To be in: András Kertész, Csilla Rákosi, and Edith Moravcsik, eds, *Current Approaches to Syntax – a comparative handbook*.

# Grammaticality judgements in the light of corpus linguistics

Geoffrey Sampson

Sussex University

## 1 Languages do not have grammars

Natural human languages do not have grammars. There is no particular set of word-sequences (whether finitely or infinitely large) that constitutes "all and only" the grammatical sentences of a particular language or a particular speaker's idiolect. Consequently there are no things which could appropriately be called grammaticality judgements: there is no "grammaticality" for them to be judgements about. The grammatical habits of a language-community could be compared to the tracks that might develop in grassland among settlements of a pre-modern society lacking the institution of private land ownership. There will be some broad, well-trodden highways, corresponding to the most usual sentence structures, other narrower tracks, corresponding to less-common turns of phrase, and other cases again where one or two people have walked and scarcely left a disturbance in the grass. Even if no-one has ever yet happened to walk from point *A* to point *B*, nothing prevents someone doing so tomorrow. And, if he or she finds that route convenient, perhaps others will follow – the new route might eventually become a major thoroughfare.

Logically there must exist a distinction at any given moment between routes which have at some past time been taken by at least one person, and routes which no-one has ever walked. And similarly, since only a finite number of English-speakers have ever lived, there must be a distinction between those strings of English words which have been uttered at least once in the history of the language, and those which have never yet been uttered. But these are not interesting distinctions of principle. New sentences are constantly being assembled and uttered, and although some of these will conform perfectly to patterns found in many previous utterances, others will deviate from prior experience, in minor or perhaps in major ways. As John Taylor (2012: 285) summarizes the findings of his corpus research, "Speakers ... are prone to *innovate* with respect to previous usage, using words in ways not already sanctioned by previous experience". For an innovative utterance to work, its hearer(s) need to grasp more or less what the speaker intends by it. But, evidently, hearers often do.

A map of the kind of grassland territory I have described would include the broadest, most-frequented tracks, but it would need to impose some essentially arbitrary cut-off between paths well-defined enough to chart and routes too occasional to be recorded. Likewise, a grammatical description of a natural language will identify the best-established sentence structures, but could not hope to cover every structure that some speaker has occasionally used, or might use in the future. Any grammar must limit itself to describing usage down to some frequency threshold, and that threshold will be governed by practical issues such as the quantity of time and manpower available for compiling the description. There is no "natural" place to set the threshold. "Starred sentences" are a myth.

A radical point of view can be persuasive only with a concrete example, so let me give one. My wife and I recently addressed the problem of one of our cats stealing the other's food by buying a new type of feeding-stations with lids that open and close automatically under the control of the individual pet's microchip. The makers, Sureflap, are an English firm founded recently by a Cambridge physicist, and the manual provided is well written. So I was initially surprised to encounter a section headed "Learning your pet into the feeder" and beginning "When learning your pet into the feeder, make sure all other pets are kept away". (It explains how to get the mechanism to respond to a particular pet.) Surely these word-sequences are not English? – *learn* does not take an animate object, or an *into* phrase. But the activity described is novel, and the writer has used English in a novel way to refer to it. I might have preferred to write "Teaching the feeder to recognize your pet" – but that would not be quite right, because the change to the feeding-station is instantaneous, brought about by a single press of a button, it is the cat which has to be gradually taught to exploit its resulting behaviour. Perhaps there would be some other form of words which would have been faithful to that reality and yet deviated less from established usage; but the manual writer chose the words I quoted, and he or she is doubtless as much an English native speaker as I am, so who am I to say the wording is not English? It did not seem so previously, because no English-speaker had found occasion to use *learn* that way. But now someone has had a reason to use *learn* with that grammar, and I and other native speakers can certainly understand what is intended. If Sureflap prospers, in years to come probably no-one will bat an eyelid at this way of using *learn*.

Many theoretical linguists have a concept of "grammaticality" according to which, at a given time, some fixed (though infinitely numerous) class of word-sequences are "grammatical" in a given idiolect, though from time to time the rules of the language or idiolect change so that new word-sequences become grammatical. They would describe the *learn your pet* usage as one that is currently ungrammatical (for most speakers) but which may be destined to become grammatical, under the influence of things such as the Sureflap manual. I do not believe in this concept of "grammaticality". Putting words together in novel ways in order to express novel

ideas is part of competent language behaviour. Perhaps the *learn your pet* example seems a rather extreme case – but it might strike us that way only because it was encountered in print, and it could well be that innovation, or major innovation, occurs most often in speech. In any case, grammar that departs from prior norms is normal.

The idea that some sequences of words of a language are "grammatical" and the rest "ungrammatical" is not a self-evident one. So far as I am aware it never occurred before Noam Chomsky (1957: 13) defined a "language" as "a set ... of sentences" and stated that the "fundamental aim in the linguistic analysis of a language L is to separate the *grammatical* sequences ... from the *ungrammatical* sequences ..." – though from then on it was accepted by the discipline remarkably uncritically. In our book *Grammar Without Grammaticality*, Anna Babarczy and I have pointed out that earlier grammarians, even formal grammarians, discussed the grammatical structures which do occur in a language without feeling a need to contrast them, explicitly or implicitly, with "starred sentences" (Sampson and Babarczy 2014: 1–6), and we argue that they were correct in that. Chomsky was misled, at the outset of his career, by a false analogy between human languages and computer programming "languages". A programming language, such as Java or C, really is defined by a fixed, clearcut generative grammar – if it were not, it could not be used with computers. But human languages are very different kinds of thing from programming "languages", in this (and other) respects. (For the intellectual influences which moulded Chomsky's thinking as a young man, in particular his collaboration with the mathematician Marcel-Paul Schützenberger, see Sampson 2016a.)

