make cobwebs out of their own substance. But the bee takes the middle course; it gathers its material from the flowers of the garden and field, but transforms and digests it by a power of its own" (Francis Bacon, Novum Organum, 1620/2004, 153)

5.1 The Dialectological Investigation of Dialect Variation, in Brief

Contra the Neogrammarians, who introduced the concept of grammatically blind and lexically exceptionless sound laws—the effects of which can be obscured by analogy or borrowing—late nineteenth-century dialectologists explicitly took the idiosyncrasies of individual lexical items into account (as per Jakobson's axiom "in reality each word has its own particular history" (Jakobson 1908, 6)). This tendency to concentrate on isolated linguistic forms is a manifestation of atomism, that is, the inclination not to embed in the grammar the phenomena under study. Findings are typically not interpreted in the context of existing linguistic theories, which has led to the marginalisation of dialectology in the wider field of linguistics.

A related (disputable) trait of traditional dialectology is its tendency not to embed the phenomena being investigated in the verbal repertoires of the speakers and the speech community under study. The sociodialectological approach to dialect variation does provide this possibility, but with its tendency to treat linguistic variables in isolation rather than properly embedding variation in language structure, the sociolinguistic approach to language variation and change unmistakably inherited some of the features of dialectology. Related to this is the fact that the construction, testing, and revision of theories of language change does not appear to have been given much priority in the sociolinguistic approach to dialect variation, the biggest and probably most important exception to this generalization being the work of William Labov (1994, 2001, 2010).

5.2 On the Formal Theoretical Investigation of Dialect Variation

Muysken's (2014) "brief history of 20th century linguistics" only mentions "de Saussure: structure, Chomsky: deep structure." Generative theory characteristically focuses on the meta-grammar, that is, the parameters, principles and constraints that are assumed to govern the way grammars work. This hypothetical meta-grammar is claimed to be biologically programmed, and to manifest itself in a necessarily abstract way as the common core of natural languages. The meta-grammar, usually referred to as Universal Grammar (UG), can be seen as a kind of decision tree, whose nodes are called parameters. Every individual language can be uniquely defined as a specific constellation of choices made for the respective parameters. The decision tree representing UG is built in such a way that it reflects an essential trait of natural language, namely modularity, that is, the fact that the different parts of grammar (syntax, phonology, etc.), though interrelated, are internally autonomous to a certain degree. Certain instances of linguistic change can be a side effect of the modular organization of language, which may make it possible for abstract principles to interact.

Not all adherents of formal theory seem to be sufficiently aware of the fact that "heterogeneity and variation are not abnormalities but part of the normal condition of language" (Kiparsky 1988, 370), although there is a growing interest in geographical (inter-systemic) and quantitative (intra-systemic) variation among formal linguistic theorists. There is, however, no standard view of language variation in formal theoretical models. Several generativists have tried to understand the smallest differences between dialects as manifestations of universal principles underlying the organization of language systems. The smallest difference (at the level of the language as a system shared by the members of a community; cf. Chomsky's (1995) E-language) is thus explained on the basis of the highest common denominator (I-language, language as a cognitive commodity). Dialect features are thus sometimes explained as different instantiations of language universals or as instantiations of different language universals. But how does it work concretely?

From the 1980s onward, generative syntax was dominated by Principles and Parameters (P&P) theory, which looked at UG as an invariant system of abstract principles, some of which permit at most a specified degree of variation within a given language. Originally, this notion of variation referred to differences between languages (macro-parametric variation), but the approach came to be applied to cross-dialectal variation (micro-parametric variation). From this line of research, deeper insights into the universal set of parameters were expected, in terms of their form as well as the substantial variation they allow.

Whereas in P&P variation resides in the computational system, in the Minimalist theory of generative syntax variation is located in the lexicon: all relevant parameters are encoded in the feature specification of individual items in the lexicon. Barriers sketches a specific model to account for syntactic micro-variation, which "avoids the tendency found in much generative work to explain syntactic variation by syntactic principles exclusively" (2013, 24). In Barriers' model there is a role for cognition, body ("brain, oral tract, etc.") and society ("groups, contact, history, etc."). This is in line with the Minimalist view, according to which variation is not only located in the syntactic module, but also in other linguistic and non-linguistic dimensions. Barriers' model also in principle has room for frequency of usage and conventionalization, which in Barriers' perception are relevant to the question of "why certain syntactic variables are sensitive to sociolinguistic specialization while others are not" (2013, 23).

While the concepts of rule and derivation had faded into the background in nonlinear phonological theory, they were entirely abandoned in Optimality Theory (OT). In OT, a set of constraints is assumed which determine the way in which the surface structure is allowed...
to deviate from lexically-underlying representations. In principle, all constraints are both universal (although their relative importance is language-specific) and "soft," as they can all be violated by conflicting constraints. The only generative capacity of the model resides in a function labeled GEN ("generator"), which projects an unlimited set of possible output candidates from a single lexical input form. All output candidates are rated according to their success in complying with the ranked constraints set; the candidate that best satisfies the relevant high-ranked constraints is selected as the optimal one. (It is never perfect, since it will always violate certain relevant lower-ranked constraints.) Constraints come in three kinds. Faithfulness constraints require the phonetic output form to be maximally identical with the underlying form, whereas markedness constraints make the phonetic output conform to prosodic and articulatory requirements. A third type of constraints guards the alignment of prosodic and grammatical structures.

