

Fred Karlsson

Origin and maintenance of clausal embedding complexity¹

In: Geoffrey Sampson, David Gil and Peter Trudgill, eds., *Language Complexity as an Evolving Variable*. Oxford University Press, 2009, 192-202.

1. Introduction

There are many well-known parameters of grammatical complexity: number of terms (e.g. cases or genders) in a grammatical subsystem; more or less intricate elaboration of phrases by premodification and postmodification; restricted versus repeatable nominalization processes; number of affixes in word-forms; simplex versus iterative compound formation; number and depth of embedding of clauses in sentences, etc.

The purpose of this paper is to analyze one aspect of grammatical complexity, that of clausal embedding complexity, by which we mean the possibilities of repeatedly embedding subordinate clauses in various positions in their main clauses. After a presentation of some clausal embedding complexity constraints recently established for 'Standard Average European' (SAE) languages like English, German, Swedish and Finnish, we shall inquire what the origin of these constraints is and whether they have been stable or fluctuating over time. Particular attention will be paid to center-embeddings, the theoretically most interesting type of clausal embedding.

2. Constraints on clausal embedding complexity

Clausal embedding complexity increases with repeated embedding of clauses in initial, central or final position. Initially-embedded clauses (IEs) have either nothing to their left, or at the most a subjunction of the superordinate clause (also cf. Dryer 1980: 154-155 and Quirk et al. 1989: 1037). Center-embedded clauses (CEs) have superordinate clause constituents to their left (other than subjunctions) and right. Final embeddings (FEs) have no material rightwards that immediately belongs to the superordinate

clause. IE and FE manifest tail-recursion, CE is nested recursion.

Karlssohn (2007a, b, c, 2008) has empirically established constraints, i.e. quantitative limits and qualitative restrictions, on clausal embedding complexity in many European languages. His data were (i) complex sentences in the British National Corpus (BNC), the Brown corpus, and the LOB corpus, (ii) computerized extraction of complex embedding patterns in Finnish, German, and Swedish, and (iii) consultation of more than one hundred corpus-based syntactic and stylistic descriptions of European languages, especially Latin and older variants of German, both well-known for being syntactically complex.

Here (1-5) are some constraints operational in Standard Average European languages. 'I' stands for initial clausal embedding, 'C' for clausal center-embedding, 'F' for final clausal embedding, and the raised exponent expresses the maximal degree of embedding of a sentence, e. g. I² is double initial embedding as in sentence (6). Expressions like C-2 indicate type and embedding depth of individual clauses; e. g. C-2 is a center-embedded clause at depth 2.

- (1) I²max: the maximal degree of initial embedding is two (some 100 instances found, as in (6)).
- (2) Qualitative I²-constraints: Double initial embedding strongly prefers a) written language, b) an *if*-clause as higher embedding, c) a sentential subject, i.e. a *what*-clause as lower embedding, and d) finiteness. Cf. (6).
- (3) C³max-w: in written language, the maximal (utterly rare) degree of multiple center-embedding is three (thirteen instances retrieved, (7)).
- (4) C²max-s: in spoken language, the maximal (utterly rare) degree of multiple center-embedding is two (less than five instances retrieved, (8)).
- (5) Only-postmodifying-self-embedding: only postmodifying (i.e. typically relative) clauses allow self-embedding (7, 8), i.e. repeated embedding of

precisely the same type of clause.

- (6) [Main [I-1 If [I-2 what he saw through security] did not impress him] Tammuz ...] (BNC)
- (7) [Main Der Ritter von Malzahn, [C-1 dem der Junker sich als einen Fremden, [C-2 der bei seiner Durchreise den seltsamen Mann, [C-3 den er mit sich führe,] in Augenschein zu nehmen wünschte,] vorstellte,] nötigte ihn ...] (von Kleist, *Michael Kohlhaas*)
- (8) [Main A lot of the housing [C-1 that the people [C-2 that worked in New Haven] lived in] was back that way.]

