

Kaius Sinnemäki

On Halliday's Distinction between Embedded and Hypotactic Clauses

Abstract

Functional Grammar, as presented in Halliday (1994), divides subordinate clauses in embedded and hypotactic clauses. Hypotaxis is characterized as being more like parataxis than like embedding. In this short paper, I identify embedding and hypotaxis according to Halliday's criteria and test their behavior in written Finnish. My aim is to determine whether hypotaxis behaves like embedding or like parataxis with respect to maximally complex clauses. The results show that all subordinate clauses are constrained in much the same way and thus behave alike in this respect. As a by-product, I also discuss the nature of maximally long coordinated clauses.

1. Introduction

At the time of writing my master's thesis on complex clauses under Prof. Fred Karlsson's supervision, I got interested in the way Systemic Functional Grammar (Halliday 1994) divides subordinate clauses in hypotactic and embedded clauses.¹ I wanted to know more about the characteristics of these two clause-types and why some linguists have felt the distinction necessary.² I understood that the distinction was mainly semantic in nature, but somehow I was not convinced that it was justified.

¹ I am grateful to Prof. Fred Karlsson, Dr. Kari Pitkänen, Dr. Matti Miestamo, and two anonymous referees of JL for comments on earlier versions of this paper which I presented at the *Symposium of Syntactic Functions – Focus on the Periphery*, organized by the Linguistic Association of Finland (SKY) in November 14–15th 2003. This short paper has been written as part of a project Natural Language Complexity at the University of Helsinki, sponsored by the Finnish Academy, whose grant is gratefully acknowledged.

² The distinction to these two clause-types is not critical to the Systemic Functional Grammar as a whole but it is rather central to the analysis of syntactic organization in the theory. In fact, this distinction has aroused a heated discussion within the theory,

In a nutshell, **hypotaxis** is characterized as a relationship between clauses, an interdependency relationship in which neither of the clauses is a constituent part of another, whereas **embedding** is a nominalizing device, a mechanism of rankshift in which a clause comes to function as a constituent part of another (Halliday 1987: 73, 1994: 242). For example, the subordinate clauses in the sentences *I couldn't come because I was sick* and *John thought that Alice wouldn't come* are in a hypotactic relation to the main clause; they are dependent on it but not constituents of it. Embedded clauses, on the other hand, have undergone a shift in rank from clause level to group level; they function as groups or parts of groups,³ in the following functions only (Halliday 1994: 242):

- head of a nominal group (e.g. *It is obvious that Jeff wrecked the car*)
- post-modifier in a nominal group (*I saw the car that Jeff wrecked*)
- post-modifier in an adverbial group (*He came earlier than we had expected*)

In less theoretical parts of his scholarly writings, Halliday makes two rather strong claims about the characteristics of hypotaxis and embedding. Halliday (1987: 74) asserts that “[H]ypotaxis is more like parataxis than it is like embedding; and both are characteristic of spoken rather than written language” whereas Halliday (1985: 84) states that “[I]n mathematical terms, the hypotactic relation is one of iteration, whereas embedding is one of recursion.” These characterizations are not central to the distinction between the clause-types. However, they provide us with interesting viewpoints which are not found in Halliday (1994) but which are well worth commenting. In the following, I show that complex clauses in written Finnish provide evidence against Halliday’s claims. I admit my data is limited in scope, especially since it does not contain material from spoken language, but I believe the data shows clearly enough that hypotactic clauses are not as different from embedded clauses as Halliday claims.

resulting in two opposing camps: “Sydney Grammar” holds on to the distinction whereas Cardiff Grammar rejects it altogether (e.g. Butler 1993: 259–292).

³ In Systemic Functional Grammar group is a technical term for a clausal element which is similar but slightly different from that of *phrase* in Constituent Grammar.