In OT analyses, variation resides in the constraint ranking. Inter-systemic variation is commonly described through subtle differences in constraint ranking (e.g., Herrgen’s 2005 analyses of cross-dialectal variation in word-final [t] deletion in modern southern and southwestern dialects of German). In this "multiple grammars" scenario, every variety is viewed as a categorical grammar of its own. Intra-systemic variation is typically analyzed as partial ordering of constraints. The oldest variant of this type of analysis is based on the concept of floating constraints. This theoretical variant (represented in e.g., Nagy and Reynolds 1997) has meanwhile disappeared from the stage. The second variant is unranked constraints, the situation in which several constraints entirely coincide, producing multiple optimal candidates as output, although not every logically possible ranking emerging from this scenario will necessarily yield another winner (e.g., Anttila 1997, 2002). A closely related way to capture intra-systemic variation is by associating a relevant domain of the language to a different constraint ranking, a concept known as "co-phonology." The third variant is essentially different from the other two.

In the case of continuous ranking (Zubritskaya 1997) and stochastic OT (Boersma and Hayes 2001), the constraint ordering is not discrete but normally distributed, and the tails in the distributions of neighboring constraints overlap, which accounts for the variation (see Figure 5.1). The range of the intervals between neighboring constraints can vary.

Harmonic Grammar (HG) is a family of models in which constraints, which can be language-specific, are assigned a specific weight, rather than an absolute position in the constraint ranking (Smolensky and Legendre 2006; Pater 2008, 2014); lower-ranked constraints can "join forces" to match the violation of higher-ranked constraints. Within HG the same range of theoretical variants apply as in OT: a stochastic HG model is discussed in Boersma and Pater (2016), for example.

A mechanism known as Lexicon Optimization guarantees that input forms do not necessarily need to be identical in different dialects. Nevertheless, unlike Minimalist syntax, formal phonology does not operate on the basis of the assumption that variation resides in the lexicon. In all mainstream theories of phonology, variation is located in the computational system, whereas in OT and related declarative models it lies in the constraint ranking.

![Figure 5.1](image)

**Figure 5.1** Constraint ranking on a continuous scale with stochastic evaluation (adapted from Clark 2005, 213).
Generative studies of language variation are often based on data from existing grammatical descriptions, an approach which is known as the "armchair method." Alternatively, and even more in line with generative research traditions, the data consist of intuitions concerning the well-formedness of linguistic output, the source of which is typically a single native speaker (commonly the researcher, a methodological decision which is understandable given the fact that syntactic phenomena in particular tend to occur infrequently in spontaneous speech). This is what is referred to as "internal evidence." In this view, "external" evidence concerns data from actual language use, and diachronic data—as well as data from geographical, social and stylistic variation—are relegated to the domain of E-linguistics. In generative studies of language variation, explanations are preferably I-linguistic in nature, and they are never historical/diachronic.

5.2.1 Weaknesses

Markey (1986) distinguished three "explanatorily adequate basic epistemologies" of language change:

I. People (as communicative beasts) do things to language systems [...]  
II. Languages qua systems do things to people (The Whorf-Sapir Hypotheses).  
III. Languages qua systems have a life of their own apart from people.  

Markey (1986, 16-17)

The first perspective can be termed the materialistic position, as opposed to position II, which is idealistic in nature. This latter position may make people deliberately change aspects of their language use (e.g., avoiding sexist language use) with the aim and conviction that doing so may cause others to change their ideologies. Position III, finally, can be labeled the biologicist view. Language variation and change are in part social phenomena. Hence, a type I epistemology is called for. Formal theories largely fall under the type III epistemology, and they are claimed to be psychological theories. Generative theory conceives of language as a cognitive object. It operates in a fundamentally different dimension than the socio-dialectical approach to language variation, since "the central dogma of sociolinguistics is that the community is prior to the individual. [...] language is seen as an abstract pattern located in the speech community and exterior to the individual" (Labov 2010, 7). The ultimate object of sociolinguistic research is the community grammar; a speech community is defined as a group of people who share the same set of linguistic norms. Apart from well-formedness judgements (as data), norms do not play a role in formal theory, let alone sociolinguistic prestige.

Until recently there were only few theoretical variation studies in which the position of the language variety under scrutiny seemed to matter. In this respect the world seems to be changing, since some researchers who approach language variation from a formal theoretical perspective have come to realize that they abstract away from the fact that dialects "change fast, due to rapid changes in their sociolinguistic context—factors involved include but are not limited to the lack of codification, the erosion of the dialect landscape due to social changes, and the low status of dialect varieties" (Bennis and van Oostendorp, 2013, 676).

