Does adults' speech complexity increase with age?

In 2001 I analysed data on a cross-section of British speech which seemed to show that adult speakers' utterances tend to become more grammatically sophisticated as they move through life from youth through middle age towards old age. This struck me as a socially-significant finding, so I published it. However, on the basis of a larger sample of much higher-quality data, I now believe that my interpretation in 2001 was entirely wrong. The following paper is under consideration by a journal, but since I seem to have done my best to mislead the linguistic community I thought I should place the new paper on the Web without delay, and with a prominent link from my home page.

In Sampson (2001) I presented evidence that utterances in everyday British speech increase in grammatical complexity (in the schoolroom sense of proliferation of subordinate clauses) as speakers proceed from youth through middle age towards old age. (I shall refer to this paper as "DCC".)

If correct, this finding would harmonize with various currents of thought about language development. Leonard Bloomfield (1935: 46) saw it as self-evident that "there is no hour or day when we can say that a person has finished learning to speak, but, rather, to the end of his life, the speaker keeps on doing the very things which make up infantile language-learning". Fred Karlsson (2009) has shown that the development of civilization in the West went hand in hand with increase in grammatical complexity of written language, and Guy Deutscher (2000) has documented the emergence of complement clauses, in an ancient language which previously lacked them, under the pressure of new communicative needs in society. Ngoni Chipere (2003, 2009) has shown that adult speakers in present-day Britain differ fairly widely in how far they have mastered the ability to operate successfully with clause complexity.

On the other hand there is also an influential trend in linguistic thought which runs counter to these ideas. Generative linguists assert the existence of a "critical period" in language acquisition (an idea introduced by Eric Lenneberg 1967), that is a point in any speaker's biography when he or she switches from being a language learner to being a mature language user, attaining "a 'steady state' ... not changing in significant respects from that point on"; and they believe that "To a very good first approximation, individuals are indistinguishable (apart from gross deficits and abnormalities) in their ability to acquire grammar". (Quotations from Chomsky 1976: 119, 144.) The generative concept of "Universal Grammar" (e.g. Pinker 1994) implies that a radical expansion of the range of logical structures expressible in a language, such as described by Deutscher for early Akkadian, should not be possible.

Not all the ideas belonging to one of these trends directly contradict all ideas belonging to the other. Chomsky's "steady state" concept of language acquisition is primarily about the acquisition of rules of grammar, some of which are formally recursive, whereas DCC was discussing how far speakers exploit this property of recursivity once a relevant rule has been acquired. But there is a clear harmony among the various strands of thought within either trend, and a clear disharmony between the two trends. One can think in terms of a contrast between "lifelong learning" and "steady state" pictures of language ability.

And this contrast is not of interest solely to linguistic theoreticians. As DCC pointed out (p. 58), grammatical complexity is associated with various types of logical precision, which are at least difficult, if not impossible, to achieve without it. Karlsson's and Deutscher's work suggests that the sophistication of 21st-century societies might have been unachievable if human languages had not developed and exploited mechanisms of clause subordination. If adult humans grow in their ability to handle grammatical complexity as they age, this would seem to be a socially-significant fact which deserves to be generally appreciated.

Unfortunately, the evidence put forward in DCC was not strong – my phrase, p. 70, was "not overwhelming". The present paper re-examines the claim, using evidence that was not available in 2001 and which is of much higher quality than what was available then.

DCC was based on grammatically-annotated extracts totalling about 74,000 words from the "demographically-sampled speech" section of the original British National Corpus (Burnard 1995), which contains transcriptions of about 4.2 million words of the casual everyday speech of a cross-section of UK residents in the early 1990s (I shall refer to this resource as BNC94ds). In its day, BNC94ds was widely seen as "the state of the art" in spoken-language corpora. But it had serious weaknesses, both in terms of the information it contained about speakers and in terms of the accuracy of the transcriptions.

With respect to the former, quite often demographic information about particular speakers was missing. Thus, 13.92% of words in BNC94ds were spoken by individuals for whom information about age is missing – and age is not the least satisfactory category in this respect: as much as 38.1% of wordage was by speakers whose socioeconomic status was unspecified. Perhaps worse, when information was provided it was sometimes clearly wrong. One female speaker's occupation was shown as "doctor", and her social class was given (in terms of a standard classification based on occupation) as "DE: unskilled or partly skilled". Unskilled and partly skilled people do not work as physicians. In this case, demographic information supplied in the corpus is self-refuting; it is not so easy to detect that an age specification must be wrong, but there is no reason to expect that category to involve a lower incidence of error. As for transcription quality: an academic contact who witnessed the transcription process at first hand described to me a scenario of low-level clerical workers transcribing sound recordings under time pressure, in a process which prioritized quantity over quality of output. The transcriptions often read that way, with frequent absurdities. A discussion of unsatisfactory child-minders includes the sequence *unless you've low and detest children*, which is meaningless, but *you've low and detest* is obviously an error for the common turn of phrase *you loathe and detest*. The exchange *Did you want to have a shower with daddy? — Umm yes* looks plausible, but the question is attributed to a three-year-old boy and the answer *yes* to his 34-year-old father. These are not isolated flaws.

