Internal Variability in Competence

Pieter A. M. Seuren, Nijmegen

0 Introduction

To the non-linguist one of the most obvious facts about natural languages is the availability of ranges of, often subtly, differing ways of expressing the same linguistic meaning, where the differences correlate with values on non-linguistic parameters. Every speaker of a language knows that certain features in his or other persons' speech indicate, not the linguistic meaning carried by the message, but associations with geographical area, with social class, with kinds of interactional situation, with age, with a profession, with a religion, etc. The differences show themselves in lexical selection, choice of grammatical construction, choice of pronunciation, — i.e., choices throughout the grammar. Even though most speakers are not aware, and cannot make themselves aware, of the precise features carrying the variable associations (except, sometimes, for a few prominent features which then acquire a special symbolic value$^1$)), the associations are nevertheless registered by listeners.$^2$ To know a language does not only mean to have the ability of converting meanings into well-formed strings of symbols and vice versa, but also to be able to distinguish between standard, substandard, formal, informal, dialectally and sociolectally marked forms of speech. An ideal speaker-listener would have command, both active and passive, of all varieties occurring

* Thanks are due to all those who were kind enough to discuss with me the ideas expressed in this paper. They are too numerous to be mentioned individually. Yet an exception must be made for Marianne Ebertowski, without whose unrelenting criticism this paper might never have seen the light.
in the community where the language in question is spoken. Obviously, such ideally competent people hardly occur in real life. But equally obviously, real people hardly ever have a linguistic competence which is entirely homogeneous: we all have active and passive command of a fairly wide range of varieties. A foreign learner whose competence is restricted to one variety only cannot be taken to know the second language perfectly. It is thus clear that variability poses a problem for any theory aiming at a reconstruction-by-hypothesis of linguistic competence.

In spite, however, of this very obvious fact theoretical linguistics, as it has developed since the beginning of the century, has never paid serious attention to variability phenomena. Dialectology and a primitive form of sociolinguistics existed well before theoretical linguistics appeared as a separate subject (with, say, de Saussure). But the rapidly growing prestige of the new theoretical discipline, both in Europe and across the Atlantic, made the study of dialects and sociolects recede into a limbo of disrespect, away from the spotlights. In some exceptional cases a serious reference was made in theoretical linguistic literature to variability phenomena, but no attempt was made at integrating them into a comprehensive linguistic theory. On the contrary, in more recent linguistic literature, especially in the various forms of transformational grammar, it has become customary to appeal to some undefined and often extremely crude notion of “dialect” in order to escape from awkward consequences that facts sometimes have for theories. This is the more surprising since it has been in the context of transformational grammar that the reconstruction-by-hypothesis of linguistic competence became the central concern.

New and powerful developments over the past fifteen years, especially in sociolinguistics, but also in the study of communicative interaction (pragmalinguistics), and to a lesser extent in dialectology, have re-established the status and respectability of variability studies. But theoretical linguistics has remained unaffected. Theoretical problems arising in the context of serious up-to-date work in linguistic variability have, on the whole, been ignored. The new developments have been met with supreme indifference. Any attempt at integrating variability studies into theoretical studies has, so far, come from practitioners of the former, never the latter. The result is that theoretical linguistics, with all its sophistication, has been unable so far to provide any clear definitions for such indispensable notions as “standard language”, “dialect”, “sociolect”, “slang”, “hypercorrection”, “diglossia”, to mention just a few.

The present paper is an attempt at improving upon this unsatisfactory situation. The native speaker’s linguistic competence will be taken as a central notion, but it will be stripped of the well-known idealization of homogeneity (CHOMSKY 1965: 3). A system of internal variability will be outlined, with variable features correlating with values on extra-linguistic parameters. A critique will be given of two important methods, proposed in recent years, of describing variability phenomena: the variable rule (LABOV 1972b:93–101) and the implicational scale (DECAMP 1971; BICKERTON 1971).
1.1 Competence and idealization

Over the past twenty years, theoretical linguistics has been largely competence-based. The central effort has been to establish, by approximation and hypothesis, what sort of mental structures and processes must be postulated to explicate the intuitive notion "knowledge of a language" or "command of a language". The criteria by which to evaluate the various hypotheses put forward have been drawn from observed facts (often, somewhat presumptuously, called "data"), from comparisons with rival theories, from constraints imposed by psycholinguistics and semantics. The object of investigation has thus been to establish what it is to have command of a language. The mental object thought necessary to explain those aspects in native speakers' verbal behaviour which show their command of the language, has been called "competence". Taken this way, i.e., as a technical synonym for commoner expressions such as "knowledge" or "command" of a language, there can be little doubt that there must be something like competence. COOPER (1975: 28) eloquently formulates precisely this point:

"Some writers, peculiarly, have doubted if there is a genuine distinction between competence and performance. But of course there is. One can and must distinguish between the ability to do something and actually doing something."

It is in this simple sense of ability to produce and understand utterances of a language that we will be using the term "competence".

Over the years, the notion of competence has widened a little. Whereas in the early years of transformational grammar competence was considered to be the speaker's capacity to produce and perceive well-formed (so-called "grammatical") utterances, each with a surface structure and a transformational derivation assigned by the grammar or syntax, it soon became clear that to be competent in a language involves more than just the handling of well-formed utterances, but must include also the ability to understand utterances, or, more generally, to assign meanings to utterances. To know a language implies knowledge of what the utterances of that language are about in actual use (given sufficient contextual and other background information). Thus various proposals were put forward to account for the knowledge of meanings, and competence models were enlarged with semantic components, a level of semantic representation, and what not. What interests us here is that we should realise that to know a language also implies the ability to recognise (and control) certain linguistic features as correlated with some geographical area, some social class, or some kind of interactional setting (formal, informal). A competent speaker of a language picks these connotations up, and has a considerable control over what variable features to select when he produces his utterances. Therefore, if the notion of competence is to be taken seriously at all, it calls for a model that allows for internal variability. We must drop that aspect of the well-known idealization of the native speaker/listener (CHOMSKY 1965: 3) which puts him in a completely homogeneous speech-community. Speech-communities are never completely homogeneous and speakers know that.
This does not mean that we propose to do without any idealization at all, only
that the type of idealization of the native speaker/listener should be brought a
step or two nearer reality. When the semantic parameters came to be added to
(or, as the case was, came to be seen as central to) competence models, this brought
about essential modifications with respect to the older, purely syntactic, models.
In just the same way the further requirement that competence models must take
internal variability into account cannot but have far-reaching consequences for
previous theory. Given a monolingual but internally varied speech-community,
our idealization will consist in our considering the linguistic competence of a
native speaker/listener who has at his complete command, both active and passive,
the whole range of variable features, plus their social meaning correlates, operative
in that speech community.5)
Clearly, this idealization is unrealistic: it is hardly imaginable that a speaker of, say,
English has a complete knowledge of all the minute dialectal, sociolectal and
interactional variations that occur throughout the entire part of the world where
English is the native language. Much more realistic is the assumption that real
individual speakers are competent in isolated bits of entire spectrum, and that
their active competence is a proper subpart of their passive competence. But our
idealization has the advantage of being close enough to what we may reasonably
expect reality to be, to enable us to describe real speakers in terms of the parts
of the spectrum they have active or passive command of. The idealization pro-
vides the terms for a specification of the details in which one speaker’s competence
diffs from another’s, and this is something the old idealization was unable to do.
In a wider perspective, the question will be taken up again in section 2.2.

1.2 Competence and psychological reality

A further difference between our notion of competence and what is
commonly found in the literature on transformational grammar consists in the
ontological interpretation of competence theories. Here, too, we aim at closer
and more direct links with actual performance. In transformational literature
it is customary to claim psychological reality for competence models on the one
hand, but to deny, on the other, that they could be taken as process models.6)
In this view a grammar, or competence model, does no more than characterize
what a good speaker’s competence amounts to in terms of possible output, with-
out any implication as regards the procedures followed during the actual pro-
duction or understanding of utterances.
Let us speak of taxonomic models when nothing more is intended than a precise
characterisation of output. Taxonomic models make no claims about the inner
workings of the apparatus or organism producing the output. If competence models
are taxonomic, they can have no psychological reality. Hence, the combination
of the psychological reality claim with the notion of competence models as taxo-
nomic models leads to contradiction.
Taxonomic models are certainly useful. They call for theories or models for the apparatus producing the output (or processing the input) described by them, and at the same time they impose essential constraints on such theories. These theories will have reality value, or a truth-value, with regard to the inner workings of the apparatus or organism, under some formula of interpretation. The formula can be weaker or stronger, according to the amount of physical or procedural detail to be specified. A weak formula of interpretation might claim no more than that the system contained in the theory must somehow, no matter how, correspond to structures and procedures occurring in and followed by the organism. A strong formula will be more specific as to precisely how the theory corresponds to the organism. The stronger the formula, the more vulnerable the theory, since the correspondence claim lays the theory open to confirmation or disconfirmation by other disciplines (such as psychology) concerned with the same organism. And as the correspondence claim gets more specific, as a result of the formula getting stronger, the risk of disconfirmation from outside automatically increases.

