

Take empiricism seriously! In support of methodological diversity in linguistics

A commentary on Geoffrey Sampson: Grammar without Grammaticality

Antti Arppe

University of Helsinki
Finland

Juhani Järvikivi

University of Turku
Finland

Address all correspondence to:

Antti Arppe

Department of General Linguistics
University of Helsinki
FIN-00014, Helsinki, Finland
Tel. +358-9-19129312
Fax. +358-9-19129307
Email: antti.arppe@helsinki.fi

Geoffrey Sampson has for quite some time now argued for an empirical basis of linguistics, as he has done in earlier works (e.g., Sampson 2001) as well as in this paper under discussion. It seems, however, that Sampson has hitherto considered only corpus-based evidence as sufficiently empirical, and he sees little or no role, not only for intuition, but also for experimentation, because he sees no difference with the researcher's own intuition and that of an (external) informant, characterizing this as “merely treat[ing] the informant's intuition rather the linguist's as the source of authority.” (2005: 28). In his present paper, it now seems that Sampson has fallen into a pessimistic stance not far from defeatism, when he asserts that not even corpora can provide reliable evidence on rare linguistic items and phenomena. And if one considers corpora as the only reliable source of empirical evidence in linguistics, and intuition as the only but in effect unreliable alternative, this is, of course, a conclusion that one could possibly reach. However, in our opinion Sampson's analysis is at times incoherent and often fallacious, and his view of the range of methods and sources of evidence available to linguistic research in general is too restricted, to say the least.

Based on his corpus analysis of several rare linguistic items Sampson asks in his paper the following question: “[H]ow someone constructing a grammatical description is to distinguish between well-formed but rare constructions, which the grammar should generate, and performance deviations, which it should exclude? What evidence could this distinction be based on?” (p. 12) Sampson then argues at length that “speaker's intuitions” are not reliable, at least for less common linguistic phenomena, as “speakers cannot be expected to make reliable judgments about the status of the unusual word strings.” (p. 14) Disregarding why this should be the case *a priori*, one should take note that Sampson mostly implies that “speaker's intuition” is that of the linguistic researcher himself/herself, often criticized as “arm-chair linguistics”, not the intuition of independent informants other than the researcher. Sampson provides examples of extensive intuition-based research on the one and same linguistic phenomenon, namely the position of subordinate clauses in questions in English, firstly yielding contradictory results by several different researchers over time, and secondly that these intuition-based results are furthermore in contradiction with those observed in corpora. In addition to Sampson, for instance Gries (2002: 27) has shown, in a case study of English genitive alternation that “informed linguists' intuitions on (syntactic) phenomena are inadequate. In this respect, we could not agree more with Labov's ([1970] 1978: 199) urging, which Sampson cites, that “linguists cannot continue to produce theory and data at the same time”. Nor do we disagree with Sampson's own formulation that “[I]f linguistics is to be

taken seriously as a knowledge-generating enterprise, it simply must base itself on empirical, interpersonally-observable data.” (p. 17).

In this critical commentary of Sampson’s paper we would like to make two general points: We would first like to argue that Sampson’s view about what exactly constitutes empirical data for linguistics is too narrow and simplistic, because it is not only based on an incorrect analogy with natural sciences but also because his argument on the unreliability of intuitive evidence is fallacious, and thus his presuppositions do not warrant the conclusion that he is consequently forced to make. Secondly, we would like to argue that in practice the basis of empirical evidence available to linguistic research is far broader than Sampson assumes.

Especially when one looks outside the confines of theoretical “core” linguistics “proper” as it has been largely conceived in the latter half of the 1900s, for instance in its many “hyphenated” subfields, e.g. sociolinguistics, psycholinguistics or field linguistics and language typology, one discovers that the range of different empirical evidence types and associated methods is and has been in fact for some time both wide and varied, and includes, in addition to introspection, elicitation (paraphrasing, completion and sorting tasks), corpus data (synchronic and diachronic, standard vs. dialectal, formal vs. colloquial, spoken and written, published and internet), psycholinguistic experiments (off-line and on-line forced choice/selection and off-line categorical and graded acceptability judgment/rating tasks, on-line reading and reaction time measurement), language acquisition and learning data (both for mother tongues and other languages), neurolinguistic measurements, and of late increasingly cross-modal studies (e.g., cross-modal lexical decision interference and story-picture selection tasks, truth-value judgments), just to name a few.

