118. The Grammar of Varieties/Varietätengrammatik

1. Linguistic variability and its analysis

Natural languages such as Greek, Russian or Tagalog are not very well-defined entities; they include numerous varieties - dialects, registers, sociolects, historical variants -, which share a number or regularities while differing in others. A speaker's linguistic knowledge reflects this heterogeneity in varying degrees. In childhood, he (or she) normally learns to replicate one such variety, the speech habits of his social environment, 'up to perfection', i.e., such that other members of this social group accept the speaker as a 'native', as one of theirs. In the course of life, he is regularly faced with many other varieties of 'his' language and gathers knowledge about these. This knowledge may be less 'perfect', it may also be asymmetric with respect to comprehension and production (many speakers understand other dialects perfectly well but cannot reproduce them perfectly well); but it may also be felt more perfect than his original knowledge, for example if dialectal pronunciation and grammar are 'improved' in school. This variability faces the linguist who takes reality seriously with three descriptive tasks:

A. The characteristic properties of the individual varieties must be described. These properties include phonological, morphological, syntactical regularities, the specific lexical repertoire, but also peculiarities in communicative behaviour (for example the

choice of appropriate address forms or the rules of turn taking).

B. The relationship between the individual varieties must be accounted for. This task is difficult for at least three reasons. First, the descriptive tools of modern linguistics are usually designed for homogeneous, idealised forms of language; hence, additional devices must be created to adapt these tools to the comparison of various varieties. Second, the range of properties in which varieties may agree or differ precludes the application of a uniform method; phonological variation can normally not be described with the same instruments as differences in forms of address. Third, the differences between two varieties are often not categorical but gradual, i.e., varieties do not differ in the presence or absence of a particular rule but in its weight, as reflected, for example, in the more or less of its application.

C. It must be described how linguistic variation correlates with extralinguistic factors. These include, for example, social class, geographical distribution, the specific communicative situation, the medium (written or spoken), or development over time, be it of an individual ('language acquisition') or a speech community ('historical change').

One way to deal with these tasks is *variety grammar*. It is easy to apply, and if appropriate empirical evidence is available, it allows a precise modelling of virtually all types of grammatical variation. It is less appropriate for lexical variation and not apt for pragmatic aspects of variability. In what follows, we will only be concerned with syntax; but it should be clear that this analysis can easily be extended to other parts of grammar, for which precise rules are available.

Modern linguistics has provided the researcher with several types of grammars which allow a precise description of syntactical rules. They include, for example, phrase structure grammar (context free or context sensitive), transformational grammar, dependency grammar, categorical grammar. Irrespective of their respective advantages and disadvantages, about which opinions are strongly at variance and which will not to be discussed here, they have two features in common. First, they are explicit and precise; this is in remarkable contrast to most of traditional school grammar, which still dominates empirical research on languages. Second, they were developed for homogeneous

languages, hence they are not fit for the tasks A-C mentioned above. Variety grammars try to overcome this inadequacy while keeping the explicitness and precision of a formal grammar. The core idea is to take such a formal grammar, say a transformational grammar, as a constant, stable base of comparison and to restrict it in such a way that the specific properties of individual varieties are precisely captured; this is done by assigning probabilistic values to the rules of the underlying common grammar. Thus, the rules of all varieties are the same; what differs from variety to variety is the 'weight' of these rules; this weight is described by the probability with which a rule is applied in a particular variety, including the borderline cases that the probability is 1 (i.e., the rule is obligatory) and 0 (i.e., the rule does not show up at all). This procedure is sufficiently flexible to account for categorial distinctions between varieties as well as for very fine-grained differences. In section 2, this idea will be explained in more detail; sections 3-4 are devoted to various types of probabilistic weighing. These parts are fairly theoretical; but variety grammar is in first place a versatile instrument for empirical analysis; in the concluding section 5, we will therefore briefly discuss some aspects of its practical application.