2. Preliminary observations and the hypothesis

Considering Halliday's first claim, I think he possibly wanted to emphasize that hypotactic relationship does not involve rankshift of the clause. In that sense, hypotactic relation resembles paratactic relation between clauses. Earlier corpus-based studies also support his conclusion that written English is marked by complexity in the nominal group whereas spoken English is marked by complexity in the clause complex (e.g. Ellegård 1978). However, I am amazed that Halliday's examples from written and spoken language are not always genuine instances but his own rewordings.⁴ Moreover, he occasionally bases his calculations of lexical density and number of clauses, which are used to enhance the point that embedding characterizes written language and hypotaxis characterizes spoken language, partly on these fabricated examples. Therefore, the calculations do not convincingly reflect real differences in spoken and written modes but rather Halliday's rewordings and intuitions about the differences between the two modes. There is a danger that his intuitions have been a priori influenced by too broad generalizations. In this article, I will not pay more attention to his claims on spoken vs. written language but will concentrate on the claim that hypotaxis is more like parataxis than embedding because it resembles his second claim.

I'm not quite sure how I should understand Halliday's second statement, especially concerning the difference between recursion and iteration. However, I assume that for Halliday iteration is potentially unlimited concatenation at the same rank level. Thus, a chain of clauses in which all clauses are either paratactic or hypotactic should be able to continue endlessly. However, he also sees recursion as unlimited (Halliday 1987: 73), which is not helpful at all, making the allusion to iteration vs. recursion seem vacuous. The only difference between recursion and iteration would be that clauses are concatenated at different rank level in recursion but at the same rank level in iteration. From a mathematical point of view both subordination and coordination are unlimited, but earlier corpus based studies (e.g. Karlsson 2004, Sinnemäki 2004) have shown that people hardly ever use sentences in which the descending level of subordination in a chain of subordinate clauses is more than -4 (when main

⁴ This is illustrated by a quote from Halliday (1987: 61) "2B is my attempt at a somewhat less "written" version; while 2C is another step nearer to speech."

clause is 0). As for coordinated clauses, it is easier to see that they might be concatenated unlimitedly.

I assume that Halliday's claims may be taken as predictions on text occurrences, although this might be a wrong assumption. Yet, I see no other way for evaluating the behavior of hypotactic clauses in actual language use. In order to do this, I combine parts of the two statements in my hypothesis which I will test with data from written Finnish: in maximally complex clauses, hypotactic clauses behave more like paratactic clauses and not like embedded clauses. What I mean by maximally complex is the possible upper limit of consecutively connected finite clauses, whether embedded, hypotactic or paratactic.

3. Methods and data

In order to test the hypothesis about the nature of hypotaxis, I specified two reference points as operational definitions: one for embedding and the other for parataxis. Firstly, I defined the reference point for parataxis as the chain of coordinated and juxtaposed main clauses (Halliday's 1994 criteria for parataxis). I call this a paratactic chain. As an example, the length of the paratactic chain in the sentence *Jack ate cake, Mary washed the dishes, but Mike slept in his bed* is three (clauses). In order to find maximally long paratactic chains, I consulted a machine-readable corpus of the Finnish Language Bank. The sample corpus for this analysis consisted of roughly 23 million words of machine-readable newspapers, magazines and prose in Finnish. I carried out this analysis by first marking the coordinating conjunctions automatically and then printing a list of instances with five or more conjunctions in journalistic writing and those with ten or more in prose. Finally, I analyzed these instances manually (roughly 35 pages) to detect both asyndetic and syndetic coordination.⁵ Secondly, I defined the reference point for embedding as the chain of clauses that according to Halliday's (1994) criteria were embedded. This is called an embedding chain. For example in (1) (from *Suomen kuvalehti* 1987), all subordinate clauses are embedded.

Having set the reference points, I chose the sequence of hypotactic clauses (according to Halliday's criteria), a hypotactic chain, as the variable

⁵ The automatic analysis detected only syndetic coordination but the manual analysis included asyndetic coordination as well. The result was thus slightly suboptimal but does not affect the main conclusions.

whose values I compared to the reference-points. Sentence (2) (from Lander 1997) is an example of a hypotactic chain, in which all subordinate clauses are in a hypotactic relationship to the matrix clause. Finally, I also checked the maximal values of mixed chains which included both hypotactic and embedded clauses.