In theoretical linguistics the distinction between taxonomic models and causal theories has, by and large, been neglected. In any case, claims denying procedural validity to competence theories have not been accompanied by suggestions as to how competence should then be taken to really work. On this issue present-day linguistics is confused and unclear. CHOMSKY & KATZ (1974:364) wish to leave no doubt that linguistic theory is meant to be about the inner workings of the mind:

“Early physicists had no opportunity to look at the microstructure of matter, but like present-day linguists had to leave the ‘skulls of their subjects intact’. Nonetheless, they were able to determine a great deal about the microstructure of matter from its behavior.”

But on the other hand we read (CHOMSKY 1965:139–40) that

“It seems absurd to suppose that the speaker first forms a generalized Phrase-marker by base rules and then tests it for well-formedness by applying transformational rules to see if it gives, finally, a well-formed sentence.”

Chomsky considers it absurd to regard “the system of generative rules as a point-by-point model for the actual construction of a sentence by the speaker”. This would be “totally misconceiving its nature”. We thus find two incompatible views expressed by the same author (Chomsky). A not very convincing answer is given in CHOMSKY & KATZ (1974), where the view is put forward that a (generative) grammar should be seen as an “idealized” component of a more comprehensive theory of performance, the latter consisting largely of “perceptual strategies”. The performance theory would then be a process model or causal theory in our sense. They write (pp.359–60):
"Thus suppose it were discovered, say, by neurophysiological investigation or by psychological study, that all the linguist's data (and more) can be better explained by assuming that the organism has a system of perceptual strategies not involving the principles of generative grammar in any manner. The linguist who postulates a dag ( = descriptively adequate grammar) as 'true of' the organism will be unperturbed. Linguists who take the realist position, claiming that a dag actually describes the speech mechanism at work, might well abandon their formerly held comprehensive performance theory, with its idealized components and its specific principles and properties."

As long, however, as we are not provided with more specific information as to what is meant by saying that a system of perceptual strategies involves the principles of generative grammar, or that the grammatical component is idealized, this does not help very much. Clearly, a comprehensive theory of language use, or speech, involves a great deal more than just the grammar: the grammar deals with sentence types, but what occurs in speech is utterance tokens. Moreover, we do not expect of the grammar that it specifies how expressions are linked with (refer to) things in the world. And other aspects of language use can be mentioned which are not part of the domain of grammar in any sensible sense of the term. But it is perfectly reasonable to expect a grammar to provide a specification of a mechanism of sentence construction and perception as performed by humans during speech. A theory of grammar containing descriptive and constructive devices which cannot possibly be assumed to operate during speech, is thereby disqualified. An appeal to "idealization" will then only rescue the theory if a formulation is provided of the principles converting the "idealized" grammar to a grammar cast in realistically operative terms. In general it seems that, to the extent that theoretical linguistics is under the influence of Chomsky, there is a reluctance to accept the normal consequences of a realist position. The result is confusion and unclarity. The realist, or mentalist, position in linguistics proposed in Katz (1964) has been accepted in some ways, but not in others. What remains is a certain halfheartedness about the question of psychological reality.

The competence theory developed here is based on a seriously realist view. That is, we consider it to be a reasonable requirement that the linguist should specify in what ways the various elements of this theory should be interpreted in a process model of speech production and comprehension. When he gives formation rules for the structural description of representations at some level (deep structure, shallow structure, etc.), he should make it clear whether he intends these to be interpreted as descriptions of actual processes or as taxonomic descriptions of the possible output of processes operating before that output during production. When he gives transformational rules, then, if these are not to be taken as descriptions of actual processes, it must be made clear how they should be interpreted. There is, of course, nothing "absurd" about the view that transformations represent actual production processes. On the contrary, it imposes constraints on the transformations
in so far as the input must be somehow recoverable from the output for comprehension to be possible.

Unfortunately, psychology provides us with relatively few results that have a direct bearing on the selection of grammatical theories. To the extent that psychologists have investigated the possibilities of transformational grammar for a process model, their reports have been discouraging. However, most of the objections disappear when grammatical processing is assumed to take place "underground", i.e., as a routine procedure with a semantic input and a phonetic output.7) The requirement of psychological reality has not so far led to many actual constraints on grammatical theory. But there is no reason why linguists should not undertake the task of formulating the constraints emanating from a seriously realist view of grammar. BRESNAN (1978) contains an attempt at doing this. Bresnan proposes a lexicalist-interpretivist theory of grammar (the merits of which are not at issue here), with full acceptance of the constraints imposed on linguistic theory by the condition of psychological reality. Yet her account still suffers to some extent from the halfheartedness observed in current linguistic thinking. She renders our distinction between taxonomic and process models as a distinction between "two research objectives", viz. "the grammatical characterization problem and the grammatical realization problem" (p. 1). But the problem of how "to characterize the grammar, that is to represent the language user's knowledge of language" (p. 1) is precisely the problem we are facing and for which we require a seriously realist theory. She furthermore adheres to the view (p. 3) that psychological reality "in the broad sense" is achieved when a linguistic concept contributes to the understanding of the behaviour involved in giving linguistic judgements such as are commonly used in grammatical theory construction and presented as "facts" or "data". This seems an unnecessary complication. The distinction between taxonomic characterization of output on the one hand, and process models on the other, seems entirely sufficient and clear. It is also, in fact, the distinction Bresnan operates with.

Given the scarcity of actual constraints provided by psychology or neurophysiology, it would seem hazardous for the linguist to venture more than the weakest possible formula of interpretation. This will involve nothing more than the claim that the structures and processes defined in his theory describe actual structures and processes occurring in speech. The claim does not involve completeness: the linguist's grammar may be limited to just production (but then the possibility of comprehension will impose strict constraints), or it may not contain precise details of feedback control, etc. Without any undue commitment to particular grammatical theories or processing details, we can tentatively limit the class of viable grammatical theories in terms of the following schema:
This schema leaves open the possibility of processing for the purpose of monitoring during production or comprehension (MARSHALL & MORTON 1978). That is, it may well be that, for example, during production a particular PR put out by G is fed back into G to yield an SR. If this SR matches the SR originally fed into G, production can carry on through PA to yield an acoustic signal. Furthermore, one must keep in mind that only CR and AS are open to awareness, and that the processing from CR to AS and vice versa takes place "underground", i.e., as a routine procedure which is not open to introspection or to controlled interference (see note 7). These aspects, however, are not the prime concern in this paper: what counts here is that the processing units S and G will have to allow for certain forms of internal variability. S will have to allow for differences in lexical selection or registration under identical CR or SR input. G will have to allow for differences in rule application (syntactic, morphological, and phonological rules), not only as regards the actual rules applied but also as regards the order in which the rules are applied.8)

1.3 Competence and parameters of variability

When studying variability phenomena in spoken utterances we shall not be concerned with those variations that are not bound up with what can reasonably be considered to be the system of the language in question, but result from factors that are quite independent of any linguistic system or grammar. Variation of utterances with the same linguistic meaning is not random or arbitrary. All kinds of linguistic or anyway acoustic features correspond more or less regularly with all kinds of parameters, and listeners do, in one way or another, register these correspondences. Information about age, sex, emotional state and many other parameters is carried by the spoken message, along with information about the geographical and sociological identity of the speaker as well as his evaluation of the situation features involved. We shall limit our attention to variations whose registration or active command can be said to involve knowledge of the language. To know how to distinguish male from female speakers is not part of our knowledge of English. But to know how to distinguish British from American English certainly is. The variations which have to do with knowledge of the language.
will have to be accounted for in an adequate theory of linguistic competence, i.e., in the processing units S and G. We shall have to organize the lexical and grammatical rules in such a way that vast multitudes of choices can be made by the organism according to some present standard or stylistic ideal.

We shall thus be concerned solely with *linguistically relevant* variations. But here the question naturally arises of what exactly the parameters are which allow for linguistically relevant variation. There is a fairly general consensus, though not often expressed, that there are at least three parameters of systematic linguistic variation: the *geographical parameter* (G), according to which dialects are distinguished; the *sociological parameter* (S), according to which the so-called sociolects are distinguished; and the *interactional parameter* (I), where distinctions are made between different kinds of speech situations. CEDERGREN & SANKOFF (1974), as well as KLEIN (1974), add as a fourth parameter the *diachronical development* (D) of a language.

