In: Kertész, A., Rákosi, Cs. (eds.): *New Approaches to Linguistic Evidence. Pilot Studies / Neue Ansätze zu linguistischer Evidenz. Pilotstudien.* Frankfurt am Main etc.: Lang, 2008.

# Conservatism vs. innovation in the debate on data in generative grammar[1]

András Kertész and Csilla Rákosi

> *"[…] much work in the generative tradition of grammar is fatally undermined by its oversimplified assumptions about the patterns that the data which it is attempting to model actually exhibit."*
>
> *(Featherston 2007: 270)*

## 1. Introduction

In Kertész & Rákosi (2008a) we summarised what we called the standard view of linguistic data (SVLD).[2] This view involves two antagonistic positions concerning the nature of linguistic data exemplified by generative and corpus linguistics, which nevertheless share some basic assumptions.

In footnote 19 of Kertész & Rákosi (2008a) we called attention to two further branches of the current discussion on data and evidence in linguistics which we did not go into there. One was the debate published in the special issue of *Corpus Linguistics and Linguistic Theory*, the other was the issue of *Theoretical Linguistics* entitled *Data in generative linguistics*. In Kertész & Rákosi (2008b) we continued our state-of-the-art analysis by considering the former branch. So as to arrive at a relatively comprehensive picture of current views, in the present article we will supplement our previous considerations with an analogous analysis of the latter debate.

In analogy to Kertész & Rákosi (2008b), we raise the following problem:

---

[1] Our research was supported by the Research Group for Theoretical Linguistics of the Hungarian Academy of Sciences at the Universities of Debrecen, Pécs and Szeged. It was also supported by the grant OTKA T 049139. We are grateful to George Seel for improving our English.

[2] For the English summary of (SVLD), (SVAPS), (SP1) and (SP2) see Kertész & Rákosi (2008b).

(P) In what respects do the contributions to the debate on data in generative grammar as published in Sternefeld (2007) go beyond
  (a) (SVLD), and
  (b) the views summarised in (SP1) and (SP2) in Kertész & Rákosi (2008a)?

So as to avoid repetitions of former insights and a lengthy presentation of the views we want to evaluate with respect to (P), we will proceed as follows. Our objective will not be to present a detailed analysis of all arguments put forward in the debate. We will not take sides between the claims of Featherston's target article and the commentaries by the other contributors, either. Just as in Kertész & Rákosi (2008b), here, too, we will concentrate on highlighting those ideas which seem to motivate a methodological turn in linguistics and provide the points of departure for the elaboration of a possible metatheoretical model of linguistic data and evidence.[3]

## 2. The target article

In the target article Featherston criticises the way introspective data are handled in the research practice of generative linguistics. However, his aim is not restricted to merely pointing out the deficiencies of this practice; rather, he puts forward suggestions for its improvement. His criticism centres on the following:

> "The most fundamental part of the problem is that a significant number of linguists are still, in spite of all the warnings to the contrary, using as the basis of their work what we might call *linguist's judgements*. In the worse [sic] case these are introspective judgements given by linguists themselves as the data base of their own theoretical work, on the basis of a single example sentence, not checked against the intuitions of other independent informants, and often idealised to a dichotomy of good or bad." (Featherston 2007: 272f.; emphasis as in the original)

Featherston narrows down this general criticism to a series of more specific claims:

– It is inadequate if linguists use their own unconfirmed introspective judgements so as to obtain linguistic evidence. Individual judgements are *subjective* and may be influenced by the linguist's theoretical concerns as well. The objectivity of data cannot be secured this way (Featherston 2007: 277f.).

[3] So as to document our evaluation of the particular authors' claims, we will make extensive use of quotations in the footnotes. Quotations will appear in the main text of our paper only if their analysis is especially relevant for our line of reasoning.

- For similar reasons it is inadequate, too, to take judgements uncritically from the literature. Data referred to in the literature are often *dubious* because they are based solely on the given author's linguistic intuition (Featherston 2007: 278).
- Isolated examples from languages which are little studied and seem to be obscure are an insufficient source of data, too, because their reliability *cannot be checked* (Featherston 2007: 278).
- Linguists often defend themselves by referring to their own idiolect when confronted with the fact that a particular grammaticality judgement of theirs was incorrect. This tactic is, according to Featherston, mistaken. It is irreconcilable with the desideratum of building syntactic theory on objective observations as well as with the generalisability of grammaticality judgements and via this, with the idea of universal grammar (Featherston 2007: 278f.).
- At the present stage of the development of generative syntax, introspective judgements do not provide an adequate empirical basis for linguistic theorising. What is needed is, therefore, "the new qualities and quantities of data which have become available" (Featherston 2007: 279).
- Uncontrolled and inconsistent idealisations may lead to oversimplifications and loss of relevant information. This applies especially to cases in which different degrees of grammaticality are idealised to binarity in well-formedness (Featherston 2007: 280).
- The idealisation to binarity often leads to *inconsistency* between theoretical assumptions and the applied methodology because many linguists "frequently use multiple intermediate judgements at the same time as assuming a model of grammar that has categoricity as one of its enabling assumptions" (Featherston 2007: 281). A further problem with binarity is that it allows only constraints whose infringement results in full ungrammaticality – which is an uncontrolled idealisation as well, clearly reducing the explanatory power of the theory (Featherston 2007: 280f.).

So as to overcome these shortcomings of introspective judgements and to arrive at an adequate empirical basis for theory development, Featherston proposes new methodological principles:

> "We might summarise the essential requirement of better data in one word: control. This factor is what makes evidence gathered experimentally more valuable than any single example. Experimental control excludes many irrelevant factors from affecting the results, so that one can be more certain that the effects one finds are due to the factors that one is interested in." (Featherston 2007: 281f.)

