

Syntactic recursion and iteration

Fred Karlsson

1 Introduction¹

The nature and origin of syntactic recursion in natural languages is a topical problem. Important recent contributions include those of Johansson (2005), Parker (2006), Tomalin (2006; 2007), and Heine and Kuteva (2007). Syntactic recursion will here be discussed especially in relation to its cognate concept of iteration. Their basic common feature is plain structural repetition: “keep on emitting instances of the current structure, or stop”. Their main difference is that recursion builds structure by increasing embedding depth whereas iteration yields flat output structures which do not increase depth.

My focus here is on the types of recursion and iteration, and on what empirically determinable constraints there are on the number of recursive and iterative cycles of application.

Recursion comes in two subtypes, nested recursion (= center-embedding) and tail-recursion, the latter covering left-recursion and right-recursion. There are six functionally different types of iteration: structural iteration, apposition, reduplication, repetition, listing and succession.

It will be empirically shown that multiple nested syntactic recursion of degrees greater than 3 does not exist in written language, neither in sentences nor in noun phrases or prepositional phrases. In practice, even nesting of degree 2 is extremely rare in writing. In speech, nested recursion at depths greater than 1 is practically non-existing, thus partly confirming an early hypothesis of Reich (1969). Left-branching tail-recursion of clauses is strictly constrained to maximally two recursive cycles. Right-branching clausal tail-recursion rarely transcends three cycles in spoken language and five in written language. On constituent level both left- and right-branching is less constrained (especially in written language), but e.g. left-branching genitives rarely recurse more than two cycles ([[[*Pam*'s] *mum*'s] *baggage*]).

2 Types of sequential arrangement

There are many types of rules and restrictions on the sequential arrangements of words, constituents and clauses. Most important in any language are the principles of *basic linearization* of obligatory elements. These are word order restrictions and tendencies governing the arrangements of grammatical elements and simple canonical constituents, clauses and larger structures, e.g. (for English typical pragmatically unmarked declarative sentences):

- a grammatical subject precedes its main verb,
- a grammatical object follows its main verb,
- a subcategorized locative adverbial follows its main verb.

This obligatory basic linearization of word classes, constituents, heads and dependents, grammatical functions etc. defines the co-occurrence and word-order properties of the grammatical elements and constructions. This core grammar is the foundation of the optional complexity-introducing types of sequential arrangement such as iteration and recursion. Every natural language has basic linearization principles which also determine its classification in word order typologies. These rules are mostly local and simple precedence statements.

A special type of basic linearization is that of *lexicalized multiword expressions*: phraseology

(1) and lexical bundles (2), i.e. expressions frequent in a certain register (Biber et al. 1992: 993):

(1) *you know*

(2) *I was going to*

Every language has lexicalized multiword expressions, at least lexical bundles. Most languages have semantically non-compositional idioms and phraseology (however, Pirahã is claimed by Dan Everett, p.c., to have no idioms nor lexical bundles).

Minimal canonical structures may be optionally elaborated by *free modification*, e.g. by adding adjectival or genitival premodifiers to nouns, intensifiers as premodifiers to adjectives, free adjuncts to verbs, and non-complement clauses subordinate clauses to main clauses:

- a determiner precedes its nominal head,
- a determiner precedes an adjectival modifier of the same head,
- an adjectival modifier precedes its nominal head,
- a case-marked genitival NP precedes its nominal head,
- an *if*-clause prefers initial embedding position,
- a *when*-clause may be embedded initially or finally.

There are also *information-packaging* principles such as inversion, preposing, postposing, extraposition, and “fat things last”, which interact with the principles of basic linearization and free modification, adapting the orders of canonical arrangements to fit textual and pragmatic needs. These three mechanisms together define grammatically well-formed and situationally appropriate sentences such as (3).

(3) *Nat’s fat cat sat on the mat when it spat out a rat.*

Recursion and *iteration* are also types of sequential arrangement, but layered on top of basic linearization, free modification and information-packaging by invoking repetition: “keep on emitting instances of the current structure, or stop”. Their main difference is that recursion builds structure by increasing embedding depth whereas iteration yields flat output structures, repetitive sequences on the same depth level as the first instance.

Recursion and iteration are the main topics of this paper and they will here be analyzed in conformance with Wilhelm von Humboldt’s insightful early agenda:

- (4) “Ist aber auch die [grammatische] Freiheit an sich unbestimmbar und unerklärlich, so lassen sich dennoch vielleicht ihre Grenzen innerhalb eines gewissen ihr allein gewährten Spielraum auffinden; und die Sprachuntersuchung muß die Erscheinung der Freiheit erkennen und ehren, aber auch *sorgfältig ihren Grenzen nachspüren.*” (von Humboldt 1836: 81; italics added/FK)
‘Even if [grammatical] freedom in itself would be impossible to determine and explain, it might still perhaps be possible to find its limits within the confines of a certain latitude reserved for it, and language study must recognize and respect the phenomenon of freedom, but also *carefully try to determine its limits.*’

In this spirit, it is relevant to ask both what types of recursion and iteration there are, and what empirically determinable constraints (if any) there are on the number of repetitive cycles in both recursion and iteration. I first turn to iteration.

3 Six types of iteration

On both formal and on functional (partially semantic, partially psycholinguistic) grounds, six types of

iterative sequential arrangement can be distinguished: 1) structural iteration, 2) apposition, 3) reduplication, 4) repetition, 5) listing, and 6) succession. They all concatenate elements without additional depth-increasing structure building. The latter feature distinguishes them from recursion, which creates hierarchic structures governing the linear strings it produces, typically by repeatedly subordinating an element to another instance of the same type.

