Paradigmatic scale building in OT: The case of strong verbs in German

ANATOLI STRIGIN

ZAS Berlin

1. INTRODUCTION

Optimality Theory (OT), introduced by Prince and Smolensky (Prince & Smolensky 2004) as a grammar model in generative phonology, has recently gained popularity in other domains of linguistic research as well, but work on pure morphology is still infrequent and there is no agreed upon framework for representing realisations of morphosyntactic features via rules in the vein of Matthews (1991) and Stump (2001). Some proposals already exist, e. g. Russell (1999), MacBride (2004), Yip (1998), Xu (2007), but in general the properties of OT morphology remain largely unstated.

The paper presents something like a feasibility study on tentative properties that might be needed in OT to solve problems in morphology. It examines the phenomenon of paradigmatic scale building among strong verbs in German and suggests some extensions to the standard concept of a constraint in OT which could help to explain this phenomenon. It also provides a tentative explanation of the phenomenon itself in the framework thus modified.

2. AN OBSERVATION IN THE GERMAN LANGUAGE EVOLUTION AND A PROBLEM

The evolution of the verbal system of the German language shows an interesting phenomenon of PARADIGMATIC SCALE BUILDING.

Recall that in the Germanic languages we have the so-called **strong verbs,** which are typically those which mark their past tenses by means of a regular change of their root vowel (mostly *ablaut*), e. g. *nehm-en – nahm – ge-nomm-en* (*take – took – taken*). They are contrasted with **weak verbs**, i. e. are those verbs that form their preterites and past participles by means of a dental suffix - an inflection that contains a /t/ or /d/ sound or similar, e. g. *leb_en* - *leb_t_e*– *ge_leb_t* (*live* - *lived*). Most verbs in the early stages of the Germanic languages were strong. However, almost all new verbs in Germanic languages are weak, and the majority of the original strong verbs have become weak, too.

The transition from strong verb forms to weak verb forms involves characteristic places in a verb paradigm at which strong verbs differ from weak verbs. The following table is a summary of the relevant characteristics of some strong verbs which differentiate them from the weak verbs and have therefore to be changed during the transition (Bittner 1996).

Index of the characteristic s	The characteristic relative to the root as exemplified by the infinitive
A	The change of *e* to *i* plus the absence of an ending in the Imperative.Sg
B	Vowel change in 2 and 3 Per.Sg in PresTns
C	The ending in the PretTns 1 and 3 Per.Sg is either not dental or absent
D	Vowel change (*Ablaut*) in PretTns
E	Vowel change (*Umlaut*) in PretTns.Conj
F	-*en*-ending (and *Ablaut*) is the second participle (Part.II)

Table 1: strong verb characteristics

The first intriguing observation in (Bittner,1996) is that the transition involves not all strong verbs at once. Some have already replaced all their strong forms by the weak ones, some have retained all of them. The second observation is that such forms are not replaced all at once, and there are strong verbs which are so to say "on the way" to their new weak status. There are about 50 strong verbs in German which are under transition at present. They have completed the transition at a different number of positions, as of now. The third observation is that these transitional classes are not random.
Bittner (1996) argued that the alphabetic ordering of the indices in the table indicates an implicative scale (see (Guttman1944). If we consider the pure strong verbs and the weak verbs as the two poles, then it is possible to order the six characteristics (where applicable) so that the loss of a characteristic predicts the loss of all the characteristics ranked before it:

- if V is a pure strong verb, then it has all of A ... F;
- if V is a weak verb, then it has none of F ... A;
- if V has a strong form at characteristic with index *i*, it has strong forms at all the characteristics with indices that follow *i*. and if V substituted the weak form for the strong form at its characteristic with index *i*, it has already done so for the preceding characteristics.

Vowel change was not obligatory for the past participle, and the implication applies in principle whenever there was an *ablaut* in PPII.

Consider the following examples in Table 2. The mark ✓ shows the forms where the transition already occurred.

Infinitive	*geben* *give*	*fahren* *drive/go*	*heben* *pick up lift*	*brennen* *burn*	*schinden* *flay*	*melken* *milk*	*leben* *live*
A	gib	✓fahre	✓heb	✓brenne	✓schinde	✓melke	lebe
B	gibt	fährt	✓hebt	✓brennt	✓schindet	✓melkt	lebt
C	gab	fuhr	hob	✓brannte	✓schindete	✓melkte	lebte
D	gab	fuhr	hob	brannte	✓schindete	✓melkte	lebte
E	gäbe	führe	höbe	brannte	schünde	✓melkte	lebte
F	gegeben	gefahren	gehoben	gebrannt	geschunden	gemolken	gelebt

Table 2: strong verb transitions

Optimality Theory (OT), introduced by Prince and Smolensky (Prince & Smolensky 2004) as a grammar model in generative phonology, has recently gained popularity in other domains of linguistic research as well, but work on pure morphology is still infrequent and there is no agreed upon framework for representing realisations of morphosyntactic features (in the vein of Matthews (1991) and Stump (2001). How can these characteristics of the transition of strong German verbs be explained in the framework of OT-based theories of grammar, grammar learning and language evolution?

Among the different flavours of OT only the stochastic OT (StOT, Boersma,1998) provides a well spelt-out approach to variation. To follow the line of explanation which uses stochastic OT e. g., the one adopted in Jäger (2003) the explanation task can be formulated more precisely:

- find a set of constraints which allows both the forms of strong and weak verbs to surface as alternative realisations of some morphosyntactic features (learning state zero)
- find a ranking on it which allows the current state of affairs (CS)
- find a ranking on it which allows all the strong verbs which now have weak forms to have there original strong forms (initial evolution state)
- show how the learning algorithm of StOT, combined with a model of evolution, could lead from the initial evolution state to the current state of affairs observing the implicative scale.

There are some problems with the realisation of the task in the standard version of StOT. The paper attempts to solve these problems by suggestion some extension to the standard and provide an explanation of the phenomenon in the modified framework of OT. The solution proposed may also prove to have general validity for the development of the OT framework.

3. The problems in the realisation of the task and a sketch of a solution

3.1 Bits of OT morphology

An OT system defines a relation GEN which specifies all the possible underlying/overt representational structure pairs for any language. Such structures may correspond in a rather loose sense to meaning and form in general, but usually reflect the relative position wrt. observable part of the data, depending on the domain being described (OF for Overt Form and UF for Underlying Form will be used in the sequel instead of overt/underlying structure). Following Russell (1999), MacBride (2004), Yip (1998), Xu (2007) it will be assumed that morphological information associated with inflectional affixes is introduced via realisation constraints (RC).

