

Tanulmány

Ágnes Abuczki

On the notion of grammaticality¹

Abstract

The monograph reviewed is concerned with methodological issues on the one hand, and raises fundamental theoretical and metatheoretical questions on the other, such as (1) what language is like, (2) how grammar unfolds in a child's mind and (3) what counts as suitable and reliable data in linguistic theories. The authors aim at contributing to the resolution of several empirical problems and current linguistic debates. The present paper provides a chapter-by-chapter overview of the book and an evaluation of their main ideas, focusing on the issue of the applicability of several data types as well as the restrictions and limitations of applying various methods and data in linguistic research.

Keywords: grammaticality, empirical linguistics, linguistic data

1 Background and general overview

The book under review, a collection of research papers written by Geoffrey Sampson (University of South Africa) and Anna Babarczy (Budapest University of Technology and Economics), is a volume published in the series of Trends in Linguistics Studies and Monographs (Mouton de Gruyter). The work presented in the monograph is the outcome of the collaboration of the authors at Sussex University in England, where Geoffrey Sampson was Professor of Natural Language Engineering, and Anna Babarczy was a Research Fellow.

One of the major ideas in the book is that many linguists have accepted a fundamentally inaccurate understanding of the nature of grammatical structure. As the authors point it out in the Preface, “the book sets out to explain and give evidence for a distinctive view of the nature human language, a view which contrasts sharply with assumptions that are taken for granted by many or most linguist publishing today” (v). The most provocative thing in the book is that the authors entirely reject the intuition-based generative paradigm (Chomsky 1957, Chomsky 1976) and fundamentally question the validity and relevance of such relatively widely accepted generative concepts as well-formedness and ill-formedness, as used by followers of prescriptive linguistics. It is suggested in several chapters that linguists should be concerned with “a realistic kind of descriptive linguistics”, rather than prescriptive grammar (Sampson & Babarczy 2014: 92). Their perspectives are predominantly computational linguistic; accordingly, they provide plenty of structurally annotated samples of real-life English usage. In their case studies the authors contrast various conflicting views about language

¹ Review article on: Sampson, G. & Babarczy, A. (2014). Grammar Without Grammaticality: Growth and Limits of Grammatical Precision. Berlin & Boston: Mouton de Gruyter.

structure, innateness theory, as well as the usability and acceptability of data provided by language production that can be captured in corpora, versus data provided by using linguistic introspection. Looking at corpus data in various languages as well as diachronic changes in language use, the authors come to the conclusion that it is worthless to try to define and talk about grammaticality since languages cannot be described and explained using "a common logic" (8), by which the authors mean something like universal structural rules.²

As the authors put it, the central aim of the book is "to establish a truth about language" (Sampson & Babarczy 2014: vi). This goal can scientifically be considered rather naïve as it intends to find "truth", a term generally avoided in theories of science.

In short, the book contains a great number of thought-provoking papers, presented in a logical and reader-friendly way. Grasping its notions is stimulated by a series of metaphors and analogies. From the point of view of its target audience, understanding various established linguistic theories is a pre-requisite to reading the book. Therefore, the volume can be recommended to graduate and PhD students of linguistics as well as linguists working in any field, fascinated by theoretical controversies, meta-theoretical and methodological issues.

2 The structure of the book

The book starts with a Preface where readers can find the aims of the book as well as the origins of the various chapters in the monograph. Most of the chapters are based on research papers previously published in journals or conference proceedings (written by either one of the authors or co-authored by both of them),³ while some others (Chapters 1, 5, 9 and 14) are new materials. Some of the fifteen chapters primarily focus on theoretical issues, while others discuss the details of empirical research. Most of the chapters can be read independently of the others. Concerning the structure of the book, the order and sequentiality of its chapters, readers may have the impression that it is a bit ill-proportioned at a few places, with too many (sometimes superfluous) quantitative details in some of the empirical studies (such as in Chapters 7 or 12). Chapter 14 on William Gladstone also stands out among the rest of the chapters as a bit out of place since it is not an indispensable element in the line of reasoning of the authors. Lastly, the final chapter (Chapter 15) sums up the far-reaching implications of all their studies.

3 Synopsis

3.1 Introduction

In the Introduction (Chapter 1) the authors describe the theoretical landscape in current linguistics and express their dissatisfaction with the fact that even today one can find a great number of papers in modern linguistics journals that contrast grammatical and ungrammatical word-sequences by defining several rules, although no one has ever managed to clearly define the concept of grammaticality and plot the boundary between grammatical and ungrammatical so far. The title of the book also suggests that the authors acknowledge that all languages have

² Moreover, the authors claim that grammar encodes "thought-structures" rather than "logical structures" (9).