2. Variety space, overall grammar and probabilistic weighing

A *variety space* is an ordered set of varieties under investigation. Suppose someone wants to study the syntactical properties of
- love letters as compared to business letters (factor *register* with values r_1 for love letters and r_2 for business letters);
- around 1900, 1950 and 2000 (factor *time* with values t_1, t_2 and t_3);
- in Boston and in Oxford (factor *space* with values s_1 and s_2).

This yields a three-dimensional variety space with $2 \times 3 \times 2 = 12$ varieties; for instance, (r_1, t_2, s_t) is the variety of love letters of Bostonians around 1950. This may not be a particularly interesting variety space; but this is a matter of the researcher's individual preferences, on the one hand, and of what looks scientifically meaningful and rewarding, on the other.

An *overall grammar* is a formal grammar of whichever type, which covers *all* syntactic rules that show up in at least one of the va-

rieties in the variety space. Thus, it does not yet discriminate between varieties and is therefore not very meaningful in itself; it says what is possible in the variety space but not what is the case in individual varieties. To this end, its rules must be appropriable restricted; this is done by probabilistic weighing. The idea is best explained with (context free) phrase structure grammars. In such a grammar, certain rules can be applied alternatively, namely those rules which rewrite the same symbol. Thus, a context free grammar may contain, for example, the following rules for noun phrases:

(1)   1.   NP → N
        2.   NP → DET N
        3.   NP → DET ADJ N

i.e., an NP can be expanded as a bare noun *{books),* as a noun preceded by a determiner *(the books, some books)* or as a noun preceded by an adjective and a determiner *(the old books).* One of those rules must be applied in a derivation. All alternative rules form a sort of 'rule block', often indicated by brackets:

(2)
$$NP \rightarrow \left\{ \begin{array}{ccc} N & & \\ DET & N & \\ DET & ADJ & N \end{array} \right\}$$

Note that (2) is not a rule but a conventional abbreviation for three rules. Rule blocks may contain an arbitrary (but finite) number of rules; it is useful to include rule blocks with just one rule (which, since there is no alternative, is then obligatory).

Whenever the lefthand symbol, here NP, shows up in the course of a derivation, then one of the rules of the rule block *must* be applied. But the individual rules may vary in their likelihood: there are rules which are very likely in a particular variety, whereas others are rare, and this may vary from variety to the other. This fact can be covered by assigning probabilities to the rules of a block. The probability of some event, here the application of a rule, is normally expressed by a real number between 0 and 1, where 0 stands for 'does never happen' and 1 stands for 'happens in all possible cases'; values in between refer to intermediate stages between these extremes. The rules within a block can thus have different probabilities; but they inevitably must add up to 1, since one of them must be applied. In other words, the entire block has the probability 1 ('obli-

gatory'), which can be distributed in various ways over its rules, including the borderline case that it contains only one rule (which then gets the probability 1). Consider, for instance, the NP-example in (1). In some variety A, rule NP -> N may have the probability 0.2, rule NP -> DET N may have the probability 0.3, and rule NP -> DET ADJ NP may have the probability 0.5, whereas in variety B, the values may be 0.7, 0.1 and 0.2, respectively. In other words, noun phrases are of the same type in A and B, but in B, more complex NPs are preferred. In some variety C, the values might be 0.8, 0.2 and 0.0, respectively; that means that in this variety, NPs tend to be very simple, and the 'most complex' possible pattern DET ADJ N does not exist at all - it has probability zero.