3.1 Structural iteration

The most important type of iteration is structural iteration which is overwhelmingly most frequent. It has no clear semantic task and this too sets it off from the other five types of iterative sequential arrangement which are functionally motivated special cases.

The main manifestation of structural iteration is coordination, with or without explicit conjunctions. The gist of coordination and all other types of structural iteration is “optionally add an instance of the same structural type as the current one, with (syndetically) or without (asyndetically) a coordinating conjunction”. Prototypical structural iteration works much like the Kleene star, “*”, in formal language theory.

All major constituent types (NP, AP, PP etc.), grammatical functions (subject, object, adverbial etc.) and clause types (relative clauses, *if*-clauses, infinitive clauses etc.) may be iteratively coordinated, on the same level of structure. Also many free modifiers may be iteratively coordinated, e.g. adjectival modifiers (but not determiners) and case-inflected genitival NPs as premodifiers of nominal heads. It is thus an inherent property of grammatical elements and construction types whether they are iterative, recursive or neither or both.

By far the two most important and frequent iterative structures are coordination of clauses in sentences and of NPs in clauses. There are no bounds on the number of constituent-level conjuncts, or on how many coordinated clauses there may be in a single sentence. For example, in Hermann Broch’s (1961: 138–144) novel *Der Tod des Vergil* there is a 1077-word sentence extending over six pages, with 81 finite and 38 non-finite clauses, and several hundred instances of coordination on modifier, constituent and clause level.

By contrast, observe that depth-increasing recursion with multitudes (e.g. 100, or even 10) of successively embedded subordinate clauses never is the central mechanism underlying such unusually long sentences. Thus, the maximal clausal embedding depth of Broch’s gigantic sentence is a modest 3, and this is reached only once. In this sentence there are only two instances of subordinate clauses at depth 2.

To pick another example, consider the famous chapter of *Ulysses* where Joyce describes the stream of consciousness of Molly Bloom in eight huge paragraphs without punctuation. One sentence is 12,931 words long. The first ten pages of this chapter (Joyce 1934: 723–732; 5,500 words) contain one instance of recursive final (right-branching) clausal embedding reaching depth 6 (“F” = final, the numbers as in F-2 indicate embedding depth):

(5) [... *she said yes* [_{F-1} *because it grigged her* [_{F-2} *because she knew* [_{F-3} *what it meant* [_{F-4} *because I used* [_{F-5} *to tell her a good bit of* [_{F-6} *what went on between us ...*]]]]]]]] (Joyce 1934 [1922]: 729)

But such depths are very untypical. The abnormal length of Joyce’s 12,931-word sentence is caused by extensive use of coordination. One more indication of the prevalence of iterated coordination is the fact that the conjunction *and* or its equivalent is one of the most common words in many languages.

Clause chaining is another variant of iterative clause linkage, closely related to asyndetic coordination. Likewise, certain types of verb serialization are basically iterative in nature (Sebba 1987: 111).

3.2 Apposition

Appositions are typically iterated NPs on the same level of depth. Their defining feature is mutual coreference. This property sets appositions off from coordination as a special, semantically motivated type of iteration. (6) is a genuine fivefold Swedish appositive structure (coreferential nominal heads in boldface).

(6) *Ej långt därifrån sitter ett **föremål** för mängdens kärleksfulla hyllning,*
not far away sits an object for crowd's loving ovation

*Henrik Gabriel **Porthan***

Henrik Gabriel Porthan,

*akademiens yppersta **lärare**,*

Academy's most magnificent teacher

*den fosterländska **häftaforskningen** **fader**,*

the national historical research's father

*en **man** som gaf sitt namn åt ett helt tidevarfi vår kulturhistoria.*

a man who gave his name to a whole epoch in our cultural history

'Not far away sits the **object** of the loving ovation of the crowd, Henrik Gabriel **Porthan**, the most magnificent **teacher** of the Academy, the **founder** of national history, a **man** who gave his name to a whole epoch in the history of our culture.'

The number of cycles in iterated apposition is unconstrained but numbers in excess of a handful are rare. This property too sets apposition off from structural iteration for which the number of cycles is truly unconstrained and easy to document by concrete examples.

3.3 Reduplication

Syntactic reduplication, often called repetition, is a motivated type of iteration, where a word (normally a content word) is repeated in order to express meaning modulations such as intensification, augmentation, repetition, diminution, or iterative or continuative action. Usually the construction is asyndetic, but it may be syndetic as well, especially with adverbs, comparatives, and verbs. Here are some originally Swedish examples from Lindström (1999), in direct translations all of which qualify equally as idiomatic English (and all of which can be found on the internet):

(7) a. *There were **cars cars cars**.*

b. *a **long long** way*

c. ***small small** creatures*

d. ***much much** better*

e. *played **better and better***

f. *came **closer and closer***

g. *They **walked and walked and walked**.*

In his corpus-based study of syntactic reduplication in 1.7 million words of spoken and written Swedish, Lindström (1999) noted the predominance of reduplicated pairs with two identical terms. Triples are rare, quadruples extremely rare. Therefore reduplication is indeed the most apt term for this type of iteration. The incidence of asyndetic reduplication (without conjunctions) in Swedish is one per 3000 words, higher in spoken than in written language, and of syndetic reduplication one per 5000 words. Of 620 Swedish instances of reduplication only five were triples. These are clear preferences.

On the other hand, they do not strictly rule out longer chains (which are easy to attest e.g. by googling).