He suggests a set of simple *minimum criteria* that should be applied in any research in generative linguistics, and that facilitates the systematic control of da-

ta. He suggests two "essentials" that he considers necessary so that judgement data can work as valid evidence at all, and two further "desirables" whose fulfilment may improve the evidence at issue:

- The first essential requirement is the use of multiple informants. Featherston considers this to be the most important criterion, because it secures the independence, the accuracy (noiselessness) and the reliability of the data by keeping the effects of performance errors at the minimum.[4]
- According to the second essential requirement, multiple lexical variants should be used, the main advantage of this being that they facilitate the control of lexis-specific and content-specific factors.[5] Since the use of multiple lexical variants leads to the repetition of the judgement process, it contributes to the accuracy of the data as well, because the effects of performance errors can be reduced by averaging over judging events (Featherston 2007: 291).
- The first desirable is the appropriate formulation of the task given to the informants. The main maxim is that reference to the informant's own production is to be avoided. Informants should judge the linguistic input, not their production because "this prevents issues of personal prestige or prescriptively defined 'correctness' playing a role" (Featherston 2007: 292).
- The second desirable says that instead of the binarity of grammaticalness, informants should be allowed to use an open-ended multi-point scale of grammaticality. This is an essential component of the new methodology

---

4 "If the responses of a group of a dozen of independent informants produce a given pattern, then this cannot simply be the wishful thinking of the author. […] the judgement patterns of a group of informants are replicable: we have an objective phenomenon to investigate. It is important that syntacticians should show scientific detachment and be seen to be producing theory to fit the facts, rather than the opposite. […] A study which uses data from multiple informants is thus more intellectually rigorous than one which does not." (Featherston 2007: 283)
"Judgements are fundamentally noisy, and show some variability both between informants and across judging events by the same person. […] both these sorts of variability can be evened out, if we obtain multiple independent judgements. The errors cancel each other out and the judgements cluster around a mean, which we can take to be the 'underlying' value, free of the noise factor. Multiple informants thus deliver more accurate data." (Featherston 2007: 284)
"[…] the only reliable source of perceived well-formedness is the judgements of a group, in which the slips and noise cancel each other out." (Featherston 2007: 289)

5 "This has two positive consequences; firstly, if ten lexical forms can be found, then the effect is much less likely to be dependent on a particular lexicalization […]; second, the search for lexical alternatives will make the researcher aware of all those examples which do not work." (Featherston 2007: 291)

> Featherston proposes, which he calls "thermometer judgements".[6] This step makes it possible to retain the original, "raw" gradient character of the judgements and leaves the task of the idealization to "good" and "bad" structures to the linguist (Featherston 2007: 293).

After having summarised the central tenets of the target article, as a next step we have to examine in what way and to what extent they deviate from (SVLD).

(SVLD)(a) says that not all linguistic theories can be considered as empirical, but only those that rely on the appropriate type of data. Only one data type may be regarded as relevant and legitimate: in the generativist view it is introspective data, while according to corpus linguists it is corpus data. Featherston clearly transcends the antagonism of the two views because he considers both corpus data and introspective data as legitimate,[7] although he expressly prefers one kind of experimental data, namely grammaticality judgements gained under controlled circumstances.[8] Further, he praises cooperation with other branches of cognitive science, above all, with cognitive psychology (Featherston 2007: 271f.).

His attitude towards (SVLD)(b) – which says that it is solely the origin of the data that decides what counts as the appropriate type of data – is more difficult to evaluate. On the one hand, he does not condemn introspection as an illegitimate data source *per se*; that is, in this respect he breaks with (SVLD)(b). On the other hand, however, he clearly rejects a special kind of introspection, that is, the use of the linguist's own linguistic intuition, deeming it an "extremely poorly controlled experiment" (Featherston 2007: 282). A further difficulty results from the inconsequent judgement of the acceptability and usability of data consisting of one person's linguistic intuitions. According to the "essentials", data stemming from single individuals' introspection do not fulfil the minimal requirements of controllability; that is, they have to be regarded as "invalid". At the same time,

---

6 "Our methodology, *thermometer judgements*, differs chiefly in allowing informants to use a linear scale, not a magnitude scale. […] The name thermometer judgements derives from the similarity of our judgement scale to the temperature scale, which is an open-ended scale with two reference points, freezing point and boiling point. Our method also has two reference items to fix the location and amplitude of the scale, setting the lower one at twenty and the upper one at thirty to keep informants away from zero, where distortion can occur. This method thus rather resembles the simple seven-point rating scale, but has no minimum division and no maximum or minimum scores. It thus allows informants the freedom to express their intuitions with minimum constraint." (Featherston 2007: 296f.)

7 "This focus is in no way intended to belittle the value of corpus data or make out that this data type is any less relevant." (Featherston 2007: 271)

8 "[..] we do not wish to condemn introspective judgements as a data type." (Featherston 2007: 277)
See also Featherston (2007: 277).

however, he does not forbid the use of uncontrolled data but he regards them only as "weak", which "will […] be valued less highly" (cf. Featherston 2007: 282).

According to (SVLD)(c) both camps (i.e. generative linguists and corpus linguists) regard their own data-handling procedure as unproblematic. They believe that a few general methodological rules are sufficient to control the reliability of the data. At first glance, Featherston seems to reject this attitude, for the essence of his target article is the claim that the way introspective data are gathered and evaluated in the practice of generative linguistic research is highly problematic. His standpoint is, however, double-faced in this respect as well, since he argues that the data-handling of generative linguistics can be made unproblematic via the application of a few simple methodological rules. Although he assumes the "validity" of data to be gradual (cf. Featherston 2007: 283), in connection with his own proposals he does not raise and scrutinise possible weaknesses or sources of error etc. He takes it for granted, for example, that a (small) group of informants is enough to secure the independence and the reliability of data, that "systematic differences across groups of speakers are small and rarely of any significance" (cf. Featherston 2007: 284), or that the influence of "issues of personal prestige or prescriptively defined 'correctness'" (cf. Featherston 2007: 292) can be avoided simply by the appropriate wording of the tasks. The most serious objections can be lodged to his unwarranted claims that groups always produce clear patterns and individual deviations from these are due to "performance errors" which are always balanced out through the errors of the other participants in the given experiment (cf. Featherston 2007: 285ff.). These claims raise at least two problems. First, they might prevent the linguist who accepts them as background assumptions from interpreting data in a way that contradicts them, and they might tempt one not to raise or even to ignore alternative hypotheses or explanations. Second, an overhasty and uncritical adoption of these methodological principles makes it impossible to reflect on their tenability and hinders the elaboration of a methodology which could serve as a series of guidelines on the interpretation and the statistical evaluation of the data as well as on the identification of the factors which may influence or even distort the results.

(SVLD)(d) says that the relationship between the data and the hypotheses of the theory is unidirectional, and determined by general, strict methodological rules. Featherston's stance with respect to this background assumption needs differentiated evaluation.

First, he seems to accept it. He assumes that linguistic theory formation proceeds in such a way that, as a first step, the data have to be gathered and analysed and then, as a second step, a theory should be developed that accounts for them:

> "A theory is always a theory about a set of data. […] Data is a precondition for theory, and the quality of a theory can never exceed the quality of the data set which it is based on." (Featherston 2007: 314)

He criticises the practice of generative linguists which tendentiously manipulates the data so that they fit the prefabricated assumptions, while data which are at variance with the hypotheses at issue are either labelled as irrelevant or modified in the light of the theory (Featherston 2007: 271, 279). This seems to be a naïve inductivist position according to which the hypotheses of the theory have to stem directly from the data.[9] This attitude could be called "the data-dependency of the theory".[10] This is the exact reversal of the thesis maintaining the theory-dependency of data, also stated in (SP1).