Applying the nomenclature to the problem in question, we consider a set GEN_{GerV} of pairs (UF,OF) where UF is a conjunct of some semantic information associated with a root, e. g. [leb], and a complex of some more or less easily semantically expressible morphosyntactic features, e. g. [Inf] or [pl] (for infinitive and plural). OF will be some representation which is explicit enough for our purposes to reflect the relevant characteristics of the verb form. e. g. /leb/+/en/ for *leben*. Thus ([leb]+ [Inf],/leb/+/en/) is a member (UF,OF) of GEN_{GerV}. This can be basically enforced by some realisation constraint of the kind

[Inf] => /-en/

which applies conditionally and is vacuously satisfied by any UF which does not contain [Inf], i. e. is a material implication by nature. But this cannot be material implication, of course, since we have no inference rules in OT. We shall return to the interpretation of this form of RC in OT.

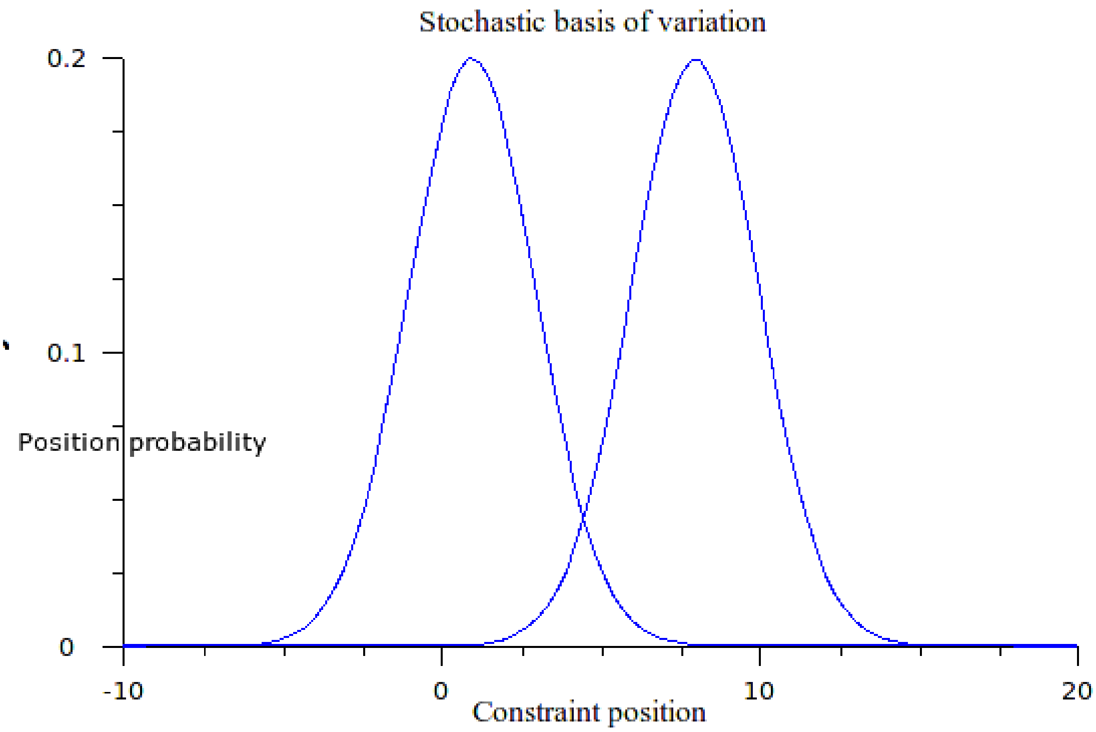
Simplifying we shall assume that the realisation space for all morphosyntactic features contains only the forms which are relevant to the discussion.

3.2 The problems in the realisation of the task

The exact formulation of an OT grammar depends on the ranking scale adopted. The original OT grammars employ an ordinal scale for constraints, in which case practically no OF variation is possible. Stochastic OT (StOT) grammars use an interval scale and can much better represent variation. Evidently the learning problem in case of German verb transition should be formulated so that the variation of rivalling OFs be reflected.

A grammar in StOT is defined by assigning ranks to the constraints on an interval scale under a normally distributed probability function, i. e. treating constraints rankings as continuous random variables. The rank of a constraint C is thus a point together with a normal probability distribution function, with the same variance for all constraints but possibly with a different mean. The point defines the mean of the function showing the most probable location of the constraint in the constraint ranking. But the actual location of the constraint is determined for every use of the grammar anew by sampling this constraint (Jäger 2003, Boersma 1998). The sampling is done by generating a point for each constraint by some random process with the same normal probability distribution. The stochastic element is the basis for the treatment of the form variation.
Variation in the StOT is treated by locating the constraints responsible for the variation close enough so that the probability of their actual positions being swapped is high enough. Consider the probability distributions for the positions of two constraints in (1).

(1)



At the position where the two curves cross the probability for each constraint that it is located there is the same. But it is rather low, so we may expect that there will be little variation and the leftmost constraint will be dominating, if the domination relation is computed from left to right. If the overlap were more, the probability of the variation would be higher.

Since we have realisation constraints to manage the morphology in our model, we need at least two realisation constraints for one UF to obtain two variants, say RC1: UF => A and RC2: UF => B. A is the preferred realisation, if either the probability of RC1 to be ranked higher than RC2 is much higher than of the inverse case, or there are either one or a pair of constraints which are ranked above RC1 and RC2 and are related to these RCs or to each other in the same way.

Adopting this framework to explain the observations without modifications leads to problems.

Problem 1. The first problem in devising a solution becomes evident if we consider the simple case where the realisation constraints are solely responsible for the variation. If UF is a morphosyntactic feature, the variation equally affects all the verbs since realisation constraints apply to verb-independent UFs. An individual pace of transition for the strong verbs at this position is therefore impossible. Yet this is the first observation we should explain.

Problem 2. The second problem concerns the scale. Suppose we have RC1: UF => A and RC2: UF => B and RC3: UF1 => C and RC4: UF1 => D. How do we express the conditional constraint on grammar G that if RC2 is ranked higher than RC1 in G then RC3 is ranked higher than RC4 in G? Actually, since the underlying theory is stochastic, a more elaborate statement is required: the probability of this conditional constraint must be very high.