We can use this technique to model the NP acquisition of a second language learner who passes through a series of 'learning stages', say the six stages V_1 - V_6. They constitute a very simple, one-dimensional variety space. The overall grammar is a fragment of a context-free grammar which includes all NP rules observed in these six varieties. Let us assume that it contains the three rules mentioned above and two more rules:

(1')   ...
        4.   NP → DET N   ADJ
        5.   NP → DET N   ADV

Rule 4 generates NPs such as *the books old,* as might well appear in the language of a learner; rule 5 generates structures such as *the books there,* A variety grammar - or more precisely, a fragment of a variety grammar - could then look as follows:

(3)

	V_1	V_2	V_3	V_4	V_5	V_6
NP →N	0.9	0.6	0.3	0.2	0.2	0.2
NP →DET N	0.1	0.3	0.3	0.3	0.3	0.3
NP →DET ADJ N	0	0	0	0	0.4	0.4
NP →DET N ADJ	0	0.1	0.3	0.4	0	0
NP →DET N ADV	0	0	0.1	0.1	0.1	0.1
Whole rule block	1	1	1	1	1	1

This is a precise account of a whole bundle of developmental processes. In the first learning stage, V_1 bare nouns are predominant, with a small share of NPs such as *the book;* more complex NPs are still absent. In V_2, we note the first NPs with an adjective; but this adjective is placed after the noun; moreover, determiners become more frequent. This development continues in V,: bare nouns are

increasingly rare, adjectives are found more often (though still in 'wrong' position), and we note the first occurrences of a new pattern. There is hardly any change from V_3 to V_4. From V_4 to V_5, there is only one development - but a very salient one: the learner all of a sudden got the 'right' position of adjectives within an NP. There is no change from V_5 to V_6, i.e., the learner may have reached the endpoint of this acquisitional process. Note that this final variety may still be very different from the 'target variety', i.e., the language to be learned.

The example is very simple: there is only one dimension of variation with a limited number of varieties, and only a few NP structures are considered; but it should be clear that variety space as well as overall grammar can easily be made much richer. It should also be clear that continuous as well as abrupt changes (e.g., complete dropping of a rule) can be modelled with a degree of precision that is only restricted by the available data.

3.   Probabilistic weighing I: Phrase structure grammars

3.1. Probabilities

In what follows, we shall informally sketch some basic ideas of probability theory (for a thorough introduction, see e.g., Milton and Arnold 1990). Probabilities are numbers associated with the possible outcomes of repeatable incidents. Such an incident is, for instance, the toss of a die: it has six possible outcomes, which are mutually exclusive and one of which *must* happen: 'one' or 'two' or 'three' or 'four' or 'five' or 'six'. Usually, one is not only interested in the likelihood of these 'elementary events' but also in combinations, such as the likelihood of having 'one or two' or 'not five', and so on. The set of elementary events is called *sample space* O; in the case of tossing a die, O = {o_1, o_2, o_3, o_4, o_5, o_6}, where o_1 is the event that the die falls on 1, and so on. The set of all elementary and complex events is called *event space* F. F is usually considered to be the power set of O; thus, the subset {o_1, o_3, o_5} is the event that the die falls on an odd number. The probability of an event is expressed by a real number p between 0 and 1, where 1 is 'the certain event'. The event {o_1, o_2, o_3, o_4, o_5, o_6}, that is, the sample space itself, includes *all* possible outcomes; hence, the sample

space gets the probability 1. This entire probability of 1 may distributed in different ways over the elementary events. If the die is not loaded, there is no reason to assume that one outcome should be preferred over another; hence, a priori considerations lead to the assumption that each elementary event has the probability of 1/6. How do we get from there to the likelihood of a complex event such as {o_1, o_3, o_5}, i.e., the chance that the die falls on an odd number? It is apparently the same probability as getting an even number, i.e. 1/2, and this value is reached by adding the probability of the elementary events: $p(o_1) + p(o_3) + p(o_5) = 1/2$. The probability of getting a five is 1/6, the probability of getting *no* five, i.e. of the event {o_1, o_2, o_3, o_4, o_6} is $1 - 1/6 = 5/6$.

