Chapter 13

THE MOTIVATION FOR SYNTACTIC COMPLEXITY
IN
LANGUAGE

1. Introduction

Several characterizations of sentential complexity (e.g. Mithun 1984 and 1988) have suggested that it is connected with the (supposed) lengthened temporal processing afforded by literacy or ritual language (Chafe 1982). Chafe (1982.37-39) notes that

... we write more than 10 times more slowly than we speak ... In writing, then, our thoughts must constantly get ahead of our expression of them in a way to which we are totally unaccustomed when we speak. As we write down one idea, our thoughts have plenty of time to move ahead to others. The result is that we have time to integrate a succession of ideas into a single linguistic whole in a way that is not available in speaking. In speaking, we normally process one idea unit at a time. That is apparently about all we have the capacity to pay attention to, and if we try to think about much than that we are likely to get into trouble ... In writing we have time to mold a succession of ideas into a more complex, coherent, integrated whole, making use of devices we seldom use in speaking ... I will say that written language tends to have an 'integrated’ quality which contrasts with the fragmented quality of spoken language ... The fragmentation of spoken language shows up in the stringing together of idea units without connectives ... Integration refers to the packing of more information into an idea unit than the rapid pace of spoken language would normally allow.

A possible implication of these statements is that complexity is the result of literacy, but Chafe (1982.50) is careful not to allow this interpretation:¹

¹ Chafe (1982.49-50) suggests that ritual language provides a cognitive freedom in a manner similar to written language, which allows ritual language to maintain a similar complexity:

... I suspect ... that ritual language, like written language, has a permanence which colloquial language does not. The same oral ritual is presented again and again, not verbatim, to be sure, but with a content, style, and formulaic structure which remain
To find these integrative devices missing in a language without a written tradition might suggest, of course, that they are features which arise in a language precisely because of writing. But such a conclusion would have to be treated with a large dose of skepticism, since there are numerous other unwritten languages [in addition to Seneca, PWD] in which at least some of these same features exist.

Neither of these scholars addresses the issue of whether the formal resources of clause combining convey content, that is, that it might mean something to exploit complexity in place of simplicity, and Mithun (1988) perceives the loss of content in the presence of sentential complexity. If we are able to ‘predict’ the appearance of complexity, or at least if we are able to identify some language characteristics which are co-present with its appearance and which are implicated in its emergence, the ‘problem’ of sentential complexity may seem to have been solved, i.e., to have been understood, and thus no further questions need be asked. If understanding a phenomenon is attained by providing an explanation for its existence, then the question of Why does clause-linkage exist?, and the rule predicting its presence constitute the issue of sentential complexity and its resolution. But, if the pattern of sentential complexity is cast in terms of What does it mean to frame some content in a sententially complex way in place of a concatenation of sentences?, then we must not seek a pattern in terms of a rule which predicts the existence of clause-linkage in language, but we must seek to find pattern in the semantics of the alternate expressions.

Lehmann (1988), more directly than others, addresses the question of content of complexity, drawing from his typology of clause linkage a semantic scale of

```
compression —————— elaboration
```

which he explains in this way (Lehmann 1988.217-18):

In a functional framework, clause linkage may be viewed as either representing two states of affairs so tightly interconnected that they form one complex state of affairs (compression), or on the contrary analyzing one state of affairs as composed of two (elaboration ...).

constant from performance to performance ... As a result they contain language that has been formalized and polished, even over many centuries, contrasting with the spontaneity and roughness of conversation. We might then expect to find in ritual language something like the integration of written language, as opposed to the fragmentation of spoken.
Lehmann does not provide this semantics more detailed discussion. Matthiessen & Thompson see a similar semantics in clause linkage (1988.317):

... hypotactic clause combining is best understood as a grammaticalization of the Nucleus-Satellite relations which characterize the rhetorical organization of certain types of written discourse ... if hypotaxis in English is a grammaticalization of rhetorical relations then it follows that the grammar of clause combining may differ radically from one language to another ... If the basic approach to clause combining taken in this paper is correct, then the interesting cross-linguistic issue is how and to what extent the grammar of clause-combining in a given language reflects the rhetorical organization of discourse in that language ...

The similarity between Lehmann and Matthiessen & Thompson is that the nucleus-satellite relation may be taken as one degree of the compression which Lehmann perceives. One difference between them is that Matthiessen & Thompson see that compression as a characteristic of (written) discourse.

