

EVA EJERHED

A Swedish Clause Grammar and Its Implementation

Abstract

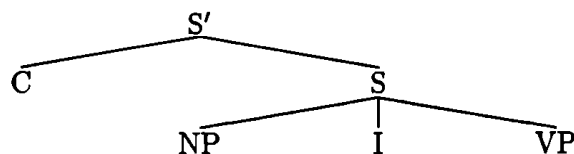
The paper is concerned with the notion of clause as a basic, minimal unit for the segmentation and processing of natural language. The first part of the paper surveys various criteria for clausehood that have been proposed in theoretical linguistics and computational linguistics, and proposes that a clause in English or Swedish or any other natural language can be defined in structural terms at the surface level as a regular expression of syntactic categories, equivalently, as a set of sequences of word classes, a possibility which has been explicitly denied by Harris (1968) and later transformational grammarians. The second part of the paper presents a grammar for Swedish clauses, and a newspaper text segmented into clauses by an experimental clause parser intended for a speech synthesis application. The third part of the paper presents some phonetic data concerning the distribution of perceived pauses (Strangert and Zhi 1989, Strangert 1989) and intonation units (Huber 1988) in relation to clause units.

1 What is a Clause in Linguistic Theory?

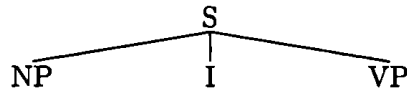
In traditional grammar a clause is defined as a unit consisting of a subject and a predicate. The terms *suppositum* and *appositum* were used in scholastic grammar to denote the syntactic functions of these two basic parts of a clause. Traditional grammar makes a distinction between main clauses and dependent clauses.

In current transformational grammar as presented by Radford (1988), three types of clauses are recognized (see (1)).