Suppose now the die is loaded, such that some outcomes are more likely than others. Then, a priori considerations are no longer helpful. But by tossing the die many times, we can empirically determine how often the six possible elementary events occur. If, for example, o_5 occurs n times in m trials, then n/m is called the relative frequency f of o_5, in brief $f(o_5)$. This value stabilises with the number of trials, and we can *interpret* it as the probability p of o_5. Under this statistical interpretation, the probability of an event is the limit of its relative frequency, which is empirically determined in some finite experiment. Suppose, for instance, that many tosses of a loaded die have lead to the following relative frequencies:

(4)

event	o_1	o_2	o_3	o_4	o_5	o_6	O
relative frequency	1/18	1/9	1/9	1/9	1/9	½	$\Sigma = 1$

Note that these values, as resulting from empirical observation, are not probabilities but relative frequencies: they are *interpreted as probabilities*. This statistical interpretation of probability underlies virtually all empirical investigations, including those of variety grammar. It is not without mathematical problems, but we shall not enter this discussion here (see, e.g., Stegmüller 1973).

In variety grammar, the events are not tosses of a die but applications of rules. In order to assign them probabilities, two questions must be answered: (1) What is the sample space? (2) How can we empirically determine the probabilities of a given sample space? The second question will be addressed in section 6. The answer to the first

question depends on the type of grammar; we begin with context free grammars.

3.2. Suppes type weighing

Context free phrase structure grammars are an attempt to formalize the traditional notion of constituent structure grammar. They consider a language to be a (normally infinite) set of sentences, and they describe these sentences as well as their structures by 'generating' them, i.e., by deriving them systematically from a 'start symbol', usually called S. S is replaced by some string of symbols, these symbols in turn are replaced by other strings of symbols, until this process cannot be further continued. Replacement possibilities are explicitly defined by a set of rules. These rules must be formulated such that the symbol strings which are eventually derivable from the start symbol correspond precisely to the possible sentences of the language to be described. A more precise definition is as follows:

(5) A context free grammar G is a quadruple $<V_T, V_N, S, R>$, where
- V_T is a finite set of symbols, the *terminal vocabulary* (roughly speaking, the 'words' of the language);
- V_N is a finite set of symbols, the *nonterminal vocabulary* (roughly speaking, the 'syntactic categories');
- S is a distinguished element from V_N, the 'start symbol' or 'axiom';
- R is a finite set of sequences of the form $A \rightarrow x$, where $A \in V_N$, $x \in (V_T \cup V_N)^+$, where $(V_T \cup V_N)^+ = (V_T \cup V_N)^* - \{0\}$ (R includes the 'replacement rules'; A is some syntactic category and x an a nonempty string of symbols of whatever sort).

This is best illustrated by a simple example of a context free grammar which produces correct sentences of English - not all sentences, of course, but infinitely many:

(6) VT = {the, some, lions, pigs, mice, love, see, dance, dangerous, tiny, there, here, and, or}

VN = {S, NP, VP, CON, DET, N, ADJ, VI, VT, ADV}

R = {r_1, ..., r_{22}}, where

r_1:	S	→	S CON S
r_2:	S	→	NP VP
r_3:	S	→	NP VP ADV
r_4:	VP	→	VT NP
r_5:	VP	→	VI
r_6:	NP	→	DET N

r_7:	NP	→	DET ADJ N
r_8:	NP	→	DET N ADV
r_9:	DET	→	the
r_{10}:	DET	→	some
r_{11}:	CON	→	and
r_{12}:	CON	→	or
r_{13}:	N	→	lions
r_{14}:	N	→	pigs
r_{15}:	N	→	mice
r_{16}:	VT	→	love
r_{17}:	VT	→	see
r_{18}:	VI	→	dance
r_{19}:	ADJ	→	dangerous
r_{20}:	ADJ	→	tiny
r_{21}:	ADV	→	here
r_{22}:	ADV	→	there

These rules allow, for example, the following derivation:

(7)

