

Conservatism vs. innovation in the (un)grammaticality debate¹

András Kertész and Csilla Rákosi

“Corpus linguistics doesn’t mean anything.”
(Chomsky in: Andor 2004: 97)

“Linguistics will not move forward healthily until the generative approach is bypassed as the pseudoscience it is, with its remaining practitioners retiring from academic life and new recruits to the discipline ignoring its ethos, assumptions, and alleged results.”
(Sampson 2007b: 122)

1. Introduction

In Kertész & Rákosi (2008a) we started our analysis of current discussions about data and evidence in linguistics by referring to the fact that *the standard view of the analytical philosophy of science* played a decisive role in the development of theoretical linguistics. One of the central ideas of the standard view of the analytical philosophy of science was that *empirical* theories must consist of statements which can be tested on the basis of a special kind of data, namely, *evidence*. According to this idea – to put it in a very simplified manner, as a first approximation – the intuition underlying the notion of evidence was the following:

- (E) Evidence
 - (a) is objective,
 - (b) serves as a neutral arbiter among rival hypotheses/theories,
 - (c) is expected to justify hypotheses/theories,

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.

- (d) is immediately given,
- (e) is primary to the theory,
- (f) is reliable.

We summarised the basic tenets of the standard view of the analytical philosophy of science from the point of view of the data/evidence problem in linguistics as follows:

- (SVAPS) (a) There is a clear distinction between “the context of discovery” (covering the creative, cognitive, social, historical etc. aspects of scientific theory formation) and “the context of justification” (the reconstruction and the evaluation of the results of the discovery process).
- (b) The task of metascientific reconstruction is the justification of scientific theories.
- (c) Basically, the analytical philosophy of science involves two main trends, namely inductivism (verificationism/confirmationism) and hypothetico-deductivism.
- (d) Evidence, as characterised in (E), is attributed special significance in the justification of hypotheses and theories.

Against this background, we showed that in the eighties-nineties of the twentieth century with respect to the treatment of the data/evidence problem, basically two camps emerged: on the one hand, the advocates of corpus data and on the other hand, generativists in favour of introspective data. Despite the considerable differences, these views share a series of assumptions which we called *the standard view of linguistic data* (SVLD):

- (SVLD) (a) Not all linguistic theories can be considered as empirical but only those that rely on the *appropriate type of data*. Only certain data types may be regarded as relevant and legitimate: in the generativist view it is introspective data, while according to corpus linguists it is primarily corpus data.
- (b) It is solely the *origin of the data* that decides what counts as the appropriate type of data.
- (c) Both camps regard their *own data-handling procedure as unproblematic*. They believe that a few general methodological rules are sufficient for controlling the reliability of the data.
- (d) They deem the relationship between the data and the hypotheses of the theory *unidirectional*, and *determined by general, strict methodological rules*. They explicate it either as *induction* from data to hypotheses, or as *deduction* from hypotheses to data.

- (e) *Evidence* is considered as an empirical datum playing a pivotal role. Its significance lies in its special function. Since it constitutes a directly (that is, without reference to any theoretical framework) given special subset of data, it is qualified as reliable. Thus, it can be treated as an unquestionable fact suitable for the justification (verification, confirmation, or falsification) of hypotheses.

We analysed a series of current approaches to linguistic data and evidence which we consider to be progressive because they try to transgress the limits of (SVLD). We showed that these approaches tackle the following problem:

- (P1) (a) What types of data do linguistic theories use, and what types of data should they use;
 (b) what data do they consider as evidence, and what data should be considered as such; and
 (c) what functions do they attribute to the latter, and what function should evidence be attributed to?

However, we argued that there is no avoiding the examination of the role meta-scientific considerations play in the solutions that have been proposed to (P1). Thus, we tried to find an answer to the following question:

- (P2) What is the role of metascientific reflection in the proposed solutions to (P1)?

We showed that in comparison with (SVLD), *all* the approaches, which we analysed in the paper mentioned, are characterised by *two kinds of double-facedness*. According to the first they are still close to (SVLD) in several respects, while they also attempt to go beyond its tenets in that they yield relevant new insights and shed fresh light on the nature of linguistic data and evidence:

- (SP1) (a) In accordance with (SVAPS), linguistic evidence is still interpreted as a specific subset of data which is not revisable in the theory at issue. It is supposed to comprise statements which are true with certainty and whose main task is to test the hypotheses.
 (b) In opposition to the standpoints in (SVLD), the new proposals we surveyed raise a series of innovative ideas. These progressive insights are, however, not acknowledged by all authors; and even if they are accepted by several of them, then in different forms, so that they cannot be generalised. Despite this, the following insights can be highlighted which appear in one or more contributions and are –

in comparison to (SVAPS) and (SVLD) – of a clearly innovative nature:

- The pluralism of linguistic data is acknowledged instead of preferring just one type of data.
 - Linguistic data are considered to be theory- and problem-dependent.
 - The source of data is not held to be decisive; rather, it is treated only as one of the relevant factors among others, such as the structure, the complexity, the directness, the abstractness of data etc.
 - The relationship between data and theory is assumed to be cyclic rather than linear.
 - It is admitted that data very often do not support or refute the hypotheses unanimously, but rather, generate contradictions.
- (c) Nevertheless, neither the relation between (a) and (b), nor the criteria which might facilitate the application of the assumptions summarised in (b) have been tackled.

The second double-facedness concerns the attitude of these approaches to metascientific reflection. On the one hand, all the authors realise the necessity of metascientific reflection. On the other hand, they do not recognise that the untenability of the standard view of the analytic philosophy of science necessitates a search for an alternative metascientific perspective:

- (SP2) (a) The views we have investigated *realise* the need for metascientific reflection on the structure and function of linguistic data and evidence.
- (b) This, however, does not go beyond the linguist's naïve (self-)reflection. Although the continuous, partly implicit, partly explicit confrontation of (SVAPS) with everyday research experience suggests the unacceptability of the former, the necessity to develop alternative metascientific tools has *not* been raised.

Our conclusion was therefore that in order to arrive at an adequate account of linguistic data and evidence there is no avoiding the elaboration of a new metascientific approach. Consequently, the progressive components we revealed in the views we analysed should serve as points of departure for the development of such a novel metascientific approach to linguistic data and evidence. The details of the latter have been outlined in Kertész & Rákosi (2008b).