## 2 What could be evidence for grammaticality?

If the existence, in human languages, of a distinction between grammatical and ungrammatical word-sequences were more than a dogmatic article of faith, it would have to have some observable correlates. The question is what these might be.

One way in which grammaticality could very convincingly be shown to be a real property would be through successful construction, for one or more human languages, of generative grammars as envisaged by Chomsky (1957), which came close to the ideal of covering all the sentence structures actually used in practice while failing to cover numerous word-strings that are never uttered or written. But that goal was abandoned long ago. The last serious and well-known attempt at a generative grammar for English that I know of was Stockwell et al. (1973), and these authors summarized the conclusion they eventually reached about the feasibility of the task by quoting the seventeenth-century grammarian James Howell (spelling modernized):

> the English ... having such varieties of incertitudes, changes, and idioms, it cannot be in the compass of human brain to compile an exact regular syntaxis

thereof

Maurice Gross (1979) wrote interestingly about the failure of a more-than-ten-year effort by a team of linguists to produce a satisfactory generative grammar for French. The generative linguist Cedric Boeckx has claimed (2006: 220) that the "early goal" of "a grammar that would describe all the well-formed sequences and none of the ill-formed ones ... was abandoned as soon as it was formulated", as if the goal was given up for some principled reason (though the truth, I believe, is that attempts faded away after various linguists had discovered from experience that the task was impossible). In fact some work continued after the 1970s – the "Alvey Natural Language Tools Grammar" of English was being developed at the University of Cambridge Computer Laboratory into the early 1990s[1] – but no linguists today, I believe, see developing a generative grammar for a human language as a realistic goal.

One factor which those who believe in grammaticality point to as explaining the futility of trying to compile adequate generative grammars is the Chomskyan contrast between "competence" and "performance", and this might offer an alternative approach to finding an empirical basis for "grammaticality". Even if a native speaker's mental linguistic competence implies some limited, definable class of word strings as "all and only" the grammatical sentences of his language, the utterance-patterns produced in practice will be affected by various extra-linguistic performance factors, such as slips of the tongue, memory limitations, and so forth, and these might interfere so unsystematically with the class of sentences defined by competence that the class of utterances actually produced will not be definable by any finite system of generative rules.

It is tempting to dismiss this move as one which renders the idea of grammaticality unrefutable, and hence a mere empty dogma. That may be correct; but it could be unfair. If performance factors interfere with linguistic competence in ways that are essentially random from a grammatical point of view, then one might expect to find a frequency difference between structures that are grammatical in a language, and those which are ungrammatical but occur nevertheless as a consequence of "performance factors". Structures of the former type ought to recur repeatedly in a sufficiently large corpus, while the latter would be "one-offs" or at least recur very rarely. Some sort of discontinuity ought to be visible in the spectrum of frequencies of different grammatical structures, and observation of such a discontinuity would provide evidence that "grammaticality" is a real property despite the fact that its extension cannot be defined.

In the early decades of generative linguistics there was no practical way to research this kind of issue. Nowadays, with the availability of treebanks (electronic corpora equipped with syntactic annotations), one can do so fairly easily.

1 See <www.cl.cam.ac.uk/users/ejb/anlt-gram.pdf>.

In *Grammar Without Grammaticality* Anna Babarczy and I examined construction frequencies in an English treebank, and found no discontinuity: constructions occurred in our data at all frequencies from very common down to one-offs. The tracks in the grassland turned out to vary smoothly from very wide to evanescent, as it were, with nothing akin to the contrast we find in modern societies between recognized public rights of way and occasional routes taken by trespassers. The grammaticality concept seemed to have no leg to stand on.

We dealt in that book with most of the criticisms that had been levelled against this point of view as expressed in earlier publications of mine. One criticism, though, did have some force: the treebank we had used (the SUSANNE Corpus) was one for whose compilation we were ourselves responsible, and this appeared to create a danger that we might have built into our data the conclusion that we wanted to prove. There was a good reason for using SUSANNE: the project which created it gave an unusually high priority to precision and refinement of the annotation scheme (as has been noted by independent third parties, e.g. Lin 2003: 321), while some other treebank projects have instead prioritized quantity of material annotated. Nevertheless, I accept that the danger appeared to exist (though I do not myself believe it was real).

Another criticism which I would also have accepted as fair (though no-one made this criticism to my knowledge) was that research of this kind ought not to consider exclusively English. Apart from the fact that it is unhealthy for research on general linguistics to limit itself to looking at a single language, English with its very limited system of grammatical inflexion could be a language in which the bounds of grammaticality are unusually vague.

And of course for research on frequencies, a larger data-set will always be better than a smaller one. The SUSANNE Corpus contains only 130,000 words, so that a discontinuity occurring anywhere on the frequency spectrum below 1 in 130,000 words would be invisible in that data-set.

Accordingly, for this chapter I have looked at the same issue using a treebank compiled by researchers unconnected with myself, larger than the SUSANNE Corpus, and representing a language other than English: namely version 2.0 of the NEGRA Corpus of German newspaper prose,[2] compiled at the Universities of Stuttgart and the Saarland and containing about 20,000 sentences comprising 355,000 "tokens" (words, punctuation marks, etc.). I thank Tania Avgustinova of the University of the Saarland for making this resource available to me.