3.4 Repetition, listing, succession

The iterative concept of plain repetition (or repeat, in the terminology of Biber et al. 1999: 1055) concerns the special, psycholinguistically motivated case of speech planning. The speaker repeats (often involuntarily) some word or syntagm, especially words belonging to closed grammatical classes, in order to gain time to complete planning the message. The outcomes of such repetitions are formally ungrammatical.

- (8) a. *Yeah, it, it, it is it's it's it's it's good.* (Biber et al. 1999: 1055)
b. *And and / if if if you know they had none at all.* (British National Corpus, FMD 321)

The corpus data of Biber et al. show that the likelihood of the repetition decreases sharply with the number of words repeated. The vast majority of examples contain a single repeat. There are extremely few instances of three or more repeats.

Listing (enumeration) is an iterative mechanism predominantly used in restricted lexical taxonomies, see (9a) and (9b):

- (9) a. *Smith, Brown, Jones ...*
b. *Monday Tuesday Wednesday ...*
c. *one two three ...*

A special kind of listing is succession (9c), which underlies numerosity (Heine and Kuteva 2007: 271). Given any number n , the successor function enables production of $n + 1$. Repetition, listing and succession of course have no limits.

4 Syntactic recursion: historical and conceptual background

Heine and Kuteva (2007: 264–266) discuss several important conceptual distinctions relating to recursion: direct recursion (modeled by rules like $A \rightarrow A B$) vs. indirect recursion ($A \rightarrow B, B \rightarrow A$), counting recursion (yielding AABB) vs. mirror recursion (yielding ABBA), and simple recursion (with only once application cycle) vs. productive recursion (with an unlimited number of cycles). The difference between *nested recursion* and *tail-recursion* is the most important one from the viewpoint of the present paper. Nested recursion and tail-recursion have often been presumed to be unrestricted (or productive in Heine and Kuteva's terminology). For tail-recursion this is by and large true, but in Humboldt's spirit (1) it will be demonstrated below that there are quantitative constraints especially on nested syntactic recursion of clauses, noun phrases and prepositional phrases.

The use of nested and tail-recursive rules in grammatical description is ancient because such rules were formulated already around 400 BC, in Pāṇini's Sanskrit grammar (Bhate and Kak 1993: 79) and in Tholkaappiyar's Tamil grammar (Baladunarman, Ishwar, and Kumar 2003).

In recent linguistic theory recursion was reintroduced around 1950, probably first by Bar-Hillel (1953), under influence from the theory of recursive functions in mathematical logic (Tomalin 2006: 64). In Chomsky's PhD dissertation (1975 [1955-1956]) recursion is discussed in several connections, e.g. as the source of sentential infinity and as for its proper locus in the grammar. However, Chomsky's (1956: 65; 1957: 21–23) original published claims concerning the alleged unconstrainedness of nested recursion in English syntax are sketchy and rely only on examples invented by Chomsky himself.

It is surprising to note that prior to the current millennium the deeper nature and the various types of recursion in natural languages have not been discussed at great length in any brand of

grammatical theory (however, see Halliday 1963 for an early exception). Parker (2006) surveys the prevailing conceptual and terminological variation. Table 1 provides an overview of some relevant concepts and terminology in a mainly generative and computational framework.

Table 1. Basic conceptualization and terminology of recursion.

Type of rewrite rule ($\alpha, \beta \neq 0$)	Mode (orientation) of recursion	Relation to iteration	Linguistic term
$A \rightarrow A \alpha$	left-recursion	tail-recursion = convertible to iteration ²	left-branching = initial embedding
$A \rightarrow \alpha A$	right-recursion	tail-recursion = convertible to iteration	right-branching = final embedding
$A \rightarrow \alpha A \beta$	nested recursion	full-blown recursion, not convertible to iteration	nesting, center-embedding

A standard view of recursion in the formal sciences is the top-down one that a function is recursive if it is defined in terms of successively simpler versions of itself such that a termination condition (variably called the baseline, base case, or degenerate case) guarantees the process to halt. A recursive function is not well-formed if its termination is not guaranteed (Brookshear 2007: 246). A well-formed recursive procedure is designed to test for the termination condition before requesting further cycles of activation. Material introduced by any lower recursive cycle is always contained (embedded) in the material introduced by the immediately higher cycle. Every applicative cycle increases the recursive depth by 1. Thus, the depth of sentence (5) is 6.

The typical generative-linguistic conception of recursion on the contrary emphasizes that syntactic recursion lacks termination conditions: this is the very property that guarantees the infinity of sentence length and structural elaboration.

Nested recursion (center-embedding) is recursion par excellence. A nested structure X^n is embedded within a structure of the same type, X^{n-1} , with non-null material (excluding subordinating conjunctions, cf. endnote 3) of X^{n-1} both to the left and to the right of X^n . Nested recursion consumes memory resources and presupposes storage of pending material and return addresses on a stack in order to secure proper unwinding. This sketchy description is equally true of human language processing and its counterparts in formal language theory and computational linguistics.

Nested recursion of precisely the same type of element (e.g. two relative clauses, two *if*-clauses, two NPs) is called *self-embedding*. Formally, left-recursion and right-recursion can always be converted to iteration, i.e. to repetition (one more cycle) of the same procedure, without anything pending on the stack (Aho, Sethi, and Ullman 1986: 53).

Structures like “*if if if S₁, then S₂, then S₃, then S₄*”, i.e. counting recursion (Christiansen and Chater 1999: 168) were introduced into the public theoretical discussion already by Chomsky (1956: 65) and claimed to be unrestricted as for recursive depth. For reasons mentioned in endnote 3, such sentences are here regarded as initially-embedded, not as center-embedded. Karlsson (2007a) demonstrated that there is a strict empirically determinable quantitative constraint on repeated initial embedding of clauses, viz. (10) to be discussed shortly.