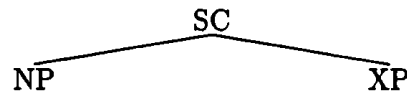
(1) (a) Ordinary Clauses



(b) Exceptional Clauses

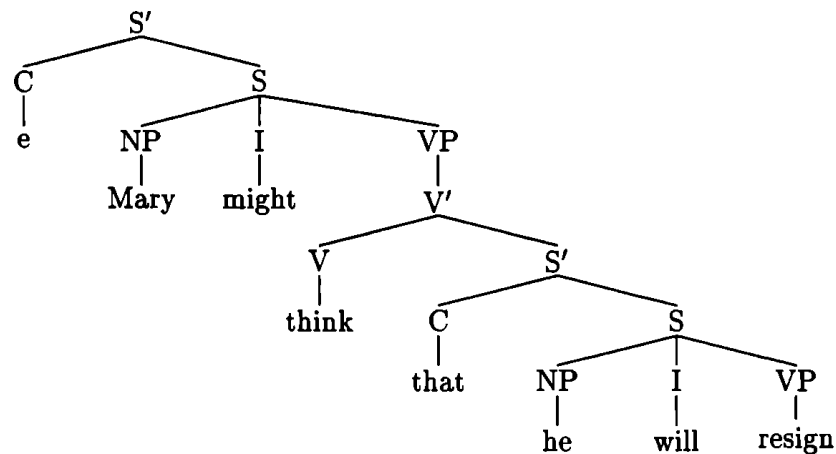


(c) Small Clauses

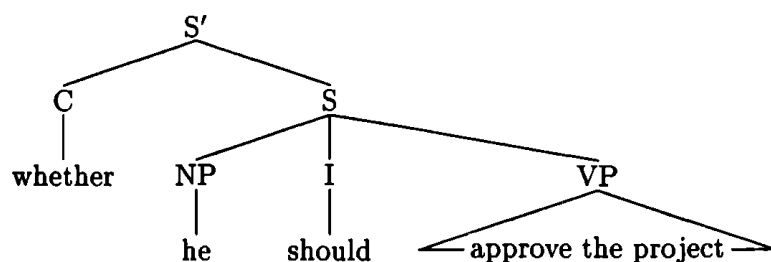


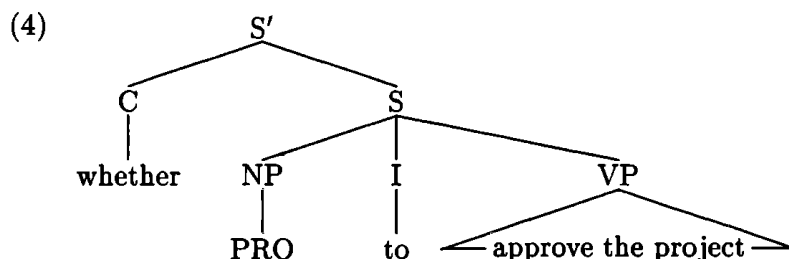
According to Radford (1988) “the three Clause types differ principally in that Ordinary Clauses contain both I and C, Exceptional clauses contain I (=infinitival to) but not C, and Small Clauses contain neither C nor I. Moreover, both Exceptional Clauses and Small Clauses are highly restricted in their distribution: for example, Exceptional Clauses typically occur only as the Complements of certain specific types of verbs; and Small Clauses occur mainly as the Complements of a subset of Verbs and Prepositions . . .” It should be noted that I here is tense, modal, or infinitival to, and C is complementizer. Examples of ordinary clauses are given in (2), (3) and (4) below.

(2)



(3)





In computational linguistics, there is no single answer to the question of what a clause is, since this depends on the particular grammatical theory chosen in a given computational framework.

In order to illustrate one particular and explicit notion of clause, or more precisely predication, in computational linguistics, I want to quote an interesting study by Henry Kučera (ms, 1985) on the computational analysis of predicational structures in the Brown Corpus.

He considers a predication to be, first of all, any verb or verbal group with a tensed verb that is subject to concord (for person and number) with its grammatical subject. These verbal constructions he calls finite predications. In addition to that, he also includes in his analysis non-finite predications, consisting of infinitival complements, gerunds and participles. What he did in his study was to identify and classify all the predications, which were 145,287 in all the 54,724 sentences of the Brown Corpus.

Table 1 shows for each genre in the corpus, the mean sentence length (words

Genre	Words per Sent.	Pred. per Sent.	Words per Pred.
A. Press, report.	20.81	2.65	7.85
B. Press, edit.	19.73	2.74	7.20
C. Press, reviews	21.11	2.65	7.96
D. Religion	21.23	2.90	7.32
E. Skills	18.63	2.60	7.17
F. Pop. lore	20.29	2.82	7.20
G. Belles lett.	21.37	2.94	7.27
H. Misc.	24.23	2.82	8.59
J. Learned	22.34	2.87	7.78
K. Fiction, gen.	13.92	2.41	5.78
L. Mystery/detect.	12.81	2.29	5.59
M. Science fict.	13.04	2.23	5.85
N. Adv./Western	12.92	2.30	5.62
P. Romance	13.60	2.45	5.55
R. Humor	17.64	2.84	6.21
CORPUS	18.49	2.65	6.97

Table 1:

per sentence), sentence complexity (predications per sentence), and mean predication length (words per predication).

Table 2 below shows that whereas sentence length varies a great deal between a mean of 21 words per sentence in informative prose (INFO) and 13 words per sentence in imaginative prose (IMAG), sentence complexity does not vary that much between genres: 2.80 versus 2.38 predications per sentence.

Measure	INFO	IMAG	CORPUS
Words/Sent.	21.12	13.55	18.49
Pred./Sent.	2.80	2.38	2.65
Words/Pred.	7.54	5.69	6.97

Table 2:

Table 3 below shows how the finite (F) and non-finite (NF) predications were distributed in the genres of informative and imaginative prose.