In discrediting intuition as reliable linguistic evidence, Sampson is lead to purge linguistic research of the concern for a distinction between *grammaticality* and *ungrammaticality*, considering this a futile and misguided exercise, as “...the concept of ”ungrammatical“ or ”ill-formed“ word-sequences¹ is a delusion ...” (p. 1). Claiming for an analogy with the saying “There are no strangers – only friends I have not met”, as he puts it, he argues that the distinction between grammatical and ungrammatical has no empirical basis and should be abolished. His argument is based on two observations: Firstly, that there are unarguably

¹ We have opted to use the more general term *sequence* instead *string* throughout this text when referring to a set of words output in some particular order to constitute an utterance, though we have, of course, retained wordings in citations as they are. In our view, *string* is overtly particular to computer science and computational linguistics.

many cases where we cannot decide based on intuition alone whether a certain sequence of words is grammatical or not – thus making intuitive evidence unreliable and unscientific. Secondly, that it is hard, if not impossible, to come up with a sequence of words that is not altogether potentially usable and understandable in some situation or context.

However, we would like to argue that neither of these conclusions follows from the premises that Sampson offers. Let us take as a case in point the utterance “*Am what I doing is worthwhile*” that Sampson presents (p. 17), one which he has heard a single time. Sampson takes this to be an example of the kind of *mistake* that has been argued is never produced. If this is an *incorrect* and thus an *ill-formed* token of speech, then we may (following Sampson) also deem it *ungrammatical* (which it is, of course, but this is just to make a point).² This example is enough to make clear that there are a great number of cases (limited only in terms of types of structures) where we can in fact be quite certain of whether a particular sequence of words is a correct or incorrect (and thus well-formed or ill-formed, *ergo* grammatical or ungrammatical to follow Sampson) structure in a particular language. This brings out the first fallacy in Sampson’s argument against intuition based evidence: From the fact that we cannot in *every* instance decide upon whether a particular structure is correct or not, it does not follow that we cannot do so for *any* instance. This example also shows that Sampson’s argument against the notion of grammaticality cannot rest on whether one can imagine a potential use, situation or context where a particular sequence of words could be uttered or useful (p. 5), or the lack of an ability to imagine a structure that would or could not be usable in some situation or another. Furthermore, it also shows that whether a particular sequence of words is *judged* correct or incorrect (i.e., grammatical or ungrammatical) cannot be decided upon by whether one can *understand* what the sequence/utterance is meant to mean in a particular situation/context.

In addition, Slips of the Tongue, i.e., output *errors* that are produced by normal native speakers unintentionally and spontaneously (and sometimes unconsciously) are another well-documented case in point. These *errors* are not random, but instead follow the way the language system is organized (Fromkin 1971, Cutler 1982, Niemi and Laine 1997). On many occasions the speaker in fact corrects the error – therefore even the corrections can tell us

² Contrary to his own advice in his article, Sampson seems to be quite happy to take a single spatio-temporal linguistic occurrence, i.e., one utterance, which he even deems incorrect (“a mistake”) solely based on his own intuition, as a counter-example to a theoretical claim about possible production errors.

quite a lot about how our linguistic system is organized (Hokkanen 2001). Thus, Sampson's analogy with "There are no strangers – only friends I have not met" (p. 1) breaks down. So, if one can understand (or imagine a use for) a particular uttered sequence in context in no way determines whether the sequence is in fact correctly produced or not (as Sampson himself so conveniently shows). To put this another way around, the ability to deem a particular sequence as a mistake and at the same time know what is meant by the utterance in fact presupposes that also the correct structure(s) for expressing that meaning is clearly known. Therefore, judgments of comprehensibility and correctness pertain to distinct perspectives in the study of linguistic usage, the first to the semantic intelligibility of an utterance in some pragmatic context and fulfilling a communicative need/purpose, the second to a more structural normlikeness of the same utterance, implying membership in a linguistic community.

It thus also follows (against the practice of intuition-based linguistics) that 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.

As a summary of our critique on Sampson's conceptual treatment of the notion of grammaticality, from the fact that we cannot intuitively decide (or agree upon) whether *every* instance of language use represents correct or incorrect types (structures), it does not follow that we cannot make such a decision on *any* instance of language use. Thus, if we cannot *always* rely on intuition, this does not entail that we can *never* do so. Similarly, the fact that we cannot be certain of whether a particular instance is grammatical or not, should not be generalized to mean that this applies to any and all such instances. Neither does it follow that intuitions are always reliable – they certainly are not so. As Wasow and Arnold (2005) show, extragrammatical factors influence intuitions, especially in the case of complex and/or infrequent structures.

