Papers from the Ninth Scandinavian Conference of Linguistics

Stockholm, January 9-11, 1986

ed. by

Östen Dahl

Institute of Linguistics
University of Stockholm
S-106 91 Stockholm, Sweden

Fred Karlsson
Department of General Linguistics
University of Helsinki

PROCESS GRAMMAR

1. Introduction

Grammars in the standard sense are autonomous descriptions of the language system, or an idealised competence, and take no stand on how the grammar is used or could be put to use in the processes of language production and comprehension. A similar conceptual distinction between grammar and parser is common in computational linguistics. On the other hand, it is not unusual that autonomous grammars are used for hypothesis generation in psycholinguistic research. The most celebrated example is the Derivational Theory of Complexity widely tested in the 1960s. In this context it was soon noted that serious problems arise when one tries to directly incorporate wholly autonomous grammars into process-oriented theories.

The purpose of this paper is to explore the alternative possibility of incorporating processing considerations into grammar theory. Thus, we pursue a different goal than the common one of trying to realise a performance-independent competence grammar as a parser.

A process grammar (PG) formalism should be designed to directly reflect certain basic performance properties (section 2). Process grammars are thus not primarily intended to express "optimal generalizations" holding over and between various types of syntactic and morphological structures. Such generalizations are typical inmates of autonomous grammars.

PGs are not intended as substitutes for autonomous grammars. The aims of these two enterprises are partly different. The basic ontological premise of autonomous grammar is language seen as a selfcontained system: Process grammars are models of (some aspects of) language use. Therefore, process grammars are alternatives to process (or performance) <u>interpretations</u> of autonomous grammars.

Neither is PG a substitute for psycholinguistics. Even if PGs are models of the processes of language production and comprehension, they are formulated at a macrolevel that is basically linguistic in nature. Process grammars can be seen as an attempt to investigate the linguistic implications of the basic results of experimental psycholinguistics such as those of Marslen-Wilson (overview 1984) for word-form recognition, or those of Frazier, Clifton, & Randall (1983) for gap-filling phenomena.

Much relevant work of this type has been done in computational linguistics. Topical contributions are e.g. Kaplan's (1972) early work on the psychological interpretation of augmented transition networks, Marcus' (1980) deterministic parser PARSIFAL, Lexical-Functional Grammar (Kaplan & Bresnan 1982), Ades & Steedman's (1982) augmented categorial grammar, Johnson-Laird's (1983) left-corner parser based on psycholinguistic considerations, and the government-binding based model of Berwick and Weinberg (e.g. 1984). The present PG approach differs from these mainly in insisting upon a less strict (but still principled) demarcation between grammar and parser.

2. Basic properties of process grammar

The subsequent discussion is confined to language comprehension, especially to morphology and syntax. No stand will be taken on the bidirectionality issue. Furthermore, we shall remain neutral in regard to the differences between spoken and written language comprehension. The supposition is that there is some linguistically and processually relevant level where spoken and written language comprehension share representations. Nine central properties of PGs will be considered. An example grammar will be presented in section 4 and an example of an analysis in section 5.

(i) A process grammar should basically work <u>from left to right</u> (i.e. in real time) and be able to <u>assign structure immediately and incrementally</u> as a function of all the structural and, optimally, semantic information available with ongoing sensory input. In particular, a PG should not presuppose availability of the whole sentence before it can start working. Thus, some structure can normally be assigned already to the first input word. This property is an essential one and it constitutes an important difference in comparison to autonomous grammars. The rules of the latter normally presuppose that long(er) chunks (maximally, whole sentences) be seen before structure can be assigned. There is undisputable psycholinguistic evidence that this property of incremental left to right interpretation is pervasive in language comprehension (cf. the work of Marslen-Wilson and his associates on word form recognition, in particular Marslen-Wilson's (1976) discussion of this property in regard to sentence comprehension).

Structure is built incrementally as a function of current input, previously interpreted input, pending (i.e. not yet fully analyzed) words and larger structural subparts, and rules and strategies of specific types (cf. section 4).