S	(start symbol)
NP VP ADV	(by rule r_3)
NP VT NP ADV	(by rule r_4)
DET N VT NP ADV	(by rule r_6)
DET N VT DET ADJ N ADV	(by rule r_7)
some mice VT DET ADJ N ADV	(by rule r_{15})
some mice love DET ADJ N ADV	(by rule r_{16})
some mice love the DET ADJ N ADV	(by rule r_9)
some mice love the dangerous N ADV	(by rule r_{19})
some mice love the dangerous lions ADV	(by rule r_{13})
some mice love the dangerous lions there	(by rule r_{22})

After the application of r_{22}, no further rule is applicable, and the derivation is terminated. In many but not in all cases, another rule could have been applied, resulting in somewhat a different structure. It is these alternatives which are probabilistically weighed. Our sample grammar can be broken down into ten rule blocks with alternative rules, one for each of the ten nonterminal symbols. Each of these blocks is considered to be a sample space, hence it gets the entire probability 1. Empirical investigation of corpora which are representative for some variety must show how this probability is distributed over the individual rules which constitute the entire block. This simple way of probabilistic weighing was independently developed by several mathematicians. It was Patrick Suppes who first applied it to the study of natural language. Therefore, it is called here 'Suppes type weighing'. We give a precise definition:

(8) A probabilistic grammar G_S of type Suppes is a quintuple $<V_T, V_N, S, R, p>$, where $<V_T, V_N, S, R>$ is a context free

grammar, and p is a function on R into the real number such that

(a) for all $r \in R$, $p(r) \; 0$

(b) for all $A_i \in V_N$: if there are m many sequences $x_j \in (V_T \cup V_N)^+$ such that $A_i \to x_j \in R$, then

$$\sum_{j=1}^{m} p(A_i \to x_j) = 1.$$

Thus, the function p assigns a value equal or larger than 0 to each rule (condition (a)); condition (b) guarantees that the probabilities of all rules within a block (i.e., those which expand the same symbol) adds up to 1; m is the number of rules within a block.

A *variety grammar,* type Suppes, is a set of Suppes type grammars with the same underlying context free grammar (the 'overall grammar') and varying probabilistic weighings, one for each variety.

In a Suppes type grammar, the probabilistic value of a rule does not 'look beyond its block', i.e., it is the same, independent on what happens in other blocks. This makes probabilistic weighing simple, but may be empirically inadequate. According to our sample grammar (6), an NP can be introduced in 'subject position', i.e., by rule r_2, or in 'object position', i.e., by rule r_4. It may now be that the further expansion of such an NP by rules r_6 - r_8 is sensitive to this distinction, because, for instance, NPs in subject position might on the average be much simpler than NPs in object position. There are some ad hoc measures to solve this problem within this type of probabilistic weighing, for example by assuming different rules for 'object NPs and 'subject NPs. But this is not very intuitive, it leads to an inflation of syntactical categories, and it does not work for recursive rules, such as r_1. A better way are probabilistic weighings which take into account potential dependencies between rule applications. This will be shown in section 3.3. First, however, we will briefly discuss two potential misunderstandings.

The probabilistic weighings considered here relate to *rules,* not to *sentences* (or to the derivations which generate these sentences). The probability of a sentence is always close to zero and hence completely uninteresting. A somewhat more realistic grammar than (5) may contain, for example, 200 rules in 50 blocks. Then, the average probability of a rule within block is 0.25. Suppose now that 40 rules altogether are applied in a derivation (not every block must be represented since its 'lefthand symbol' may never be in-

troduced and hence need not be expanded). Then, the probability of the resulting sentence is 1 divided by 0.25^{40}, a ridiculously small number, in contrast to the probabilistic weight of each rule, which is substantial. The second misunderstanding is somewhat hilarious but still occasionally found in the literature. It is the argument that the speaker has no 'counter' in his or her head that could tell him how often to apply a rule. This is a curious misunderstanding of the notion of probability; a die does not count, either, how often it has fallen, or has to fall, on 5. The circumstance that certain rules are applied by certain speakers with a certain probability is simply a descriptive fact, just as the fact that the probability to marry someone from the same town is, say, 0.31, or the probability to die before age 90 is, say, 0.84; here, too, no one calculates whether he or she is within the marriage quota or whether the time has come to kick the bucket.