No genuine triple initial embeddings nor any quadruple center-embeddings are on record ('genuine' here means sentences produced in natural non-linguistic contexts, not sentences produced by professional linguists in the course of their theoretical argumentation). Only some 130 double CEs were retrieved among the tens of millions of words checked, indicating that even C² is a rarely used structural option. Here are three more of these genuine double CEs (see Karlsson 2007c for details):

- (9) a. [Main A number of speeches [C-1 into which a great deal of thought and preparation on a level a great deal higher [C-2 than is common in modern politics] have gone] are not reported at all ...]
- b. [Main ... for that matter, [C-1 when one compares Swann and Jeunes filles, [C-2 - in which the theme of homosexuality remains latent -] with the shoddiness of the later volumes,] one is inclined to wonder ...]
- c. [Main It was not [F-1 until he was an old man [F-2 that one day his son, [C-3 who, [C-4 as was the way of the world,] had left the shamba] explained to him [F-3 that ...]]]]

The overall rarity and the numerous constraints on multiple IE and CE prompt the conclusion that multiple clausal initial and center-embeddings are not fully recursive in

SAE. Particularly important is the observation that multiple center-embedding, i.e. multiply nested clausal recursion, is practically absent from ordinary spoken language (4). Only a handful of genuine spoken C²s like (8) are on record.

Note that C³max-w and Only-postmodifying-self-embedding taken together license triple relative self-embedding which does indeed occur (7).

Final clausal embedding is less constrained but not totally unregimented. In an extensive empirical study Karlsson and Sinnemäki (to appear) found evidence in SAE languages for a tendency called F³⁻⁵max:

- (10) F³⁻⁵max: Syntactically simple genres (such as spoken language) avoid final clausal embedding in excess of degree three, complex genres in excess of degree five.

As sentence (11) from the Brown corpus with eight finally-embedded clauses shows, F³⁻⁵max may be overstepped, but such sentences are perceived as complex. F³⁻⁵max thus is a stylistic preference. In what follows, written language only will be treated, and therefore we refer to (10) by the abbreviation F⁵max-w.

- (11) [... it seems ... [that ... it is unlikely [to bear much relation ... to the ... need of [ensuring ... [that there may be adequate opportunity given to the staff [to do [what they can [to get the man [to stand again ...]]]]]]]]]]]]]]]]]]

3. Origin of clausal subordination

Typical everyday spoken language consists of brief utterances containing few constituents and less than ten words. The utterances of early spoken language or 'protolanguage' must have been even briefer. Formally marked embedding of finite sub-clauses, hypotaxis, arose much later than such simple proto-clauses, but also later than clausal coordination (parataxis) and later than non-finite constructions based on infinitives, participles and gerunds, all of which presuppose the existence of a (finite)

main clause verb. Deutscher (2000) substantiates this claim for Akkadian, Itkonen (1966) for Finno-Ugric.

Before the advent of writing in the third millennium BC, the major expository genre was oral narrative, demonstrated to be aggregative and paratactic rather than subordinating. Phrases are strung together into loosely conjoined shallow sequences (Lord 1960). This pattern of ancient additive structure is found across the world. As a case in point, Leino (1975) reports that the famous *Sampo*-section (402 lines of verse, some 1300 words) in the *Kalevala*, sung by Arhippa Perttunen to Elias Lönnrot in 1834, contains only three subordinate clauses, all 'when'-clauses embedded at depth 1 immediately below the main clause. For comparison, 1300 words of current written Finnish would contain some 60 finite subordinate clauses (Hakulinen et al. 1980).

Non-finite clausal embedding, especially final embedding such as the infinitive constructions in (11), certainly dates back to preliterate times in many languages. As for the origin of finite clausal embedding, there are preliterate languages lacking finite subordinate clauses, e. g. Inuktitut, which is in the very process of obtaining clausal embedding along with the development of the native press (Kalmár 1985). But finite subordinate clauses do occur in some preliterate languages. Mithun (1984) notes the existence of such clauses in oral texts in Mohawk, Gungwingu (Australian), Kathlamet (Chinookan) and Tlingit (Na-Dené). However, their incidence is much lower than in English.

Even if some basic finite embedding patterns date to preliterate times, there is ample evidence e. g. from Semitic, Indo-European and Finno-Ugric languages that the emergence of more elaborate grammaticalized patterns of finite clausal subordination is related to the advent of written language, especially to the conventionalization of various written registers. Proof of this development is provided e.g. by Deutscher (2000) for Akkadian, by Lehmann (1974) for subordinate clauses in Vedic, by O'Neil (1976) for subordinate clauses in Old English, by Harris (1988) for concessive clauses in English and Romance, and by König and van der Auwera (1988) for subordinate clauses in Dutch and German.