Fortunately, BNC94ds has since been supplanted by a far superior research resource, the 11.4 million transcribed words in the speech section of the 2014 British National Corpus ("BNC14s"), on which see Love et al. (2017). As one concrete index of this superiority, instead of 13.92% and 38.10% of wordage spoken by people of unknown age and social class respectively, in BNC14s the corresponding figures are 0.74% and 3.39%. (All these figures are taken from the BNC14 *User Manual and Reference Guide*, version 1.1, p. 7.) But also in other ways less easily quantified, BNC14s is a great improvement on its predecessor. No-one with more than a passing acquaintance with the two resources would have any difficulty, I believe, in agreeing that the newer one has been compiled to an altogether higher standard of scholarly accuracy and system.

BNC14s consists of 1251 files each containing a transcription of a recorded conversation (together with the *User Manual*). To assemble data to test my 2001 claim, I discarded the few BNC files where some speaker's age was unknown, together with all files involving more than three speakers (p. 46 of the *User Manual* warns that attribution of individual utterances in those files to their respective speakers may be unreliable). This winnowing left 782 files, far more than I could use; not knowing whether there is any logic in the sequence of file names (each of which consists of four alphanumeric characters beginning with S, for Speech), for the sake of unbiased sampling I permuted them into a random order and worked through them in that order, extracting from each file a continuous sequence of about a thousand words beginning and ending at reasonably natural breaks. Knowing from experience that speech at the beginnings and (to a lesser extent) the endings of such recordings is often preoccupied with the recording process itself, I chose extracts ending roughly a thousand words from the end of the BNC14s files, where there was a good chance of encountering unselfconscious conversation about diverse topics of interest to the speakers rather than to the corpus compilers. These extracts I equipped with annotations identifying their clause structure; clauses were classified as finite, surrounded by square brackets, or non-finite, surrounded by round brackets, but not further subclassified.

As I progressed, I deviated from the file sequence in a number of ways. Speakers in their twenties were heavily over-represented, so after a while I rejected further files involving such speakers (though they are still over-represented in the eventual file-set used in the work reported below). Conversely, child speakers are woefully under-represented in BNC14s, so I ensured that I annotated at least one extract involving any under-eighteen speaker in the winnowed file-set; and I trawled through the files that had been rejected for having too many speakers, searching for children likely to be below a putative critical age, reasoning that even if transcribers confused the utterances of different adults, they would not be likely to confuse adults' with children's speech. (However, this last decision yielded only two further child speakers.) Paucity of child data is a weakness of BNC14s, but because the present research is about whether complexity increases in adult life, this weakness matters less for our purposes than it might for other kinds of research.[1]

I also rejected files with a high incidence of speech marked by transcribers as too unclear for the wording even to be guessed. (Where wording was marked as unclear but the transcribers offered a guess, I assumed their guess was correct.) There was a quandary here: avoiding transcriptions including any unclear material at all might bias the overall sampling of the speaker population (it could be that some categories of speaker produce more unclear wording than others), but imposing a grammatical analysis on sequences of unknown words is a worthless exercise. I compromised by rejecting files in which ten per cent or more of utterances contained passages marked as unclear.

I eventually annotated a total of 170 extracts containing a total of 174,303 words excluding "ums and ers" – well over twice the wordage used for DCC. The number of different speakers is 174, about thirty per cent more than in the DCC data. (This contrast between the two proportions comes about because speakers recur from file to file in the newer corpus but not in BNC94ds.) Since my annotated database was prepared in order to examine relationships between grammar and age, I refer to it as the "Grammage database". To be clear: a BNC14s file is an unannotated transcription of a conversation comprising thousands, sometimes tens of thousands, of words; a

1 The corpus compilers' policy about child data is unclear to me. At one point in text SRYY speaker p0189 asks *do they need like child data?* and p0192 replies *no actually you're not meant to record anyone that's less than eighteen*, going on to imply that when child speech is recorded that has happened accidentally because *what can I do about that?* [...] *the children are everywhere*. On the other hand, no fewer than 58 of the 782 winnowed files are conversations between the same three speakers, a 39-year-old female teacher and her children of nine and seven; thus there is quite a lot of child speech in BNC14s, though very few child speakers. I find nothing in the *User Manual* to explain this contradiction. Unfortunately, representativeness is not improved by repeated sampling of the speech of one or two members of an under-represented category.

