Temporality in speech – Linear Unit Grammar

Anna Mauranen
University of Helsinki

Language is usually modelled through a predominantly synoptic perspective; even if the object of analysis is spoken language, we tend to look at extracts where the analysis of parts makes use of the whole. Holistic analyses can be very good for capturing realities of language in many respects, but in the case of modelling temporal aspects of processing they fall into the trap of unrealistic hindsight. The experience of speech is time-bound: a hearer will go on what he or she has heard at any given point, and will anticipate what may follow. The predictions will be either confirmed or rejected in rapid succession, as speech moves on. The time window for working memory is very brief, and processing focuses on continuously changing input. Models of this process must take into account this dynamism, and they need to take on board the fact that language must be continually processed even while utterances are still incomplete. Most models of language structure are based on completed units; this tends to lead to a hierarchical view of language, embodied in most grammars. The reality of temporally progressing speech is nevertheless fundamentally linear along the dimension of time.

Keywords: chunking, dynamic grammar, language processing, language in working memory, temporality

1. Introduction

This paper seeks to model spoken language in a linear manner. To put it in another way, the modelling is based on the temporal order of spoken language as it unfolds, and the linear order of reading written text. The analysis is based on short processable chunks, and the kinds of elements that make up these chunks. To capture dynamism, the analysis proceeds in two steps. The initial step assigns chunk boundaries, attending only to junctures at which a natural boundary of a processable chunk might seem to lie. It is only after this stage that the nature of the chunks is analysed by paying attention to the internal properties of each chunk. The resulting descriptive model is fundamentally linear (Sinclair & Mauranen 2006). As the different kinds of elements are analytically processed further, they can be seen to
form units of meaning, which would seem to be largely compatible with more traditional models of grammar. Linear Unit Grammar, which is based on the linear, sequential chunking of language as we experience it (as explained in more detail in the next section) therefore also serves to bridge the gap between language as we encounter it and grammars as we know them.

2. Linear processing of speech

Speech is fleeting. As we hear utterances unfold, our memory rapidly loses its hold on preceding material. As we speak, we similarly lose track of what we just said. This can be seen in terms of what Christiansen & Chater (2015) call the ‘Now-or-Never bottleneck’ of information processing: “if linguistic information is not processed rapidly, it is lost for good” (Christiansen & Chater 2015: 3).

A listener’s immediate experience of an extract of spontaneous speaking may thus look something like the following:

(1a) they|cannot|do|do|anything|because|there|are|no|union|of|er|people|working|for|private|enterprises|mhm|hm|mhm|hm|mhm|yeah|its|kind|of|strange|like|we|have|you|can|think|about|this|maybe|we|will|have|the|ten|seven|years|but|now|yeah|there|is|no|anything|like|that

This extract from an academic seminar (ELFA 2008; see below) may not appear highly accessible to a reader used to more sophisticated conventions of orthography with its word boundaries, punctuation, and other conventional markings we tend to insert into written renditions of speaking. Yet if we look at it for a while, things begin to fall into shape, and it becomes possible to discern individual words. Adding word boundaries to the transcript helps our interpretation, even though this is not an accurate reflection of a hearer’s experience:

(1b) they|cannot|do|do|anything|because|there|are|no|union|of|er|people|working|for|private|enterprises|mhm|hm|mhm|hm|mhm|yeah|its|kind|of|strange|like|we|have|you|can|think|about|this|maybe|we|will|have|the|ten|seven|years|but|now|yeah|there|is|no|anything|like|that

A reader encountering this may experience the word boundaries helpful, and speakers of course use many resources for signalling them. In addition to pauses, stress, intonation, glottal stops, vowel and consonant patterns, along with other devices depending on the language serve to punctuate the sound stream for us better than a simple sequence of sounds. Aware of the boundaries, a reader might also surmise there are more than one speaker involved, which a listener in the situation would of course already know.
We tend to help a reader interested in spoken language analysis by transcribing speech by a variety of conventionalised systems, with considerable differences with regard to their inclusion of detail. In broad transcription the text could look like this:

(1c) 〈/S1〉 they cannot do do anything because there are no union of er people working for private enterprises 〈/S4〉 〈S1〉 [mhmm-hm] 〈/S1〉 〈S6〉 [mhmm mhmm yeah] 〈/S6〉 〈S3〉 [it’s kind of] strange 〈/S3〉 〈S6〉 [like] 〈/S6〉 〈S3〉 [we] we have 〈/S3〉 〈S1〉 you can think about this [er in your essay] 〈/S1〉 〈S4〉 [maybe we will] have the ten seven years [but now] 〈S1〉 [yeah] 〈/S1〉 there is no [anything like that] 〈/S4〉

While the extract in the shape of this last rendition may be the most familiar, and perhaps at first glance appear the most approachable to the analyst, it is actually a poorer reflection of what is likely to be a hearer’s experience than the previous two. A hearer experiencing the discourse unfold, assuming for a moment one not actively involved in the dialogue, does not experience such large wholes all at once. There are two major features that make representations of the above kind (1c) unrealistic: the forward view and the backward view. Both go beyond the limits of the working memory. The forward view is obviously impossible in its literal sense, as there is no way a hearer could get a peek at what they will hear in a few moments. Clearly, the way we cope with this is by prospecting ahead on the basis of what has already passed in the discourse, which means that a good proportion of processing what we hear consists in confirming or rejecting our prospected hypotheses. This in itself takes space in the working memory, thus both facilitating and taxing it at the same time.