4. Origin and development of I²max

In view of what was said in Section 3 concerning the overall emergence of embedding, it is clear that the appearance of multiple clausal embeddings (especially finite ones) must be sought in the realm of written language. The age of this phenomenon can be no more than 5000 years because Sumerian started to be written around 3000 BC.

As already noted, no genuine triple (or deeper) INITIAL CLAUSAL EMBEDDINGS have been attested from any period of time. Interestingly, the oldest genuine instance of I² we have found is so old as to be from Hammurabi's Code of Law (Middle Babylonian, 1800 BC). The sentence starts with the subjunctions *šumma ištu* 'if after' and here is its (somewhat shortened) literal gloss (Roth 1997: 92):

- (12) [Main [I-1 If [I-2 after the sheep and goats come up from the common irrigated area [F-3 when the pennants [c-4 announcing the termination of pasturing] are wound around the main city gate]] the shepherd releases the sheep and goats into a field ...] the shepherd shall guard the field ...]

Another notable fact about (12) is that the lower clause I-2 in the double initial embedding contains a further final finite embedding F-3 which contains a further non-finite center-embedding c-4 (non-finiteness indicated by lower-case 'c' while upper case signals finiteness). The complex initial embeddings typical of law language ever since are likely to have their origin in Hammurabi's Code of Law (Simo Parpola, p.c.). Thus, the border of constraint I²max (1), still operational today, was reached already in Akkadian at the latest when the Akkadian tradition of writing was 500 years old.

5. Origin and development of C³max-w

As for the origin of multiple center-embedding, Thomsen (1984: 244) and Frayne (1993: 285) provide a most interesting Sumerian instance from c. 2000 BC of triple center-

embedding which Deutscher (2000: 234*** (check)) has analyzed in detail:

(13)

lú	Dabrum-ak-e	Utu-Hegal		Main
people	Dabrum.GEN.ERG	Utu-Hegal		
	bar			C-1
	account			
	lugal			C-2
	king			
	Enlil-e	á	šum-a	C-3
	Enlil.ERG	strength	give.SUBORD	
	i-me-a			
	P.is.SUBORD			
	i-zu-a-ak-eš(e)			
	PREF.knew.SUBORD.GEN-on			
Tirigan-ra	šu nu-ni-ba			
Tirigan.to	did not release			

'The people of Dabrum did not let Tirigan escape, because they knew that Utu-hegal was a king to whom Enlil had given strength.'

This means that C³max-w was reached already in Sumerian, at a point of time when written language had existed at most 1000 years.

Here is a Latin C³ from Cicero's *De haruspicum responsis*, pointed out by Nágelsbach (1846 [1963]: 645) to 'stretch the limits of Latin grammar to the extreme':

(14)

Postea	vos, patres conscripti, ...	M
then	you Senators	
	huic furiae,	C-1
	from this madman	
	si diutius in hac urbe,	C-2

if longer in this city
 quam delere I-4
 which destroy
 cuperet, C-3
 wanted
 maneret,
 would stay
 vox interdiceretur,
 vote cancel

decrevistis ...

decided

'After this you, Senators, decided to cancel this man's right to vote if he would stay longer in this city which he wanted to destroy.'

Sentence (14) is close to falsifying our empirical generalization C³max-w (3) because the C-3 of (14) contains a fourth level of embedding, *quam delere*. But this clause is initially embedded (I-4) in its matrix clause C-3.

A sum total of thirteen examples of C³ are at hand in our corpus, from all times. This tiny amount certainly indicates that the construction is a marginal one. Here is a C³ from the 20th century (pointed out by Geoffrey Sampson, p.c.):

- (15) [Main In an excellent article ... Salvini draws a parallel between the way [C-1 in which the spoken Latin of the men [C-2 with whom Gregory of Tours, [C-3 whom he has no reason [f-4 to mention,]] must have mixed] eventually became Old French ...,] and the comparable direct development of pre-Romanesque painting ...] (L. Thorpe, *Gregory of Tours: The History of the Franks*, Penguin Books, 1974: 39)

Even if the existence of C³ s is incontestable, it must be stressed that many of the attested examples (see Karlsson 2007c for the whole material) are so convoluted as to be almost incomprehensible. They overload the short-term memory resources by having

(too) many unresolved syntactic dependencies pending, and by stretching these unresolved dependencies over the introduction of (too) many new referents, which also taps available memory resources. In this regard, note that the deepest embedding C-3 of both (7) and (15) contains only pronouns referring to already introduced referents, thereby considerably easing the processing load.