Grammage file is a 1000+ word extract from such a file, with annotations identifying its structure of finite and non-finite clauses.[2]

To quote Jane Edwards (1992: 139), "The single most important property of any data base for purposes of computer-aided research is that *similar instances be encoded in predictably similar ways*." Any linguist has a rough idea about finite and non-finite clauses, but to make statistical analysis meaningful it was important to annotate according to an explicit prescriptive scheme that leaves as little as possible to analyst's judgement despite the endless variety of wording found in natural speech. As one example: when a participle is used with an adjectival or nominal function it is often very debatable whether such a use of that particular participle is sufficiently established for the word to count as an adjective or noun, or whether it must be analysed as a non-finite clause; rather than decide case by case I relied on the part of speech information in a specific dictionary (Roger Mitton's electronic version of the *Oxford Advanced Learner's Dictionary*, Mitton 1986). Thus, in S2LD.1290–1294 *... my mum was getting like so scared and so worried ...* , "adjective" is listed in the dictionary as a part of speech available for *worried*, but *scared* is listed only as a verb form, so the latter is analysed as a one-word past-participle clause here but *worried* is not treated as a subordinate clause. This may seem an odd contrast, but the point is that the analysis is predictable: the same word is treated in the same way wherever it occurs.

I annotated Grammage files according to the SUSANNE scheme (Sampson 1995), ignoring everything in that scheme except the rules for placing finite and non-finite clause boundaries. This scheme was used not only for the sake of comparability with the data used in DCC, but also because compared with some other grammatical annotation schemes it prioritizes tight definition and predictability. (Lin 2003: 321 remarks that "Compared with other possible alternatives ... [t]he SUSANNE corpus puts more emphasis on precision and consistency"; research is available (Sampson and Babarczy 2008) on how closely the scheme approaches perfect predictability.) However, in a few ways the SUSANNE scheme had to be adapted. The most important adaptation related to speech-editing phenomena, and needs to be explained in some detail to enable the reader to assess what the findings below mean.

The SUSANNE scheme contained elaborate rules for annotating the structure of speech edits, but these had been "road tested" only on material from BNC94ds. Those transcriptions contained only fairly few and simple speech-editing phenomena, whereas BNC14s has a considerably higher incidence of such phenomena which are often relatively complex. If we can assume that the speech of British people in general did not grow strikingly more incoherent over the twenty years between the corpora, the probable explanation is that the 1994 transcribers, consciously or unconsciously,

2 In due course I intend to place the Grammage files, including analysis software and documentation, on the internet; I have not yet done this.

tended to see their task as doing what the compilers of *Hansard* do: recording not so much what speakers actually say as what they seem to be trying to say. The 2014 transcribers by contrast appear to have done a good job of faithfully logging each false start, um and er, and trivial solecism produced by speakers. It was not practical to apply the subtle SUSANNE speech-edit annotation scheme to the resulting material. Instead, I simply surrounded all sequences that the speaker had effectively "edited out" with angle brackets. When computing depth statistics, all words within angle brackets will be ignored. This paper is not a study of speech errors; we are interested in the structure of wording that speakers both say and mean to say.

The clearest cases of speech editing are where the speaker embarks on a phrase or clause, then before completing it has a second thought and substitutes alternative wording. If the medium were print, the earlier word(s) would be deleted and invisible to the reader. Spoken words cannot be "deleted", but by replacing them the speaker appears to intend them to be ignored. In other cases, a construction is abandoned before it is complete, without replacement; again I angle-bracketed it as effectively withdrawn by the speaker. And very often, a speaker utters a word or phrase, hesitates, then decides that the word(s) are indeed what he or she wants to say, so repeats them and continues – in this case the first instance of the repeated wording is angle-bracketed out although it is not necessarily "incomplete" in any way.

Incompleteness may be straightforwardly grammatical, but there are also cases where, say, a postverbal object or complement is optional according to the rules of grammar, but in context it is clear that the speaker's motive in embarking on the clause was to specify that phrase, so if the clause terminates before the phrase is finished, that clause should be treated as abandoned. For instance, SHBY.89 begins:

maybe it's not compatible with my well anyway we have ...