## 3 The NEGRA experiment

German is arguably a particularly suitable language for this investigation. In the first

2 See <www.coli.uni-saarland.de/projects/sfb378/negra-corpus/>.

place, having much more inflexion than English, it makes grammatical structures relatively explicit and hence, perhaps, more open to clearcut delimitation. But, also, it is a socially "disciplined" language in a way that English has never been. English might be described as an anarchic language: no organ of the British State has ever claimed a right to regulate the language, and I believe the same is true in other English-speaking countries. National and provincial governments in German-speaking Europe, on the other hand, regard themselves as "owning" the German language. One very noteworthy way in which this attitude manifested itself recently was that these governments jointly imposed a set of language reforms (on which see Johnson 2005), including regulations which instructed schoolteachers precisely how they should in future mark down deviations from the new norms in pupils' written work. The reforms led to a series of controversies fought out in law courts, in a way that simply could not happen in Britain. Between 1997 and 1999 German orthography in the province of Schleswig-Holstein changed three times, with schools legally required to modify their teaching accordingly (op. cit.: 111–15).

A striking feature of these reforms was that many of the changes were of a kind whose English-language equivalents would be below the radar for most literate English speakers. The school I attended in England in the 1950s would certainly by 21st-century standards be seen as traditional in its syllabus and teaching methods, and it put a great deal of effort into training pupils to write well, but at no time in my schooldays did I hear any suggestion that there were specific rules for the use of punctuation marks such as commas – in English there are no definite rules. Pupils were expected to absorb good punctuation practice tacitly through the activities of reading, writing, and having their prose criticized. German, on the other hand, has explicit syntax-based rules for punctuation. One feature of the 1996 reforms, for instance, was to make commas newly optional before co-ordinating conjunctions (Johnson 2005: 72). If a contrast between "grammatical" and "ungrammatical" structures were a reality for any human language, one would surely expect it to be a reality for German.

Furthermore, all material in the NEGRA Corpus is taken from a single publication, the *Frankfurter Rundschau* daily newspaper. For many research purposes this would be a disadvantage, making the corpus less representative than it might be, but for present purposes it is if anything a positive factor. Newspaper publishers commonly require their journalists to conform to a set house style, so a single-newspaper corpus offers a specially stiff test of my contention that there are no definite boundaries between the grammatical and the ungrammatical.

The NEGRA Corpus exists in alternative forms, of which I have used the one called <negra-corpus.penn>. To give an impression of the nature of the analysis, Figure 1 shows the structure assigned to the short sentence *Schade jedoch, daß kaum jemand daran teilhaben kann.* (This is NEGRA sentence 5, which translates as "All the same, it's a pity that scarcely anyone is able to share in it.")

FIGURE 1 ABOUT HERE

Where a node label contains a hyphen, the symbol after the hyphen identifies the function of the constituent within its containing tagma. The function-tagset has 45 members. Symbols preceding hyphens, or occurring as sole node-labels, are drawn either from a 57-member set of wordclass tags, or from a 25-member set of tagma-type labels. For the full tagsets the reader is referred to the NEGRA documentation, but the tags occurring in Figure 1 are defined in Table 1. (Because German grammar differs from that of English, the wordtag categories in particular often have no direct English equivalents, and since the NEGRA documentation defines them in German I have taken the liberty of modifying some of the definitions in order to clarify the categories for English-speaking readers.)

TABLE 1 ABOUT HERE

In order to investigate the grammaticality issue, we can treat a parse-tree as a set of productions (pairings of a mother-node label with a sequence of daughter-node labels), and examine the statistical distribution of distinct production types over the corpus as a whole. In registering productions I have modified the raw NEGRA data in three ways. Most importantly, although functiontags are for convenience shown in the corpus as parts of node-labels, logically (as the NEGRA documentation rightly says) they apply not to nodes but to the lines linking mother to daughter nodes – they label the relationship between a constituent and its containing tagma. Consequently, in listing productions I include functiontags where they occur in the labels of daughter nodes, but delete them from the labels of mother nodes. Secondly, NEGRA leaves root nodes unlabelled, but for practical convenience I have given root nodes a distinctive label of their own, namely "O". And lastly (although there happens not to be an example in Figure 1), some NEGRA parse-trees include pairs of nodes bearing "trace" labels showing where a constituent occurs in a different surface-structure tagma from the tagma in which it plays a logical role. NEGRA trace labels are of the form **T1**, **T2**, etc., but the only significance of the digits is to define the correct pairing of traces in trees containing more than one pair. If two productions differ only in that one has **T1** where the other has **T2**, this does not make them "different grammatical constructions" in any meaningful sense. Therefore in my investigation all trace labels are rewritten as **T0**.

The tree of Figure 1, then, comprises five productions:

O → S $.
S → ADJD-PD ADV-MO $, S-SB
S → KOUS-CP NP-SB VP-OC VMFIN-HD
NP → ADV-MO PIS-NK

VP → PROAV-MO VVINF-HD

The production O → S $. is the most frequent in NEGRA as a whole: it is instantiated 12,750 times. At the other end of the frequency scale, there are almost 19,000 *hapax legomena* – production-types instantiated just once each. (A few of these are listed, with the wording realizing them, their NEGRA sentence numbers, and English translations, in Table 2. In this table, immediate constituents are single words except as indicated by square brackets, e.g. in sentence 4878 the label AP-MO applies to the phrase *zu schlecht*, the label VVPP-HD to the word *erschlossen*.)

TABLE 2 ABOUT HERE

Figure 2 plots production-type frequencies against the numbers of different production-types instantiated at the given frequencies (i.e. the frequencies of frequencies). Both axes are scaled logarithmically. To make the figures at higher frequencies meaningful, zero figures are averaged with the nearest non-zero figure. (Precise details of the derivation of Figure 2 from the NEGRA data can be read off from the software used, available at <www.grsampson.net/SNegFofs.html>.)