The present article will continue the argumentation put forward in our previous analysis and thus provide further motivation for the new framework. It focuses on a debate which was not considered in Kertész & Rákosi (2008a). The discussion was published in a special issue of the journal *Corpus Linguistics and*

Linguistic Theory entitled *Grammar without grammaticality*. The reason why we decided to devote a separate article to it is twofold. First, the scope of this debate is considerably narrower than that of the papers we analysed previously, in that it is restricted to the questions of whether grammaticality is a well-motivated notion in linguistics and whether introspective data are empirical. Second, although in the contributions to this debate there are similarities with the papers we analysed previously, there are, as we will see, very important and far-reaching differences as well. In particular, while *all* of the works that were the subject matter of our previous paper are both close to (SVLD) in certain respects and break with it in other respects, in the (un)grammaticality debate the picture is more differentiated. *Some* contributions are *clear manifestations* of (SVLD). *Others* are similar to the papers we analysed in Kertész & Rákosi (2008a) in that they are double-faced in the sense (SP1). Moreover, although they are double-faced in a second sense as well, that is, regarding their relationship to metascientific reflection (cf. (SP2)), some of them make bolder suggestions to outline possible new metatheoretical perspectives.

Here we do not want to put forward a comprehensive analysis of this controversy. We will not evaluate the arguments pro and contra Sampson's target article. Rather, our main aim is to highlight those aspects of the debate that indicate progressive ideas, with special attention to insights differing from those we have already discussed at length in Kertész & Rákosi (2008a) and which might pave the way for the elaboration of an adequate metatheoretical model of linguistic data and evidence. We raise the following problem:

- (P) In what respects do the contributions to the (un)grammaticality debate as published in Stefanowitsch & Gries (2007) go beyond
- (a) (SVLD), and
 - (b) the views summarised in (SP1) and (SP2)?

Therefore, we will proceed in two simple steps. In Section 2 we will show in what way one part of the contributions represents (SVLD). In Section 3 we will highlight those suggestions in the rest of papers which we claim to be progressive and which may serve as points of departure for the future development of a suitable metascientific approach to linguistic data and evidence. Finally, in Section 4 we will summarise our findings.

2. The prevalence of (SVLD)²

In the debate there are three articles which are clear manifestations of (SVLD): Sampson's (2007a) target article with which the special issue begins, Stefanowitsch's (2007) contribution and Sampson's (2007b) reply to his critics.

The main thesis of Sampson's target article is relatively simple. The author summarises it in his reply as follows:

"I urged that a key tenet of generative linguistics may be a delusion. That doctrine is that the class of all possible sequences over the vocabulary of a language can be divided into two subsets, a set of grammatical sequences and a set of ungrammatical sequences, and that we can in principle establish where the dividing-line falls. I suggested that it might be more realistic to think in terms of a three-way classification, between (1) sequences for which speakers have already found uses, (2) sequences for which uses have not yet been found but which will one day find a use, and (3) sequences that will never have a use; and I suggested that the boundary between (1) and (2) + (3) might in principle be knowable but is of no intellectual interest, while that between (1) + (2) and (3) is not in principle knowable even approximately." (Sampson 2007b: 111)

Let us see, why this thesis is a clear manifestation of (SVLD).

First, in accordance with (SVLD)(a), Sampson considers *solely* that kind of linguistics as empirical that makes use of the type of data he thinks to be appropriate. In particular, Sampson clearly follows the requirement imposed by the standard view of the analytical philosophy of science that empirical theories must consist of statements based on or controllable by *objectively testable data* which can be regarded as *directly given*.³ Against this background, he puts forward his view that generative linguistics is not part of empirical linguistics, extremely sharply.⁴

² 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.

³ Objectivity is interpreted as interpersonal. See (E) above.

⁴ By way of illustration, see for example the second epigraph to this paper and the following quotations selected from numerous similar passages:

"For mainstream generative linguists, the undemanding nature of their research style is part of its attraction (cf. Postal 1988). You make up some research findings out of your head, publish them in a forum controlled by similarly non-empirical linguists, and after a few years you have got yourself a career." (Sampson 2007b: 122)

"It is startling to find 20th- and 21st-century scientists maintaining that theories in any branch of science ought explicitly to be based on what people subjectively 'know' or 'intuit' to be the case, rather than on objective, interpersonally-observable data." (Sampson 2007a: 14)

Second, (SVLD)(b) says that it is solely the *origin of the data* that decides what counts as the appropriate type of data. Sampson expressly rejects data types which cannot be regarded as *completely free of subjective factors*. From this he concludes that data in whose origin intuition has played any role have to be excluded from linguistic theories:

“What makes a theory empirical is that it is answerable to interpersonally-observable data. An utterance of ‘fourth floor’ by a department store assistant is an empirical datum; a speaker’s feeling that a given word-string is good or bad (acceptable or unacceptable, grammatical or ungrammatical, ...) is not an empirical datum, because another person cannot check that same feeling (though the other person may have a feeling of his own about the string, which may coincide or otherwise).” (Sampson 2007b: 115)

In this vein he judges not only introspective data of linguists that make use of the linguists’ own linguistic intuition untenable but he seems to deem experiments based on speakers’ intuitive grammaticality judgements unreliable as well. Although in his reply to the criticisms he admits that “people’s reports of their linguistic intuitions or introspections are themselves intersubjectively observable data, and it would be possible to construct empirical descriptions of those reports” (Sampson 2007b: 118), he believes that such data contain several factors which make them “chaotic, or even contradictory” (Sampson 2007b: 119). He builds his opinion upon the following arguments (cf. Sampson 2007a: 14ff.):

- The judgements of individual speakers differ considerably from each other. This is especially true of more complicated constructions that do not belong to core grammar.
- There are often significant dialectal or sociolectal differences in connection with the acceptability of a given linguistic form.
- Native speakers might be influenced by the grammar lessons heard at school, by their own “folk theories”, or by preferring forms that they think to be more prestigious.
- Linguists are often strongly influenced by the theory they want to evaluate with the help of the data. Consequently, they run the risk of misjudging the acceptability of the investigated forms in order to get support for their theory and to avoid its falsification.
- Linguists may judge a linguistic structure ungrammatical only because they simply do not think of possible uses in special contexts or in special senses.
- Corpora may contain linguistic forms which are in conflict with the linguist’s grammaticality judgements and refute his/her expectations or even firm convictions.