Problem 2 is connected with Problem 3 which concerns the evolution of the grammar. Learning can be iterated, with the learned grammar serving as a generator for a new training set. This iterated learning model (ILM) of language evolution was discussed by Kirby and Hurford (Kirby & Hurford 2002). Put simply, ILM is a learning loop: at each iteration of the learning scenario speakers choose some meanings and produce signals to convey them. The resulting meaning-signal pairs serve as training sets for learners. Having learned the language, the learners become speakers and produce signals in their turn, replacing the previous speakers as the language source. If the grammars learned do not exactly correspond to the grammars of the previous speakers, the language has a chance to evolve with time. If the learning task consists in correctly reflecting the distributions of the variants, a language evolves by changing these distributions. Recently, Optimality Theoretic (OT) systems were exploited to provide models of language learning to model language evolution. Some of these studies suggest that there may be stable states in the language development (Jäger 2003), and that language evolution is directed towards such states.

The transition of strong verbs to weak verbs is the kind of development that probably can be modelled in this way: strong verb forms go out of existence, weak forms remain as a stable state. The question is why and in which way this happens.

Problem 3. There is no obvious reason why the relative positions of the RCs should change with time, leading to the a change in the distribution of variants, instead of fluctuating, or that this change should proceed in on the observed scale. Some experiments suggest that stem change is no more difficult to learn than affixation in principle, but there is a type frequency effect on learning (Bybee & Newman 2007).

3.3 A first sketch of a solution

Let us start on the assumption that we have only the realisation constraints, two per morphosyntactic feature and consider RC1 and RC2 for some feature which account for the two approximately equally frequent variants. They have therefore a high swap probability and are located very near to each other on the ranking scale. Note that if we use a learning algorithm which is very good at coding the frequency distribution of the realisations of the variants into the constraint positions like the one used in (Jäger 2003), we could be almost certain that we get the correct current state of the grammar. This starting point is characterised by all the problems, however.

Let us start solving Problem 3 first.

Starting again with the simple case above suppose we introduce an additional constraint - call it CWF - which prefers the weak form realisation of some morphosyntactic feature F. Suppose it is ranked below both of the realisation constraints. The variation potential is as it was. If, however, CWF is frequently ranked above the realisation constraint with which it conflicts, the variation in the speech of such a speaker becomes skewed. In the next generation the variants do not have the same relative frequencies. In the course of the evolution, ceteris paribus, the dispreferred variant will disappear. Again, this process affects all the verbs.

In view of the fact that at the beginning of the drift strong forms of strong verbs prevail absolutely what could be a reason to rank the decisive constraint higher then RCs? Arguably, if such a constraint was operating on a greater level of generality than RCs, this generality could be reason enough for it to be ranked higher then all RCs in the learning state zero. Then, if the strong verbs were predominant at IS, the RCs for strong forms can become ranked higher than the prohibiting the CWF constraint during learning. With weak verbs growing in numbers their cumulative effect would also grow, so at some point it would neutralise the cumulative effect of the strong forms on moving strong form RCs up the ranking and finally outweigh it and move them down the ranking. It would then become almost impossible for the strong realisation constraints to outrank CWF. Of course, as long as CWF concerns only one feature its higher degree of generality is rather dubious. Note moreover that it is the cumulative effect of the forms of all the verbs which is important, since RCs do not mention individual verbs.

The next problem in connection with this solution which needs explanation then is the increase of weak verbs over time.

Presumably, already in Old High German, new verbs, which were derived from verbs, adjectives or nouns, took weak forms predominantly and weak verbs were predominantly derived. Why couldn't new derived verbs take strong forms? It's common wisdom in German linguistics that they did not get them.

Now consider the situation in which a new verb is derived from a [N+]-root. This verb has to get its forms assigned. Since such verbs were formed from *one* base, there was not a base pair to establish a change, in contrast to the strong verbs. And since there was no established vowel change, there was no evidence for any vowel change which could help to re-rank RCs. Hence strong RCs for F would remain dominated by CWF. And weak forms would function somewhat like a default form.

So an introduction of a new derived verb would be accompanied by employing a default weak form for which there is no contradicting evidence, weak forms would accumulate with the increase of the number of derived verbs and could erode the strong form paradigm.

Assuming this line of reasoning may prove to be fruitful, the major remaining problem is then Problem 2.

4. A more detailed treatment of the German strong verb transition

4.1 Learning morphology: rules of exponence

Two views of what exactly is learned when learning morphology have to do with the learning of lexical items. The first is to view morphology as learned on an individual sign basis: what is learned is a word as a mapping between the bundle of its core meaning cum morphosyntactic properties and its form. No analysis is performed on words. This view does not incorporate any generalisations, hence it is not well suited to model grammars. The second view is based on the notion of morpheme, so that a learner has to arrive e. g. at a correct morpheme for the preterite in a similar way that s/he has to arrive at a correct lexical entry. But this view, too, is known to have its theoretical problems. And though it possibly seems to be easily integrated into the learning model, it is unclear, e. g. how to represent vowel shift as an OF_i in a morpheme, since no notion of a morpheme covers comparing two forms. And at least at present (in CS) *ablaut* has become so irregular that it would probably present difficulties, if considered as an instance of template morphology.

An alternative view would consider the notion of a morpheme as derivative at best, and operate with mapping rules linking UF and OF. Such rules, known as realisation rules or rules of exponence (Matthews 1991, Stump 2001), could be considered as recording more or less strong tendencies in mapping morphosyntactic structures to OFs. They are largely verb independent, but inasmuch as they need to consider pairs of forms to establish if there is a vowel change, they are sometimes also bound to individual roots. The vowel change in the root of the strong verbs is predominantly *ablaut* and therefore can be shown to have some systematicity when projected onto Proto-Indo-European (Coetsem 1993, Watkins 2000). But German *ablaut* is not considered here to be an infix or a case of template morphology. Whatever similarities in the *ablaut* context are perceived by a learner, they are used as heuristics at best. The exponence rules for German verbs do not substitute vowels in roots. This is the single important basic assumption made in the paper.