(1)	<p>Sattui, 'It happened' <i>että minulle kertoi kokemuksistaan eräs työtovereistani,</i> 'that to me told of his experiences one of my colleagues' <i>joka äskettäin oli johtanut valtuuskuntaa Suomeen ja oli saanut sen kunnianosoituksen,</i> 'who recently had lead a delegation to Finland and had received the mark of honor' <i>että tasavallan presidentti otti valtuuskunnan vastaan.</i> 'that the president of the republic received the delegation.'</p>	<p style="text-align: right;">M</p> <p style="text-align: right;">nominal group (fact) -1</p> <p style="text-align: right;">postmodifier of nominal group -2</p> <p style="text-align: right;">postmodifier of nominal group -3</p>
(2)	<p><i>Ymmärrän kyllä,</i> 'I well understand' <i>että filmi saattoi olla pelottava,</i> 'that the film might have been scary' <i>jos ei tiedä,</i> 'if one doesn't know' <i>mistä on kysymys.</i> 'what it's about'</p>	<p style="text-align: right;">M</p> <p style="text-align: right;">Projection -1</p> <p style="text-align: right;">Enhancement/ adverbial projection -2</p> <p style="text-align: right;">projection -3</p>

For the analysis of the embedding and hypotactic chains, I used a sample of two million words, a sub-corpus of the larger 23 million word sample. This analysis was a part of my master's thesis (Sinnemäki 2004) in which I studied the constraints on right-branching clauses in Finnish. Firstly, I marked all the subordinating conjunctions and relative pronouns automatically. Secondly, I printed a list of those with three or more marks (roughly 120 pages) which I then manually analyzed in order to detect genuine instances of increasing chain length. As the two examples above illustrate, the longest chains of embedding and hypotaxis were typically right-branching clauses.

4. Results of the chain-analyses

Since individual idiosyncrasies surface easier in prose than in journalistic writing, I have divided my results into these two genre categories in the corpus. Table 1 presents the results of the analyses.

	journalistic writing	prose
embedding chain	3	3
hypotactic chain	4	5
mixed chain	4	7
paratactic chain	17	132

Table 1. Maximal lengths of the chains

The maximal length of the embedding chain was 3 clauses in both journalistic writing and prose. This was a rather clear upper limit, although there were only twelve instances of embedding chains that were 3 clauses long.⁶ Paratactic chains, on the other hand, do not seem to obey any limit, since the lengths do not cluster around certain values. However, we must see behind the figures in more detail with the help of table 2.

	5–7	8–12	13–17	18–26	42–132
magazines	7	0	1	0	0
newspapers	5	0	1	0	0
prose	12	9	4	8	3

Table 2. Distribution of paratactic chains in the corpora

The maximal length of the paratactic chain in newspapers was 17 clauses. However, this instance was a long poem which is clearly a quotation from another genre. In addition, there were no paratactic chains whose length varied between 8 and 16 clauses. In magazines, the longest paratactic chain was 13 clauses, but there were no paratactic chains whose length varied between 8 and 12 clauses. Thus, although the maximal length

⁶ The percentual share of embedding chains whose length is three clauses was 2.7% of all mixed chains whose length was three clauses. The corresponding share for hypotactic chains was 35.1%. This indicates that hypotactic clauses—or at least their sequences—are more frequent than embedded clauses not only in spoken language but also in written language.

of the paratactic chains were around 15 clauses in journalistic writing, there were only 2 paratactic chains longer than 7 clauses.