Archaic Latin syntax before 300 BC was basically paratactic (Devoto 1968: 92). Clausal embedding below depth 1 was not established in Latin before 100 BC (Lindskog 1896). Relative center-embedding next to the antecedent was consolidated by Cicero who treated center-embedding from a stylistic and rhetorical point of view in *De oratore* (55 BC). If Cicero was the first Latin writer to use C³, the time extension from the earliest beginnings of Latin writing to (14) is some 500 years, a magnitude similar to that of reaching I²max in Akkadian and C³max-w in Sumerian.

By laying down rules for sentential composition Cicero completed the doctrine of PERIODS, a cornerstone of Western rhetoric and stylistics initiated by Aristotle, which has greatly affected the later development of sentence composition, especially the use of multiply center-embedded clauses, in Ancient Greek, Latin and later European languages. Here is a direct quotation from the relevant section in Aristotle's *Rhetoric*, cited after Fowler (1984):

(16) "By a period I mean a portion of speech that has in itself a beginning and an end, being at the same time not too big to be taken in at a glance. Language of this kind is satisfying and easy to follow. It is satisfying, because it is just the reverse of indefinite; and moreover, the hearer always feels that he is grasping something and has reached some definite conclusion; whereas it is unsatisfactory to see nothing in front of you and get nowhere. It is easy to follow, because it can easily be remembered; and this because language when in periodic form can be numbered, and number is the easiest of all things to remember. That is why verse, which is measured, is always more easily remembered than prose, which is not: the measures of verse can be numbered. The period must, further, not be completed until the sense is complete: it must not be capable of breaking off abruptly..."

A typical periodic sentence contains at least one center-embedding, a detour from the overriding sentence scheme, brought to structural completion and the satisfaction of 'full meaning' by the latter part of the main clause. The master of periodic sentences was Livy (59/64 BC-13AD), along with Cicero a stylistic icon for centuries. Cicero and Livy were the ones who set the model for the use of multiple center-embeddings in written language. W. Kriebel described the principles of their sentence composition already in his doctoral dissertation *Der Periodenbau bei Cicero und Livius*, defended at the University of Rostock in 1873. Here is a typical example of a C² by Livy:

- (17) [Main Nunc, [C-1 quod temeritatem temeritate alia luerit,][&C-1 tristioeremque rem, [C-2 quam necesse fuerit,] fecerit,] leniter castigat.] (Livy, *Ab urbe condita*; cf. Kriebel 1873: 15) '(M) Now, (C-1) as he [Masinissa] had paid one thoughtlessness with another, (&C-1) and worse the situation (C-2) than was necessary (&C-1) had made, (M) he [Scipio] mildly criticized [Masinissa]'.

Cicero's orations are known to have been carefully rehearsed and imprinted in long-term memory by ingenious retrieval techniques (van Dijk and Kintsch 1983: 368), thus differing from the conditions of natural everyday discourse. This stresses the special nature of these unusually complex constructions.

From the Renaissance to the 1900s there are hundreds of grammars, stylistic manuals, and scholarly monographs treating periodic sentence structure, e.g. Boivie's (1834: 99-102) Swedish and Becker's (1870: 418-23) German grammars. This is the historical source of the fairly uniform patterns of clausal subordination in present-day Standard Average European. In the words of Blatt (1957: 49): "Above all the limpidity and transparency of the Latin period was the big model."

Over the centuries one can spot occasional uses of C²s and every now and then of C³s in written European languages. Kracke (1911) studied the development of center-embedding in German texts from the 9th to the late 18th century, with no less than 58 authors in his corpus. He spotted twenty C²s, the earliest one in Heliand from around 830. Erman (1913: 43-44) found four C²s in his large Old High German corpus. Admoni

(1980) conducted an extensive study of the syntactic development of New High German literary language from the period 1470-1730. In his material of more than 500 particularly complex sentences there are one C³ and some fifty C²s.