Group	Type	No.	Pred. per Sent.	Percent
INFO	F	68,157	1.91	68.09%
	NF	31,935	0.89	31.91%
		100,092	2.80	100.00%
IMAG	F	34,329	1.81	75.96%
	NF	10,866	0.57	24.04%
		45,195	2.38	100.00%
CORPUS	F	102,486	1.87	70.54%
	NF	42,801	0.78	29.46%
		145,287	2.65	100.00%

Table 3:

What Kučera considers as the main result of his study is the lack of correlation between sentence length and sentence complexity, and it is indeed surprising.

Kučera's study was concerned with finding, counting and classifying predications units (verbal groups) in the Brown Corpus. It was not concerned with what would have been an even more difficult goal, that of finding entire clause units, in the sense of demarcating their beginnings and endings. There is an obvious relation between predications and clauses, in that a reasonable definition of clause, I think, would be one in which there is one predication, in Kučera's sense of the term, per clause.

In Ejerhed (1988), which is a computational linguistic study of clauses in English, done in collaboration with Ken Church when I visited ATT Bell Laboratories 1986–87, I used a definition of clause that differed somewhat from the one considered in the previous paragraph. In my definition of clause in English,

only finite and to-infinitival predications are criterial for clausehood. Other infinitival predications, gerunds and participles are not taken to imply the presence of a clause unit.

Another feature of my definition of clause that was used in parsing clauses in unrestricted text, is that the opening of a new clause always implies the closure of the previous clause unit, whether or not this unit is complete with subject and predicate, or complete with respect to the argument structure of its predicate. To illustrate this no-nesting of clauses, the sentence in (2) is reproduced in (5) below with clause boundaries inserted where the clause parsers described in Ejerhed (1988) would place them.

(5) [Mary might think] [that he will resign]

There are several reasons for the move to adopt the hypothesis that clauses do not nest, at a very superficial level of syntactic structure.

The first reason is that the hypothesis makes possible an exceedingly simple definition of, and recognition algorithm for, clauses: a clause can be defined as a set of permissible sequences of word classes by means of a regular expression, i.e. by using the operations of concatenation, union and Kleene star on elements that are word classes.

That such a simple definition of clauses, or sentence forms as he called them, was possible, was something Harris considered, but rejected in the following passage from Harris (1968:31–32):

... in English a *wh*-clause can be away from its noun (usually if no other noun intervenes):

Finally the man arrived whom they had all come to meet.

In describing sentences, one can still say that there is a constituent, even though with non-contiguous parts: the subject above is MAN with adjoined THE on the left and WHOM ... after the verb on the right.²³ But the difficulty lies in formulating a constructive definition of the sentence. For if we wish to construct the sentence by defining a subject constituent and then next to it a verb (or predicate) constituent, we are unable to specify the subject if it is discontinuous, because we cannot specify the location of the second part (the adjunct at a distance). At least we cannot specify the location of the distant adjunct until we have placed the verb constituent in respect to the subject; but we cannot place the verb in respect to the subject as a single entity unless the subject has been fully specified.²⁴

²³ And one can specify that it can be at a distance primarily if no noun intervenes.

²⁴ To the extent that such problems did not arise, it would be possible to define sentence forms as short sequences of morpheme classes (or word classes), each class being expandable by a certain neighborhood of other classes (my emphasis EE).

The sentence discussed in the passage above would be parsed as indicated in (6), given the clause grammar of Ejerhed (1988).

- (6)
 [Finally] [the man arrived] [whom they had all come to meet]

The second reason for the hypothesis that clauses do not nest has to do with performance considerations, i.e. observational data from studies in psycholinguistics and phonetics.

For a review of the clausal hypothesis in psycholinguistics and studies relating to it, the reader is referred to Flores d'Arcais and Schreuder (1983:14–19). They present the clausal hypothesis as a view of sentence comprehension that is characterized by two major features. First, clauses are taken to be the primary units of normal speech perception. Incoming material is organized in immediate memory clause by clause; the listener or reader accumulates evidence until the end of a clause. Second, at the end of a clause, working memory is cleared of surface grammatical information and the content of the clause is represented in a more abstract form. They point out that these two major properties of the hypothesis are logically independent.