Grammatically, a postmodifying phrase is optional after *compatible*, suggesting that *with my* should be angle-bracketed as an incomplete prepositional phrase, leaving what precedes as a complete clause; but for the speaker it is clear that the purpose of the clause was going to be identification of what "it" (a digital camera) was possibly incompatible with, so not only the *with* phrase but the entire clause beginning *maybe* is enclosed in angle-brackets.

Where a construction is abandoned before being completed, speaker and hearer may not necessarily think of it as withdrawn. Speaker 0058, a 23-year-old female graduate who appears to be working at Cambridge University, has a habit of utterances that end prematurely, as in:

SACQ.26 *... putting your hand up and arguing is a better*
SACQ.50 *... stopping someone from speaking isn't the best way of*

– leaving it to her hearer to complete the thoughts silently. But a predictable analytic scheme cannot ask the analyst to make judgement calls about which cases of objectively incomplete wording were intended to stand, so these examples too are angle-bracketed. (This, despite the fact that both quoted examples begin with a complete present-participle clause; both of these are functioning as subjects of *is*-clauses which as wholes are incomplete, so they are part of what is angle-bracketed out.) Even when the only "incompleteness" in a clause is that what seems to have been its last word has been marked as truncated, and it might have been obvious to the hearer and is obvious to the analyst what word was intended, for the sake of analytic predictability the clause is treated as incomplete.

There are of course some cases where a speaker's wording seems so thoroughly confused that even the possibility of marking parts of it to be ignored does not allow what remains to be seen as a well-formed structure. But this is not as frequent as some discussions of performance versus competence might lead one to expect. I have been surprised by how commonly wording that remains after angled-out material is deleted expresses a grammatically-complex but perfectly well-formed proposition.

The nature of the Grammage files is best explained by showing an example. Figure 1 is the beginning of file S5U8, a conversation among three people of different generations of a family, all British though the younger two live in Cork, and the conversation was recorded there. Speakers in BNC14s are identified by codes comprising "S" for Speaker followed by four digits; but it is confusing for files and speakers within files both to have code names beginning with S, so in Grammage files S for "Speaker" is replaced by p for "participant" or "person". Lines of the file, representing individual utterances, begin with filename and BNC utterance number, after which the utterer is identified by a single random capital letter – it is easier to follow the logic of a conversation between A, K, and B than p0475, p0416, and p0417; the first line of each Grammage file is a "cast list" showing the equivalences between these single-letter codes, unique within the file, and pXXXX codes, unique across the set of 170 files.

Figure 1:

```
A: p0475    K: p0416    B: p0417
S5U8.1135   B:    [ but I think [ NAME .. does want ( to slow down a
     bit ) ] .. but < I think .. > I'm not sure [ whether he would ]
     but he does .. want ( to ) ] .. [ I dunno [ whether he will ] ] **
     [ I don't know ]
S5U8.1136   A:    ** yeah .. yeah
S5U8.1138   A:    'mm
S5U8.1137-1139    B:    [ but he talks every now ] < and even though
     I'm not sure if 'erm .. > [ cos he keeps ( going 'oh [ with the
     kids it's only a few years and there's all the .. ( ferrying
     around ) [ I think ] ] ) ]
S5U8.1140   A:    'mm 'mm 'mm
```

Cumbersome XML structures in BNC files are replaced by user-friendly codes, for instance XML indications of anonymized names are shown as `NAME`; short and long pauses become `..` and `...` respectively; nonlinguistic vocalizations become `**`. BNC14s takes care to show where one speaker's utterance interrupts another's, splitting the interrupted wording into separate "utterances" even within a grammatical constituent; A's first *mm* in Figure 1 actually occurred between B's *even though* and *I'm not sure*, which are grammatically continuous with one another (although as it happens they are part of a grammatical structure that is never completed). For present purposes, grammatical structure matters but timing of utterances does not, so grammatically-continuous material is rejoined in the Grammage file, with its first and last BNC utterance numbers linked by hyphen. BNC14s transcription rules require transcribers to represent filled pauses – ums and ers – by one of a fixed range of alphabetic imitations including *mm*, *oh,* and *erm*, and in Grammage files these are prefixed by a prime to enable software counting word-depths to ignore them as non-words.

B's *I think .. I′m not sure whether* is a case of speech-editing where initial wording is replaced on second thoughts by different wording; *and even though I′m not sure if erm ..* is the beginning of a subordinate clause which is abandoned without replacement.