4.2 Representing UF and OF of the German verbs morphology

To keep the task of the paper manageable we shall make some simplifications in GEN. The first simplification concerns the characteristic positions. They will be cut to those involving vowel change. So we do not consider position C. The second simplification concerns the treatment of verb morphology.
Suppose we have the following ingredients of UF:

- the set of 7 semantic (root) pointers {[*geb*], [*fahr*], [*heb*], [*bren*], [*schind*], [*melk*], [*leb*]}
- the set of semantic pointers of morphosyntactic features {[V],[Inf], [Imper], [Sg], [Per],[Pres],[Pret], [Conj],[PII]}

A representation of a verb at UF is an expression of the form [1]+[2] with [1] a root pointer and [2] a feature pointer. The roots without any feature cannot be realised to play a role in the acquisition, so each verb root is accompanied by at least the [V] feature. And we assume that the feature [Inf] is the rendering of some basic form of the feature [V], hence [V] is viewed as being normally realised by *-en*[1]. We assume that the prefix *ge-* is added automatically to the form realising past participles and ignore it. On the OF side we assume that there are three possible endings, -/*e*/, -/*en*/, -/*t*/ and one suffix, -/*t*/- (in writing OFs we will drop the slashes). There are, of course, more of them, but we assume that there is no need to distinguish any two endings except with respect to their being dental or not, so we use *-e* for the non-dental and *-t* for those containing a dental, but retain the *-en* ending type.

A feature combination in [2] of [1]+[2] will be rendered by the concatenation of their names, e. g. [ImperSg] for imperative singular. We shall represent all the weak verbs cumulatively by a single verb - here *leben* (*live*).

A difficult question is whether vowel change should have a special treatment as a realisation possibility, and if yes, how to represent it.

For the present investigation it is sufficient just to note that a change in the vowel of a root can only be established when two roots are compared, so treating vowel change as a feature realisation involves keeping track of at least two roots. We need not specify here precisely which roots are being compared, it is enough to register this fact by assuming that such a change is rendered by a pair of roots. Similarly, there is no need to associate this pair of roots with any specific morphosyntactic feature, as long as it is associated with some feature combination. And lastly, since we re not interested in the

1 It seems that German children start learning their language by making the assumption that *-en* is a kind of default ending to show that a word is a verb, marking perhaps [+V-N] with it. Later they add [+distance] to [+V-N], so that [Inf] may just 'inherit' the ending (Bittner 2003). OT has not investigated the concept of a dynamic or changing GEN as yet, so we cannot incorporate this insight into our treatment.

exact spelling of the change we may simply render it by some mnemonics, i. e. let it be *vc* for vowel change.

Table 3 is then a simplified tabulation of the possible realisations of the features we are interested in. Transitions from strong to weak forms are indicated by denoting the initial state by IS and the current (or projected) weak state by CS.

zero	-e	-en	-t	-t- -zero	-t- -e	-t- -en	-t- -t	vc zero	vc -e	vc -en	vc -t	vc -t-zero	vc -t- -e	vc -t- -en	vc -t- -t	Features
		IS CS														Inf
	CS								IS							ImperSg
			CS								IS					2pSgPres 3pSgPres
					CS			IS								3pSgPret
	CS								IS							3pSgConj
		IS		CS						IS						PII

Table 3: a realisation matrix (UF to OF) for German verbs

4.3 Constraints

The following known constraint types will be useful in the treatment:

- realisation constraints, RC;
- output-output faithfulness constraints
- structural constraints.

They will be supplemented by a new type of constraint which will be called paradigm structure condition constraints.

Realisation constraints

We postulate that there are constraints for each cell of Table 3 which require that this cell is a legitimate (UF,OF) pair. All in all there are thus 6 X 16 = 96 constraints. They will be referred to in the usual table notation, $C_{i,j}$, where *i* is the row number starting from top (row 1) and *j* is the column number starting from left (column 1). Note that zero vowel alternation is still a vowel alternation formally, so it is comparable with the other vowel alternations. These constraints have the form of implications, A => B. Since we do not explicitly indicate which vowel changes to which, the table and the constraints defined by it are, of course, inadequate as they stand. They should be expanded by the contextual information <base, vowel_change> indicating which vowel change is meant and which base is intended. This information is of course provided the moment the lexical item is chosen. For instance, if the item in question is [*geb*]+3pSgPret the contextual information is <[geb],<C_e_C,C_i_C>>. We skip the formalisation of the details of this process.

Output-output constraints.

This kind of constraint was discussed by Benua (1995). We shall need the implementation which requires that a verb root in any paradigm form have the same segments as the verb root of the base of comparison. This requirement is very much like the maximal identity constraint on two outputs by Benua, and will be dubbed MAX-OO constraint. Clearly only the zero vowel change roots satisfy it, i. e. the pairs (*i,j*) where 1 <= *i* <= 6 and 1<= *j* <= 8 (the pairs violating this constraints are greyed out in Table 3.

Structural constraints

The two structural constraints express our understanding of the interplay between laziness and informativeness. We shall require that every feature have some realisation (MAX-SRUCT). And we shall require on the other side that in general there should not be any structure in OF apart from the root (*STRUCT). These constraints are usual, but in this paper their role is probably only to ensure that the dark grey area of Table 3 is not used, because it fails *STRUCT more than the other possibilities.

Note that these constraints are hardly an interesting theory of the German verb inflection. But whatever the constraints in such a potential theory, some of them will have a similar effect. So we are simply assuming an extensional equivalence in a very restricted area of modelling.

4.4 Constraint interaction and individual constraints

Let us start by examining how a possible grammar would look like in case of the current state of the language, CS. Consider only the realisation constraints, RC, first, e. g. $C_{3,12}$: 2pSgPres ==> *vc+-t*. To get all the correct realisation variants the corresponding realisation constraints $C_{i,j}$ should dominate their rivals. This is impossible, however, since weak verbs and strong verbs require different realisations for some feature complexes at UFs. To continue the example, $C_{3,6}$: 2pSgPres ==> *-t* contradicts $C_{3,12}$. So if we accommodate the weak verb ending by letting $C_{3,6}$ dominate $C_{3,12}$, vowel change will be entirely improbable, contrary to the observed state of things. Allowing a hight probability for the vowel change for some strong verbs would allow it for all. This problem requires a solution. And we need a more general solution, applying to all the inputs at once.

Suppose MAX-OO is the highest ranked constraint and both *STRUCT and MAX-STRUCT are ranked lower than the realisation constraints, so that we can ignore them. Then *all* vowel changes are ruled out. So if any strong verb has one of its forms replaced by a weak form due to this interaction, all will have done so. This general rule does not prescribe any kind of realisation, but favours those for the weak verbs, making then a kind of defaults.

On the other hand, since vowel change is assumed to be detected on the individual level, we could attempt to introduce INDIVIDUAL RCs which are actually RCs parametrised by verb roots. Thus we will have constraints like, e. g. $C[geb]_{3,12}$: 2pSgPres ==> *vc+-t* for *geben* (*give*) and $C[nehm]_{3,12}$: 2pSgPres ==> *vc+-t* for *nehmen* (*take*).