In the Romance languages, the first complex periodic sentences did not appear until Dante in the 1300s (Blatt 1957: 37).

6. Origin and development of F⁵max-w

As for the emergence of multiple final embedding, the earliest relevant data known to us concern Ancient Greek. Webster (1941) studied the development of sentence complexity in Ancient Greek, especially of FE. F³ was used already by Homer around 700 BC but he has only one instance of it. This shallow syntactic depth of the *Iliad* and *Odyssey* is in good keeping with current theories of their origin in oral narrative. F⁴ was first used by Herodotus, Sophocles, and Thucydides around 450 BC, and F⁵ by Xenophon a century later. Demosthenes and Plato both have one F⁶ and, as an extreme, Dinarchus (360-292 BC) one F⁹. Overall, even F³ was rare in Ancient Greek (Schwyzer 1950: 710). One may conclude that F⁵max-w (10) of final embedding in written Ancient Greek was reached in less than 500 years.

Meadler (1905: 39) and Otto (1928: 810) found F⁶ in Classical Latin (Plautus, Cicero, Tacitus), just like in Classical Greek. Erman (1913) analyzed the embedding depths of 3302 Old High German sentences with at least two sub-clauses, demonstrating that F⁵max for written language had been reached and even surpassed in three instances:

depth	N
2	2747
3	478
4	70
5	4
6	2
7	1
sum	3302

Table 1. Frequency of clausal embedding depths (> 1) in Old High German (Erman 1913: 474).

Holm (1967) investigated the development of written style in Swedish. The medieval Icelandic Sagas contain sub-clauses, preferably in final position and normally just one. Due to Latin influence, Renaissance Swedish contained F⁴ and a growing share of center-embedding. The hundreds of texts from 1100-1960 analyzed by Holm contain one C², several instances of F⁴, a few of F⁵, and one of F⁶. In his history of Swedish, Bergman (1968: 123) gives as an extreme example of Latin-inspired periodic sentence structure one F⁷ from 1712.

Hiltunen (1990) studied the development of the syntactic structure of English law language from Anglo-Saxon in the 7th century, finding that sentences became more complex towards the 11th century. F⁵ occurred.

7. Conclusion

Evidence from many language families indicates that non-finite clausal subordination and initial stages of finite clausal subordination existed already in preliterate languages. The more complex forms of (especially finite) clausal embedding arose as part of large-

scale grammaticalization along with the advent of written language and the consequent gradual conventionalization of written registers (Givón 1979: chapter 5) . Of course there is a well-known underlying reason for the greater embedding complexity of written language: it enables more extensive and time-consuming processing in short-term memory of longer syntactic dependencies and more simultaneous discourse referents than in spoken language. This new property gradually pushed the complexity limits of clausal embedding to I^2_{max} , C^3_{max-w} and F^5_{max-w} which, as we have demonstrated, were reached already in the ancient languages and have not since been surpassed (including the flexibility of final embedding). The complexity maxima set by these three constraints were reached over a period of at most 500 years for I^2_{max} in Akkadian, for C^3_{max-w} in Latin, for F^5_{max-w} in Ancient Greek and Latin, and over at most 1000 years for C^3_{max-w} in Sumerian. Available data for older forms of English and German similarly indicate that it took some 500 years to reach these maxima. This seems to confirm Comrie's (1992: 202) hypothesis that the languages of 4,500 to 6,500 years ago might not have been noticeably different in complexity from those today, at least as regards this particular aspect of syntactic complexity.

It goes without saying that these developmental times limits are provisional and set the upper time limit based on currently available documentation.

Latin syntax was affected by Greek and both of these by ancient Near Eastern languages, especially Akkadian. Latin strongly influenced other written Western languages, especially as concerns periodic sentence structure, for more than 1500 years and imposed the same embedding limits - which, of course, must have common roots in biological properties of the human organism, especially restrictions on syntactic and discourse management in short-term memory.

It is a well-known fact that, mainly due to Latin influences, German and English were syntactically most complex in the 17th century and Swedish in the 19th century (Admoni 1974: 30-55, Jespersen 1967: 118, Lindstedt 1927, Engdahl 1962). After the downfall of the Latin-based school system around the year 1800 the patterns of clausal embedding have become successively simpler, especially in the 20th century, as a result of systematic language planning in communicative school teaching and phenomena like the Plain English movement.