A final note is in order concerning the ontological nature of recursion. Halliday (1963: 85–13) is clear on interpreting recursion as an integral part of language structure. A structural element is repeated in depth, a series of such elements thus forming a progression. This ontological commitment also is at the heart of generative grammar, where recursion has always been regarded as *the* very structural feature that caters for the presumed infinity of language (i.e., a category on the left-hand side of a rewrite rule reoccurs on the right-hand side, cf. Table 1).

But there are other views. For Linell (2005: 75), in spoken interaction long utterances with multiple embeddings are emergent results of local decisions of moving from one turn-constructive unit to another. They are not part of the grammar per se. The grammar contains just a core syntax plus methods for moving to the next turn-constructive unit. Similarly, Harder (2006) treats recursion as a semantic and functional option of use, a possibility allowed by the whole system, arising as a more or

less epiphenomenal consequence or side effect when a functional upgrading operation such as reusing a *that*-clause after reporting verbs happens to concern a structural type that is already present in the input. Clausal recursion as such is not built into the fabric of grammar.

Heine and Kuteva (2007: 265) opt for an intermediate agnostic position. According to them, recursion is not a property of language but rather the product of a theory designed to describe language structure: "... we are not maintaining that language, or language structure, has recursion but rather that there are certain properties to be observed in language that can appropriately be described in terms of a construct [of formal recursion]".

These views need not be fully contradictory. Properties of languages can be described by formal mechanisms without necessarily renouncing the ultimate functional motivation of the properties. Allowing multiple (even if not fully unconstrained) right-recursion in core grammar seems to be more or less equivalent to Linell's basic embedding patterns + methods for moving to the next turn-constructional unit.

On the other hand, every theoretical concept is a hypothesis about a phenomenon in some more or less real world and the question whether the concept properly represents the "God's-truth Ding an sich", or is a plain "Hocus Pocus construct", in the sense of Householder's (1952) famous and much-debated distinction, can be left open for philosophical discussion.

5 Constraints on clausal recursion

Clausal embedding recursion means repeated embedding of clauses in initial, central or final position of the superordinate clause. Initially-embedded clauses (IEs) have either nothing to their left, or a subordinating conjunction of the superordinate clause.³ Center-embedded clauses (CEs) have superordinate clause constituents to their left (other than subordinating conjunctions) 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.

Karlsson (2007a; 2007b) has empirically established *constraints*, quantitative limits and qualitative restrictions, on clausal embedding recursion 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 100+ corpus-based syntactic and stylistic descriptions of European languages, especially Latin and older variants of German.

Here (10-14) are some constraints operational in many European languages (at least in Danish, English, Greek, Finnish, Latin and Swedish). "I" stands for initial clausal embedding, "C" for clausal center-embedding, and the raised exponent expresses maximal degree of embedding of a sentence, e.g. I² is double initial embedding. Expressions like C-2 indicate type and embedding depth of individual clauses. "-w" is short for "written language" and "-s" for "spoken language".

- (10) I²max: the maximal degree of initial clausal embedding is 2; 100 instances found, as in (15).
- (11) Qualitative I²-constraints: Double initial clausal 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. See (15).
- (12) C³max-w: in written language, the maximal (very rare) degree of multiple clausal center-embedding is 3; thirteen instances retrieved, (16).
- (13) C²max-s: in spoken language, the maximal (very rare) degree of multiple clausal center-embedding is 2; only two instances retrieved, (17).

- (14) Only-postmodifying-self-embedding: only postmodifying (i.e. typically relative) clauses allow self-embedding (16, 17).
- (15) [[_{I-1} If [_{I-2} what he saw through security] did not impress him] Tammuz ...] (BNC)
- (16) a. *Der Ritter von Malzahn, dem der Junker sich als einen Fremden, der bei seiner Durchreise den seltsamen Mann, den er mit sich führe, im Augenschein zu nehmen wünschte, vorstellte, nötigte ihn ...*

b. <i>Der Ritter von Malzahn,</i>	Main
the rider from Malzahn	
<i>dem der Junker sich als einen Fremden</i>	C-1
whom the Junker himself as a stranger	
<i>der bei seiner Durchreise den seltsamen Mann,</i>	C-2
who upon his journey the unusual man	
<i>den er mit sich führe,</i>	C-3
whom he with himself would bring	
<i>in Augenschein zu nehmen wünschte,</i>	C-2
by appearance to judge wanted	
<i>vorstellte,</i>	C-1
introduced	
<i>nötigte ihn ...</i>	Main
forced him ...	

'The rider from Malzahn, to whom the Junker introduced himself as a stranger, who upon his journey wanted to judge by appearance the unusual man whom he would bring with himself, forced him ...' (Heinrich von Kleist, Michael Kohlhaas)

- (17) [A lot of the housing [_{C-1} that the people [_{C-2} that worked in New Haven] lived in] was back that way.] (Reich and Dell 1976)

No genuine triple initial embeddings nor quadruple center-embeddings are on record, suggesting that $I^2\text{max}$ and $C^3\text{max-w}$ are firm restrictions on respectively left-branching clausal tail-recursion and nested clausal recursion. Only some 130 double center-embeddings were retrieved among the tens of millions of words checked. 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 at least in the European languages here considered, if "fully recursive" is taken to mean that ordinary language use would provide instances of multiple initial clausal embedding or nesting degrees beyond the confines of $I^2\text{max}$ and $C^3\text{max-w}$.