The backward glance is heavily constrained by working memory capacity. The limits of the working memory would not get a listener through the first turn in the above extract. Of course, the maximum number of elements included in the seven or so units that working memory can hold at any point (Miller 1956) depends on what processable units are like. Thus, if meaningless strings of, say, unrelated sounds or letters are presented, the working memory capacity fills up very quickly, but if we can rely on words or other highly conventionalised units of meaning, we can expand the length of the unit held in memory somewhat. Nevertheless, in continuous speech new elements need to be continually processed which push the earlier ones out of the way. It is therefore not realistic to assume a listener has access to long stretches of the preceding conversation. We need to process all elements really fast, as suggested by the Now-or-Never Bottleneck model. The
usual linguistic models do not incorporate this aspect of language processing sufficiently. While for example Conversation Analysis and Interactional Linguistics emphasise the temporal progression of speech events (see, e.g., Auer 2005), they nevertheless perform analyses by looking at whole, several-minute extracts at once.

The shortcomings of established descriptions indicate that we need better ways to depict the experience of continuous speech, which can feed into models of grammar and bridge the gap between experienced language and analysed language. Above all, we need models that match what we know about human information processing, its efficiency and speed in coping with fleeting and complex input, and we need to come to terms with approaching grammar from this perspective.

A seminal work in efforts to account for the on-line nature of spoken language is Brazil’s (1995) *Grammar of Speech*, which sets out from the premise that meanings are made in language as utterances unfold, not after they are finished. A key realisation in Brazil’s analysis is the incremental nature of unfolding speech, and consequently the notion of the increment. Brazil analysed a longish spoken narrative essentially in terms of ‘telling increments’, by which a speaker tells the story. Despite their centrality, telling increments do not exhaust Brazil’s analytical instruments. Among the others are units similar to what Biber et al. (1999: 225) call “non-clausal material in conversation”, consisting of text signalling elements such as discourse markers.

The model presented in this paper, Sinclair & Mauranen’s (2006) *Linear Unit Grammar*, shares the point of departure with Brazil. It develops Brazil’s central premises like constructing meaning in unfolding language and the incremental nature of utterances, but goes on to widen the scope to include conversation and written language, and thereby makes more general claims. Linear Unit Grammar (LUG) is essentially a theoretical model for conceptualising grammar. At the same time, it reaches out to the experiential world to help understand how language works. Above all, it bridges the gap between the way language is experienced and the way it is depicted in grammars. Rather than presenting yet another grammatical description of English, it takes us from the stream of language as it flows into our experience and shows how it connects to the complexities of grammars that traditionally have been based on static, frozen language.

Linearity is probably more obvious in speech than in writing, because the fleeting nature of on-going speech and the difficulty in recalling verbatim what was said a few minutes earlier in a conversation are part of our everyday experience. Written text is nevertheless also organised in a linear fashion on the page or screen. Even though our reading process includes digressions, achieving meaning requires some minimum amount of linear reading at a time; the predominant mode of reading for meaning takes place by following the order of the text. Linearity is
therefore at the heart of language. Conventional grammars, by contrast, are hierarchical. Thus, if language is seen as horizontal, descriptive grammars stand in a vertical relation to it. Grammars incorporate meaning in a paradigmatic way, as a set of mutually exclusive choices; meaning, in this line of thinking, is understood to arise from the relation of the choice that was made, in relation to those that could have been made but were not. In LUG, meaning arises in the co-selection of constituents. This highlights syntagmatic relations over paradigmatic, and requires only minimal hierarchy.

Linear and temporal models of grammar, also sometimes called dynamic models, show awareness of the problem of a synoptic view of the sentence, which progresses from the abstract to the concrete in a top-down, hierarchical manner towards the actual words uttered. Such sentence models depend entirely on the completed artefact and cannot be applied while it is still in progress. Dynamic models of syntax, notably Dynamic Syntax by Kempson et al. (2001) and Syntactic Carpentry by W. O’Grady (2005, 2008), develop grammatical descriptions that proceed from left to right. However, they utilise far more hierarchy than either LUG or Brazil’s work, build their models on sentences, and use constructed, not naturally-occurring language (or “used language”, as Brazil (1995) calls it). A development of Brazil’s work that also addresses temporality and is in this sense parallel to LUG, is G. O’Grady (2010), but like Brazil, G. O’Grady does not really address spontaneous speech in the way that LUG does. On the other hand, it focuses on intonation, thus capturing something that is yet to be addressed by LUG research. All the data LUG draws on, both speech and writing, is naturally-occurring – even if the speech is represented in transcribed form. This was motivated largely by the unavailability of recordings of the oldest extracts used in the development. These were included because they had advantages like not showing transcriptional finesses like upper and lower case spelling or speaker changes, which may seem alienating to a present-day reader but which at the same time revealed how much we can perceive in seemingly primitive transcripts.

LUG does not directly threaten more traditional grammars, because it operates in the space between text (spoken or written) and grammar. The final output of its analysis is a synthesis that can, in principle, feed into hierarchical grammars; they can furnish more detail in a number of ways, according to the model chosen. The first stages of LUG, by contrast, work on data that is normally inaccessible to other grammars, notably spontaneous speech; yet it is robust enough to embrace written text as well.