From this Sampson concludes that *all* data based on intuition are totally unreliable. He is not disposed to accept the use of introspection even as a secondary source in order to complete or control results obtained on the basis of corpus data.⁵ What is observable and objectively given, is almost solely corpus data, that is, for example, the presence or the relative frequencies of different linguistic constructions etc.

He sets up, however, very strict requirements even for the use of corpus data in so far as he forbids the reliance on intuition not only in the process of the production of data, but in the process of their interpretation and evaluation as well. From the fact that grammaticality cannot be directly observed – because there is no connection between (observable) frequency of constructions and their grammaticality – he concludes that research concerning questions of grammaticality must not be part of linguistics.⁶

Third, Sampson clearly regards corpus data as *unproblematic* exactly as (SVLD)(c) says. He seems to be convinced that a few general methodological rules are sufficient for the control of the data and that objectivity (interpreted as interpersonality) and observability can be easily guaranteed. While he fiercely criticises introspective data and the grammaticality-ungrammaticality distinction, he rejects every objection put forward by his critics concerning the reliability and the usability of corpus data. For example, he characterises his own corpus by simply declaring that

“[...] the SUSANNE annotation scheme approaches the limit of humanly-possible grammatical-annotation precision. Thus, grammatical diversity in the SUSANNE treebank is likely to be genuine, rather than a product of inconsistent annotation behaviour by analysts.” (Sampson 2007a: 7)

Fourth, in the vein of (SVLD)(d), Sampson prefers the inductive method to the deductive one. Obviously, he assumes that the inductive investigation of corpus data will immediately yield hypotheses about particular properties of linguistic units. Thereby, deduction plays no role. In his view, there is no feedback between the hypotheses and the data, either; the relation between corpus data and

⁵ “My target article said that if we have a language-description based on empirical data, then an intuition-based description would be redundant; Hoffmann objects that when two people witness a crime, the police do not take just one of their statements. That is a false analogy: the two witnesses are logically equivalent in status. A better analogy is that, if we have used a ruler to measure the precise length of a line, we will not be interested in asking a person to estimate the length by eye.” (Sampson 2007b: 120)

⁶ “I believe that the concept of ‘ungrammatical’ or ‘ill-formed’ word-sequences is a delusion, based on a false conception of the kind of thing a human language is.” (Sampson 2007a: 1)

hypotheses is assumed to *be unidirectional*: the data – conceived of as unquestionable facts – have to confirm the latter.

Finally, Sampson does not define what he means by evidence. Despite this, we may risk the claim that he accepts (SVLD)(e) as well, because he attributes characteristics to corpus data which correspond to the ones stipulated for evidence by this requirement of the standard view.

As we have seen, Sampson urges the rigid and total rehabilitation of the tenets of (SVAPS). This means that not only the allowable data types and methodological tools but the scope of possible topics of empirical investigations into language, too, have to be narrowed down.

Stefanowitsch adopts a similar stance. Since he does not set forth his views but restricts himself to the elaboration of a case study aiming at reinforcing Sampson's hypotheses, we do not reflect on his paper.

In the next section, as already indicated, we will not go into an evaluation of the arguments put forward by Sampson's critics and Sampson's counter arguments. Rather, we will pick out those ideas which significantly go beyond (SVLD) and which may pave the way for a novel metascientific framework capable of handling the problem of linguistic data and evidence in a more adequate way than (SVLD).

3. Beyond (SVLD)

3.1. Linguistic data and reflective equilibrium

Among the critics it is Geoffrey Pullum who suggests *a comprehensive metascientific perspective* that differs significantly from (SVLD). Pullum analyses Sampson's arguments against the notion of (un)grammaticality and in favour of corpus data, and tries to show that these arguments are basically untenable.

Pullum's conclusion is a clear rejection of (SVLD)(a) in that he accepts neither the exclusion of introspective data on the basis of defending the sole application of corpus data, nor the exclusion of corpus data by arguing for the sole application of introspective data.

As regards corpus data, he seems to hold this data type to be the least problematic and most reliable.⁷ But he excludes neither experimental techniques nor introspection from the range of possible data sources although he thinks that

⁷ "I think corpus-based confirmations of syntactic claims can be enormously convincing." (Pullum 2007: 38)

"Corpus use is in a different league. I think computer searching of corpora is the most useful tool that has been provided to the grammarian since the invention of writing. Time and again I have found it to be extraordinarily important as an investigative aid." (Pullum 2007: 39)

these are not as reliable as corpus data.⁸ This clearly indicates that Pullum rejects (SVLD)(b) because he does not attach crucial significance to the origin of data. He mentions three reasons for this decision. First, he deems excluding questions about grammaticality and narrowing down the scope of linguistic investigations to relative frequencies of linguistic forms and similar phenomena unacceptable because he characterises linguistics as research into the “norms implicit in linguistic practice” (cf. Pullum 2007: 41). Second, he considers introspective data – even though their shortcomings cannot be completely eliminated – as *indispensable*.⁹ Third, he emphasises the presence of intuition, that is, of subjective factors in connection with *all* data types.¹⁰

This means at the same time that an especially important difference between Sampson’s and Pullum’s standpoint lies in their relationship to (SVLD)(c). While Sampson treats corpus data as entirely unproblematic and fully reliable, for Pullum *all data types are more or less problematic*:

“[...] I don’t think the how-does-it-sound-to-you-today method can continue to be regarded as a respectable data-gathering technique. [...] For quite a long time now, this head-tilting grammatical investigation, this divination by consulting the inner ear, has been discrediting theoretical syntax. *The trouble is, a switch to naive or absolutist reliance on other techniques produces little improvement.*” (Pullum 2007: 38f.; emphasis added)