³ For simplicity, I will consistently refer to the authors of the individual chapters in plural, as "the authors", irrespective of the fact if the particular chapter under discussion was written by one or two of them, since it is a joint book compiled and edited by both of them.

grammar but not grammaticality. Moreover, they claim that most contemporary grammars seem to have holes when they are checked against real-life usage; or, as they put it, “all grammars leak” (Sampson & Babarczy 2014: 6). Concerning their terminology, the authors prefer to talk about grammar encoding “thought-structures” rather than “logical structures”. Quoting Riau Indonesian and Old Chinese language, the authors argue that they find it implausible that all languages follow a universal logical pattern. After providing the theoretical background for their consequent studies in the monograph, their objective is formulated: “the main point of the book is not to argue against the grammaticality concept, but to present a range of concrete positive findings about grammar that emerge when one gives that concept up” (21).

3.2 The bounds of grammatical refinement

Chapter 2 attempts to answer the question how refined English grammar is by applying a comprehensive annotation scheme, the SUSANNE scheme to samples of the English language registered in the SUSANNE corpus. An experiment on the reliability of the annotation of the SUSANNE corpus is being fully described in this chapter. In short, various expert analysts, independently of each other, applied the SUSANNE scheme to the same samples of data. Essentially, this means using inter-annotator agreement as a means to measure the objectivity of the annotation. Their experiment reveals frequent problematic areas concerning annotation, and leads to three conclusions:

- (a) human fallibility is more significant than definitional limitations;⁴
- (b) structural ambiguity is often pragmatically nonsignificant;⁵
- (c) functiontagging is especially problematic.⁶

Provided that the authors reject using introspection, invented sentences and linguistic intuition, the question can be raised why inter-annotator agreement (together with its shortcomings and discrepancies) is considered as a criterion for the objectivity of data collection and analysis methods. Annotation itself always involves some subjective judgements and relies on interpretation, therefore, it can be questioned to what an extent the annotations of the corpus, especially the functional ones, are reliable. The authors fail to elaborate in detail on the problem areas of inter-annotator agreement and the extent of subjectivity in annotation. However, the good thing about their experiment is that it raises crucial questions and has practical implications, too.

⁴ According to the authors, it is clear that “the limitation on annotation predictability is due far more to human fallibility than to the limited precision of the scheme” (Sampson & Babarczy 2014: 47). They claim that real language is too anarchic to permit them to fit examples into the pigeonholes of any scheme, however well-defined (48).

⁵ The second conclusion of the authors is that structural ambiguity does not always correspond to important meaning differences; therefore, the over-refined annotation of these structural differences is often pragmatically nonsignificant (49).

⁶ Thirdly, the authors admit that assigning functional categories to clause constituents, for example, the relationships between verbs and their arguments in English, is especially problematic (51).

3.3 *Where should annotation stop?*

The third chapter of the monograph overviews parsing methods in order to survey indeterminacy in the most comprehensive way. It is the SUSANNE structural annotation scheme for English which is in the focus of the study. It is described as aiming at rigorous explicitness and maximum completeness of detail. The authors list *trainability*, *reliability*, *independence* from linguistic theorizing and *predictability* as the criteria of ideal annotation schemes. However, they also illustrate a few problematic cases and structures where annotation, in simpler terms, choices from the available categories, is difficult. In conclusion, the authors express that they do not believe that a single, correct answer can be given to the question how refined a general-purpose structural annotation should be. At the same time, they seem to have no doubt in the reliability and usability of the SUSANNE Treebank, and claim that its scheme approaches the limit of humanly-possible annotation precision. However, inter-annotator disagreement (outlined in Chapter 2) might have resulted from shortcomings and the over-refined nature of the scheme. It is claimed that some kind of order and classification on raw language data is far better than nothing. All things considered, this final idea is quite weak compared to their promising plans and aims outlined in the beginning of the chapter.

3.4 *Grammar without grammaticality*

Chapter 4 is an edited version of an earlier paper of Sampson, published in 2007 as a “target article” for a special issue of *Corpus Linguistic and Linguistic Theory*. The issues addressed in the study include the notion of grammaticality (Chomsky 1957) as well as the question whether grammar expresses logical structures. The authors claim that the concept of “ill-formed” word-sequences is a delusion, based on a false conception of what human language is like.

Their argumentation is supported by several anecdotal analogies, metaphors and comparisons, such as the one between ungrammatical sentences and unmet friends, or the metaphor of language as an unfenced prairie, having a huge network of paths. The authors reject the notion of un-/grammaticality (and the existence of deviant sentences) using, among others, the analogy between sentences that can be judged ungrammatical and friends one has not met yet.⁷ Secondly, quoting examples from idiolects and socially-deprecated varieties of English (that would be considered unacceptable/ill-formed in high-prestige discourse), the authors put forward their idea that “unfamiliar does not imply ungrammatical” (66). Instead of trying to identify the boundary between grammatical and ungrammatical sequences, it is proposed that it may be a more conducive approach to consider two boundaries: one between familiar and unfamiliar sets of sequences, and another one “between sequences destined never to have a use, and those which will in due be useful” (75). In short, the authors believe that there is no sharp boundary between grammatical and ungrammatical structures. However, it must be noted that it IS worth discussing the notion of grammaticality, we just need to bear in mind that its concept and definition is changing across languages and over time. Moreover, what weakens the force of the argument of the authors is that the “attacked” example sentences invented by linguists can also easily be seen as unmet friends (similarly to novel sentences captured in corpora, as suggested by the authors).