Of special importance is the fact that multiple nested (= fully productive) clausal recursion (i.e. multiple center-embedding) is practically absent from ordinary spoken language, as stipulated by constraint (13) $C^2\text{max-s}$. Multiple nesting cannot therefore reasonably be considered a central design feature of language, as claimed by Hauser, Chomsky, and Fitch (2002).⁴ On the contrary, empirical facts prove that Reich (1969) and Reich and Schütze (1991) were on the right track in claiming that the limit on clausal nesting is 1. This is practically true for spoken language, whereas written language allows two cycles and extremely marginally three.

$C^3\text{max-w}$ and Only-postmodifying-self-embedding taken together license triple relative self-embedding, as in (16).

Final clausal embedding is less constrained but not totally unregimented. In an extensive empirical study Karlsson (ms.) found evidence in several European languages for a tendency called $F^3\text{max}$:

quarry		
<i>beschäftigt</i>	<i>gewesenen Arbeiter</i>	PP-1c; NP-1c
employed	have been workers	
<i>vorgeschossenen</i>	<i>Arbeitslöhne Ersatz</i>	zu leisten.
prepaid	salaries	remuneration to pay
		PP-0; NP-0

'May the court rule that the defendant remunerate me for the salaries prepaid by me on his behalf to the workers that were employed in the quarry that he had taken over to run.'

Sentence (20) involves clauses that have been nominalized to premodifying adjectival participles and which carry along their original complements as dependents of the participle in the form of PPs with their requisite NPs. In (20) the PP-0 and NP-0 (*für die ... Arbeitslöhne* '(for) the ... salaries' at basic level (depth 0) are interrupted by a nominalization (*von mir ... vorgeschossenen* 'by me ... prepaid'), where the PP-1c and NP-1c (at the same depth as PP-1a, NP1-a and PP-1b, NP-1b) (*an die ... Arbeiter* '(for) the ... workers' (nested PP and NP recursion, degree 1) are interrupted by another nominalization (*in dem ... gewesenenen* 'in the ... been'). This nominalization contains the further nestings PP-2 and NP-2 (*in dem ... Steinbrüche* '(in) the ... quarry' (nested PP and NP recursion, degree 2), interrupted by a third nominalization (*von ihm ... übernommenen* 'by him ... overtaken'), which contains no less than two more nested PPs and NPs, at depth 3, PP-3a and NP-3a *von ihm* 'by him' and PP-3b and NP-3b *zur Bearbeitung* 'to run'.

Sentence (21), from *Kaseler Zeitung* 1.5.1889, also manifests triple nested recursion of prepositional phrases and noun phrases. As for NPs, the nesting sequence is [*die ... [der ... [das ... [dem Kloster] Hospiz] Insel] Badereisenden*].

(21) a. *Für die nach der durch das von dem Kloster Loccum erbaute Hospitz in weiteren Kreisen bekannt gewordenen Insel Langeoog kommenden Badereisenden hat sich...*

b.	<i>Für die</i>	PP-0; NP-0
	For the	
	<i>nach der</i>	PP-1; NP-1
	after the	
	<i>durch das</i>	PP-2; NP-2
	through the	
	<i>von dem Kloster Loccum</i>	PP-3; NP-3
	by the monastery Loccum	
	<i>erbaute Hospitz</i>	PP-2; NP-2
	built hospital	
	<i>in weiteren Kreisen bekannt gewordenen Insel Langeoog</i>	PP-1; NP-1
	in larger circles known become island Langeoog	
	<i>kommenden Badereisenden hat sich ...</i>	PP-0; NP-0
	coming seaside guests has REFL	

'For the seaside guests coming to the island Langeoog which has become known to larger circles through the hospital built by the Loccum monastery there has ...'

Of course triple center-embedding of NPs and PPs as in (20) and (21) is extremely convoluted and rare and confined to restricted, formal, stylistically marked, older (obsolete?) genres of writing. It must be emphasized that no genuine examples of triple nesting of PPs and NPs are at hand from the twentieth (not to speak of the twenty-first) century. Even double center-embedding of NPs and PPs is utterly rare so one more genuine example deserves presentation. Here (22) is an Ancient Greek NP from Plato's *The Sophist*.⁵

- (22) a. *tà gâr tês tôn pollôn psychês ómmata*
- b.
- | | | |
|--|-------------------|------|
| <i>tà</i> | <i>gâr</i> | NP-0 |
| DET | because | |
| | <i>tês</i> | NP-1 |
| | DET | |
| | <i>tôn pollôn</i> | NP-2 |
| | DET of many | |
| | <i>psychês</i> | NP-1 |
| | soul's | |
| <i>ómmata</i> | | NP-0 |
| eyes | | |
| 'because the eyes of the soul of many' | | |

Insofar as this small corpus (20, 21, 22) can be trusted, the favourite locus of multiple center-embedding of NPs and PPs seems to be after the determiner in the NP. Two quantitative constraints (where again “-w” is short for “written language”) can be postulated at least for several European languages:

- (23) NP³max-w: In written language, the maximal (very rare) degree of multiple center-embedding of NPs is 3.
- (24) PP³max-w: In written language, the maximal (very rare) degree of multiple center-embedding of PPs is 3.

Multiple nesting of PPs presupposes the same amount of nested NPs. PPs and NPs thus form a peculiar tightly bonded recursive pair of layered nestings. No genuine examples are at hand of nested NP³s (with nouns or pronouns as headwords) that would not involve at least one PP.

What about the present-day limits on multiple NP- and PP-nesting, especially in spoken language? The weirdness of the above genuine obsolete German examples from the nineteenth century, and the total absence of similar genuine contemporary examples, certainly indicates that the current limit is below 3. On the other hand it is clear that at least NP-nesting at depth 1 is a living option for example in Finnish (25).