(ii) Due to the complexity of natural language, full-scale models have to be modular, roughly along the traditional dimensions of phonology/graphemics, morphology, syntax, semantics, pragmatics, inference on common-sense knowledge, etc. Optimally, PGs should allow for <u>interactive parallel processing</u> where all modules can feed each other as directed by ongoing input. Currently, little more than lip-service can be paid to this desideratum. However, the present model explicitly provides for certain interactions between morphology and syntax. Assigned syntactic structure is used for eliminating word-form ambiguities, and each word-form immediately comes to the attention of the syntactic rules.

(iii) In temporarily non-deterministic situations at intermediate points in the sentence there are three ways of resolving ambiguities. All possible analyses may be pursued, just one may be chosen on heuristic grounds, or the decision may be delayed until sufficient information is available for a unique analysis (if there is one). We shall opt for a combination of the latter two alternatives. First, evidence from experiments with garden-path sentences of the well-known type "The horse raced past the barn fell" show that there exist more likely and less likely interpretations at least for some structural ambiguities. In such situations, we shall allow the PG to pick the more likely analysis, at the expense of bein forced to back up if the selected route turns out to be inappropriate (cf. Frazier, Clifton & Randall (1983) for some recent discussion). It is unclear whether the notion "more likely" can be sufficiently constrained to render it theoretically interesting. In the worst case, it amounts to no more than crude heuristic guessing.

But the normal way of coping with temporary ambiguities will be <u>delayed processing</u>. The decision is postponed until sufficient information becomes available. However, delayed processing may be effected in several ways. A standard method is to store away the ambiguous word/constituent in a HOLD register. A different approach is adopted here. When a word enters the syntactic processor, it will be assigned <u>all functional interpretations</u> compatible with the current situation. Later processing may discard some of the previously assigned multiple interpretations (cf. section 5). This makes structural ambiguities fall out with the same or even a lesser amount of processing than what is invoked in a fully disambiguated clause (also cf. Church & Patil, ms.).

No <u>look-ahead</u> is allowed. It is strictly speaking inconceivable that the human language processor would invoke look-ahead. The same effect is always possible to achieve by delayed processing.

A central requirement of a realistic process grammar is that it should <u>not</u> turn out a proliferation of interpretations having spent an amount of processing time proportional to the number of interpretations (as is typically the case if all alternatives are separately pursued). After all, temporary ambiguities do not slow down language comprehension, and structurally ambiguous sentences often are assigned just one interpretation (which is picked fast). These are essential facts that should be reflected in the design of a PG.

(iv) Frequently occurring, <u>prototypical structures</u> should get fast and reliable structural interpretations. Processing may be slower and involve more checking for complex, untypical syntactic structures. Of course, this kind of property explains why the finite verb interpretation is favored over the reduced relative clause interpretation in garden-path structures of the above type.

(v) <u>Redundancy</u> is accorded a much more prominent position than in current autonomous grammar theory. E.g., processing should be faster towards the end of a clause than in the beginning since early constituents narrow down the range of structural alternatives for later constituents. This property will be directly incorporated into our PG model.

Furthermore, the canon of striving for "optimal" or "significant" generalizations will not be followed. E.g., since there are no positive indications that people would have performance analogues of linguistic rules relating actives and passives, or related sentences with and without <u>there</u>-insertion, the PG will not be required to contain such rules. Of course, this does not mean that these relations would not exist, or that they should not be captured in autonomous grammars. It just means that no such rules are directly involved in comprehending instances of actives, passives, etc.

(vi) PG uses only one level of structure, <u>surface structure</u>.

(vii) A prominent position is accorded to <u>morphology</u> the role of which has been much underestimated in current syntactic theory. A PG is supposed to work with a full morphology and a large lexicon, not just with a handful of previously selected forms of a handful of lexical items. This upgrading of the significance of morphology is necessary since it is the only way of effectively bringing out all potential temporary ambiguities there are in random sentences fed to the PG, and of realistically simulating their syntactic analysis. By large lexicon we mean one containing at least 10,000 lexemes. Full morphology, on the other hand, means that the PG should be capable of correctly analyzing all word forms of all lexemes in the lexicon. There should also be convenient facilities of extending the lexicon by adding new lexical items the morphological behavior of which is predicted by productive rules of the morphological processor.

Precisely such a lexicon and morphology is at the heart of the Finnish PG we have been experimenting with (Karlsson 1986). Currently, this lexicon contains 12,000 lexemes all the forms of which are correctly analyzed, with multiple analyses, if any. A realistic PG syntax must stand on a solid morphology.