The second key issue addressed in Chapter 4 is the usability of intuition as a data source. Apparently, the authors dismiss the idea that an average language user has access to the as-

⁷ It can be noted here that would add here that certain realizations of language use, for instance, jokes and intentional misleading, can neither be analysed using the notions of well-formedness and grammaticality.

pects of the grammar of a language; therefore, they deny the reliability of grammaticality judgements. They build their opinion that introspective data is controversial and unreliable on the grounds of the following arguments:

- (a) On the one hand, owing to different backgrounds in education and speaking different dialects and sociolects, the judgements of average language users may considerably differ from each other, resulting in chaotic findings.
- (b) On the other hand, concerning linguists' intuition, linguists might be influenced by the theory they evaluate using the data.
- (c) As opposed to the above mentioned risks of misjudgements, corpus data contain forms and structures which can be in conflict with grammaticality judgements of either naïve speakers or linguists.

Throughout their book, the authors are preoccupied with the systematic description of the usage of language, claiming that grammars and lexicons both rely on past and present experiences that can be easily captured in corpora; and that is why corpus linguistics can be a more fruitful approach to the study of language than “prescriptive linguistics”, as the authors call the opposing camp. In their closing words in Chapter 4, instead of further prompting the definition of “psychologically real” grammars, the authors believe that corpus-based grammatical description is the most that scientific linguistics can realistically hope to attain.

3.5 Replies to our critics

Chapter 5 provides a survey of the responses to their 2007 “target article” (the original version of Chapter 4), complemented with their novel reflections on the acceptance and criticism of their paper. They outline what the academic community makes of their position by quoting and replying to both hostile and supportive critics. On the one hand, they respond to the criticism of Pullum (2007) and Müller (2008). On the other hand, they also refer to works that offer arguments in favour of their ideas, such as Stefanowitsch (2007) who believes that sequences commonly judged as ungrammatical often diminish from real-life usage not because of grammar, but because they mean things that people simply do not want to say. Following this discussion, the remaining part of Chapter 5 is essentially a critique of linguistic nativism and grammaticality implied by Universal Grammar (Chomsky 1965).

3.6 Grammatical description meets spontaneous speech

In Chapter 6, before the authors look at common difficulties that arise when tagging annotations to spontaneous speech, first of all, they stress the primacy of spontaneous speech (over both written texts and introspection) to linguistic research, and propose that “if one wants to examine aspects of language that are innate rather than acquired, they are surely more likely to be found in the biologically-natural domain of speech than the relatively artificial domain of writing” (Sampson & Babarczy 2014: 119).

To provide background for their claim, the authors argue that as a result of the increase of computing capacities as well as methodological innovations in corpus linguistics, research interest shifted from the study of written to spoken interaction, which allowed linguists to raise fundamentally different kinds of research questions. Corpus linguists, among others, revealed that spoken language greatly differs from written text insofar as not all sentences are well-formed, sentence boundaries are often unclear, and different principles and maxims guide

spoken interaction. It is presented through a variety of examples that using the CHRISTINE Corpus and the demographically sampled speech section of the British National Corpus as data sources, a greater variety of grammatical structures⁸ can be found than only using a written corpus. However, doing speech annotation admittedly has its particular difficulties,⁹ which are also described in Chapter 6.

In contrast with generative models that describe the competence of ideal speakers and do not explain what is going on behind sentence boundaries; in a performance-based framework, like that of the authors, the analyst needs to be faced with the output of average language users, not only ideal speakers who actually do not even exist. As long as one cannot define what “error” is, one should not rely on the notions of ungrammatical or ill-formed sentences. However, if one can identify a set of several common patterns of language use with the help of applying frequency-based statistics, one can understand deviation from frequent patterns. As illustrated by several examples in the monograph, it is interesting to read how speakers deviate from frequent patterns of usage and behaviour.

