On three types of dialect variation and their implications for linguistic theory. Evidence from verb clusters in Swiss German dialects

Guido Seiler

Abstract

Dialectology and typology are both concerned with the empirical investigation of the variability of grammars. In the present paper I will argue that three basic types of dialect variation must be distinguished. The distinction between diatopic, conditioned and free variation reveals many more different systematic arrangements of structural options than a purely non-geographical enumeration of phenomena would. Therefore, the distinction proposed here allows us to make more precise statements concerning the typologically relevant issue as to which grammatical features can combine, and how they combine. The study is based on evidence from Swiss German verb clusters and their geographical distribution. I will show that transitions from one to another pattern are gradual, yet follow a certain systematicity. From these facts, specific requirements for an architecture of grammar emerge: a grammar must allow for variable outputs and preference directions. I will propose a way how these requirements can be satisfied within an Optimality theoretic setting.

1. Introductory remarks: Dialectology and typology

At first glance, typology and dialectology seem to be two linguistic subdisciplines between which not much interaction is possible.[1] Traditionally, typological work has indeed not influenced dialectological work much, and vice versa. In recent times, however, the picture has changed. The contribution that dialects and dialect variation may make to

linguistic typology is now increasingly being acknowledged (Kortmann 2002). To begin with, I would like to address the question in what respects dialectology and typology are similar and in what respects they differ, in order to determine the conceptual basis for a closer interaction between the two disciplines. Dialectology and typology have in common that they are both concerned with the empirical study of the variability of grammars. The leading questions of dialectology and typology can be defined as follows:

(1) Leading questions of dialectology:
– What structural properties distinguish one dialect from another?
– How is a structural property spread across space?
– What structural properties co-occur in the same areas?

(2) Leading questions of linguistic typology:
– What structural properties are common to all languages?
– What structural properties are variable between languages?
– In which ways do structural properties depend on each other?

The points in (2) can be traced back to one basic question:

(2a) What are possible grammars? What grammars are attested in the world's languages?

The benefit of dialectology for linguistic typology is obvious in at least two areas: first, dialects simply provide more systems of grammar which have not been taken into account so far in a satisfying way – and therefore they add to our knowledge of linguistic structure. Second, dialects are to a large degree free from standardization, a fact that is crucial for the typological picture of Europe, which has mainly been based on standardized languages. However, I want to draw attention to an additional argument based on the specific potential of dialectology as a geographical discipline.

A specific feature of dialectology is full coverage of the areal continuum. Given this empirical base, we might expect that the minimal units ('atoms') of contrasts between grammars can be isolated. If we take seriously the finest gradations found across the areal continuum, a much larger number of attested grammars can be uncovered – a desired result for the empirical foundation of the investigation of cross-linguistic variability. Therefore, the type of evidence provided by the study of dialect variation is somewhat complementary to that provided by wide-range cross-linguistic

comparison. However, both types of evidence enrich our knowledge of attested grammars and call for an account within any theory allowing and restricting the variability of grammars. The study of dialect variation cannot and need not answer all relevant questions of linguistic typology (and vice versa). Instead, both dialectology and typology play their part in a more general enterprise: the investigation of the full range of cross-linguistic variation.

In the present paper, I intend to give an idea of the types of variation we can expect across closely related dialects, and what their relevance is for linguistic typology and theories of grammar. The study is based on evidence from verb clusters in the dialects of German-speaking Switzerland. My central claim is that an in-depth analysis of the geographical distribution of patterns uncovers a multiplicity of area-specific systems of grammar, each of which is acquired by the local speakers, and therefore must be treated as an attested and thus possible grammar.

The paper is organized as follows. I will first illustrate how far-reaching and subtle cross-linguistic variation can be, using as an example word order patterns (section 2). In section 3, I will focus on the ordering of elements within verb clusters of Alemannic (Swiss German), and present the evidence provided by full coverage of the geographical area. In section 4, I will propose an empirical distinction between three basic types of dialect variation, and I will show that this distinction is crucial for the introductory question of what are attested and thus possible grammars. From the geographical facts and their consequences specific requirements for a theory of grammar can be derived. In section 5, I will propose a way how these requirements can be satisfied, focusing on Optimality Theory as a promising formal device for the modeling of the different types of variation found in section 3.

2. The limitations of microscopic variation

The facts and issues in the field of constituent order are well known from the typological literature (Greenberg 1966; Lehmann 1973; Vennemann 1973; Hawkins 1983; Dryer 1992; see also the surveys in Comrie 1989, ch. 4; Whaley 1997, ch. 5–6; Croft 2003, ch. 3.2–3.4). Put in most general terms, head–modifier or modifier–head patterns are related to each other across different syntactic categories according to correlations, implicational

universals, dominance, and harmony. These relations define the possible feature combinations in a particular language and thus the possible positions a particular language can take on a scale from maximally head-initial to maximally head-final languages. Croft lists the following serialization parameters in his very recent survey (2003: 72, based on Lehmann 1973; Vennemann 1973):

(3) Serialization parameters (Croft 2003: 72):

Head-final:	Head-initial:
object–verb	verb–object
subject–verb	verb–subject
verb–auxiliary	auxiliary–verb
adverb–verb	verb–adverb
verb–subordinator	subordinator–verb
purposive–verb	verb–purposive
sentence–qualifier	qualifier–sentence
noun–postposition	preposition–noun
genitive–noun	noun–genitive
relative clause–noun	noun–relative clause
adjective–noun	noun–adjective
demonstrative–noun	noun–demonstrative
numeral–noun	noun–numeral
adverb–adjective	adjective–adverb

In what sense, then, can cross-dialectal variation supplement our knowledge of possible variation in head/modifier patterns? Within Alemannic, only one of the parameters above is variable, the order of verb and auxiliary in subordinate clauses. This does not seem to be very much. However, it is part of a broader domain concerning the serialization of heads and modifiers, namely the order of elements in verb clusters. Consider the following sample of verb clusters collected at two locations in Switzerland: Guttannen in the West (Bernese Highlands) and Thusis in the East (Grisons district):

(4) Guttannen (Bernese Highlands):

a.	*…ob*	*si*	*das*	*Auto*	*schon*	***hed***	***zahld***
	whether	she	the	car	already	has	paid

'(I have no idea) whether she has already paid for the car'

b. *…ob* *är* *äis* ***wil*** ***hiraten***
whether he ever wants get_married
'(Well, I don't know) whether he ever wants to get married'

c. *I* *han* *erscht* *mit* *vierzg* ***glehrt*** ***fahren***
I have only with 40 learned drive
'I learned to drive only when I was 40'

(5) Thusis (Grisons district):

a. *…ob* *sie* *das* *Auto* *scho* ***zalt*** ***het***
whether she the car already paid has

b. *…ob* *är* *amal* ***hürata*** ***wetti***
whether he ever get_married wants

c. *I* *han* *erscht* *mit* *viarzig* ***faara*** ***glärnt***
I have only with 40 drive learned

(4&5)

	Guttannen: 1–2	Thusis: 2–1
a.	AUX–V	V–AUX
b.	MOD–V	V–MOD
c.	V1–V2	V2–V1

Our first impression is that there is a distinction between strictly head-initial ordering of elements in the West and strictly head-final ordering of elements in the East. However, the picture is much more complex, due to three factors. First, the three head categories auxiliary, modal verb and lexical verb behave differently in many areas between Guttannen and Thusis. In Zurich German, for instance, the most widespread pattern is head-final with auxiliaries (*zalt hät* 'paid has') but head-initial with modal verbs (*wett hürate* 'wants get_married') and lexical verbs (*gleert faare* 'learned drive'). Second, two-verb clusters and three-verb clusters behave differently. Three-verb clusters are never head-final even in the Eastern dialects. Here, too, *ha wele gaa* 'have wanted go' (1–2–3) is common whereas **gaa wele ha* 'go wanted have' (3–2–1) does not occur. Third, there is not only variation between speakers or dialects, but also intrapersonal variation, i.e., very frequently individual speakers accept several variants (1–2 or 2–1) side by side.