LUG works in stages, and the step-by-step manner of analysis is a vital feature. In this it departs radically from virtually all other grammars, which perform their analysis in a single pass. The first step, chunking, is intuitive, which is of course highly untraditional and has, not surprisingly, therefore raised criticism
Anna Mauranen (e.g. Mason 2008). The pros and cons are discussed in the next section. The second phase of LUG analysis is based on a set of rigorous principles determined by the model. In the following, I shall present the LUG model, starting from the chunking phase in the next section, and then moving on to the principal analytical categories. As in Brazil (1995) and Sinclair & Mauranen (2006), I use a transcribed extract of naturally-occurring speech to show how the analysis is done and what the main categories are. This means that audio materials are not made use of, which can be seen as a problem in defending the analytical solutions. However, the purpose of this paper is essentially to illustrate and discuss the principles of Linear Unit Grammar and its prospects rather than to present empirical work on a new kind of data. There are also preliminary supportive results from an earlier study by Cheng et al. (2008), which carried out an independent analysis of one of the extracts used in LUG, in terms of tone units. Another source of support comes from an unpublished pilot study done in Helsinki based on extracts from the ELFA (English as a Lingua Franca in Academic Settings) corpus. The present data is drawn from the ELFA corpus, which consists of a million words of naturally-occurring academic speech from university contexts (seminars, lectures, presentations) and conferences (presentations, discussion sections) recorded and transcribed in four Finnish universities. All situations use English as a lingua franca, but English is not the object of study in any of them. For a more detailed description, see the ELFA project webpage (www.helsinki.fi/elfa).

3. Chunking

The point of departure for chunking is the notion of ‘chunk’. It is a pre-theoretical term, and defining it precisely is not even attempted in LUG or at present. The chunking process can nevertheless be described and tested, for instance by asking speaker groups to perform chunking on a text sample (Sinclair & Mauranen 2006), by an independent analysis of tone units (Cheng et al. 2008), or by analysing rephrasing, repeating and dysfluency behaviour (Mauranen 2012a). Moreover, Smart (2013) performed an inter-rater reliability test of assigning boundaries on a modest scale, which nevertheless produced very high correlations, as did a more ambitious effort by Carey (2013). All these kinds of evidence have yielded high levels of agreement with LUG analysis. Of course, inter-rater reliability testing on a small scale is not very impressive, but in combination the independent sources yield stronger support for the feasibility of the model than just a single experiment would do, even if with more participants.

The criticism that Mason (2008) levels at chunking is that it should rest on a more ‘objective’ basis, and he therefore proposes using a multi-word unit algorithm
to derive chunks. This is not what LUG is about. A number of scholars have used the term ‘chunk’ in connection with other terms that comprise phraseological or multi-word units of many kinds: fixed expressions, formulaic sequences, multi-word units, constructions, etc. These can all be seen as products of the chunking process. What we are concerned with in LUG is not the product in the first instance, but the process. The first outcome is the chunk boundary, as will be seen below. Even though in some cases it is indeed possible to identify certain chunks as performing similar functions in different texts (see, e.g., Mauranen 2009/2012; Carey 2011, 2013, 2015), the point about intuitive determination is not a matter of language-internal regularities. Chunking, as part of LUG, is a matter of how people deal with the fleeting nature of the language experience; the process is very similar to what happens in human information processing in general.

Humans chunk incoming information as part of normal perceptual processes. In this, linguistic information is no different from any other. How large the chunks are and what exactly contributes to their boundaries is less clear. It looks like linguistic boundary marking is multi-faceted, with a variety of factors – grammatical, phonological, semantic or cognitive – coming together to bring about a sense of the ‘natural’ chunk boundary that we seem intuitively to respond to. The usual elements of redundancy that natural languages make use of are likely to operate in chunk boundary markings so that more than one element converge on the same boundary, and there is evidence that this is what is taking place (e.g. Monschau et al. 2004). At the same time, the ‘staggering’ that is also found between boundary markings generated by different systems (Mukherjee 2001) need to be borne in mind. Plausible linguistic sources are phonological, tonal and intonational patterning, pausing, syntactic boundaries, phrases, breathing patterns, punctuation marks, etc. Undoubtedly paralinguistic and non-linguistic situational sources play a role, too. A likely cognitive source is the limited capacity of working memory. We can only hold very few items in working memory at any time, in Miller’s (1956) classic model about seven units, plus or minus two. This comes close to the five to seven words that seem the upper limit before a chunk boundary.

4. Provisional Unit Boundaries (PUBs)

The first step of chunking analysis involves assigning boundaries – ‘provisional unit boundaries’ (PUBs). It is important to focus attention on only boundaries in the first instance, and only after they are in place to look inside them. Here is an excerpt, a discussion carried out as part of a graduate seminar, that will be used throughout this paper to illustrate the analysis (supplemented by very few other
examples from the same corpus for illustrating particular points). At this stage, speaker changes are not indicated:

(2) but er actually i know erm it's common as a saying in Poland that you er men should have a son er build a house and plant a tree yeah yeah we got the same thing to be a real man yeah do you have this sort of yes expressions i've never heard anything like that (excerpt from ELFA)

Assigning PUBs consists of inserting boundaries where they seem ‘natural’; to capitalise on intuition, they should be put in quickly, spontaneously, without dwelling too long on deciding. At points of uncertainty, boundaries are better inserted than shelved.