(viii) The core of the syntax is not an ordinary context-free phrase structure grammar but pattern-action rules assigning dependency-oriented <u>functional descriptions</u> similar to those of traditional grammar. Thus, for each word it must be decided whether it is a head or a modifier. Allowed heads, in Finnish, are SUBJECT (S), OBJECT (O), PREDICATE COMPLEMENT (PC), and ADVERBIAL (ADVL), ADVL with semantic subgroups. If a word is a modifier, its head must be indicated. The set of Finnish modifier labels includes NPREMOD (premodifier to a noun in S, O, PC, or ADVL function), NPOSTMOD, APREMOD, APOSTMOD, PRPMOD (prepositional complement), and PSPMOD (postpositional complement).

The use of functional head/modifier structure gives the opportunity of capitalizing on the following important restriction, here called the Head Restriction:

> <u>Head Restriction</u>
> In a simplex clause, there is at the most one occurrence of each of S, O, IO, PC, and MAINPRED.

As soon as one of these categories has been conclusively identified, no more instances of the same category are to be expected later in the current clause. This provides a way of modelling how the number of potential interpretations decreases towards the end of the clause. The repertoire of functional head labels available for consideration by the PG at any given moment will be precisely the complement of the set of conclusively identified designated heads listed above. - Ordinary phrase structure grammar does not make possible the use of such implications since there is no upper bound on the number of NPs and AdvlPs occurring in a clause.

(ix) A PG should be capable of assigning structure also to (somewhat) <u>degenerate input</u>, especially to sentences with heavy ellipsis (e.g. consisting only of one word), but also to sentences violating ordinary norms but still comprehensible. Such instances often turn out to be analyzable by PG rules because these are not "rules" in the strict sense of describing optimal norms. Rather, they are norm-neutral inductions on the current situation.

Once more, this brings out the difference between autonomous grammar and process grammar. The former normally describe idealized norm systems, the latter relate more directly to behavior (which often violates optimal norms).

3. Workspace

The central elements in the presumed mental workspace of a PG are word slots, structure slots, and the grammar board.

There is a word slot for each input word. Appended to each word slot there is a bipartitioned structure slot. One part of the structure slot (the M-slot) contains morphological information, i.e. a set of morphological categories and features provided by the morphological processor. The other part (F-slot) contains function labels (one or more) assigned by the rules on the board. The contents of the function slot may change during analysis but only in a regimented way. Some initially assigned labels may be discarded upon turning out to be incompatible with the evolving analysis. Change of an assigned functional label to another is allowed only in genuine garden-path sentences.

The grammar board contains the grammar, or rather, by a somewhat loose metaphor, a clone of the full grammar. One of the basic ideas is that the contents of the grammar clone on the board may change during the evolving analysis. When a sentence is to be analyzed, all word slots are emptied and a full fresh clone of the grammar is put on the board.

Actions enforced by the grammar rules lead to insertions of function labels into function slots, but they may also lead to expulsions of already assigned function labels both from the function slots and from the grammar board. The grammar clone may thus be in different "states" depending on what functional categories have been conclusively identified. E.g., if a certain word is identified as S, all previous assignments of S in the function slots of earlier encountered words are safely discarded. Likewise, all instances of S in all rules on the grammar board may now also be safely discarded since no more subjects are to be found in this clause. The grammar clone thereby enters a new state. Note that the grammar clone is reduced after each conclusive identification of a head falling under the Head Restriction, thereby decreasing the number of alternatives subsequently available (also cf. the Shift-Reduce Parsers discussed by Pereira (1985)).

4. Rules, constraints, and processes

A PG contains a set of designated heads, viz. those covered by the Head Restriction, and a designated initial state with a fresh clone of the full grammar, i.e. the set of morphofunctional mapping rules under (ii) below, on the board. There are seven types of rules, constraints, and processes.

(i) Morphological analysis by which all necessary morphological categories and features are inserted into the M-slots of each word.

(ii) Morphofunctional mapping rules (MFMR) that constitute the core of the grammar. MFMRs are simple mappings from

morphological categories onto functional categories, much like certain types of rules found in traditional grammars. Henceforth, examples will be drawn from English. A subset of typical English mapping rules looks as follows. For each category (to the left of the colon), an unordered set of all potential functional labels is given (to the right of the colon).