Such constraints may help to reflect the process by which the general rule and the individual rule diverge. The idea is simple: such an individual constraint may come to dominate MAX-OO, whereas the general rule $C_{3,12}$ can be still dominated by it. Thus, strong verbs in general will have a weak realisation form for their 2pSgPres, whereas *gibst* (you give) preserves the strong form, cf(1).

(2)

[geb]+2pSg Pres	$C[geb]_{3,12}$	MAX-OO	$C_{3,12}$	$C_{3,6}$
-t *gebt*	*!		*	
☞ *vc+-t* *gibt*		*		*

Individual RCs help to solve the first problem. It is unclear, however, at this stage, how they can come into being and how they get promoted.

The basic operations in StOT which is responsible for the final ranking are promoting each constraint from a set by some small predefined amount ε and demoting the constraints from its complement set by the same amount[2]. To deal with the implicative constraints we shall treat them constructively and modify this algorithm somewhat.

Assume that in the learning state zero we have a constraint ordering including (3).

(3) MAX-OO >> $\{C_{i,j}\}$ >> $\{C[root]_{i,j}\}$

This ordering reflects the generality hierarchy of the application of the constraints: an OO-constraint is ranked initially higher than the exponence rules for morphosyntactic features and these are ranked higher than individual lexical constraints.

If a grammar predicts an incorrect optimal item, then standardly the constraints favouring this item are demoted by ε and those disfavouring it are promoted by ε. An implicative constraint has a range of application however for which it is irrelevant; these are items which satisfy this constraint vacuously. It seems to be incorrect to let such cases promote the constraint. Consider a lexical realisation rule like $C[geb]_{3,12}$. This constraint is vacuously true of all other roots and of all feature combinations. Suppose that for such cases nothing happens. So no item except the one containing [geb]+3pSgPres as UF leaves a trace on the position of this constraint. This seems to be more in line with the intended interpretation of the implicative constraints. So this will be the way they are handled in learning.

Now lets us make another assumption, which is important for the individual constraints. Assume that there can be only one unique individual constraint per verb root and feature combination. At this level then there can be no variation. If a verb has variants, this is either because of an interaction between two general exponence rules or because of an interaction between an exponent rule and a lexical realisation constraint.

So $C[geb]_{3,12}$ is the only lexical constraint for this UF. Since almost all the observed instances of [geb]+3pSg are strong, as long as MAX-OO dominates it, $C[geb]_{3,12}$ will be successively promoted until it dominates MAX-OO strongly enough to make the probability of their swapping be very small, PROVIDED THE FREQUENCY OF THIS OF IS HIGHT ENOUGH.

But what is this 'high enough'? To obtain a plausible viewpoint consider what happens to MAX-OO. Every strong form demotes MAX-OO, every weak form promotes it. As long as the proportion of strong and weak forms are balanced, MAX-OO remains in

2 We assume here that all constraints are applicable, so there is no truth-value gap.

place, allowing $C[geb]_{3,12}$ after some exposure to be ranked higher.

This line of reasoning applies to all features and all roots. It makes the prediction that strong forms with individually high frequency of occurrence RELATIVE TO THE OVERALL PROPORTION OF STRONG FORMS VS WEAK FORMS have more chances to survive.

Note that all the strong exponence rules may outrank MAX-OO, if there are considerably more strong forms than weak forms cumulatively. Presumably then there could have been a time when strong forms could function as default, outranking MAX-OO and the rules for the weak forms. This period would be characterised less by the derivation of weak verbs than by the derivation by, e. g. prefixation to strong forms or by coining new roots which could get their basic forms by analogy, provided the *ablaut* were still transparent. The advent of weak verbs would gradually shift the rules for strong forms downwards to positions below weak form rules, making the latter the default case.

5. THE ROLE OF RULES IN PARADIGM FORMATION AFTER WOLFGANG WURZEL

The treatment discussed above makes crucial use of two constructs in one, exponence rules and implications. Exponence rules are based here on the notion of an implication, which forms the structure of such rules. Implication also underlies another notion of a morphosyntactic rule – the rule as a part of mechanisms forming inflexional classes. Such rules state that if some class of items fall under realisation rule RC1 it may be expected that elements of this class also fall under realisation rule RC2.

Are rules of this kind an interesting entity? Some linguists claim that they are. Wurzel (Wurzel 1989, Wurzel 1998) appealed to them to explain some properties of paradigms. We shall use his ideas on implicative relations in a paradigm to form the basis of the other construct which will be used to tackle Problem 3, i. e. the existence of the transition scale.

Wurzel investigated the formation of inflectional classes in different paradigms – nominal verbal, etc,- in different languages. An inflectional class is defined through rule application. A subset of elements of some syntactic category the inflections of which are generated by the application of a specific set of rules constitutes an inflectional class if any other member of the category differs in its inflection with respect to at least one rule. A good example would be the set of German feminine nouns ending in /-ung/, which have uniform inflexions.

Wurzel argued that inflectional classes manifest clusterings of rules which tend to define more natural (in his terms) classes as contrasted to those defined by an arbitrary rule collection. Wurzel distinguishes system-dependent and system-independent naturalness. We will consider only system-dependent naturalness here. It comprises *inflectional class stability* and *system congruity*. We shall briefly consider these notions starting with the latter.

Rules and clusters defining inflectional classes are called *structural properties* of the inflectional system by Wurzel. Thus, for instance, *Hund* (*dog*) and *Zirkus* (*circus*) have basic form inflection: *Hund-e*, *Zirkuss-e*, whereas *Radius* has stem inflection, viz. *Radi-en*. The type of inflection – i. e. the type of a morphosyntactic rule - is a structural property of the nominal inflectional system of German.

Wurzel proposes that some structural properties are system defining. They define the “quality” of the inflectional system of a language as a whole. Such system-defining structural properties correlate with

- the number and relative size of the inflectional classes in which they are realised (the number of paradigms in which they occurs)
- the extent to which they are realised in these classes (i.e. the number of forms in which they occur)

System congruity in the sense of Wurzel is "... the degree of agreement of a paradigm (and thus of an inflectional class), of a partial paradigm, of an inflectional form, or of a morphological marker with the system-defining structural properties of the language in question" (Wurzel 1989, p.80). In a language where structural properties conflict system-defining properties represent "... the unmarked variants of competing structural traits, in contrast with the residual marked variants" (Wurzel 1989, p.81). From the point of view of morphological change in the language system-congruent properties tend to be stable attractors, inducing the change and giving it direction, but being rather resistant to the change themselves.