It is vital to follow linear progression: boundary assignment is to be done in view of the discourse up to that point, without looking ahead. The end of a chunk becomes evident at the latest when something comes up that is incompatible with its continuation, which is immediately clear as the new item begins.

The suggested PUBs for the extract are here. While some boundaries may be contested by some readers, usually surprisingly few are controversial.

(3) but/ er/ actually/ i know/ erm/ it's common as a saying/ in Poland/ that/ you/ er/ men should have a son/ er/ build a house/ and/ plant a tree/ yeah/ yeah/ we got the same thing/ to be a real man/ yeah/ do you have this sort of/ yes/ expressions/ i've never heard anything like that

Once the boundaries are in place, we can look inside them. The chunks serve as input to the next stages of the analysis.

5. Types of chunks

When we proceed to analyse the elements enclosed by the chunk boundaries, a set of principles is applied in a cyclical process. Each cycle works on the output of the previous one, and applies a limited set of distinctions, mostly a dichotomy. The first cycle consists of a basic distinction between two principal types: elements concerned with what is being talked about (men should have a son; I've never heard anything like that), and those concerned with managing the discourse (actually; yeah; but). The distinction is rough and hard to pin down in precise terms, but recognisable, and indeed somewhat similar distinctions are recognised at least implicitly in models of discourse (e.g. Chafe 1994) or grammars that take discourse on board (e.g. Biber et al. 1999; Carter & McCarthy 2006).

Elements of the first type deal with the kinds of matter that is variously known under such names as ‘topical matter’, ‘content’ or ‘message’. In LUG
terms, they increment shared experience among interlocutors. They take forward the ‘contents’ that individuals contribute to the conversation. We call these *message-oriented elements* (‘M’) for lack of a better term. Elements of this kind are not always complete, although in the basic case they are. They may look fragmentary at times, but their vital characteristic is that they take the ‘topic’ forward. These are also the elements that seem to be best retained from conversations; they are the building blocks for people’s spontaneous informal summaries of what was said.

The second main element type focuses on social interaction and its management: *organisation-oriented element* (‘O’). These help participants deal with real-time aspects of conversation, most fundamentally to maintain cooperation: managing turn-taking, changing topic, shifting footing, manoeuvring situations. The fundamental distinction between the O and the M elements lies in their orientation towards either *action* (O) or *message* (M). At this stage, we can see a sequential alternation of O and M elements, as in any binary code. Each element is either an M or an O. The example text appears as a structured pattern of units of discourse organisation and units of topic incrementation (Table 1; speaker changes reinstalled).

The assignment of M and O follows the principle of linearity in that it is done in the light of the discourse up to that point. In other words, each element value is assigned at the point of discourse that has been reached in a given chunk, without going beyond it. Since chunk boundaries are in place at this stage of the analysis, it is easier to maintain linearity than in the case of assigning chunk boundaries. The stepwise method helps keep up linearity: it is easy to appreciate the advantage of focusing on simply the boundary marking in the first phase, and it even facilitates the analysis of the elements. In this way, the analysis also stays closer to the experience of online speech, which inexorably passes on in time, compared to a ‘synoptic’ approach, which draws on whole extracts at a time and moves comfortably back and forth.

5.1 Types of ‘O’ elements

Once the elements within the chunk boundaries have been identified as being of one of the principal types, each type can be analysed in greater detail. This can be done in any order, so let us start with the O type, where fewer subtypes have been identified so far. O elements are of two kinds: those concerned with interaction (*er; actually; yeah*) and those concerned with text (*but; that; and*). The two types are not always easy to keep apart, as these elements can be expressed with very similar wordings, and because O elements may play interactively tinged roles while organising the text. The general principle is nevertheless as clear as such
distinctions tend to be in natural language: we have the possibility of using elements of language in unusual and untypical ways at any point.

The first kind orient to managing the shifting situation – framing and focusing moves, controlling timing, holding and yielding the floor, etc. Elements that specialise in tasks like this are called ‘OI’, where ‘I’ stands for ‘interactive’. Thus, they are elements involved in organising the on-going interaction.

The other type is concerned with organising the discourse as text. These O elements bring the M elements into a network of relationships with each other. Their role is quite aptly captured in Firth’s (1968) terms as replacing sequence with order. What this order does, going beyond sequence, is help establish textual relations such as exemplification or contrast, and also to enable handling larger and smaller structures in text: set up anticipation, position sequences in their wider