- (25) a. *tämä edellisen hallituksen ehdotus*⁶
- b.
- | | |
|--|------|
| <i>tämä</i> | NP-0 |
| this.DET.NOM.SG | |
| <i>edellisen</i> | NP-1 |
| previous.A+GEN.SG | |
| <i>hallituksen</i> | NP-1 |
| government.N+GEN.SG | |
| <i>ehdotus</i> | NP-0 |
| proposal.N.NOM.SG | |
| 'this proposal by the previous government' | |

Native intuitions tell that NP-1 in (25) could be interrupted by another short NP inflected in the genitive case and still remain grammatical and acceptable:

- (26) a. *tämä edellisen pääministeri Lipposen hallituksen ehdotus*

b.	<i>tämä</i>	NP-0
	this.DET.NOM.SG	
	<i>edellise+n</i>	NP-1
	previous.A+GEN.SG	
	<i>pääministeri</i>	NP-2
	Prime Minister.N.NOM.SG	
	<i>Lippose+n</i>	NP-2
	Lipponen.N+GEN.SG	
	<i>hallituksen</i>	NP-1
	government.N+GEN.SG	
	<i>ehdotus</i>	NP-0
	proposal.N.NOM.SG	
	'this proposal by the previous government (led by) Prime Minister Lipponen'	

As a tentative generalization I therefore submit (27), where “-s” is short for “spoken language”:

- (27) NP²max-s: in spoken language, the maximal (very rare) degree of multiple center-embedding of NPs is 2.

Conclusive resolution of this issue, including the limits of PP-nesting in speech, must await additional detailed corpus study of a sufficient number of languages.

As for *left-recursion* on constituent level, prepositional phrases by their very nature do not have it. NPs are left-recursive, in the form of repeated genitives, where the previous genitive modifies either the subsequent one or the combination of this and the head noun – this is *submodification* and *stacking*, respectively, in the terminology of Huddleston and Pullum (2003: 446–447). This feature distinguishes recursive genitives from iterated phrases, which are structurally parallel. The Brown, LOB and ICE-GB corpora (3 million words) show that submodifying genitives are rare, brief, and semantically restricted (28a, 28b, 28c). Only some 20 instances were found of submodifying genitive chains of depth 2. Stacked genitives are even rarer.

- (28) a. *Hilary's lawyer's secretary*
 b. *Burnside's horse's belly*
 c. *Pam's mum's baggage*

Left-recursive open-endedness on constituent level emerges in formal written genres of Finnish where there is a productive construction based on nominalization of verbs and multiple NPs inflected in the genitive or oblique cases. As in (20) and (21), the nominalized verbs act as premodifying adjectival participles in NPs, with their original complements as properly case-inflected dependents of the participle. The postpositional phrase in (29) comes from a regulation adopted by the Finnish government. There are five left-recursive NPs inflected in the genitive case, several of which have adjectivized verbs with their own left-branching case-inflected NPs as dependents or premodifiers, creating an overall depth at word level of 11.

- (29) a. *valtion virkamieheltä virallisesti vaadittavasta kielitaidosta annetun lain täytäntöönpanosta annetun asetuksen 7 pykälän 1 momentin mukaan*
- b.
- | | |
|--|-------------------------|
| | <i>valtio+n</i> |
| | state+GEN.SG |
| | <i>virkamiehe+ltä</i> |
| | official+ABL.SG |
| | <i>vaadi+tta+va+sta</i> |

require+PSS+PCP+ELA.SG
kielitaido+sta
 language proficiency+ELA.SG
anne+tu+n
 give+PSS.PCP+GEN.SG
lai+n
 law+GEN.SG
täytäntöönpano+sta
 implementation+ELA.SG
anne+tu+n
 give+PSS.PCP+GEN.SG
asetukse+n
 regulation+GEN.SG
 7
 seventh+GEN.SG
pykälä+n
 paragraph+GEN.SG
 1
 first+GEN.SG
momenti+n
 subsection+GEN.SG
mukaan
 according_to
 ‘according to subsection 1 of paragraph 7 in the regulation which was issued concerning the implementation of the law which was issued concerning the language proficiency which is required of a state official’

Right-recursion of phrase-level constituents is most productive in the interplay between multiple PPs and NPs contained in one another in postmodifying structures such as (30). No clear upper limits are in evidence but recursive PPs in excess of degree 5 are rare. However, the complex (genuine) Swedish NP (31), again from a regulation issued by the Finnish government, contains no less than ten right-recursive PPs, labelled a, b, c ... in (31b), and headed by the prepositions *om* ‘concerning’, *av* ‘of’, and *i* ‘in’.

(30) [in the soil [on the slopes [of the rock [by the river [in ...]]]]]

(31) a. *statsrådets förordning om ändring av ikraftträdandebestämmelsen i statsrådets förordning om ändring av ikraftträdandebestämmelsen i statsrådets förordning om ändring av statsrådets förordning om utveckling av landsbygden*⁷

b.
statsråd+et+s
 government+DEF+GEN.SG
förordning
 regulation
om ändring (a)
 concerning change
av ikraftträdandebestämmelse+n (b)
 of implementation decision+DEF.NOM.SG
i (c)
 in

statsråd+et+s
 government+DEF+GEN.SG
förordning
 regulation
om ändring (d)
 concerning change
av ikraftträdandebestämmelse+n (e)
 of implementation decision+DEF.NOM.SG
i (f)
 in
statsråd+et+s
 government+DEF+GEN.SG
förordning
 regulation
om ändring (g)
 concerning change
av (h)
 of
statsråd+et+s
 government+DEF+GEN.SG
förordning
 regulation
om utveckling (i)
 concerning development
av landsbygd+en (j)
 of rural area+DEF.NOM.SG

‘the government’s regulation concerning change of the implementation decision in the government’s regulation concerning change of the implementation decision in the government’s regulation concerning change of the government’s regulation concerning the development of rural areas’⁸

7 Conclusion

No evidence for nested syntactic recursion of degrees greater than 3 is at hand, neither on clause level nor on phrase level. Two interesting questions cannot be addressed here: Why is the maximal degree 3 and not 2, 7 or some other number? Why is the maximal degree the same on phrase and clause level? Motivated answers to these questions would presuppose careful consideration of referent processing and short-term memory capacity restrictions which transcend the scope of this contribution.