Phonetic evidence for the segmentation of speech (in perception as well as production) at the level of clauses, as structurally defined units, will be discussed in the last section of the paper, after a presentation and illustration of a structural definition of Swedish clauses.

2 A Swedish Clause Grammar

This grammar for Swedish clauses has the same structural units as targets as the grammar for English clauses in Ejerhed (1988), modulo the difference between the two languages, i.e. finite (tensed) clauses and infinitival clauses introduced by *att* are clauses. In addition, there are three types of clause fragments: verb phrase fragments, noun phrase fragments and adverb fragments.

In an appendix to this paper, there is a Swedish newspaper text from April 1984 which has been segmented into clauses and clause fragments, labelled to the right according to the type of unit in the grammar that they instantiate. The categories that are criterial to the identification of a clause or clause fragment according to the grammar, have been labelled underneath.

GRAMMAR

Main clause (mc)

- | | |
|--------------|----------------------------|
| 1. mc-noninv | (COORD) NP' VFIN ... |
| 2. mc-inv | (COORD) VFIN (SADV) NP ... |
| 3. mc-coord | COORD VFIN ... |

Subordinate clause (sc)

- | | |
|-------------|----------------------------|
| 4. sc-comp | (COORD) (PREP) COMP ... |
| 5. sc-coord | COORD (SADV) VFIN/VSUP ... |

6. *sc-nocomp* (COORD) NP' (SADV) VFIN/VSUP ...

VP-fragment

7. *mc vp-fragment* VFIN ...

8. *sc vp-fragment* (SADV) VFIN/VSUP ...

NP-fragment

9. (COORD) (COMP) NP' ...

10. NP' COORD NP'

ADV-fragment

11. (COORD) PP/ADVP/SADV*

A few words on the notation used in the grammar are required. For readability, concatenation is represented simply by juxtaposition. Union (i.e. alternatives) is represented by /, and the special case where something alternates with nothing (i.e. optionality) is represented by (). Kleene star is represented by *, which has scope over /. The three dots ... should be read as a variable over any word class.

- COORD is the category of coordinating conjunctions, *och, eller, men*.
- NP is a non-recursive noun phrase consisting of any prenominal modifiers plus head noun. NP does not include any postnominal modifiers. For the concept of such a noun phrase as applied to English, see Church (1988).
- NP' consists of a non-recursive NP followed by postnominal modifiers that are non-clausal, i.e. prepositional phrases PP, or adverbs ADV. Thus, NP' = NP PP/ADV*
- VFIN is the category of finite verbs, active or passive, and VSUP is the category of supinum forms of verbs occurring after the auxiliary *hava*. Because finite forms of *hava* can be optionally deleted in subordinate clauses in Swedish, it is necessary to allow occurrences of VSUP in such cases to count as finite.
- COMP is the category of subordinating conjunctions, including *att* as infinitive marker.
- SADV is the category of sentence adverbs, *inte, ofta, aldrig*.
- ADVP is the category of adverbial phrases.
- PREP is the category of prepositions.

Each of the regular expressions 1 through 11 constitutes an alternative definition of clause or clause fragment. The way that these alternative definitions interact in the processing of a text is very important. In cases where two or more alternative analyses compete, *the regular expression that matches the longest substring wins*. This can be illustrated by considering how the first sentence of the text in the appendix is processed. The sentence is repeated below with numbers indicating linear positions in the string of words.

(7) 0 Allting 1 verkar 2 så 3 okontrollerat 4
 NP VFIN ...

The regular expression 9, NP-fragment, matches the string of words from 0 to 1.

The regular expression 7, VP-fragment, matches the string of words from 1 to 4.

The regular expression 1, non-inverted main clause, matches the string of words from 0 to 4. This is the expression that matches the longest substring, and it wins over the alternative analyses of the string from 0 to 4.