As regards parameter D, it is known (e.g. LABOV 1972a: 26, 30, 160—82) that there are communities where systematic correlations can be observed between age groups and the application of certain linguistic (phonological) rules, and such correlations have been interpreted as reflecting "the mechanism of linguistic change". This interpretation may well be correct in the cases studied, but many questions remain. First, it is not clear how general this phenomenon is. It is imaginable that there are speech communities where no significant correlations with age groups occur, or where such correlations do not reflect linguistic change. Moreover, and more importantly, it is not clear that we have to do with an independent parameter. It may well be that significant correlations with age, where they occur, can or should be accounted for in terms of the parameter S. It is an empirical question whether or not it is true that wherever there are systematic linguistic correlations with age groups in a community, the age groups in question will be distinguished on independent grounds as sociological groups. Given these uncertainties, it seems safer not to consider D as a separate, independent parameter of linguistic variation. The model to be developed below is such, in any case, that D can be incorporated if that is desirable.

We shall thus operate with the parameters G, S, and I. As regards G, it is common knowledge that some structural linguistic variations correlate regularly with geographic areas. The term *dialect*, though still undefined in theoretical linguistics, has become part of the everyday stock of terms denoting linguistic phenomena. It might be objected that this form of variation could be eliminated from the theory by stipulating that a speech community is never larger than a "dialect". But this would not be adequate, since every speaker has at least passive competence in more than his local dialect. Furthermore, many dialectal varieties are limited to certain restricted social classes and/or certain restricted interactional settings. Then, it is not at all clear a priori what the boundaries should be of a dialect area. In some sense, one particular neighbourhood in a big city can be said to have its own dialect, in that certain features are characteristic for that and only that neighbourhood. But in another sense half a province can be said
to have its own dialect, for exactly the same reason. In general, if we drop the
general umbrella notion of "language of a speech community", banning all forms
of geographic variation, then what remains as our object of investigation is an
ill-defined set of nesting and overlapping speech communities, which are obviously
related but for whose interconnections linguistic theory will be unable to ac-
count.

Other structural variations correspond with social groups, in particular with social
classes. Since Labov started his sociolinguistic investigations in the early sixties,
it has become abundantly clear that there are systematic correlations between
certain structural linguistic features and certain sociological parameters. We now
use the term *sociolect*, although, again, there is no good definition for it in theo-
retical linguistics.

The third axis of systematic linguistic variation is given by the situational or inter-
actional setting of the speech-event. This parameter, which we have dubbed "I",
is specifically studied in the newly developing field of *pragmalinguistics*. Even
more than dialectology or sociolinguistics, this new discipline suffers from the
neglect, or incapacity, of theoretical linguistics in providing an integrative frame-
work.

### 2.1 Implicational scales

The implications of variability phenomena have provoked a lively interest
among sociolinguists. Two perspectives, in particular, have been developed for
their interpretation. The first is based on observations made by Decamp (1971),
and confirmed by Elliot, Legum & Thompson (1969) as well as by Bickerton
(1971). According to these observations, variable features can be ordered hierar-
chically in such a way that for each individual speaker the prediction holds that
if he has feature \( a \), he will also have the features \( b, c, d, e, f, \ldots \); if he has feature
\( b \), he will also have the feature \( c, d, e, f, \ldots \); but not necessarily \( a \); if he has feature
\( c \), he will also have \( d, e, f, \ldots \), but not necessarily \( a \) or \( b \), etc. The terms "implicational scales" or "scalogram analysis" have been adopted to describe such an
arrangement. It is sometimes possible to correlate an implicational scale with
some extra-linguistic parameter, such as social status. Thus, feature \( a \) could be
ranked lowest on a social scale, \( b \) a little higher, \( c \) a little higher still, etc. Spea-
kers can then be ranked along the social scale according to the starting point of
their linguistic implicational scale: a speaker who has \( c, d, e, f, \ldots \), but neither
\( a \) nor \( b \), will be ranked higher than a speaker who has \( b \) plus its implicational set,
but not \( a \).

There can be little doubt that the existence of such scales is an important fact
for the study of linguistic variability, which has to be accounted for in any serious
competence model. The competence model presented below, in fact, predicts the
occurrence of implicational orderings of features for groups of speakers. The re-
levance of implicational scales has been questioned by Fasold (1970:555–6):
"Would it not be possible to ransack a body of data of any kind and find several elements which would happen to be amenable to scalogram analysis?"

The answer is clearly negative. Suppose there is a random distribution of a large number of variable features over a large enough number of individuals, then it would be surprising if any set of features could be found that would allow for implicational ordering for all the individuals in the population. When such sets are found regularly, the distribution is not random: there must be some system. In that case it makes sense to ask what the system is and what causes the implicational arrays. Apparently, Fasold anticipates this answer, for he continues:

"Such studies may have some interest sociologically, however, as a way of discovering which language features are considered ‘worse’ or ‘better’ than others by linguistically naive members of the speech community."

Another criticism formulated by FASOLD (1970: 552–4) cuts harder wood. Scalogram analysis is based on binary decisions: a speaker does or does not have a certain feature. There is abundant evidence, however, from the work by Labov and others, that many features do not come in such an all-or-nothing way, but co-occur with alternatives in the speech of the same individuals. In fact, a great deal of Labov’s work is based on a systematic analysis of the proportions in which certain designated features occur in the speech of individuals of different socio-economic groups. Fasold quotes STOLZ & BILLS (1968), who recognised the problem and admit that “thresholds were set post hoc to give the optimal fit between the data and the scalogram model”.

An immediately obvious move to counter this objection would be to limit scalogram analysis to cases with a very high threshold, say near or at 100%. This would presumably save the relevance of this analysis but limit its applicability. More essential, however, is the answer that implicational scales are nothing but taxonomic arrays, a selective form of data presentation. They do not provide a model but impose empirical constraints on models. Any adequate theory will have to account not only for implicational phenomena, but also for Labovian proportional variability. And it is, as yet, an open question whether proportional variability is normal for the system underlying speech output, binary variation being no more than marginal. Or whether the all-or-nothing cases are closer to the idealized system, proportional variability being the result of (no doubt systematic) external factors.

### 2.2 Variable rules

The second perspective for the interpretation of variability phenomena is presented by the concept of variable rule. This concept deserves a detailed discussion because it has gained widespread acceptance as a rule category in competence. I shall try to show that the assumption of variable rules in competence theories is misguided even though the statistical functions which they incorporate have a high taxonomic relevance.
Variable rules predict text profiles, i.e., distributions of variable features over texts. The notion of text profile has some intuitive reality. We may say of a given text that it is of an overall formal character but that the speaker slips into occasional colloquialisms, that there is a surplus of learned words, etc. We spot hypercorrect usage, foreign accents, regional accents, archaisms, etc. But it is still largely unknown how specific text profiles come about. A partial determinant consists in what is called "attitudes". It has been established (Giles, 1973, 1977; Giles, Taylor & Bourhis 1973; Simard, Taylor & Giles 1976) that attitudes in the social-psychological sense\(^9\) are responsible for certain features in texts. Positive feelings towards an interlocutor will cause a certain amount of accent convergence, whereas negative feelings may cause accent divergence. Not only such short-term attitudes are at work: long-term attitudes concerning socio-political issues or similar matters may determine or influence style selection in certain types of interaction. But on the other hand it is an obvious and undeniable fact that certain features will occur, either throughout or proportionally, no matter what the speaker's attitudes are. That is how we recognise dialectal or sociolectal accents. Socio-demographic parameters undeniably play an important role in the formation of text profiles.

It is not obvious that for the purpose of reconstruction-by-hypothesis of linguistic competence we need rules or principles that predict text profiles.\(^10\) In one sense such a requirement quickly leads to absurdity. If we require of competence theory that it predict for every speaker on every occasion what the profile will be of the text that will be produced, then the whole long-term and short-term attitudinal make-up of each speaker would have to be specified as part of his linguistic competence. We would have to say that a young man, who speaks differently after his getting married from the way he spoke before (because of his different outlook on life and its values) has undergone a shift in linguistic competence. Or that a person who thinks sherry parties are a capitalist institution and therefore makes a point of frequenting them and using inappropriate language, should be taught better English (whereas, of course, he uses his competence in the language to provoke certain effects). Clearly, defined this way, "competence" becomes an impossible notion, expanded beyond tolerable limits.

But this, of course, is not the way in which variable rules predict text profiles. In variable rule theory (VRT), as developed in Labov (1969) and Cederberg & Sankoff (1974), competence is taken as an apparatus which is sensitive to a number of parameters, roughly in the following way. VRT is based on Labov's discovery that correlations can be established between the frequency of occurrence of given variable features and (a) certain linguistic features in the text environment, (b) the socio-demographic description of the speaker, and (c) the position of the text on a formal-informal scale. It takes the familiar distinction between optional and obligatory rules as a starting point. The terminology, however, is changed: optional rules become variable rules, and the old obligatory rules are now called "categorical". Given the correlations mentioned, a probability value \(p\) can be assigned to each optional/variable rule. The value of \(p\) is the
result of a complex function which takes the socio-demographic description of the individual speaker as a basic element, modified by linguistic features, contextual style, and other possible factors (such as attitude). The frequency of a particular feature in a text is then determined by the $p$-value of the rule that brings this feature about.