3.7 Demographic correlates of speech complexity

The case study presented in Chapter 7 empirically tests the idea of correlation between utterance complexity and demographic properties of speakers, such as age, sex and social status. In order to measure the grammatical complexity of speech, each speaker (in the selected conversations of SUSANNE, the spoken part of British National Corpus) is assigned an “embedding index” which shows the degree of embedding of various individual words uttered by the particular speaker in the sample. As for the demographic features and the classification of speakers, the British National Corpus (henceforth: BNC) involves reliable information about the sex and age of its speaker. However, concerning classification in terms of social class and regional accents within Great Britain, BNC is rather inaccurate.¹⁰ The authors aimed at finding correlations with each of the demographic variables by grouping speakers into categories and applying the F statistic to compare variance among the categories. The statistical tests carried out on their dataset did not confirm the correlation between grammatical complexity and social class, neither between complexity and region. Regarding the sex variable, British females may on average utter more complex sentences than British males, and this is probably not just a random effect.¹¹

From all the statistics, the results on the age variable are the most striking. Using six age bands: (1) speakers up to 15 years, (2) 16-24, (3) 25-34, (4) 35-44, (5) 45-59, (6) 60 and above, a correlation was found between age and utterance complexity. After presenting this

⁸ According to the authors, one can find a greater variety of grammatical structures and a larger number of repairs and instances of nonstandard usage in speech than in writing; therefore, one can discover the true messy and anarchic nature of language using spoken corpora.

⁹ Difficulties in annotation involve wordtagging, especially tagging discourse items, speech repairs and words used in nonstandard grammatical functions. Distinguishing difference of dialect from performance error (131) also sets an enormous task for the annotator. Unclear segments are evidently not accurately transcribable either.

¹⁰ In order to eliminate the inaccuracies in the demographic taxonomy BNC, the research reported in Chapter 7 uses a revised, five-way regional taxonomy (distinguishing speakers as coming from Southern England, Northern England, Wales, Scotland or Northern Ireland) and a four-way social classification (with professional, skilled non-manual, skilled manual and partly skilled or unskilled categories).

¹¹ “Conventionally, a probability of less than one in twenty is taken as good reason to think that a statistical phenomenon is real rather than a random effect” (146). It is for this reason why the authors say that the result may not be just a random effect.

correlation, the authors look at the age variable in more detail, and contrast two ideas about first language acquisition (henceforth: FLA): (1) the “critical period” hypothesis held by linguists believing in Universal Grammar, and (2) the lifelong learning hypothesis. Apparently, the authors aim to illustrate that FLA is a continuous, life-long process. It is argued in Chapter 7 that the ability to use grammatical subordination is a significant component of learning to use one’s mother tongue; therefore, embedding indices plotted against age can be a useful marker of the FLA scenario. Their figures give them grounds for claiming that “increase in average grammatical complexity of speech appears to be a phenomenon that does not terminate at puberty, but continues throughout life” (Sampson & Babarczy 2014: 152). Indeed, their data presented convincingly suggests lifelong growth in complexity of usage (as they put it, “lifelong learning”), which they claim to support their view of language as having grammar but lacking grammaticality (153).

Besides presenting their research material and the consequent findings, another essential methodological issue, the difficulty of quantification might have also been addressed in this chapter. Among the other problem areas listed, it would also be useful to discuss the limitations of using certain statistical tests. For instance, it could be argued that correlation does not prove the existence of causation since significant effects can be caused by external factors as well (cf. Schegloff 1993). It should always be kept in mind that variables may co-vary but one can never be a hundred percent sure why; there might not be interaction between them; it might as well be the result of a third, unknown factor.

3.8 *The structure of children’s writing*

Chapter 8 presents the findings of an analysis on the process of writing skills acquisition based on the CHRISTINE and LUCY corpora as well as material published by a research project on pupils’ use of oral and written language.¹² On the whole, their various annotated samples were divided into three categories representing the uses of language in different modes and ages: (adult) speech, (adult) published writing and child writing. The question could be raised why there are two categories for the analysis of adult language use (that is, adult speech and adult writing), while child speech as a category is missing. Supposedly, the authors aimed to uncover the trajectory of the transition from one stage to the next by investigating if child writing can be seen as intermediate between spontaneous speech and published writing.

Their data showed that structures in writing commonly consist of more lexical items and involve deeper branching (i.e. recursion) than structures in speech. In this respect, children’s writing can be seen as transitional between conversation and published writing. It was presented based on their data that children earlier use relative clauses than simpler constructions in writing. Their sample is small and might be unrepresentative, therefore, unreliable; however, their finding leads the reader to the research question addressed in the next chapter: what permits or encourages the early written usage of relative clauses.

3.9 *Child writing and discourse organization*

Since the authors were sceptical about the nativist idea of having a fixed grammatical programme, they looked more closely at “mistakes” and syntactic patterns used in the child writ-

¹² For the purposes of their case study, that part of the research material was used where children of both sexes, aged from nine to twelve were asked to write essays on a choice of open-ended topics.

ing material described in Chapter 8. In sum, the following conclusions were drawn from their analysis:

- (a) Contexts conducive to relative clause/participle clause constructions are most frequent in child writing, and least frequent in spontaneous speech.
- (b) Relative clauses/participle clauses are most prone to be selected by adult writers, and least prone to be selected by speakers.
- (c) Unlike in speech, the position of the antecedent heavily affects clause type selection in written modes.
- (d) Syntactic errors are similarly rare in both writing and speech (182).