Pullum (2007: 39) comes to the conclusion that “a more sensible set of ideas about the epistemology of grammar is needed”. The key point of this *methodological turn* consists in abandoning (SVLD)(d). The relationship between the data and the hypotheses of the theory is no longer regarded as a *unidirectional* process, *determined by general, strict methodological rules*. Rather,

“[...] the process of settling on a certain set of claims about the rules of grammar is still one of *iterated cross-checking*. The linguist tentatively formulates some proposals about the rules and checks their consequences against *intuition* (the linguist’s own, or those of other speakers, or other grammarians). Those are cross-checked against the evidence of

⁸ “Occasionally people suggest that survey work – administering questionnaires about grammaticality – would be an appropriate empirical basis for syntax, and I do know of cases where well-conducted surveys have produced very useful results.” (Pullum 2007: 39) He also quotes another work of his according to which “[t]he evidence we use comes from several sources: our own intuitions as native speakers of the language; the reactions of other native speakers we consult when we are in doubt [...]” (Huddleston & Pullum 2002: 11).

⁹ “Speakers’ judgments of grammaticality are, under suitably optimal conditions, an important source of evidence concerning the content of those norms, though in practice they are by no means infallible, or even largely reliable.” (Pullum 2007: 41)

¹⁰ “It may well be that Sampson believes linguistics should be reconstructed on an entirely judgment-independent basis, making linguistic analysis operationalizable, even mechanizable. If so, I do not agree.” (Pullum 2007: 41)

what *corpora* contain. The latter may quite often change the linguist's mind about the former. Sometimes it will be clear that the current proposal about the rules should be *revised*. Other times a cluster of rules may start to look stable enough that it seems preferable to *rethink* the issue of grammaticality for certain sequences. That might mean *another* corpus check to re-stimulate intuitions or provide a corrective. And sometimes things may be rather more subtle: the linguist finds that a distinction was missed, and while the rule is right about *one* class of cases, it is wrong about *the others*, for which *a different* generalization holds." (Pullum 2007: 40; emphasis added)

Pullum suggests the application of the method of "reflective equilibrium" to the treatment of linguistic data. The main idea behind the former goes back to Goodman (1955) and is generally regarded as a radical break with the standard view of the analytical philosophy of science. Although now the notion of reflective equilibrium is made use of primarily in ethics, originally Goodman (without using the term) introduced its mechanism to justify induction and it was therefore of a clearly epistemological nature. The basic aspects of reflective equilibrium can be summarised in the following points (see also Daniels 2003):

- The method of reflective equilibrium aims at establishing *coherence* between our judgements about particular instances, our assumptions concerning the rules and principles underlying these instances, and the theoretical hypotheses expected to account for them.
- This is achieved by a *cyclic process* of continuously revising our previous beliefs in the light of later ones in the course of feedback mechanisms until at least temporary coherence is established.
- This is possible only if we continuously *reflect on* all the pieces of knowledge at our disposal and the relations among them. Accordingly, reflexivity is an indispensable constituent of this mechanism.
- The coherence view associated with the method of reflective equilibrium is in sharp contrast with the foundationalism of the standard view of the philosophy of science. *There are no privileged beliefs* one accepts at the outset. Rather, it is the overall process of achieving reflective equilibrium that decides what beliefs can be assumed to be acceptable at a certain stage of inquiry.
- Accordingly, in principle all beliefs are *revisable*.
- The *rules of inference* can be justified by bringing them into reflective equilibrium with particular cases of inference assumed to be acceptable.
- However, it is not only the rules of inference that can be corrected by achieving their coherence with inferential practice, but *inferential practice* can also be revised in the light of the tentatively accepted rules of inference. This means the constant, complex revision of all components of the system of our beliefs.

Pullum argues for the application of this metatheoretical model of linguistics as follows:

“The goal is an *optimal fit* between a general linguistic theory (which is never complete), the proposed rules or constraints (which are quite as conformant with the general theory as we would like), the best grammaticality judgments obtainable (which are not guaranteed to be veridical), and facts from corpora (which may always contain errors). *All are revisable, at every point.* [...] There are no one-answer-fits-all solutions to the problems of knowledge acquisition in our subject.” (Pullum 2007: 39; emphasis added)

It is very important to see how and to what extent Pullum’s proposal goes beyond (SVLD)(e) as well. Since all beliefs can and should be revised in the light of other beliefs with which they are to be brought into reflective equilibrium, all kinds of data are assumed to be problematic at the outset. It follows from these tenets that evidence is not something that is qualified as such by some prefabricated methodological principles. Rather, what information counts as evidence for or against certain hypotheses can be decided only relative to the given state of reflective equilibrium.

Despite the definite and firm break with (SVLD), however, there are also remnants of the latter in Pullum’s considerations:

- The text contains traces of the terminology of the standard view of the analytical philosophy of science, because Pullum speaks for example of “intersubjective checkability”, “potential falsifiability”, “actual facts of usage” (39ff.). It is clear that the original (and widespread) meaning of these terms cannot be reconciled with the assumptions and spirit of “reflective equilibrium”.
- Goodman’s model is a coherentist approach. Despite this, Pullum treats corpus data as “facts” which “lend an empirical aspect to the project”. Therefore, this point needs further clarification.

3.2. The middle course between the extremes

Meurers’ argumentation is similar to Pullum’s standpoint in several respects but there are some remarkable differences as well. As regards (SVLD)(a), Meurers definitely votes against the extremes of (SVLD), too:

“Arguably the biggest mistake of generative linguistics was to throw out most sources of empirical evidence, from rejecting the use of corpora to the downplaying of historical linguistic evidence or observations about variation in language use. Replacing the narrow generative perspective with a view that is narrowly corpus based, in which the only evidence considered is the language which happens to be used in a given corpus, would amount to replacing one half-blind extreme with another. *Any single methodology can*

mislead, overemphasize some aspects and hide other aspects from view. So while the use of corpus data has much to contribute towards empirically grounding theoretical linguistic research, *the key to a successful linguistic science arguably lies in a plurality of methods.*" (Meurers 2007: 54; emphasis added)

He advocates a broad range of possible data sources such as introspection, corpora and their quantitative analysis, psycholinguistic experiments, perception experiments, neurolinguistic imaging techniques etc.