Let us try and get a clearer idea of what is involved in the problem we are discussing. The problem is: should we assume this notion of variable rule as part of our competence theory? Labov (1969:759) gives a positive answer, but he has not repeated it in later work. Cedergren & Sankoff (1974:334, 353) assert emphatically that variable rules should be considered part of each speaker's individual linguistic competence. They infer from statistical regularities to competence rules, and we wish to see if this inference is tenable, reasonable, or otherwise defendable. It seems that the answer to this question is intimately connected with the level of abstraction or idealization we want to maintain for our notion of competence, and with the nature of this idealization. Cedergren & Sankoff speak about each speaker's individual competence, which is reasonable enough since it is individual speakers who constitute our empirical problem, not some idealization. It would be unwise to exclude the possibility that each individual speaker has mechanisms attached to or associated with the rules of his linguistic competence, which are unique or have unique qualities. But we want to know what sense it makes to say that a description of competence is a description of the language of a community, or that acquisition of competence is a process of linguistic socialization.\footnote{11}

It is customary in linguistics, and we have not diverged from that custom, to consider a level of idealization that captures the notion "knowledge of a language". It is important to realize that this is not just an idealization of the sort often found in the physical sciences, where idealization is necessary in the theory because of the contamination of nature. Ours is an idealization that captures a bit of social reality. And this imposes certain criteria on our selection of the kind of idealization we will work with, quite apart from the ordinary fact that we have to idealize in the theory because of the contamination of nature. It was on account of criteria of linguistic social reality that we decided that variability phenomena have to be built into an adequate theory of linguistic competence, and that we selected the parameters G, S and I as the main parameters of variability. To know a language completely is, as we have seen (section 1.1), a practical impossibility. But to know a language, or dialect, or sociolect, or any other identifiable kind of variety, perfectly is much less unrealistic: it means that complete linguistic socialization has taken place within certain limits defined on independent, external grounds (geographical, sociological, interactional).

The question now is whether at this level of abstraction, which is meant to represent the social reality of language, we need a descriptive device of the type represented by the variable rules. That is, do we need to incorporate a probability function of a variable-rule-type into all or some optional rules of the grammar, letting the values for the various parameters be filled in by the individual speakers?
A positive answer implies the view that the idiosyncrasies of text profiles are the result of a processing mechanism, the variable rules, which is part of the reality of language and has to be acquired if command of a language is sought. A negative answer implies that there is no such mechanism and that it is the realities of human life — social, cultural, geographical, biological, biographical —, which account for idiosyncratic text profiles. In this view text profile characteristics are part of the “contamination” of the data, and any regularities found in the contamination zone will be non-linguistic.

A number of authors have commented upon the question of variable rules as part of competence. Bickerton's criticism (1971) mainly centers on two points: the unlikelihood of any mental apparatus keeping track of the frequencies of occurrences of features in texts, and the fact that in Labov's data presentation peculiarities attributable to individuals are not detectable, since they are "drowned" in group averages. His first point, however, is mainly supported by rhetoric (p.461), and he is given short shrift by Cedergren & Sankoff (1974:335), who say that this claim "ignores the extensive psychological literature documenting probabilistic aspects of mental processes". The second claim was true in 1971, but may mean no more than that a rearrangement in Labov's data presentation would show the desired correlations. He also expresses the intuitive judgment, shared by a great many linguists, that a speaker selects his variable features to a large extent according to "his own peculiar experiences and attitudes" (p.483). We have seen, that even though attitudes as well as socio-demographic and biographical descriptions strongly influence text profiles, that provides no argument against the principle of variable rule theory, since all these factors can be built into variable rules as separate variables.

Fasold (1978) notes that "most proponents of variable rules in linguistics have not made detailed attempts to integrate variable rules into linguistic theory, beyond adapting the notational conventions of generative phonology and answering those critics who claim that variable rules are unlearnable" (p.85). He devotes specific attention to the theoretical implications of variable rule theory, and defends the view that linguistic environment factors favouring or inhibiting the application of a variable rule and their relative weight or strength vis-à-vis each other, should be included in a competence description and be reckoned to be part of a speaker's knowledge of his language, but not the probability values that can be associated with each factor. His argument is supported mainly by data on final t/d-deletion in East Harlem Puerto Rican English, and the New York Black English phenomenon, extensively discussed in Labov (1972c:78–85), of the variable deletion of postconsonantal word-final alveolar stops. This latter rule is subject to two linguistic environment constraints: a following consonant favours the rule more than a following vowel, and a preceding morpheme boundary inhibits the rule. Moreover, the former constraint is stronger, or heavier, than the latter. Hence deletion of -ed in e.g. I missed my bus is more frequent than deletion of the phonetically identical segment -t in e.g. There is mist on the river.
The complex machinery, he argues, of variable rules is not needed in the reconstruction of competence. All that has to be assumed is competence knowledge of the relative weights of the constraints: "The precise probability values are not important, and any estimation procedure available now or in the future would be equally valuable, provided that the estimates achieve a reasonably good fit with the observed data." (p.94) This is in fact a simplicity argument: no assumptions should be made that are not necessary. We notice that, in this view, a grammar still predicts text profiles, but rather less strictly and more as a side-effect.

It is clear that this notion of variable rule implies a specific claim about linguistic variability: if linguistic competence includes variable rules of this type, and if linguistic competence is the result of socialization, i.e., the assimilation of principles controlling areas of behaviour within a given group, then the relative weights of constraints in variable rules must be constant within the community for which linguistic competence is defined. For a rule with different relative weightings is a different rule, and the rules of a speech community must be shared by all, i.e., must be identical for each member of the community. Now Labov tells us (1972c:81), speaking of final t/d-deletion in Black English, that "the weighting of the variable constraints is not general — in fact, it varies from group to group, and from one age level to another in a meaningful manner." Given this, the Fasold-type variable rule does not take us very far in describing internal variability. It is far too restrictive as regards the notion of speech community. As has been said above, our aim is to reconstruct or describe the competence of people whose active and passive command of their language enables them to produce and even more so to comprehend utterances that are marked for different areas, social groups and interactional situations, without having to assume that for each variety they use a different language with a different grammar. Fasold's grammar will describe competence for extremely limited groups only, and no unified competence description can be achieved in his theory for larger groups. The same conclusion is formulated in Kay (1978) for all varieties of variable rule theory. He writes (p.82):

"The theoretical utility of variable rules in modeling the normal heterogeneity of speech communities would appear limited, since the heterogeneity frequently entails interaction of linguistic and social constraints. Variable rules will no doubt remain a valuable statistical tool for the analysis of the differential strengths of linguistic and social variables in determining linguistic behaviour in subsections of speech communities that are homogeneous and stable with respect to the relevant linguistic constraints on rule application."

It thus appears impossible to describe sociolinguistically meaningful variation within a speech community in terms of rules whose application is subject to probability constraints. In the following section a system will be proposed where variation is described in terms of rule alternatives, each alternative being marked for a socially meaningful connotation.
There is one further respect, not so far noticed in the literature, in which variable rules fail to do justice to the facts. VRT ties all variation down to application or non-application of particular rules, and thus cannot deal with variation which is due to a difference in order of application of otherwise identical rules. A clear instance is provided by French, where a socially significant difference is found between, e.g.:

(1) C'est les étudiants qui ont commencé la bagarre.

and

(2) Ce sont les étudiants qui ont commencé la bagarre.

Both mean “it is the students who started the riot”, but (1) is colloquial, though geographically and socially unmarked, and (2) is formal standard. The grammatical difference can be defined as a difference in the order of application of the rules of ce-Placement and Number Agreement. Starting from an underlying structure of the following simplified form:

(3) the x's (\(\_x\) started the riot) be the students

extraposition of the relative clause gives:

(4) the x's be the students (\(\_x\) started the riot).

As regards the main clause, two operations have to be performed: ce-Placement and Number Agreement. In (1), ce-Placement has taken place first, making the subject singular so that Number Agreement will yield the singular est (as, analogously, is the case in all varieties of English). But in (2) the two rules have been applied in the opposite order, and the verb is therefore plural (as in German, Dutch or Italian). Clearly, the mechanism of variable rules is incapable of dealing with cases such as this. We need a competence model allowing for different rule sequences, each marked for a different social meaning.12

These considerations make one wonder whether variable rules should not be rejected as a competence category. Perhaps the aim of predicting text profiles was too ambitious. As we have seen, it is not a priori clear that any probabilistic machinery is needed: text profiles may be “contaminated” material. In general terms we have to do with an apparatus whose output is subject to variation according to variable input from a number of independent factors. We must account for the fact that variable features do not always occur in an all-or-nothing way. That is, we must be able to account in principle for style mixtures, but not necessarily for the precise mixture we find.