3.3. Derivation weighing grammars

If the probability with which a rule is applied depends on whether some other rule has been applied before, then there must be a device to keep track of what is applied when in a derivation. This can be done by a *control word.* Each rule in the grammar is supposed to have a label, say $r_1,\ldots, r_{22}$ in our sample grammar. A control word is a string of such labels, in the order in which the rules are applied. There are two small complications. The first line in the derivation is not brought about by a rule; for the sake of generality, it is useful to assume an 'empty rule', called r_0, which is responsible for the first line and hence is the first symbol in the control word. Moreover, it must be determined in which order the symbols in a given string are replaced, because otherwise, several control words could lead to the same result. A simple way to achieve this is to adopt the principle 'from left to right', as we have tacitly done in derivation (6). The control word of (6) is then: $r_0 r_3 r_4 r_6 r_7 r_{15} r_{16} r_9 r_{19} r_{13} r_{23}$. The set of all control words of a given grammar is called its *control language,* and probabilistic restrictions (and, incidentally, other restrictions, as well) can now be defined for this control language, rather than for individual rules.

Such an approach provides the linguist with a very mighty instrument which makes it possible to account for all sorts of dependencies between rules. For most practical

purposes, however, it rapidly gets to complicated; therefore, we do not go into details here (see Klein/Dittmar 1979, 48-56). A reasonable compromise is to look only at effects 'neighbouring rule effects', for example at the influence of a rule in the immediately preceding step in a derivation. This is done by considering all 'digrams', i.e., all strings such as r_1r_1, r_1r_2, r_1r_3 etc, that can occur in the control language. Many such digrams are automatically excluded; in our sample grammar, r_1r_2 and r_1r_2 are possible, whereas r_2r_2 is not. Impossible digrams get the value zero; the probabilities of possible digrams must be empirically determined by corpus analysis. Whether such an approach is useful for the study of variability or not, is an empirical issue.
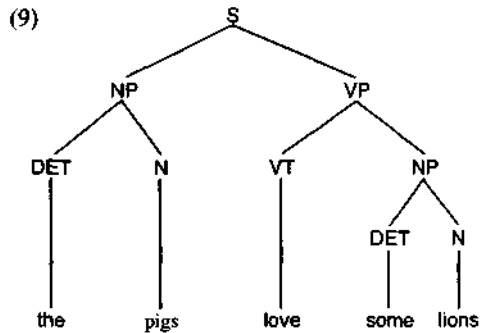
3.4. Context sensitive grammars

In context free grammars, a symbol A in some string can be replaced by x, if there is a rule A -> x; it does not matter what precedes or follows this symbol A. This is different for context sensitive grammars, whose rules have the general form u A v -> u x v, where, as before, A is a symbol from the nonterminal vocabulary and x is a nonempty string of symbols; u and v are strings of symbols, too; but they can be empty (alternatively, the notation A -> x/u_v is used for A is replaced by x in context u_v'). It is easy to see that context free grammars are simply a special case of context sensitive grammars, namely those in which u and v are always empty.

Context sensitive grammars can be probabilistically weighted in much the same ways as described in section 3, the only difference being that not all rules which expand the same symbol form the sample space but *all rules which expand the same symbol in the same context*. Otherwise, everything remains the same; in particular, they can also be used as an overall grammar in a variety grammar. An advantage of context sensitive rules is the fact that they in some cases, even Suppes type weighings can account for rule dependencies. Consider, for example, the potential asymmetry between subject-NP and object-NP discussed in section 3.2. Since subject-NPs only appear in context _VP, whereas object-NPs only appear in context VT_, it is easily possible to assign different probabilities to their expansion in these positions. This cannot be generalized, however; recursive rules such as r_1, for example, cannot be treated in this way, although their applicati-

on probability normally diminishes with each application in some derivation.