The methodological construct of the "ideal speaker/hearer" is diametrically opposed to the concept of "inherent variability," the insight that variation is a deep property of (living) language. Moreover, most dialect speakers have (at least receptive) knowledge of related varieties at their disposal. As King (2013, 448) points out, "it is often difficult to tease out what sort of knowledge the speaker is drawing on. Is knowledge of the dialect in question entirely separate from knowledge of more prestigious varieties?"
Generative theory is strictly synchronic. That is not incompatible with language variation as such, but it does not align with variation as a synchronic reflection of an ongoing process of language change.

Formalists often call dialectological and sociolinguistic research "descriptive." In generativist rhetoric this term is often used as an invective, whereby studies that do not address any formal-theoretical questions are dismissed. In Chomsky's philosophy, explanatory adequacy is determined through "evaluation metrics," including economy and simplicity, both of which are relevant to language acquisition. In this regard, however, language change is out of the picture, since it is the diachronic issue par excellence.

Formal theory is deeply committed to categorical reasoning. "A probabilistic approach," King (2013, 452) argues, "would go against the basic assumption [...] that grammar and use are modularly distinct." Gradience is problematic and relegated to phonetics and diachrony. Linguistic systems are conceived as closed systems without, as it were, "frayed hems." The restrictive operationalization of the concept of dialect variation as cross-dialectal variation is also problematic. Like all linguistic phenomena, cross-dialectal variation is in principle approached as the presence or absence of a discrete feature. The only refinement concerns potential differences in the structural conditioning of the feature at issue in the dialects concerned.

There is also an asymmetry in accessibility: most "theory-driven" studies are only accessible to those who are sufficiently familiar with the theoretical matrix. Given the expiration date of many generative proposals, it is by no means self-evident that acquiring the necessary familiarity is worthwhile. By contrast, most studies of a specific phenomenon in a specific language area that are not primarily theory-driven are accessible to every linguist who is sufficiently familiar with traditional linguistic and phonetic terminologies and classifications (although these are being gnawed at in modern typological research, as in e.g., Haspelmath, 2007). The obvious bridge across this divide is the construction of generally accessible databases of linguistic data; a sustainable approach of this sort enables "computer-aided armchair linguistics" (Fillmore, 1992). More on this recent development, in which many theoreticians actively participate, is given below.

5.3 Why Should They Collaborate?

Like Bacon's (1561-1626) empiricist "men of experiment," who collect data without altering them, dialectologists tend to be at best theoretically shortsighted. The generativist "reasoners" weave theoretical webs out of their own material; many of them are methodologically challenged, especially data-wise. Yet together they can make headway. Together they can gather the best material and convert it into something that is superior to the original material.

There are several ways in which the study of (both inter- and intra-systemic) dialect variation can profit from the types of theories that have been developed in formal linguistics. The following considerations can be made with regard to specific dialect features:

- the theory can inform the decisions underlying the selection of dialect features to be studied, although the selection will typically not solely be based on considerations of a strictly linguistic nature;
- in-depth structural analysis, whether or not in a formalized fashion, can be indispensable when deciding whether or not a phenomenon constitutes a case of quantitative variation. Variation in itself is usually linguistically structured, and is necessarily part of the larger structure of the linguistic system. In these respects, formal theory can offer analytical
depth—as Rooryck (2014) puts it, “theoretical linguistics renews the toolbox of descriptivists”—although dialect research usually has more to offer than mere description. With respect to theory as a tool or heuristic device, Labov (1997, 146–147) ponders, “The chief value of formal models, I believe, is to draw the attention of empirical investigators to undetermined relationships and unanswered questions that they may have overlooked. Once such questions have been raised, and clearly formulated, the chief purpose of the model has been achieved. It may then fruitfully be dissolved and replaced by other models, which will reveal new aspects to be investigated. The cumulative character of the enterprise lies not in the models, but in the gradual development of our knowledge through further inference and investigation.”

Linguistic analysis can counteract the “atomistic” approach to dialect features that is typical of dialectology in that its practitioners have had a tendency “to treat linguistic forms in isolation rather than as parts of systems or structures” (Chambers and Trudgill, 1998, 33). Linguistic analysis, indeed, can help to embed the features in the structure of the grammar at large, by specifying the conditions on their distribution, linking the features at hand to related phenomena, and so on.

- sometimes theory can help elucidate the *raison d’être* of a specific dialect feature, that is, of why there is variation at this specific point in grammar (insofar as it was not borrowed from another dialect or language). It can, in other words, help to reveal the “innere Kausalität” (“inner causality”; Moulton 1961) of a phenomenon, and thus help to answer questions regarding the actuation and constraints, the transition, and the embedding of a change in the linguistic structure (Weinreich, Labov, and Herzog 1968; Labov 1972).