The plot of Figure 2 extends smoothly and log-linearly from a handful of high-frequency productions on the right to large numbers of hapaxes and low-frequency production-types on the left. I see no hint of discontinuity. There is no suggestion of a large gap or U-shaped trend indicating bimodality. Reading from right to left, where do "grammatical" productions end and ungrammatical oddities of performance begin?

FIGURE 2 ABOUT HERE

Christopher Culy (1998) responded to the earlier version of my argument, which used English-language data, by claiming that it is possible to devise probabilistic generative grammars which yield frequency-of-frequency distributions not dissimilar to Figure 2. But (apart from the fact that Culy's plots did not look *very* like Figure 2, or like the SUSANNE-based plot which he had seen), as a would-be defender of "grammaticality" Culy was missing the point. If it were possible to produce successful generative grammars for English, German, or other human languages, they would be excellent evidence in themselves for the reality of "grammaticality". We would not care what plots of frequency data looked like. It is worth examining data like those of Figure 2 just because, as everyone now accepts, producing accurate generative grammars has turned out to be a hopeless task. If frequency data had revealed a discontinuity between high-frequency grammatical constructions and one-off or rare oddities of performance, that might have been a way to show that grammaticality is a real thing even though no-one can capture it in a generative grammar. But there is no discontinuity. So far as I can tell, we have no evidence of any kind for grammaticality. It really is just an unsupported dogma.

It might be, of course, that other evidence, of kinds that have not occurred to me, is waiting to be uncovered and put forward. But linguists who defend the grammaticality concept never put such evidence forward. In my experience they show little interest in empirical evidence – they rarely work with corpora. They see it as enough to assert that native speakers of a language have "grammaticality intuitions".

## 4 Where do grammaticality judgements come from?

If there is no such thing, in reality, as an "ungrammatical sentence", where do the reactions that linguists call "grammaticality intuitions" or "grammaticality judgements" come from? There is no doubt that speakers who are asked "Can you say XYZ in your language?" often feel able to give a yes or no response (though there are also plenty of cases where they feel puzzled to know how to answer). What is more, there are often clear patterns in the responses to related word-strings – informants do not just answer yes or no at random, as one might perhaps expect they should if the "grammaticality" concept corresponded to nothing at all in reality. Linguists sometimes point to research establishing such patterning (e.g. Schütze 1996) as confirmation that, despite appearances, intuition-based grammar development is a genuine empirical science.

But someone who is asked "Can you say XYZ?" has to decide how to interpret the question – its meaning is not self-evident. There are plenty of ways to interpret such a question which do not involve postulating a distinction between "well-formed" or "grammatical" word-strings and ill-formed strings. For instance, one obvious interpretation would be "Can you imagine circumstances in which you might want to say XYZ?" If XYZ asserts something which is obviously false or absurd, the answer is likely to be no – not because XYZ is "ungrammatical", but because people do not normally want to assert absurdities.

This point has been overlooked from a very early stage in the history of generative linguistics. Noam Chomsky's *Aspects* (Chomsky 1965) claimed that the innate machinery of language competence included a complex system of "selection rules" and "subcategorization rules", whose function was to disallow sentences such as the English examples:

> the boy may frighten sincerity
> John amazed the injustice of that decision
> the book dispersed

These sentences are certainly odd, but not because they are "ungrammatical". Sincerity, and the injustice of a decision, are abstractions, which as such feel no

emotions; it is obviously impossible to frighten or amaze an abstraction. To "disperse" means for an aggregate of physically-separate items to spread apart and become sparse; a book is not such an aggregate, it is a single physical object, so it cannot disperse. These examples conform perfectly to the norms of English grammar, contrary to Chomsky's statement (1965: 76) that they "deviate ... from the rules of English". They correspond in my grassland analogy to broad, well-trodden highways. It is just because the examples are grammatically normal that we can easily see what they mean – and hence see that they assert absurdities.

It is true that, if the absurdity of a proposition is *too* obvious, linguists will not be tempted to call a sentence asserting the proposition "ungrammatical". We could not imagine wanting to say *My daughter is 500 years old*, or *I ate a lorry*, but because anyone can immediately see why these assertions are absurd they do not look for explanations in terms of grammar. On the other hand, one has to ponder a little about the precise meaning of *disperse*, or about the fact that *sincerity* refers to an abstract character-trait rather than to a bearer of the trait, in order to see that the propositions in question are impossibilities. Only brief pondering is needed, but many linguists have not done even that much before leaping to the conclusion that there must be something about the English language which disallows those examples.

After the publication of *Aspects*, selection and subcategorization rules became part of linguists' standard grammatical apparatus for many years. I do not know how far they are still accepted by linguistic theorists today, but I do know that invalid leaps from "speakers don't say XYZ" to "there must be some principle or mechanism in speakers' minds that renders XYZ ungrammatical" are as widespread today as they ever have been.

Bernd Heine and Heiko Narrog's *Oxford Handbook of Linguistic Analysis* (Heine and Narrog, 2nd edn 2015) is a standard collection on (mainly) the grammatical aspects of modern theoretical linguistics from one of the world's leading academic publishers. It is reasonable to assume that the approaches espoused by its various authors are representative of the best-established styles of linguistic research in the early 21st century. Yan Huang's chapter in that *Handbook*, "Neo-Gricean pragmatic theory of conversational implicature", remarks that "we can say *They summered in Scotland* [but] cannot say **They falled in Canada*", and he explains this by postulating a universal linguistic principle which he calls "pre-emption": the fact that *fall* has the verb sense "drop down" blocks it from being used as a verb similar to *summer* meaning "spend the relevant season".

