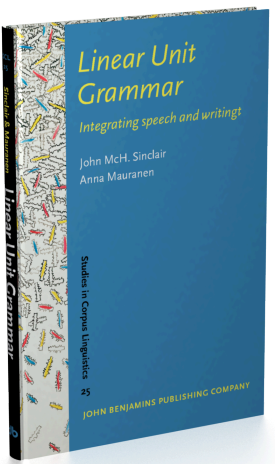
Studies in Corpus Linguistics
25

Preamble

https://doi.org/10.1075/scl.25.03pre



Pages xi–xiii of
Linear Unit Grammar: Integrating speech and writing
John McH. Sinclair and Anna Mauranen
[**Studies in Corpus Linguistics**, 25] 2006. xxii, 185 pp.

Preamble

You are encouraged to work through the following introductory activity as a preliminary to the presentation of Linear Unit Grammar. Please look at Figure 1.

theheadmasterofharrowtellsannmcferranwhyhehasl
etthetvcamerasintoaschoolfullofodditiesbarnabylen
on30thheadmasterofharrowschoolleansoverhisdesk
therearemoreimportantthingsinlifethanstrawboaters

Figure 1.

We expect that before you are consciously aware of having thought about what Figure 1 is, you will have formulated a number of hypotheses about it and will have decided on a number of tactics for dealing with it, and will be first of all aware of the preliminary results of all this involuntary activity. It is a written verbal communication. There are traces of English words in it. It is a string of characters without spaces or punctuation or even capital letters.

Perhaps almost immediately, given your linguistic training, competence in English and expectations of this book, you will guess that it is a piece of written English rendered solely as a sequence of characters. You are then able to formulate a stratagem for dealing with it, and you settle on the "traces of English words" that was one of your immediate reactions. You decide that you will attempt to turn it into a string of English words, and then see if it makes sense. Perhaps some investigation that you are conducting subliminally reports at this point that there are signs of coherent phrases here and there, strengthening your hypothesis.

Not everyone will respond in the same way to Figure 1, and we would be interested to hear from anyone who has a strikingly different experience. The two points that we hope most readers will agree on are (a) that most of your reaction is involuntary, and the hypotheses and strategies are formulated with little or no conscious intervention; (b) you begin, mentally, to add word spaces, capital letters and punctuation in order to make sense of the passage. There is no need to copy it out and add all these features, because once you have the idea it will become almost immediately readable. Occasionally you may have to backtrack or read a string of characters several times before the word spaces become obvious, because there are some specific difficulties in this passage — one of the reasons why it was chosen. The proper names are slightly unusual, there are some words and phrases that may not

be familiar to many readers — like *straw boaters* at the end. You may not be aware of Harrow school, an old English school, where hats called straw boaters are, or were, worn in the summer. On the helpful side there are repetitions like *the headmaster of harrow* and a cliché, *there are more important things in life than*, which, once noticed, explains almost a fifth of the whole passage.

The passage begins easily, because the most frequent word in English, *the*, heads the text. *Headmaster* is recognisable; in English it can be written in three different ways, *head master, headmaster* and *head-master*, and with or without initial capitals. In the original of this passage, curiously, the two instances are spelt differently, as a single word and as two separate words. *Of* is the second commonest word in English, fairly easy to spot; *harrow* may not be readily recognised, but *tells* is clear enough, so *harrow* must be something you can be headmaster of. If you saw the word *arrow* there you would have to revise your guess or have an *h* left over. The same goes for the *s* following *tell.* Perhaps at first its role as the present singular inflection is not obvious, and *san* could be the start of a name; the run of consonants *nnmcf* is a little off-putting, and indeed the whole string until *why* is encountered may be a puzzle. The clue is that many Scottish family names begin with *mc* or *mac*, signifying "son of". *McFerran* or *Mcferran* are plausible names, though unfamiliar to the authors; the original text reads *mcferran* which is certainly short of at least an initial capital. The occasional inconsistencies in the original text show that users of a language encounter routinely the same kinds of problems that the reader of Figure 1 is faced with, though usually diluted with plenty of unproblematic text. Once Ann Mcferran has been identified, the rest falls into place, and the first line is all but deciphered, because *he* and *has* are easy to spot. The line break splits *let*, but with *the* following it can be reconstructed quickly. *TV* is almost universal, and supported by *cameras*. By now the decipherer can polish off *into a school full of* without much perplexity, but the last few characters in the line may need a moment; *oddities* is not a common word, and *Barnaby* is not a common name, nor is *Lenon* unless you award it a double *n* in the middle and think of the Beatles.

The beginning of line 3 is distracting unless Mr Lenon has been identified, but 30^{th} need not detain us, and the next phrase has occurred before, here with the helpful word *school* following for those not so familiar with UK institutions. By now *leans over his desk* should be easy because, unconsciously, we have become trained in this medium even over such a short passage, and headmasters have desks anyway. The last line is a quote from Mr Lenon, which we have already construed. (This extract is the beginning of a piece in The Times of 10th June 2001).

The task of separating words in a piece of ordinary printed matter is an unfamiliar one for most readers, but one that we adapt to readily, presumably working out ad hoc tactics as we go along. The keys to efficient performance include:

(a) the ability to apply a hierarchical model to the linear string — in this case to postulate, correctly, that the passage consists of a string of word tokens, and that a placement of word boundaries will make the passage instantly legible and understandable.
(b) the ability to *prospect*, to look ahead for features that will help the interpretation of a difficult passage or settle a question of alternatives
(c) the ability to hold provisional interpretations in mind and to abandon them if they are superseded by more plausible ones — otherwise to continue with them perhaps without resolution — like how exactly to spell *mcferran*.

In the rest of this book we will be applying essentially the same techniques to a later stage of the process of "making sense of" text. We will not artificially hamper ourselves as we have done in this illustrative exercise, but will use the same arguments, rely on similar perceptions, knowledge and abilities in the reader, and chart the progress of text from its fairly raw state in a range of situations to something that makes reasonably good sense.