---

Table 1. Types of chunks: M and O

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A:</td>
<td>but O</td>
</tr>
<tr>
<td>2.</td>
<td></td>
<td>er O</td>
</tr>
<tr>
<td>3.</td>
<td>actually</td>
<td>O</td>
</tr>
<tr>
<td>4.</td>
<td>i know</td>
<td>M</td>
</tr>
<tr>
<td>5.</td>
<td>erm</td>
<td>O</td>
</tr>
<tr>
<td>6.</td>
<td>it’s common as a saying</td>
<td>M</td>
</tr>
<tr>
<td>7.</td>
<td>in Poland</td>
<td>M</td>
</tr>
<tr>
<td>8.</td>
<td>that</td>
<td>O</td>
</tr>
<tr>
<td>9.</td>
<td>you</td>
<td>M</td>
</tr>
<tr>
<td>10.</td>
<td>er</td>
<td>O</td>
</tr>
<tr>
<td>11.</td>
<td>men should have a son</td>
<td>M</td>
</tr>
<tr>
<td>12.</td>
<td>er</td>
<td>O</td>
</tr>
<tr>
<td>13.</td>
<td>build a house</td>
<td>M</td>
</tr>
<tr>
<td>14.</td>
<td>and</td>
<td>O</td>
</tr>
<tr>
<td>15.</td>
<td>plant a tree</td>
<td>M</td>
</tr>
<tr>
<td>16.</td>
<td>B:</td>
<td>yeah O</td>
</tr>
<tr>
<td>17.</td>
<td>C:</td>
<td>yeah O</td>
</tr>
<tr>
<td>18.</td>
<td>we got the same thing</td>
<td>M</td>
</tr>
<tr>
<td>19.</td>
<td>to be a real man</td>
<td>M</td>
</tr>
<tr>
<td>20.</td>
<td>D:</td>
<td>yeah O</td>
</tr>
<tr>
<td>21.</td>
<td>E:</td>
<td>do you have this sort of M</td>
</tr>
<tr>
<td>22.</td>
<td>C:</td>
<td>yes O</td>
</tr>
<tr>
<td>23.</td>
<td>E:</td>
<td>expressions M</td>
</tr>
<tr>
<td>24.</td>
<td>F:</td>
<td>i’ve never heard anything like that M</td>
</tr>
</tbody>
</table>
contexts. These O elements are ‘textual focused’, or ‘OT’. In contrast to OI, which look ‘outwards’ from the unfolding text and are thus ‘situation focused’, as they keep relating the discourse to the text-external world of interaction, the OT look ‘inwards’ to the text, the M elements, indicating and constructing complex interrelations between them. From the point of view of the M, we could characterise the O roles as ‘first-level ordering’ (OT) and ‘second-level ordering’ (OI), as the OI as it were organise the whole interaction and thus go beyond the text. Table 2 shows how these two element types appear in our example.

Table 2. Types of organisational elements: OT and OI

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 1. | A: | but
| 2. |   | er
| 3. |   | actually
| 4. |   | i know
| 5. |   | erm
| 6. |   | it's common as a saying
| 7. |   | in Poland
| 8. |   | that
| 9. |   | you
| 10. |   | er
| 11. |   | men should have a son
| 12. |   | er
| 13. |   | build a house
| 14. |   | and
| 15. |   | plant a tree
| 16. | B: | yeah
| 17. | C: | yeah
| 18. |   | we got the same thing
| 19. |   | to be a real man
| 20. | D: | yeah
| 21. | E: | do you have this sort of
| 22. | C: | yes
| 23. | E: | expressions
| 24. | F: | i've never heard anything like that

This extract has more OI than OT. In the analyses of a variety of text and conversation types in Sinclair & Mauranen (2006), this was found to be the case in casual conversations, but also, like here, in academic polylogic discussions (see
also Mauranen 2012b). The distribution of element types is sensitive to text type: in some genres, like written prose, the reverse distribution tends to be the case, with many OT and very few or even no OI at all.

As seen in the excerpt, O elements tend to be short, just two or three words, and in this case all of them happen to consist of just one item. OI are er(m) (lines 2, 5, 10, 12), actually (line 3), yeah (lines 16, 17, 20), yes (line 22), and OT are but (line 1), that (line 8), and (line 14).

All the OI help keep the discourse going, with er or actually signalling among other things continuation and intention to speak or keep the floor, yeah and yes doing little beyond showing listenership. Typical OI are also that’s right, just very briefly, as you say. At discourse level, these ‘convergent’ cases can be contrasted with ‘divergent’ ones like well, I don’t know, I don’t think, not really, which signal an opposing view or shift direction by, for example, changing the topic or making a new opening (Mauranen 2009).

Some elements in themselves are not clearly convergent or divergent, but the interpretation is contextual, as is of course typical in discourse. Thus I think, I mean, it seems are OI that can precede divergent discourse but do not have to. Similarly, elements mostly acting as OT (coming back to this; just a comment) can also undergo ‘flipping’ and count as OI, or even M elements. This is part of the context-bound nature of the analysis. Most of the time items are used conventionally, but speakers can exploit their potential flexibility, which a robust grammar must accommodate.

Even though OI and OT assume organising roles, this does not imply that they take the back seat in constructing discourse. Quite the contrary, O elements are vital in creating the whole that constitutes linguistic activity and, in essence, meaning. As Firth (1957) suggested, total language activity incorporates also other situational features, which implies that the role of OI in particular is pivotal as mediating between textual and extra-textual contexts.