The status of the implementation of this particular clause grammar for Swedish is that it is in the process of being implemented. What that means, is that I do not yet have a running program for Swedish that automatically decides the location of boundaries between clauses and clause fragments in unrestricted text. This is an ambitious and long range goal, and the biggest problem in developing such a program is lexical. Each word in a text has to be labelled with a unique syntactic category (including information about the form of the word) before any matching against the regular expressions in the grammar can take place. The category label assigned to a word has to be the one that is correct for the word in its context of occurrence.

A successful approach to the problem of automatically assigning unique and correct syntactic categories to English words in context is probabilistic (Church 1988, DeRose 1988, Eeg-Olofsson 1985). This is one of several approaches that will be applied to Swedish in the context of a joint corpus based research project between the universities of Stockholm and Umeå (Källgren, Ejerhed) that will start in the fall of 1989.

Another approach to the disambiguation of the syntactic category and form of a word in context is rule based, constraint based or heuristic, and the disambiguation between alternative analyses of a word is done as an integrated part of the parsing of a text, rather than as a separate subroutine completed before parsing begins. A version of this approach has been applied to Swedish with successful (95% correct) results (Brodda 1983, Källgren 1984a, 1984b).

Fred Karlsson claimed in his paper at this conference, on the basis of his recent research on disambiguation, that more than 60% of the consecutive words in a Swedish text are at least two-way ambiguous, as compared with 45% in English according to DeRose (1988), and 11% in Finnish. Karlsson's figure for Swedish tallies with what is reported in Allén (1970:XV, XXV): 645,000 out of the 1,000,669 words of the Swedish corpus Press-65 were homographs, and that amounts to 64.5%.

What I have by way of implementation at this time is a modification of the finite state parser for Swedish, described in Ejerhed & Church (1983), Ejerhed & Bromley (1985), and Ejerhed (1986). Subject to the limitations of its lexicon, which is currently being expanded, the modified parser, in its parsing of orthographic sentences as input, is capable of identifying and assigning constituent structure to substrings that can be put in direct correspondence with the 11 different clauses and clause fragments enumerated in the new clause grammar described here.

3 Phonetic Data concerning Clause Boundaries

There are two recent phonetic studies of spoken Swedish, based on recordings of several different speakers reading the same texts aloud. One is by Eva Strangert (Strangert and Zhi 1989, Strangert 1989) and the other by Dieter Huber (1988).

Strangert's research project, which is still going on, studies perceived pauses in 2 texts of a total of 810 words read aloud by 10 different speakers at 3 different speech rates, and the acoustic and grammatical properties of such pauses. The first of the two texts is identical to the text in the appendix of this paper. Acoustically, a perceived pause can be signalled in several different ways: by final lengthening, a special fundamental frequency contour, silence, and/or voice quality irregularities. Strangert and Zhi (1989) reports findings primarily concerning these acoustic properties of the pauses perceived by two different judges. Strangert (1989) is also concerned with the distribution of the perceived pauses in relation to the following kinds of boundaries: paragraph, sentence, clause and phrase.

Using the definition of clause presented in this paper, I have segmented the two texts used in Strangert's study and found that they consist of a total of 115 units that are clauses or clause fragments. The number of perceived pauses at these 115 clause boundaries is presented in Table 4 below, for which I am indebted to Eva Strangert. A perceived pause is here a pause judged by both of the two judges to be present in the speech of at least 5 of the 10 speakers. For the purposes of this table, all clause boundaries have been included, whether they are sentence internal, or happen to coincide with sentence boundaries or paragraph boundaries. In Strangert (1989) these three boundary conditions are treated separately.

Speech rate	Number of clause boundaries with perceived pauses	Percent (N = 115)
Fast	57	50
Normal	78	68
Slow	97	84

Table 4: The frequency of clause boundaries where pauses were perceived.