Second, in his argumentation there are hints at the considerable refinement of this picture. For example, according to Featherston, the main advantage of using multiple lexical variants is that it allows for the control for content-specific elements (cf. Featherston 2007: 291). He emphasises the importance of theoretical considerations in planning experimental settings in order to exclude the influence of irrelevant factors (cf. Featherston 2007: 281f.). This is clearly the opposite direction compared to the progress from data towards theory, because it proceeds from theoretical considerations towards possible data. The *pattern matching technique* (cf. Featherston 2007: 301) also requires a bidirectional relationship between data and theory, since it involves the testing of already existing hypotheses with the help of different sets of conditions, e.g., sets of data. A further indication of a subtler attitude towards the relationship between data and theory is that he considers the fine differentiation of data as useful because "it sometimes gives us clues about the correct structural analysis" (cf. Featherston 2007: 299) which is a much more moderate stance than rigid naïve inductivism and makes the idea of the "data-dependency of theory" an important and progressive step.

---

9 We use the term "naïve inductivism" in the following sense:

> "The image of science presented by NI [= naïve inductivism] is the following. Science is a 'bottom-up' affair. It starts with observatonal facts. In a strong version of NI, facts are prior to theory; facts are gathered independently of any theoretical considerations. But ones gathered they can be used to infer hypotheses, laws or theories. The reporting of these facts can be made more precise with the development of observational techniques and instruments; and the stock of observational facts will increase over time. It is upon this allegedly firm foundation that the edifice of scientific knowledge is commonly said to be built, mainly using inductive inference to get from observational facts to generalizations or theory. […] Often NI is presented as *the* one and only method of science." (Nola & Irzik 2005: 208; emphasis as in the original)

10 This idea appears in the following passage as well:

> "[…] much work in the generative tradition of grammar is fatally undermined by its oversimplified assumptions about the patterns that the data which it is attempting to model actually exhibits." (Featherston 2007: 270)

Third, the idea of falsificationism is explicitly present in his paper (cf. Featherston 2007: 311).

There are several considerations in Featherston's argumentation which might allow to transcend the rigid dichotomy between "the context of justification" and the "context of discovery" as maintained in (SVAPS). As we know, the standard view of the analytical philosophy of science regarded the "context of justification" as the only legitimate perspective of the philosophy of science and this perspective was also adapted by the methodology of generative linguistics (see Kertész & Rákosi 2008a). In opposition to this view, Featherston mentions many elements of linguistic theorising that require that *the generation of hypotheses* should be focused on – that is, the process of what counts as the "context of discovery" in the terminology of the analytical philosophy of science. This insight could be a significant step towards a break with (SVLD) and (SVAPS). That this is not so, is first of all due to his uncritical acceptance of data gained as results of experiments. He conceptualises this kind of data at several points of his writing as firm, unshakeable evidence. He also points out that one of the reasons why generativism may be criticised is that on the one hand there are many alternative possibilities to analyse the same data, while on the other hand there are no accepted mechanisms leading to a decision between the alternatives.[11] In spite of this, he did not realise that it is the uncertainty and fallibility of *data* that allows alternative accounts of them, each of which may be plausible in one way or another. He formulates the desideratum to reveal those mechanisms which facilitate the decision between the alternatives. Nevertheless, the elaboration of such mechanisms remains within a methodology which is the combination of naïve inductivism and strong falsificationism. The latter cannot be evaluated as the transgression of the boundaries of (SVLD)(d).

(SVLD)(e) considers evidence as an empirical datum that is directly (that is, without reference to any theoretical framework) given, that is reliable and that is treated as an unquestionable fact suitable for the evaluation of hypotheses. Featherston seems to accept the background assumption that evidence is a special subset of data capable of deciding between alternative hypotheses. He assumes that, as a norm to be followed, evidence should be the means of testing hypotheses and should, therefore, be objective in the sense of (SVAPS), i.e. intersubjectively given. However, in another respect he also goes beyond (SVLD)(e), because he does not question the evidential status of introspective data entirely, although they cannot be regarded as spatiotemporal events. There is a second ambiguity in his standpoint. His argumentation aims at showing how fallible, uncertain and questionable the evidence is that generative linguists in their everyday research practice actually use and he seems to accept the gradual

[11] "It cannot be a healthy position for the theory of grammar to be in, when for many phenomena, multiple widely varying analyses are possible and the field offers no procedures for deciding between them." (Featherston 2007: 310)

reliability of data.[12] Nevertheless, although he suggests ways of improving reliability and assumes that fallibility and uncertainty should be diminished, at several points he claims that data gained under controlled circumstances must be regarded as unquestionable facts whose truth is certain.[13]

In sum, we have seen that on the one hand, Featherston goes beyond (SVLD)(a). He tries to improve the way introspective data are handled in generative grammar, mainly by continuing what was initiated in Schütze's seminal work. On the other hand, his attitude towards (SVLD) is ambiguous. Besides progressive elements of great significance his views also contain remnants of (SVLD).

## 3. The commentaries

### 3.1. Introductory remark

Most of the commentators are practicing generative linguists. Accordingly, their reaction to Featherston's target article is characterised by two features right from the outset. First, they do not deny that the data handling technique of generative grammar needs improvement. Second, they advocate the methods they have been using for decades and also argue against the particular suggestions proposed by Featherston. Therefore, all of the commentaries are double-faced in the sense of being both close to (SVLD) in several respects and transgressing its boundaries at the same time. Moreover, not all of them touch on all the five tenets of (SVLD).[14]

---

12 "[…] we are not demanding that every syntactician carry out elaborate experiments all the time. Less time-consuming ways of gathering data are also valid and useful, but the detail in the results is proportional to the effort put into obtaining them. Linguists who use very informally obtained data must expect to be contradicted by others using more reliable and more detailed information." (Featherston 2007: 270)

13 "If all syntacticians took this step we would find a marked improvement in the linguistic quality of work in the field, since theoretical work would be accounting for *empirical facts*." (Featherston 2007: 308; emphasis added)
"If we obtain the judgements of twenty-five informants, we can reasonably assume the findings to be valid for the whole population of speakers and not just for a single person. Furthermore, such findings become *hard facts* which need accounting for […]." (Featherston 2007: 309; emphasis added)

14 We have not analysed Newmeyer's (2007) contribution because he does not comment on the problems that are relevant for us at all.