He opposes (SVLD)(b) as well because he thinks that the origin of the data cannot be regarded as a decisive factor. First, he states that the *sine qua non* of empiricalness is that the theory has to make "testable predictions" (cf. Meurers 2007: 49). The falsifiability of the hypotheses maintained by linguistic theories, however, is in his view closely related to the issue of grammaticality in that the latter is necessary to decide which sentences are permitted by the given theory. Second, he argues that grammaticality and the presence of individual linguistic constructions are by no means independent:

"[...] it is not as though all linguistics can do is observe people use and change language in unrestricted ways – there is a constraining topography underlying language use, and studying it can result in linguistic theories that make scientific predictions about possible language use and change." (Meurers 2007: 50)

That is, even corpus linguistics cannot be restricted to the study of relative frequencies. Third, relying on the linguist's linguistic intuition is unavoidable in the case of *all* data types because she has to select the data relevant to her research question (that is, she has to decide which utterances contain the structure investigated) and she has to make a decision whether a given utterance belongs to the language studied (that is, whether it is a correct sentence or not) (cf. Meurers 2007: 51f.). From this Meurers concludes that – as opposed to the view held by Sampson who, as we have seen, thinks that corpus data are totally free of subjective elements – there is no substantial difference between introspective data and corpus data:

"An arguably relevant difference between the use of intuition by the generative linguists Sampson discusses and his own use of intuition is the fact that his intuitions are about real-life examples from corpora, whereas the intuitions he discusses as problematic are based on handcreated example sentences." (Meurers 2007: 52)

As for (SVLD)(c), Meurers discusses the advantages and problematic points of both introspective and corpus data in greater detail and more deeply than Pullum. On the one hand, he clearly prefers corpora because they "play a major role as source of examples for theoretical linguistics" (Meurers 2007: 52). According to him, one reason for this is that since corpus examples are natural and contextualised, they may be helpful in the evaluation of examples obtained by

introspection. Another reason is that corpus data may exhibit a wide variation of lexical, syntactic, semantic and contextual properties, thus making it possible to arrive at more adequate characterisations of relevant properties of the linguistic construction investigated. Nevertheless, Meurers emphasises the limits of corpus data as well. The main problem is that corpora are finite and therefore they cannot provide evidence for all generalisations which the researcher wants to test with respect to the constructions in question. Another problematic issue is representativeness: corpora cannot always be helpful in deciding whether a construction is a rare phenomenon or not. Furthermore, grammaticality judgements, obtained by introspection, are needed for the evaluation of corpora. They are indispensable means of testing the predictions of syntactic theories. But they exhibit serious drawbacks as well. Introspective data as used by generativists may lead to unreliable results because of their isolation from any context and the subjective elements they contain. However, the latter can be considerably reduced by reliance on the experimental techniques. Another problem is that with the help of grammaticality judgements only a small part of the subject matter of linguistics can be captured.

Meurers' attitude towards (SVLD)(d) is the exact opposite of Pullum's. Pullum applied the notion of reflective equilibrium to the mechanism of evaluating linguistic data, clearly transgressing the borders of (SVLD). Meurers, however, shares the tenets of the hypothetico-deductive branch of the standard view of the analytical philosophy of science. He considers the relationship between the data and the hypotheses of the theory as *unidirectional*, that is, as (deductive) falsification.

As regards the question of what constitutes evidence, Meurers seems to remain close to (SVLD)(e). He considers evidence as an empirical datum that is intersubjectively testable (that is, observable), and qualified as reliable (gained by, for example, "well established and validated experimental methods", cf. Meurers 2007: 53), which can be treated as an unquestionable fact suitable for the evaluation of hypotheses (as he puts it, for "validation of linguistic theories"). With this, however, his contribution becomes inconsequential in so far as when characterising psycholinguistic experiments as "examples of 'interpersonally-observable evidence'" (cf. Meurers 2007: 53), he seems to disregard his arguments for the subjective factors inevitably present in all data types.

3.3. Grammaticality without invented sentences

Foster (2007) starts her contribution with the observation that in naturalistic corpora there are many ungrammatical sentences to which, however, due attention has not yet been devoted. Naturalistic ungrammatical sentences are important sources for different kinds of research such as, for example, language produc-

tion, language loss and language learning. Therefore, the distinction between grammaticality and ungrammaticality cannot be dispensed with (cf. Foster 2007: 73):

“What kind of evidence is needed in order to develop and test [...] the part of the grammar which describes ungrammatical sentences? Since this grammar is to form the basis of a parsing system, its description of ungrammaticality must reflect the kind of ungrammaticality that actually occurs in language. *This means that naturalistic ungrammatical sentences will be needed as evidence rather than imagined ones.*” (Foster 2007: 77; emphasis added)

Accordingly, Foster has a specific and intricate attitude to (SVLD)(a). On the one hand, she maintains that it is not invented sentences but naturalistic sentences documented in corpora that should serve as the basis for the investigation of the manifestations of language use mentioned, whereby sentences must be judged in context, not in isolation. Therefore, she definitely rejects the way generative grammar treats data on the basis of judging the grammaticality of *invented and isolated sentences* – at least, in investigations aiming at the elaboration of cognitively realistic models of grammar (cf. Foster 2007: 74). On the other hand, she also rejects Sampson’s practice of narrowing down linguistic research to the observation of relative frequencies of linguistic constructions and admits that reliance on intuition cannot be completely excluded. Quite the contrary is true: grammaticality judgements are indispensable when working with corpora. She emphasises that the decision as to whether a sentence is grammatical or contains an error is, at least to a certain extent, unavoidably subjective (cf. Foster 2007: 80).

Foster considers the origin of data to be a crucial factor in judging the acceptability of data types. That is, she remains within the limits of (SVLD)(b), although her standpoint is in a certain respect more subtle than that of (SVLD). She regards corpora the only possible data source in investigations of language processing. Despite this, she does not rule out grammaticality judgements entirely but she thinks that linguistic intuition cannot be treated as an independent data source. Rather, grammaticality judgements should be based on naturalistic data within context.

