Degraded Acceptability and Markedness in Syntax, and the Stochastic Interpretation of Optimality Theory*

Ralf Vogel, University of Potsdam

draft, 5th May 2004, to appear in Fanselow, Féry, Schlesewsky & Vogel (eds.), *Gradience in Grammar*, OUP, 2005, comments very welcome

1 Gradience and categoricity in generative syntax

Conceiving *grammaticality* as *gradient* poses problems for those traditional conceptions of grammar which assume that linguistic expressions can only be either grammatical or ungrammatical. That a sentence is, for instance, "grammatical to 75%" is a nonsensical statement from this point of view. In this tradition, generative grammar assumes the native speaker's linguistic competence to be the system of rules with which, among other things, she classifies sentences as either grammatical or ungrammatical. A linguistic theory is explanatorily adequate to the extent that it successfully models this underlying knowledge.

As a consequence of the categorical view on grammaticality, generative models rarely reflect gradience. The minimalist program (Chomsky, 1995) for generative syntax knows "converging" and "crashing" derivations, which conform to grammatical and ungrammatical structures. But

*I want to thank my collaborators Jutta Boethke, Stefan Frisch and Marco Zugck, without whom the empirical research presented in this paper would not have been undertaken. For fruitful comments and helpful suggestions I further thank Gisbert Fanselow, Caroline Féry, Doug Saddy, Joanna Blaszczak, Arthur Stepanov, and the audiences of several presentations of parts of this work at the Potsdam University. This work has been supported by a grant from the *Deutsche Forschungsgemeinschaft* for the research group "Conflicting Rules in Language and Cognition", FOR-375/2-A3.

there is no way of talking about slightly ungrammatical or nearly perfect structures in terms of derivation. Measures for gradience could be added in principle – for instance, one could state that acceptability decreases with the number of derivational steps. However, such a program can hardly be applied successfully, as the number of derivational steps often only reflects an author's assumptions about syntactic structures in general – for instance, the number of universal functional categories that one assumes.

A popular strategy is attributing gradience to "external" factors. A version of this argument has been given by Chomsky himself already in his early work (Chomsky, 1964, 1965): We can distinguish *degrees of grammaticalness* by the type of grammatical rule that is violated. A violation of a semantic selectional rule causes weaker ungrammaticality than a violation of a syntactic subcategorisation rule.

However, the "?" need not necessarily be interpreted in such a way. Linguists regularly experience gradience in data when they contact a larger amount of informants. That two native speakers fully agree on all expressions of their language is more the exception than the rule. The examples that grammarians are interested in are often quite complicated structures, and disagreement on such structures is even more likely.

For that reason it has become common practice to let informants give more fine-grained judgements along a scale. The resulting subtle differences among structures, however, are often ignored by the grammarian. The bounding line between 'grammatical' and 'ungrammatical' is located somewhere on this scale, and so the multi-valued scale is mapped onto the binary distinction that a categorical grammar can deal with.

To give an example, Josefsson (2003) reports a survey that she did with about 30 Swedish native speakers on the possibility of pronominal object shift in Swedish. She gave her informants a five-point scale with the values "o.k. – ? – ?? – ?* – *". For the statistical analysis, she correlated the judgements with natural numbers ranging from "o.k."= 4 to "*"= 0. Josefsson further assumed that grammatical sentences have at least an average acceptability value of 1.5. This decision is not of particular importance in her analysis. However, it appears to be a purely normative decision. She could as well have proposed 2.0 or 2.5 as the boundary. How can such a decision be justified independently?

It is not at all clear how such scales are to be treated, and whether they are really scales. A number of issues need clarification:

- What is the 'meaning' of the question mark? Is it

1. "I am uncertain about the grammaticality status of this sentence", or is it
2. "This sentence has degraded grammaticality"?, or is it
3. "I don't like this sentence for some reason"

- Given that each of these answers is possible: How can we distinguish these three kinds of judgements empirically?
- What does it mean for a sentence to be exactly in the middle between 'grammatical' and 'ungrammatical'? Is this a category of its own?
- Is it legitimate to treat acceptability judgements as interval scaled variables, rather than ordinal, or even nominal variables, in the statistical analysis?
- As soon as more than one informant is involved: how can we ensure that all of them use the given scale in the same way? – this is crucial for the question of which point in the scale divides between grammatical and ungrammatical.
- What is the ontological status of the scale? Does it exist independently, and are "?", "??" and "?*" therefore *categories of grammaticality*? Or are they only used to *mark acceptability contrasts* between two sentences?

An answer to these questions requires a theory of acceptability judgements. Theoretical linguists rarely explicate their point of view on this. Interpreting the '?' as uncertainty could simplify the problems somewhat, as we could still assume a categorical grammar.

But we would still have to exclude that the gradience that we observe results from inherent properties of the grammar, instead of being the result of 'random noise'. If, on the other hand, phenomena of gradience are systematically correlated with grammatical properties, then the whole categorical view on grammar is called into question. I think that this is indeed the case. And I will show this with a case study that I present step by step in this article.

More recent variants of "explanations" in terms of non-grammatical factors attribute variation and gradience in grammaticality judgements to "performance". Abney (1996) already remarked that such a line of argumentation takes the division between competence and performance more seriously than it should be taken:

> "[...] Dividing the human language capacity into grammar and processor is only a manner of speaking, a way of dividing things up for theoretical convenience. It is naive to expect the logical grammar/processor division to correspond to any meaningful physiological division – say, two physically separate neuronal assemblies, one functioning as a store of grammar rules and the other as an active device that accesses the grammar-rule store in the course of its operation. And even if we *did* believe in a physiological division between grammar and processor, we have no evidence at all to support that belief; it is not a distinction with any empirical content. [...]" (Abney, 1996)

Gradedness can even be used as a criterion for determining whether a constraint belongs to competence or performance: Constraints that belong to performance cause degraded acceptability, rather than plain ungrammaticality. This would immunise the competence/performance distinction against any empirical counter-evidence, which would make even clearer that the distinction is only made for theoretical convenience.

Manning (2003) argues along very much the same lines. Emphasising, like Abney, that the generative grammarian discourse centred around the notion of competence is very limited in its scope, he calls for the application of probabilistic methods in syntax:

> "[...] Formal linguistics has traditionally equated structuredness with homogeneity [...], and it has tried too hard to maintain categoricity by such devices as appeal to an idealised speaker/hearer. [...] The motivation for probabilistic models in syntax comes from two sides:
>
> - Categorical linguistic theories claim too much. They place a hard categorical boundary of grammaticality where really there is a fuzzy edge, determined by many conflicting constraints and issues of conventionality versus human creativity. [...]
> - Categorical linguistic theories explain too little. They say nothing at all about the soft constraints that explain how people choose to say things (or how they choose to understand them).
>
> [...]"