The study of Huber (1988) is concerned with intonation units in recordings of 3 newspaper texts read aloud by 4 different speakers of Swedish, a total of 2.2 hours of connected speech. He defines the concept of intonation unit in purely acoustical terms, related to fundamental frequency only, and devises a method of automatically segmenting connected speech into such intonation units. The advantage of this segmentation procedure is that it makes no reference to either higher level linguistic information concerning syntax, or to lower level physiological information concerning pausing, breathing, phonation onset or offset etc. He arrives at a total of 1664 intonation units in the accumulated text material (3 texts, 4 speakers). Table 5 shows the grammatical correlates of the 1664 intonation units, averaged across four speakers and three texts. For the exact definitions of the grammatical units, see Huber (1988:78). Of interest here is that he defines as sentences "graphic sentences that begin with a capital letter and end with a full stop (or some other mark of 'final' punctuation)". And he defines as clauses "units of linguistic organisation smaller than the sentence and consisting of at least one subject and one finite verb".

Grammatical Unit	Number of intonation units	Percent
SENTENCE	299	18.2
CLAUSE	662	39.7
SUBJECT	83	4.8
VERBPHRASE	76	4.5
ADVERBIAL, init.	35	2.0
ADVERBIAL, final	141	8.5
PARENTHETICAL	132	8.0
MISCELLANEOUS	238	14.3
Total	1666	100.0

Table 5: Frequency of intonation units corresponding to different grammatical categories.

Unfortunately, these figures cannot be directly related to the notions of clause and clause fragment discussed in this paper, because the definitions of the grammatical categories do not agree. However, it is likely that we can equate monoclausal *sentences* (which accounted for 63.6% of the 1-IU-per-sentence that occurred) with a subset of main clauses (Rules 1–3 in the Swedish clause grammar), *clauses* with either a subset of main clauses (in the case of multiclausal sentences) or a subset of subordinate clauses (Rules 4–6), and *initial adverbials* with adverb-fragment (Rule 11), and these three categories together account for 60% of all intonation units. It is also likely that *subject* corresponds to NP-fragment, and *verbphrase* to VP-fragment on the basis of the illustrative examples of these categories in Huber (1988:83–85). If so, close to 70% of Huber's intonation units would correspond to a clause or clause fragment in the sense of the present paper. In order to establish the exact extent to which the notions of clause and clause fragment proposed here correlate with the intonation units found in Huber's study, a separate study is being undertaken in collaboration with Huber.

Acknowledgement

The work on this paper was done while the author was a member of the Speech group headed by Bertil Lyberg, Department of Research and Development, Swedish Telecom, Stockholm, as well as of the Department of Linguistics, University of Umeå. I am indebted to Swedish Telecom in Stockholm for the use of the resources of its Speech Lab, and to Eva Strangert in Umeå for collaboration on pauses.

References

- Allén, S. 1970. *Nusvensk frekvensordbok baserad på tidningstext. 1. Graford, homograftkomponenter*. Data linguistica, 1. Almqvist & Wiksell, Stockholm.
- Brodda, B. 1983. An experiment with heuristic parsing of Swedish. *Proceedings of the First Conference of the European Chapter of the Association for Computational Linguistics*:66–73. Pisa.
- Church, K.W. 1988. A stochastic parts program and Noun Phrase parser for unrestricted text. *Proceedings of the Second Conference on Applied Natural Language Processing*:136–143. Association for Computational Linguistics, Austin, Texas.
- DeRose, S.J. 1988. Grammatical category disambiguation by statistical optimization. *Computational Linguistics*, 14(1):31–39.
- Eeg-Olofsson, M. 1985. A probability model for computer aided word class determination. *ALLC Journal*, 5(1 & 2):25–30.
- Ejerhed, E. 1986. A finite state parser for Swedish with morphological analyzer and semantics. *Proceedings of SAIS-86*. Institutionen för Datavetenskap, Linköpings universitet.
- Ejerhed, E. 1988. Finding clauses in unrestricted text by finitary and stochastic methods. *Proceedings of the Second Conference on Applied Natural Language Processing*:219–227. Association for Computational Linguistics, Austin, Texas.
- Ejerhed, E. and H.J. Bromley. 1986. A self-extending lexicon: description of a word learning program. F. Karlsson [Ed.] *Papers from the Fifth Scandinavian Conference of Computational Linguistics*:59–72. Publication No. 15, Department of General Linguistics, University of Helsinki.
- Ejerhed, E. and K.W. Church. 1983. Finite state parsing. F. Karlsson [Ed.]. *Papers from the Seventh Scandinavian Conference of Linguistics*:410–432. Publication No. 9, Department of General Linguistics, University of Helsinki.
- Flores d'Arcais, G.B. and R. Schreuder. 1983. The process of language understanding: A few issues in contemporary psycholinguistics. G.B. Flores d'Arcais and R. Jarvella [Eds.]. *The Process of Language Understanding*:1–41. Wiley, Chichester.
- Harris, Z. 1968. *Mathematical structures of language*. Wiley, New York.
- Huber, D. 1988. Aspects of the communicative function of voice in text intonation—Constancy and variability in Swedish fundamental frequency contours, Department of Computational Linguistics, University of Göteborg, Department of Information Theory, Chalmers Institute of Technology, Göteborg, and Department of Linguistics and Phonetics, University of Lund.