Therefore, a simple division between dialects with head-initial and head-final verb clusters does not hold, since many intermediate feature combinations occur. More generally speaking: on the one hand, the amount of variation we can expect across closely related dialects is rather restricted

with regard to the cross-linguistic potential of variation defined by the parameters listed in (3); in our example, only one parameter shows variation (V–AUX / AUX–V). On the other hand, within this restricted field of variation, we are able to contribute something to the precise formulation of the finest graduations with regard to that parameter. Although the variation found in our dialects affects only a small subsection of a scale from maximally head-initial to maximally head-final languages, it uncovers what the minimal distances are between systems of grammar – which is, in my view, a substantial contribution to our understanding of the *full* range of cross-linguistic variation, too.

3. The geography of Swiss German verb clusters

3.1. The experiment

Swiss German verb clusters have attracted some attention by syntacticians (e.g. Lötscher 1978; Wurmbrand, forthcoming; Zwart 1996). However, our information has been only haphazard so far. Now, we have completely new evidence covering the entire area. The data and maps I am presenting here are the outcome of a research project on Swiss German dialect syntax at the University of Zurich.[2] The goal of the research project is to explore the geographical distribution of syntactic construction types over German-speaking Switzerland. 2700 informants at 350 measuring points were sent four written questionnaires (7.7 informants per measuring point in average). The questionnaires contain translation tests, fill-in tests and multiple choice tests. The sentences in question are embedded into a situational context (see Bucheli Berger and Glaser 2002 for further details concerning the research design). All questions on verb clusters are multiple choice tests, since we intended to avoid a suggestive effect by the Standard German pattern given in translation tests. In multiple choice tests we asked for acceptable variants as well as for the preferred variant. The relevant example sentences for AUX/V, MOD/V and V1/V2 have been provided in (4a-c) and (5a-c) above. For AUX/MOD/V, the example sentences are provided in (6) an (7):

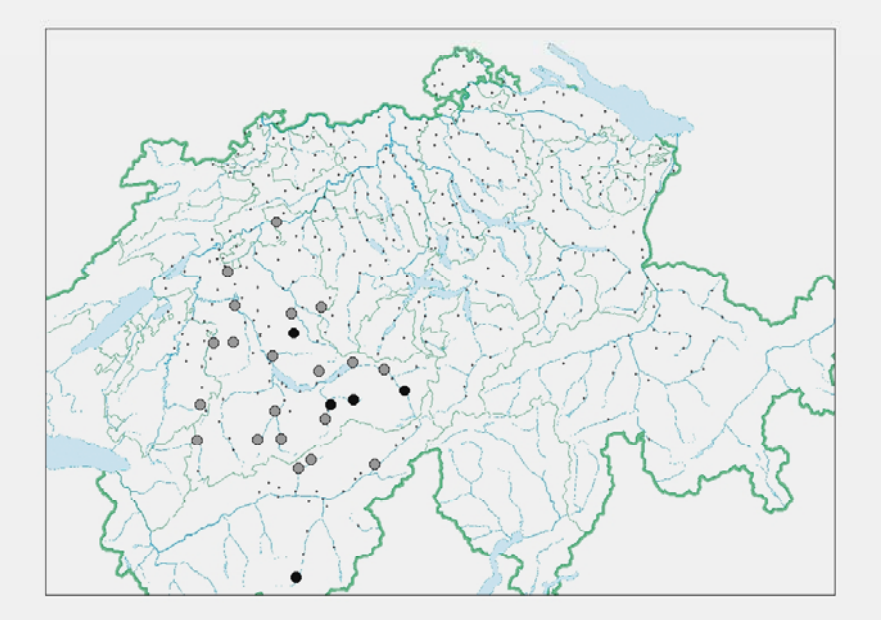
(6) a. *S Telefon hät grad glüütet, woni* ***han welle gaa***
the phone has just rung when=I have wanted go
'The phone just started to ring when I wanted to leave.'
b. *S Telefon hät grad glüütet, woni* ***han gaa welle***
c. *S Telefon hät grad glüütet, woni* ***welle han gaa***

(7) a. *Er isch grad choo, woni* ***ässe han welle***
he is just arrived when=I eat have wanted
'He just arrived when I wanted to eat.'
b. *Er isch grad choo, woni* ***ässe welle han***
c. *Er isch grad choo, woni* ***welle ässe han***

The six combinations which are logically possible in three-verb clusters are split into two questions in order not to confuse the informants. Since only a subset of the logically possible combinations was presented to the informants in each of the questions (6) and (7), a remarkable suggestive effect resulted: for instance, 671 informants accepted *ässe han wele* 'eat have wanted' if it was suggested, but only 88 informants independently wrote down the structural equivalent of that pattern in question (6), *gaa han wele* 'go have wanted'. I decided to restrict my analysis to the cases where this variant has been given independently, since there is no doubt that this is clear positive evidence that this pattern really exists for the respective speakers.

3.2. The dimension of the head category

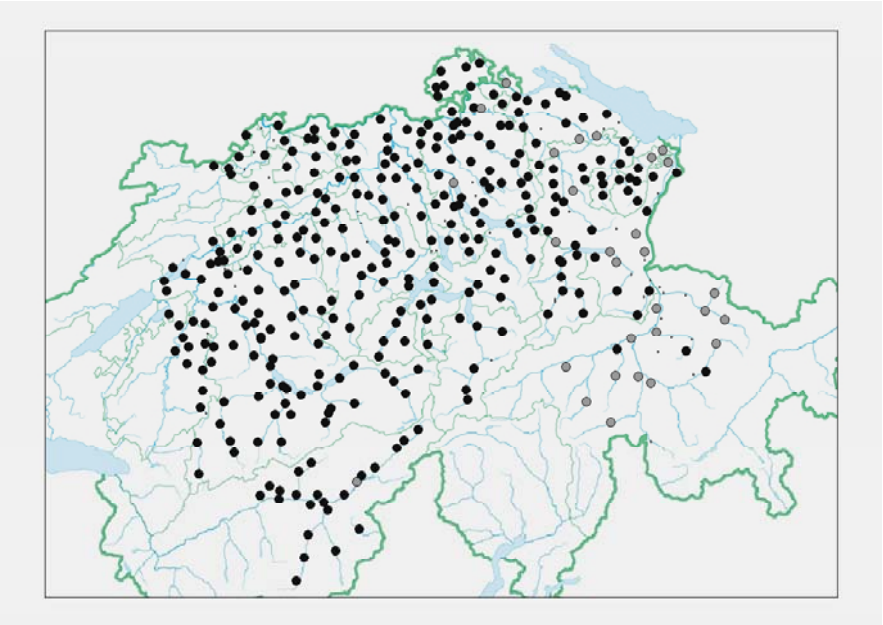
As I already pointed out, the general impression is a contrast between head-initial dialects in the West and head-final dialects in the East. The sample examples (4–5) from Guttannen (Western; Berne district) and Thusis (Eastern, Grisons district) suggest that two-verb clusters pattern alike across different head categories (auxiliary, modal verb, lexical verb). However, this holds true only for the western- and easternmost dialects. In fact, many more combinations are found. Let us now have a more detailed look at the geographical facts.



Map 1. Black dots: AUX–V preferred by ≥ 50%; grey dots: AUX–V preferred by ≤ 49%

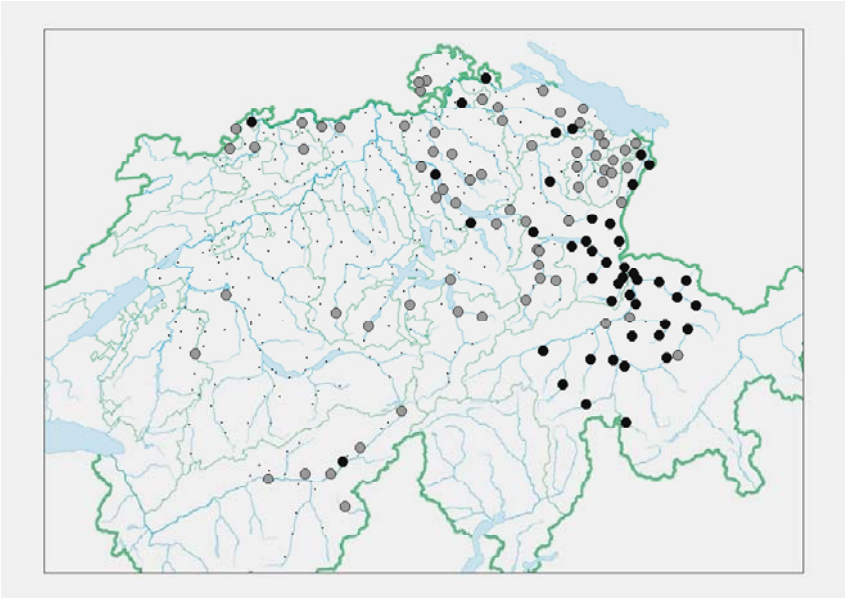
The geographical extent of AUX–V is illustrated in map 1. Each dot represents a measuring point where at least two informants[3] prefer the ascending variant *hät zalt* 'paid has'. If these persons are a majority of the informants at the location, the dot is black, if they are not, it is grey. As expected, ascending ordering of elements is limited to the Western part of our area. However, the picture changes dramatically when we take into account two-verb clusters with modal verbs.