Sternefeld (2001) provides further detailed argumentation against the traditional competence/performance distinction within generative grammar. One of his topics is that structures with central embeddings are very often degraded in their acceptability. The problem for the competence/performance distinction is: Why should the computational system of I-language, the competence system, be able to produce structures that the parser is unable to compute efficiently? Why is the parser unable to use the computational system of I-language in processing?

With Kolb (1997), Sternefeld claims that the description of I-language as "computational system" makes it impossible to distinguish it theoretically and empirically from the "processing system", i.e., performance. Both have the same ontological status as generative, procedural systems. Therefore, Sternefeld proposes with Kolb that competence should be understood as a *declarative axiomatic system*, comparable to formal logics. Computational procedures, however abstract they may be conceived, are then part of the performance system. A derivation can be seen as a proof for a particular structure, interpreted as a theorem of the algebraic system of I-language. The performance system, however, includes not only these derivational procedures, but also, for instance, all the psychological restrictions that are known to influence linguistic behaviour, and anything else that is usually subsumed under the term "performance".

A research program that restricts itself to an investigation of competence in this sense would not be able to formulate anything of empirical relevance. In other words: The linguists' focus of interest is and has always been performance. Abney, Manning and Sternefeld each argue from different perspectives for abandoning the competence/performance distinction in the traditional sense. In particular, they all show that it is useless for the investigation of a number of empirical phenomena, including gradient acceptability.

Perhaps, the difficulty to relate the numerical, statistical results of psycholinguistic experiments, corpus studies or other more advanced empirical methods to a categorical understanding of grammaticality is the reason why the results of such empirical studies only rarely find their way into the grammar theoretical work of generative syntacticians.

On the other hand, an acceptability judgement is a very artificial task, even if done by a linguist who is used to it, or especially in that case. Furthermore, we cannot tell to what extent judgements are influenced by purely *normative* preferences. A grammaticality judgement can even be interpreted as experimentally enforced normative action.

There is no "innocent" empirical method as such. No empirical method gives us direct insights into the "pure grammar". It would therefore be

naïve to trust only in one single method. A theory has a more solid foundation if it has been verified for different empirical domains. Effects that can be observed in psycholinguistic experiments should also be observable in corpus studies and vice versa. The empirical predictions of a theory of grammar should not be specific to only one empirical method or domain.

2 Markedness in Syntax

An important feature of many empirical methods is their *relational* way of achieving data about syntactic structures. A typical design for a psycholinguistic experiment uses minimal pairs. An example is the pair in (1): Free relative clauses (FR) are clauses that stand for non-clausal constituents. They have the syntax of relative clauses, but miss a head noun. The initial *wh*-pronouns of argument FRs are sensitive to the case requirements of both the FR-internal verb and the matrix verb. When the two cases differ, we observe a conflict: one of the two cases cannot be realised. This leads to ungrammaticality in (1-b):[1]

(1) Case matching in argument free relative clauses in German:

a. Wer uns hilft, wird uns vertrauen
[who-NOM us-DAT helps]-*NOM* will us-DAT trust
"Whoever helps us will trust us"

b. *Wer uns hilft, werden wir vertrauen
[who-NOM us-DAT helps]-*DAT* will we-NOM trust
"Whoever helps us, we will trust him"

Experiments usually test for *contrasts* between minimally different expressions. In our example, the theory of case matching in argument free relative clauses (Groos and van Riemsdijk, 1981; Pittner, 1991; Vogel, 2001) is confirmed if (1-b) is judged as grammatical less often than (1-a) to a statistically significant degree. This is indeed the result of a speeded grammaticality judgement experiment by Boethke (in preparation). Structure (1-b) has significantly less often been judged as acceptable than (1-a).[2]

[1]The case required by the matrix verb appears slanted and attached to the FR in the glosses.

[2]For sake of completeness, a brief description of the esperiment design: Each of the 24 participants – students of the University of Potsdam – saw eight items of each of the conditions. Test items were FRs with the four possible case patterns with nominative and dative. The experiment included four further conditions which will be introduced later – so the experiment included eight conditions altogether. The test items of this experiment have been randomised and mixed with the test items of three other experiments

This result is unproblematic for a categorical grammar. However, the experiment included two further conditions:

(2) a. Wem wir helfen, werden wir vertrauen
[who-DAT we-NOM help]-*DAT* will we-NOM trust
"Whoever we help, we will trust him"
b. Wem wir helfen, wird uns vertrauen
[who-DAT we-NOM help]-*NOM* will us-DAT trust
"Whoever we help, he will trust us"

The acceptability rates for these two structures are between those for the two structures in (1). All contrasts except for the one between (2-a,b) were statistically significant:

case of *wh*-phrase:	case required by matrix verb: NOMINATIVE	DATIVE
NOMINATIVE	87% (1-a)	17% (1-b)
DATIVE	62% (2-b)	71% (2-a)

Table 1: Acceptability rates for the structures (1) and (2) in the experiment by Boethke (in preparation)

A categorical grammar here faces the problem of mapping this observation to its dichotomous grammaticality scale. How can we independently justify where we draw the boundary? If we state that (2-b) is ungrammatical, then we state that the observed contrast between (2-b) and (1-b) is crucial, but all the others are not. Likewise, if we treat (2-b) as grammatical, we ignore the contrast between (1-a) and (2-b). No matter how we decide, the difficult task is finding arguments for our decision to ignore some contrasts while using others. But most impor-

which served as distractor items. The sentences have been presented visually word by word on a computer screen, one word at a time, each word was presented for 400 ms. Subjects were asked to give a grammaticality judgement by pressing one of two buttons for grammatical/ungrammatical, within a time window of 2500 ms.
I want to emphasise that this experiment lead to gradient acceptability (see below) without asking for it. In questionnaire studies with multi-valued scales and magnitude estimation based experiments gradience is already part of the experimental design. One could argue that the subjects only give gradient judgements, because they are offered this option. In the experiment described here, the gradience results from intra- and inter-speaker variation in repeated measuring. Subjects found different structures differently often acceptable.

tantly, there is *no* way of accounting for *all* contrasts with the grammatical/ungrammatical dichotomy only.

This shows that the decision between a categorical or a gradient conception of grammaticality is also an empirical matter.[3] If empirical methods show an intermediate acceptability status under such controlled conditions, it is very likely that the factor that caused this intermediate status is grammar-internal. At least, this should be the null assumption.

A theory of grammar that has the potential to deal with gradience more successfully is *Optimality Theory* (Prince and Smolensky, 1993). It departs in a number of ways from classical generative grammar. It is *constraint-based*, which is not strikingly different, but the constraints are *ranked* and *violable*. Different structures have different *violation profiles*.

One important departure from traditional grammars is that the grammaticality of an expression cannot be determined for that expression in isolation. An expression is grammatical, if it is *optimal*. And it is optimal if it performs better on the constraint hierarchy than all possible alternative expressions in a competition for the expression of a particular underlying input form.