But we need no linguistic principle to explain the difference between the *summered* and *falled* examples. That difference arises because there are (or at least have been) recognized, established social institutions, among people whose circumstances allow(ed) it, of spending whole summers, or whole winters, away from home in places with pleasanter weather: hence "to summer" and "to winter". In spring and autumn

the weather is not extreme, so there was never an institution of spending those seasons away from home. In Britain we call the season following summer *autumn* rather than *fall* as in American English, but (although "pre-emption" would be irrelevant to *autumn*) we too do not say things like *They autumned in Canada. They falled ...*, or *they autumned ...*, are not "ungrammatical": they are just not customary, for a good non-linguistic reason. (If a new custom arose of spending the autumn season away, presumably for some non-weather-related reason, then Americans and Britons probably would begin using *to fall*/*to autumn* in that sense. Certainly no linguistic "pre-emption principle" would stop them doing so.)

In the examples discussed above, there were at least good reasons why speakers find them odd, though linguists were wrong to locate the oddness in the structure of the language. In other cases it seems that linguists have answered "no" to the question "can you say XYZ?" merely because they were not imaginative enough to think of circumstances in which they might want to say it.

In Sampson and Babarczy (2014: 81–2) we discussed a case where, beginning in the 1970s, linguist after linguist asserted that a particular type of clause is ungrammatical in English, as a preliminary to proposing theoretical mechanisms that could be used to disallow it, despite the fact that the clause-type is in reality perfectly normal and commonplace. Again this approach remains alive and well in the 21st century. Another chapter in Heine and Narrog (2015) is Vilmos Ágel and Klaus Fischer's "Dependency grammar and valency theory", which notes a difference between the grammars of English and Hungarian: in Hungarian, but not in English, it is acceptable (they say) for the verb *lie* (tell an untruth) to take a complement clause expressing the content of the lie. What sort of "unacceptable" would that be? I googled *lied that* and was offered "about 328,000 results", beginning with *Have you ever lied that you had a boyfriend ...* and *Kelly Baker lied that a young member of her family had cancer ...* There were a few irrelevant examples using *lied* as the noun meaning a type of song, but the great majority directly refuted Ágel and Fischer's claim.

Linguists who make ungrammaticality claims that are at odds with observable usage sometimes explain this in terms of idiolect differences: perhaps the usage is grammatical for many fellow-speakers, "but I can't say it". That is analogous to someone who always walks home from work by one route, ignoring another street which would serve equally well to get home by, and who explains this by saying "I can't walk down that street". Of course he *could* walk that way, he has just got into a different habit. What is described as an "ungrammaticality judgement" would in this case be better described as a perception of novelty.

Undoubtedly there are many other factors that may give rise to the feelings which theoretical linguists think of as "grammaticality judgements". In William Labov's famous account of Philadelphian positive *any more* (Labov 1975: 35–6), what seemed to be going on was that his informants firmly believed that a turn of phrase which they

regularly and systematically used was impossible and meaningless, because they knew that it was not current in a higher-prestige majority dialect. And there must be other relevant factors again. But none of these factors, so far as I am aware of them, justifies a model of language which divides the class of all possible strings of words of a language into a set of "grammatical" sheep and a complementary set of "ungrammatical" goats.

## 5 Is anyone a true believer?

The case against "grammaticality" seems overwhelming. But that raises a large question about how the concept can have been accepted so widely for so long. Charles Hockett (1968) argued a contrary point of view, but although he was one of the best-known linguists of his day his book was politely ignored. Another leading name in mid-20th-century linguistics, Fred Householder (1973: 371), discussed how difficult it is in practice to come up with clear cases of "starred sentences" in English, but again his objection achieved no traction. Yet it was not as if the arguments against grammaticality are particularly convoluted or depend on obscure data. Most of the points I have made above could have been made by any linguist at any time. My analysis of the NEGRA data did depend on the availability of electronic treebanks, which are a fairly new thing; but I believe few linguists will have been surprised by the findings summarized graphically in my Figure 2.

Indeed I have the impression that even generative linguists whose work depends on a belief in grammaticality know quite well that the picture of human language offered by Figure 2 is a fair one. My reason for saying that is that, although I have been putting forward the view argued in this chapter for thirty years now, beginning with Sampson (1987), and plenty of linguists have argued against my view, so far as I am aware no defender of "grammaticality" has ever voiced the kinds of objection that I would expect them to make, if they believed that the model of language expressed in a diagram like Figure 2 was misleading.

Defenders of grammaticality could argue that the treebanks I have used as data sources are too small to provide good evidence against the grammaticality concept. Or they might suggest that the fuzzy continuity of a plot like Figure 2 merely shows that the parsing schemes embodied in the relevant treebanks – the shapes of the parse-trees assigned to particular sentences, and the range of grammatical categories from which node-labels are drawn – are ill-chosen. On the face of things, these might seem quite reasonable objections. Creating treebanks is such a labour-intensive activity that reliably-analysed treebanks are inevitably smaller than one would like, but the consequence is that a hypothetical discontinuity in frequency statistics would be invisible in the SUSANNE treebank unless it occurred at a higher point on the frequency spectrum than 1 in 130,000 words, and even in the NEGRA Corpus it would need to occur somewhere above 1 in 355,000 words. These frequencies are not all that

low. Someone might well urge that a discontinuity is there, but that to observe it one would need a treebank of millions of words, or hundreds of millions of words. Furthermore, correct labelled parse trees for language examples are not self-evident. A treebank development project has to evolve an explicit parsing scheme for itself; normally such a scheme will try to agree with consensus views about grammatical structure wherever possible, but that principle is far from enough to settle every question that arises, and the parsing schemes of different treebank projects differ in many respects. (There are features of the NEGRA parsing scheme for German that I find surprising – but if one wishes to avoid the suspicion of circularity arising from using one's own analyses as data, the only alternative is to use someone else's analyses.) It seems quite possible that a statistical discontinuity between "structures which are grammatical" and "performance oddities", which would emerge clearly from a "correctly" analysed treebank (whatever "correct" might mean in this context), could be blurred if the same language samples are analysed in line with an "incorrect" parsing scheme.