Therefore, we disagree categorically with Sampson that “there is no reason to assume that patterns in a speaker’s intuitive grammaticality judgments reflect realities of his language.” (p. 13) And we consider, as incorrect and misinformed, Sampson’s assertion that “different individual speakers of the same language regularly differ when asked about the grammaticality or well-formedness of particular strings of words – though the fact that they do is by now so well-known that it is not worth quoting examples here...” (p. 13). We base our view on extensive research proving otherwise (e.g., Keller 2000, Featherston 2005, Kempen and Harbusch 2005, Sorace and Keller 2005, Arppe and Järvikivi 2006, Bresnan 2006). These counterexamples are valid, of course, on the condition that such observations of speaker’s intuitions have been collected – not haphazardly in impromptu settings – but in a scientific manner, something which Sampson, too, calls for consistently. This methodological reliability and validity is achieved by systematically observing the judgments of a substantial number of speakers in a carefully controlled setting in order to rule out potentially confounding factors, inconsistencies and occasional errors and slips, exactly tokens of the type that Sampson presents as ungrammatical utterances. In fact, this is exactly the type of experimental research presented by Cowart (1997), which Sampson himself cites (p. 13). As Cowart says in his own words (1997: 26-27) “there is a stable phenomenon of sentence acceptability; we find that for all of the syntactic phenomena considered here, native speakers of native American English exhibit stable, clear-cut differences across sentences”, and further continues that “[t]he evidence … also demonstrates that there are practical experimental methods by which the phenomenon of sentence acceptability can be measured and assessed.”

Therefore, experimentation, exemplified by the studies mentioned above, is the very answer to the plea for a reliable research method for the judgment of grammaticality that Sampson presents the linguistic community. With experimentation, often in comparison with corpus-based analysis, one can discern linguistic phenomena which are rare but fully acceptable, from others which are also rare but considerably less acceptable. In our own research of factors influencing the choice of a synonym pair in Finnish (Arppe and Järvikivi 2006), in which we explicitly compared corpus-based evidence with two types of experimental judgments, we found out that *rareness* in a corpus and *acceptability* in a rating experiment (with 45 informants) do not go hand in hand. A form which was rare in both relative and

absolute terms in a corpus when compared with its alternative (2 vs. 24³) was considered quite acceptable, in relation to the other forms studied, whereas another relatively rare form with considerably more observations in absolute terms in comparison to its alternative (10 vs. 34) received the lowest acceptability ratings by far of the six different forms studied. In short, with experimentation we were able to distinguish between acceptable and unacceptable rare items, which is exactly what Sampson asks for. Furthermore, Featherston has shown that language users are actually able to make graded judgments, along a continuum, on *all* linguistic structures *regardless* of their degree of well-formedness, i.e. not only on the "best" structures which occur frequently or rarely in a corpus, but also concerning forms which do not occur at all (2005: 204-205). Kempen and Harbusch (2005), too, have made a similar general observation.

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.

In general, it is a mischaracterization that the generative linguistic community (which Sampson has been tempted to characterize explicitly as "unempirical" in 2001: 11) is inherently an intuition-based scientific enterprise. It is largely true that generative grammar has traditionally been methodologically encapsulated and relied in practice almost solely on intuitive evidence (e.g., Wasow and Arnold 2005) – often contrary to claims that this were not the case. However, starting with Schütze's landmark methodological work (1996), generative linguists are presently very much testing their hypotheses against empirical evidence, be it experimental or corpus-based (see, e.g., a collection of such work in Kepser and Reis 2005). In this collection of the 26 studies, half (13) employed two, or in a few cases, even more⁴ different empirical data types and methods. Rather than the (lack of) use of

³ In a substantially larger reference corpus the ratio for the same two items was 8 vs. 88, i.e., consistently of the same magnitude.

⁴ The number of data types and methods is in many cases difficult to specify exactly, as a study might incorporate, possibly quite comprehensively, results from research undertaken by others, or as a single

experimental evidence as such in current linguistic research, the trouble is mainly, as Wasow and Arnold (2005) remark, that this evidence does not yet feed back to theoretical linguistics sufficiently enough to modulate linguistic theory construction and testing.