### 3.2. From the multidimensionality of grammaticality judgements to the combination of methods

Bornkessel-Schlesewsky & Schlesewsky (2007) agree with Featherston's criticism of the treatment of data in generative linguistics and admit that the prevailing practice "may lead to completely incorrect analyses of particular constructions or even entire languages" (Bornkessel-Schlesewsky & Schlesewsky 2007: 320). However, they do not deem his proposals tenable, and argue for the combination of methods which might reduce the shortcomings of the individual methods and provide information which could not be obtained otherwise.

Bornkessel-Schlesewsky & Schlesewsky clearly break with (SVLD)(a) because they think that relying on a single data type is not satisfactory. What is needed is the combination of different data types and the search for correlations between the patterns the latter show (cf. Bornkessel-Schlesewsky & Schlesewsky 2007: 320ff.).

Consequently, they reject (SVLD)(b) as well. They argue that Featherston's methodological innovations are not able to identify and dissociate the variety of linguistic, psychological and other factors which influence the judgement process. On the basis of judgements alone, it is impossible to reveal the role and structure of grammatical competence. This means that they regard judgements as an unsatisfactory data source, because they do not allow for a differentiation between the consequences of the violation of rules of grammar, the violation of violable grammatical principles, the effect of processing difficulties (the need for reanalysis, for example) or of individual psychological factors (cf. Bornkessel-Schlesewsky & Schlesewsky 2007: 326). Despite this, they do not fall back into (SVLD)(b) because they do not suggest another source which, serving with allegedly reliable data, could be considered as the only legitimate source.

Their attitude towards (SVLD)(c) is negative in a similar way, mainly because they emphasise the inherent multidimensionality of judgements. This means that the latter are exposed to various linguistic and extralinguistic influences which make them subject to interpretation problems.[15] In order to reduce the effects of irrelevant factors, they propose methods which provide more complex data, taking into consideration brain activities or temporal parameters. They emphasise, however, that none of these data types suffices alone, but only in

[15] "[…] judgements are inherently multidimensional in that they incorporate a range of linguistic and extralinguistic influences, which cannot be teased apart by only considering the final judgement itself. Judgements of the gradient and time-insensitive type therefore give rise to the same interpretation problems as those faced by introspection." (Bornkessel-Schlesewsky & Schlesewsky 2007: 319)

combination with others.[16] They mention cases which exemplify their conclusion, according to which

> "[w]hile the output of such a combination cannot provide us with 'the grammar' itself, it will certainly constrain what we might consider part of the grammar and what not." (Bornkessel-Schlesewsky & Schlesewsky 2007: 331f.)

They seem, however, to judge results of neuro- and psycholinguistic experiments to be certain or at least more reliable than results gained by purely linguistic methods. They do not raise questions about their interpretability and reliability as they did in connection with judgement data.[17] This undermines the alleged multidimensional character of their methodology and points towards the clear dominance of neuropsychological methods and the inferior status of other methods – and via this, makes their relationship to (SVLD)(c) indefinite.

Bornkessel-Schlesewsky & Schlesewsky abandon (SVLD)(d). This is clearly witnessed in the summary of their criticism of Featherston's stance:

> "[…] it seems to us that his approach brings with it the danger of replacing what might be referred to as 'theory-driven imperialism' with a new kind of 'data-driven imperialism'. In other words: while experimental data may be superior to intuitions in terms of their reliability, they still require interpretation. Indeed, they allow for just as much misinterpretation as intuitions." (Bornkessel-Schlesewsky & Schlesewsky 2007: 320)

They seem to assume that there are several cycles of production, interpretation and comparison of different data types and the raising, the evaluation, the refinement or the rejection of possible explanations.

Bornkessel-Schlesewsky & Schlesewsky's attitude towards (SVLD)(e) is less clear than it was towards the other aspects of (SVLD). On the one hand, they seem to hold the view that empirical data are unavoidably complex and multidimensional. They contain different factors which make it impossible to reveal the structure of grammar directly. Data give rise to interpretation problems, which means that they yield only indirect information about linguistic competence and allow only for uncertain, more or less reliable hypotheses. This is especially true of judgement data which stem from the interaction of a series of linguistic and extralinguistic factors that considerably influence the result of the judgement process. On the other hand, however, they use the term 'evidence'

---

16 "From an empirical perspective, there cannot be 'one perfect method' for the investigation of linguistic knowledge. Rather, it is important to recognise the limitations of individual methods and to capitalise upon the insights that can be gained by their combination." (Bornkessel-Schlesewsky & Schlesewsky 2007: 331)

17 "These findings therefore show that brain activation patterns can provide a differentiation between markedness and grammaticality even though no difference is observed in overt judgements." (Bornkessel-Schlesewsky & Schlesewsky 2007: 329)

solely in connection with neurological data, and seem to consider them as certain.[18]

### 3.3. Individual judgements as "a veritable goldmine of linguistic discovery and explanation"

Dikken, Bernstein, Tortora and Zanuttini think that grammaticality judgements of individuals are the only means with the help of which one can study linguistic competence, that is, Chomskyan I-language.[19] They stipulate that the task of generative linguistics is to study individuals' linguistic competence. Consequently, one must not presuppose that every individual possesses the same I-language. Rather, we have to investigate the "micro-variants" of individual speakers.[20] Consequently, they judge Featherston's methodological proposals mistaken because they lead to the loss of relevant information:

> "What one typically finds is that there is linguistically significant variation among speakers with respect to their judgements on individual sentences in the questionnaire. Averaging the informants' responses to a mean value will obliterate individual differences. The problem with this is that all potentially interesting points of variation are then cast aside as 'noise', and the net result is lots of gray averages. Universal Grammar this becomes Universal Gray, and that would hardly be reflective of the real patterns that micro-analysis would allow one to identify." (Dikken, Bernstein, Tortora & Zanuttini 2007: 343)

---

18 "In essence, the activation level of this region therefore appears to directly mirror the underlying grammatical system. In fact, the correspondence is so close that even the relative strengths of the different constraining factors behave just like they would be expected to on the basis of theoretical accounts […]." (Bornkessel-Schlesewsky & Schlesewsky 2007: 329)

19 "Because the object of investigation is the knowledge of an individual, the researcher could work with a single native speaker and characterise his/her knowledge of language, or grammar. […] More than five decades of research in generative linguistics have shown that the standard generative methodology of hypothesis formation and empirical verification via judgement elicitation can lead to a veritable goldmine of linguistic discovery and explanation." (Dikken, Bernstein, Tortora & Zanuttini 2007: 336)

20 "We agree that it is desirable to work with data elicited from as large a number of informants as possible. However, the generative syntactician's reason for desiring many (versus one) informants would differ from that given by Featherston: for the generative syntacticians, the more informants you have, the more data from individual grammars you have, which gives you the potential to find micro-variants you might otherwise not have found […]. However, despite the desirability of working with data elicited from a group of informants, we do not believe that the field should *require* that every piece of research be based on the judgements of at least ten or twelve speakers." (Dikken, Bernstein, Tortora & Zanuttini 2007: 339; emphasis as in the original)

They raise the methodological principle that the factors which might influence the outcome of grammaticality judgements such as unexpected semantic or pragmatic interpretations, can be avoided by choosing and consulting "trained informants", whose judgements are generally more reliable because they understand better what the linguist asks them for (cf. Dikken, Bernstein, Tortora & Zanuttini 2007: 349f.).