Taking into account all the above findings, the authors conclude, although child speech was not investigated in the study, that the stage of child writing cannot be seen as syntactically intermediate between speech and adult writing. The data of Chapter 8 and 9 do not imply any innate programme under which grammar unfolds in a child's mind in a fixed sequence (183). Therefore, their findings are interpreted as counter-arguments to the idea of having "innate grammatical knowledge", as claimed by nativists (183).

3.10 Simple grammars and new grammars

All throughout the book, the authors aim to undermine the idea that all languages are cut to a common, universal logical pattern. This idea is supported in Chapter 10 by referring to two specific classes of language, namely, the classes of pidgins and creoles which lack structural mechanism and useless irregularities. Instead, these languages are relatively simple in terms of abstract categories and relations they encode in their grammars. To explain this phenomenon, the authors quote McWhorter (2001) referring to the lives of languages, and claim that new languages, such as pidgins, tend to start as simple and later become grammatically more complex. The authors agree with one of McWhorter's ideas about creoles, claiming that if a language is creole then it is simple. On the other hand, they do not concede to his claim from the other direction, that is, they disagree with the idea that if a language is simple, it must be a creole or pidgin (Sampson & Babarczy 2014: 185).

As an example, the authors turn to the case of Old Chinese and argue that it has never been a creole on the grounds that it does not lack inflexional affixation, which would be typical of creole languages in McWhorter's diagnostics. Although the majority of Old Chinese words were simple roots, it is clear that there was some derivational affixation. Since this phenomenon (the presence of derivational morphology) is in contrast with the third defining property of creoles in McWhorter's diagnostics (namely: the lack of semantically-unpredictable derivation), it excludes Old Chinese from the group of creole languages.¹³

In conclusion of Chapter 10, the authors assert their idea that structural simplicity does not imply a history of creolization (200). European languages are claimed to be rich in logical apparatus because they are old. In relation to the case of Old Chinese versus creole languages, the authors, as sceptics of the idea of Universal Grammar, criticize the theory of generative grammar on the grounds that generativists' data is predominantly taken from present-day languages.

¹³ Instead, it is proposed that Old Chinese was probably a "regular language", tracing ancestry from an unwritten Proto-Sino-Tibetan language, undergoing normal processes of language change.

3.11 The case of the vanishing perfect

The authors argue all through the book that grammatical constructions and the semantic concepts they express are cultural creations rather than universal items in the mind fixed by genetic inheritance. The case of the vanishing perfect in present-day English is presented as further evidence that there is no such thing as grammaticality. In order to support their claims, the authors look in detail at the loss of the perfect aspect in English.

The material of their case study in the demographically-sampled CHRISTINE Corpus which represents the conversational language use of a cross-section of the United Kingdom population. Using their material, a corpus-based examination was carried out on the overall system of verb qualifiers in spoken English. Their data seems to demonstrate the tendency that Southern English and Irish English speakers systematically refrain from uttering Perfect constructions.¹⁴ The authors argue that languages can become not only more complex over time but can also become less complex by losing existing constructions.¹⁵ This phenomenon makes them think that there is no universal system dictating the logical or semantic architecture of languages (217).

3.12 Testing a metric for parse accuracy

Chapter 12 is entirely of technical nature, not really essential in the line of reasoning of the book as a whole. Briefly, it is set forth that one of the major tasks in natural language processing is automatic parsing.¹⁶ Chapter 12 presents the results of an experiment comparing the accuracy of two metrics using a set of sentences from the SUSANNE Treebank. The two metrics compared are the widely used Grammar Evaluation Interest Group (henceforth: GEIG) metric and the metric the authors use, the leaf-ancestor (henceforth: LA) metric¹⁷ (Sampson 2000). Their testing results supported their initial assumption that the LA metric parses more accurately than the GEIG metric. The reason why the LA metric might have been ignored so far is that the algorithm calculating it is much more computation-intensive than the GEIG metric. However, this should clearly not be a problem with the computing capabilities presently available. The authors believe that the main reason why GEIG metric must have been widely approved is simply the fact that it has had authority behind it.

3.13 Linguistics empirical and unempirical

The central question in Chapter 13 is which data types are considered empirical in linguistic research. To start with, the authors raise the provocative question: “how can so many linguists have got thing so wrong?” (Sampson & Babarczy 2014: 237). In order to check how far the discipline of linguistics has moved towards empiricism, the authors carried out a survey on

¹⁴ The authors also call attention to their opinion that it would be misleading to think of Past and Perfect constructions as alternative ways of expressing similar ideas. In addition to the marginal use of the Perfect aspect, uttering got without auxiliary is also common in these two regions. Provided that the Perfect is vanishing from Southern English speech, the authors expect to soon find the same tendency in writing as well. They actually managed to find an advertisement including an example of the use of a Past form with just in the sense “just now”. However, a single counter-example is not convincing enough.