The extant examples of nesting of degree 3 and even those of degree 2 are so few, convoluted, and almost exclusively confined to written language as to practically falsify the hypothesis of the importance of unlimited syntactic nesting (multiple center-embedding), thus also confirming the experimental results of Christiansen and Chater (1999) achieved by connectionist modeling concerning the marginality of multiple center-embedding.

No more than two spoken instances of double clausal center-embedding have been documented (Karlsson 2007b). Double NP and PP center-embedding have, to the best of my knowledge, not been recorded in speech at all. Thus, in speech, nested recursion in the domains of clauses and constituents is for all practical purposes confined to just one cycle, degree 1, which is what Heine and Kuteva (2007: 268) call simple (as opposed to productive) recursion.

Karlsson (2009) demonstrates in philological detail that multiple nesting arose along with the advent of written language. This is also compliant with known psychological facts: multiple nesting

taps short-term memory resources and such resources are more readily available in writing and reading than in real-time speaking and listening.

Multiple tail-recursion is real both in its rightwards and leftwards manifestations especially in written language, but not fully unconstrained, as shown above e.g. for multiple left-branching of clauses ($I^2\text{max}$) and for genitives (triple genitive chains are extremely rare). One more interesting question will be left open here as for its proper explanation: Why is clausal left-branching strictly constrained to two cycles, but left-branching of certain premodifiers (such as nominalized participle constructions) is not?

In the invitation to the April 2007 recursion conference organized by Dan Everett at the University of Northern Illinois, a complex sentence by the late Clifford Geertz was singled out in the call for papers, with the implication that this sentence (32a) is particularly relevant for the conference topic. (32b) is the precise analysis of this sentence in terms of clause structure (C = finite center-embedded, F = finite finally-embedded, f = non-finite finally-embedded).

- (32) a. *The notion that the essence of what it means to be human is most clearly revealed in those features of human culture that are universal rather than in those that are distinctive to this people or that is a prejudice that we are not obliged to share...*
- b.
- | | |
|--|------|
| <i>The notion</i> | Main |
| <i>that the essence</i> | C-1 |
| <i>of what it means</i> | C-2 |
| <i>to be human</i> | f-3 |
| <i>is most clearly revealed in those features of human culture</i> | |
| <i>that are universal</i> | F-2 |
| <i>rather than in those</i> | |
| <i>that are distinctive to this people or that</i> | F-2 |
| <i>is a prejudice</i> | |
| <i>that we are not obliged</i> | F-1 |
| <i>to share...</i> | f-2 |

The analysis shows that (32b) complies with all constraints on clausal embedding outlined in Section 5, in particular with the constraints $C^3\text{max-w}$, $F^{3-5}\text{max}$ and Only-postmodifying-self-embedding. There is a double center-embedding in (32b), where the lower center-embedding C-2 contains a further non-finite final embedding f-3 (the infinitive construction *to be human*). The upper center-embedding C-1 contains two more final embeddings F-2 + F-2. The predicate complement *a prejudice* contains two more final embeddings, F-1 + f-2. No NPs nor PPs in (32b) violate $NP^3\text{max-w}$, $PP^3\text{max-w}$, nor $NP^2\text{max-s}$.

Thus, Geertz's sentence is structurally interesting, but fully compliant with all the constraints on clausal and phrasal recursion outlined in this paper.

Notes

1. I thank three anonymous referees for valuable comments and suggestions for improvement.
2. Tail-recursion is always convertible to simple iteration (Aho, Sethi, and Ullman 1986: 53).
3. Note that I follow Quirk et al. (1989: 1037) in interpreting clauses like I-2 in sentence (15), embedded immediately after a subordinating or coordinating conjunction, as *initially-embedded* rather than as center-embedded. The main argument is that subordinating and coordinating conjunctions are not syntactically as tightly integrated and real constituents in their clauses as ordinary full constituents (and relative pronouns) are.
4. This is further support for Pinker and Jackendoff's (2005) criticism of Hauser, Chomsky, and Fitch's (2002) claim.

5. This example was helpfully provided by Martti Leiwo.
6. The NP (25) can be verified by googling on the internet.
7. The source of (31a) is *Asetus/Förordning 778/2004* and the Finnish original is: *Valtioneuvoston asetus voimaantulosäännöksen muuttamisesta valtioneuvoston asetuksessa maaseudun kehittämistä annetun valtioneuvoston asetuksen muuttamisesta annetun valtioneuvoston asetuksen voimaantulosäännöksen muuttamisesta.*
8. One referee aptly remarks that the English translation of this NP "is completely Martian, and nothing like anything that would ever be produced by even the most constipated English-speaking bureaucrat. It sounds like an insane utterance!"