OT thus determines grammaticality in a relational manner. This is reminiscent of what is done in the empirical investigations described above. It should therefore be possible to systematically relate empirical gradience to relative optimality of violation profiles.[4]

OT is based on two types of constraints, *markedness* and *faithfulness* constraints. Markedness constraints evaluate intrinsic properties of candidates, while faithfulness constraints evaluate how similar candidates are to a given input. As there are infinitely many possible input specifications, there is the same rich amount of competitions. Grammatical expressions are those that win, i.e., are optimal, in at least one of these competitions.

Candidates which are good at markedness, i.e. relatively unmarked candidates, are not as much dependent on the assistance of faithfulness constraints than relatively marked candidates. This is schematically illustrated in the tableaux in (3).

(3) a. *Grammar with low ranked faithfulness*:

[3]Featherston (to appear) provides more arguments in favour of this position.

[4]The first author who explored this feature of OT systematically, was Frank Keller (Keller, 2000, and further work). See below for a brief discussion of his approach.

cand1	M1	M2	F
☞*cand1*		∗	
cand2	∗!		∗

cand2	M1	M2	F
☞*cand1*		∗	∗
cand2	∗!		

b. *Grammar with highly ranked faithfulness*:

cand1	F	M1	M2
☞*cand1*			∗
cand2	∗!	∗	

cand2	F	M1	M2
cand1	∗!		∗
☞*cand2*		∗	

(M1, M2: markedness constraints; F: faithfulness constraint; cand1, cand2: input specifications; *cand1, cand2*: output candidates; ∗ = constraint violation; ∗! = fatal violation; ☞= winning candidate)

Candidate *cand1* performs better than *cand2* in the hierarchy of markedness constraints "M1 ≫ M2". Therefore, we can say that *cand1* is less marked than *cand2* in the language at issue. This does not tell us anything about the grammaticality of *cand2*, however. But we know that *cand1* is grammatical, irrespective of the grammaticality of *cand2* – provided, as we assume for the sake of the example, that there are no further constraints and candidates to consider. The faithfulness constraint F, if ranked low, cannot assist candidate *cand2*, and so *cand1* wins the competitions for both inputs 'cand1' and 'cand2'. Highly ranked faithfulness gives higher priority to input preservation, and therefore *cand2* wins its own competition.

Irrespective of the fact that both *cand1* and *cand2* are grammatical under highly ranked faithfulness, we can still derive that *cand2* is the more marked structure from the violation profiles of the two structures, when we abstract away from particular inputs, i.e., leave out the faithfulness constraints.

An OT grammar, interpreted this way, not only tells whether a structure is grammatical, it also determines its relative markedness compared to other structures. This second property is particularly interesting for the predictability of gradience. Markedness can be seen as the correlate of gradience within the OT grammar. Because markedness is one of the key concepts of OT, nothing substantial needs to be added to account for gradience. In our abstract example, the prediction would be that the less marked *cand1* receives higher acceptability, is easier to process, is more frequent etc.

Let me illustrate this with the real-language example in (1) and (2). Simplifying my own account (Vogel, 2001, 2002, 2003*b*), we can assume the following constraints to distinguish the four structures:

(4) **Realise Case (RC):** An assigned case must be realised morphologically.
Realise Oblique (RO): An assigned oblique case (e.g., dative) must be realised morphologically.
S<O: The subject precedes the object(s) of its clause.

Under the ranking illustrated in Table 2, the violation profiles are such that the relative markedness of the candidates matches the relative acceptabilities given in Table 1.[5]

rank	*candidate*	RO	RC	S<O
1.	(1-a)			
2.	(2-a)			*
3.	(2-b)		*	
4.	(1-b)	*	*	*

Table 2: Comparison for the relative markedness of the structures in (1) and (2)

Groos and van Riemsdijk (1981); Pittner (1991) and Vogel (2001) offer three different approaches to case conflicts in German FRs. Interestingly, these authors also differ in the grammaticality judgements they present. In particular, they agree that the two patterns in (5) are grammatical. (5-a) is a so-called 'matching' FR, both verbs assign the same

[5]For the relative markedness of these four candidates the constraint RO is unnecessary. That it is the constraint that excludes candidate (1-b) can be seen from the fact that it does not improve, if the FR is postponed, which avoids a violation of S<O:

(i) *Wir werden vertrauen, wer uns hilft.
we-NOM will trust [who-NOM us-DAT helps]-*DAT*
"We will trust whoever helps us"

Note also that, strictly speaking, we have no evidence for the contrast between (2-a) and (2-b), because their acceptability rates (71% vs. 62%) did not differ to a statistically significant degree. If we interpret this result such that the two structures are equally marked, then RC and S<O would have to be ranked on a par in order to mirror this in our model.

case, accusative, no conflict arises. In (5-b), two different cases are assigned, nominative and accusative, but the *wh*-pronoun '*was*' is ambiguous for these two cases, so at the FR is matching at the surface, and this is obviously sufficient.

(5) a. Ich lade ein, wen ich treffe
I invite [who-ACC I-NOM meet]-*ACC*
"I'll invite whoever I meet"
b. Ich kaufe was mir gefällt
I buy [what-NOM me-DAT pleases]- *ACC*
"I'll buy whatever pleases me"

While (6-a,b) are both grammatical in Vogel's (2001) dialect "German A", Pittner (1991) only classifies (6-a) as grammatical (Vogel's (2001) "German B"). Both patterns in (6) are classified as ungrammatical by Groos and van Riemsdijk (1981) (Vogel's (2001) "German C").[6]

(6) a. Ich lade ein, wem ich begegne
I invite [who-DAT I-NOM meet]- *ACC*
"I'll invite whoever I meet"
b. Ich lade ein, wer mir begegnet
I invite [who-NOM me-DAT meets]- *ACC*
"I'll invite whoever meets me"

Pittner (1991) offers an explanation for the contrast she sees in (6) in terms of the case hierarchy "nominative $<$ accusative $<$ dative, genitive, PP": a case may only be suppressed in favour of another case that is higher on the case hierarchy, in particular, accusative can be suppressed in favour of dative, but not in favour of nominative. In (Vogel, 2001), I capture this with the following OT constraint:

(7) **Realise Case (relativised)** (RCr):
An assigned case must be realised morphologically by its case morphology or that of a case that is higher on the case hierarchy.

I assume a further constraint that we may informally call "1to1" here (cf. Vogel, 2001):

(8) **1to1**:
Case assigner and case assignees are in 1-to-1 correspondence.

[6]Note that (5-b) and (6-b) do not differ in the case conflict configuration. The *wh*-pronoun '*was*' is ambiguous for nominative and accusative. It is therefore the correct realisation for both of these cases, and the case conflict is, obviously, resolved.