Wurzel believes that nether system congruity nor inflectional class stability are reducible to each other. The starting observation about inflectional class stability is that it is not determined in general which inflectional rules apply to a given word (e. g. why not have *Radiuse* instead of *Radien*?). This can be learned on an item-by-item basis, of course. But there is another possibility. Some extra-morphological or morphological properties can be used to tag words to define a set of applicable rules (define an inflectional class) and thus facilitate learning: given such a class, one learned inflection would imply simultaneously learning a number of others. For example, a phonological characteristic tagging a morphological inflectional class of Russian nouns is their nominative singular ending in /a/, of the type *sobak**a*** (*dog*): they all have uniform inflection.

Such regularities have the form of implications. The individual inflectional classes are usually characterised by a series or chains of such implications. The more chains are involved in the inflectional class, the more concise can be its definition, since many of its properties can be derived.

The chains of implications defining inflectional classes are called *PARADIGM STRUCTURE CONDITIONS* by Wurzel (we shall abbreviate the term by PSC). Some paradigm structure conditions are predominant in the system where they compete with others in that they define inflectional classes with more words. Inflectional classes are stable, if they follow the pattern of the predominant paradigm structure conditions.

While these observations do pinpoint some regularities in the organisation of the morphosyntactic system of the language, they are not explanatory, but are rather in need of explanation by themselves. However, Wurzel goes on to investigate the nature of the paradigm structure conditions, and his discussion will provide the next step in our investigation.

According to Wurzel (1998), such chains of implications are to be viewed rather as defaults defined not on forms, but on markers which are mapped to forms. Thus, e. g. a very general default in German nominal paradigm specifies that a feminine noun should preferably have zero ending in the genitive singular, in a self-explaining notation [Fem] => [⊘/Gen.Sg]. Note that in this formulation we have a close resemblance of the defaults to the RC-constraints, e. g. [N]+[genSg] => [⊘]. But the primary interpretation of the default is, of course, propositional. Which is not the case with RC. RCs are not propositional implications, because an OT grammar has no inference devices to support,

e. g. modus ponens. The interpretation of the implicative constraints was specifically defined in the context of OT.

The form of the defaults induces a certain hierarchy on defaults and on the the set of infections in the paradigm. At the bottom of the hierarchy we have totally idiosyncratic inflection assignments. An example of this would be the specification for *nehmen* (*take*), requiring that the imperative singular have the vowel -*i*- instead of -*e*-. So our parametrised RCs are idiosyncratic inflexion assumptions by that characterisation.

The second layer of the hierarchy consists of inflexions which conform to very general defaults the preconditions of which refer only to features, like [Fem] => [⊘/Gen.Sg]. An example closer to our subject matter is the generalisation that the first an the third person plural of German verbs end in either -*en* or -*n.* In our case all general RCs are this kind of default.

The third layer defaults and inflections conform to rules which make reference to realisations which are assigned according to some default rule, and not merely to features. To stay close to our problem, there is the generalisation that a verb which changes its root vowel -*e*- to -*i*- in singular imperative, will change it the same way in the third person singular,

(4) (<C_e_C,C_i_C/[ImperSg]>) => (<C_e_C,C_i_C/[3perSg]>)

cf. *nehmen* - *nimm* – *nimmt*[3].

Now if we attempt to translate this into the language of constraints we face a difficulty. Since constraints are not defaults, if we try to use implicational constraints we should allow antecedents to be constraints in their turn. A constraint modelled on (4) would express that a verb which changes its root vowel in its 3persSg must show *ablaut* in the preterite, i. e. it would be something like the following constraint in our nomenclature, which does not distinguish the vowels:

(5) ([ImperSg] => vc) => ([pret] => vc)

If these are paradigm structure conditions (PSC-Constraints), the question is how to incorporate them in our account of the scales of verb classes. They need an interpretation.

6. Explaining the scales

6.1 Rendering PSCs in OT

Suppose we provide in addition to the constraints introduced so far also the constraints of the form (5) to spell out the paradigm structure conditions for our Table 3. How should the new kind of grammar work? OT grammar is not in the possession of an inference engine, so [pret] => vc cannot be formally deduced as a constraint for realisations of [pret]. And anyway, this is not the intended interpretation of the type of implicational constraints.

When a constraint like (5) is tested, first its antecedent is tested, then its consequent. But what would testing an implication in the antecedent amount to? An item with [ImperSg] should have a vowel change relative to its base. If we test the antecedent for some item which does not have [ImperSg], the implication is vacuously true and hence irrelevant. In this case we assume that the testing device – whatever it should be – knows that the

3 See Bittner (1996) for more on a treatment of verb inflection scale in Wurzel's framework.

implication is irrelevant.

Assuming that the antecedent is true, how do we proceed now with the testing of the consequent? We cannot test its implication for the same item we tested the RC of the antecedent for, because if this item contained [ImperSg] it cannot contain [pret]. And it does not improve things to begin with the consequent, since we still have the disagreement of the antecedents of the component RCs.

The idea we want to express is the same as in Wurzel's framework: if a generalisation is observed to the effect that there is a correlation between two forms, it may be used to facilitate learning by introducing a rule which applies to a number of verbs. Since learning is done by some learning algorithm, such generalisations should be expressed in terms which the algorithm can use. So the double implication constraints – PSCs – are meta-generalisations which are not used by the grammar directly, but may be formed and used by the learning algorithm. We thus introduce a totally different kind of constraint and therefore we need to provide two guidelines. The one is on how the learning algorithm learns them, the other is how it uses them. We shall start with the second guideline.

Suppose our grammar has (5) at its disposal. To regulate the behaviour of a learning algorithm we will propose the following. As we postulated for the case of realisation constraints, whenever the antecedent is not satisfied, the learning algorithm simply ignores the constraint and does not move it in any direction. The same applies to the PSC-Constraints, with one change. An RC in the antecedent of a PSC defines an inflexion type. Whenever the item does not belong to the inflection type defined by the antecedent, it is irrelevant for the operation of the learning algorithm wrt the PSC-constraint.

If the item satisfies the antecedent, the RC is promoted. The role of the PSC constraint in this case is to enforce that the RC in the consequent is promoted, too. In other words, any time an antecedent RC is promoted, all the RCs which are consequents in some PSC with this RC antecedent constraint are promoted, too. If it is demoted, noting happens to the consequent RCs.