5.2 Types of ‘M’ elements

Let us now turn to the other, and more diverse main category of elements. These are M or ‘message oriented’ elements, which increment the evolving topic. Their greater diversity as compared to O can partly be accounted for by the incompleteness of many M. This is especially the case in spontaneous conversation: not all communicative elements are successful, and some are just fragments; many serve to rephrase speakers’ first formulations, to modify, clarify or expand them. Clearly they may be analysable, and analysts or hearers can come up with interpretations of their import or meaning, but judging from the way speakers keep reformulating, the elements are not to their satisfaction.
The basic element is simply called ‘M’. It is a straightforward, mainly ‘grammatical’ sequence that does not need anything else to complete it (*men should have a son; we got the same thing; I’ve never heard anything like that*). Grammaticality must here be seen quite broadly, covering speech as well as non-standard and simply unusual text types, as represented in the wide variety of spoken and written text types that LUG was developed on. An M often shows full clause structure, but sometimes an individual word or a single nominal group is sufficient as a complete unit (examples from ELFA: *both; the genre; not just western Christianity; transport media plug-in*).

Other M types are not equally complete or independent. Some follow an already complete M, providing an addition or specification to it. These are ‘message supplements’ (‘MS’), which characteristically though not always are specifying elements, such as adjuncts of time or place (*in Poland* line 7 in Table 3; *to be a real man* line 19). Further M types include those that seem to start something, but fail to complete an M, like false starts (examples from ELFA in (3a) and (3b) below).

Another element type that does not lead to completion is seen in line 4 (Table 3): an ‘incomplete M’ symbolised by ‘M–’ (‘M dash’). Like a core M, this is an M in its own right, yet it raises a strong expectation of another, completing element. An M– is thus not fully independent, but anticipates something else to follow. The anticipation is not always fulfilled (like *I know* in line 4), although in most cases it is. Thus, the M– in line 21 (*do you have this sort of*) is fulfilled by *expressions* (line 23) almost immediately. *Expressions* exemplifies a ‘completion M’ (‘plus M’, ‘+M’), which supplies appropriate completion material to the M–. These two types normally appear in pairs, although as we already saw, in the fortuities of conversation it may happen that an anticipated completion of an M– fails to materialise. The conversation can simply drift to something else. O elements can come between M– and +M, so they need not be adjacent (line 22 in Table 3).

There is one pair of elements in lines 13 and 15 that is almost like the M– and +M pair just discussed, but not quite: the first element is MS, *build a house*. This clearly supplements the preceding M, *men should have a son*, which in itself neither requires nor anticipates a supplement. However, as the supplement is uttered, we anticipate something else to follow: the utterance is now neither complete nor finished. The OI (*and* line 14) followed by a +M (*plant a tree* line 15) completes this, as interlocutors B and C also recognise, B by acknowledging it (*yeah* line 16) and C with a longer supportive utterance (lines 17–19).

As seen in Table 3, the ‘core’ M are self-sufficient, MS add specifications of place (line 7) or other additional information (lines 13, 19). The interposition of O elements (*and* line 14, OI; *yes* line 22, OI) does not interfere with the continuity of the incrementation of shared experience by M.
It seems that minor, mainly monosyllabic repeats or false starts, often called dysfluencies, do not interfere with incrementation continuity either. Let us look at just two additional examples from the ELFA corpus, one illustrating word search (3a), the other a change of tack (3b). The first, like you in line 9, is MF (‘message fragment’), while the second could be seen rather as an MA (‘message adjustment’).

(3a) and i i wi- wi- will get back to it
(3b) okay some very sm- detailed thing

In (3a), the MF exemplify very brief false starts possibly resulting from word search or planning, and the analysis would be as in Table 4.
Here a few false starts appear in a row, but this does not seem to reduce the intelligibility; listeners were clearly able to take in the main import of what was being said in view of how the discussion followed smoothly on from here. Also, the final +M is nearly complete without the preceding minor attempts at sorting out the beginning. It is possible that such very brief false starts are simply not held in the working memory – possibly somewhat longer ones are not retained either, but simply recognised for what they are and erased immediately. Likewise, in Table 5 where (3b) is analysed, the message adjustment does not hamper intelligibility. Message adjustment (MA) is an M that gets modified in progress. As here, there is an interruption, but no false start that could be disregarded, because the M unit continues without repeating the beginning (some very). Probably in cases like this where the elements are adjacent the extra load on working memory is negligible. The extracts are so short here that they fit in comfortably within working memory span in any case, but they illustrate the point.

An independent M that can also be followed by another element that does not make a new departure of its own is ‘MR’, ‘message revision,’ which reworks an M by a slight reformulation like a small expansion or rephrasing. This is a recurrent feature in conversations, even though it does not appear in our main example. I shall therefore present a final example from the same corpus (4), both to illustrate this element and at the same time also to give a new extract that shows several of the elements introduced above in a new context.

(4) A: there are still shops, streets with shops and
    B: yeah
    C: no they are they are just
    D: oh yeah
C: located outside the city centre
D: yeah but outside yeah
C: they have better roads there you can go there with a car easily so

I present the passage analysed as a whole below in Table 6, but the main point of interest is line 2, streets with shops. The element revises the phrasing that the speaker first gave. Such reformulations, often with slight expansions, are common in speech, like the changes of tack that we saw in Example (3b). As speech is temporally irreversible, speakers can only reformulate by adding to what they already said. In line 7 we see a re-start (they are just) following a false start in line 6 (they are), with enough shared substance to talk about an MR here. Perhaps the most unexpected MR is that in line 14, where a different speaker (D) repeats part of the message that C just produced. As we see from this, LUG does not adhere to speaker shifts, but takes the discourse as an evolving, joint product, which draws on the contributions of the participants as it increments shared knowledge among the participants.