High rank of this constraint has the effect that FRs are disallowed, and lose against an unfaithful candidate. In German, this unfaithful winner is a structure that I call "correlative" (CORR):

(9) Wer uns hilft, dem werden wir vertrauen
Who-NOM us-DAT helps that one-DAT will we-NOM trust

Here, the case conflict is avoided by the insertion of an additional resumptive pronoun ("*dem*"). The differences between the judgements given in the three papers can be described in terms of OT grammars that use the same hierarchy of markedness constraints, and differ only in the rank of faithfulness (see Table 3).

RO $\gg$ **F** $\gg$ RCr $\gg$ RC $\gg$ 1to1	Vogel (2001), German A
RO $\gg$ RCr $\gg$ **F** $\gg$ RC $\gg$ 1to1	Pittner (1991)
RO $\gg$ RCr $\gg$ RC $\gg$ **F** $\gg$ 1to1	Groos and van Riemsdijk (1981)

Table 3: Different rankings of faithfulness based on identical rankings of markedness yield three variants of German reported in the literature, which only differ in their 'tolerance'

If the rank of **F** is not absolutely determined, but allowed to vary between RO and 1to1, then there is no need to assume that varying judgements result from different grammars, as long as variation and gradience are based on the same hierarchy of markedness constraints. Faithfulness can be interpreted as a '*floating constraint*' in the sense of Reynolds (1994) and Nagy and Reynolds (1997). Floating constraints are ranked within a particular range in the constraint hierarchy. They are exceptional, constraints in general do not float. The motivation for the introduction of floating constraints is the same kind of problem observed here, variation within a speech community, or a family of closely related dialects.

In our example, the variable rank of F can be interpreted to mirror the individually varying level of "error tolerance" within a speech community.

In Stochastic Optimality Theory, *all* constraints occupy a particular rank only with a certain probability, and can potentially be floating. It therefore provides even more flexibility to make adequate predictions for empirical investigations. This approach is the topic of section 4.

3 Markedness in OT

Markedness shows up in two ways in an OT model, and these need to be distinguished. First, an OT comparison only of structures that are optimal

in some competition, as sketched in Table 2, results in a relative markedness ranking only of grammatical structures. This is all we need to predict the relative frequencies of different structures in a corpus, as ungrammatical structures do not occur in even very large corpora to a measurable degree. It can be used in the same way to predict relative acceptabilities in experiments. That an OT markedness grammar outputs the correct relative acceptabilities/frequencies in such a comparison could be a criterion for the empirical adequacy of a model.

Second, suboptimal structures, the losers of single OT competitions, are more marked than the winners. Many of these output candidates do not win in any competition of a language. Of course, it is the nature of OT that suboptimal structures also differ in their violation profiles. Keller (2000) relates this conception of suboptimality to degraded acceptability.

3.1 Relative Markedness of Winners

An understanding of markedness in the first sense underlies the common sense usage of this term in the linguistic literature. Expressions are classified as grammatical, ungrammatical, and 'marked', which usually means, informally speaking, "not ungrammatical, but not perfect either", but rarely ever "ungrammatical, but better than other ungrammatical structures".[7]

The possibility of a comparison of all winners in the OT grammar of a particular language with respect to their relative markedness is an important feature that distinguishes OT from ordinary models of generative grammar. There, all grammatical structures are equal in the sense that the only criterion for grammaticality is the possibility of assigning them a well-formed structural description. This notion of 'well-formedness' is not abandoned in OT. All winning structures of single competitions are well-formed in this sense. But these winning structures are not equal. They are assigned different violation profiles by the OT grammar, and these, in principle, are accessible for comparison. The result of such a comparison is a scale of relative markedness which should ideally conform to the gradience that we observe.

The first one who exploited this idea to account for gradience in syntax, as far as I know, was Müller (1999). However, much of the crucial work in his proposal is done by a subsystem of constraints which works

[7] To me, this formulation even has the flavour of a logical contradiction. ungrammatical structures can by definition not be better than other structures.

differently in accounting for grammaticality than in accounting for degraded acceptability. That is, he uses slightly different grammars for the two tasks.

This is not the case in the proposal that I developed above. I only use the constraint types that are already there, markedness and faithfulness constraints. Faithfulness plays a crucial role in selecting the winners of single competitions, but cannot, by definition, play a role in the relative comparison of these winners, as they are winners for different inputs. Müller, on the contrary, selects the constraints that are responsible for gradience in an ad hoc manner from the set of markedness constraints.

In a similar vein, Keller (2001) and Büring (2001) propose differences among markedness constraints. Roughly speaking, they should be distinguished by the effect of their violation. Irrespective of their rank in the constraint hierarchy, markedness constraints are claimed to differ in whether their violation leads to ungrammaticality or only to degraded acceptability. These three authors have in common that they propose that markedness in the traditional sense must be added to the OT model as a further dimension of constraint violation. They did not find a way of accounting for it within standard OT. This is surprising insofar as the traditional conception of markedness is the core of OT. However, I think that I showed a way out of this dilemma that can do without these complications.

3.2 Markedness as Suboptimality

Markedness in the second sense that I mentioned above, as an artefact of the OT model, is a much more problematic concept, and one might wonder whether it has or should have any empirical consequences. A single OT competition only knows winners and losers. Müller (1999) already argues against the conception of suboptimality proposed by Frank Keller (cf. Keller, 2000): In many OT analyses, the second best candidate is simply the candidate that is excluded last, and very often this candidate is plainly ungrammatical and much worse than other candidates which have been excluded earlier.

Take the case of a candidate $cand1$ that is excluded early in competition A only because of highly ranked faithfulness, but wins another competition B that has the appropriate input. Such a structure would certainly be judged better in an experiment than a candidate $cand2$ that is excluded lately in both competitions, but does not win any competition in the language at hand, and is therefore ungrammatical. Because of faithfulness,

structures are assigned different violation profiles in different competitions. Being suboptimal in one competition does not mean being suboptimal in all competitions. Consider our abstract example in (3). Under highly ranked faithfulness, $cand1$ is better than $cand2$ in one competition, and in the other competition it is the other way around. On which of these two competitions shall we base or empirical predictions?

A competition in an OT model is a purely technical device which shall not be identified with a comparison task in a psycholinguistic experiment. The only possible way to derive empirical predictions from a standard OT model also for the comparison of ungrammatical structures seems to me to be the meta-comparison for markedness sketched above that abstracts away from single competitions, and therefore from faithfulness.

A powerful enhancement of OT that tries to relate grammar theory and empirical linguistics is *Stochastic Optimality Theory* which will be discussed in the next section.

4 Stochastic Optimality Theory – how to make grammar fit observations

The foundations for Stochastic OT have been laid down by Boersma (1998) and Boersma and Hayes (2001). The most important difference to classical OT is that constraints are ordered at an infinite numerical scale of "*strictness*". The relative rank of constraints is expressed by their distance on this scale, rather than simply by domination. The "rank" of a constraint, furthermore, is not a fixed value, but a probabilistic distribution. A constraint has a particular rank only with a particular probability. At evaluation time, a certain amount of noise is added, the probabilistic distributions of two constraints might overlap and the grammar can have different rankings at different times, though these rankings might differ in their probabilities.