PSC constraints are constraints, and they should be ranked with respect to the other constraints. But since their interpretation is rather operational and they cannot be considered to be satisfied or failed by a data item. They thus play no role when the grammar is used for parsing or production. Moreover, it seems that their ranking cannot directly reflect the distribution of forms. It is therefore unclear how this ranking would come about and whether it is relevant at all.

Suppose we rank them at the position of their antecedent. When its antecedent is demoted to the point where it is almost never considered, because it is dominated by a more viable realisation constraint, the corresponding PSC constraint should be largely irrelevant and not applied. In this way PSCs may degrade, leading to the restructuring of the paradigm.

Now let us turn to the first guideline. How does a learning algorithm learn PSC constraints? While this is a research question by itself, at least some speculative remarks are in order here.

6.2 PSCs and the scales

If there is an implication chain, then the consequent of the last PSC constraint profits from all the preceding elements in the chain. Viewed on the ranking scale, it is quite plausible that this RC in the long run will be ranked higher than the antecedent of the first element of the chain, and its replacement will take longer:

(6) ([ImperSg] => vc) => ([2pSgPres] => vc)

(7) ([3pSgPres] => vc) => ([2pSg Pret] => vc)

(8) ([3pSg Pret] => vc) => ([3pSgConj] => vc)

(9) ([3pSgConj] => vc) => ([PPII] => vc+-en)

(10) ([3pSgConj] => vc) => ([PPII] => zero-vc+-en)

It seems that perhaps the existence of (9) and (10) is not the desired state of things, though suggested by the data. A way out might be to condition the form of the past participle on the parametrised lexical constraints, hence to obtain e. g. either (11) or (12). We thus re-introduced lexical parametrisation as an exception handling device.

(11) ($C[nehm]_{3,12}$ => vc) => ([PPII] => vc+-en)

(12) ($C[geb]_{3,12}$=> vc) => ([PPII] => zero-vc+-en)

Actually, (12) is simply ($C[geb]_{3,12}$=> vc) => ([PPII] => -en). The simultaneous existence of the two kinds constraints implies that the ending *-en* is bound to be very resistant to the transition to the weak form, since both constraints drag it along with the strong lexical RC constraints from the antecedent. This is true of all the other lexical constraints of this kind as well, and the effect is cumulative.

These considerations suggest that there may be different factors involved in the emergence of the hierarchy. The mechanism of the PSC constraints creating the effect by itself does not explain the particular ordering on the scale, however. Note that since PSC constraints are grafted onto the general skeleton of the OT framework, they must conform to the hypothesis of the richness of the base, and we must accept the PSC constraints for all RCs. Part of the solution of why some particular ordering occurs may lie in the relative frequencies of the items conforming to RCs. But this is surely only a part of the whole picture.

Consider now a tentative scenario how the development of the German verb paradigm could proceed in the model developed so far. The initial state IS of the evolution of the grammar could look like (13). Here the constraints the relative ordering of which is irrelevant here are written in set notation, e. g. $\{C_{i,j}$ ...PSC$\}$.

(13) $\{C_{i,j}$ ...PSC$\}_{strong}$>> $\{C_{i,j}$ ...PSC$\}$$[root]_{strong}$>> $\{C_{i,j}$ ...PSC$\}$$[root]_{weak}$>> {MAX-OO}>>...

>> $\{C_{i,j}$ PSC$\}_{weak}$>>...

Since we assume that weak verbs were relatively few at IS, the ordering $\{C_{i,j}$ ...PSC$\}$$[root]_{strong}$>> $\{C_{i,j}$ ...PSC$\}$$[root]_{weak}$ seems to be justified. Remember that we assumed that PSC constraints are connected to the corresponding antecedents in their ranking.

The ordering characteristic of the current state CS would then presumably look like (14).

(14) $\{C_{i,j}$ PSC$\}_{weak}$>> $\{C_{i,j}$ ...PSC$\}$$[root]_{weak}$>> $\{C_{i,j}$ ...PSC$\}$$[root]_{strong}$ >> {MAX-OO}>>..

.>>$\{C_{i,j}$ ...PSC$\}$$[root]_{strong}$>> $\{C_{i,j}$ ...PSC$\}_{strong}$>>...

It seems that at this stage to have $\{C_{i,j}$ ...PSC}[root]$_{weak}$ is superfluous, since they would merely replicate $\{C_{i,j}$ PSC}$_{weak}$, whereas their presence was essential as long as they were exceptions. This is just one of quite a number of issues to be resolved in the extended framework, though.

The split of $\{C_{i,j}$ ...PSC}[root]$_{strong}$ during the evolution of IS to CS can be probably traced back to different frequencies of the comparable forms of different strong verbs. The general shift in which $\{C_{i,j}$ PSC}$_{weak}$ became promoted is due to the cumulative effect of the growing number of weak verbs with time. As noted, derived weak verbs had no second base to be able to use $\{C_{i,j}$ ...PSC}[root]$_{strong}$ as defaults. The corresponding constraints were irrelevant. This is then the proposed solution to Problem 3.

7. A Summary and Conclusions.

The paper sketched an extension to the standard version of StOT which allows to explain (if only in part) the emergence of implicative scales in paradigms. Three interesting properties of the example studied in this paper – the transition of the strong verb inflection to the weak (regular) inflexion - were explained by the interplay of three constraint types – lexical constraints, realisation constraints and paradigm structure conditions. Another important underlying assumption was that we used the gradual learning algorithm of StOT. But in fact, probably any learning algorithm which is sensitive to the distribution frequencies of the training data would produce similar results.

The relevant property of the individual lexical realisation constraints is their immunity towards the cumulative effect of frequency distributions. This allows them to be very resistant against re-ranking, in virtue of which they are a conservative factor in the paradigm structure. On the contrary, the sensitivity of the general realisation constraints to the cumulative effects is responsible for the general paradigm shifts. The implicative properties of both kinds of constraint were made precise by redefining them in the OT context.

The emergence of the inflection classes was shown to be in part a consequence of the paradigm structure conditions which were implemented as a new kind of constraint. Whether this kind of constraint adds new generative power to the OT framework was not investigated.

In general, paradigmatic phenomena in morphology seem to be amenable to the treatment in the extended StOT framework, which is also suitable to implement realisational morphology. Whether there is a simpler approach to these two issues remains to be seen.