So far as I am aware, proponents of grammaticality have not made either of these objections, reasonable though they might appear. The counter-arguments I have encountered have been pitched at a more abstract, aprioristic level. This could just be because the objectors are aware that making good the objections I have outlined would involve work with detailed empirical language data which they would find uncongenial. However, I suspect that a main reason is that, in their hearts, even believers in grammaticality know that data like Figure 2 are in fact a fair representation of how human languages are. They accept that, however much one were to increase the quantity of data examined, and whatever reasonable parsing scheme one were to use, the picture would not change radically.

## 6 How is the myth maintained?

So one is left with the question why so many academics have accepted such an implausible concept of grammar. This question is answerable, but to answer it one must draw attention to issues which professional academics commonly prefer to leave undiscussed.

In the first place, theoretical linguistics as practised during my lifetime has turned into a subject in which recklessness about scholarly accuracy is treated as acceptable. A striking example (to which I have drawn attention before) relates to Jerry Fodor's 1975 book *The Language of Thought*, which is regarded by many linguists as a key work showing that shared innate structuring applies not merely to the area of language mainly considered by Noam Chomsky, namely grammar, but even to vocabulary, which on the face of things seems the most obviously culture-specific and diverse area. Fodor argued at length and with apparent conviction that the vocabularies of all human languages are fundamentally alike. Yet in his 200-plus pages he only ever once

mentioned an example drawn from any language apart from English: he quoted the French word for "dog" ... and got it wrong, writing *chein* rather than *chien*. Nor is this kind of casual inaccuracy restricted to the Chomskyan, generative school. Currently, one of the best known opponents of that school is the "cognitive" linguist Vyvyan Evans. His 2015 book *The Crucible of Language* offers a picture of how meaning works in human language which explicitly contradicts Fodor and the generative school. Again Evans is explicitly making claims about universal features applying to all human languages, but again he offers few non-English examples. Among the handful of such examples in his 320 pages, Evans quotes the German word for "cat" ... and gets it wrong, writing *Kätze* rather than *Katze*. (I have discussed elsewhere, Sampson 2016b, how Evans mistakes features which characterize the English language in particular for universal properties of language in general.)

It is as if writers like these are telling us "Sure, we can't be bothered to look at languages other than our own or to get elementary facts right; but you have to agree with what we tell you about the general nature of human language, because we are university professors". And in the academic world as it has evolved in the decades around the millennium, this attitude is not entirely absurd. A high proportion of readers of books like Fodor's or Evans's are students, who find such books listed as course homework by their teachers, and know that to get the marks they need for their degrees they must learn to reproduce what the books say in their essays and exam scripts.

I have quoted two specific examples of this style of linguistics, because serious criticism is not persuasive without concrete examples of what one is criticizing. But this general style of work is all too common in theoretical linguistics, and Fodor and Evans are probably not much more blameworthy than many others. Departments of linguistics contain plenty of academics doing solid work on the description of particular languages and language-families. But much of what is published under the heading of general "linguistic theory" scarcely counts as scholarship in the traditional sense at all. It might be better seen as an academic branch of show business, where the aim is not to get things right but to be famous and admired. It is hard to see such work as entitled to more respect than we accord to film stars' or pop singers' pronouncements (which are often reported as significant by the media) about political or current-affairs issues.

But there is also a more specific reason for widespread acceptance of the grammaticality concept, which relates to changes in university management.

During my own career I have seen the circumstances of university employment radically transformed. When I entered the profession in the 1960s, staff in leading British universities were paid essentially to teach, and they engaged in research (if they did engage in research) mainly for personal satisfaction. There were no pressures pushing people towards one intellectual position rather than another.

Today, all university staff are made aware that advancement in the profession, or perhaps even keeping the job one is in, depend crucially on publishing research and on inducing outside bodies to sponsor one's research. There is great pressure to choose fields which provide good opportunities for these things. (I describe the British case, as the one I am familiar with, but I believe universities elsewhere have gone through similar changes, sometimes earlier than in Britain and sometimes more recently.)

In linguistics, this pressure creates a powerful motive for believing in grammaticality. If a language is defined by rules of grammar which imply a clearcut distinction between "grammatical" and "ungrammatical" sequences of words, and which can be inferred from judgements about the status of individual word-sequences, then this creates a fertile research field. A high proportion of "theoretical" articles in linguistics journals, for many years now, have dealt with formalizing grammatical rules to define some aspect of the structure of some language, or with what such rules suggest about universal features or constraints applying to the grammars of all human languages. On the other hand, if there is no grammaticality and human languages are not defined by specific sets of grammar rules, then much of theoretical linguistics becomes an empty non-subject. If languages do not have grammar rules, then there is no room for universal properties applying to the rules of all languages. There is just much less to say about human language in general than theoretical linguists suppose, and hence less scope for proposing projects that might appeal to research sponsors. It is not so much that present-day theoretical linguistics consists of theories which are false. Depending on precise wording, some of them may be false, but many will be neither false nor true: there is nothing in the real world to which they apply, accurately or otherwise.

Small wonder then if, for many 21st-century linguists, grammaticality simply cannot be allowed to be unreal. If it is unreal, much of the discipline, and many university posts for linguists, lose their *raison d'être*. To imagine that considerations like this do not have a powerful influence on climates of academic opinion would be naïve. There is no implication of conscious dishonesty here. People are good at not permitting dangerous ideas to rise to the conscious surface of their mind. When I became convinced of the emptiness of theoretical linguistics, I shifted academic affiliation and spent the second half of my career until retirement in a department of computer science, a subject which does not raise troubling problems about whether the material one is teaching students is worth their while to learn. But not all academic linguists are prepared for that kind of intellectual retooling. So they resist recognizing the truth about "grammaticality".