The body of work of stochastic OT in syntax is still rather small, and most of it has been carried out by Joan Bresnan and her group at Stanford University. Let me introduce only one example. Bresnan & *al.* (2001) study the influence of person features of agent and patient on the choice of voice in English and Lummi. They analysed the parsed SWITCHBOARD corpus, a database of spontaneous telephone conversations spoken by over 500 American English speakers. The analysis revealed the absence of full passive (with *by*-phrases) if the agent of the transitive verb is first or second person, while they found an albeit small number of full passives

with third person agents. This difference, though numerically small, is statistically significant. Table 4 displays their figures.

action:			#Active:	#Passive:	% Active:	% Passive:
1,2	→	1,2	179	0	100.0	0.0
1,2	→	3	6246	0	100.0	0.0
3	→	3	3110	39	98.8	1.2
3	→	1,2	472	14	97.1	2.9

Table 4: English person/role by voice (full passives) in the parsed SWITCHBOARD corpus, from Bresnan & *al.* (2001)

English exhibits as a tendency what a language like Lummi has as a categorical rule: Passives are avoided for structures with first and second person agents. and they are more likely to occur with first and second person patients than with third person patients.[8] Observations of this sort are evidence for a position that unites functional and formal linguistics within Optimality Theory under the slogan of the "stochastic generalisation":

> "[...]The same categorical phenomena which are attributed to hard grammatical constraints in some language continue to show up as statistical preferences in other languages, motivating a grammatical model that can account for soft constraints. [...]" Bresnan & *al.* (2001)

Bresnan & *al.* (2001) show that Stochastic OT "... can provide an explicit and unifying theoretical framework for these phenomena in syntax." The frequencies of active and passive are interpreted to correspond to the probabilities of being the optimal output in a stochastic OT evaluation.

The most important constraints that are used in that account are $*\mathrm{Obl}_{1,2}$, which is ranked highest and bans *by*-phrases with first and second person, $*\mathrm{S}_{Pt}$, which bans patients from being subjects, i.e., penalises passives, $*\mathrm{S}_3$, which penalises 3rd person subjects, and $*\mathrm{S}_{Ag}$, which penalises Agents as subjects. The latter two constraints are ranked on a par and overlap a bit with the higher ranked $*\mathrm{S}_{Pt}$, which in turn overlaps a bit with the higher ranked $*\mathrm{Obl}_{1,2}$.

The rarity of passives is mirrored by the high rank of $*\mathrm{Obl}_{1,2}$. Is it really the case that the rarity of passives with first and second person

[8] In Lummi, sentences with first or second person objects and third person subjects are ungrammatical. Likewise, passive is excluded if the agent is first or second person.

by-phrases is the result of a grammatical constraint, or is it not rather the result of the rarity of the communicative situation in which such a passive would be appropriate? Not all instances of infrequency have a grammatical cause. It seems that a constraint system that is designed to directly derive frequency patterns runs into the danger of interpreting properties of the "world" as properties of the grammar. I will discuss this problem in more detail below.[9]

5 A case study – continued

One problem for Stochastic Optimality Theory that has often been noticed (cf. Boersma, this volume) is that different tasks seem to require different stochastic OT grammars. In particular, corpus frequencies and relative acceptabilities might not always go hand in hand.

The studies on German argument free relative clauses that have been introduced briefly above already are another example in case. The experiment by Boethke (in preparation) altogether included eight conditions, four FRs in the four different case configurations, and their correlative counterparts.[10]

NOM-NOM		DAT-DAT		DAT-NOM		NOM-DAT	
FR	CORR	FR	CORR	FR	CORR	FR	CORR
87	95	71	91	62	92	17	90

Table 5: Mean acceptabilities for FR and CORR in different case configurations in %

One prediction is that the correlative structures have a higher acceptability rate than their FR counterparts, because they avoid any case realisation problems by the additional resumptive pronoun. This expectation is met, as can be seen in Table 5. All contrasts are statistically significant, except for the least problematic context, NOM-NOM. This could be due to the fact that FRs in this context are too good already, and so there might in fact be a difference, but it cannot be detected with this method of measuring.

Secondly, the contrast between the FRs in the contexts DAT-DAT and DAT-NOM was not significant either, contrary to all other contrasts. This

[9] See also (Boersma, this volume) for more discussion of problems of this kind.

[10] The abbreviations for the case patterns here and below have the following logic: in "CASE1-CASE2", CASE1 is the case of the *wh*-pronoun, CASE2 is the case assigned to the FR by the matrix verb.

is perhaps due to an equal rank of the constraints RC and S<O. Both of these seem to be minor problems.

However, we also carried out a corpus study on the same structures, and this study yielded different results precisely in these two problematic cases (Vogel and Zugck, 2003). We used the "COSMAS II" corpus of written German of the Institut für Deutsche Sprache (IDS) Mannheim. Samples of 500 randomly chosen sentences containing the *wh*-pronouns "*wer*" and "*wem*" have been generated, the FR usages of these instances have been sorted out and counted. The results which are relevant for our discussion are shown in Table 6.

Case pattern	FR	CORR
NOM-NOM	274 (89.8%)	31 (10.2%)
DAT-DAT	1 (5.6%)	17 (94.4%)
DAT-NOM	33 (34.4%)	63 (65.6%)
NOM-DAT	0 (0%)	5 (100%)

Table 6: Results of a corpus investigation (Vogel and Zugck, 2003)

About 90% of the instances in the NOM-NOM context are FRs. This is remarkably different from the result of the acceptability judgement experiment where CORR had a slightly higher rate, but the difference to FR was not significant. We therefore would have expected equal frequencies for the two structures at best, but not such a high preference for the more marked FR. The second difference concerns the contrast between the DAT-DAT and the DAT-NOM context: FR is used significantly less often in the DAT-DAT context. In the experiment, these FRs have a higher acceptability rate, though this was again not statistically significant. A formulation of this problem in terms of standard OT requires the following steps. First, we need a new constraint:

(10) **Avoid Redundancy (*Red)**: Avoid meaningless elements that have a purely grammatical purpose (so-called "function words").

This constraint penalises the CORR structures because of the additional resumptive pronoun, a pure function word without contribution to the meaning of the clause. Typologically, the inclusion of this constraint predicts the existence of languages that have FRs, but no CORRs – this has explicitly been denied by me in (Vogel, 2001). But this is an empirical matter, and such languages might be found in the future.

Depending on how we interpret the results, *Red must either be ranked lower than 1to1 in grammaticality judgements, because FRs are judged as worse than CORR in the experiment, or equal with 1to1, because this tendency was not significant for the NOM-NOM context. For ease of presentation, we deliberately decide to give a clear ranking, and assume that the observed contrast was only accidentally not significant. Because FR is only more frequent in the NOM-NOM context, the effects of *Red must be restricted to that context. We do this by adding a constraint conjunction of 1to1 and S<O, "1to1&S<O":

(11) **1to1 & S<O:** No simultaneous violation of 1to1 and S<O.