¹⁵ Languages may lose existing means of expressing grammatically-encoded conceptual distinctions (217).

¹⁶ Automatic parsing refers to the development of software that takes in chunks of prose and outputs correct grammatical analyses of that prose (218).

¹⁷ The essence of LA assessment is that their metric evaluates the parsing of an individual terminal element in terms of the similarity of the lineages of that element in candidate and gold-standard parse trees (220).

the types of publications of the journal *Language*. They examined the 1950 volume of the journal, and then two out of every five volumes from 1960 up to 2011, and imposed a three-way classification on each of the papers in their sample: (1) evidence-based, (2) intuition-based, and (3) not applicable.¹⁸ In the authors' definition, "evidence-based" research relies on interpersonally-observable authentic data in order to support its claims (255). Papers classified into the "evidence-based" category tend to refer to their findings as empirically checked and objectively tested beliefs. As a further possible criterion in the classification, the authors say that evidence-based research often uses computational methods (252). The authors use the category of "intuition-based papers" to refer to, as they put it, "speculative" works relying on either intuition or "elicitation from informants" (255).

The findings of their literature survey match their view of the development of linguistics. Before the appearance of generative linguistics, one could come across a high proportion of "observation-based"¹⁹ papers" (263). As it can be expected, the lowest percentages of evidence-based papers occur during the years between 1965 and 1970. After the early 1970s, we can find more and more evidence-based papers. The availability of electronic corpora made it possible for linguistic research to advance from intuitive to evidence-based, more scientific techniques. Although there had been considerable innovations over the past three decades, linguistics is still lagging behind other scientific disciplines in terms of empirical methodology.

According to the authors, the problems underlying the insufficient level of empiricism in linguistics may spring from three reasons: (1) a major misconception is that linguistics is not an empirical field of study/science; and therefore, (2) the linguists in question may not have been looking at the facts; third of all, (3) speculative, intuition-based theorizing provides easier reading than discussion based on a large amount of empirical data and its statistical evaluation. Chomsky's belief in Universal Grammar and his rejection of descriptive empiricism (Chomsky 1964, Chomsky 1976) must have indeed discouraged work trying to construct generative grammars accountable for real life data. Against this background, the rather extreme idea is also put forward in the monograph that questions of grammaticality in particular, and generative linguistics in general should not be considered as part of empirical linguistics.

Although unempirical linguistics is claimed by the authors to be being written and read even today, this is hoped to be only a temporary situation, since the accessibility and the convenience of advanced computing technologies has enlarged the range of issues which can be quickly searched and objectively tested. What is missing from the account of the authors is the explanation of what exactly they mean by the notions of facts, real life data and unempirical linguistics.

3.14 William Gladstone as linguist

Chapter 14 is of historical interest, therefore, it is a bit out of place and irrelevant in the line of argumentation of the book. Its aim to reinterpret a yet another widely held claim and clarify a misinterpretation in linguistics connected to the works of William Gladstone who published studies on the vocabulary of Homeric Greek. Gladstone has repeatedly been mistaken for believing that Greeks of the Homeric period were colour-blind, although he was not saying that.

¹⁸ There were a large number of cases where it was difficult to assign a label to the paper in question and where the question of origin of data was irrelevant due to the different nature of the research, such as discussions or evaluations of different methods.

¹⁹ By observation-based papers, the authors supposedly mean papers assigned to their category of evidence-based papers.

After clarifying this misconception, the authors list the positive aspects of his writings, and highlight that his claim, namely, that exotic languages are systematic in their own way, is in fact a subtler form of the Sapir-Whorf hypothesis. The authors assume that his reviewers, claiming that Gladstone believed in Homeric colour blindness, make it appear as if no one before Sapir and Whorf proposed that language differences may reflect socially-determined differences in perception. The lesson of this story that the authors try to convey in this chapter is that we should be more alert to the danger that bright ideas put forward by less known names may be similarly shuffled aside even today.

3.15 Minds in uniform

The closing chapter is based on an intriguing keynote presentation by Sampson, and nicely sums up the far-reaching implications of the previous studies (in linguistic theorizing, empiricism and politics), welcoming cultural diversity and relativism. The authors entirely reject the existence of biological, cognitive and cultural universals guiding language acquisition and language change. Moreover, they make a provocative claim by saying that modern generative linguistics is junk, and it serves the function of establishing a pseudo-intellectual justification for the trivialization of cultural differences between communities. It is concluded that languages are cultural inventions, which human civilizations freely create without any biological constraints. This sounds like a strong exaggeration since, for instance, our sentences cannot be any long or complex to be correctly understood or remembered. No evidence is found to support a model of human cognition according to which we are limited to a fixed set of constructions and behaviours. If we believed in the existence of cognitive universals, we would run the risk of mistaking the accidental features of our own culture for cultural universals, which might easily result in a discriminatory mindset. Instead of maintaining such disingenuous theories, the authors suggest that academics should help people grasp the diversity of human cultures.