8. Appendix. Definitions of OT

An OT system starts with defining a relation GEN which determines all the possible (UF,OF) combinations. It also specifies a finite universal set CON of constraints on GEN. A grammar of a particular language L, G_L, is based on a certain ordering of these constraints. Classically the constraints in G_L are linearly ordered in a ranking. The constraints in the ordering may contradict each other, but are violable. In a basic setting constraints are binary, i. e. a constraint may be either satisfied or violated (a constraint *C* acts here like a predicate). However, since constraint can be in principle violated by an (OF,UF) datum a couple of times, a constraint is better viewed as a function from GEN into the natural numbers. Thus an OT-grammar assigns a finite sequence of natural numbers to every pair in GEN – its VIOLATIONS VECTOR.

Now given two data piece with the same UF, a datum d_1 IS BETTER THAN d_2 under G_L, $d_1 \leq_{GL} d_1$, if its violation vector precedes that of d_2 under the lexicographic ordering on number strings, e. g. if the violation vector of d_1 is (1123798), and that of d_2 is (1131107), then d_1 is better than d_2. The third constraint is the decision point in this example.

An OT grammar G_L defines items from GEN as belonging to the language only if they are OPTIMAL wrt. G_L. Different notions of optimality may be involved.

Item d_1=(UF,OF) is SPEAKER OPTIMAL, if there is no better item d_2=(UF,OF'). Item d_1 =(UF,OF) is HEARER OPTIMAL, if there is no better item d_2=(UF',OF).

An item d_1=(UF,OF) is UNIDIRECTIONALLY OPTIMAL iff it is speaker optimal.

An item d_1=(UF,OF) is BIDIRECTIONALLY OPTIMAL iff

(i) there is no speaker optimal item d_2 =(UF,OF') which is also bidirectionally optimal

(ii) there is no hearer optimal item d_2=(UF',OF) which is also bidirectionally optimal.

Directionality of optimisation is a parameter of OT learning models as well as the use of the learned OT grammars.

REFERENCES

Beckman, Jill N, Lora Walsh Dickey & Suzanne Urbanczyk (eds.). 1995. Identity effects in morphological truncation. *University of Massachusetts Occasional Papers in Linguistics*.

Benua, Laura. 1995. Identity effects in morphological truncation. In: Beckman, Jill N, Lora Walsh Dickey & Suzanne Urbanczyk (eds.), pp. 77-136.

Bittner, Dagmar. 2003. Aspectual interpretation of early verb forms in German. In: Bittner, Dagmar & Natalia Gagarina (eds.), pp. 3-25.

Bittner, Dagmar & Natalia Gagarina (eds.). *Acquisition of aspect*. ....ZAS Papers in Linguistics, 29, Berlin.

Bittner, Andreas. 1996. *Starke "schwache" Verben, schwache "starke" Verben*. Tübingen: Stauffenburg Verlag.

Blutner, Reinhard & Henk Zeevat (eds.). 2003. *Pragmatics in OT*. Basingstoke: Palgrave MacMillan.

Boersma, Paul. 1998. *Functional phonology: formalizing the interactions between articulatory and perceptual drives*. Ph.D. dissertation, University of Amsterdam.

Bybee, Joan & Jean E. Newman. 2007. Are stem changes as natural as affixes?. In Bybee, Joan (ed.), Ch. 7, pp. 148-166.

Bybee, Joan. 2007. *Frequency of use and the organization of language*, Oxford: Oxford University Press.

Cangelosi, Angelo & Domenico Parisi (eds.). 2002. *Simulating the evolution of language*. London: Springer.

Coetsem, Frans van. 1993. A*blaut* and reduplication in the Germanic verb, Heidelberg: Winter Verlag.

Fabri, Ray, Albert Ortman & Teresa Parodi (eds.). 1998. *Models of inflection*. Tübingen: Max Niemeyer Verlag.

Guttman, Louis. 1944. A basis for scaling qualitative data. *Americal Sociological Review* 9, pp. 139-150.

Jäger, Gerhard. 2003. Learning constraint sub-hierarchies. The bidirectional gradual learning algorithm. In: Blutner, Reinhard & Henk Zeevat (eds.), pp. 251-287.

Kirby, Simon & Jim R Hurford. 2002. The emergence of linguistic structure: An overview of the iterated learning model. In: Cangelosi, Angelo & Domenico Parisi (eds.), p. 121--148.

Lapoint, Steven G., Diane K. Brentari & Patrick Farrell (eds.). 1998. Stanford, CA: CSLI Publications.

MacBride, Alexander Ian. 2004. *A constraint-based approach to morphology*. Ph.D. dissertation, UCLA.

Matthews, Peter H. 1991. *Morphology*, 2nd edn, Cambridge: Cambringe University Press.

Prince, Alan & Paul Smolensky. 2004. Optimality theory: constraint interaction in generative grammar, Malden, MA: Blackwell Publishing.

Russell, Kevin. 1999. Mot: sketch of an ot approach to morphology. [ROA-352], Rutgers Optimality Archive, http://roa.rutgers.edu/.

Scheffler, Tatjana, Joshua Tauberer, Aviad Eilam & Laia Mayol. (eds.). 2007. Proceedings of the 30th annual Penn linguistics colloquium. *University of Pennsilvania Papers in Linguistics*.

Stump, Gregory T. 2001. *Inflectional morphology: a theory of paradigm structure*, Cambridge: Cambridge University Press.

Watkins, Calvert (ed.). 2000. *The American heritage dictionary of Indo-European roots*. Boston, New York: Houghton Mifflin.

Wurzel, Wolfgang Ullrich. 1989. *Inflectional morphology and naturalness*, Dordrecht: Academic Press and Kluwer.

Wurzel, Wolfgang Ullrich. 1998. Drei ebenen der Struktur von Flexionsparadigmen. In: Fabri, Ray, Albert Ortman & Teresa Parodi (eds.), pp. 225-243.

Yip, Moira. 1998. Identity avoidance in phonology and morphology. In: Lapoint, Steven G., Diane K. Brentari & Patrick Farrell (eds.), pp. 216-246.

Xu, Zheng. 2007. An optimality-theoretic account of full and partial identity of forms. In: Scheffler, Tatjana, Joshua Tauberer, Aviad Eilam & Laia Mayol. (eds.), pp. 421-434.

Author's address: Dr. Anatoli Strigin, Centre for General Linguistics, Typology and Universals Research (ZAS), Schützenstrasse 18, 10117 Berlin, Germany
E-mail: strigin@zas.gwz-berlin.de