This constraint should be ranked on on top of 1to1 in order to take effect independent of that constraint. Clause-initial FRs which are not the subject of the main clause violate this constraint, and therefore cannot profit from the effects of the lower ranked *Red. The same holds for FRs which violate RC.

Conjoined constraints should be ranked higher than their constituent constraints, hence, 1to1&S<O should also be ranked higher than S<O. In fact, the constraint can fully take over the job of S<O. So we will rank 1to1&S<O in place of S<O, which will be ranked lowest.

We can now state which constraint rankings we need to model the results of the experiment and the corpus study. The two methods differ in two rerankings which have been marked with frames in (12):

(12) **Judgement ranking:**

RO ≫ RCr ≫ $\boxed{\text{RC} \gg \text{1to1\&S<O}}$ ≫ $\boxed{\text{1to1} \gg \text{*Red}}$ ≫ S<O

Corpus ranking:

RO ≫ RCr ≫ $\boxed{\text{1to1\&S<O} \gg \text{RC}}$ ≫ $\boxed{\text{*Red} \gg \text{1to1}}$ ≫ S<O

How can we account for these contradictory rankings with a single OT grammar? Boersma (this volume) discusses similar mismatches between relative corpus frequencies and relative acceptability in experiments in terms of the comprehension/production asymmetry.

In our case, this would imply that the correlative structure is easier to parse, hence preferred in the experiment, but it is avoided in production, because it is, so to speak, "hypercorrect". Indeed, the most plausible reason why the CORR structure is avoided in the NOM-NOM context is that the resumptive pronoun appears totally superfluous:

(13) Wer Hunger hat, (der) soll eine Banane
[Who-NOM hunger has]-NOM (the one-NOM) shall a banana

essen
eat
"Whoever is hungry shall eat a banana"

The FR is the subject of the main clause in (13), it is clause-initial and therefore occupies the typical subject position and the FR pronoun bears nominative, the case of subjects. These two hints suggest the correct analysis to the parser already, the resumptive pronoun "*der*" provides no new information.

In our corpus study (Vogel and Zugck, 2003), we also counted the average length of the FR in the FR and CORR structures we found in the NOM-NOM context. The result, which was highly significant, was that the number of words between the FR pronoun and the first word of the main clause is 6.02 in the case of the FR, and 12.04 in the case of CORR.

This can be seen as additional evidence for the redundancy theory: The longer the FR, the harder it is to keep track of the first word of the FR, i.e., the more advantageous it is to double it with a resumptive pronoun.

However, Boersma's line of reasoning also implies for our case that the CORR structure is easier to comprehend in principle. More generally, the more function words a sentence contains the easier can it be comprehended. I doubt that this is correct. When we try to understand a clause we are first of all interested in the content words. Function words are meaningless, by definition, and it appears much more likely to me that an "inflation" of function words makes comprehension more difficult rather than easier.

The advantage of the CORR structures in the experiment could be more task-specific: It is easier to judge their grammaticality, precisely because there is less implicit grammatical information than in the FR structures. This is rewarded in the experiments. The participants might also be more accurate in their judgements.

The experimental test sentences have been presented visually, word by word. Only one word at a time was displayed. This is different from the "presentation mode" of a newspaper text, where the full text is always available to the reader. It might be a consequence of such an experimental design that constraints that evaluate the morphological properties of words play a more important role than they "usually" do. This could be responsible for the task specific constraint "lifting" of RC and 1to1 – constraints which evaluate case morphology – that we observe in the grammaticality judgement task. If this explanation is on the right track then the variation in constraint ranking is systematic, not probabilistic: it affects only constraints which are particularly useful or useless in the task

at issue.

The above discussion might suggest that corpus data are more trustworthy, or more realistic than experimental data. That is, the constraint ranking that we need for the corpus would be the "real" grammar ranking, and the judgement ranking is derived from it. A couple of objections against such a point of view have to be made.

First, the constraint *Red is problematic. As already mentioned above, assuming that it is universal and freely rerankable would predict that there is a language that has only FRs, but not CORR structures. Such a language has not been attested thus far. Second, it is, in general, not the case that redundancy of the kind observed here, as an extension of syntactic structure, leads to ungrammaticality. The transformation of an ordinary question into a cleft question is another instance of such an extension. It has no effect on grammaticality:

(14) a. What do you want to buy?
b. What is it that you want to buy?

(14-b) expresses the same as (14-a), it only does so in a more complicated, perhaps less elegant way. Nevertheless, it clearly violates *Red more often. Third, *Red is sort of counterproductive in grammaticality judgements. Here, we *prefer* redundancy. Thus, this constraint very much looks like a *task-specific* constraint – which should perhaps not be included in the constraint hierarchy of "grammar proper".

Consider the study by Bresnan & *al.* (2001) again. Passive is chosen extremely rare in general in spoken English. Does this mean that passive is judged as ungrammatical by speakers 97% of the times? Corpus frequencies reflect preferences for the *use* of particular structures, and so does the corpus ranking proposed in (12). Frequencies reflect the grammar itself only in an *indirect* way. The rarity (not: absence!) of a structure S in a corpus could be given at least three different explanations:

1. S is the rare winner of a frequent competition.

2. S is the frequent winner of a rare competition.

3. S is *hypercorrect*, and avoided for stylistic reasons.

Bresnan & *al.* (2001) seem to treat all infrequency in terms of explanation 1. They use a model for OT syntax, where in fact passive and active always compete within the single competition for the expression

of a particular meaning. This has the somewhat counterintuitive consequence that passives are suboptimal, and hence ungrammatical, most of the times.

An alternative model that uses faithfulness constraints, and gives a syntactic specification for active/passive in the input, could first of all derive that a passive wins in a competition where passive is specified in the input.[11] The low frequency of the passive would then not be the result of the passive being a rare winner, but of the passive being rarely chosen as input. This is a totally different issue. The reason why passive is more rarely chosen is, of course, its higher markedness. But such an explanation makes no intrinsic claim about the grammaticality status of alternative structures, as does the stochastic evaluation in the model used by Bresnan & *al.* (2001).

In the same way, our constraint *Red could be seen as a criterion for the choice of particular inputs, but not as a constraint that evaluates the candidates for this input.

In fact, if we reconsider Bresnan & *al.*'s (2001) statistically significant finding that passives are even less frequent with third person patients, and do not occur at all with first and second person agents, we see that we cannot tell what this significance is evidence for: under the assumption that subjects are more likely to be topics, and first and second person are more likely to be topics, too, this finding could simply be due to the fact that the *contexts* where first and second person have a lower information structural status are extremely rare – a stochastic OT grammar based on this finding would take a property of the environment within which a grammar is applied for a grammatical constraint.