Map 2 shows the extent of preference for ascending ordering of MOD–V. Surprisingly, ascending ordering with modal verbs is much more widely accepted than with auxiliaries, since MOD–V covers almost the entire area. Note, however, that in the easternmost part the acceptance for MOD–V significantly decreases.

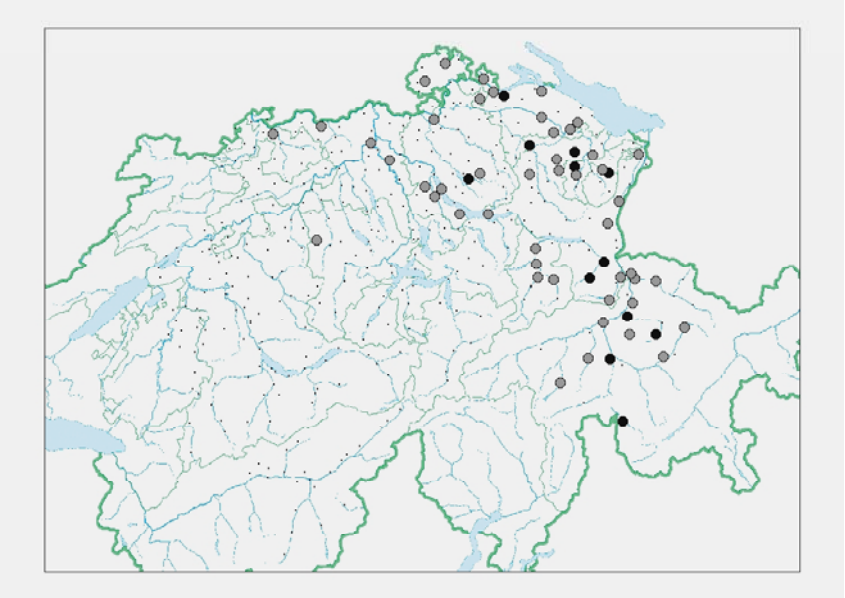


Map 2. Black dots: MOD–V preferred by ≥ 50%; grey dots: MOD–V preferred by ≤ 49%

Map 3 shows the distribution of preference for V–MOD, thus, it is the mirror image of map 2. Even more clearly, it illustrates that descending ordering of verb and modal verb is preferred by more and more speakers the further we move eastwards. The same holds true for clusters of two lexical verbs, as we see on map 4. Head-final ordering in clusters of two lexical verbs is even more restricted than head-final clusters of V–MOD, both geographically (fewer affected measuring points) and quantitatively: in total, only 246 persons prefer V_2–V_1, whereas 540 persons prefer V–MOD.



Map 3. Black dots: V–MOD preferred by ≥ 50%; grey dots: V–MOD preferred by ≤ 49%

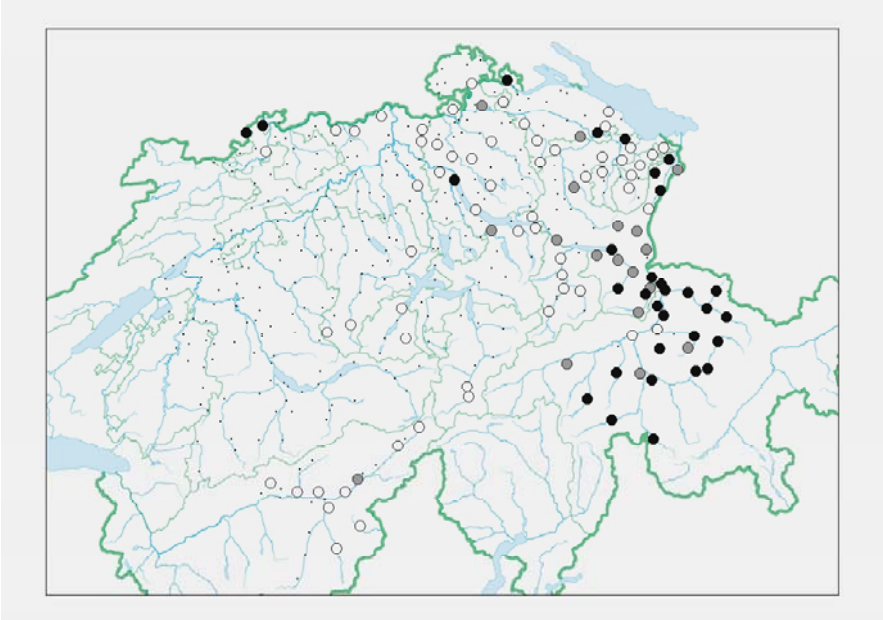


Map 4. Black dots: V_2–V_1 preferred by ≥ 50%; grey dots: V_2–V_1 preferred by ≤ 49%

Let us sum up the observations so far. On the one hand, the geographical extents of head-initial (or head-final, respectively) two-verb clusters do not coincide across the different head categories. Ascending ordering of elements is most widespread with auxiliaries, and it is most restricted with lexical verbs. On the other hand, the geographical distribution of patterns accordingly shows that the acceptance of head-initial patterns decreases from West to East.

3.3. The preferential dimension

In what follows, I will pick out one of the variants discussed above, V–MOD, and show that the preferential dimension from accepted via preferred to obligatory fits the general pattern.



Map 5. Black dots: V–MOD obligatory for ≥ 50%; grey dots: V–MOD preferred by ≥ 50%; white dots: V–MOD accepted by ≥ 50%

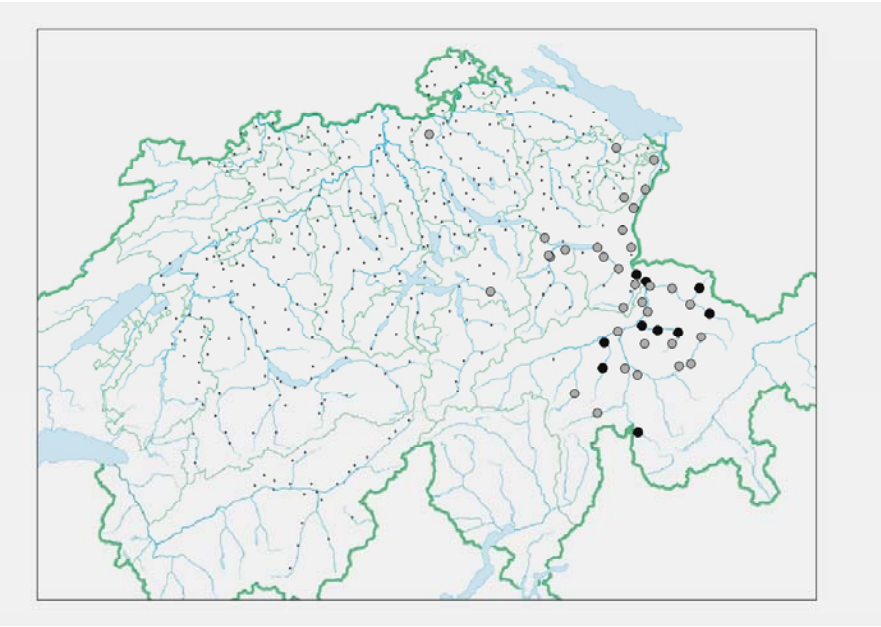
Map 5 gives the relevant data. The map is based on the results from question (6) above. The cartographic representation of the results aligns darker symbols with greater prominence of V–MOD on the preferential dimension. The symbolization focuses on the answers given by the

majority of informants at each measuring point, minority answers are not taken into account. The white dots represent measuring points where 50% or more of the informants accept the head-final variant *ob er emal* ***hürate wett*** 'whether he ever get_married wants'. The grey symbols express that at these locations the head-final variant is not only accepted but also preferred by 50% or more of the informants. At the points with black dots, for 50% or more of the informants this variant is even obligatory, i.e., it is the only possible variant. Note that obligatoriness is a subset of preference which is a subset of acceptance. Thus, every black symbol is – 'underlyingly' – grey and white as well, but not vice versa.