4. Probabilistic weighing II: Transformational grammars

Phrase structure grammars generate structures - trees whose end nodes, the 'leaves' of the tree, correspond to a sentence of the language to be described. Transformational rules turn such a tree into another tree. We shall here not review here the linguistic merits of these rules but only discuss how they can be integrated into a variety grammar.
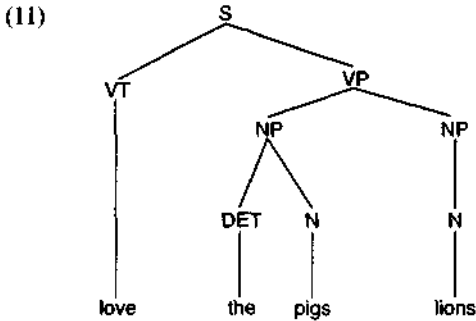

A transformational rule has two components. The first of these, the 'structural description' (SD) specifies to which trees the rule applies; the second component, the 'structural change' (SC) specifies how a tree which matches the SD is modified. We may think of the SD as a sort of cut through the nodes of a tree. The sample grammar (5) can generate, for instance, the tree shown in (9). This tree matches the SD,: NP - VT - DET - N, as indicated by the broken line, but it does not match the SD₂: NP - VT - DET - ADJ - N:

(9) A SD may be more complex; for example, it may indicate that a 'subtree' is identical to another 'subtree', or it may state that the tree must contain a specific nonterminal element; these refinements are not relevant to our present concern, however; all that matters is that the SB precisely indicates what the transformation applies to.

The SC indicates what happens with the various subtrees thus cut off. In (9), these are the subtrees dominated by the first NP, by VT, by the second DET and by the second N. They may be moved to a different position, replaced by another subtree or even omitted. A full transformational rule could then look as follows:

(10) SD: NP – VT – DET – N
      SC: 1   2   3   4 → 2 1 0 4

This means: the first subtree is moved into the position of the second subtree, the second subtree is moved into the position of the first subtree, the third subtree is omitted (i.e., replaced by the empty subtree 0), and the fourth subtree is left as it is. When applied to (9), the result is as follows:

(11)

(11) Clearly, rule (10) is not a top candidate for a meaningful transformation; in reality, all of these rules must be linguistically well-motivated; but here, we are only interested in the principle. A precise formalisation of transformational rules is extremely difficult, and we will skip it here. It is usually assumed, that transformational rules can be optional or obligatory; in the latter case, it must be applied whenever a given tree fits the SD.

The probabilistic weighing of transformational rules is straightforward; we only sketch the basic ideas (for details, see Klein and Dittmar 1979, 57-63). Much in parallel to the 'rule blocks' of a phrase structure grammar, *all rules with the same SD are taken together as a 'T-block'.* Each such T-block is considered as a sample space. The entire probability of 1 associated with such a T-block is then distributed over the various possible SC, which a tree falling under the SD may undergo. This leads into a problem when there is only one SC for a given SD. Then, this SC automatically gets the probability value 1. This means that the rule is obligatory, although this may not be desirable for linguistic reasons. This problem can be avoided if each T-block also includes, as one possible SC, an 'identical transformation', i.e., a transformation which does not change anything. The alternative is then 'change' or 'don't change', and this alternative can then be probabilistically weighed as any other alternative.

5. Variety grammars in empirical research

Probabilistic weighing of grammars has not been invented by linguists but by mathematicians who were interested in certain formal properties of these grammars (Habel 1979 gives an excellent account of this background). Their first and still very elementary application to problems of linguistic variability is due to Suppes (1972). The notion of 'variety grammar' with various types of weighing was elaborated in Klein (1974) and subsequently applied in a number of comprehensive empirical investigations, such as Klein/Dittmar 1979; Senft 1982; Tropf 1983 or Carroll 1984). There are no comparable studies of variability afterwards. But more recently, probabilistic grammars started to play an important role in other fields of linguistics, in particular in automatic parsing (see, e.g., Manning/Schütze 1999).