Table 6. Analysis of passage (4)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>A: there are still shops</td>
<td>M</td>
</tr>
<tr>
<td>2.</td>
<td>streets with shops</td>
<td>MR</td>
</tr>
<tr>
<td>3.</td>
<td>and</td>
<td>OT</td>
</tr>
<tr>
<td>4.</td>
<td>B: yeah</td>
<td>OI</td>
</tr>
<tr>
<td>5.</td>
<td>C: no</td>
<td>OI</td>
</tr>
<tr>
<td>6.</td>
<td>they are</td>
<td>MF</td>
</tr>
<tr>
<td>7.</td>
<td>they are just</td>
<td>MR</td>
</tr>
<tr>
<td>8.</td>
<td>D: oh yeah</td>
<td>OI</td>
</tr>
<tr>
<td>9.</td>
<td>C: located outside the</td>
<td>+M–</td>
</tr>
<tr>
<td>10.</td>
<td>er</td>
<td>OI</td>
</tr>
<tr>
<td>11.</td>
<td>city centre</td>
<td>+M</td>
</tr>
<tr>
<td>12.</td>
<td>D: yeah</td>
<td>OI</td>
</tr>
<tr>
<td>13.</td>
<td>but</td>
<td>OT</td>
</tr>
<tr>
<td>14.</td>
<td>outside</td>
<td>MR</td>
</tr>
<tr>
<td>15.</td>
<td>yeah</td>
<td>OI</td>
</tr>
<tr>
<td>16.</td>
<td>C: they have better roads</td>
<td>M</td>
</tr>
<tr>
<td>17.</td>
<td>there</td>
<td>+M</td>
</tr>
<tr>
<td>18.</td>
<td>you can go there</td>
<td>M</td>
</tr>
<tr>
<td>19.</td>
<td>with a car</td>
<td>+M</td>
</tr>
<tr>
<td>20.</td>
<td>easily</td>
<td>+M</td>
</tr>
<tr>
<td>21.</td>
<td>so</td>
<td>OT</td>
</tr>
</tbody>
</table>
All the other element types in the extract have already been discussed above, apart from +M – (line 9), which is a combination of +M and M –. In other words, it is both prospected by the M – in line 7 and it anticipates continuation, which is interrupted by er in line 10.

Some subtypes of M elements have not been discussed here, but what I regard as the most central ones have been included. For the rest, the reader is encouraged to consult Sinclair & Mauranen (2006). As is often pointed out in LUG, the analysis is context-sensitive and therefore what might look like category flips, i.e. certain elements appearing in one function in one context, and another elsewhere, can happen. I don’t know is a case in point. Additionally, some analyses can have alternative interpretations. In Table 6, the final so (line 21) can arguably also be seen as an OI, indicating the speaker’s readiness to hand over the floor.

Some further developments have been suggested to M elements in LUG. One is concerned with written dialogue: Smart (2013) analyses message board discussions, and adds an M element type modelled on Brazil’s (1995) ‘suspension’ and the notion of ‘challenge’ from discourse analysis, which enables him to deal with cases where the expectations set up in linear processing are not met. This is potentially an important consideration for speech processing as well, because the progress of prospection presumably relies on both success and failure in making guesses about what is to follow. Surprises are always possible, and they make us alter prospections already formed. Brazil’s notion of suspension was also adopted by Carey (2015) in modelling speech samples from the ELFA corpus and SELF (Studying in English as a Lingua Franca) data (www.helsinki.fi/englanti/elfa/self.html). He developed the MS to include one that accounts for a temporary discontinuity in the progression of speech. In this way, MS would consist of two kinds: those that add something to an already complete M (an ‘extension’) and those that add information to an M– in progress (a ‘suspension’).

In all, M type elements are attempts, successful or not, to increment shared experience. What have been illustrated here cover most of the types identified so far in LUG, although not quite all of the minor kinds. M elements tend to be analysable by other, conventional grammars, and while it is possible to go deeper into subtler distinctions in LUG as well, it is not its first priority. What LUG has to offer is not another breakdown of well-established units of language, but a linear perspective on the way we look at these elements in relation to each other.

6. Synthesis – linear units of meaning

Following our analysis of the chunks and the types of elements, the final stage reverses the direction: instead of an analysis, it makes a synthesis. The final step brings together the outcomes of the earlier steps. It also helps serve as a bridge
between LUG and other grammars. Moreover, it also narrows the gap between
descriptions of spoken and written text, although unconventionally starting with
speech. I shall draw on the last example above (4) for illustration.

The synthesis results in linear units of meaning (LUM), *topic increments* that
update the virtual world of shared experience. LUMs come fairly close to units
generally recognised in grammars, such as clauses and phrases. They incorporate
the rudiments of hierarchy in that they recognise two kinds of relationship, one
of which can constitute a component in the other: the *endocentric* forms ‘textual
objects’, single entities like noun phrases that must combine with something else to
set up a communicative act (*streets with shops*). The *exocentric* relationship forms
‘textual incidents’ by relating two separate entities such as subjects and predicates
(*there are streets with shops*). Textual objects are thus typical components of tex-
tual incidents. It is the textual incident that updates and changes the virtual world
of shared experience. Importantly, textual incidents can also include non-textual
objects. Such exocentric relationships relate textual objects with contextual, or situ-
tational, elements – as a metal container and a label on it saying *Tea*. *Tea* is a textual
object, and the container a situational object; together they form a textual incident.