3 Socially meaningful alternative rules system

Instead of building variability into single rules, as in VRT, we may envisage a competence model with large numbers of alternative rules at given stages in the derivation of sentences, each rule being marked for some social meaning.
We assume that a speaker aims at some specific style target, such as formal or informal, dialectal or standard, and will instruct his production grammar to make every choice according to the preset style target. Ideally, a speaker will then produce a one hundred percent pure and uncontaminated output. Since we must also assume, however, that speakers are often less than fully competent in many of the styles they wish to achieve, and that moreover the setting of style targets itself is often unclear and shifts from one moment to another, due to unclear and shifting attitudes, we must expect outputs that are highly contaminated. A socially meaningful alternative rules system (SMARS) therefore operates with the idealization of fully competent speakers and uncontaminated outputs, and attributes proportional variability to defective learning and attitudinal ambiguities. To the extent that proportional variability phenomena result constant or uniform across a particular speech community, non-linguistic social factors are invoked, such as common school training, culturally generated attitudes with respect to certain interactional situations (experimental sessions, for example), with respect to individual or class identity, and the like. A SMARS-type competence theory does not predict text profiles.

Let us revert to the example mentioned above of postconsonantal word-final deletion of $t/d$ in New York Black English Vernacular (BEV). LABOV (1972c:80) gives precise data for eleven individual speakers, members of an adolescent group called "Jets", for $t/d$ dropping in four different linguistic environments:

1. Without preceding morpheme boundary, before consonant 90%
2. Without preceding morpheme boundary, before vowel 49%
3. With preceding past tense morpheme boundary, before consonant 46%
4. With preceding past tense morpheme boundary, before vowel 7%

Fig. 2 Percentages of final $t/d$ deletion for 11 Jet members. From LABOV (1972c:80)

LABOV comments (p.81) that the deletion is virtually uncontaminated (95%) in environment 1, if one member is excluded who deleted in only 60% of the cases in the environment, and very little throughout.

Instead of formulating one deletion rule, we can formulate three, each with a different linguistic environment feature:

a. $/ C \overline{mb} \rightarrow wb C$

b. $/ C \overline{mb} \rightarrow wb V$

c. $/ C mb \rightarrow wb$

("wb" for word boundary, and "mb" for morpheme boundary; "\overline{mb}" for no morpheme boundary). Rule (a) is very general in English anyway (LABOV 1972c:78–9), and we may assume that for the speakers in question it is marked as belonging to "proper English". Rules (b) and (c), however, have a very different status. We see a drop of well over 40% when one of the two requirements of rule (a) is not fulfilled (viz. \overline{mb} or rightmost C), and a cumulative drop of over
80% when both requirements are violated. This is accounted for by the assumption that rules (b) and (c) are marked areally (the parameter G), sociologically (S), and/or interactionally (I) for South-Central Harlem, Blacks and informal speech, respectively, - plus the assumption that the speakers were striving for some sort of English that was acceptable to socially higher-classed outsiders (they were "accommodating their accents" in the terms of GILES, TAYLOR and BOURHIS (1973)). Rules (b) and (c) can be taken to be stigmatized, in the sense that has become current in sociolinguistics, but belonging to the core of the competence spectrum commanded by the speakers in question.

This explanation is, of course, an oversimplification. On pp. 81-2 Labov gives some comparative data on final t/d deletion for the Jets and another such group, the Cobras, not more than three blocks away. All Cobras but one apply rule (c) significantly more frequently, and a majority apply rule (b) with greater frequency, some of them coming up to nearly 100 percent. Without further information it is hard to say what is going on here. It is possible that the rules (b) and (c) are associated with Cobra-speech rather than with Jet-speech, but it is equally possible that the difference is due to group-specific attitudinal differences in the sense that the Cobras care less about proper speech than the Jets. A closer look at the data will help, no doubt, but we must let the matter rest here.

We have already isolated three parameters of variability: G, S and I. We now assume that each alternative in SMARS is marked for a certain value on each of the three parameters. Let the following be an elementary marking system. We assign the value "O" for minimal marking. On G this means "not associated with any particular area"; for S: "not associated with any particular class or group"; for I: "not associated with any particular class of interactional settings". Let us use the value "1" to indicate, for G, the association with some particular area; for S, the association with substandard speakers; for I, the association with informal settings. A stronger marking "2" applies only to S and I. For S: the association with superstandard speakers (such as the language of the Court, or of the aristocracy), and for I: the association with specifically formal situations. The strongest marking "3" applies likewise only to S and I, and is reserved for specific sociological groups (members of a church, a political party, a profession, etc.) on S, and for specific communicative settings (ceremonies, rituals, telephone conversation, etc.) on I. This gives us the following elementary marking table:

<table>
<thead>
<tr>
<th>G</th>
<th>S</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>general</td>
<td>general</td>
</tr>
<tr>
<td>1</td>
<td>specific area</td>
<td>lower class</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>upper class</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>special group</td>
</tr>
<tr>
<td></td>
<td></td>
<td>general</td>
</tr>
<tr>
<td></td>
<td></td>
<td>informal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>formal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>special setting</td>
</tr>
</tbody>
</table>

Fig. 3 Elementary marking table for the parameters G, S and I.
In terms of this table we shall use the notation "O,O,O" to indicate the marking on the parameters G, S and I, in that order.

This elementary marking system will, of course, be insufficient as a comprehensive descriptive system for socially meaningful rule alternatives. For one thing, certain lexical selections are more substandard than others: not all rule alternatives that are marked 1,1,1 are equivalent from a social meaning point of view. There are degrees of intensity, and an adequate notation will have to take these into account. Then, there may be a group-specific social ranking which is expressed in rule alternatives. In such case the alternatives will have to be marked on a social scale within the group. This is easily done with the help of the colon: whereas \( \alpha, 3, \beta \) is a rule marking for a special group on S, we can express, say, lower standing within the group in the following way: \( \alpha, 3:1, \beta \). And there are no doubt more inadequacies and insufficiencies. Yet, the elementary marking system is good enough to help us in defining some important sociolinguistic notions.

In interpreting the marking system, the following two points must be observed. First, the strongest marking for each parameter ("1" for G, and "3" for S and I) is a variable, not a constant, marking in the sense that a specification must be provided as to what area, what group and what setting. Then, the markings must be read in progressive restriction: if G is marked "1", then the marking values for S and I are valid for the area given for G. Likewise, the value for S restricts the scope of I. Thus, e.g., a rule marked 0,2,1 will produce a variable feature that is non-dialectal, upper class, and informal for that upper class.

In general it is individual rules that are given SMARS markings. They can be lexical selection rules, or rules in the syntax, the morphology, the morphophonology or the phonology. But occasionally we have sequences of rules that are SMARS-marked. This is the case when variable output is due to the application of different rule sequences rather than different single rules. A syntactic example taken from French has been given above, and a few references to phonological cases are given in note 12. Limiting ourselves to the values "O" and "1", we consider the following rule-types:

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>S</th>
<th>I</th>
</tr>
</thead>
</table>
| 4. | 0 | 0 | 0 | general standard rule  
| 3. | 1 | 0 | 0 | standard dialectal rule  
| 2. | 1 | 1 | 0 | stigmatized dialectal rule  
| 1. | 1 | 1 | 1 | stigmatized socio-dialectal rule

Figure 4 helps us to define the notion of "style". When a speaker aims at an informal, lower class, dialectal style (we shall use brackets to indicate styles: (1,1,1)), he will maximize the selection of the rules marked 1,1,1. Since, however, rules of that marking will not always be available when a selection is to be made, he will, for the
remaining cases, maximize the selection of rules marked 1,1,0. Again, for a great many selections no 1,1,0 rule will be available. He then maximizes rule selections of the type 1,0,0. And for the remaining cases he will select the minimally marked rules 0,0,0. When he achieves his preset style target perfectly, he will produce an uncontaminated, homogeneous style of the type (1,1,1). If, however, he aims at the style (1,1,0), he will maximize the selections of the rules marked 1,1,0; then of those marked 1,0,0; and finally of those marked 0,0,0. If his style target is the standard dialect, he will start at maximal rule selection of 1,0,0 rules, and fill in the remainders with rules of the 0,0,0 type. We thus see a hierarchy emerging which reminds us strongly of the implicational scales of DECAP and BICKERTON, discussed in section 2.1. Given an adequately socialized speaker staying within a stable stylistic range that belongs to the core of his competence spectrum, SMARS predicts implicational scales as were found by these authors and others. In general, styles are defined implicationally. A style (α, β, γ) will always involve a maximization of the rules carrying the same marking, plus an implicational scale specifying what order successive maximizations are to follow.\(^{13}\)