## 7 Grammaticality as caboose

Academic research should be about establishing truth, however, not about providing

careers for academics. Grammaticality is an academic counterpart of the "caboose" which once brought up the rear of every goods train on the American railway network. (The British term was "guard's van", but American trains still had cabooses when guard's vans had become history in Britain.) By the 1980s, as I understand it, cabooses had in many cases lost their original safety function, and for a while they served only to create jobs for the railwaymen who rode in them. Eventually, like comparable make-work arrangements in other industries, cabooses were swept away. It is high time for grammaticality, grammaticality judgements, and associated areas of theoretical linguistics to follow the caboose into history.

References

Boeckx, Cedric. 2006. Review of Postal, *Skeptical Linguistic Essays. Journal of Linguistics* 42.216–21.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.

Culy, Christopher. 1998. Statistical distribution and the grammatical/ungrammatical distinction. *Grammars* 1.1–13.

Evans, Vyvyan. 2015. *The Crucible of Language: how language and mind create meaning*. Cambridge: Cambridge University Press.

Fodor, Jerry A. 1975. *The Language of Thought*. New York: Crowell.

Gross, Maurice. 1979. On the failure of generative grammar. *Language* 55.859–85.

Heine, Bernd and Heiko Narrog, eds. 2015. *The Oxford Handbook of Linguistic Analysis*, 2nd edn. Oxford: Oxford University Press.

Hockett, Charles F. 1968. *The State of the Art*. The Hague: Mouton.

Householder, Fred W. 1973. On arguments from asterisks. *Foundations of Language* 10.365–76.

Johnson, Sally. 2005. *Spelling Trouble? Language, ideology and the reform of German orthography*. Clevedon, Somerset: Multilingual Matters.

Labov, William. 1975. *What is a Linguistic Fact?* Lisse: Peter De Ridder Press.

Lin Dekang. 2003. Dependency-based evaluation of Minipar. In Anne Abeillé, ed., *Treebanks*, 317–29. Dordrecht: Kluwer.

Sampson, Geoffrey. 1987. Evidence against the "grammatical"/"ungrammatical" distinction. In Willem Meijs, ed., *Corpus Linguistics and Beyond*, 219–26. Amsterdam: Rodopi. A version reprinted in Sampson, *Empirical Linguistics*, ch. 10. London: Continuum.

Sampson, Geoffrey. 2016a. Rigid strings and flaky snowflakes. To be in *Language and Cognition*, vol. 8 no. 3.

Sampson, Geoffrey. 2016b. Review of Evans 2015. *Language and Dialogue* 6.329–36.

Sampson, Geoffrey, and Anna Babarczy. 2014. *Grammar Without Grammaticality: growth and limits of grammatical precision*. Berlin: De Gruyter.

Schütze, Carson T. 1996. *The Empirical Base of Linguistics: grammaticality judgments and linguistic methodology*. Chicago: University of Chicago Press.

Stockwell, Robert P., et al. 1973. *The Major Syntactic Structures of English*. New York: Holt, Rinehart and Winston.
Taylor, John R. 2012. *The Mental Corpus: how language is represented in the mind*. Oxford: Oxford University Press.

| | |
|---|---|
| wordtags: | |
| $. | sentence-final punctuation |
| $, | comma |
| ADJD | adverbial or predicative adjective |
| ADV | adverb |
| KOUS | conjunction introducing finite subordinate clause |
| PIS | indefinite pronoun functioning as NP |
| PROAV | preposition–pronoun portmanteau[3] |
| VMFIN | finite modal verb |
| VVINF | full infinitive verb |
| tagma-tags: | |
| NP | noun phrase |
| O | root node |
| S | sentence |
| VP | non-finite verb phrase |
| function-tags: | |
| CP | complementizer |
| HD | head |
| MO | modifier |
| NK | noun kernel modifier |
| OC | clausal object |
| PD | predicate |
| SB | subject |

Table 1

3 In the NEGRA documentation, this category is listed as "PAV", but in the treebank itself it appears as "PROAV".

NP → PIAT-NK AP-NK ADJA-NK NN-NK
*keine* [*größere* **T0**] *rechtliche Qualität* (Sentence 700)
no greater legal status [than ... ]

VP → AP-MO VVPP-HD PP-MO
[*zu schlecht*] *erschlossen* [*für die Schüler*] (Sentence 4878)
too poorly accessible for the pupils

PP → ADV-MO APPR-AC ART-NK ADJA-NK NN-NK CAP-MNR
*rund um den ehemaligen Schrottplatz* [*nördlich des Höllweges und westlich des Park-and-ride-Platzes*] (Sentence 19114)
all round the former scrapyard north of the Höllweg and west of the park-and-ride area