4 Evaluation of the key ideas

As for one of the most central issues in contemporary linguistics research, that is, the applicability of data types, the authors address the restrictions and limitations of applying various methods and data, especially those methods which solely rely on linguists' intuitions and grammaticality judgements. It is claimed by the authors that giving grammaticality judgements is an unnatural kind of linguistic activity, and using grammaticality judgements and own intuitions in linguistic theorising are unreliable sources of data because grammaticality judgements are, in their view, mere individual "feelings" (in the form of subjective interpretations). To support their views, the authors list problems with metalinguistic tasks and raise research questions that can only be answered by using annotated corpora, rather than introspection. These questions tackle the differences in spoken and written language, adult and child language as well as variations in language use grasped by frequencies. In order to grasp these variations, the authors advocate the use of authentic data that builds on actual language production (versus subjective introspection). According to the authors, the only acceptable data source is corpus data because it is natural, contextualised and entirely free of subjective elements.

Although the authors attempt to contribute to the resolution of several empirical problems and current linguistic debates, their standpoint is not in line with dominant current linguistic

views. Sampson and Babarczy, the authors of the book under review, regard corpus data as the only relevant and legitimate scientific data type in linguistics. This can be considered a conservative view which is in contrast with the dominant tendencies in contemporary theoretical linguistics. As Kertész & Rákosi (2008: 62) point it out: “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”. Using the taxonomy of Kertész & Rákosi (2008), Sampson & Babarczy (2014) can be described as advocates of corpus data, and their views are manifestations of the “standard view of linguistic data and evidence” (Kertész & Rákosi 2008), as opposed to more progressive views of linguistic data and evidence, such as that of Pullum’s (2007) who suggests a more comprehensive metascientific perspective and the method of “reflective equilibrium” which can be achieved by the cyclic process of revisiting former beliefs in the light of new data and new ideas. Rather than combining both inductive and deductive methods, which is a dominant practice in contemporary linguistic research, Sampson & Babarczy (2014) restrict empirical linguistic and linguistic theorizing to rigid inductivism, manifested in the observation of corpus data only. Therefore, they exclude the use of introspective data from legitimate linguistic research, unnecessarily limiting the realm of research.

As a counter-argument to the views of the authors concerning the preferred avoidance of introspection, including its rejection as scientific data, it should be noted here that, first of all, identifying sentence and utterance boundaries, annotating various grammatical and discourse aspects of spontaneous speech data also involves reliance on introspection and judgements.²⁰ Even though the authors forbid reliance on intuition in the process of data analysis and evaluation, some subjectivity in interpretation is inescapable and objectivity cannot be entirely guaranteed in corpus annotation, even if it employs a relatively theory-free system. Broadly speaking, introspection and interpretation are essentially inescapable in both the comprehension and production of language. At the same time, carefully planned and rigorously controlled experiments using native speaker intuitions and judgements can provide both stimulating starting points for empirical linguistic research as well as the basis for arguments in linguistic theories. Besides this, both the native speakers’ intuitions and people’s reports of their linguistic intuitions can be treated as observable data, and it is feasible to formulate descriptions of those reports. Intuitive data may in fact seem chaotic but they can be controlled by developing such carefully designed questionnaires that can be systematized and evaluated. Arppe & Jarvikivi (2007) claim that every manifestation of linguistic behaviour is a relevant datum. Moreover, a corpus (any corpus) does not necessarily mirror reality. Only a well-balanced corpus (in terms of the number, age, origin and background of its speakers) may reflect language use in a reliable way. Yet again, it is culture-dependent and almost impossible to define what counts as a well-balanced corpus. Introspective data can be the subject of analyses, and grammaticality judgments are reconcilable with corpus-based methods. The origin of data is not of decisive importance (Kepser & Reis 2004, Meurers 2007); therefore, it is encouraged to use a variety of data sources such as introspection, corpus analysis, psycholinguistic and neurolinguistic experiments.