Annotating casual speech requires sensitivity to grammatical habits which sometimes differ from those of standard written language. In standard English, *because* is a subordinating conjunction introducing a cause or reason clause, but (as has been noticed by Stenström 1998), in speech *because*, or commonly *cos*, while it can be used that way, can alternatively introduce a main clause, meaning something like "incidentally" or "furthermore". The fact that NAME *keeps going* (i.e. saying) *oh with the kids it′s only a few years ...* is not the reason why he *talks*, so the clause introduced by *cos* is not parsed as subordinate to the earlier clause.

However, clearly, no analyst can have comprehensive knowledge of non-standard usages. The first few times I encountered usages like:

SWTX.1577 *yeah you just kind of you go from one to another to something else so*
SZBR.586 *oh I 'll ask NAME cos NAME 's .. NAME 's fourteen in January so*

I took *so* to be intended to introduce a new main clause which was abandoned after that first word, and I angle-bracketed it out accordingly. But I later found so many instances of *so* used this way in different files that I infer it may be a novel usage with which I happened to be unfamiliar. Having begun to annotate the usage as just described, though, I continued to do that with fresh instances: consistency of annotation is the first priority.

Co-ordination is an aspect of grammar where Grammage annotation practice deviates from the SUSANNE scheme. In that scheme, second and subsequent conjuncts are treated as structurally subordinate to a first conjunct (whether the conjuncts are clauses or other elements): co-ordinations have a structure like [*A* [*B*] [*and C*]]. This works well when it is clear which stretches of wording are intended as co-ordinations, and in writing this is normally clear: punctuation delimits sentences, and co-ordinations do not normally spill across sentence boundaries. But speech is not in general clearly divided into sentences, and utterances often begin with a co-ordinating conjunction, such as *but* at the beginning of S5U8.1135 in Figure 1; so for Grammage purposes it is necessary to adopt an annotation practice which avoids making the clause depth of words depend on unanswerable questions about sentence boundaries. The Grammage rule allows *A B and C* to be treated either as [*A*] [*B*] [*and C*] or as [*A B and C*] indifferently; either way the clause-depth of the words will be the same. (Another deviation from SUSANNE is that contractions such as *don 't*, *there 's* are treated as single words rather than split to reflect their derivation from *do not*, *there is*. This occasionally leads to odd results, as when at S2LC.665 *what do you think 's in that lorry?* the prepositional phrase *in that lorry* has to be annotated as a finite clause, since the verb of the clause has been reduced to a suffix to *think* in the main clause, and the subject WH-fronted to the beginning of the main clause. But the Grammage analysis, while odd, is predictable.)

Various other practices were adopted to ensure that Grammage annotations are maximally predictable and consistent, and these are discussed in the documentation file which will accompany the Grammage files online.

Complicated bracketings are difficult to check when displayed linearly as in Figure 1. The Grammage files were proof-read via a software system which uses indenting to display clause depth, thus the material of Figure 1 appeared as Figure 2.

Here, non-finite clauses are shown with curly rather than round brackets, since these make a better visual contrast on screen; plus-signs with left brackets mean that the bracket does not open a new clause, but resumes a clause opened on a higher line, after a subordinate clause has intervened. Angle-bracketed wording is omitted.

Figure 2:

```
S5U8.1135 B:
  [but I think
  [  [NAME .. does want
  [  [  {to slow down a bit
  [+.. but I'm not sure
  [  [whether he would
  [+but he does .. want
  [  {to
+..
  [I dunno
  [  [whether he will
+**
  [I don't know

S5U8.1136 A:
** yeah .. yeah

S5U8.1138 A:
'mm

S5U8.1137-1139 B:
  [but he talks every now
  [cos he keeps
  [  {going 'oh
  [  {  [with the kids it's only a few years and there's all the
..
  [  {  [  {ferrying around
  [  {  [  [I think

S5U8.1140 A:
'mm 'mm 'mm
```

As an example of the concept of mean clause depth, consider a short utterance containing a finite main clause which includes a subordinate present-participle clause:

no [*I don't like* (*wasting food*)]

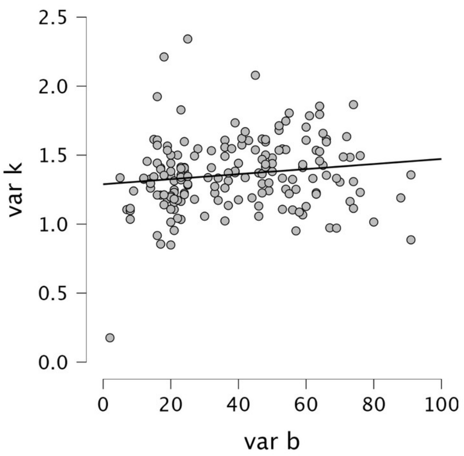
We can compute its mean depth by counting one word *no* at depth zero (outside any clause), three words *I don't like* at depth one, and two words *wasting food* at depth two. Then the mean depth would be (0 + 3 + 4 =) 7 ÷ 6 = 1.167. The most direct way to use Grammage data to try to reproduce the DCC findings will be to look for a correlation between speakers' ages and the mean depths, computed this way, of the sets of words spoken by the respective speakers, ignoring words between angle brackets and truncated words, which the speaker is deemed to have "edited out". Within the statistical tables generated in the Grammage study, mean depth measured in this simple way is labelled variable *k*; in due course we shall consider alternative depth measures.