N:	S O IO PC NPREMOD PRPMOD
A:	NPREMOD PC
DET:	NPREMOD
AD-A:	APREMOD
PRON:	S O IO NPREMOD PRPMOD
V:	FINMAIN INF NONFINMAIN NPREMOD NPOSTMOD
PRPa:	ADVL NPOSTMOD
PRPb:	ADVL IO NPOSTMOD
NOM:	S O IO PC NPREMOD PRPMOD
GEN:	NPREMOD
ACC:	O PRPMOD
ING:	NPREMOD NONFINMAIN S O PC PRPMOD
PASTTNS:	FINMAIN
3rdP:	FINMAIN
(etc.)	

I.e., nouns may be subjects, objects, indirect objects, predicate complements, nominal premodifiers, prepositional complements, etc., adjectives may be nominal premodifiers or predicate complements, determiners may only be nominal premodifiers, intensifiers (AD-A) are adjectival premodifiers, verbs are finite main, infinitives, nonfinite main verbs (in composite verb forms), some prepositions (PRPa) are either heads of adverbials or nominal postmodifiers (e.g. in), others (PRPb) may additionally be heads of indirect objects (to), the nominative allows the same functions as nouns, the genitive is only NPREMOD, the accusative is O or PRPMOD, ing-forms may be nominal premodifiers, parts of composite verb forms, or S, O, PC, PRPMOD, just as nouns and nominatives (but not IO), etc.

The MFMRs are fairly uncontroversial and the reader is free to make changes and/or additions to the repertoire if that is called for. Such amendments would not change the gist of the argument. - The mapping rules are much like ordinary autonomous rules and there is nothing inherently processual in their format.

(iii) The basic process of PG is intersection of all MFMRs triggered by the morphological categories of the word-form being analyzed. The intersection is performed in regard to the current state of the grammar board. The outcome of intersection is the set of all functional labels currently allowed. Thus, given the above full grammar clone, intersection for the word-form beautiful with the morphological categories A, NOM, SG, yields the functional possibilities NPREMOD, PC. The outcome is inserted into the F-slot.

(iv) Depending on the current situation, constraints on F-slot insertion may occasionally rule out some alternatives produced by intersection. An English F-slot constraint would state "PC iff a finite copula has been passed", i.e., an N or A cannot be PC in initial preverbal position. These constraints do not affect the grammar board, they only restrict what goes into particular F-slots in particular situations.

(v) Expectancy introduction and termination, i.e. "flag raising" and "flag lowering". Certain pieces of structural information introduce or terminate expectancies concerning function assignment. E.g., PRPMOD is a relevant function for nouns only if a preposition has been passed, and PC is a possible function only if a copula-type verb has been passed. Constraints on F-slot insertion often depend upon expectancies or their absence.

(vi) F-slot and grammar updating takes place after every intersection when one of the designated functions covered by the Head Rule has been uniquely identified. Thus, when S is identified, all occurrences of S in the F-slots of previously analyzed words should be discarded. Furthermore, the grammar clone should be updated so as to contain no more instances of S in any MFMR. The grammar is thus reduced.

Important information in this regard is obtained when the finite verb is identified. If it is intransitive and not a copula, all occurrences of O, PC, and IO in all previous F-slots and the whole grammar may be eliminated. A set of special updating constraints have to be stated for enforcing these effects.

Eliminating individual instances of functions from F-slots and the grammar is not a repetitive sequential search process. An analogy from LISP is helpful here. If one sees all occurrences of a certain function in F-slots and the grammar as containing a pointer to the same location, all updatings can be done by one operation of changing the contents of this location (say, to NIL).

(vii) Patterns are wellformedness constraints on strings of F-slots. They rule out syntagmatically impossible sequences of function labels. E.g., if NPREMOD was inserted in the F-slot of a N, and the next word turns out to be a finite verb, a preposition, or nothing (in final position), a pattern tells that i.a. "NPREMOD FINMAIN" is not a permissible sequence, thereby ruling out NPREMOD. Patterns may also be used for assembling constituents and building up more complex clause and sentence patterns.