- Karlsson, F. 1989. The resolution of morphological ambiguities. Paper presented at the Scandinavian Conference of Computational Linguistics, Reykjavik, June 26–28, 1989.
- Kučera, H. n.y. Computational analysis of predicational structures in English. Brown University, Providence, R.I. (unpublished).
- Kučera, H. 1985. The analysis of the English verbal group. Paper presented to the ICAME Sixth International Conference on English Language Research on Computerised Corpora, Lund (unpublished).
- Källgren, G. 1984a. HP-systemet som genväg vid syntaktisk märkning av texter. *Svenskans beskrivning*, 14:39–45, Lunds universitet.
- Källgren, G. 1984b. HP—A heuristic finite state parser based on morphology. A. Sägvall Hein [Ed.]. *De nordiska datalingvistikdagarna 1983*:155–162. Centrum för Datorlingvistik, Uppsala Universitet.
- Radford, A. 1988. *Transformational Grammar: A First Course*. Cambridge University Press, Cambridge.
- Strangert, E. and M. Zhi. 1989. Pause patterns in Swedish: A project presentation and some data. *Fonetik-89. Speech Transmission Laboratory Quarterly Progress and Status Report*, 1:27–31. KTH, Stockholm.
- Strangert, E. 1989. Pauses, syntax and prosody. Paper presented to the Nordic Prosody V meeting in Turku, Finland, August 23–25, 1989 (to appear).

Department of Linguistics
University of Umeå
S-90187 Umeå
Sweden
EJERHED@SEUMDC51 (Bitnet)

Text A1

The text is divided into paragraphs by consecutive numbering. The paragraphs are divided into orthographic sentences by sentence final punctuation marks. The sentences are divided into non-recursive clauses or clause fragments marked by [], and each such unit is labelled according to the Swedish clause grammar presented in this paper.

Paragraph 1	
[Allting verkar så okontrollerat.]	mc-noninv
NP VFIN	
[Det tycks]	mc-noninv
NP VFIN	
[som om ingen längre håller i styret.]	sc-comp
COMP	
[Framför allt]	adv-fragment
P NP	
[verkar läget vara okontrollerat inne i Tripoli]	mc-inv
VFIN NP	
[där ungdomar i femtonårsåldern på något sätt	sc-comp COMP
har fått tag i skjutvapen.]	
Paragraph 2	
[Det sade en ung spanjor]	mc-noninv
N VFIN	
[som var en av de 113 personer]	sc-comp
COMP	
[som lyckades komma ut ur Libyen	sc-comp
COMP	
med den första flygningen]	
[sedan USA bombade Tripoli och Bengazi	sc-comp
COMP	
i början av veckan.]	
[Den unge spanjoren fanns ombord	mc-noninv
NP VFIN	
på det reguljärplan från Libyan Airlines]	
[som kraftigt försenat landade på	sc-comp
COMP	
den internationella flygplatsen utanför Rom	
sent på torsdagen.]	
Paragraph 3	
[Planet återvände aldrig till Tripoli	mc-noninv
NP VFIN	
på torsdagskvällen.]	