Although the boundaries between different degrees of prominence are not always clear-cut – partly due to topographical reasons – the overall picture is striking: V–MOD is for the most part not acceptable in the Western areas, with the exception of the city of Basel in the Northwest and the Valais district in the Southwest. Moving from West to East, we see that the density of locations accepting V–MOD increases. Similarly, more and more locations prefer or even obligatorily choose V–MOD, such that in the easternmost parts of the area most measuring points show black symbols. These results nicely mirror the general pattern we have already discussed above: the prominence of head-final ordering of elements in verb clusters increases from West to East. Hence, the gradual geographical transition is reflected on two dimensions: across different head categories as well as on the preferential dimension.

3.4. Three-verb clusters

Ascending ordering of elements in three-verb clusters (1–2–3) occurs in almost all parts of our area, i.e., even in most of the Eastern dialects where two-verb clusters commonly are head-final. Only in the easternmost dialects the most embedded verb is fronted: *gaa ha wele* 'go have wanted' (3–1–2); see map 6. Note that the 3–1–2 variant has not been suggested in question (6). The informants independently gave us 3–1–2 as an additional variant.



Map 6. Black dots: V–AUX–MOD given by ≥ 50%; grey dots: V–AUX–MOD given by ≤ 49%. Informants independently gave the variant *gaa ha wele* 'go have wanted' (V–AUX–MOD) which was not suggested in question (6).

3.5. How many grammars?

Let us now summarize which combinations of serialization patterns in verb clusters we have detected. First, I have shown that the ordering of elements in Western dialects is strictly ascending (1–2–3), but the more we move eastwards the more the tendency for ascending ordering weakens. Second, the ordering of elements is sensitive to the category of the head. Auxiliaries tend to be placed at the right edge of the cluster. This tendency is much weaker with modal verbs and almost absent with lexical verbs as heads of a cluster. The cross-cutting of these two dimensions defines the contrasts between grammars I–IV in table 1 below. Furthermore, only in the easternmost dialects do we find a 3–1–2 pattern, whereas 1–2–3 is the common pattern for the rest of the Eastern area, which defines the contrast between grammars IV and V:

Table 1. Five grammars from West to East (disregarding quantities and preferences). Shaded cells represent ascending patterns.

	West ◄				► East
	I	II	III	IV	V
AUX,V	1–2	2–1	2–1	2–1	2–1
MOD,V	1–2	1–2	2–1	2–1	2–1
V_1,V_2	1–2	1–2	1–2	2–1	2–1
AUX,MOD,V	1–2–3	1–2–3	1–2–3	1–2–3	3–1–2

4. Three types of dialect variation

I have shown that the transition from one to another option of a geographical variable is *gradual*. Suppose a scenario where the prominence of a pattern (e.g., ascending order in verb clusters, but the general picture holds true for other cases as well) gradually decreases in the areal continuum from centre to periphery. Graduality, then, is reflected in four respects. First, the prominence of the pattern decreases with regard to the relative density of affected measuring points. Second, the percentage of speakers at one measuring point using that pattern is higher towards the centre and lower towards the periphery. Third, the pattern in question can be obligatory in the centre, but the farther we move away from there, it becomes optional. If two options co-exist, it can be the case that the pattern in question is preferred in or towards the centre, and it is dispreferred in or towards the periphery. Fourth, the pattern is subject to more systematic or contextual restrictions in the periphery than in the centre. Remember for instance that ascending order in two-verb clusters is possible (or even required) with all categories of heads in the West, but the further we move eastwards, the more is ascending order restricted, first to clusters with modal verbs and lexical verbs as their heads, then to clusters with lexical verbs only, and, finally, ascending order is excluded (or dispreferred) in the easternmost dialects.

What can we conclude from these facts? A simple division between the areas where a pattern occurs and those where it does not is unsatisfactory. Instead, we need a much more sophisticated way of describing the geographical distribution of grammatical variables. Areas can be distinguished not only according to the existence or non-existence of a pattern, but also according to the relations of a pattern to its competitors. In

the first place, it has to be determined whether a pattern is obligatory or not in a given dialect. It is obligatory if no other realizations of the feature in question occur in the same dialect. For instance, 1–2 order in two-verb clusters is obligatory if 2–1 is excluded in that particular dialect. If, however, two options co-occur in one and the same dialect (e.g., both 1–2 and 2–1), we have to find an answer to each of the following questions:

(i) Do the options x and y co-exist in one (or more) individual's competence (intra-individual variation), or is only one of the options, x or y, possible for an individual (inter-individual variation)?

(ii) What is the preference direction between two options? Even if a speaker allows x and y, usually the two options are not equally preferred. It is normally the case that speakers prefer one of the variants, although the other variant is grammatical, too.[4]

(iii) Are there systematic (or contextual) factors guiding the variation between x and y? The fact that two options x and y are co-present in one speaker's grammar does not necessarily mean that the choice between x and y is random. Very frequently, one option is more preferred [??] or even required under certain systematic circumstances. In our example, 1–2 and 2–1 orders of two-verb clusters are both present in the majority of Swiss German dialects. However, there is a functional arrangement between them insofar as 1–2 is most expected with a lexical verb as the head and least expected with an auxiliary.

Having determined the ways how two (or more) options of a geographical variable can combine across different dialects, the consequences for our notion of 'dialect variation' are remarkable. It is necessary to distinguish between (at least) three major types of dialect variation. The first distinction is to be made between *diatopic variation* and types of grammar-internal variation. In a very general sense, diatopic variation simply means all kinds of geographical contrasts between grammars. The classical case is that two options x and y of a geographical variable are in complementary geographical distribution. However, I propose that geographical contrasts with regard to the status of an option within the respective system of grammar must be included in the definition of diatopic variation, too. Thus, if e.g. two areas A and B can be distinguished, A with obligatory x, and B with x in competition with y and used only under certain conditions, there is a diatopic contrast between the grammars A and B, too. Similarly, A and B can be different solely in their preference for one or another option, but not in their inventories of devices.

Note that in diatopic variation grammars can be involved within which no internal variation occurs. 'Variation' is discovered only through the comparison of grammars. The distinction between diatopic variation – referring to contrasts between two or more grammars – and all other types of variation found within one system of grammar is absolutely crucial. It is the latter I shall address in the following. From the mutual relationship of the two options x and y co-present in one grammar, two basic types of internal variation can be derived. Where two options coexist, their distribution is either totally random, or it underlies certain (syntactic, semantic, pragmatic) conditions. In *free variation*, two or more options coexist and are randomly distributed. Thus, we are dealing with grammars containing true optionality. I am convinced that completely free variation exists, though it is perhaps not very common. For this infrequency two factors are responsible: first, it is often the case that distributional asymmetries between two options can be isolated (see below). Second, there can be a preferential asymmetry between the variants.[5] Although in this case the variation is biased, I propose to include such cases in the notion of free variation as long as no functional or contextual factor can be isolated constraining the variation.

In our investigation, informants were asked for the preferred ('most natural') variant when they were offered a set of options in multiple choice questions. In most cases, the informants did not have any problems in selecting one of the variants as the preferred one. Some informants, however, sporadically gave us two preferred variants. Of course, the design of the questionnaire is somewhat suggestive in asking for the 'most natural variant' (singular). On the other hand, if informants really wanted to indicate two variants as preferred, they had the opportunity to do so. For instance, 36 informants explicitly indicated that both *wett hürate* 'wants get_married' and *hürate wett* 'get_married' are equally natural for them, despite the fact that our questionnaire forced the informants to a decision. I interpret this behaviour as evidence for the existence of completely free variation.