This final phase of synthesis applies again a staged approach: altogether eight
operations are applied one after the other as the synthesis takes shape. Here only a
few are illustrated, together with the outcome.

At the outset, all OI (*er, yeah*) are shelved, and kept for later as notes. The focus
in this somewhat skeletal synthesis is on structural relevance, and OI lose theirs
immediately. In our example (Example (4) above) the OI do not play a very major
role in the interaction, apart from contributing to the fluent construction of the
conversation, at some point co-constructing consensus as well (lines 12 and 15 in
particular). In more conflict-driven conversations OI assume more central and
divergent roles.

A number of operations are performed on M elements; M– and +M pairs
are joined up (for instance *they are just* with *located outside the* and *city centre*,
together with *outside*, lines 7, 9, 11, 14). In principle MS are added to the nearest
preceding M, although there are no instances in our present example. The third
operation merges MR with those they repeat or reformulate: a reformulation of
(*there are still*) *shops* (line 1) expanded as *streets with shops* (line 2); in the same
way *they are* (line 6) and *they are just* (line 7) are joined. We can retain the final OT,
move it a little and put the other combinations in place. The example now reads:

there are still streets with shops –
they are just located outside the city centre –
they have better roads there so you can go there with a car easily

Some added dashes for clarity are all that is needed to make this a readily read-
able passage that is analysable by any conventional grammar. The dashes could of
course also easily be replaced by more conventional punctuation. An intriguing outcome of the synthesis is the smoothly progressing, coherent text that was collaboratively produced by the interlocutors. This is not likely to be an intended consequence of a spontaneous conversation, even academic discussion; we would not even expect such an outcome from a formal meeting or a session of a task force. If and when they produce joint formulations, these are suggested by participants as tentative formulations, with amendments being put forward by others. But as we see here, co-construction of meaning seems to be at work in ordinary, seemingly wandering discussion. LUG thus offers support for analyses of co-construction in conversation (McCarthy 2010; Clancy & McCarthy 2015) from an alternative perspective. Moreover, speaker changes do not seem to play a crucial role in the jointly constructed text.

7. Conclusion

LUG makes a contribution to the analysis of spoken interaction with its strict focus on a linear, temporal perspective on unfolding talk. It seeks to address the challenge of bringing together what we know from the nature of processing speech and how we are used to modelling language. While one is severely restricted by the temporal dimension, the other still largely retains a hierarchical position in its analysis.

At the same time, LUG is not confined to the analysis of spoken language, but straddles speech and writing. This paper did not draw examples from written text, unlike Sinclair & Mauranen (2006) or Smart (2013), as it is in speech that the specific features of the model are best seen. Nevertheless, interactional aspects of speech are not fully taken on board yet in this somewhat skeletal description. Although a structural emphasis tends to downplay interactive aspects of communication, this is not what LUG aims at. Quite the contrary, the model recognises the crucial impact that interaction and the situational context have on grammatical description, as well as the complexities of capturing a number of facets of verbal interaction embedded in multimodal social situations, for instance those ensuing from the simultaneously collaborative and competitive nature of interaction. Those are not sufficiently described by looking at language only, but linguistic models must prioritise the workings of language. In LUG, OI specialise in managing interaction, and are vital in reaching out to the situation of context. Their outward orientation and focus on the situation connect them immediately to interaction. They are thus also indispensable to integrating grammar and discourse.

One of the future directions of developing LUG that have come up in the present paper, albeit only in passing, is the interactional potential that LUG holds. It provides a supportive but nevertheless novel perspective on co-construction,
intersubjectivity and distributed cognition that starts from the observation that the collaborative outcome of a piece of conversation is not much dependent on how the turn-taking progresses. Several participants can co-produce coherent text, and from the viewpoint of the text the individual contributions are unimportant – while for instance in Conversation Analysis participants’ actual turn-taking behaviour has been a central point of interest.

Even more crucial to the present focus is what LUG can contribute to the understanding of cognition and chunking. To begin with, the process of chunking in on-line language processing is in need of further cognitive validation. Inter-rater reliability on a larger scale than hitherto is needed to supplement the sources already used, and more work is under way: for instance, a research group at Helsinki, CLUMP (Mauranen et al. 2015), is engaged in testing the plausibility of the chunking process by various means, including experimental methods, which in the pilot phase have combined soundtrack and transcript. The unpublished findings from the pilot study suggest a high level of consensus in chunking. In the next phase this research will involve more study participants, and continue with sound. It is important that further research on chunking processes proceeds with sound included.

Linear Unit Grammar has not been developed or tested by other scholars very much yet (but see Mauranen 2009, 2009/2012; Carey 2011, 2013, 2015; Smart 2013). Yet as the chunking of language has become a point of interest among brain researchers (e.g. Giraud & Poeppel 2012; Christiansen & Chater 2015) and since dynamic approaches to language are simultaneously gaining ground (see papers in this issue), we can expect more research in this domain in the near future. The potential of taking linear, fast processing on board is great for developing better-fitting models of language.

Primary data


References


All rights reserved


All rights reserved


**Author’s address**

Anna Mauranen  
P.O. Box 3  
Fabianinkatu 33  
FI-00014 University of Helsinki  
Finland

anna.mauranen@helsinki.fi