We can now also define the notion of "standard language", as opposed to "dialect" or "sociolect" or the like. We say that a speech community is unilingual if all the members of that community have in common a set of rules (0,0,0) (curly brackets are used to indicate sets of rules of a type), containing lexical as well as syntactic, morphological, morphonological and phonological rules. The larger the set \{0,0,0\}, the more clearly we have to do with a unilingual speech community. One of the standard problems mentioned in the literature when attempts are made to define the notion “standard language” against "dialect", is the alleged fact that so often national boundaries stand in no relation to local dialects: the dialects are said to form a perfect continuum, unperturbed by some state boundary that may run right across. This is said to be typically the case between Holland and Germany. Why then, one asks, should one dialect be called Dutch and an adjacent very similar dialect be called a dialect of German? The answer is that speakers West of the boundary have been socialized, through school training, newspapers, cultural contact, etc., in a community characterized by a very large set of linguistic rules that are minimally marked for, at least, the piece of land enclosed by the state boundaries of the Netherlands. East of the boundary, on the other hand, speakers have been socialized in a community with an equally large, but different set of minimally marked linguistic rules. Speakers on the Dutch side of the border will, at least in comprehension, be competent in unmarked, general standard Dutch, and not or much less in general standard German. Analogously, on the German side of the border speakers are competent judges of general standard German, but not of Dutch. Moreover, although it is often said that the dialects form a smooth continuum, this is, in fact, not so. In a standard dialect style of a Dutch dialect near the border, i.e., (1,0,0), the Dutch \{0,0,0\}-fund is used to fall back on for the cases where no 1,0,0 rule is available, whereas on the German side it is, of course, the different German \{0,0,0\}-fund. Repeated personal observation has confirmed this.
More interesting, and more controversial, is the Southern border of the Netherlands, separating it from Belgium. Belgium, clearly, does not form a unilingual speech community: the Southern part belongs to the French speaking community, but the Northern part speaks a language that can only be classified as Dutch. Yet, although there clearly is a set of rules \( \{0,0,0\} \) from every section of linguistic competence, and common to the Netherlands and the Dutch speaking part of Belgium, commonly called the Flanders, this set is not large enough to enable speakers to produce a pure general Netherlandic style. In other words, in many cases during the production of sentences speakers will find no \( 0,0,0 \) rule among the required alternatives. When the style target is "minimally marked style", the speakers can do nothing in those cases but fall back on rules that are marked geographically, but for the widest possible area. This means that, although according to our definition the Netherlands plus the Flanders form a unilingual speech community, there is no single standard language in the sense of a pure style \( (0,0,0) \): the most neutral, unmarked, general styles that can be achieved are two varieties of \( (1,0,0) \), the Flemish and the Dutch of the Netherlands. Accordingly, we must say that there are two standard languages in one unilingual community, a standard Flemish Dutch and a standard Netherlands Dutch. Recognition of this fact is not always easy. In the Flanders in particular there is some resistance to the idea that Flemish speakers have their own standard. But I think the social reality of the language clearly points that way.

In fact, this situation is very common. English, of course, is the prototype case. There is a valid notion "English speaking world", defined by the fact that there is a fairly large set of rules \( \{0,0,0\} \), taken from all sections of competence, and common to whoever speaks English. Yet there are many standard English languages, since there is no single \( (0,0,0) \). There always has to be some areal colouring for any least marked style. The British ideal of least marked styles is "received pronunciation" (RP), but RP puts the speaker squarely inside England, and not, for example, Scotland or Ireland. There is a standard American English, a standard Australian English, etc.

A standard language is thus equated with the least marked style available in a community, and such a community does not have to be co-extensive with a unilingual speech community. How standard languages come about is not a simple question to answer. Apparently, organized communities have a tendency towards developing a linguistic style which is as neutral as possible as regards internal divisions or distinctions. For that to happen rules will have to acquire new markings, and it is largely a question of social psychology what mechanisms lead to group-neutrality. An appeal to reference groups is certainly insufficient, though perhaps necessary. Groups migrating from the homeland will, as a rule, have a different sociological make-up from the population at home. To the extent that such groups form organized communities, or even societies, the need for a new group-neutral standard style will arise. Clearly, due to sociological and social-psychological developments each society will have periods of flux, where an existing standard collapses (for example, when it has become core competence of a
specific social group), or where there is a chaotic collection of least marked styles, each style carrying a fair but always different bit of marking. Periods of considerable social upheaval, such as the collapse of the Roman Empire in the fifth Century, are always followed by an equally considerable restructuring of the norms for least markedness. Dialects, just like standard languages, are styles, but without the property of minimal markedness. That is, for every dialect style there is at least one other style which the speakers have at least passive command of and therefore does not have to belong to their core competence, and which is less marked. For dialects the difference consists in a limitation of the geographical scope. Sociolects are like dialects, except that the marking difference consists in a sociological restriction, possibly compounded with a geographical restriction. Dialects, therefore, are to be defined in direct correlation with a given area, which is smaller than the area covered by the available standard. We all know of isoglosses, and it has often been said that the idiosyncratic spread of individual isoglosses disturbs the notion of "dialect". This, however, does not follow. Typically, an isogloss defines the specific area for one single (usually phonological or lexical) rule with the G-marking "1". We now define a (standard) dialect for area A as the style (1,0,0) resulting from the maximal application of all rules 1,0,0 whose specific area includes A, — with an automatic fall-back on 0,0,0 rules whenever no alternative marked for an area including A is available. A is thus at most the intersection of all isogloss areas covering A. If A is smaller than that intersection, then the dialect of A is identical with the dialect of that intersection. In such a case it makes better sense to speak of the dialect of the intersection, rather than just of A. In precisely the same way we can define a style (1,1,0) or a style (1,1,1) for some area A. Often, as we know, but certainly not invariably, there is no standard dialect (1,0,0) for a given area, usually when the area is in or near the cultural and political centre of the speech community. In many capitals standard speakers are without any dialectal markings, but substandard speakers are marked for dialect. In such cases there is a (1,1,0), and probably also a (1,1,1) dialect.

Amsterdam, for example, is the intersection of a number of 1,1,0 isoglosses, and we speak accordingly of the Amsterdam dialect, which is the style (1,1,0) for that city. One feature of this dialect is the use of the relative pronoun wie in all cases where the standard has die, i.e., after antecedents that take the common gender article de. Thus: de uren wie ik gemaakt heb, versus standard de uren die ik gemaakt heb (the hours that I made). Moreover, informal standard has the relative pronoun wat where the formal standard has dat, after antecedents taking the neutre article het: het geld wat ik ontvangen heb, versus formal standard het geld dat ik ontvangen heb (the money that I was paid). There thus seems to exist a generalised rule alternative selecting relative wie/wat for standard die/dat, marked 1,1,0 for Amsterdam. Within Amsterdam, however, there is a small area in the centre where wie is generalized as the relative pronoun throughout: het geld wie ik ontvangen heb. This is a different rule alternative: wie for relative pronoun, marked 1,1,0 for the few blocks in the centre where this is done. If we
therefore want to describe the dialect of that small area in the centre, the latter
rule alternative will replace the former, but all the other features that are general
for Amsterdam as a whole will remain, unless, of course, they are affected in a
similar manner. We thus have a model that predicts the possibility of setting up
implicational scales in the sense of DECOMP and BICKERTON.

The notion of idiolect falls entirely outside the scope of this theory. We started
off by saying that we wish to capture language, with its variable phenomena, as a
bit of social reality, so that by definition the otherwise valid notion of idiolect is
irrelevant for our purpose. Idiolects are part of social reality only in the trivial sense
that they are known or at least recognised within usually restricted groups of per-
sonal acquaintances. But they cannot be part of social reality in the sense that we
are considering here and which implies that a newcomer in a community must as-
similate a set of rules in order to be fully socialized. For linguistic socialization in
particular it is contradictory to say that full socialization is dependent upon the
assimilation of idiolects. We see in fact that the marking system in SMARS does
not allow for idiolectal variation: there is no parameter for "individual person".

An important and inevitable question arises in this context: if SMARS embodies
a certain idealization as we have said it does, what counts as data and how do we
assign rule markings? Clearly, LABOV'S correlative data form a very powerful data
base, and we shall discuss a few examples. But a lot will depend, apart from the
actual data provided, on our interpretation of the data. This is simply inevitable,
and of course a very common phenomena in science, the human sciences in parti-
cular. We can feel fortified by the consideration that we have undeniable and
irresistible intuitions of sociolinguistic markings for variable features. In this respect
the situation is just marginally better than in established theoretical linguistics,
where the data problem has not been solved in a fully principled way, and where
in practice individual and uncontrolled intuitions or judgements replace data.