If two options coexist in one system of grammar and show distributional asymmetries according to functional or contextual principles we are dealing with what I will call *conditioned variation*. The distribution of 1–2 and 2–1 in two-verb clusters is a case of conditioned variation for most speakers of Zurich German: both patterns are present in principle, but clusters with a modal verb as their head are 1–2 whereas clusters with an auxiliary are 2–1. A simple statement like 'there is variation in Zurich German verb clusters'

is misleading, since the variation is not random but conditioned by one specific factor, namely the category of the head. Conditioning can be 'hard' or 'soft'. It is 'hard' if no optionality is involved. For a total of 1542 informants of our investigation MOD–V is obligatory (*V–MOD) and V–AUX is obligatory (*AUX–V). However, for 309 informants the category of the head defines a 'soft' preference direction: with modal verbs (but not with auxiliaries), both orderings are acceptable, but they prefer MOD–V over V–MOD. Thus, whereas optionality is always involved in free variation, this is not necessarily the case in conditioned variation.

The consequences of the distinctions proposed here are crucial for the introductory question of what are attested and thus possible grammars. A mere enumeration of grammatical features found in a larger area without determining their mutual relationships defined by their specific geographical extents is insufficient. Having accepted this, a number of consequences follow:

(i) Since grammatical patterns combine in the areal continuum, the number of grammars (= specific combinations of features) is higher than the number of geographically variable features. If there are two options of a geographically variable feature, e.g. 1–2 and 2–1, the number of grammars is higher than only two. At least we expect three grammars: one with 1–2, one with 2–1, and one where 1–2 and 2–1 coexist. Furthermore, if several types of conditioned or free variation between 1–2 and 2–1 can be isolated, the number of attested grammars increases even further.

(ii) The slightest differences between grammars arise as contrasts in preference for an option and/or constraints on the use of an option. Thus, the inventory of grammatical devices can be identical in two hypothetical adjacent dialects A and B, i.e., there are no grammaticality contrasts with respect to a given set of devices. However, grammar A can be different from grammar B in preferring x over y generally, or under certain systematic (functional, contextual) circumstances.

(iii) Optionality and preference directions are not due to 'performance errors'. Instead, grammars with optionality and preference directions occupy a specific position in the areal continuum. We have seen that the preferential dimension from obligatory via optional-preferred and optional-dispreferred to ungrammatical is reflected in space. If optionality and preference were merely accidental performance noise we would expect a random geographical distribution of the levels on the preferential dimension – which is, however, not the case.

(iv) A grammar with variable features is not a random mix of two (or more) 'consistent' neighbouring grammars (Seiler 2003: 153). It is tempting to trace back (free or conditioned) variation within one speaker's grammar to the co-presence of two parallel grammars both equally accessible to the speaker. Thus, the parallel grammars approach treats internal variation like diglossia (examples: Kroch 2001: 720, Lightfoot 1999: 92, Zwart 1996: 245). At first glance, a geographical picture with transition zones seems to support this approach. It is, however, false. First, diglossia and grammar-internal variation are sociolinguistically different facts, and we need a formal instrument in order to account for this difference. Second, we need a 'meta-grammar' telling us the right arrangement of the two co-present grammars. Third, if two co-present grammars are needed in order to account for one variable feature (i.e., one feature with two options), the number of parallel grammars exponentially increases with every variable feature – which is a highly undesired result (2^n grammars are needed for n variable features; see Bresnan and Deo 2001: 39). And fourth, most strikingly: the transition from one to another option follows a certain geometry. Take the transition from Western 1–2 to Eastern 2–1. Going from West to East, it is first AUX–V that switches to V–AUX, then MOD–V to V–MOD, then V_1–V_2 to V_2–V_1, and finally AUX–MOD–V to V–AUX–MOD. Every location occupies a specific position on the transition from 1–2 to 2–1. It is part of the internal grammatical competence of the speakers at all locations which verb clusters can be head-initial or head-final in their local dialect.

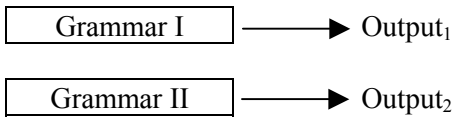
(vi) Consequently, at each point in the areal continuum a grammar exists that is acquired by the local speakers (Seiler 2003: 193–195). Of course, grammars of several points can be identical.

(vii) Therefore, we need a means of description that is powerful enough to capture maximally distinct grammars, but also precise enough to capture the minimal gradations between grammars. In particular, optionality and preferences must be explicitly accounted for within one grammar.

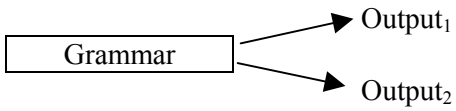
(viii) A promising instrument in order to satisfy these requirements is Optimality Theory (OT). In the next section, I will outline an OT analysis of the most common patterns found in Swiss German verb clusters, including all major types of diatopic, free and conditioned variation.

5. Variable output grammars and Optimality Theory

Within the generative tradition, internal variation and optionality have been traced back to the co-presence of two (or more) mutually incompatible grammars (Kroch 2001; Lightfoot 1991; Lightfoot 1999):



In the previous section, I have argued that the competing grammars approach is not satisfactory for empirical as well as for principled reasons. The alternative to competing grammars must be one grammar allowing for different outputs. I will call this type of grammar 'variable output grammar':



In particular, our model has to give an account for (i) consistently ascending ordering of elements in Western dialects, (ii) the conditioning of variation by the category of the head, (iii) optionality and (iv) preference directions.

What can be the internal architecture of a variable output grammar? Within the context of Optimality Theory (OT; Kager 1999; McCarthy 2002), several promising proposals have been made in order to derive variable outputs, such as floating constraints (Nagy and Reynolds 1997), tied constraints (Löhken 1997, ch. 2; Anttila 2002: 231), or continuous ranking (Boersma 1997; Bresnan and Deo 2001; Bresnan, Dingare and Manning 2001; Hayes 2000); for a survey of these and other types of variable output grammars see Paolillo 2002, ch. 10. In what follows, I will demonstrate how an analysis with tied constraints works, and I will close with a short remark on continuous ranking. I will focus on one locus of variation, namely the order of elements in two-verb clusters with a modal verb as their heads – i.e, the contrast between grammar II (MOD–V) and grammar III (V–MOD) from table 1 above.

I assume that different serialization patterns of verb clusters are output realizations of a more abstract feature structure which determines only the

head-complement relationships of the involved verb forms (AUX, MOD, V), but not their linear ordering. Therefore, a function is needed in our theory that maps the feature structure to subsequent positions. I formulate this function in terms of OT constraint ranking. The function assigns all possible output candidates (i.e., all logically possible orderings of verb clusters) a value 1 for 'grammatical' or 0 for 'ungrammatical'. In classical OT the value '1' can be assigned only once – namely to the 'winner' marked by '☞'. The winner is identified by a hierarchy of well-formedness constraints. Constraints are in competition and are violable in principle, but violations of higher ranked constraints are more serious. Schematically:

Table 2. Schema for output candidate evaluation in OT.

(Input)	Constraint 1	Constraint 2
Candidate A	*!	
☞ Candidate B		*

The constraints I am suggesting are strictly surface-oriented. First, I propose a constraint family *HEAD-RIGHT. *HEAD-RIGHT constraints guarantee left-headedness of a structure [head [complement]] and penalize [[complement] head]. However, since most languages exhibit variation in left- or right-headedness, a general constraint *HEAD-RIGHT is not sufficient. Instead, we need specifications of the domains where left-headedness is required. The relevant *HEAD-RIGHT constraints for present purposes are:

(8) *HEAD-RIGHT(aux): An auxiliary as a head precedes its (verbal) complement.

(9) *HEAD-RIGHT(mod): A modal verb as a head precedes its (verbal) complement.

(10) *HEAD-RIGHT(v): A lexical verb as a head precedes its (verbal) complement.

It is probable that for each of these constraints a mirror constraint exists (*HEAD-LEFT), competing with the *HEAD-RIGHT constraints. However, we do not need them for the present analysis. I propose that Eastern head-final structures in two-verb clusters (V–MOD, V–AUX) are not due to

constraints such as *HEAD-LEFT(mod), or *HEAD-LEFT(aux), respectively, but due to a constraint requiring that the predicate nucleus, i.e. the most embedded lexical verb, is positioned at the left edge of the cluster:

(11) V_{nuc}-LEFT: The most embedded lexical verb is positioned at the left edge of the verb cluster.