(DCC pointed out, p. 61, that if a speaker produces very few words, his or her depth index must necessarily be low; I found that this effect disappears once speakers produce at least sixteen words. Because of the relatively large number of Grammage words which have to be discounted as speech edits, it seemed safer to double this low threshold, and I excluded from the new statistical analysis any speaker who produced fewer than 32 words. But this proved to apply only to one speaker, the 52-year-old female p0199, who uttered seventeen words in text SRYY; this speaker is ignored in computing the statistics discussed below.)

Figure 3, produced by the JASP statistics package, plots *k* values against age for the 173 remaining speakers.[3] The initial impression from Figure 3, unsurprisingly, is of great variation among speakers. However, the line of best fit through the data-points does show an underlying rising trend. Although the slope is not steep, and the coefficient of correlation r is only 0.144 (where zero means that variables are completely uncorrelated and one or minus-one means perfect correlation), the upward trend is statistically significant at the $p < 0.05$ level ($p = 0.029$), assuming a one-tailed test – which is arguably justified here, since it seems unreasonable to imagine that syntactic complexity might systematically *reduce* over a speaker's lifetime. (I suspect nevertheless that some theorists of statistics might insist that only a two-tailed test is appropriate in this context, in which case the correlation would not achieve statistical significance.)

3 JASP has been developed by statisticians at the University of Amsterdam and elsewhere as a free, user-friendly alternative to commercial packages such as SPSS.

Figure 3:



This finding agrees with the conclusion of DCC. It is the more striking, since someone seeing the scatter of data-points in Figure 3 without being shown the line of best fit might well take the overall trend to be negative (downward from left to right), in view of the data-points high in the upper left-hand area. (The two speakers with highest *k* scores are p0444 at *k* = 2.342, a 25-year-old female graduate working as a Pearson marketing executive, and p0252 at *k* = 2.212, an 18-year-old male sixth-former.) But this appearance is caused by the fact that outlying data-points – provided they are not too far from the main bulk of points – draw the eye and hence are mentally given undue weight. Finding the line which minimizes the sum of squared vertical distances between itself and the various data-points shows that, when all points are given equal weight, the trend is in fact positive.

However, Figure 1 contains data-points for speakers of all ages including children. (The point deep in the lower left-hand corner is for p0418, a two-year-old most of whose utterances are single words.) It is entirely uncontroversial that young children's speech tends to be grammatically simpler than that of mature speakers. What was remarkable in the DCC findings was that the expected increase in complexity between infancy and maturity seemed to continue rather steadily through middle age and towards old age.

If the concept of a critical period for language acquisition is right, then DCC (p. 70) notes that the literature on this concept implies that most children in modern Britain will have completed it by age 13. So I also plotted variable *k* against age omitting under-13s; there are 164 speakers of 13 years and above. To save space I do not show this plot, but the slope of the line of best fit is gentler; *r* reduces to 0.056. Even when two variables are uncorrelated in a population, there will almost always be some detectable slope, positive or negative, in a finite sample drawn from that population; the question is whether the correlation is strong enough to be unlikely to have occurred by chance. The answer in this case is no: even on a one-tailed test the p value is 0.238, far above the $p < 0.05$ level which is conventionally regarded as the borderline of statistical significance.

A line of best fit is straight by definition – it is produced by a mathematical manipulation guaranteed to yield a straight line. But there is no reason to assume that a relationship between grammatical complexity and increasing age must be linear, and the "critical period" idea suggests that it will not be. Because of the high inter-speaker variation, plots of data-points for individual speakers cannot tell us much about this. DCC addressed the issue by plotting mean values for successive age-ranges – it was particularly the fact that these points appeared to fall rather close to a straight line that gave the paper such impact as it achieved.[4] Figure 4 plots *k* values using the same age-ranges as DCC. (To be clear, the data-points of Figure 4 are not averages over speakers' means: they are averages over all words uttered by speakers within the respective age-ranges, so taciturn speakers have less weight than talkative speakers.)