With respect to the study of changes in the usage of a dialect feature:

- an explanation of the phenomenon at issue may be the basis for predictions about possible future changes (say, the fate of a given dialect feature in the course of processes of structural dialect loss), provided that the account is grounded in a general theory;
- formal theory can sometimes also provide interpretable insights into the structural relationships between different features (elements, structures, or processes) of a dialect. The nature of these relationships is part of what distinguishes language varieties from each other, and they can play a role in processes of linguistic change. For example, one of the features which set apart the Ripuarian dialects of Dutch (spoken in the extreme southeastern part of the language area) is dorsal fricative deletion (DFD). In lexical morphemes with rhymes consisting of a short vowel followed by a dorsal fricative + /t/, the fricative can be deleted. As a result of compensatory vowel lengthening, non-low vowels develop a schwa offglide. Examples are:

  (1) nau(t)  ~  nat (st. Dutch)  nacht  “night”
  (2) ze:at  ~  zegt  “says”

(2) shows that DFD applies to inflected forms, but it does not apply to derived forms, as in:

  (3) jawi:ct  ~  jawi:ct  gewacht  “weight”
  jawi:cti  ~  jawi:sti  gewichtig  “weighty”

The deletion of word-final [t] (WFD) in clusters is one of the most widespread non-standard phenomena in the Dutch language area. In Ripuarian dialects of Dutch, variable
WFtD is fully productive following obstruents, and it can affect every word-final [t], that is, /t/ and /d/ (Hinskens 1992, 244–248). Examples are:

(4a) *kasept ~ kasep recept “prescription”

(4b) *wirokt ~ wirok werkt “works”

(4c) *a lef knik ~ a lef ein lief kind “a sweet child”

WFtD applies to [t], which is part of a lexical representation (4a), as well as to affixal [t] (4b), (4c) contains a word-final [t] in a grammatical function that is lacking in standard and other dialects of Dutch; here [t] is as pronominal affix marking neuter nouns that are not lexically expressed (Hinskens and Muysken 1986; Hinskens 1992, 180–181). This [t], too, can fall prey to WFtD. (4b, c) show that WFtD is blind to morphological structure, which is one of the properties of post-lexical processes.

In view of the coexistence of DFD and WFtD in Ripuarrian dialects of Dutch, we are presented with the question of what happens to Ripuarrian variants of items such as nacht “night” and licht “light,” which in principle constitute input for both phenomena. DFD and WFtD are disjunctive: they cannot simultaneously apply on the same word. In words with this structure, DFD and WFtD bleed each other (Koutsoudas, Sanders, and Noll 1974), that is, one destroys the input of the other. Traditionally in Ripuarrian dialects, words of this type systematically show DFD; both processes apparently apply in accordance with the “Elsewhere condition” (Kiparsky 1973). This condition says that, whenever a given form obeys the structural description of two different rules, the more specific rule applies. In such cases, the more general rule is blocked, but it does apply elsewhere. Indeed, the structural description of DFD, viz. /Vct/, forms a proper subset of that of WFtD, viz. [[son][t/d]]low. In Ripuarrian dialects, the latter process applies exceptionlessly elsewhere. Moreover, as a lexical rule, DFD will apply before post-lexical WFtD.

If processes of dialect leveling proceed gradually in linguistic respects, then this analysis would lead to the prediction that DFD, which is conditioned on several levels, will be given up in favor of WFtD, which is automatic and exceptionless, and, therefore, easier to acquire, and hence more resistant. For the dialects at issue this amounts to the prediction that variants such as [nacht], [licht], “night,” “light,” will gradually be replaced by variants such as [nacht], [licht].

Recordings of both elicited and conversational dialect use were made using a stratified random sample of 27 male speakers of the Ripuarrian dialect of Rimbou; the speakers represent three different age groups. For DFD each of these speakers realized, inter alia, four words (a) in which in the Ripuarrian dialects of Dutch form part of the potentially DFD-susceptible set, (b) in which /Vct/ is in absolute final position, and (c) in the realization of which no other dialect features play a role, such that linguistic covariation effects are excluded. In the realization of these four words, WFtD is in principle possible. Each individual realization can be characterized as (DFD, WFtD). (1,1), that is, application of both DFD and WFtD is impossible. With respect to the three possible realization types: (1,0) is a realization of the type [nacht], whereas (0,1) is a realization of the type [nacht], (0,0), the realization type in which neither the dorsal fricative nor the final [t] are deleted, did not occur in the recorded data. This part of the database hence consists of 108 observations (4 words × 27 speakers): 4 realizations are lacking. The distribution of the remaining 104 realizations is summarized in Table 5.1.
Table 5.1 Findings: loss of DFD, territorial gain for WFD.

<table>
<thead>
<tr>
<th>(DFD/WFD)</th>
<th>Older</th>
<th>Middle</th>
<th>Younger</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,0)</td>
<td>35</td>
<td>30</td>
<td>31</td>
<td>96</td>
</tr>
<tr>
<td>(0,1)</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>33</td>
<td>36</td>
<td>104</td>
</tr>
</tbody>
</table>

There is an apparent-time decrease in the use of DFD variants, a decrease that is completely absorbed by WFD. According to the outcomes of a significance test ($\chi^2=4.95$, df = 2, .05 < p < .10), the data show tendential support for the hypothesis that DFD undergoes dialect loss but WFD does not. On the contrary, WFD gains from DFD's loss of ground.