From this it clearly follows that in contrast with (SVLD)(c) Foster does not deem corpus data unproblematic. She argues for the importance of revealing and controlling the subjective elements which unavoidably occur in the elaboration and use of corpora. One of the dangers that one has to face is the *theory-dependency* of grammaticality judgements:

“The concept of ‘ungrammatical’ is, of course, a slippery one: it can be used in an everyday and in a theoretical sense, and coming up with a definition of the term will depend on what kind of linguistic information one believes should fall under the remit of grammar.” (Foster 2007: 79)

Nevertheless, after a detailed analysis of the problems stemming from the subjectivity coded in grammaticality judgements Foster introduces an astonishingly simple definition of ungrammaticality intended to be theory-independent. This states that a sentence is ungrammatical if two conditions are satisfied: the sentence contains an error and in this sentence every individual word is well-formed.

As for (SVLD)(d), Foster's model does not presuppose a unidirectional relationship between data and hypotheses of the theory but seems to rely on both induction and deduction. On the one hand, the hypotheses of the theory, that is, the rules of the grammar have to be based on generalisations over naturalistic data. On the other hand, it is naturalistic corpora that are intended to test the hypotheses. However, she does not allude to possible corrections, "iterated cross-checking" (cf. Pullum 2007: 40), that is, cyclic argumentation in order to handle the uncertainties arising, for example, from the subjective factors of grammaticality judgements.

Finally, for Foster evidence is basically problem-dependent because data from different sources are needed to investigate special areas of linguistic performance.

3.4. Introspection and corpus data as corroborating evidence

Citing Penke & Rosenbach (2004: 480), Hoffmann (2007) takes for granted that linguistics is an empirical science. He agrees with Sampson that *all* linguistic data have to be interpersonally observable and in this sense objective. To these assumptions he adds the requirement according to which all linguistic data must be capable of being measured by "valid and reliable" methods (see Hoffmann 2007: 87), and that the main goal of any kind of science is the achievement of explanations. These statements seem to indicate the uncritical acceptance of (SVLD)(a). Hoffmann, however, shares the views of Pullum and Meurers in so far as he explicitly rejects (SVLD)(a) maintaining that

"[...] introspection and corpus data should in fact be seen as 'corroborating evidence'."
(Hoffmann 2007: 89)

He shows with the help of a short case study that not only must both data types be regarded as fully legitimate, but there are cases when they are able to complete each other. This leads to results which could not have been achieved by relying on only one of them (cf. Hoffmann 2007: 93ff.). This strategy helps us to grasp the full complexity of linguistic phenomena:

"As usual, real life is much more messy than is dreamt of in our philosophy: languages have highly frequent grammatical constructions shared by a great number of speakers, but

there are also Dunster constructions which only a few speakers employ. While these are examples of perfectly grammatical structures which can be identified using a road-atlas compilation corpus approach, I have argued that carefully designed introspection experiments can yield further insights into the phenomenon of grammaticality: while there are clear-cut examples of ungrammatical structures [...], there are also constructions which occupy a middle position and exhibit a cline of acceptability which depends on their processing effort [...]. [...] at least these days we have various scientific means of describing and explaining these shades of grammaticality. One way of successfully doing so is by treating corpus and introspection data *not as mutually exclusive approaches but as corroborating evidence.*" (Hoffmann 2007: 96; emphasis added)

In contrast to (SVLD)(b), he does not evaluate the origin of data as decisive. Data may stem from corpora or be acquired by introspection and share two important characteristics. First, their use is inevitably based on subjective factors to a certain extent since introspection is needed to decide whether a sentence is grammatical or contains performance errors or is unacceptable for some reason. Second, the objectivity and reliability of the data used as evidence in linguistic research has to be secured by the application of a strict methodology. Contrary to Sampson he believes that carefully planned experiments processing grammaticality judgements of individuals are able to guarantee the same level of objectivity and intersubjective controllability as corpus linguistic methods.

From this we could draw the conclusion that – as opposed to (SVLD)(c) – neither data type can be considered as unproblematic. However, Hoffmann's stance resembles at this point Meurer's because he seems to think that the methodological tools of present-day linguistics secure total reliability and objectivity.¹¹

Hoffmann does not transgress (SVLD)(d), although he turns against Sampson, claiming that linguistic inquiry must not stop at collecting and documenting data, but should strive to explain them. He thinks that, accordingly, it is not enough to register the variability of certain constructions as Sampson does, but it is necessary to find explanations for their variability. Nevertheless, he seems to stick to the hypothetico-deductive methodology in connection with the

¹¹ He writes, for example, the following:

"Note that in this passage Sampson does not mention the fact that Cowart suggests imposing the standards of psychological experiments on introspection data, i. e., the selection of a representative sample of subjects, careful design of experimental materials, randomization of the order of stimuli, the use of fillers/distractors and the employment of statistical analysis. He also doesn't discard Cowart's approach because it doesn't yield valid, reliable and objective data. No, Sampson simply claims that the results from such studies are irrelevant. [...] Furthermore, in all of the above publications, the authors stress the potential confounding effect of subject's forming an implicit hypothesis about the experiment. Yet this is a standard problem in psychological testing which can easily be avoided by using more fillers than experimental stimuli (which consequently also reduces the effect of prescriptive grammar rules)." (Hoffmann 2007: 92)

relationship between data and theory. Similarly, his concept of evidence is explicitly identical with that of (SVLD)(e).

3.5. Methodological diversity in linguistics

Arppe & Järvikivi (2007: 100) enumerate a broad range of empirical data sources and corresponding methods: introspection, elicitation, many different sorts of corpus data, various types of psycholinguistic experiments, data from language acquisition and learning, neurolinguistic measurements, cross-modal studies etc. This means that – unlike (SVLD)(a) – they judge all data types typically made use of in linguistic research as basically acceptable.

They disagree with (SVLD)(b) as well because they do not think that the origin of the data is decisive. There are several facets of their argumentation that bear witness to this stance. First, language is an extremely complex phenomenon. Thus, every manifestation of linguistic behaviour is a relevant datum requiring special treatment which cannot be investigated in isolation but together with a series of connected phenomena:

“It should be obvious from this kaleidoscope of different representations and characteristics of human language that it is a multimodal phenomenon which we can expect to understand fully and comprehensively only by combining multiple methods and multiple sources of evidence, by scientists and practices from multiple disciplines.” (Arppe & Järvikivi 2007: 106f.)