This constraint accounts for V–MOD, V–AUX of the grammars III, IV and V, *and* for V–AUX–MOD of grammar V in a unified way (cf. table 1 above). The reason for rejecting right-headedness constraints in Eastern dialects lies in the fact that only two-verb clusters consistently show descending order of elements. Surprisingly, in three-verb clusters of grammars III and IV ascending order appears. Obviously, three-verb clusters underlie specific conditions blocking descending order. However, they are irrelevant in two-verb clusters. The constraints *DISCONT and *BRANCH-LEFT account for these conditions:

(12) *DISCONT: Discontinuous constituents are not allowed.

In the easternmost dialects (grammar V), *DISCONT is outranked by V_{nuc}-LEFT which favours *gaa ha wele* 'go have wanted', containing discontinuous *gaa…wele* 'go…wanted'. In grammars III and IV, *DISCONT dominates over V_{nuc}-LEFT, such that *gaa ha wele* is penalized. However, *gaa wele ha* 'go wanted have' would satisfy both V_{nuc}-LEFT and *DISCONT, but it is ungrammatical in all our dialects. I propose that branching complements are more likely to follow their heads than to precede them:

(13) *BRANCH-LEFT: Branching complements must not precede their heads.

On the basis of the proposed constraint set, we are able to derive the patterns found in grammars I–V from table 1 above. Let us start with grammars II which is different from grammar III only with regard to MOD/V clusters:

(14) Grammar II: *ha wele gaa* 'have wanted go'
gleert faare 'learned drive'
wett hürate 'wants get_married'
zalt hät 'paid has'

Ranking: {*BRANCH-LEFT, *DISCONT, *HEAD-RIGHT(v), *HEAD-RIGHT(mod)} >> V_{nuc}-LEFT >> *HEAD-RIGHT(aux)

Table 3. Candidate evaluations for grammar II

	*BRANCH-L	*DISCONT	*HD-R (v)	*HD-R (mod)	V_{nuc}-L	*HD-R (aux)
☞ *ha wele gaa*					*	
ha gaa wele				*!	*	
wele ha gaa		*!			*	*
wele gaa ha					*	*!
gaa ha wele		*!		*		
gaa wele ha	*!			*		*
☞ *gleert faare*					*	
faare gleert			*!			
hät zalt					*!	
☞ *zalt hät*						*
☞ *wett hürate*					*	
hürate wett				*!		

The leftmost column lists the candidate sets. Constraints separated by a broken line are unranked, i.e., their mutual ranking cannot be determined on the basis of our data (for tied, i.e. crucially unranked constraints see below). Similarly, grammars I, III, IV and V can be expressed by reranking of the same constraints:

(15) Grammar I: *ha wele gaa* 'have wanted go'
gleert faare 'learned drive'
wett hürate 'wants get_married'
hät zalt 'has paid'

Ranking: {*BRANCH-LEFT, *DISCONT, *HEAD-RIGHT(v), *HEAD-RIGHT(mod), *HEAD-RIGHT(aux)} >> V_{nuc}-LEFT

(16) Grammar III: *ha wele gaa* 'have wanted go'
gleert faare 'learned drive'
hürate wett 'get_married wants'
zalt hät 'paid has'

Ranking: {*BRANCH-LEFT, *DISCONT, *HEAD-RIGHT(v)} >> V_{nuc}-LEFT >> {*HEAD-RIGHT(mod), *HEAD-RIGHT(aux)}

(17) Grammar IV: *ha wele gaa* 'go have wanted'
faare gleert 'learned drive'
hürate wett 'get_married wants'
zalt hät 'paid has'

Ranking: {*BRANCH-LEFT, *DISCONT} >> V_{nuc}-LEFT >> {*HEAD-RIGHT(v), *HEAD-RIGHT (mod), * HEAD-RIGHT (aux)}

(18) Grammar V: *gaa ha wele* 'go have wanted'
faare gleert 'learned drive'
hürate wett 'get_married wants'
zalt hät 'paid has'

Ranking: {*BRANCH-LEFT, V_{nuc}-LEFT} >> {*DISCONT, *HEAD-RIGHT(v), *HEAD-RIGHT(mod), *HEAD-RIGHT(aux)}

As for the contrast between *wett hürate* 'wants get_married' (grammar II) and *hürate wett* 'get_married wants' (grammar III), the relevant constraint pair is V_{nuc}-LEFT and *HEAD-RIGHT(mod). In grammar II, *HEAD-RIGHT(mod) dominates V_{nuc}-LEFT, which accounts for the grammaticality of *wett hürate* and the ungrammaticality of **hürate wett*. In grammar III, V_{nuc}-LEFT dominates *HEAD-RIGHT(mod), therefore *hürate wett*, **wett hürate*. However, we expect that both *hürate wett* and **wett hürate* are grammatical in a grammar where the two constraints are unranked. We have already dealt with unranked constraints in the tableaux above: no mutual ranking of two or more constraints could be determined on the basis of the evidence, since the two constraints in question are not in conflict. In other words, all possible rankings of these constraints lead to the same results. However, if two conflicting constraints are unranked, different rankings of these constraints lead to different results; in other words, unranking becomes crucial. Following McCarthy (2002: 7), I will call conflicting constraints

which are unranked 'tied constraints'. Non-conflicting unranked constraints vs. tied constraints can be defined as follows:

(19) Non-conflicting unranked constraints:
There is at least one output candidate satisfying all unranked constraints.

(20) Tied constraints:
(i) all candidates display at least one violation, (ii) at least two candidates display the same number of violations, and (iii) no other candidate displays a lower number of violations.

In deriving variation from tied constraints I follow Anttila 2002 and Löhken 1997 (ch. 2). Anttila (2002: 231) proposes that a grammar containing tied constraints allows for different rankings which are equally accessible in actual candidate evaluations. Schematically:

(21) Grammar:
... >> {Constraint 1, Constraint 2} >> ...

Actual candidate evaluations:
(i) ... >> Constraint 1 >> Constraint 2 >> ...
(ii) ... >> Constraint 2 >> Constraint 1 >> ...

Table 4. Schema for tied constraints.

	(Input)	(Others)	Constraint 1	Constraint 2	(Others)
☞	Candidate A		*		
☞	Candidate B			*	

Thus, if the transition from grammar II (*HEAD-RIGHT(mod) >> V_{nuc}-LEFT) to grammar III (V_{nuc}-LEFT >> *HEAD-RIGHT(mod)) is gradual, we might expect that a grammar exists where the two constraints in question are tied, which predicts that both *hürate wett* 'get_married wants' and *wett hürate* 'wants get_married' are available; however, no multiple outputs are found in other verb clusters – which is a desired result since it is not the idea to write a grammar where 'anything goes', but a grammar where variation occurs within strictly defined limits:

(22) Grammar VI: *ha wele gaa* 'have wanted go'
gleert faare 'learned drive'
wett hürate/*hürate wett* 'wants get_married/ get_married wants'
zalt hät 'has paid'

Ranking: {*BRANCH-LEFT, *DISCONT, *HEAD-RIGHT(v)} >> {V_{nuc}-LEFT, *HEAD-RIGHT(mod)} >> *HEAD-RIGHT(aux)

Table 5. Grammar VI with variable outputs of MOD/V clusters due to tie between V_{nuc}-LEFT and *HEAD-RIGHT(mod).