As far as concerns the dichotomy into grammatical and ungrammatical linguistic phenomena, it rather appears that these are rather two opposite end points on a gradient continuum. Some linguists (Sorace and Keller 2005, Kempen and Harbusch 2005) argue that there would be a fixed juncture along this continuum, which would separate the grammatical phenomena from the ungrammatical ones. However, as Sorace and Keller explicitly admit, their argument with respect to this dichotomy rests on the assumption that there are cross-linguistically universal rules or constraints which would constitute hard constraints, the satisfaction of which would be the minimum requirement for grammaticality or well-formedness. The universality of such hard constraints would entail that the requirements for grammaticality would be uniform for all languages. However, some others (e.g., Featherston 2005) argue that there is no “uniform single level of well-formedness that triggers [or excludes] the [production of] output, and that judgments are particular to each linguistic phenomenon. Featherston shows that the distributions of experimental ratings for variants, and the associated averages, maxima and minima, of different linguistic phenomena, within the same experiment by the same group of informants, can vary in absolute terms in comparison to each other. However, only the variant forms with the very highest ratings per each set of (related) phenomena are observed in corpora; variant forms with lower ratings within such a set do not appear, even though such ratings are higher or equal than the best ones for some other set. Thus, the lack of a clear juncture on a continuum, for instance concerning grammaticality, does not imply that there is not a clear distinction between the opposite ends of such a continuum.⁵ In other words, if we cannot pinpoint exactly (and universally) where grammatical (correct) ends and ungrammatical (incorrect) begins, it does not follow that the distinction is not real, i.e., that there are no (un)grammatical structures. Therefore, instead of scrapping a dichotomy between grammatical and ungrammatical, which Sampson argues for, we would neither regard as essentially any better the trichotomy suggested by him (p. 10).

experiment may be analyzed via two, clearly distinct perspectives (e.g., linguistically cued visual target choice and associated eye-movement), thus providing two types of data.

⁵ This common-sense point has been made earlier, for instance by Esa Itkonen.

To conclude, while we agree with Sampson that since Chomsky much of generative linguistics has been methodologically encapsulated (i.e., intuition-based; Wasow and Arnold 2005, and a myriad of others), Sampson's own attempt to solve the problem by taking the other extreme (naïve empiricism) leads to a pessimistic, unfruitful, and in the end, impossible conclusion. Sampson is right in arguing that linguistics cannot progress as long as intuitive evidence is (i) relied upon as the only type of evidence and (ii) deemed to have a privileged status over and above any other type of evidence. This is in fact the apparent reason for the fact that linguistics has not progressed much during its history that extends back over 3 or 4 millennia⁶. However, Sampson's extreme inductivist empiricism relying on (seemingly) only strict observation of spatio-temporal language use leads not only to a practical cul-de-sac (as he actually shows himself) but also to conceptual difficulties. Contrary to what he says should be done, what he in fact does himself is a par excellence example of the use of introspection/intuition in validating the correctness/incorrectness of particular instances of language use. Therefore, in constructing and testing linguistic theories both inductivist and hypothetico-deductive paradigms are necessarily needed (Allan 2003).

The message that we want to drive home is therefore: “Take empiricism seriously!” Language as an object of inquiry exhibits a continuum of distinct phenomena ranging from the physical (phonetic) via biological and psychological to social domains. As the object of inquiry should determine the methods of inquiry (e.g., Itkonen 1983), linguistics should therefore naturally conform to the principle of “methodological diversity” rather than to “methodological monism” which Sampson argues for. Sampson’s plea for the unity of science is based on a simple analogy between linguistics and natural sciences – a form of reductionism – and totally ignores the fact that their objects of inquiry are on the whole quite different. However, we should not forget that e.g., physics, the quintessential role model of natural science for many in the human sciences, is in the way it is practiced not so different, as many phenomena can best be explained by combining quite different conceptions of the object of research. Take for instance, light, which can be studied, and understood, as either particles or waves. To our minds the only plea for the scientific unity that is relevant to the present discussion is in the words of Wasow and Arnold (2005): “[L]inguistic inquiry should be subject to the methodological constraints typical of all scientific work.” And to end with

⁶ Linguistics, at least as a descriptive if not also as an analytic activity, can be considered to trace back to the first lexicographical and grammatical descriptions of Sumerian by Babylonians no later than in the second millennium BCE (overviewed in Black 1984).

an optimistic note, Wallace Chafe has argued for empirical diversity in the stead of empirical sectarianism so eloquently that we cite him here at length (the emphasis at the end is ours):

“But I continue to believe that one should not characterize linguists, or researchers of any kind, in terms of single favorite tie to reality. The term “corpus linguist” puts the emphasis on one tie to reality that has been neglected by many contemporary linguists, I believe to the great detriment of the field: a tie that must be vigorously pursued if our understanding of language and the mind is to enjoy significant progress. But there is a complementary danger in implying that that is all a linguist should do, of pitting corpus linguists against introspective linguists or experimental linguists or computational linguists. *I would like to see the day when we all be more versatile in our methodologies, skilled at integrating all the techniques we will be able to discover for understanding this most basic, most fascinating, but also most elusive manifestation of the human mind.*” (Chafe 1992: 96).