But the upper limits of clausal embedding complexity have remained the same since the advent of written language.

References

Admoni, Wladimir G. (1974). *Die Entwicklungstendenzen des deutschen Satzbaus von heute. Linguistische Reihe*. Band 12. München: Max Hueber Verlag.

Admoni, Wladimir G. (1980). *Zur Ausbildung der Norm der deutschen Literatursprache im Bereich des neuhochdeutschen Satzgefüges (1470-1730). Ein Beitrag zur Geschichte des Gestaltungssystems der deutschen Sprache*. Bausteine zur Sprachgeschichte des Neuhochdeutschen 56/4. Berlin.

Barnes, Jonathan (1984). *The Complete works of Aristotle. The Revised Oxford Translation*. Vol. II. Bollingen Series LXXI: 2. Princeton, NJ: Princeton University Press.

Becker, K. F. (1870). *Ausführliche deutsche Grammatik als Kommentar der Schulgrammatik*. Zweiter Band. Zweite neubearbeitete Ausgabe. Prag: Verlag von Friedrich Tempsky.

Bergman, Gösta (1968). *Kortfattad svensk språkhistoria*. Stockholm: Bokförlaget Prisma.

Boivie, P. G. (1834). *Försök till en svensk språklära jemte inledning, innehållande allmänna grammatikan*. Omarbetad upplaga. Upsala: Palmblad & C.

Comrie, Bernard. (1992). Before complexity. In Hawkins, John A. and Gell-Mann,

Murray (ed.), *The evolution of human languages. Proceedings of the workshop on the evolution of human languages, held August, 1989 in Santa Fe, New Mexico*, 193-211. Redwood City, Calif.: Addison Wesley Publishing Company.

Deutscher, Guy (2000). *Syntactic change in Akkadian: the evolution of sentential complementation*. Oxford: Oxford University Press.

Deutscher, Guy (2002). The Akkadian relative clause in cross-linguistic perspective. *Zeitschrift für Assyriologie und vorderasiatische Archäologie* 92: 86-105.

Devoto, Giacomo (1968). *Geschichte der Sprache Roms*. Heidelberg: Carl Winter.

van Dijk, Teun A. and Kintsch, Walter (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Dryer, Matthew Syngé (1980). The positional tendencies of sentential noun phrases in universal grammar. *The Canadian Journal of Linguistics* 25 (2): 123-96.

Engdahl, Sven (1962). *Studier i nusvensk sakprosa. Några utvecklingslinjer*. Uppsala: Skrifter utgivna av Institutionen för nordiska språk vid Uppsala universitet, 11.

Erman, K. B. (1913). Beziehungen zwischen stellung und funktion der nebensätze mehrfacher unterordnung im ahd. *Zeitschrift für deutsche Philologie* 45.1-46, 153-216, 426-84.

Fowler, R. L. (1984). Aristotle on the period. *Classical Quarterly* 32: 89-99.

Frayne, Douglas R. (1993). *The Royal Inscriptions of Mesopotamia — Early Periods. Volume 2. Sargonic and Gutian Periods (2334-2213 BC)*. Toronto, Buffalo and London: University of Toronto Press.

Givón, T. 1979. *On understanding grammar*. New York: Academic Press.

Hakulinen, Auli, Karlsson, Fred and Vilkuna, Maria (1980). *Suomen tekstilauseiden piirteitä: kvantitatiivinen tutkimus*. Helsinki: Publications of the Department of General Linguistics, University of Helsinki, No. 6.

Harris, Martin. 1988. Concessive clauses in English and Romance. In Haiman, John and Thompson, Sandra A. (ed.), *Clause combining in grammar and discourse*, 71-99. Amsterdam: John Benjamins B.V.

Hiltunen, Risto (1990). *Chapters on legal English. Aspects past and present of the language of the law*. Annales Academiæ Scientiarum Fennicæ B 251. Helsinki: Suomalainen Tiedeakatemia [Academia Scientiarum Fennica].

Holm, Gösta (1967). *Epoker och prosastilar. Några linjer och punkter i den svenska prosans stilhistoria*. Lundastudier i nordisk språkvetenskap A:17. Lund: Studentlitteratur.