Table 2

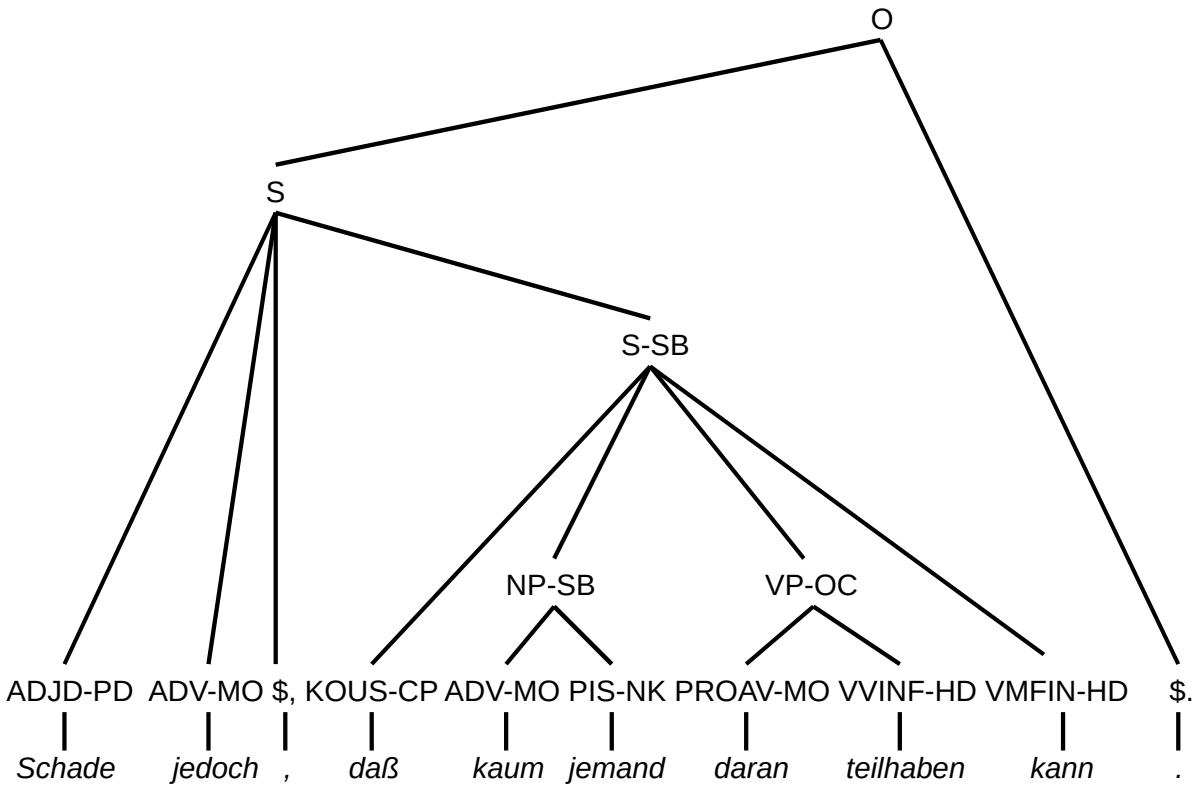
O
S
S-SB
NP-SB
VP-OC
ADJD-PD ADV-MO $, KOUS-CP ADV-MO PIS-NK PROAV-MO VVINF-HD VMFIN-HD $.
Schade jedoch , daß kaum jemand daran teilhaben kann .

Figure 1

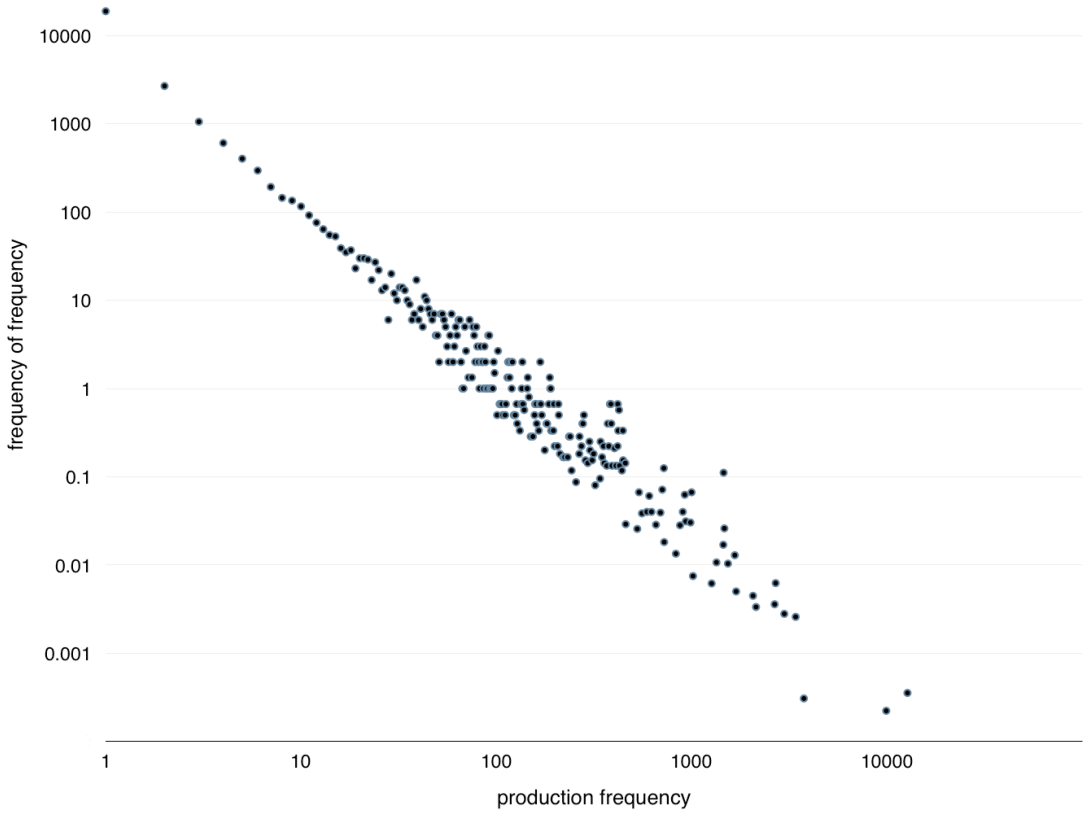


Figure 2

Sampson AGji fc