Figure 4:

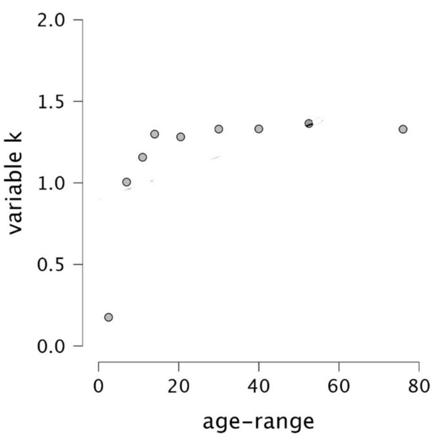


Figure 4 places successive data-points at distances along the *x*-axis corresponding to the midpoint of the respective age-range (e.g. the point for 16–24-year-olds is at 20.5 years). The nine ranges, with numbers of speakers within those ranges in brackets, are as follows:

4 Though this appearance was somewhat illusory, since distances (as opposed to sequence) along the *x*-axis of Figure 5.2 in DCC had no defined meaning.

under 5	(1)
5–8	(5)
9–12	(3)
13–15	(6)
16–24	(50)
25–34	(15)
35–44	(20)
45–59	(39)
60+	(34)

Figure 4 does not approximate a straight line. To my eye Figure 4 suggests a scenario closer to the critical period idea, with an inflexion point at the 13–15 years age-range, than to the lifelong learning concept. And the issue mentioned in footnote 4 does not apply here: distance along the *x*-axis is meaningful.

(One might object that the leftmost point in Figure 4 is not very robust, representing just 74 words from one two-year-old; but we know that if child speakers were more fully represented in our sample, the leftmost point in the plot could not be far above zero – infants do not begin by speaking in clauses.)

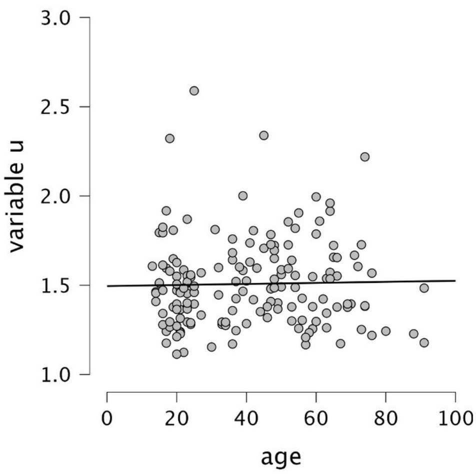
The way we have been measuring the grammatical depth of utterances is not the only way it might be measured. As DCC pointed out (p. 61), variation in mean word depths is necessarily damped by the fact that any clause at a depth $n > 1$ implies the existence of a clause at depth $n - 1$; the idea of an utterance comprised wholly of a clause at depth two would be nonsensical. This could help to explain why, even if there is some departure from perfect horizontality in the part of Figure 4 to the right of the inflexion point, any rise is very gentle. When variation in a variable of interest is at risk of being swamped by variation among irrelevant variables, one standard approach is to measure not the simple mean of its values, but its root mean square; and that could be done in the Grammage case.

Again, one might argue that depth zero should not be seen as the end-point of a scale on which depths one, two, and so forth are lower points. Words at depth zero, outside any clause structure, should perhaps be seen as doing communicative tasks which are different in kind from the task of building up propositional content: they are things like vocatives, *yeah* or *no*, exclamations, turn-retaining words such as *well …*, and so forth. Variation in clause depth might be brought into sharper focus by ignoring depth-zero words in utterances and averaging over only words within (completed) clause structures. There is a complication, in that words with the same non-logical functions can and often do occur also in the middle of clauses, so if depth-zero words are ignored in depth counts, ideally clause-medial words with similar functions ought also to be ignored. That is not possible with the Grammage files, since nothing in the annotation identifies the latter word-tokens. But it is certainly possible to count depth in Grammage files ignoring depth-zero words.

And one might feel that non-finite clauses ought not to count for as much as finite clauses in computing mean clause depth. A finite subordinate clause contains all the kinds of logical structuring found in main clauses, but most non-finite clauses are inherently simpler – for instance, most contain no separate subject. A sequence such as [*he's gonna* (*do it*)] is annotated as containing a non-finite subordinate clause because the sequence is a contraction of *he is going to do it*, where *to do it* is explicitly an infinitival clause, but *gonna* feels little different from an auxiliary verb such as *will*; *he will do it*, or *he'll do it*, are each regarded as single clauses. Karlsson (2009: 201) notes that finite subordinate clauses typically represent a more sophisticated, later stage of language development than non-finite clauses. Differences among mean clause depths might be more striking if non-finite clauses were lower-weighted, or even zero-weighted, relative to finite clauses.