Let us consider an example. LABOV (1972a: 239) provides correlative data on
the pronunciation of -in for standard -ing in gerundives for white New York adults.
His data is presented as in figure 5:

![Fig. 5](image)

Class and style stratification of -in for
-ing for white New York City adults.
Socio-economic class scale: 0–2,
3–6, 7–8, 9.A, casual speech; B, care-
ful speech; C, reading style. From
LABOV (1972a:239)
The question now is: what is the SMARS marking for the rule alternative giving *-in* as against standard *-ing*? Assuming that speakers will aim at accepted neutral standard pronunciation in the highly monitored style C, we can accept that the dramatic decrease of *-in* frequencies for all classes in that style shows that *-in* selection is non-standard. We know, moreover, anyway, that *-ing* is standard. But in what ways is *-in* non-standard? Does the high percentage in style A for the classes 0–2 justify an S-marking "1"? It seems not, since lower class marking for this rule alternative cannot explain the notable percentage in style A for 7–8 and the 10% of class 9. A reasonable interpretation seems to be that we have a dialectal marking "1" (i.e., on G), and a 0-marking for S. The differences for style A can then be explained as a result of attitudinal tendencies: higher-class speakers tend to strive for standard even in casual speech, but lower-class speakers show a much weaker tendency to select the standard style target for casual speech situations. The argument can be carried over to the parameter I: here too we can assume a 0-value. The percentages for style B again suggest a pull towards the standard, which is relatively stronger for lower classes than for the upper end of the socio-economic scale. Yet, I-values are more complex: we shall discuss them in a moment.

Whether this interpretation is correct is hard to say on the basis of these data alone. But it is interesting to compare the New York data on *-in* for *-ing* with those collected by TRUDGILL (1974), and quoted in LABOV (1972a: 242), for Norwich in England.

![Graph](image)

**Fig. 6** Class and style distribution of *-in* for *-ing* in Norwich. D: reading of words lists. From LABOV (1972a: 242) and TRUDGILL (1974: 92)

Here working class and middle class are clearly distinct. But, as in New York, the higher social strata still show a fair percentage in style A. The same reasoning, therefore, applies: this rule alternative is dialectal, but sociologically unmarked. The systematic differences between working and middle class for style A can be attributed to institutionalized attitudes towards the standard. We notice, moreover that the A-percentages for all classes are significantly higher than in New York, which points at attitudinal differences. Then, the middle classes show the sharpest decline between the styles A and B, whereas for the working classes it
comes after B. This suggests a stronger dialect loyalty among working class people than among the middle classes: the former come out frankly with this dialect feature even in careful speech. But they apply a different style target, the standard, when reading. On this reckoning, the I-marking for -in in Norwich must again be "0". On this interpretation the SMARS values for this rule alternative are identical for New York and for Norwich, but institutionalized attitudes towards dialect features differ.\(^{14}\)

Attitudes must be an important factor in historical change of SMARS markings. If, at some later time, the Norwich middle classes show no or almost no -in-pronunciation even in style A, we may infer that the rule is then clearly marked "1" on S. Its marking will then be "1,1,0", which is, as we have seen, a typical dialect marking in capitals or other urban centres. It looks as though such a development is, in fact, beginning to take place for the feature -in versus -ing in Norwich.

This way of reading the graphs implies, in practical terms, that rule alternatives are considered to be non-minimally marked for the parameter G when their effects show up in the speech produced by all speakers of a given area and only these. If all speakers of a given area and only these produce a certain feature, it gets the marking "1" for G and "0" for S. Proportional differences between speakers of different socio-economic status are attributed to attitudinal factors, as the proportional variations in the speech of one single individual are attributed to attitude shifting. If, however, a feature does not occur at all in the speech of some socio-economically defined groups, but does occur in the speech of other groups, then the rule underlying that feature will be assigned a non-minimal marking ("1\(^a\), "2\(^a\), or "3\(^a\) for S, next to the "1" already assigned to G. For these two parameters, therefore, the criterium is fairly clear: it is the active competence as reflected in (interpreted) output that determines the values.

For the interactional parameter, however, this criterion cannot be applied, since it is normal for competent speakers to be actively competent in both formal and informal styles. Here we may think of a different approach to an optimal interpretation of the available data. For this parameter we might look at the general style characteristics of the text from which the data is drawn. Suppose we have a rule alternative already marked "1" for both G and S, but with the I-value still undetermined ("\(\alpha\)". Suppose we have a text displaying a style target (1,1,\(\alpha\)), that is, we know that the speaker tends to maximize the use of rules marked "1" for G and S. Now we observe that this and other similar texts are used indiscriminately for both formal and informal speech. (There are, of course, many independent diagnostic features that help in making this distinction, such as the use of first or second names, formal and informal pronouns of address, etc.). That is, we find that systematically the 1,1,\(\alpha\) rules used in the text and therefore contributing to the stylistic profile show constant frequencies for casual and careful speech styles (but tend to fall away in those styles, such as reading, where we may assume that the dialect is avoided). In such a case we will be tempted to conclude that the style of the text is (1,1,0), and not (1,1,1). It then follows that the rule alternatives determining the style character of the text in question will also be marked

25
It is important to keep in mind that these assignments are in fact hypotheses, not "instrument readings". In cases such as the one described the hypothetical "0" marking for I will turn out to be wrong if it is found, for example, that the speaker in question is not typical for the community and quite generally displays unadapted or deviant behaviour.

In the actual cases of figure 6 we assigned the values "1" and "0" to G and S, respectively, for the reasons given. We tentatively assigned "0" also to I, given the fact that middle class speakers in Norwich still show -in behaviour in style B, albeit minimal, and in view of the fact that for working class speakers the frequencies hardly diminish. But in order to have greater certainty we should really know if this distributional pattern can be generalised for all or almost all 1,1,1 rule alternatives in the texts in question, and for all speakers falling under the same socio-economic description.

The assignment of SMARS values to the rule alternatives for which a Labovian diagram is available is far from a simple matter. These diagrams are hard to read, and there can be little doubt that they are the result of the interplay of a variety of factors, only some of them being linguistic. In the informal discussions around these diagrams the distinction between the linguistic competence factors and other, external, factors is usually not made or not observed consistently. It is the purpose of this paper to isolate the linguistic component from the others. In order to do that properly, we must develop the art of finding the most reliable interpretation of the tables and diagrams, and the interpretations ventured above for the figures 5 and 6 are nothing but a, probably clumsy, attempt.

The problem would be simpler if comparable rule alternatives always showed the same statistical pattern for a given community. Unfortunately, this is not so. LABOV (1972a:113) gives a chart for the pronunciation of aspirated initial t for standard θ (as in thing), also for the nine social classes in New York City.

The pattern here is the same as we had in figure 5, with the exception that there is sharp division (as in Norwich) between the working classes (0-4) and the middle classes (5-9). Does this mean that attitudes are not defined with respect to the dialect as a whole, but rather with respect to either individual rule alternatives or bunches of them? If so, we can no longer say that the sharply distinct or diffuse stratification in a Labovian graph is a function of a SMARS marking, a style, and attitudes with respect to markings and styles. We will then have to build in a further factor, such as, perhaps, the degree of awareness a speaker has of the specific rule marking. But it is also possible that the elementary marking system used so far is too rough to capture finer distinctions. In any case, if we allow attitudes and awareness factors to play a role in the interpretation of Labovian graphs (as Labov, Trudgill, and others, do) then this only makes sense if it is clear what speakers are aware of and what they have attitudes towards. This is precisely what we aim at clarifying.

A last example may be of use. A very typical pattern in Labovian diagrams is the so-called cross-over pattern (LABOV 1972a:245):
Rule alternatives showing this pattern can be interpreted as having a SMARS marking "0,0,0", and as having an alternative rule in every non-standard style of that community. That is, it will not be used in any non-standard style for lack of a marked alternative: it occurs just in the style (0,0,0). This is precisely what makes it a prestige feature or rule alternative. Labov interprets the probability value of a prestige feature as a function of socio-economic class, style and an index of linguistic insecurity. In our theory the p-value of any feature is a function of a SMARS marking, a style, an attitude with respect to the marking and the style (with "socio-economic class" as a determinant of attitudes), and possibly other factors such as awareness.

LABOV and others have used the term *hypercorrect* for the cross-over phenomenon. WINFORD (1978) prefers to reserve this term for what has traditionally been called hypercorrection, and coins *supercorrection* for the cross-over phenomenon. I shall follow Winford, and take hypercorrection in the traditional sense. Thus, we find forms like *I plays*, *I works* in many English speaking decrrealizing areas (especially the Carribbean), or *Hoxford* for *Oxford* in an h-dropping dialect (such as Caribbean English).