5. An example

We are now prepared to see how a PG for English goes about analyzing a simple sentence. Let the grammar G contain the MFMRs just discussed, which are cloned onto the board when the process begins, and let the sentence be "Eve shook the tree in the garden". Each word-slot occupies one line and consists of an identifying integer, the word-form, its M-slot, and its F-slot (separated by --).

Morphology yields i.a. the features N, NOM, SG for the first word "Eve", and intersection of these provides the set S, O, IO, PC, NPREMOD, PRPMOD. Now the constraints on F-slot insertion are consulted, informing i.a. that no copula has yet been noted. Therefore PC is ruled out. Furthermore, no preposition has been passed so as to be expecting a PRPMOD, and thus PRPMOD is discarded. Despite these reductions, the function of "Eve" was not uniquely identified, and therefore the grammar cannot be updated. The workspace thus contains an unchanged G clone and one word slot:

```
1.   Eve      N NOM SG --   S O IO NPREMOD
```


Note how all relevant possibilities, and only these, are left pending. "Eve" may be an active subject, a passive object, a topicalized IO with a stranded preposition, or a premodifier to a following noun. The correct interpretation remains unknown so long as only this one word has come to the attention of the processor.

The next word "shook" is analyzed yielding i.a. the features V, PAST, MONOTRANSITIVE, and intersection of these gives the unique function MAINPRED. Since a designated head was identified, the next step is F-slot and grammar updating. MAINPRED is turned off in all other F-slots and in G. The feature MONOTRANSITIVE informs that no PCs are IOs are to be found in this clause, and these labels are thus turned off in F-slots and G by an updating constraint. Since MAINPRED is not an auxiliary, NONFINMAIN is also ruled out from V in G. Since what has been seen of the verbal group makes clear that MAINPRED is active, a pattern discards O in preverbal F-slots (here, Y-movement phenomena are supposed not to be included in G). Another pattern examining F-slot syntagms discards NPREMOD in front of MAINPRED. But now only S remains in F-slot 1. Since a designated head was disambiguated, all occurrences of it are turned off in G and other F-slots.

After morphological and syntactic analysis of "Eve shook" the situation in the F-slots is thus:

```
1.   Eve     N NOM SG --              S
2.   shook   V PAST MONOTRANS --      MAINPRED
```

and the state of the updated grammar clone is:

```
N:        O NPREMOD PRPMOD
A:        NPREMOD
DET:      NPREMOD
AD-A:     APREMOD
PRON:     O NPREMOD PRPMOD
V:        NPREMOD NPOSTMOD
PRPa:     ADVL NPOSTMOD
PRPb:     ADVL NPOSTMOD
NOM:      O NPREMOD PRPMOD
GEN:      NPREMOD
ACC:      O PRPMOD
ING:      NPREMOD O PRPMOD
PASTTNS:
3rdP:
(etc.)
```

G has been considerably reduced and subsequent processing will be speeded up since there are less alternatives to choose from. In the current situation, an A can only be NPREMOD, a V only NPREMOD (premodifying ing) or NPOSTMOD (postmodifying infinitive), a N only O, NPREMOD, or PRPMOD, etc.

Now "the" is analyzed yielding morphological DET and functional NPREMOD in F-slot 3. Next "tree" is analyzed. Intersection of N, NOM yields O NPREMOD PRPMOD, but since no PRPMOD expectancy has been activated, a constraint on F-slot insertion again discards PRPMOD. The situation now is:

```
1.   Eve     N NOM SG --              S
2.   shook   V PAST MONOTRANS --      MAINPRED
3.   the     DET --                   NPREMOD
4.   tree    N NOM SG --              O NPREMOD
```

O was not yet uniquely identified ("tree" may be NPREMOD of a compound!). Therefore G remains unchanged. Next "in" is analyzed yielding the feature PRPa. Since no noun is potentially expecting a postmodifying PP, a constraint on F-slot insertion discards NPOSTMOD, i.e. "in" is uniquely ADVL, and an expectancy for NPOSTMOD is activated. But again a pattern tells that NPREMOD in slot 4 turned unfeasible since the next word was a preposition. This in turn makes O unique in slot 4 and therefore O is expelled from G. Now we have the situation:

1.	Eve	N NOM SG --	S
2.	shook	V PAST MONOTRANS --	MAINPRED
3.	the	DET --	NPREMOD
4.	tree	N NOM SG --	O
5.	in	PRPa --	ADVL

and G reduced to:

N:	NPREMOD PRPMOD
A:	NPREMOD
DET:	NPREMOD
AD-A:	APREMOD
PRON:	NPREMOD PRPMOD
V:	NPREMOD NPOSTMOD
PRPa:	ADVL NPOSTMOD
PRPb:	ADVL NPOSTMOD
NOM:	NPREMOD PRPMOD
GEN:	NPREMOD
ACC:	PRPMOD
ING:	NPREMOD PRPMOD
PASTTNS:	
3rdP:	
(etc.)	

i.e., no more designated heads are expected to occur in this clause. G could not be further reduced regardless of what the remaining words and constituents are.

The PRPMOD expectancy remains active. The next word "the" is treated as above. Finally "garden" enters the processor. Intersection of N, NOM turns out NPREMOD and PRPMOD. Now the PRPMOD expectancy makes this interpretation highly probable, and NPREMOD is conclusively ruled out since the sentence was terminated. The final analysis is thus:

1.	Eve	N NOM SG --	S
2.	shook	V PAST MONOTRANS --	MAINPRED
3.	the	DET --	NPREMOD
4.	tree	N NOM SG --	O
5.	in	PRPa --	ADVL
6.	the	DET --	NPREMOD
7.	garden	N NOM SG	PRPMOD

The functional structure assigned is fairly flat. Of course, additional pattern rules are needed for specifying attachment preferences and for collecting subconstituents. Note, e.g., that the familiar ambiguity of the final ADVL has not been resolved. It might be argued that this is not a matter of syntax alone.

Our account has been simplified in several respects. In particular, it has not been shown how lexical ambiguities should be treated. Neither have any comments been offered on the recalcitrant problems of e.g. long-distance dependencies,

ellipsis, conjunction, and intertwined clauses. However, we hope to have shown at least the general feasibility of integrating processing considerations in grammar theory.

References

Ades, A. N. & Steedman, M.J. 1982. On the Order of Words. Linguistics and Philosophy 4, 517-558.

Berwick, R. & Weinberg, A. 1984. The Grammatical Basis of Linguistic Performance. Cambridge, Mass.: The MIT Press.

Church, K. & Patil, R. ms. Coping with Syntactic Ambiguity or How to Put the Block in the Box on the Table. Ms. MIT, Laboratory for Computer Science.

Frazier, L., Clifton, C. & Randall, J. 1983. Filling Gaps: Decision Principles and Structure in Sentence Comprehension. Cognition 13, 187-221.

Johnson-Laird, P. 1983. Mental Models. Cambridge: Cambridge University Press.

Kaplan, R. M. 1972. Augmented Transition Networks as Psychological Models of Sentence Comprehension. Articifical Intelligence 3, 77-100.

Kaplan, R.M. & Bresnan, J. 1982. Lexical-Functional Grammar: A Formal System for Grammatical Representation. In Bresnan, J. (ed.), The Mental Representation of Grammatical Relations. Cambridge, Mass.: The MIT Press, pp.173-281.

Karlsson, F. 1986. A Paradigm-Based Morphological Analyzer. In Karlsson, F. (ed.) 1986: Papers from the V Scandinavian Conference of Computational Linguistics. Department of Linguistics, University of Helsinki, Publications 15.

Marcus, M. P. 1980. A Theory of Syntactic Recognition for Natural Language. Cambridge, Mass.: The MIT Press.

Marslen-Wilson, W. D. 1976. Linguistic Descriptions and Psychological Assumptions in the Study of Sentence Perception. In Wales, R.J. & Walker, E. 1976: New Approaches to Language Mechanisms. Amsterdam: North Holland, pp. 223-229.

--- 1984. Function and Process in Spoken Word Recognition. In Bouma, H. & Bouwhuis, D. (eds.) 1984: Attention and Performance X. Control of Language Processes. New Jersey: Lawrence Erlbaum Associates, pp. 125-150.

Pereira, F. C. 1985. A New Characterization of Attachment Preferences. In Dowty, D. R., Karttunen, L., & Zwicky, A.M. 1985: Natural Language Parsing. Psychological, Computational, and Theoretical Perspectives. Cambridge: Cambridge University Press, pp. 307-319.