The major questions are to find out if the findings of the two or more data collection methods are convergent or contradictory, and what might explain this phenomenon (either similarity or difference). What is more, there might be inconsistencies in research findings even when

²⁰ The annotation of spontaneous speech data also involves reliance on introspection and judgements on the side of the annotator or the linguist when selecting labels to tag.

using only one data type. Principally, linguists should not only collect and document data but also find explanations for their variability. A potential model that is able to account for inconsistencies is the *p-model of plausible argumentation* (Kertész & Rákosi 2012) which is a complete model of linguistic theorizing. Their concept of plausibility refers to the belief that linguistic data are most of the time uncertain or plausible to some extent rather than true with certainty. The p-model defines a datum as a statement originating from a direct source (including various sources such as corpora, linguistic intuitions and experiments) which functions as a starting point in the argumentation process. Plausible argumentation consists of *argumentation-cycles* where one retrospectively re-evaluates former decisions from another perspective or in the light of new information, therefore making the process cyclic and prismatic (Kertész & Rákosi 2012). Rather than advocating the substitution of intuition with observation and corpus collection, it would be a more fruitful and more innovative approach for Sampson & Babarczy (2014) to consider the reflective combination of two or more data types and apply the constant re-evaluation of their findings in order to maximize the plausibility of the hypothesis to be supported.

As a further weakness of the book under review (Sampson & Babarczy 2014), it must be mentioned that the concept and criteria of *evidence* are not clearly defined in any of the chapters, but it can be assumed that the authors only accept corpora as scientific data in linguistics, and only approve corpus-based results as explicit linguistic evidence. It may also be noted here that the authors do not seem to consistently distinguish the concepts of data and evidence, although it would be indispensable to define based on what criteria they distinguish these concepts. Kertész and Rákosi (2012) offer a potential solution to this problem of data and evidence in linguistic theorising since their p-model of plausible argumentation clearly defines the relation of data and evidence. In their view, evidence is not an objective subcategory of data; instead, any datum can serve as evidence for a hypothesis in an argumentation process with the condition that it is a premise of a plausible inference making the hypothesis plausible. In the understanding of Kertész and Rákosi (2012), data and evidence both have crucial functions in every stage of linguistic theorizing, therefore, their role cannot be restricted to testing a hypothesis. If the views and practices of Sampson & Babarczy (2004) concerning the applicability of data types were supplemented with these notions of data and evidence according to the p-model of Kertész & Rákosi (2012), their work could be more exhaustive and more progressive.

In summary of the reviewed book (Sampson & Babarczy 2014), it must be highlighted that all of its chapters are very informative and original, at times, even provocative, contributing to both methodological and metatheoretical discussions, although not in the most innovative way. Above all, the book underlines the idea that the nature and development of languages as well as the status of data in linguistics are still unresolved, controversial issues that present scientific challenges even today. Readers interested in language and empirical problems in linguistics can find the monograph interesting and use it as a springboard for further research.

References

- Arppe, A. & Järvikivi, J. (2007): Take empiricism seriously! In support of methodological diversity in linguistics. *Corpus Linguistics and Linguistic Theory* 3, 99-109.
- Chomsky, N. (1957): *Syntactic Structures*. 's-Gravenhage: Mouton.
- Chomsky, N. (1964): *Current Issues in Linguistic Theory*. 's-Gravenhage: Mouton.

- Chomsky, N. (1965): *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1976): *Reflections on Language*. London: Temple Smith.
- Kertész, A. & Rákosi, Cs. (2008): Conservatism vs. innovation in the (un)grammaticality debate. In: Kertész, A. & Rákosi, Cs. (eds.): *New Approaches to Linguistic Evidence. Pilot Studies /Neue Ansätze zu linguistischer Evidenz. Pilotstudien*. Frankfurt am Main etc.: Lang.
- Kertész, A. & Rákosi, Cs. (2012): *Data and Evidence in Linguistics. A Plausible Argumentation Model*. Cambridge: CUP.
- Kepser, S. & Reis, M. eds. (2004): *Linguistic Evidence: empirical, theoretical and computational perspectives*. Berlin: Mouton de Gruyter.
- McWhorter, J. H. (2001): The world's simplest grammars are creole grammars. *Linguistic Typology* 5: 125-166.
- 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.
- Müller, S. (2008): *Head-Driven Phrase Structure Grammar: eine Einführung* (2nd revised edition). Tübingen: Stauffenburg Verlag.
- Pullum, G.K. (2007): Ungrammaticality, rarity, and corpus use. *Corpus Linguistics and Linguistic Theory* 3, 33-47.
- Sampson, G. (2000): A proposal for improving the measurement of parse accuracy. *International Journal of Corpus Linguistics* 5: 53-68.
- Sampson, G. & Babarczy, A. (2014): *Grammar Without Grammaticality: Growth and Limits of Grammatical Precision*. Berlin/Boston: Mouton de Gruyter.
- Schegloff, E. A. (1993): Reflections on Quantification in the Study of Conversation. *Research on Language and Social Interaction* 26: 99-128.
- Stefanowitsch, A. (2007): Linguistics beyond grammaticality. *Corpus Linguistics and Linguistic Theory* 3: 57-71.