Linguistic analysis is indispensable when it comes to answering the question of whether, and to what extent, cross-dialectal similarities in processes of language change are motivated either by (universal) structural tendencies or rather by common "external," for example, sociolinguistic, factors. The study of specific dialect features can, in turn, deepen formal theory in several ways. Since all language varieties must conform to universal theories of grammar, the patterning of specific instances of variation can serve as a test for formal theories. With the exception of language acquisition research, there hardly exists a tradition in linguistics of testing theories on the basis of behavioral data. Perhaps the best-known example in sociolinguistics is Guy's (1991) study, in which he puts his exponential model of t/d-deletion, and through this model some central tenets of Lexical Phonology, to the test. Properties of specific dialect features can disprove, modify or corroborate a proposed analysis. As an example, in his Generative fonologie van het Nederlands [Generative phonology of Dutch] from 1981, Geert Booij makes specific claims regarding vowel reduction and (re)syllabification. One of Booij's claims is given in (5), as follows:

(5) 'The application of vowel reduction may result in resyllabification of the word undergoing this rule. In e.g. “pastoor” [“priest”, with the stress on the second syllable – FH], the [s] is ambisyllabic, because it is preceded by a short vowel. But after reduction the [s] is preceded by a schwa, as a result of which [s] now belongs to the second syllable only.' (Booij, 1981, 151 – my translation, FH).

The ambisyllabic nature of this claim is supported by the observation that (except for in a few interjections) lax vowels never appear in open syllables in Dutch. However plausible this might be, standard Dutch does not provide any substantive evidence for the second part of Booij's claim, although dialectology can help. A phenomenon distinguishing most Limburg and Ripuarian dialects from other varieties of Dutch is the palatal realization of /s/ and /z/ in absolute syllable-initial position before a consonant. An informal representation of this correspondence rule is given in (6):

(6) standard language     dialect  
[s]C [-] [ʃ]C  
[z]C [-] [ʃ]C  

The examples in (10) serve to illustrate this.

(7) standard language     dialect  
[s]mal [ʃ]mal' "tight, narrow"  
[z]wart [ʃ]wart "black"  

please align
So in the dialect of Dutch spoken in Rimburg, for instance, one never finds:

(8a) * boˈstɪˈtɔr  “board of directors,”

which is identical to the standard variant. Because of (6), the dialect has

(8b) boˈstɪˈtɔr

Similarly:

(8c) * goˈstɪς  goˈʃtɪς  “mental home”

However, in coda position following a vowel, an “etymological” /s/ remains [s], for example, in

(9) * meˈʃtər  meˈʃtər  “master,” “teacher”
   * poʃta  poʃta  “paste,” “pasta”

In the case of pastoor, palatalization is excluded in the variant with the unreduced vowel, thus

(10a) * poʃˈtʊər  poˈʃtʊər
   poʃˈtʊər  poˈʃtʊər

and, likewise, in the variants of the toponym Maastricht (with stress, again, on the second syllable):

(10b) * moˈʃtɪς  moˈʃtɪς
   maˈʃtɪς  maˈʃtɪς

The fact that /s/ can be palatalized when the preceding vowel is reduced implies that in these dialects it is no longer part of the first syllable. At the same time, these data imply that in the variants with the full, unreduced vowels the [s] does not belong exclusively to the second syllable. The possibility that in the latter case it is only part of the first syllable is, however, not excluded as far as these dialects are concerned. In this case, dialect data can be adduced in favor of a theoretical claim.

The study of (apparent time or diachronic) changes in the usage of a given dialect feature can adduce evidence pro or contra specific theoretical proposals. For instance, processes of chain shifting (as are e.g., taking place in North American English and Austrian German) are problematic for OT, because they involve opacity. Opacity is troublesome for OT, which is basically a “flat,” non-derivational model. In OT the various phases in a chain shift, which present themselves as many different sound changes, would not apply serially but in parallel, with the effect that all relevant vowels would be realized identically, viz. as the output of the last change in the chain. Within OT, several proposals have been advanced to solve these problems, but some, for example, “Sympathy” (McCarthy 1999), did not survive and the remaining ones—including “Turbidity,” which says that the derivation of a segment is traceable in its representation—(Goldrick 2001) are disputed.

Genuine synergy between dialectology and formal linguistic theory is visible only rarely, but there is at least one such case in the recent history of the study of the trajectory of sound change. With respect to the intensive spread of linguistic change, the so-called “Neogrammarians controversy,” that is, the distinction between Neogrammarians sound change (which is phonetically gradual, lexically abrupt, and exceptionless) and lexically
diffuse sound change (which is phonetically abrupt and lexically gradual, and hence not exceptionless; Scheutz 1988, 1608), has been the subject of a particularly fruitful exchange of ideas between Labov (1994) and Kiparsky (1995).