	*BRANCH-L	*DIS-CONT	*HD-R (v)	V_{nuc}-L	*HD-R (mod)	*HD-R (aux)
☞ *ha wele gaa*				*		
ha gaa wele				*	*!	
wele ha gaa		*!		*		*
wele gaa ha				*		*!
gaa ha wele		*!			*	
gaa wele ha	*!				*	*
☞ *gleert faare*				*		
faare gleert			*!			
hät zalt				*!		
☞ *zalt hät*						*
☞ *wett hürate*				*		
☞ *hürate wett*					*	

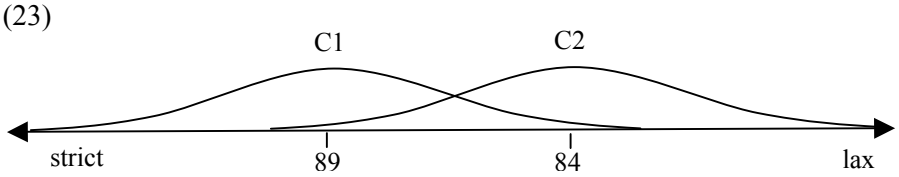
182 informants answered our questionnaires in accordance with grammar VI. Grammar VI occurs most frequently in the Central and Northern areas but its prominence decreases towards the West and the (South-) East. This is not an unexpected result: the more we move to the West, the more MOD–V is obligatory, and the same is true with V–MOD for the East.

To close this section, I will propose that one more extension is needed in our theory. So far, we have dealt with internal variation as if it were completely symmetrical, i.e., as if the two winners were equally acceptable. However, in the overwhelming majority speakers are able to rank acceptable variants according to preference. In our theory developed so far this fact cannot be reflected. We derived our results from total domination between constraints and from totally overlapping constraints. What we

need is a partial overlap of constraints, where the properties of domination and unranking are combined. Preference has in common with grammaticality contrasts (expressed by domination) that one output candidate is more acceptable than the others; it has in common with the model of a variable output grammar developed so far (with tied constraints like in grammar VI) that several (and not only one) output candidates are grammatical.

The idea of partially overlapping constraints has been elaborated within the context of Stochastic Optimality Theory (StOT; Boersma 1997, Bresnan and Deo 2001, Bresnan, Dingare and Manning 2001). StOT deviates from classical OT in two respects: first, constraints are ranked on a continuous scale of real numbers such that the distances between constraints differ. Second, there is a small amount of random perturbation at each concrete evaluation such that the actual ranking point of a constraint may be slightly higher or lower than the mean position the constraint occupies on the scale. The mean ranking position of a constraint is the peak of a probability curve of its actual rankings in concrete evaluations. If the distance between the mean ranking positions of two constraints C1 and C2 is rather small, reversed rankings can occur in actual evaluations. The closer the constraints are on the scale, the more the probability of reversed rankings increases. In (23), actual rankings with C1 >> C2 are more probable than actual rankings with C2 >> C1, though both do occur:

(23)



(taken from Bresnan, Dingare and Manning 2001: 10)

The mean ranking values are determined by the Gradual Learning Algorithm GLA (Boersma 1997; Hayes, Tesar and Zuraw 2002).[6.] For grammar VI, the GLA determined the following mean ranking values (200,000 sets of forms used for learning, 50,000 test cycles):

(24) Grammar VI: mean ranking values on a continuous scale:

114,000	*BRANCH-LEFT
114,000	*HEAD-RIGHT(v)
112,000	*DISCONT
103,392	V_{nuc}-LEFT
103,188	*HEAD-RIGHT(mod)
93,420	*HEAD-RIGHT(aux)

Ranking distance between V_{nuc}-LEFT and *HEAD-RIGHT(mod): 0.204

The ranking distance between the crucial constraints *BRANCH-LEFT and V_{nuc}-LEFT is so small that MOD–V and V–MOD are almost equally acceptable. I will now simulate a clear preferential asymmetry between the two options. Grammar VII below was learned by the GLA on the basis of input data with 80% V–MOD and 20% MOD–V. We expect that the ranking distance from V_{nuc}-LEFT to*HEAD-RIGHT(mod) is greater than in grammar VI:

(25) Grammar VII: *ha wele gaa* 'have wanted go'
gleert faare 'learned drive'
hürate wett 'get_married wants' 80%
wett hürate 'wants get_married' 20%
zalt hät 'paid has'

114,000	*BRANCH-LEFT
114,000	*DISCONT
114,000	*HEAD-RIGHT(v)
104,154	V_{nuc}-LEFT
101,846	*HEAD-RIGHT(mod)
94,000	*HEAD-RIGHT(aux)

Ranking distance between V_{nuc}-LEFT and *HEAD-RIGHT(mod): 2.308

In sum, it holds true for the entire area that head-final patterns are most expected with auxiliaries and least expected with lexical verbs as heads of the cluster. The implicative relationship AUX>MOD>V is reflected in the geographical picture, since the area with lexical verbs as cluster final heads is included in a larger area with modal verbs as cluster final heads, and the

latter is included in an again larger area with auxiliaries as cluster final heads. In OT, an explicit formulation is possible for the inclusive relation between the different head categories. All our dialects (grammars I–VII) are compatible with the ranking *HEAD-RIGHT(v) >> *HEAD-RIGHT(mod) >> *HEAD-RIGHT(aux). It is this ranking which our dialects have in common, and they differ in the ways how this ranking interacts with other constraints.[7]

6. Conclusion

In the present paper I have examined one typological parameter, the relative order of head and complement in the domain of verb clusters. It turned out that a great amount of variation is found across the dialects of German speaking Switzerland, insofar as the Western head-initial patterns gradually turn into more head-final patterns, the more we move towards the East. The prominence decrease of head-initial patterns across space can be established along different dimensions: head category (AUX, MOD, or lexical verb), density of measuring points, relative number of informants at one measuring point, and preference. As for the typologically relevant question of how many systems of grammar (specific combinations of structural features) are attested in our area, I have argued that it is necessary to distinguish between three basic types of dialect variation, in order to uncover the minimal contrasts between systems of grammar: diatopic variation (structural features are geographically distributed over different grammars), free variation (features co-occur in one grammar and are randomly distributed), and conditioned variation (features co-occur in one grammar whereby their internal distribution reflects certain systematic factors). Given the assumption that at every place at every time a system of grammar exists, the formal instruments of linguistic theory must be able to account for the minimal contrasts between grammars, even if the contrasts are only found in preference for one or another (equally grammatical) option. I have proposed a way how the similarities and differences between grammars can be captured in an Optimality theoretic setting.

Of course, there are many residual issues which I did not account for in the present paper. Some of them are the following:

(i) I only considered a subset of all possible verb cluster types. A survey of all types found in Zurich German is given by Lötscher (1978). Besides

AUX/V, MOD/V and AUX/MOD/V, Lötscher also considers MOD/AUX/V, AUX/V/V, or AUX/MOD/V/V.

(ii) The interaction of verb clusters with complements and adjuncts (objects, adverbials, etc.) is not considered here. Again, Lötscher (1978) provides some interesting data. For instance, in ascending MOD–V clusters of Zurich German the modal verb can precede a direct object:

(26) a. *wil de Joggel es gottlett* ***wott*** ***ässe***
because the (name) a cutlet wants eat
'because Joggel wants to eat a cutlet'

b. *wil de Joggel* ***wott*** *es gottlett* ***ässe*** (Lötscher 1978: 4)

It is an intriguing topic for further research to explore the conditions determining which of the above patterns are available in which dialects.

(iii) It is probable that information structure plays a part in the choice of the variants given in (26). The relevance of focus for verb clusters in general is neglected in the present paper. Schmid and Vogel (2002) give an OT analysis of focus sensitivity in the distribution of three-verb clusters. As for Swiss German, they point out that MOD–AUX–V is required if MOD is focused (Schmid and Vogel 2002: 12).

(iv) The theory outlined here does not in itself account for change. Whereas it is common sense in OT that change is to be expressed in terms of constraint reranking (cf. e.g. Löhken 1997: 85–86), it is entirely unclear what the motivations for constraint reranking are. Partly, the reason for a speaker to adopt a change lies simply in the fact that her/his neighbours (increasingly) talk that way, since changes spread across space. Possibly, spread of change can be modelled on a probabilistic basis: the more a speaker is confronted with a change, her/his constraint ranking will very slightly be adjusted towards that kind of input data, and the more the ranking is adjusted, the more changed data she/he will produce, etc. – causes and consequences of geographical spread feed each other. However, actuation, i.e. the first occurrence of a change, remains unexplained. Even so, to be more modest, the gradual nature of change including stages of variation can at least be pictured in (variants of) OT in a very plausible way, as a completely natural and expected thing.