[En väntande skara journalister fick NP VFIN officiellt beskedet]	mc-noninv
[att besättningen helt enkelt var för uttröttad.] COMP	sc-comp
Paragraph 4	
[Libyan Airlines flygning 167 tillbaka till NP ADV P den libyska huvudstaden uppsköts därför till NP VFIN någon gång under fredagen.]	mc-noninv
Paragraph 5	
[Ingen av de 113 passagerarna på den första NP P NP P NP utflygningen från Tripoli var svensk.] P NP NP VFIN	mc-noninv
[Det finns omkring 200 svenskar i Libyen] NP VFIN	mc-noninv
[varav ungefär hälften bor i huvudstaden Tripoli.] COMP	sc-comp
[Den svenska ambassaden har rekommenderat] NP VFIN	mc-noninv
[att de svenskar] COMP	np-fragment
[som arbetar i Libyen] COMP	sc-comp
[skall evakuera sina familjer] VFIN	vp-fragment
[så snart tillfälle ges.] COMP	sc-comp
Paragraph 6	
[Den unge spanjoren,] NP	np-fragment
[som ville vara anonym,] COMP	sc-comp
[talade om en skräckstämning i Tripoli] VFIN	vp-fragment
[där ingen egentligen vet] COMP	sc-comp
[vem som bestämmer.] COMP	sc-comp
Paragraph 7	
[En vild ryktesflora grasserar också NP VFIN om ledaren Muammar Gadaffi.]	mc-noninv

[Det har även under torsdagen förekommit NP VFIN skottlossning i den militärförläggningen i Tripoli]	mc-noninv
[där Gadaffi och hans familj bodde] COMP	sc-comp
[när de amerikanska bombplanen slog till COMP natten till tisdagen.]	sc-comp
Paragraph 8 [Det osäkra läget befästes NP VFIN på torsdagen ytterligare]	mc-noninv
[av att minst tre passagerarplan från Spanien, P COMP Rumänien och Jugoslavien avbröt sina flygningar till Tripoli.]	sc-comp
[Planen startade] NP VFIN [men fick återvända till sina hemorter.] COORD VFIN	mc-noninv mc-coord
Paragraph 9 [Då det gällde Libyan Airlines första utflygning] COMP [florerade också ryktena.] VFIN SADV NP [Då planet skulle ha startat återfärden COMP från Rom kl 17]	sc-comp mc-inv sc-comp
[hade det ännu inte lyft från utgångspunkten VFIN NP Tripoli.]	mc-inv
[Flera passagerare dementerade dock uppgifter NP VFIN om skottlossning i samband med starten utanför Tripoli.]	mc-noninv
Paragraph 10 [Men de bekräftade] COORD NP VFIN [att det råder kaotiska förhållanden i COMP den libyska huvudstaden.]	mc-noninv sc-comp

[De flesta passagerarna var från öststater.]	mc-noninv
NP VFIN	
Paragraph 11	
[De flesta håller sig inomhus även under dagtid,]	mc-noninv
NP VFIN	
[sade en polsk medborgare.]	mc-inv
VFIN NP	
[Ute på gatorna]	adv-fragment
ADV P NP	
[är det alldeles för osäkert.]	mc-inv
VFIN NP	
[Det finns alldeles för många ungdomar med gevär]	mc-noninv
NP VFIN	
[för att man skall kunna känna sig säker.]	sc-comp
P COMP	
Paragraph 12	
[Och ryktena om överste Gadaffi]	np-fragment
COORD NP P NP	
[och vad som har hänt honom]	sc-comp
COORD COMP	
[är lika många som fantastiska.]	vp-fragment
VFIN	