Dikken, Bernstein, Tortora & Zanuttini share (SVLD)(a) and (b), since they regard introspection not only as indispensable but as the only data source which can be relied on if one aims to reveal grammatical competence and linguistic knowledge in general.[21] They rigidly rule out the application of other data types.[22]

As for (SVLD)(c), they touch on several criticisms levelled against data handling techniques in generative linguistics. They try to show, or simply declare, that these alleged shortcomings are either completely unfounded, mistaken or do not cause serious problems and can be prevented easily – first of all, with the help of "trained informants". With this, they remain clearly within (SVLD)(c).

The authors seem to have a very special relationship to (SVLD)(d). First, they reject generalisation over grammaticality judgements by a group of individuals. Therefore, they disallow induction from data to hypotheses.[23]

Second, for the same reason, they:

> "[…] do not see how the mean value of the judgements of a group of speakers can confirm or disconfirm an individual's judgements: one's judgements are one's judgements, no matter what other speakers of 'the same language' might think." (Dikken, Bernstein, Tortora & Zanuttini 2007: 342)

---

21 "Generative syntacticians, generative linguists in general, and cognitive scientists even more generally all work with individuals and with subjective material, as that is what they must necessarily rely on, also in the construction of their 'objective' experiments and their selection of informants." (Dikken, Bernstein, Tortora & Zanuttini 2007: 338)

22 "[…] these distributional phenomena would have been entirely impossible to distill via any non-introspective, non-elicitation based data-gathering method. Corpus data simply cannot yield such a detailed picture of what is licit and, more crucially, what is not licit for a particular construction in a particular linguistic environment." (Dikken, Bernstein, Tortora & Zanuttini 2007: 336)

23 "Generalizations over a group can only be made if the individuals that constitute the group share the same linguistic knowledge. Yet there is no a priori reason to expect each individual belonging to a certain group of speakers to have set every single parameter in exactly the same way as his/her fellow group members. […] Such grammars are obviously very close, ensuring perfect mutual intelligibility, but may differ in some fundamental ways in their I-language constitutions, thus leading different speakers to analyze a particular string in different ways […]." (Dikken, Bernstein, Tortora & Zanuttini 2007: 342)

That is, they think that introspective data cannot play any role in the justification of theories. Third, although this means a clear break with (SVLD)(d), their standpoint seems to replace a questionable methodology with another one.[24]

This means at the same time that they do not accept (SVLD)(e) because they seem to give up the idea that data may confirm or refute the hypotheses of linguistic theories. Instead, they seem to advocate the paradoxical view that linguists "can gain access to the knowledge of a native speaker, which is, after all, *subconscious*" (Dikken, Bernstein, Tortora & Zanuttini 2007: 335) with the help of a *conscious* reflection with the informant on the motives of his judgements.

### 3.4. "The empirical problems of syntax do not go away with controlled experiments"

Fanselow agrees with Featherston's criticism and his proposals for increasing data quality with the help of controlled experiments, but in his view, "[a]t a purely methodological level, one gains little with the introduction of gradient judgements" (Fanselow 2007: 361). He is of the opinion that there are some relevant factors that Featherston has not considered (cf. Fanselow 2007: 354f.). He does not agree with Featherston that differences in acceptability ratings were usually due to irrelevant "noise", that is, mostly psychological factors such as inattention (cf. Fanselow 2007: 362f.). Real difficulties stem rather from the fact that acceptability judgements cannot be interpreted simply as grammaticality judgements.[25] This position is similar to Bornkessel-Schlesewsky & Schlesewsky's thesis of the multidimensionality of judgements. Fanselow's conclusion drawn from his stance about the usability of judgement averages resembles Bornkessel-Schlesewsky & Schlesewsky's standpoint, too.[26] There are, however,

---

[24] The latter appears to be a kind of "dialectical hermeneutics":
> "[…] when working with informants, it is very helpful to try to gain insight into what has moved them to accept or reject examples. To this end, it is highly advisable for the experimenter to engage in a discussion with informants. Often informants have a keen meta-linguistic sense of what is wrong with sentences they are presented with, and it can be extremely helpful for the linguist to be privy to these meta-linguistic judgments […]."(Dikken, Bernstein, Tortora & Zanuttini 2007: 349)

[25] "I find it surprising that most linguists subscribe to the view that grammaticality is only one of the factors contributing to acceptability while at the same time there has been not much research investigating the role of these factors. I find it also surprising that many (most?) syntacticians working with the results of rating experiments feed these directly into their proposals for a grammar. One can get good models for 'naturalness' if one has a constraint system that approximates the patterns of acceptability judgements gained in experiments, but such models are definitively not *grammars*." (Fanselow 2007: 363; emphasis as in the original)

[26] "I am reluctant to accept the idea that those participants who gave a mean rating below 1.5 simply failed to express their 'real' 4.72 assessment of (4) because of some noise

besides similarities, differences between Fanselow's and Bornkessel-Schlesewsky & Schlesewsky's attitude towards (SVLD), too.

In particular, at first glance Fanselow seems not to accept (SVLD)(a). He emphasises the importance and merits of native speakers' and linguists' judgements, of corpus data and of data gained from experiments (Fanselow 2007: 353f., 356) alike. He seems to oppose (SVLD)(b) as well, because he considers the decisive aspect of the acceptability of data to be their treatment rather than their origin. His paper contains, however, claims that seem not to be compatible with this standpoint. On the one hand, he stresses the significance of introspective data in the development of generative linguistics and argues for their reliability. He states that cases of "unconscious wishful thinking" can be detected easily and that disagreement about data can be cleared up, too (cf. Fanselow 2007: 354). Moreover, he calls the arguments in Postal (1974) "clear evidence" (cf. Fanselow 2007: 356) that cannot be contested. On the other hand, he refuses the "'standard' chaotic technique of data collection" (Fanselow 2007: 361), and regards not only reliance upon single judgements as insufficient but also maintains that results representing the average of acceptability judgements of a group of speakers are unsatisfactory.[27] He does not clarify why linguists' intuitions can be regarded as a data source which is more reliable than that of naïve (i. e. non-linguist) informants.