Notes

1. The present paper is a revised and extended version of a talk I gave in the Dialectology and Typology Workshop at the METHODS XI Conference, Joensuu (Finland), August 2002. I gave a talk related to the present topic at the Stanford Department of Linguistics (California) in October 2002. I am grateful to Bernd Kortmann, Freiburg (Germany), for the organization of the Joensuu workshop and the present volume, to Joan Bresnan, Stanford (California), for giving me the opportunity to discuss these issues in the Stanford syntax colloquium, and to the audiences for pointing out problems and providing insights during fruitful discussions. I thank Kathrin Würth, Jena (Germany), for many helpful comments. Of course, all mistakes are mine.
2. Project director: Elvira Glaser; Swiss National Science Foundation 1114-57121.99.
3. Single answers to multiple choice questions are not taken into account on the maps, since there is a certain risk that single answers are due to accident, e.g., if the informant indicates variant 2 but 'means' variant 1. Thus, a symbol is introduced only if at least two informants accordingly indicate preference for the same variant.
4. I have to admit that this is perhaps an artefact of our experimental design. In multiple choice questions, we asked the informants for a decision which of the acceptable variants was the 'most natural' for them. Since informants wanted to be cooperative (otherwise they would not have sent back our questionnaires) they made this decision. However, very few informants explicitly indicated that they give the same ranking to two options. Thus, they had the opportunity to extend 'most natural' over two variants, though this occurred only sporadically. I conclude from these facts that equality between variants really exists (and has to be accounted for in any model of grammar), though it is not very common.
5. I guess that a preferential asymmetry is essentially the same as a statistical bias. From our experiment, we do not have information on text frequencies but only on preferences. I stipulate that the statistically more frequent pattern is the preferred one, too – knowing that this is not more than a working hypothesis.
6. "Where do the real number ranking values in a stochastic grammar come from? Starting from an initial state grammar in which all constraints have the same ranking values (arbitrarily set to be 100.0), the GLA [Gradual Learning Algorithm] is presented with learning data consisting of input-output pairs having the statistical distribution of, say, English. For each learning datum (a given input-output pair), the GLA compares the output of its own grammar for the same input; if its own output differs from the given output, it adjusts

its grammar by moving all the constraints that disfavor its own output upward on the continuous ranking scale by a small increment, in order to make them apply more strictly, and moving all constraints that disfavor the given output downward along the scale by a small decrement, to relax their effects." (Bresnan, Dingare and Manning 2001: 11)

7. Of course, in most of the grammars I–VII, some of the three constraints mentioned here are unranked since no internal evidence for a full ranking is found. However, my point is that all rankings are *compatible* with *HEAD-RIGHT(v) >> *HEAD-RIGHT(mod) >> *HEAD-RIGHT(aux). In other words: *no reversed* rankings occur.

References

Anttila, Arto
2002 Variation and phonological theory. In: Jack K. Chambers, Peter Trudgill and Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, 206–243. Malden, MA/Oxford: Blackwell.

Boersma, Paul
1997 How we learn variation, optionality, and probability. *Institute of Phonetic Sciences, University of Amsterdam, Proceedings* 21: 43–58.

Bresnan, Joan and Ashwini Deo
2001 Grammatical constraints on variation: *'be'* in the *Survey of English Dialects* and (Stochastic) Optimality Theory. Ms. Stanford University. http://www-lfg.stanford.edu/bresnan/be-final.pdf

Bresnan, Joan, Shipra Dingare and Christopher D. Manning
2001 Soft constraints mirror hard constraints: Voice and person in Lummi and English. In: Miriam Butt and Tracy Holloway King (eds.), *On-line Proceedings of the LFG2001 Conference*. Stanford: CSLI Publications. http://csli-publications.stanford.edu/LFG/6/lfg01-toc.html

Bucheli Berger, Claudia and Elvira Glaser
2002 The Syntactic Atlas of Swiss German Dialects: empirical and methodological problems. In: Sjef Barbiers, Leonie Cornips and Susanne van der Kleij (eds.), *Syntactic Microvariation*, 42–74. Amsterdam: Meertens Instituut (Meertens Institute Electronic Publications in Linguistics). http://www.meertens.knaw.nl/projecten/ sand/synmic/

Comrie, Bernard
1989 *Language Universals and Linguistic Typology*. 2nd edition. Oxford/ Cambridge, MA: Blackwell.

Croft, William
2003 *Typology and Universals*. 2nd edition. Cambridge: University Press.
Dryer, Matthew S.
1992 The Greenbergian word order correlations. *Language* 68: 81–138.
Hawkins, John A.
1983 *Word Order Universals*. San Diego: Academic Press.
Hayes, Bruce
2000 Gradient well-formedness in Optimality Theory. In: Joost Dekkers, Frank van der Leeuw and Jeroen van de Weijer (eds.), *Optimality Theory. Phonology, Syntax, and Acquisition*, 88–120. Oxford: University Press.
Hayes, Bruce, Bruce Tesar and Kie Zuraw
2002 "OTSoft", software package, http://www.linguistics.ucla.edu/people/hayes/otsoft/
Hodler, Werner
1969 *Berndeutsche Syntax*. Bern: Francke.
Greenberg, Joseph H.
1966 Some universals of grammar with particular reference to the order of meaningful elements. In: Joseph H. Greenberg (ed.), *Universals of Grammar*, 2nd edition, 73–113. Cambridge, MA: MIT Press.
Kager, René
1999 *Optimality Theory*. Cambridge: University Press.
Kortmann, Bernd
2002 New prospects for the study of English dialect syntax: Impetus from syntactic theory and language typology. In: Sjef Barbiers, Leonie Cornips and Susanne van der Kleij (eds.), *Syntactic Microvariation*, 185–213. Amsterdam: Meertens Instituut (Meertens Institute Electronic Publications in Linguistics). http://www.meertens.knaw.nl/projecten/sand/synmic/
Kroch, Anthony
2001 Syntactic change. In: Mark Baltin and Chris Collins (eds.), *The Handbook of Contemporary Syntactic Theory*, 699–729. Malden, MA/Oxford: Blackwell.
Lehmann, Winfred P.
1973 A structural principle of language and its implications. *Language* 49: 47–66.
Lightfoot, David
1991 *How to Set Parameters: Arguments from Language Change*. Cambridge, MA: MIT Press.
1999 *The Development of Language. Acquisition, Change, and Evolution*. Malden, MA/Oxford: Blackwell.

Löhken, Sylvia C.
1997 *Deutsche Wortprosodie. Abschwächungs- und Tilgungsvorgänge*. Tübingen: Stauffenburg.
Lötscher, Andreas
1978 Zur Verbstellung im Zürichdeutschen und in anderen Varianten des Deutschen. *Zeitschrift für Dialektologie und Linguistik* 45: 1–29.
McCarthy, John J.
2002 *A Thematic Guide to Optimality Theory*. Cambridge: University Press.
Nagy, Naomi and William Reynolds
1997 Optimality theory and variable word-final deletion in Faetar. *Language Variation and Change* 9: 37–55.
Paolillo, John C.
2002 *Analysing Linguistic Variation. Statistical Models and Methods*. Stanford: CSLI Publications.
Schmid, Tanja and Ralf Vogel
2002 Dialectal variation in German 3-verb-clusters. A surface-oriented OT account. Ms. University of Potsdam, University of Stuttgart. http://www.ling.uni-potsdam.de/~rvogel/schmid-vogel02.pdf
Seiler, Guido
2003 *Präpositionale Dativmarkierung im Oberdeutschen*. Stuttgart: Steiner.
Vennemann; Theo
1973 Explanation in Syntax. In: John P. Kimball (ed.), *Syntax and Semantics*, Vol. 2, 1–50. Tokyo: Taishukan.
Whaley, Lindsay J.
1997 *Introduction to Typology. The Unity and Diversity of Language.* Thousand Oaks/London/New Delhi: Sage Publications.
Wurmbrand, Susi
forthc. Verb clusters, verb raising, and restructuring. Case to be included in the Syntax Companion. Oxford: Blackwell. http://www.arts.mcgill.ca/programs/linguistics/faculty/wurmbrand/research/files/verbclusters/Syncom.pdf
Zwart, Jan-Wouter
1996 Verb clusters in continental West Germanic dialects. In: James R. Black and Virginia Motapanyane (eds.), *Microparametric Syntax and Dialect Variation*, 229–258. Amsterdam/Philadelphia: Benjamins.