References

- Aho, Alfred V., Ravi Sethi, and Jeffrey D. Ullman
1986 *Compilers. Principles, Techniques, and Tools*. Reading, Mass.: Addison-Wesley Publishing Company.
- Baladunarman, I., S. Ishwar, and R. Sanjeeth Kumar
2003 Context-free grammar for natural language constructs. An implementation for Venpa class of Tamil poetry. In: *The Sixth International Tamil Internet Conference and Exhibition With the Theme "IT for Education in Tamil"*. Available at http://www.infitt.org/ti2003/papers/19_raman.pdf.
- Bar-Hillel, Yehoshua
1953 On recursive definitions in empirical science. *11th International Congress of Philosophy* 5: 160–165. Amsterdam: North-Holland.
- Bhate, Saroja, and Subhash Kak
1993 Pāṇini's Grammar and Computer Science. *Annals of the Bhandarkar Oriental Research Institute* 72: 79–94. Also available at <http://www.ece.lsu.edu/kak/bhate.pdf>.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan
1999 *Longman Grammar of Spoken and Written English*. London: Longman.
- Brookshear, J. Glenn
2007 *Computer Science. An Overview*. Ninth edition. Boston etc.: Pearson Addison Wesley.
- Broch, Hermann
1961 *Der Tod des Vergil*. Darmstadt: Moderner Buch-Club.
- Chomsky, Noam
1956 On the limits of finite-state description. *MIT Research Laboratory of Electronics, Quarterly Progress Report* 41: 64–65.
- Chomsky, Noam
1957 *Syntactic Structures*. The Hague: Mouton.
- Chomsky, Noam
1975 [1955-1956] *The Logical Structure of Linguistic Theory*. New York and London: Plenum Press.
- Christiansen, M. H., and N. Chater
1999 Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science* 23: 157–205.
- Engel, Eduard
1922 *Deutsche Stilkunst*. 30th edition. Vienna & Leipzig: Hölder-Pichler-Tempsky A-G.

- Halliday, M.A.K.
1963 Class in relation to the axes of chain and choice in language. *Journal of Linguistics* 2: 5–15.
- Harder, Peter
2006 Recursion in a functional-semantic grammar. In: Timo Haukioja (ed.), *Papers on Language Theory. Publications in General Linguistics* 10: 67–105. Turku: University of Turku.
- Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch
2002 The faculty of language: What is it, who has it, and how did it evolve? *Science* 298: 1569–1579.
- Heine, Bernd, and Tania Kuteva
2007 *The Genesis of Grammar. A Reconstruction*. Oxford: Oxford University Press.
- Householder Jr., Fred W.
1952 Review of Harris, Zellig S., 1951, *Methods in Structural Linguistics*. *International Journal of American Linguistics* 28: 260–268.
- Huddleston, Rodney, and Geoffrey K. Pullum
2003 *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Humboldt, Wilhelm von
1836 *Über die Verschiedenheit des menschlichen Sprachbaues und ihre Einfluss auf die geistige Entwicklung des Menschengeschlechtes*. Berlin: Druckerei der Königlichen Akademie der Wissenschaften.
- Johansson, Sverker
2005 *Origins of Language. Constraints on Hypotheses*. Amsterdam: John Benjamins Publishing Company.
- Joyce, James
1934 [1922] *Ulysses*. New York: Random House Inc.
- Karlsson, Fred
2007a Constraints on multiple initial embedding of clauses. *International Journal of Corpus Linguistics* 12 (1): 107–118.
2007b Constraints on multiple center-embedding of clauses. *Journal of Linguistics* 43 (2): 365–392.
2009 Origin and maintenance of clausal embedding complexity. In: Geoffrey Sampson, David Gil, and Peter Trudgill (eds.), *Language Complexity as an Evolving Variable*, 192–202. Oxford: Oxford University Press.
(ms.) Multiple final embedding of clauses. University of Helsinki: Department of General Linguistics.
- Lindström, Jan
1999 *Vackert, vackert! Syntaktisk reduplikation i svenskan*. Studier i nordisk filologi 77. Helsingfors: Svenska Litteratursällskapet.
- Linell, Per
2005 *The Written Language Bias in Linguistics. Its Nature, Origins and Transformations*. London: Routledge.
- Parker, Anne R.
2006 Evolving the narrow language faculty: was recursion the pivotal step? In: Angelo Cangelosi, Andrew D. M. Smith, and Kenny Smith (eds.), *The Evolution of Language: Proceedings of the 6th International Conference (Evolang 6)*, 239–246. London: World Scientific.
- Pinker, Steven and Ray Jackendoff

- 2005 The faculty of language: What's special about it?. *Cognition* 95: 201–236
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik
1972 *A Grammar of Contemporary English*. London: Longman.
- Reich, Peter A.
1969 The finiteness of natural language. *Language* 45: 831–843. Reprinted in Fred Householder (ed.), *Syntactic Theory I. Structuralist*, 238–272. Harmondsworth: Penguin.
- Reich, Peter A., and Gary S. Dell
1977 Finiteness and embedding. In: R. J. Di Pietro and E. L. Blansett Jr. (eds.), *The Third LACUS Forum 1976*, 438–447. Columbia, S.C.: Hornbeam Press.
- Reich, Peter A., and Carson T. Schütze
1991 Syntactic embedding: What can people really do? In: Barbara Brunson and Tom Wilson (eds.), *Toronto Working Papers in Linguistics* 11 (1) – 11 (2): 91–97.
- Sebba, Mark
1987 *The Syntax of Serial Verbs. An Investigation into Serialisation in Sranan and Other Languages*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Tomalin, Marcus
2006 *Linguistics and the Formal Sciences. The Origins of Generative Grammar*. Cambridge: Cambridge University Press.
- 2007 Reconsidering recursion in syntactic theory. *Lingua* 117: 1784–1800.