2. **Motivation**

Let us assume that this view is broadly correct. Then why is compression present in discourse and how? Chafe (1982) writes of ‘integration’ and presents these as examples of it (Chafe 1982.39-44):

(i) Nominalizations, e.g. “treatment rather than treat” (Chafe 1982.39),
(ii) Participles,
(iii) Attributive adjectives, e.g. “allow states to be expressed as modifiers rather than assertions” (Chafe 1982.41),
(iv) Conjoined phrases,
(v) Series, e.g. “No capital letters, definite articles, or plural markers ...” (Chafe 1982.43)
(vi) Sequences of prepositional phrases,
(vii) Complement clauses,
(viii) Relative clauses.

The notion of integration, which (i) - (viii) represent, is made more explicit with the notion of the idea unit (Chafe 1982.37):

Observation of spontaneous language has led several investigators independently to
the finding that it is produced in spurts, sometimes called idea units, with a mean length (including hesitations) of approximately two seconds or approximately six words each ... Idea units typically have a coherent intonation contour, they are typically bounded by pauses, and they usually exhibit one of a small set of syntactic structures. They are a striking, probably universal property of spoken language. It is useful to speculate that each idea unit represents a single ‘perching’ of consciousness ..., or a single ‘idea’ in that sense. If that is true, then when we speak we are in the habit of moving from one idea to the next at a rate of about one every two seconds. Perhaps that is even our normal ‘thinking rate’, if language reflects the pace of thought. Whether or not that rate applies to all thinking, it is certainly a rate we are accustomed to while we are using language — probably while we are thinking in language to ourselves as well as when we are overtly vocalizing.

If thought is not limited to that pace, i.e., if the pace of behavior is not the pace of cognition, then the pace of production would seem to provide enough time to organize sentential complexity. Perhaps, even, the accommodation of such mental calculation provides some motivation for the apparently slow pace of production.

2.1 Miller

The number six which Chafe cites is curious. Miller (1956.81) writes:

My problem is that I have been persecuted by an integer. For seven years this number has followed me around, has intruded in my most private data, and has assaulted me from the pages of our most public journals. This number assumes a variety of disguises, being sometimes a little larger and sometimes a little smaller than usual, but never changing so much as to be unrecognizable. The persistence with which this number plagues me is far more than a random accident. There is, to quote a famous senator, a design behind it, some pattern governing its appearance. Either there really is something unusual about this number or else I am suffering from delusions of persecution.

The number is seven plus or minus two. The number, for example, describes the limit placed upon the amount of information our senses may process, their channel capacity as measured by the amount of transmitted information (Miller 1956.83):

2 In 1994, the number is 4.84 for English (Chafe 1994.65). In 1987, the number was “about five or six words ... about two seconds apart” (Chafe 1987.22). In 1979, the number was “about 5 words” and the mean duration was “slightly less than 2 sec.” (Chafe 1979.164).
One bit of information is the amount of information that we need to make a decision between two equally likely [i.e., meaningless, PWD] alternatives. Two bits of information enable us to decide among four equally likely alternatives. Three bits of information enable us to decide among eight equally likely alternatives. There are two ways we might increase the amount of input information. We could increase the rate at which we give information to the observer, so that the amount of information per unit time would increase. Or we could ignore the time variable completely and increase the amount of input information by increasing the number of alternative stimuli. In the absolute judgment experiment we are interested in the second alternative. Confusions will appear near the point that we are calling his ‘channel capacity’ [Emphases mine, PWD].

Figure 1 presents some results in the discrimination of pitch, loudness, and points in a square. For the discrimination of tone, the “result means that we cannot pick more than six different pitches that the listener will never confuse” or “no matter how many alternative tones we ask him to judge, the best we can expect him to do is to assign them to about six different classes without error” (Miller 1956.84). The discrimination of degrees of loudness differs slightly, but “the important point ... is that the two answers are of the same order of magnitude” (Miller 1956.85). Similar results follow from taste intensities, visual position, hue, and brightness (Miller 1956.98):

Let me summarize the situation in this way. There is a clear and definite limit to the accuracy with which we can identify absolutely the magnitude of a unidimensional stimulus variable. I would propose to call this limit the span of absolute judgment, and I maintain that for unidimensional judgments this span is usually somewhere in the neighborhood of seven.