5.4 How Can They Collaborate?

Modern dialectology positions variation along social or social-geographical dimensions, the methodology is usually quantitative, and the phenomena and their conditioning are perceived to be probabilistic. In formal theory, by contrast, the focus is rather on psychology, the method is inductive, and languages are conceived deterministically as rule-governed systems. In view of the "yin and yang" nature of language (Bailey 1982), that is, the fact that language has both neuro-biological and socio-communicative sides, either perspective is legitimate and at the same time reductionistic, and many members of the two camps are aware of this. Referring to "autonomous phonology" as "any approach to phonological investigation that assumes that the object of phonological enquiry can be studied in its own right, relatively independently of the study of factors such as the social context in which the speakers are located," Carr (2000, 74, 84) claims that "a fully autonomous phonology is unsustainable, since the data to be accounted for cannot be divorced from social context and are inherently variable." The latter aspects are, to a large degree, matters of convention, and there is a non-obvious "relation between sociophonetic variation and UG as a natural object" (2000, 96). What role, then, does the "natural" play in the "conventional?"

The answer to this question calls for both a theoretically informed approach and more refined analyses of better data than generative theory traditionally builds upon. The types of data need to be broadened, even in those cases in which the empirical basis is confined to "internal" evidence. According to Lloret (1997, 201), regular—that is, systematic and recurrent—data "are the usual concern of formal linguistics, but the marginal [non-systematic but recurrent] data also hide relevant facts of the language and thus should be the explicit concern of formal linguistics too." Lexically diffuse sound change typically results in that kind of data.

Obviously, sharing (rich) data can bring together dialectologists and theoreticians. Ideally, digital fieldwork recordings of relatively natural speech are stored in sound archives that are preferably parts of databases, which are connected with similar databases for related dialects, enriching all databases with standardized labeling of metadata will enhance research into larger geographical domains. Once other, mainly older and comparable, and typically questionnaire-based data have also been made electronically available, cartographical tools, adapted so as to enable the mapping of both older and newer data, will facilitate genuinely diachronic research.

Simultaneously, the types of analysis need to be refined. In formal theory the language use of non-linguists has meanwhile been discovered, which has improved the reliability and the generalizability of the findings. In less than two decades, micro-variation has developed into an important object of theoretical research. Now variation in the social dimension and (in the wake of the work of pioneers such as Lightfoot, Krch, and Kiparsky) in the time dimension will have to find its way into theoretically-informed analyses.

Both for inter- and intra-systematic variation, ongoing processes of dialect leveling have far-reaching consequences regarding the question of the level at which the data can be analyzed. It is now time to abandon the position that a given phenomenon can be treated as a discrete variable, and instead to approach every non-standard phenomenon as a continuous variable. Researchers need to realize that dialect reality has become a fragmented reality; this complicates research, but approached from this angle, the findings will be more valid. And generative theory is in principle compatible with contextualized data, quantitative methods, and probabilistic relations.
Theories and models should not be expected to be able to account for 100% of the variance in the data, although the overall model should of course be as parsimonious as possible. As there are often competing internal tendencies, simple yet highly predictive accounts of dialect variation are probably an illusion. The claims deduced from formal theory should be considered as probabilistic explanations (cf. Kiparsky 1972, 222) or as predictions regarding favoring or disfavoring constraints.

5.5 How Have They Collaborated so Far? And What Has It Achieved?

For years, the Competing Grammars model introduced by Kroch (1989) has been a leading generative approach to (morpho-)syntactic variation. In this model, which is closely related to the “multiple grammars” scenario of early OT approaches to phonological variation (cf. section 2 above), on an abstract level, there is no such thing as intra-systemic variation. All variation results from the availability of two categorical grammars that differ in the setting of one particular parameter.

A selection of thorough studies inspired by formal theory and focused on syntactic variation in different languages is presented in Cornips and Corrigan (2005), a volume in which P&P theory is well represented. Yang (2003) demonstrates an elaboration of P&P theory which is explicitly constructed to deal with quantitative variation.

In generative syntactic approach to micro-variation, the accent has meanwhile shifted to Minimalism (see above). There even exists a specialized international journal that is devoted to the Minimalist-informed study of (micro- and macro-) variation, entitled Linguistic Variation, now in its fifteenth edition, and edited by Jeroen van Craenenbroeck.

Van Craenenbroeck (2014) calculates the differences in geographical spread between linguistic variables pertaining to verb cluster orders in the SAND data (see below); for each pair of verb cluster orders, a distance is calculated based on their geographical spread. On the resulting distance matrix, Van Craenenbroeck carries out a Multiple Correspondence Analysis. On the basis of the resulting dimensions, the author tests a set of syntactic micro-parameters, which had been proposed in the theoretical literature on verb clusters. It turns out that there is a clear empirical basis for several of these parameters.

Elaborating on earlier work with Smith (Adger and Smith 2005), Adger (2006) presents a model based on Minimalist syntax and Distributed Morphology in which the interpretation and checking of “uninterpretable” (i.e., purely grammatical) features allow the quantitative modeling of the patterning of instances of morpho-syntactic variation, which fits empirically established distributions astonishingly well. This approach has been refined by Nevins and Parrott (2010), who claim that their proposal is compatible with the original variable rule model. Taking an alternative tack, Bresnan (2002) has proposed OT models in which low-ranked and thus inactive constraints nevertheless play a role in shaping quantitative morpho-syntactic variation.