From this it follows that his position on (SVLD)(c) is rather vague. On the one hand, he expressly transgresses the latter:

> "[t]he empirical problems of syntax do not go away with controlled experiments. First, the difficulty in finding the appropriate syntactic model rarely lies in unclear judgments for the key empirical domains, rather, it stems from the fact that the clear data in one domain often suggest generalizations incompatible with the clear data in another domain. Second, the problem of conflicting evidence within a domain does not automatically disappear by the employment of controlled acceptability rating experiments." (Fanselow 2007: 356)

According to Fanselow, the inconsistency between data sets can emerge from several sources: from different experiments examining the same type of construction, or from the application of different empirical methods. It is the researchers' inability to keep all relevant factors constant that usually lies behind such "seemingly incompatible results". In these cases the experiments have been

---

effect […]. I think there is a better way of describing the situation: for some speakers, non-extraposed relative clauses in verbfinal structures are simply unacceptable (perhaps, because they focus on their prosodic oddity) for others, they are perfect (perhaps, because prosodic oddity plays no role in their judgements)." (Fanselow 2007: 359f.)

[27] "[…] when we observe that a certain structure comes with a certain mean acceptability, it is very difficult to tell how much of its acceptability is determined by its grammatical status. We will not find answers without experimental work." (Fanselow 2007: 363)

actually carried out on different materials (cf. Fanselow 2007: 357). Such factors may be the following: dialectal variation, social differences, the fact that "individuals may be liberal in one respect but very restrictive in another" (cf. Fanselow 2007: 360), different processing difficulties, context, prosody, presence or absence of grammatical competitors, frequency and typicality effects and further non-grammatical factors which influence the acceptability of linguistic constructions. However, he leaves the question of whether "acceptability experiments" are or will in near future be capable of controlling these factors open.

On the other hand, however, as we have already mentioned, he presupposes the existence of "sharp data", "clear cases" or "clear evidence", which are unambiguous and – to develop his metaphor – feed grammar directly:

> "I can see no unclear data the resolution of which would allow us to decide between different grammatical models (as envisaged by Featherston). Of course, there are many open issues at a level below the identification of correct syntactic models, and a number of them hinges on unsharp data, and here, controlled experiments have indeed been helpful. […] Once a syntactician decides to venture into the domain of unclear judgments, experiments for improving data quality are indeed mandatory – if only because they make it very explicit what is judged, and under what circumstances." (Fanselow 2007: 356)

In sum, we obtain a contradictory picture: while judgement data are problematic and need experimental clarification, they are in certain cases unproblematic, "sharp" and totally reliable.

Fanselow does not make his attitude towards (SVLD)(d) explicit. There are some points in his argumentation which indicate that in the case of "sharp data" he presupposes a unidirectional relationship between data and hypotheses (data confirm or refute the given hypothesis). As regards "unsharp" acceptability judgements, however, there is a more complex process needed, where with the help of a series of experiments one can identify and neutralise the influence of extragrammatical factors. As for (SVLD)(e), he thinks that "sharp" data, that is linguists' judgements, may serve as "clear evidence" which allows us to decide between competing grammatical models. "Unclear judgements" cannot be regarded as directly given, because they do not mirror grammatical competence directly but contain several extragrammatical influencing factors. Consequently, without improving their quality, they are not capable of playing the role of linguistic evidence.

### 3.5. "Deep insight into reality instead of taking more care with data"

Grewendorf charges Featherston with the underestimation of the usability, the importance and the "empirical relevance" of single introspective judgements. He also maintains that Featherston overestimates the reliability and the merits of ex-

perimental evidence collected from groups of informants (cf. Grewendorf 2007: 371, 375). His argumentation boils down to three main claims. The first of them is that the practice of relying on linguists' introspective judgements is a legitimate, empirically adequate and extraordinarily productive method.[28] Second, data collected from multiple informants are by no means as reliable as Featherston claims; they may be problematic because, for example, the task has not been formulated carefully enough and it can be misunderstood by the informants (cf. Grewendorf 2007: 372), or because the task is too complicated to be understood properly by naïve informants at all (cf. Grewendorf 2007: 377). Third, the data type preferred by Featherston is not capable of illuminating the I-language of individual speakers and is practically useless.[29] From this Grewendorf concludes that experimentally gathered judgements may be helpful only in revealing the factors which may affect the judgement process (Grewendorf 2007: 370).

He criticises Featherston at a metatheoretical level as well and charges him with naïve inductivism:

> "He [Featherston] claims that linguists need to look at the data first and develop their models afterwards. *But this is not how science works.* Data is not a pre-condition for theory except in the trivial sense that the theory is about some phenomena in the real world. It is more often the case that the theory is a precondition for the data, not only in the rather trivial sense that there are no theory-neutral data, but also in the sense that the theory makes data visible which have not been noticed before. The Principle and Parameter approach provides a good illustration of this situation in linguistics. From theoretical constructs like the ECP, a mass of data has emerged that otherwise would have never been observed. Data and serious experiments are always theory-guided rather than the other way round." (Grewendorf 2007: 377f.; emphasis added)

His criticism goes even further because he accuses Featherston of naïve falsificationism as well.[30] He argues that counter-examples do not refute theories immediately but sometimes have to be disregarded in order to make scientific pro-

---

28 "It surprises me that the field of generative syntax is claimed to produce unsatisfactory theoretical work when making use of the empirical method of individual linguistic introspection. […] The Principle and Parameter approach has permitted many subtle and revolutionary discoveries over a broad domain of phenomena in a wide range of languages such that linguistic theory for the first time in its history came close to its crucial objective of explanatory adequacy." (Grewendorf 2007: 370)

29 "The essentials suggested by Featherston should of course be requirements of good scientific practice. Nevertheless, it may eventually turn out that differences in grammaticality judgements between a group and individual linguist cannot be attributed to 'inadequate research practice' of the latter but clearly exhibit differences between I-languages. In this respect, the grammatical intuitions of the individual cannot be falsified by the results of acceptability experiments carried out with a group." (Grewendorf 2007: 376)

30 "[…] Featherston seems to advocate a naïve Popperian stand when he states that we cannot make any progress if we cannot test and discard hypotheses." (Grewendorf 2007: 378)

gress possible. That is, empirical adequacy is only one (violable) requirement among others.[31]

Referring to the idea of science "Galilean style" mentioned in Chomsky (2002: 102), he evaluates Featherston's objective to increase data quality in generative linguistics as completely mistaken:

> "Instead of taking more care with the data and collecting reliable data, scientists often just disregard the phenomena in order to find principles that seem to give deep insight into reality, where reality is taken to be represented by the abstract systems that are constructed rather than by an unstructured conglomeration of phenomena." (Grewendorf 2007: 379)

Grewendorf expressly advocates (SVLD)(a)-(c), since he regards only introspective data as acceptable and he does not raise any doubt about the reliability of this data type. He jettisons (SVLD)(d) and (e), deeming inductivism and falsificationism untenable alike. He recognises the theory-dependency of data and regards inconsistency-tolerance as a legitimate strategy. Despite this, his criticism of Featherston's proposals and the alternative he sketches are in certain respects problematic. First, an experiment based on judgements of multiple native speakers cannot be reduced to theory-independent naïve inductivism and compared to taking "a videotape of things happening out the window", nor blamed as "unstructured conglomeration of phenomena" (Grewendorf 2007: 378f.; see also footnote 9). Second, at first sight the very relationship between data and theory he alludes to in the first quotation above might go beyond the standard view of the analytical philosophy of science and might remind us of Pullum's proposal to apply the idea of reflective equilibrium to linguistics (Kertész & Rákosi 2008b: 3.1). However, on closer inspection this impression turns out to be wrong. Namely, on the one hand, the characterisation of linguistics as a "science of Galilean style" and the principle that we do not have to regard consistency between data and the hypotheses of the theory as a requirement that has to be fulfilled at every stage of scientific theorising, seem to indicate a break with (SVLD)(d). On the other hand, however, Grewendorf's refusal to take into consideration *all* data available and his neglect of the search for new and relevant data sources may easily lead to the loss of relevant information and circularity. Temporary inconsistency-tolerance must not mean simply ignoring counter-examples. One is allowed to set data apart only after examining the available possibilities for inconsistency resolution and finding them unsatisfactory or at least no better than alternative proposals. This means that advocating "science Galilean style" is not possible without a sophisticated methodology for the treatment of inconsistencies. Such a methodology has to involve, among

---

[31] "There are numerous examples in the history of the natural sciences […], which show that criteria such as simplicity, elegance, and fruitfulness are as important in the assessment of theories as are reliable data." (Grewendorf 2007: 379)

other things, tools for the reconstruction of inconsistencies, for their resolution or (temporary) tolerance etc. In addition, he also mentions the importance of the simultaneous and balanced fulfilment of principles for theory formation when he states that "[…] criteria such as simplicity, elegance, and fruitfulness are as important in the assessment of theories as are reliable data". However, in this case, too, subtle methodological considerations would be necessary so as to clarify the very relationship between these principles and data in theory formation.

### 3.6. Introspective data "do not meet the criterion of observational truthfulness"

Haider agrees with Featherston's critical remarks on the poor standards of data assessment in generative linguistics.[32] According to Haider, there is a close connection between observational and descriptive adequacy. When relevant data do not fit a hypothesis, linguists usually try to protect the latter with the help of subsidiary assumptions or even by ignoring data that conflict with it. Thus, problems with descriptive adequacy lead to the lack of observational adequacy.

He is, however, dissatisfied with Featherston's proposals. He misses "careful experimental studies", relying on an "established" data base instead of theory-driven or eclectic data selection (cf. Haider 2007: 388). Since introspection cannot be regarded as a "reliable source of evidence", judgement data may play only a heuristic role.

In particular, his central thesis is that for the justification of theories, testable data are needed, which means that data have to be independent of theories:

> "A welcome fringe benefit of a change in attitude would be *the absolute obligation to provide testable evidence* for one's favourite novel theoretical ideas. Generative Grammar is not free of post-modern extravagances that praise an extravagant idea simply because of its intriguing and novel intricacies as if novelty and extravagance by itself would guarantee empirical appropriateness. […] Contemporary papers too often enjoy *a naive verificationist style* and seem to *completely waive the need of independent evidence* for non-evident assumptions. The rigorous call for testable and successfully tested independent evidence is likely to disturb many playful approaches to syntax and guide the field eventually into the direction of a serious science. At the moment we are at best in *a pre-scientific phase* of orientation, on the way from philology to cognitive science." (Haider 2007: 389; emphases added)

---

[32] "Data handling in Generative Grammar is far from satisfactory, both in terms of observational and descriptive adequacy. […] A failure on the observational level is in most cases the misidentification of a side effect as the main effect. […] Missing the level of descriptive adequacy is missing the adequate grammatical generalizations for a given set of facts." (Haider 2007: 381f.)

As the quotation witnesses, Haider's views conform to (SVLD), although his opinion is exactly at the opposite extreme to the tenets advocated by Grewendorf.

He holds both (SVLD)(a) and (b), but he deems only experimental data legitimate and sharply rejects introspective data. He considers the latter as subjective and unscientific because there is no guarantee that one's theoretical assumptions do not influence them, and allows only independent (that is, objective) and observable facts to serve as data in scientific theories.

He seems to accept (SVLD)(c), too, because he recommends only a simple methodological rule for achieving observational adequacy, which for him is the most important condition of descriptive adequacy:

> "It is evident that careful evaluation and testing of data is the proper measure for reducing this kind of mistakes." (Haider 2007: 381)

Haider remains within the boundaries of (SVLD)(d) as well because he demands that data be independent of the theory and imposes the strict separation of the collection of the data base and the theory. He seems to acknowledge the separation of the "context of discovery" and the "context of justification" since he assigns data the role of testing the theory ("context of justification") and, as already mentioned, attributes to introspection only a heuristic role ("context of discovery"):

> "If you want to publish a paper in a psychology journal of some renown, you have to carefully explain how you gathered, organized and statistically analyzed the data before you start to develop your theoretical claims. Introspection is not considered a reliable source in cognitive science. Its role is in heuristics, but not in data assessment." (Haider 2007: 388)

From this it follows that his interpretation of the notion of evidence also remains within the boundaries of (SVLD): he speaks of "observable facts" and "independent evidence" – that is, he adheres to the standard view's requirements of objectivity, certainty and observability as we summarised in (SVAPS) (see Kertész & Rákosi 2008b). Although he remarks that linguistics is similar to psychology in so far as both of them "have to rely on secondary evidence for their primary target of explanation, namely the functioning of the mental capacity that produces that data" (cf. Haider 2007: 388), he does not explain how it is possible to treat such secondary data indirectly (that is, without any theoretical link and interpretation) as confirming or falsifying evidence.

## 3.7. Summary

We can summarise the results of our analysis of the responses to Featherston's target article as follows.