The increase in the ability to locate points in a square to 4.6 bits (in multidimensional judgments), demonstrates the influence of two dimensions.3 Results from experiments with ‘immediate memory’ (aka short-term memory) show analogous results. But Miller (1956.91) gives this caveat about equating the performances from Figure 1 with those from Figures 2 and 3

3 Miller (1956.88):

Clearly, the addition of independently variable attributes to the stimulus increases the channel capacity, but at a decreasing rate. The point seems to be that, as we add more variables to the display, we increase the total capacity, but we decrease the accuracy for any particular variable. In other words, we can make relatively crude judgments of several things simultaneously.
(Hintzman 1978.237 & 240):

What is more natural than to think that all three of these spans are different aspects of a single underlying process? And that is a fundamental mistake ... There is a real operational similarity between the absolute judgment and the immediate memory experiment. If immediate memory is like absolute judgment, then it should follow that the invariant feature in the span of immediate memory is also the amount of information that an observer can retain. If the amount of information in the span of immediate memory is a constant, then the span should be short when the individual items contain a lot of information and the span should be long when the items contain little information.

Figure 1: Some constraints on input information.
But this is not so. For example, if the medium for measuring immediate memory is sounds (e.g. English phonemes), the limit will fall into the range of seven plus or minus two. But the limit on the memory of sounds may triple if the medium is monosyllabic words (Miller 1956.92-93):

![Figure 2: An 'amount' constraint on immediate memory.](image)

![Figure 3: A temporal constraint on immediate memory.](image)
Absolute judgment is limited by the amount of information. Immediate memory is limited by the number of items. In order to capture this distinction in somewhat picturesque terms, I have fallen into the custom of distinguishing between \textit{bits} of information and \textit{chunks} of information. Then I can say that the number of bits of information is constant for absolute judgment and the number of chunks of information is constant for immediate memory. The span of immediate memory seems to be almost independent of the number of bits per chunk ...

2.2 \textit{Lashley}

The inference of chunk-like organization can be arrived at in another way. Lashley (1951) considers the problem presented by the serial order of behavior or the temporal integration of behavior (Lashley 1951.181):

\begin{quote}
Temporal integration is not found exclusively in language; the coordination of leg movements in insects, the song of birds, the control of trotting and pacing in a gaited horse, the rat running the maze, the architect designing a house, and the carpenter sawing a board present a problem of sequences of action which cannot be explained in terms of successions of external stimuli.
\end{quote}

The question was ‘Where does it arise?’ or ‘What controls it?’ Early on, there was a suggestion that it was physiologically controlled, and that it was a phenomenon, peripheral in the nervous system, “a chain of sensory-motor reactions” (Lashley 1951.188), and not centrally controlled. This view depended on ‘peripheral chains’ (Lashley 1951.182) in which speech, for example, was “a succession of vocal acts in which the kinesthetic impulses from each movement serve as a unique stimulus for the next in the series”. Introspective psychologists may have objected to this formulation in terms of peripheral control, but in their alternative, “there is implicit ... a view that verbal thought is a simple chain of central processes in which each element serves to arouse the next by direct association” (Lashley 1951.182). In another context, the succession of leg movements in an insect follow the pattern of

Figure 4: A repair to damaged serial movement.
in Figure 4. But “with the removal of the left first leg, the right first and left second alternate and the order becomes right first, left third, right third stepping together, with left second and right second advancing together, instead of alternately” (Lashley 1951.189). Cf. B in Figure 4. This change is “spontaneous”, and it appears then not to arise from “the action of predetermined anatomic paths, but is the result of the current physiological state of the various limb centers” (Lashley 1951.189). Another observation which suggests a more central control of motor patterns is this (Lashley 1951.188):

The finger strokes of a musician may reach sixteen per second in passages which call for definite and changing order of successive finger movements. The succession of movements is too quick even for visual reaction time. In rapid sight reading it is impossible to read the individual notes of an arpeggio. The notes must be seen in groups [i.e., Miller's chunks, PWD], and it is actually easier to read chords seen simultaneously and to translate them into temporal sequence than to read successive notes in the arpeggio as usually written ... Such acts ... require the postulation of some central nervous mechanism.

Lashley (1951.183) adheres to the following conception:

I believe ... that such interpretations of temporal organization [based on some mechanism other than the central nervous system, PWD] are untenable and that there are, behind the overtly expressed sequences, a multiplicity of integrative processes, which can only be inferred from the final results of their activity.

In language, Lashley notes that the ‘word arrangement’ of /rayt/ in

(1) The mill-wright on my right thinks it right that some conventional rite should symbolize the right of every man to write as he pleases.

“is obviously not due to any direct association of the word ‘right’ itself with other words, but to meanings which are determined by some broader relations” (Lashley 1951.183). And furthermore, “the readiness with which the form of an expression of an idea can be changed, the facility with which different word orders may be utilized to express the same thought, thus is further evidence that the temporal integration is not inherent [Emphasis mine, PWD] in the preliminary organization of the idea” (Lashley 1951.185).
And (Lashley 1951. 187),

There is a series of hierarchies of organization; the order of vocal movements in pronouncing the word, the order of words in the sentence, the order of sentences in the paragraph, the rational order of paragraphs in a discourse. Not only speech, but all skilled acts seem to involve the same problems of serial ordering, even down to the temporal coordination of muscular contractions in such movement as reaching and grasping.