Variety grammars are a very precise and flexible instrument which allow the modelling of virtually any kind of grammatical variability. Their application in empirical research faces the linguist with a number of tasks that have to be solved in a certain order:

(1) *Choice of variety space.* It must be decided which kind of extra linguistically defined variation is to be covered. This decision is not arbitrary; it depends on a number of heuristic considerations on what the relevant factors of variation are.

(2) *Selection of linguistic structures.* The linguistic variables to be investigated must be selected. Normally, it is not meaningful to consider full grammars; this would be by far too complex. Instead, specific structural phenomena in which the relevant varieties differ must be selected - phonological change, variation in inflectional morphology, differences in word order, in the structure of noun phrases, whatever. Again, the precise way in which these 'dependent variables' are defined is based on heuristic considerations on the one hand and the specific interests of the researcher, on the other.

(3) *Data collection.* The next crucial step is to find or to collect appropriate corpora which are representative for the varieties and the linguistic variables under investigation.

(4) *Corpus analysis.* The relative frequency of the relevant linguistic structures in these corpora must be determined.

(5) *Definition of overall grammar.* The fragment of a grammar which covers all rules represented in the various corpora must be worked out. This concerns the choice of an appropriate type of grammar, the formulation of the rules and they way in which they are combined into rule blocks.

(6) *Probabilistic weighing.* This weighing depends on the type of grammar, the type of weighing and, of course, the relative frequencies found in corpora.

Just as with any other way to account for linguistic variability, these tasks raise a number of empirical as well as theoretical problems, which we will not discuss here. There are three main advantages of this particular approach: (a) It can be used for practically any grammatical phenomenon, and it is not restricted to a particular type of grammar - provided is can be precisely defined.

(b) It allows a precise modelling of arbitrary many extralinguistic factors of variation, such as development of linguistic knowledge in the individual, in a speech community, dialectal differences, in register, communicative intention and so on - again provided these factors can be given a precise definition.

(c) It can model continuous as well as categorial differences between varieties.

In all three respects, its application is only delimited by the data available. On the other hand, it should be clear that variety grammar is not a theory of variation or even of the human language faculty. It is a clear and precise descriptive instrument, no more, no less.

6.   Literature (selected)

Carroll, Mary (1984) *Cyclic Learning Processes in Second Language Production,* Frankfurt.

Habel, Christopher (1979) *Aspekte bewertender Grammatiken,* Berlin.

Klein, Wolfgang (1974) *Variation in der Sprache,* Kronberg.

Klein, Wolfgang/Dittmar, Norbert (1979) *Developing Grammars,* Berlin/Heidelberg/New York.

Manning, Christopher/Schütze, Hinrich (1999) *Statistical Natural Language Processing,* Cambridge.

Milton, J. Susan/Arnold, Jesse (1990) *Introduction to Probability and Statistics,* New York.

Senft, Gunter (1984) *Sprachliche Varietät und Variation im Sprachverhalten Kaiserslauterer Metallarbeiter,* Bern/Frankfurt a.M.

Stegmüller, Wolfgang (1973) *Statistische und personelle Wahrscheinlichkeit,* Berlin/Heidelberg/New York.

Suppes, Patrick (1972) "Probabilistic grammars for natural language", in: *Semantics of Natural Language,* Davidson, D./Gilbert, H., eds., Dordrecht, 741-762.

Tropf, Herbert (1983) *Variation in der Phonologie des ungesteuerten Spracherwerbs,* Phil. Diss., Heidelberg.

Wolfgang Klein, Nijmegen
(The   Netherlands)