*Ad (SVLD)(a).* There is no consensus among the authors about the question which data types count as reliable in generative linguistics. Most commentators defend the legitimacy of introspective data. Some of them argue that only individual judgements are usable. There are more liberal attitudes as well that do not restrict the scope of acceptable data to judgement data. Not even these can be regarded, however, as a clear abandonment of (SVLD)(a), because they confine themselves to a few remarks vaguely indicating that perhaps the diversity of linguistic data and the corresponding methods of data collection may be legitimate and even desirable. Nevertheless, even in these cases the acknowledgement of diversity remains inconsequent in so far as they clearly prefer one data type over the others. Only one author argues against the rejection of every kind of introspective data.

*Ad (SVLD)(b).* With minor refinements, all contributors emphasise the importance of the data source in the evaluation of data. There are only a few hints at other aspects of data such as those revealed in the papers that were analysed in Kertész & Rákosi (2008a); only the indirectness and the complexity of judgement data have been touched on.

*Ad (SVLD)(c).* While Featherston (2007: 279) considers the data handling technique of generative linguistics as highly problematic and concludes that "there is little feeling of advance or questions being definitively answered", the majority of the commentators maintain more or less expressly the opposite. They claim that the data handling techniques of generative grammar are – although some modest improvement may be welcome – much less problematic and therefore there is no denying the revolutionary empirical findings of the generative enterprise. In this respect, these commentators assume a position close to (SVLD) with very little inclination to accept the significance of the problems Featherston has raised. However, several authors emphasise the need for refined data handling methods, resulting from the demand to represent and handle the complexity inherent in introspective data as properly as possible. Despite this, the latter position cannot be considered as a total break with (SVLD)(c) because they judge the weak points of the data type they prefer as less or even not at all problematic and tend to treat this type of data as certain facts. In sharp contrast to this attitude, Haider (2007: 387) labels certain features of the current data handling practice of generative grammar as "pre-scientific" and hopelessly subjective.

*Ad (SVLD)(d).* Most authors reject the idea of a linear relationship between data and theories and depict them as interrelated in a complex manner. They recognise the theory-dependency of data and the insufficiency of verification or

falsification. For all that, they do not realise the necessity of developing a meta-theoretical model that is capable of representing the complicated mechanisms of treating the uncertainty of data.

*Ad (SVLD)(e).* In some commentaries the role of evidence as confirming or disconfirming the hypotheses of generative grammar is still assumed, very much in the spirit of (SVAPS), while others clearly reject this aspect of (SVLD)(e). Despite this, there is no contribution which would completely go beyond (SVLD)(e). The reason for this lies in their ambiguous position on the interpretation of evidence as certain and directly given data. In connection with judgement data, the uncertainty and indirect nature of data is emphasised, while the preferred data type is regarded as certain and totally reliable.

## 4. Conclusions

In sum, our findings yield the following solution to (P):

(SP) (a) The contributions do not differ radically from those analysed in Kertész & Rákosi (2008a and b) because they are similarly double-faced in the sense of (SP1) and (SP2), although there is a significant difference between the former and the latter in degree and kind.

(b) The target article's most important innovative insights are:
- Both introspective and corpus data have their own role in linguistics and may be legitimate under given conditions.
- The way introspective data are used in generative grammar is anything but unproblematic.
- The generation of hypotheses should be focused on, rather than their testing.
- Introspective data are anything but the firm empirical basis of generative linguistic theory.

(c) As for the commentaries, the weight of innovative elements is clearly smaller. The most innovative insights are the following:
- Reference to the non-linear relationship between data and theory.
- Rejection of evidence as confirming/falsifying/verifying instances, weakening them in this respect and attributing to them considerably more sophisticated functions.
- Emphasis on the multidimensionality of introspective data and attempts at the combination of different methods to reveal and control influencing factors.
- The insight that inconsistencies are not only inevitable and omnipresent, but they may be important driving forces of scientific theorising.

(d) Nevertheless, even the innovative insights just enumerated are fragmentary:
  – The mechanism according to which data and theory may be interdependent has not been clarified.
  – It is not clear which data may count as evidence if all data types are influenced by factors which make them only more or less reliable but not certain.
  – The need to clarify the mechanisms of inconsistency-toleration and -resolution has not been realised.

Although Featherston intends to improve linguistic practice and his insights are undoubtedly innovative, he does not realise that one of the necessary preconditions for achieving the improvement of research practice in generative grammar is the elaboration of *a new metatheoretical perspective* which definitely breaks with the mechanical application of the standard view of the analytical philosophy of science to linguistic theories. Similarly, whether or not the commentators agree with Featherston's particular suggestions, they do not raise this issue explicitly, either. How our findings summarised in (SP) may motivate an adequate metatheoretical model of linguistic data/evidence, will be tackled in another publication of ours (Kertész & Rákosi 2008c).

## References

Featherston, S. (2007): Data in generative grammar: The stick and the carrot. In: Sternefeld (ed.)(2007), 269-318.

Bornkessel-Schlesewsky, I. & Schlesewsky, M. (2007): The wolf in sheep's clothing: Against a new judgement-driven imperialism. In: Sternefeld (ed.) (2007), 319-333.

Chomsky, N. (2002): *On nature and language*. Cambridge: Cambridge University Press.

Dikken, M. den, Bernstein, J.B., Tortora, Ch. & Zanuttini, R. (2007): Data and grammar: Means and individuals. In: Sternefeld (ed.)(2007), 335-352.

Fanselow, G. (2007): Carrots – perfect as vegetables, but please not as a main dish. In: Sternefeld (ed.)(2007), 353-367.

Grewendorf, G. (2007): Empirical evidence and theoretical reasoning in generative grammar. In: Sternefeld (ed.)(2007), 369-380.

Haider, H. (2007): As a matter of facts – comments on Featherston's sticks and carrots. In: Sternefeld (ed.)(2007), 381-394.

Kertész, A. & Rákosi, Cs. (2008a): Daten und Evidenz in linguistischen Theorien: Ein Forschungsüberblick. This volume.

Kertész, A. & Rákosi, Cs. (2008b): Conservatism vs. innovation in the (un)grammaticality debate. This volume.

Kertész, A. & Rákosi, Cs. (2008c): *Data, evidence and plausible argumentation in linguistics*. In preparation.

Newmeyer, F. (2007): Commentary on Sam Featherston, 'Data in generative grammar: The stick and the carrot'. In: Sternefeld (ed.)(2007), 395-399.

Nola, R. & Irzik, G. (2005): *Philosophy, science, education and culture*. Berlin, Heidelberg & New York: Springer.

Sternefeld, W. (ed.)(2007): Data in generative grammar (= Theoretical Linguistics 33(3)).