And finally (Lashley 1951.188):

This is the essential problem of serial order; the existence of generalized schemata of action which determine the sequence of specific acts, acts which in themselves or in their associations seem to have no temporal valence.

3. Conclusion

And here we return to Chafe and his idea units to compare them with the generalized schemata of Lashley and with the notion of chunks (Miller 1956.92-93):

I have fallen into the custom of distinguishing between bits of information and chunks of information. Then I say that the number of bits of information is constant for absolute judgment and the number of chunks of information is constant for immediate memory. The span of immediate memory seems to be almost independent of the number of bits per chunk, at least over the range that has been examined to date.

The contrast of the terms bit and chunk also serves to highlight the fact that we are not very definite about what constitutes a chunk of information.

It is this capacity to chunk and/or to integrate serial behavior into atemporal schemata (Lashley 1951.188) which underlies and allows the compaction, integration, or compression which the case of clause linkage represents. From the perspective of clause linkage and idea units, we see that chunking and integration into schemata is not a matter of either-or, but one that happens by degree. And it may be that it is precisely the indeterminacy as to what composes a ‘chunk of information’ which allows the variety in its formation in language. Chafe (1979.161-64) describes a hierarchical model in which there is a correspondence between cognitive levels (with “larger cognitive

---

units containing smaller ones”) and levels of expression in language:

<table>
<thead>
<tr>
<th>Cognitive Units</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>memories</td>
<td>stories</td>
</tr>
<tr>
<td>episodes</td>
<td>paragraphs</td>
</tr>
<tr>
<td>thought</td>
<td>sentence</td>
</tr>
<tr>
<td>focus</td>
<td>phrase</td>
</tr>
</tbody>
</table>

“If, then, in the hierarchical model memories contain episodes and episodes thoughts, we can go one step further and say that thoughts contain FOCI, expressed linguistically in PHRASES (usually ‘clauses’, but not always)” (Chafe 1979.164).

We may take the semantic componency of ‘clauses’ as basic, i.e., set the value of informational unity as equivalent to the Indo-European clause; but there is no necessity to this (Chafe 1979.164):

... people are able to segment the flow of actions into units of different size according to whether they are instructed to mark ‘the smallest actions that seem natural and meaningful’ or the largest. ‘For example, one might see a person get up from a chair, walk over to a door, close it, turn, and walk back to his chair, and mark off each segment as a discrete, meaningful action, or one might see the whole sequence as just one action — closing the door...

There is nothing which prevents the idea unit from being larger, as in clause chaining languages (Longacre 1985). Here, the integrating unit may be the chain bounded by a final verb along with its medial verb clauses. Nor is there any reason that the value of unity cannot be smaller as, perhaps, in those languages such as Makah (Wakashan), in which a principal clause may be followed by one or more absolutive clauses (Jacobsen 1979.115 & 119):

(2) ḥaqi-bits wiku-ya-w-iq ḋi-ʁiʁ
    [give-past-indic-1sg boy-art dog]
    ‘I gave the boy a dog’

(3) ṣəkyiŋ huku-t p ḡə-ʁiʁ
    [many-indic.3 bird fly]
But, in the absence of such a following clause, the principal clause will admit but one expressed PARTICIPANT, and in itself seems to be ‘smaller’ in content than the familiar Indo-European clause. That is, the projection of ‘idea unit’ upon Nootka will encompass less content than we may expect. To compose a semantic constellation which approximates the Indo-European clause, it is required to add more clauses, i.e., sentential complexity, rather than to simply add to the non-complex clause.

Viewed in this way, there are now **two** factors to complexity: (i) how one sets the extent of the idea unit and (ii) how material not included in that unity may, nevertheless, be set in relation to it. In each case, it is the ability to chunk (once to create the unity and again to relate external content to it) which enables such patterns. And the semantics of complexity is precisely the linguistic formation of that cognitive capacity, i.e., the semantics of ‘chunking’, ‘integration’, ‘nucleus-satellite’, etc.

---

5 The difference between writing and speaking is not in terms of speed or ease, but in terms of where the notion of ‘unity’ is applied. And in writing, it is probably ‘lower’ than it is in speech. Ironically, speech is the more integrated ... hence, needing fewer grammatical marks. Writing, being less integrated, requires more marks of the universal presence of integration ... less automatic.