That it is necessary to distinguish these two different explanations for the rarity of structures can also be demonstrated with the result of another corpus study that we undertook (Vogel & *al.*, in preparation), again on free relative clauses in German in the COSMAS II corpus, this time with the neuter *wh*-pronoun "*was*" ("what"). This pronoun is the same for both nominative and accusative, which has the effect that even those speakers who do not tolerate FRs with case conflicts judge such FRs with "*was*" as grammatical. A typical contrast is the one in (15):

(15) Grammaticality contrast for some German speakers:

a. Ich kaufe was mir gefällt
I buy [what-NOM me-DAT pleases]-*ACC*

[11] I argued for such a version of OT recently (Vogel, 2003*a*, to appear).

b. *Ich lade ein wer mir gefällt
I invite [who-NOM me-DAT pleases]-*ACC*

The FRs found in the randomly selected sample of 500 sentences have been counted for FR and CORR structures in the four possible combinations of nominative and accusative, where the surface form "*was*" matches both case requirements. The results are displayed in Table 7:

FR case	Main verb case	FR	CORR	Sum
NOM	NOM	34 (66.7%)	17 (33.3%)	51 (38.9%)
ACC	NOM	21 (70.0%)	9 (30.0%)	30 (22.9%)
NOM	ACC	12 (66.7%)	6 (33.3%)	18 (13.7%)
ACC	ACC	21 (65.6%)	11 (34.4%)	32 (24.4%)
Sum		88	43	131

Table 7: Frequencies of German FR and CORR with the pronoun "*was*" in a sample of 500 sentences with "*was*" (Vogel & *al.*, in preparation)

We see that the case configuration has no influence on the relative distribution of FR and CORR. It is about two third to one third throughout. The relative infrequency of the CORR structure that we found with "*wer*" in the NOM-NOM context is observed again. Furthermore, the two contexts with conflicting case requirements are totally neutralised in their effects on the choice of the construction. The case conflict does not seem to exist anymore if the pronoun is *homophonous* for the two conflicting case features.

Compare these findings with the figures that we present in (Vogel and Zugck, 2003) for the FRs with the animate *wh*-pronouns "*wer*" (nominative) and "*wen*" (accusative):[12]

[12]The two samples also differ in the syntactic positions of the FRs that have been counted. "*was*" also serves as relative pronoun in German, and it therefore was possible to include headed relative clauses which are semantically equivalent to an FR (as in "everything that ...") in the statistics, and, likewise, clause-final FRs. The studies with "*wer*" and "*wen*" only counted clause-initial FRs and CORRs.

FR case	Main verb case	FR	CORR
NOM	NOM	274 (89.8%)	31 (10.2%))
ACC	NOM	5 (25.0%)	15 (75.0%))
NOM	ACC	0 (0%)	2 (100.0%)
ACC	ACC	1 (20.0%)	4 (80.0%)

Table 8: Frequencies of German FR and CORR with the pronouns "*wer*" and "*wen*"in two samples of 500 sentences each (Vogel and Zugck, 2003)

Only the NOM-NOM context has a preference for FR with animate *wh*-pronouns. That we find the same with all FRs with "*was*" irrespective of the case pattern shows that the matching effect is indeed a surface phenomenon.

However, we make a second observation which is perhaps rather unexpected. It concerns the relative frequency of the contexts themselves. Under the assumption that nominative is more frequent than accusative in finite clauses in general,[13] we expect that NOM-NOM is the most frequent pattern, and ACC-ACC the least frequent, while ACC-NOM and NOM-ACC should be equal. This is not the case in the "*was*" sample. The context ACC-ACC is about as frequent as ACC-NOM and NOM-ACC has lowest frequency. The distribution of nominative and accusative as main verb case and FR case is listed in Table 9:

case	assigned by main verb	assigned to FR pronoun
nominative	81	69
accusative	50	62

Table 9: Cases assigned by main verb to the FR and by FR verb to the FR pronoun in a sample of 131 FRs

To calculate our expectations for the distribution of the case patterns, let us take the figures we find for the main verbs as the base.[14] Table 10

[13]All finite verbs that assign accusative also assign nominative, but there are many verbs which do not assign accusative. Independently of verb frames, all clauses must have a subject, i.e., a nominative, in German.

[14]The calculation for the four contexts is: NOM-NOM $= 81 \times 81 = 6561$; ACC-NOM $= 4050$; NOM-ACC $= 4050$; ACC-ACC $= 2500$. The percentages of these figures are then calculated relative to their sum: $6561 + 4050 + 4050 + 2500 = 17161$. These are used in Table 10.

lists the expected values, and the actual findings.

FR case	Main verb case	expected	found
NOM	NOM	50.0 (38.2%)	51 (38.9%)
ACC	NOM	30.9 (23.6%)	30 (22.9%)
NOM	ACC	30.9 (23.6%)	18 (13.7%)
ACC	ACC	19.1 (14.6%)	32 (24.4%)

Table 10: Expected and found distribution of case configurations in a sample of 131 sentences with FRs with "*was*"

The departures from the expected values for the NOM-ACC and ACC-ACC patterns are statistically significant. This result is in line with the relative markedness these patterns are assigned by the grammar. In FRs with the NOM-ACC pattern, a case that is higher on the case hierarchy is suppressed in favour of a lower one – this is a highly marked situation. FRs with the ACC-ACC pattern are much less problematic, because both cases match, there is no case conflict. That this structure has a higher frequency is therefore expected.

The two results of this study together show on the one hand that the case configuration does not decrease the preference for the FR if the form of the FR pronoun fits both case requirements. The conflict is resolved at the surface, and this is sufficient. On the other hand, we observe that potentially problematic configurations, like the NOM-ACC pattern, are significantly less frequent than we expect them to be. The conclusion must be that such patterns tend to be avoided as inputs already – even where they turn out to be unproblematic in practice. The case patterns are crucially dependent on the lexical material that is chosen, in particular, the case requirements of the chosen verbs. But the choice of lexical material is not subject to a standard OT competition. It is given in the input. Markedness is thus demonstrated to guide not only the choice of *how* things are expressed (as FR or CORR), but also of *what* is to be expressed (which verbs with which case pattern are chosen/avoided).

6 Summary

The argument that I tried to elaborate on in this paper is that the conceptual problem behind the traditional competence/performance distinction

does not go away, even if we abandon its original Chomskyan formulation. It returns as the question about the relation between the model of the grammar and the results of empirical investigations – the question of empirical verification.

The theoretical concept of markedness is argued to be an ideal correlate of gradience. Optimality Theory, being based on markedness, is a promising framework for the task of bridging the gap between model and empirical world. However, this task not only requires a model of grammar, but also a theory of the methods that are chosen in empirical investigations and how their results are interpreted, and a theory of how to derive predictions for these particular empirical investigations from the model.