For the study of phonological variation, Guy (2011) is a concise yet well-documented and highly readable historiography, which also summarizes recent work. Kostakis (2010) proposes a model based on “classical” OT in which language change is a matter of constraint demotion to account for intra-systemic variation. To this end, he introduces the concept of “Vestige constraint,” a phantom of a demoted constraint that behaves in an output-output fashion “as a sort of a receptacle for variants uttered in a linguistic community” (2010, 2476). Cardoso (2007) models developmental data in a Stochastic OT framework, whereas Pater (2014) presents an HG analysis of Canadian Raising; his model includes language-specific constraints. The Harmonic-Grammar weight associated with a given constraint is increased for every candidate where it is satisfied. The candidate with the highest sum (“Harmony”)
The impact of HG and stochastic OT seems to be growing. These frameworks serve to model both language acquisition and intra-systemic variation, typically in tandem.

In the meantime, especially in the study of phonological variation, the paradigm debate concerning cognitivist models (Usage-based Phonology, Exemplar Theory, etc.; see below) is becoming more and more important, making the theoretical part of the field richer and deeper.16

5.5.1 Net Results and Added Value

In general terms: what has the collaboration achieved so far? And what does the net balance look like?

On the negative side, the marriage between (socio-)dialectology and formal theory does not seem to be an utterly happy one, not because they still do not know each other very well—that may not necessarily stand in the way of conjugal bliss—but because both partners distrust each other and spend little time together.

On the positive side, the past 15 years have seen a sharp growth of theoretically inspired large-scale dialect geography projects. The digitization of linguistic research doubtlessly plays an important role, as does the fact that powerful servers and fast internet connections make the data broadly accessible to other researchers. The digitization consists of a chain of major technical improvements that catalyze the speed and quality of the collection and analysis of large amounts of data.

Also in theoretical syntax circles it has become clear that informally gathered intuitions are not always a satisfactory basis for syntactic theorising. It is also clear that experimental methods are sometimes necessary and may provide richer data than informal methods. It is clear too that corpus data can be valuable in various ways. Above all it is clear that questions about data are more important than is sometimes assumed” (Borsley 2005, 1479). This quote is from Borsley’s introduction to a thematic issue of Lingua (vol. 115(11), pp. 1475–1666) entitled “Data in Theoretical Linguistics.” In the eight contributions, themes such as gradience, soft constraints, and magnitude estimation are discussed. There are further encouraging signs of this type, such as Perk’s and Rosenbach’s (2007) collection of papers (cum extensive discussion) on types of evidence and argumentation in (morpho-)syntax and, for phonology, the overview of actually used and potentially relevant data types in van Oostendorp (2013). Most of these texts contain pleas for a diversification in the types and sources of data studied for theoretical purposes, which points at a general contemplation among theoreticians of the empirical basis of their work.

In phonological theory the insights have emerged that variation in the sound component of language has a range of possible loci, can accordingly be constrained in different ways, and thus cannot be approached in a standard way (Hinsakens 1998). Generally, the degree of awareness and consequently the degree of manipulability of a given phenomenon increases in accordance with the following cline: phonetic implementation < postlexical processes < phonological rules < lexicalized sound change. The productivity of sound changes decreases along the same cline. Differences in the position of a sound change on this cline can be conceived as a type of gradience, and can cause inter-systemic variation beneath the surface (Ramsammy 2014). Similarly, syntactic variation has a wide range of potential “origins” (UG, psychology, physiology, society), as Barbiers (2013) argues. Here, the degree of awareness and thus the degree of manipulability of a phenomenon generally increase in accordance with the following cline: realization of syntactic structures (e.g., doubling of Wh-elements in syntactic dependencies) < morphosyntax (e.g., verbal inflection) < lexicalized syntactic phrases (such as verbs and pronouns).

From the cross-linguistic comparison of findings from sociolinguistic research it appears that variable phenomena which occur in different languages are often influenced in the same
way by the same or similar linguistic factors (cf. Tagliamonte 2011). This implies that the internal conditioning of language variation can give important indications of possible universal constraints. A closely related insight is what Bresnan et al. (2001) label "stochastic generalization," that is, generalizations that are categorical for some language(s) but probabilistic in others. Sometimes the contextual conditioning of a variable phenomenon goes in the direction of a complementary distribution; it is easy to imagine that, say, allophony can be the end result of a gradual diachronic development, in which the preference of a variant for one type of context (e.g., more WFD before a consonant) and the dispreference for a complementary context type (less WFD before a vowel) on either side grew into a categorical distribution (no [t/d] before C, always [t/d] before V, an alternation that is comparable in some respects with one of the few aspects of French liaison that are fully understood). Teleologies of this type cannot easily be made visible in a strictly synchronic approach, but can be made so using an apparent-time approach, and more definitely still on the basis of real-time replications. At the same time, stochastic generalizations are another reason to include intra-systemic variation in the grammar; cf. "[i]f the canonical ordering and the obligatory cases are part of the competence grammar, but the quantitative preference is treated as performance, then a larger generalization is lost" (Wasow 2002, 139).