In prose, there were eight paratactic chains whose lengths varied between 18 and 26 clauses and altogether 24 paratactic chains longer than 7 clauses. In addition, I found one chain with 42, another with 102, and a third with 132 clauses, but all of them were by the same author. The longest chain was a sentence that formed a whole chapter in *Tuulikaappimaa*, a novel by Jari Tervo (1997), known for his tendency to employ stream-of-consciousness. If we focus solely on the longest paratactic chains, 17 clauses was the longest chain in journalistic writing whereas 132 clauses was the longest chain in prose. However, if we discard single oddities and instead focus on tendencies, the maximal length of the paratactic chain in journalistic writing tends to be around 7 clauses, while in prose it is somewhere around 25 clauses.

We may now compare the length of the hypotactic chains to our two reference points. As the figures in table 1 show, the maximal length of the hypotactic chain was four clauses in journalistic writing and five clauses in prose. Thus, with hypotaxis the chains were one to two clauses longer than with embedding; the differences are very small. Even when the distinction was overlooked in the case of mixed chains, the length of the chain did not approach the maximal lengths of the paratactic chains. Moreover, all the longest hypotactic and mixed chains in prose occurred in *Alastalon salissa*,⁷ a novel by Volter Kilpi (1933), who employs techniques from stream-of-consciousness style. If this novel were excluded from the sample, the maximal lengths of the hypotactic and mixed chains in prose would be identical with those of journalistic writing (four and four). All in all, the paratactic chains in journalistic writing were roughly two times longer, and in prose roughly five times longer than hypotactic chains.

5. Discussion

My analysis of the sample data shows that there are good grounds for rejecting the original hypothesis—at least as far as it concerns written Finnish—namely: in maximally complex clauses hypotaxis does not behave like parataxis but more like embedding. Neither clause-type formed longer chains than 3–4 clauses whereas paratactic chains were unlimited. In

⁷ The first 240 pages of this novel were analyzed for this paper.

relation to the feature of being constrained, all finite subordinate clauses in written Finnish seem to behave in a closely related way.⁸ Of course it is possible that subordinate clauses in written language are affected by prescriptive norms of writing. However, the maximal lengths of mixed chains in spoken language in Ikola & al. (1989) were seven clauses, but they consider chains longer than four clauses possible coding errors of the analyst. Their results thus confirm my observations.

No upper limit seems to exist for parataxis. What I mean by “upper limit” (or being unlimited) for paratactic chains is that their maximal values did not cluster around certain values as clearly as those for subordinate clauses. Nevertheless, I’m not sure whether “unlimited” should be interpreted here as “infinite”. It may be argued that such interpretations are handy theoretical hyperboles for approximating natural language with artificial languages (Tero Kainlauri, p.c.). Of course speakers or writers may produce sentences with more than 132 clauses in a paratactic chain. Yet, infinitely long sentences exist as theoretical possibilities but their actual instantiations are impossible: *infinite* is different from *much*.

It is also highly interesting to consider why maximally long paratactic chains tended to cluster around 7 clauses in journalistic writing and 25 clauses in prose. For journalistic writing there are obvious limitations imposed by space and the norms of news discourse. In general, the longer the sentence the more it begins to resemble a paragraph. *Oxford English Dictionary Online* defines paragraph as a thematic entity that is set apart from what stand before or after it. Obviously, thematic entities are preferably built up of small distinct information chunks rather than one huge chunk. Cutting the information in smaller pieces enables the hearer to decode the thematic content more efficiently. I think this tendency is exploited in artistic writing, such as *Tuulikaappimaa* and *Alastalon salissa*. They bend the norms of standard writing by bending the extent of sentence towards that of standard paragraph in order to gain text strategic effects.⁹ This is especially obvious as we consider the fact that the longest mixed chains in *Alastalon salissa* are close to the longest paratactic chains in journalistic writing.

⁸ In his study on clausal embedding complexity, Karlsson (2004) shows that the inclusion of non-finite clauses typically adds only one clause to the maximally long chains of subordinate clauses.

⁹ The norms of typical writing are bent e.g. by employing deviate punctuation (see Sinnemäki 2004: 77).