Stochastic Optimality Theory is one possible formulation of a proposal that derives empirical predictions from an OT model. However, I hope to have shown that it is not enough to take frequency distributions and relative acceptabilities at face value, and simply construe *some* Stochastic OT model that fits the facts. These facts first of all need to be interpreted, and those factors that the grammar has to account for must be sorted out from those about which grammar should have nothing to say. This task, to my mind, is more complicated than the picture that a simplistic application of (not only) Stochastic OT might draw.

Bibliography

Abney, Steven (1996). 'Statistical Methods and Linguistics.' *In* Klavans, Judith L. and Philip Resnik, eds., *The Balancing Act: Combining Symbolic and Statistical Approaches to Language*, pp. 1–26. Cambridge, Massachusetts: MIT Press.

Boersma, Paul (1998). *Functional Phonology. Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Boersma, Paul and Bruce Hayes (2001). 'Empirical tests of the gradual learning algorithm.' *Linguistic Inquiry* **32**:45–86.

Boethke, Jutta (in preparation). *Kasus im Deutschen: Eine empirische Studie am Beispiel freier Relativsätze*. Diploma thesis, Institute of Linguistics, University of Potsdam.

Bresnan, Joan, Shipra Dingare, and Christopher Manning (2001). 'Soft Constraints Mirror Hard Constraints: Voice and Person in English and

Lummi.' *In* Butt, Miriam and Tracy Holloway King, eds., *Proceedings of the LFG01 Conference*. Stanford: CSLI Publications.

Büring, Daniel (2001). 'Let's Phrase It! — Focus, Word Order, and Prosodic Phrasing in German Double Object Constructions.' *In* Müller, Gereon and Wolfgang Sternefeld, eds., *Competition in Syntax*, No. 49 in Studies in Generative Grammar, pp. 101–137. Berlin & New York: de Gruyter.

Chomsky, Noam (1964). 'Degrees of Grammaticalness.' *In* Fodor, J. A. and J. Katz, eds., *The Structure of Language: Readings in the Philosophy of Language*, pp. 384–389. Englewood Cliffs, NJ: Prentice-Hall.

— (1965). *Aspects of the Theory of Syntax*. Cambridge, Massachusetts: MIT Press.

— (1995). *The Minimalist Program*. Cambridge, Massachusetts: MIT Press.

Featherston, Sam (to appear). 'Universals and the Counter-example Model: Evidence from wh-Constraints in German.' Manuscript, University of Tübingen.

Groos, Anneke and Henk van Riemsdijk (1981). 'Matching Effects with Free Relatives: a Parameter of Core Grammar.' *In* Belletti, Adriana, Luciana Brandi, and Luigi Rizzi, eds., *Theories of Markedness in Generative Grammar*, pp. 171–216. Pisa: Scuola Normale Superiore di Pisa.

Josefsson, Gunlög (2003). 'Four myths about Object Shift in Swedish – and the Truth ....' *In* Delsing, Lars-Olof, ed., *Grammar in Focus II. Festschrift for Christer Platzack*, pp. 199–207. Lund: Wallin + Dalholm.

Keller, Frank (2000). *Gradience in Grammar. Experimental and Computational Aspects of Degrees of Grammaticality*. Ph.D. thesis, University of Edinburgh.

— (2001). 'Experimental Evidence for Constraint Competition in Gapping Constructions.' *In* Müller, Gereon and Wolfgang Sternefeld, eds., *Competition in Syntax*, pp. 211–248. Berlin: Mouton de Gruyter.

Kolb, Hans-Peter (1997). 'Is I-language a Generative Procedure?' *In GB-Blues: Two Essays*, No. 110 in Arbeitspapiere des Sonderforschungsbereichs 340, pp. 1–14. Tübingen: University of Tübingen.

Manning, Christopher D. (2003). 'Probabilistic Syntax.' *In* Bod, Rens, Jennifer Hay, and Stefanie Jannedy, eds., *Probabilistc Linguistics*, pp. 289–341. Cambridge, Massachusetts: MIT Press.

Müller, Gereon (1999). 'Optimality, Markedness and Word Order in German.' *Linguistics* **37**:777–818.

Nagy, Naomi and Bill Reynolds (1997). 'Optimality Theory and variable word-final deletion in Faetar.' *Language Variation and Change* **9**:37–55.

Pittner, Karin (1991). 'Freie Relativsätze und die Kasushierarchie.' *In* Feldbusch, Elisabeth, ed., *Neue Fragen der Linguistik*, pp. 341–347.

Prince, Alan and Paul Smolensky (1993). 'Optimality Theory. Constraint Interaction in Generative Grammar.' URL: `http://roa.rutgers.edu/view.php3?roa=537`. Rutgers Optimality Archive: ROA-537.

Reynolds, Bill (1994). *Variation and Phonological Theory*. Ph.D. thesis, University of Pennsylvania.

Sternefeld, Wolfgang (2001). 'Grammatikalität und Sprachvermögen. Anmerkungen zum Induktionsproblem in der Syntax.' *In* Bayer, Josef and C. Römer, eds., *Von der Philologie zur Grammatiktheorie: Peter Suchsland zum 65. Geburtstag*, pp. 15–44. Tübingen: Niemeyer.

Vogel, Ralf (2001). 'Case Conflict in German Free Relative Constructions. An Optimality Theoretic Treatment.' *In* Müller, Gereon and Wolfgang Sternefeld, eds., *Competition in Syntax*, pp. 341–375. Berlin: Mouton de Gruyter.

— (2002). 'Free Relative Constructions in OT Syntax.' *In Sonderheft Optimality Theory*.

— (2003*a*). 'Remarks on the Architecture of OT Syntax.' *In* Blutner, Reinhard and Henk Zeevat, eds., *Optimality Theory and Pragmatics*. Houndmills, Basingstoke, Hampshire, England: Palgrave MacMillan.

— (2003*b*). 'Surface Matters. Case Conflict in Free Relative Constructions and Case Theory.' *In* Brandner, Ellen and Heike Zinsmeister, eds., *New Perspectives on Case Theory*, pp. 269–299. Stanford: CSLI Publications.

— (to appear). 'Correspondence in OT Syntax and Minimal Link Effects.' *In* Stepanov, Arthur, Gisbert Fanselow, and Ralf Vogel, eds., *Minimality Effects in Syntax*. Berlin: Mouton de Gruyter.

Vogel, Ralf, Stefan Frisch, and Marco Zugck (in preparation). 'Case Matching. An empirical Study.' Manuscript, University of Potsdam. To appear in *Linguistics in Potsdam*.

Vogel, Ralf and Marco Zugck (2003). 'Counting Markedness. A corpus investigation on German free relative constructions.' *In* Fischer, Susann, Ruben van de Vijver, and Ralf Vogel, eds., *Experimental Studies in Linguistics 1*, vol. 21 of *Linguistics in Potsdam*, pp. 105–122. University of Potsdam.