5.6 The Roads Ahead

The last few decades have seen the rapid development and spread of "cognitivist" approaches to language, including Cognitive Grammar (Goldberg 2006) and, for the sound components, Exemplar Theory (ET; Johnson 1997) and the closely related Usage-Based Phonology (Bybee 2001). In these paradigms, which are conceived as alternatives to generative theory, lexical items and their properties (regarding form, function and usage, and including all sorts of type and token frequencies) have a pivotal position. Each realization (or "exemplar") of an item, with all its phonetic, semantic, and extra-linguistic attributes, is supposed to be stored in memory, where it is connected with other items and their many properties. From this huge multi-dimensional memory cloud, grammar emerges from the bottom up. The geographical and social distribution of the tokens (taking the place of variants) are part of the stored extra-linguistic properties.

In the cognitivist paradigm, corpora play a central role, not least as the source of lexical frequencies. For the study of dialect variation, cognitivist approaches have attractive sides. First, variability is assumed to be represented directly in memory in the shape of concrete exemplars, which are assumed also to contain social-indexical information (e.g., Docherty and Foulkes 2000). Second, the model is not based upon deterministic principles, but rather probabilistic ones; as such, it seems to match the nature of most documented instances of language variation. Third, just like adherents of cognitivist approaches to language, many sociolinguists studying language variation reject the analytical distinction between diachrony and synchrony that is applied by adherents of formal theories. Cognitivist approaches can be implemented relatively straightforwardly for the study of dynamic aspects of language such as acquisition and processes of language change.

Barbiers (2013) sketches a Minimalist model for syntactic variation that can in principle accommodate frequency effects. In generative phonology (including OT and HG), models that allow room for frequency effects are also being developed. For example, in connection with lexicalization, which is typically the last phase in the life cycle of a sound change, token frequency can be argued and demonstrated to play a central role (Hinskens 2011; Bermúdez-Otero 2012). With regard to phonological variation, several hybrid models uniting generativist/OT and ET approaches are being developed. Some of these are sketched in Hinskens et al. (2014, 13–14). In these domains, three-way traffic between dialectology, formal theory and cognitivist theory may even come into vogue; cf. Nagy (2013, 437–438).
For the study of sound change, forced alignment and automatic vowel measurement (Labov et al. 2013) is very promising, because of the precision of the method and the enormous time savings, which enable researchers to concentrate on questions of interpretation. Other than phonetics, new methods of analysis can also be imported from other disciplines, including from statistics. More advanced multivariate analyses such as path analysis, a technique to estimate both the direct effects of certain variables and the indirect effects of the same variables or others via intervening variables (as applied in Villena Ponsoda 2014), enable more careful modeling of the interplay between internal and extra-linguistic forces in processes of language change. This can add depth to any generative model of language variation.

There is a need for multi-dimensional mapping techniques, which would enable the production of maps on which static as well as dynamic data can be displayed simultaneously in geographical space, in several types of social/cultural space and in time (see Thun 2010). It would be interesting to apply multi-dimensional cartography to data relating to clusters of phenomena or even grammatical (sub-)modules, for examples, a constraint ranking (Sloos and van Oostendorp 2012), instead of specific phenomena. Such cartographical techniques can help to bring to light the gradual internal generalization of some structural phenomena in the course of its diffusion, which may well be relevant to formal theory.

Last but definitely not least, theoretically enriched dialectological research could focus on sets of questions, for example, regarding aspects of the development of diglossic into diglossic repertoires that is currently taking place all over Europe. To what extent is the development of continua between dialects and standard varieties (which the Marburg-based REDE project focuses on)14 socially motivated and supported by functional dialect loss, growing command and usage of the standard or near-standard varieties, and the like? How are those changes internally conditioned? Is it on the basis of grammatical similarities, drift tendencies, or markedness? These general questions can be addressed through answers to more specific ones such as “which types of dialect features are generally ousted ‘on the way upwards,’ and which remain?”

More light can probably be shed on the complex conditioning of language variation and thus on the relative roles of dialectological, formal and possibly cognitivist approaches from the perspective of stability. Which types of variable phenomena remain stable through time and space, and why? Is it for linguistic or extra-linguistic reasons, or both? If both, then how do linguistic and extra-linguistic forces interlock?

Acknowledgments

Many thanks to Stef Barbiers, Hans Bennis, Edoardo Cavirani, Jeroen van Craenenbroeck, Ben Humans, Peter Gilles, Pieter Muysken, and Johan Rooryck for their help and advice. They are not responsible for any flaws, nor do they necessarily agree with the views expressed here.

NOTES

1 The idea encapsulated by this phrase is Hugo Schuchardt’s in his famous Über die junggrammatiker, gegen die Lautgesetze (1885), and in a short addendum to this entitled Worte als Individuen; see Spitzer (1921). But there is nothing in Schuchardt’s œuvre that is as concise and laconic as Jaberg’s famous dictum. Schuchardt’s admirer Jules Gillieron (Jaberg’s teacher and an atlas man more than anything else; see Goebb, this volume) does not seem to have published anything that contains the idea, although it has sometimes been attributed to him.
REFERENCES