It is not too difficult to indicate the lines along which a formal definition of (traditional) hypercorrection could be developed in terms of SMARS markings. Hypercorrection, in its classical form, is the result of the following typical situation. Often, as we have seen, a standard rule (or phonological or lexical selection) has a non-standard variant. Let us say that rule $R_{1a}$, marked 0,0,0, produces effect $r_1$ in environment $E_1$: $R_{1a}[0,0,0] \rightarrow r_1/E_1$. Then rule $R_{1b}$, which is a non-standard variant, i.e., which has one or more occurrences of "1" in its marking, produces effect $r_2$ in environment $E_1$: $R_{1b}[^1\ldots^1] \rightarrow r_2/E_1$. Now it so happens that realization $r_2$ does occur in the standard, but in a slightly different (grammatical, lexical, phonological) environment $E_2$, whereas the non-standard in question has no marked alternative for $r_2$ in $E_2$. The grammatical feature (or lexical selection or phonological realization) $r_2$ thus occurs in the non-standard in question in environments ($E_2$) where it's all right also in the standard, as well as other (but not *very* different) environments ($E_1$) where the standard has $r_1$. Such a situation typically leads to stigmatization of $r_2$, with the result that the non-standard tends to create a new rule whereby $r_1$ is selected also in those cases where the standard has $r_2$. The new rule is a *hypercorrect* rule. Speakers using it mistakenly believe that it is part of the standard. The following table summarizes this process:
\[ R_{1a}[0,0,0] \rightarrow r_1/E_1 \]
\[ R_{1b}[\ldots 1 \ldots] \rightarrow r_2/E_1 \]
\[ R_{2a}[0,0,0] \rightarrow r_2/E_2 \]

There is no rule \([\ldots 1 \ldots]\) that gives \(r_3/E_2\).

Now the style \((\ldots 1\ldots)\) will have a tendency to develop:
\[ R_{2b} \text{ (mistakenly thought to be } [0,0,0]) \rightarrow r_1/E_2. \]

Fig. 8

Definition of hypercorrection

For example, \(R_{1a}\) gives -s in 3rd person singular present tense of any finite main verb. \(R_{1b}\) gives -∅ in the same environment. But -∅ also results from \(R_{2a}\), where the environment is: all other persons present tense of finite main verbs. And there is no rule in the non-standard in question yielding a different realization from -∅ in the environment of \(R_{2a}\). Now -∅ after main verbs in the present tense in finite verb position may become stigmatized, and a spurious standard rule \(R_{2b}\) may develop giving -s for all persons present tense of finite main verbs. \(R_{2b}\) is not standard, but it may, of course, become standard, either throughout the whole speech community or in only part of it, if a sufficient large number of speakers start thinking that it is standard: in no other reality than social reality does belief, provided it's widespread enough, create truth.

Let us, finally, consider the notion of diglossia, as introduced into linguistics by Ferguson (1959). Impressionistically speaking, diglossia is the phenomenon that there is no single standard style of speaking language, but either a high or a low style. In Greece, for example, we have a high variety, called katharevousa, generally used in printed texts (but not in literary texts) and rarely spoken, besides the low variety, the demotic, used for most purposes, including literature, but excluded from certain formal interactional situations. In the German part of Switzerland a high variety is provided by the standard language spoken in Germany (though with a certain allowance for local features), and the low variety is Swiss German, mostly dialectal. Such a situation can be formalized in SMARS in the following way. The speech community is unilingual, but, as in the case of English, there is no single standard style. In a diglossic situation, however, this is not due to geographic divergences, but to a division between formal and informal situations. The set of rule alternatives \(\{0,0,0\}\) is insufficient for the composition of texts: in many cases during the production of sentences speakers will find no 0,0,0 rule among the available alternatives. And when they aim at a standard style target, they must choose between a rule alternative marked "2" for I and another marked "1". A speaker may try to maximize the application of 0,0,0 rules, but where no such rule is available he must fall back either on 0,0,2 rules or on 0,0,1 rules. In the former case he will produce a high standard, in the latter a low standard. We cannot characterize the respective styles as \((0,0,2)\) and \((0,0,1)\), since that would imply an attempt at maximization of 0,0,2 or 0,0,1 rule alternatives. We must use another notational device, such as a subscripted standard style notation: \((0,0,0)_2\) and \((0,0,0)_1\), respectively.

28
4 Conclusion

A sharper focussing of the notion of competence and a less abstract form of idealization enabled us to sketch a theory of linguistic variation based on rule alternatives marked for social meaning. This theory does not predict proportional variation the way variable rule theory does, but it shows how proportional variation may occur, in which case it is due to non-linguistic factors, and not to linguistic factors, as VRT has it. Proportional variability is attributed to attitudes and to imperfect mastery of certain style targets. The theory enables us to define notions such as "standard language", "dialect", "sociolect", "unilingual speech community", "hypercorrection", "diglossia", and no doubt other notions as well, in a much more precise manner than has been possible so far in other descriptive models, including VRT.

In this theory, a grammar of a language is a socially meaningful alternative rules system (SMARS). The most immediate implication for grammatical descriptions is that different varieties of the language should be generated by rule alternatives (or rule ordering alternatives) differing in preferably minimal ways. To the extent that a grammatical description succeeds in offering one coherent rule system, with only minimal differences among rule alternatives, accounting for different linguistic varieties, the description in question gains in psychological plausibility. LABOV (1969) provides an outstanding example of such a description. See also SEUREN (to appear) for a description of tense and aspect in Sranan, where three varieties of the language are accounted for by one rule system, with minute variations accounting for the varieties.

Notes

1) See LABOV (1972a:248): "A small number of sociolinguistic markers rise to overt social consciousness, and become stereotypes. There may or may not be a fixed relation between such stereotypes and actual usage."

2) A large number of experimental studies have confirmed this. See e.g. LAMBERT (1967), LABOV (1966:405–54), LABOV (1972a:248–51).

3) BLOOMFIELD is a case in point. In his Language (1933) he shows great perceptiveness for variability phenomena and related attitudes (esp. pp. 496–9). Yet he makes no attempt at incorporating such phenomena into his general theory.

4) Germany provides a few counterexamples to this: KLEIN (1974), LUELSDORFF (1975), BIERWISCH (1976). The ideas developed in this paper have elements in common with all three, but there are also important differences. Rather than engage in a detailed critical discussion of these theoretical proposals, I prefer to offer my own ideas as an alternative, hoping that whatever is of value in each of these approaches will be put to good use by future workers in this field.

5) This is precisely the idealization put forward in BIERWISCH (1976:304): "Notice, however, that this involves an idealization of a particular sort. It would be the competence of a kind of superego of C [= speech community], that is, of a speaker who knows completely all the varieties used in C."


7) See SEUREN (1978) for a more detailed discussion.

8) For an example of socially meaningful variation where rule ordering is involved, see the French sentences discussed at the end of section 2b, and the references given in note 12.
“Attitude” is a much discussed notion in social psychology, as well as in the sociolinguistic literature more oriented towards social psychology. I shall not go into the details of this difficult notion here, but simply give a rough indication of the sense in which the term is used in this paper. “Attitude” is here used to indicate an individual’s identification (positive attitude) or dis-identification (negative attitude), with respect to a given phenomenon; of, in other words, his wish to be associated with or disassociated from that phenomenon.

Cp. KLEIN (1974: 54): “What matters in linguistics is not the determination of actual frequencies in texts, although this may be a useful preliminary. The object of our description is not finite bits of text, but rather behavioural regularities, for which the texts simply provide evidence.” (Translation mine)

KAY (1978: 72) points out correctly that in a large number of empirical studies the theoretical status of variable rules is left unclear, and that often a tacit assumption is made to the effect that they somehow belong to “community grammar”. LABOV (1972c) argues in favour of a social conception of grammar, in order not to land the descriptive linguist with the impossible task of describing idiosyncratic phenomena (ideolects) associated with individual persons. No clear answer has been given to date to the important question of “where grammars stop”. LABOV’S answers are practical (1972c: 77), but in spite of a wealth of factual information, no principled answer is given.

Examples of different rule orderings corresponding with dialectal differences are numerous in phonology. See KIPARSKY (1968: 177–9), or BROWN (1972: 165).

Creole languages are often characterized by the fact that lexical gaps or deficiencies are filled from the dominant parent language. Sranan, for example, (the Creole of Surinam) falls back on Dutch for missing lexical items. But minimization of Dutch words is a measure of the purity of a Sranan text.

TRUDGILL (1974: 93–5) notes that women use significantly less -in forms than men. His explanation is an attitudinal one, based on various social psychological considerations.

References


FASOLD, R. W., 1970: "Two models of socially significant linguistic variation". Language 46.3, 551–63.