As a conclusion, finite subordinate clauses do not form endlessly long chains in written Finnish. But as it seems, neither do coordinate clauses, although they are not constrained by as clear limit as subordinate clause are. My results raise a very interesting question for grammarians, psycholinguists and discourse analysts alike, namely: what are the factors underlying the greater difficulty of long mixed chains compared to long paratactic chains. I assume the answers lie in the grammatical (in)dependency of these clauses as well as in their information conveying functions in discourse. However, this remains an open question for future research.

Data sources

Finnish Language Bank. A sample of roughly 23 million words of Finnish electronic documents. Consisted of samples from seven newspapers (*Keskisuomalainen, Helsingin Sanomat, Aamulehti, Demari, Kauppalehti, Iltalehti, Karjalainen*), three magazines (*Tekniikan maailma, Kauppalehti-Optio, Suomen Kuvalehti*), and 82 novels and text books by various authors, published by WSOY, Otava, University of Helsinki, Finnish Literature Society, and Edita. Compiled by the Department of General Linguistics, University of Helsinki, Research Institute for the Languages of Finland [KOTUS] and CSC – Center for Scientific Computing, Finland. Available at URL: <http://www.csc.fi/kielipankki/>. Managed by the Center for Scientific Computing.

Kilpi, Volter (1933) *Alastalon salissa: kuvaus saaristosta*. Helsinki: Otava.

Lander, Leena (1997) *Iloisen kotiinpaluun asuinsijat*. Porvoo, Helsinki & Juva: Werner Söderström Osakeyhtiö.

Oxford English Dictionary (2005) Oxford: Oxford University Press. Available at URL: <http://dictionary.oed.com/entrance.dtl> (4.11.2005).

Suomen Kuvalehti (1987) A sample of roughly 0.5 million words of machine readable magazine articles in Finnish. Compiled by the Department of General Linguistics, University of Helsinki, Research Institute for the Languages of Finland [KOTUS] and CSC – Center for Scientific Computing, Finland. Available at URL: <http://www.csc.fi/kielipankki/>.

Tervo, Jari (1997) *Tuulikaappimaa*. Porvoo, Helsinki & Juva: Werner Söderström Osakeyhtiö.

References

- Butler, Christopher S. (2003) *Structure and Function: A Guide to Three Major Structural-Functional Theories. Part 2: From clause to discourse and beyond*. Studies in Languages Companion Series 64. Amsterdam & Philadelphia, PA: John Benjamins.
- Ellegård, Alvar (1978) *The Syntactic Structure of English Texts: A computer-based study of four kinds of text in the Brown University corpus*. Gothenburg Studies in English 43. Gothenburg: Acta Universitatis Gothoburgensis.
- Halliday, M.A.K. (1985) *Spoken and Written Language*. Victoria: Deakin University Press.
- (1987) Spoken and written modes of meaning. In Rosalind Horowitz & S. Jay Samuels (eds.) *Comprehending Oral and Written Language*, pp. 55–87. San Diego, CA: Academic Press.
- (1994) *An Introduction to Functional Grammar*. 2nd edition. London: Edward Arnold.
- Ikola, Osmo, Ulla Palomäki & Anna-Kaisa Koitto (1989) *Suomen murteiden lauseoppia ja tekstikielioppia* [’Syntax and text grammar in Finnish dialects’]. Suomalaisen Kirjallisuuden Seuran Toimituksia 511. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Karlsson, Fred (2004) *Constraints on Clausal Embedding Complexity in Standard Average European*. Unpublished manuscript, Department of General Linguistics, University of Helsinki.
- Sinnemäki, Kaius (2004) *Complex Right-Branching Clauses*. Master’s thesis, Department of General Linguistics, University of Helsinki.

Contact information:

Kaius Sinnemäki
Department of General Linguistics
P.O. Box 9 (Siltavuorenpenger 20A)
FI-00014 University of Helsinki
ksinnema(at)ling(dot)helsinki(dot)fi