This means that it is necessary to acknowledge the complexity of the subject matter of linguistics and, as a consequence of this, the *methodological pluralism* of linguistic research. As opposed to the methodological monism Sampson pleads for, such a methodological diversity may yield the combination of different methods and different sources of data.

Second, there is no principled difference between data types. Introspective data are empirical just like for example corpus or experimental data because our linguistic intuition is of an empirical nature, possessing the same strengths and weaknesses as other data types:

“[...] even if we can determine the correctness (or incorrectness) of certain clear cases with (what seems to approach) intuitive certainty (not conceptual certainty), it does not follow that we can do this for every case. Our intuitive ‘knowledge’ of the structures or rules (or however we want to represent regularities, standards, norms or conventions) of (our native) language is empirical, in that it is not only learnt but also adjusted by individual experience in use within a linguistic community, and contingent, in that it is not immune to (gradual) change and influences. Intuitive evidence is evidence, for sure, but it does not have a privileged status vis-à-vis other types of empirical evidence.” (Arppe & Järvikivi 2007: 102)

That is, introspective data are more or less reliable data reflecting one of the most important aspects of linguistic knowledge (questions about correctness vs. incorrectness of linguistic forms).

Third, intuitive judgements are one of the possible starting points of linguistic research and the most important guide during the construction and the testing of hypotheses. Their reliability – similarly to data stemming from other sources – has to be secured by using multiple methods (for example, investigation of speakers' self-correction, carefully designed experimental settings etc.):

“As linguistic beings we are blessed with a first-hand individual intuition of what language is in all its forms, which is a good starting point for our study of language, both as a source for hypotheses and as a guide in interpretation, but neither is intuition the one and only analytical tool we have at our disposition, and we should consider ourselves at least somewhat biased in our own, personal introspections of language. It is where intuition ends that the entire spectrum of empirical methods, each adapted to its own particular aspect of language, come to our assistance – and should be used to the fullest.” (Arppe & Järvikivi 2007: 107)

Fourth, data of different origins allow us to achieve results which could not be obtained from only one data type. For example, experimentation facilitates the delimitation of rare but fully acceptable constructions from rare but less acceptable ones. In this way, experimentation, combined with the use of corpus data, may provide a reliable research method for the judgements of grammaticality in less clear-cut cases as well (Arppe & Järvikivi 2007: 103).

Their relationship to (SVLD)(c) is as ambivalent as that of Foster's and Hoffmann's. On the one hand, they emphasise that linguists, when making decisions as to the correctness or incorrectness of the constructions they investigate, cannot dispense with their own linguistic intuition as a native speaker (cf. for example Arppe & Järvikivi 2007: 105). In this way, *subjectivity and uncertainty are inevitably coded into the use of all data types*.

On the other hand, they seem to overestimate the power of experimental and other techniques and the combination of multiple methods. Namely, the authors presuppose that these methods *yield uniform, objective and totally reliable results*:

“This type of multimethodological research in fact shows that experimental results converge with those extracted from corpora (e. g., Gries et al. 2005a, 2005b). Furthermore, Roland and Jurafsky (2002) have shown that when one controls for context effects, such as the previous discourse present in corpora in contrast to isolated sentences often used in experiments, or differences arising from the genre and topic differences of different corpora, both experimental and corpus-based observations provide uniform, consistent results – which is exactly contrary to what Sampson claims.” (Arppe & Järvikivi 2007: 104)

This is, however, very close to the standard view of the analytical philosophy of science according to which it is basically the experimental, statistical etc. methods of the natural sciences that lead to reliable findings and exclude subjective factors.

As for (SVLD)(d), the authors reject Sampson's extreme inductivism, according to which empirical linguistics must be restricted to the observation of spatio-temporal language use. Instead, they find both induction and deduction indispensable (cf. Arppe & Järvikivi 2007: 105). This means that they – like Foster and Hoffmann – break with (SVLD) as far as they presuppose *both* inductive and deductive steps in linguistic theorising. However, their approach does not integrate these two methodologies into the kind of cyclic feedback mechanism that strives to achieve the coherence among the pieces of knowledge gained from different sources by the use of different methods. Their concept of evidence is identical with (SVLD)(e), because they agree with Sampson in that linguistics as an empirical undertaking “must base itself on empirical, interpersonally-observable data” (cf. Arppe & Järvikivi 2007: 100).

4. Conclusions

In comparison with the approaches analysed in Kertész & Rákosi (2008a), the contributions to the (un)grammaticality debate can be characterised as follows.

Whereas none of the approaches we analysed there maintain all tenets of (SVLD), in the (un)grammaticality debate there are extreme manifestations of the latter. Sampson's two papers and Stefanowitsch' contribution correspond to a very rigid conception of inductivist empiricism restricted to corpus data, and these authors completely exclude introspective data from the realm of legitimate linguistic research.

All the other contributions to the (un)grammaticality debate show both remnants of (SVLD) and ideas that go beyond the latter. In this respect they are as double-faced as the papers analysed in Kertész & Rákosi (2008a). In addition, in several respects they go beyond the latter. Let us see in what respects these ideas seem to be fruitful and are, at the same time, continuations of (SVLD).

Above all, Pullum's suggestion of applying the method of reflective equilibrium to linguistic data and evidence is of *utmost importance*. It may undoubtedly provide us with an effective and substantial means to achieve the following. First, to *bridge the gap* between the two antagonistic positions mentioned in (SVLD)(a). Second, the *clear and radical rejection* of the anachronistic components of the standard view of the analytical philosophy of science. Third, to focus on the process of theory-formation and problem solving in linguistics and thus, linguistic *argumentation*. Fourth, to capture the substantial *uncertainty* of linguistic data and evidence instead of unrealistically claiming their certainty in

the sense of the standard view. Fifth, to account for the *diversity* of linguistic data and methods and the way these may *interact* in order to yield more adequate theories.

The rest of the papers also raise issues which are *highly progressive*, although to a different extent and in different ways.