Itkonen, Erkki. (1966). *Kieli ja sen tutkimus*. Helsinki: Werner Söderström.

Kalmár, Ivan (1985). Are there really no primitive languages? In Olson, David R., Nancy Torrance and Angela Hildyard (ed.), *Literacy, language and learning. The nature and consequences of reading and writing*, 148-166. Cambridge: Cambridge University Press.

Karlsson, Fred (2007a). Constraints on multiple initial embedding of clauses. *International Journal of Corpus Linguistics* 12: 107-118.

Karlsson, Fred (2007b). Constraints on multiple center-embedding of clauses. *Journal*

of Linguistics 43 (2): 365-392.

Karlsson, Fred (2007c). Genuine data on multiple center-embedding of clauses.
http://www.ling.helsinki.fi/~fkarlss/ceb_data.pdf [downloaded September 3, 2007].

Karlsson, Fred (2008). Syntactic recursion and other types of sequential arrangement.
In print for *The Linguistic Review*.

Karlsson, Fred and Sinnemäki, Kaius (to appear) . Constraints on multiple final
embedding of clauses.

König, Ekkehard and van der Auwera, Johan (1988). Clause integration in German and
Dutch. In Haiman, John and Thompson, Sandra A. (ed.), *Clause combining in
grammar and discourse*, 101-133. Amsterdam: John Benjamins B.V.

Kracke, Otto (1911). Die Entwicklung der Mittelstellung des deutschen Nebensatzes.
Giessen, Ludwigs-Universität PhD dissertation. Darmstadt: G. Ottos Hofbuchdruckerei.

Kriebel, W. (1873). *Der Periodenbau bei Cicero und Livius*. Rostock, Universität
Rostock PhD dissertation. Prenzlau: A. Mieck.

Lehmann, Winfred P. (1974). *Proto-Indo-European syntax*. Austin and London:
University of Texas Press.

Leino, Pentti. 1975. Äidinkieli ja vieras kieli: rahvaanrunouden metriikkaa.
Kalevalaseuran vuosikirja 55: 25-48.

Lindskog, Claes. 1896. Beiträge zur Geschichte der Satzstellung im Latein. Lund: *Acta
Universitatis Lundensis* 32: 1-60.

Lindstedt, Torvald (1927). Tendenser i nusvensk meningsbildning. *Nysvenska studier*
7: 1-29.

Lord, Albert B. 2000 [1960]. *The singer of tales*. Second edition. Cambridge, Mass.: Harvard University Press.

Meador, Clarence Linton (1905). Types of sentence structure in Latin prose writers. *Transactions and Proceedings of the American Philological Association* 36: 32-51.

Mithun, Marianne (1984). How to avoid subordination. In Claudia Brugman and Monica Macaulay (ed.), *Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society*, 493-509. Berkeley, Calif.: Berkeley Linguistics Society.

Otto, Walter (1928). *Handbuch der Altertumswissenschaft*. II:2. München: C. H. Beck'sche Verlagsbuchhandlung.

Nägelsbach, Karl Friedrich (1846 [1963]). *Lateinische Stilistik*. Darmstadt: Wissenschaftliche Buchgesellschaft.

O'Neil, W. (1976). Clause adjunction in Old English. *General Linguistics* 17: 199-211.

Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey and Svartvik, Jan (1989). *A comprehensive grammar of the English language*. Seventh (corrected) impression. London and New York: Longman.

Roth, Martha T. (1997). *Law collections from Mesopotamia and Asia Minor*. Writings from the Ancient World Society of Biblical Literature - WAWSBL, 6. 2nd edition. Atlanta: Scholars Press.

Schwyzler, Eduard (1950). *Griechische Grammatik auf der Grundlage von Karl Brugmanns Griechischer Grammatik*. Zweiter Band. *Syntax und syntaktische Stilistik*. München: C. H. Beck'sche Verlagsbuchhandlung.

Thomsen, Marie-Louise (1984). *The Sumerian language*. Copenhagen: Akademisk

forlag.

Webster, T. B. L. (1941). A study of Greek sentence construction. *American Journal of Philology* 62: 385-415.

1. Valuable help and advice has been provided by Guy Deutscher, John W. Du Bois, Simo Parpola and Geoffrey Sampson.