Languages are spoken and listened to, signed and watched, written and read, encrypted and decoded, studied and described, taught and learned, analyzed and generalized and imitated. Languages are represented as sounds or visual signs, scratched in a myriad of ways on a plethora of media. Languages are produced by an interaction of the lungs, vocal cords, throat, tongue, and mouth, or by the coordination of the hand(s). Languages are understood via the ears or the eyes, or even the skin, and in the end comprehended in the mind. Languages are uttered spontaneously for the instantaneous need of the moment, or recorded intentionally for the legacy of eternity. 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. 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.

References

Allan, Keith

2003 Linguistic metatheory. *Language Sciences* 25, 533-560.

Arppe, Antti and Juhani Järvikivi

(submitted 2006) Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory*.

Black, Jeremy A.

1984 *Sumerian Grammar in Babylonian Theory*, Studia Pohl: Series Maior 12. Rome: Biblical Institute Press.

Bresnan, Joan

2006 Is knowledge of syntax probabilistic? Experiments with the English dative alternation. *Pre-proceedings of the International Conference on Linguistic Evidence. Empirical, Theoretical and Computational Perspectives*, 2–4.2.2006, SFB441 “Linguistic Data Structures”, University of Tübingen, Germany, 3–10.

Cowart, Wayne

1997 *Experimental Syntax: Applying objective methods to sentence judgments*. Thousand Oaks: Sage Publications.

Chafe, Wallace

1992 The importance of corpus linguistics to understanding the nature of language. In Svartvik, J. (Ed.), *Directions in Corpus Linguistics*. Berlin: Mouton de Gruyter, 79-97.

Cutler, Anne (Ed.)

1982 *Slips of the Tongue and language Production*. Berlin: Mouton.

Featherston, Sam

2005 The Decathlon Model. In Kepser and Reis 2005a, 187–208.

Fromkin, Victoria

1971 The non-anomalous nature of anomalous utterances. *Language* 47, 27-52.

Gries, Stefan Th.

2002 Evidence in linguistics: Three approaches to genitives in English. In Brend, Ruth M., William J. Sullivan and Arle R. Lommel (eds.), *LACUS Forum XXVIII: What Constitutes Evidence in Linguistics?* Fullerton: LACUS, 17–31.

Gries, Stefan Th., Beate Hampe and Doris Schönefeld

2005a Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16–4, 635–676.

2005b. Converging evidence II: More on the association of verbs and constructions. In Newman, John and Rice, Sally (eds.), *Empirical and Experimental Methods in Cognitive/Functional Research*. Stanford: CSLI Publications.

Hokkanen, Tapani

2001 *Slips of the tongue: Errors, repairs, and a model*. Studia Fennica Linguistica (SFLIN)10. Pieksämäki: SKS.

Itkonen, Esa

1983 *Causality in linguistic theory*. London: Croom Helm

Keller, Frank

2000 *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. Ph.D Thesis, University of Edinburgh.

Kempen, Gerard and Karin Harbusch, Karin

2005 Grammaticality ratings and corpus frequencies. In Kepser and Reis 2005a, 329–349.

Kepser, Stephan and Marga Reis (eds.)

2005a. *Linguistic Evidence. Empirical, Theoretical and Computational Perspectives*. Studies in Generative Grammar 85. Berlin/New York: Mouton de Gruyter.

Labov, William

1970. The study of language in its social context. *Studium Generale* 23, 30–87. Page reference to reprint in Labov, William 1978. *Sociolinguistic Patterns*, Oxford: Blackwell.

Niemi, Jussi and Matti Laine

1997 Slips of the tongue as linguistic evidence: Finnish word initial segments and vowel harmony. *Folia Linguistica* XXXI, 161-175.

Roland, Douglas and Daniel Jurafsky

2002 Verb Sense and Verb Subcategorization Probabilities. In Stevenson, Suzanne and Paola Merlo (eds.), *The Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues*. Amsterdam & Philadelphia: John Benjamins Publishing Company, 325-346.

Sampson, Geoffrey

2005 Quantifying the shift towards empirical linguistics. *International Journal of Corpus Linguistics*, 10:1, 15–36.

2001 *Empirical Linguistics*. London/New York: Continuum.

Schütze, Carson T.

1996 *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.

Sorace, Antonella and Frank Keller

2005 Gradience in Linguistic Data. *Lingua*, 115:1, 1497-1524.

Wasow, Tom and Jennifer Arnold

2005 Intuitions in Linguistic Argumentation. *Lingua* 115:1, 1481-1496.