There seems to be consensus among the authors that *the diversity of linguistic data* has to be acknowledged. The authors mention the following reasons why this diversity is acceptable:

- problem-dependency, that is, the specific characteristics of the problem investigated profoundly influence the choice of the relevant data;
- the application of more than one data type makes it possible to explain linguistic phenomena which could not be investigated relying on one data type alone; that is, they are capable of completing each other in this sense;
- all data types are basically empirical because linguistic intuition itself is empirical too; that is, there is no “ontological” difference between them;
- different data types may reflect specific aspects of linguistic behaviour, and, accordingly, all manifestations of linguistic behaviour are potential data;
- language is such a complex phenomenon that its manifestations cannot be investigated in isolation but only as interrelated parts of “a multimodal phenomenon”. Consequently, linguists are confronted with many data types right at the outset, irrespective of what particular aspect of linguistic behaviour they strive to explain.

In connection with the diversity of data *the plurality of methods* has been emphasised in some of the papers as well (e.g. Arppe & Järvi 2007). This plurality boils down to the need to combine different data types and methods, thus conducting *multimethodological research*.

The authors seem to agree that the origin of data is not considered to be decisive because with respect to all data types linguists have to rely on their linguistic intuition. Accordingly, all data types inevitably contain subjective elements. Therefore, all data types are more or less *unreliable*. The individual data types have their own strengths and weaknesses.

The theory applied influences the applicability and evaluation of data as well, accordingly, data are *theory-dependent as well*. There have been principally two strategies proposed to remedy the consequences of their unreliability. The first one pleads for the use of several data types simultaneously because in this way one can filter out the weaknesses of the individual data types on the one hand and, on the other, achieve a coherent set of data and hypotheses which, supporting each other, secure the maximal reliability of the information at our

disposal. Another strategy tries to reduce the factors that endanger the reliability of data in that it recommends the observance of strict methodological rules such as statistical tools, controlled experimental settings etc. Of course, these two proposals do not exclude each other.

Nevertheless, the contributors do not raise certain issues which the papers analysed in Kertész & Rákosi (2008a) discuss. Two deserve mentioning. One is the problem of *inconsistency* arising from the combination of different data types. Another is the question of the specific properties of *linguistic evidence* distinguishing the latter from the rest of data.

As regards the *survival of (SVLD)*, although the authors admit the inevitable presence of subjective and uncertain factors in connection with all data types, they simply presuppose that by using tools such as statistical methods and experimental techniques the reliability of the data can be secured.

Most authors mechanically adapt the notion of evidence from (SVLD). They consider linguistic evidence as a set of intersubjective and observable data whose function is to test the hypotheses of the theory. Nevertheless, this stance seems not to be compatible with the fact that introspection is acknowledged as a legitimate data source which affects the processing of all data types

All authors, except for Pullum, presuppose that the structure of linguistic theories is basically deductive, and accordingly, formal. This is problematic because the use of induction to gain hypotheses from the data, the uncertainty of the available data, and the resolution of inconsistencies arising from the use of different data types embody factors which are clearly not formal but make the consideration of the *content* of the hypotheses indispensable.

By way of summary, we have obtained the following solution to (P):

- (SP) (a) The contributions differ from those analysed in Kertész & Rákosi (2008a) primarily in that some of them are
- explicit manifestations of (SVLD) while others are more radically innovative;
 - nevertheless, even the innovative ones adapt aspects of (SVLD). They have in common with the papers analysed in Kertész & Rákosi (2008a) the insight that linguistic data
 - are theory-dependent,
 - diverse, and
 - indicate the need to combine different methods and data types originating in different sources.
- (b) The most important innovative insights going beyond the papers discussed in Kertész & Rákosi (2008a) are
- the cyclic argumentation process labelled as reflective equilibrium,

- the subjectivity, the unreliability and the uncertainty of all kinds of data.

Now the question is, what kind of metatheoretical framework the tenets in (SP)(b), along with the findings outlined in Kertész & Rákosi (2008a), can motivate. Of course, the answer would go far beyond the scope of the present paper and therefore we have to postpone its elaboration to subsequent publications (see Kertész & Rákosi 2008b).

References

- Andor, J. (2004): The master and his performance: An interview with Noam Chomsky. *Intercultural Pragmatics* 1, 93-111.
- Arppe, A. & Järvikivi, J. (2007): Take empiricism seriously! In support of methodological diversity in linguistics. *Corpus Linguistics and Linguistic Theory* 3, 99-109.
- Daniels, N. (2003): Reflective equilibrium. In: *The Stanford encyclopedia of philosophy*.
<http://plato.stanford.edu/archives/sum2003/entries/reflective-equilibrium/>.
- Foster, J. (2007): Real bad grammar: Realistic grammatical description with grammaticality. *Corpus Linguistics and Linguistic Theory* 3, 73-86.
- Goodman, N. (1955): *Fact, fiction, and forecast*. Cambridge: Harvard University Press.
- Hoffmann, T. (2007): “Good is good and bad is bad”: but how do we know which one we had? *Corpus Linguistics and Linguistic Theory* 3, 87-98.
- Kertész, A. & Rákosi, Cs. (2008a): *Daten und Evidenz in linguistischen Theorien: Ein Forschungsüberblick*. This volume.
- Kertész, A. & Rákosi, Cs. (2008b): *Data, evidence and plausible argumentation in linguistics*. In preparation.
- Meurers, W.D. (2007): Advancing linguistics between the extremes: Some thoughts on Geoffrey R. Sampson’s “Grammar without grammaticality”. *Corpus Linguistics and Linguistic Theory* 3, 49-55.
- Penke, M. & Rosenbach, A. (eds.) (2004): What counts as evidence in linguistics? Special issue of *Studies in Language*.
- Pullum, G.K. (2007): Ungrammaticality, rarity, and corpus use. *Corpus Linguistics and Linguistic Theory* 3, 33-47.
- Sampson, G.R. (2007a): Grammar without grammaticality. *Corpus Linguistics and Linguistic Theory* 3, 1-32.
- Sampson, G.R. (2007b): Reply. *Corpus Linguistics and Linguistic Theory* 3, 111-129.

Stefanowitsch, A. (2007): Linguistics beyond grammaticality. *Corpus Linguistics and Linguistic Theory* 3, 57-71.