Variable *u* in the Grammage statistical database combines all three of these techniques. It stands for the root mean square of depths of words within completed clauses, taking only finite clauses as contributing to depth counts. Figure 5 plots *u* against age for the 164 adult speakers, in the hope that this might reveal a positive trend clearer than seen in the 13-years-and-over part of Figure 4.

Figure 5:



To the present author's surprise, it does the reverse. The correlation coefficient is much closer to zero, at $r = 0.021$, than when *k* is plotted against age (with or without under-13s); in Figure 5, the departure of the line of best fit from the horizontal is scarcely visible. The *p* value, two-tailed, is 0.785: it is much more likely than not that a sample of this size, with this degree of individual variation, would depart at least this far from the horizontal in one direction or the other, on the null hypothesis that in the population the two variables are entirely uncorrelated.

I conclude that the finding of DCC, tested against a larger body of higher-quality evidence, does not stand up. There is no evidence that older adults' speech tends to be grammatically more complex than that of younger adults. And there is even some positive evidence, in Figure 4, favouring the steady-state model of grammar development.

One might wonder, then, why I got a contrary result in 2001. The explanation I find most plausible has to do with the issue discussed above about less accurate recording of speech edits. Psychologists well know that perception is heavily influenced by what subjects expect to be the case. If the transcribers of the 1994 corpus operated in a relatively *Hansard*-like fashion, then expectations will have had more chance to override the physical facts of sound signals in determining what was recorded on paper. Even reading transcribed conversation without hearing the speech, approximate speaker ages are commonly rather obvious from topic and style of wording; that must be even more true for anyone hearing the speech. If transcribers unconsciously assume that maturer adults will tend to speak in logically subtler ways, as is not unlikely, this might well introduce enough bias into transcribing practice to account for the gentle rise through middle age seen in Figure 5.2 of DCC.

The fact that this rise disappears when tested against the newer data demonstrates the value of the meticulous recording of low-level phenomena which is manifest in the 2014 BNC speech corpus.

REFERENCES

Bloomfield, L. 1935. *Language*. Allen & Unwin.

Burnard, L. 1995. *User Reference Guide for the British National Corpus*, version 1.0. Oxford University Computing Services.

Chipere, N. 2003. *Understanding Complex Sentences: native speaker variation in syntactic competence*. Palgrave Macmillan (Basingstoke).

Chipere, N. 2009. "Individual differences in processing complex grammatical structures". In Sampson, Gil, and Trudgill (2009), pp. 178–91.

Chomsky, A.N. 1976. *Reflections on Language*. Temple Smith (London).

Deutscher, G. 2000. *Syntactic Change in Akkadian: the evolution of sentential complementation*. Oxford University Press.

Edwards, Jane. 1992. "Design principles in the transcription of spoken discourse". In J. Svartvik, ed., *Directions in Corpus Linguistics: proceedings of Nobel Symposium 82*, pp. 129–44. Mouton de Gruyter (Berlin).

Karlsson, F. 2009. "Origin and maintenance of clausal embedding complexity". In Sampson, Gil, and Trudgill (2009), pp. 192–202.

Lenneberg, E.H. 1967. *Biological Foundations of Language*. Wiley (New York).

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

Love, R., et al. 2017. "The Spoken BNC2014: designing and building a spoken corpus of everyday conversations". *International Journal of Corpus Linguistics* 22.319–44.

Mitton, R. 1986. "A partial dictionary of English in computer-usable form". *Literary and Linguistic Computing* 1.214–5.

Pinker, S. 1994. *The Language Instinct: the new science of language and mind*. Penguin.

Sampson, G.R. 1995. *English for the Computer: the SUSANNE corpus and analytic scheme*. Clarendon (Oxford).

Sampson, G.R. 2001. "Demographic correlates of complexity in British speech". In Sampson, *Empirical Linguistics*, pp. 57–73. Continuum.

Sampson, G.R. and Anna Babarczy. 2008. "Definitional and human constraints on structural annotation of English". *Natural Language Engineering* 14.471–94.

Sampson, G.R., D. Gil, and P. Trudgill, eds. 2009. *Language Complexity as an Evolving Variable*. Oxford University Press.

Stenström, Anna-Brita. 1998. "From sentence to discourse: *cos* (*because*) in teenage talk". In A.H. Jucker and Y. Ziv, eds, *Discourse Markers: descriptions and theory*, pp. 127–46. John Benjamins (Amsterdam).