

WPFG

---

working papers in functional grammar

WPFG no. 11  
July 1986

---

On the notion 'Functional Explanation'  
Simon Dik  
University of Amsterdam

On the notion 'Functional Explanation'

Simon Dik  
University of Amsterdam

WPFG 11 has appeared under the same title in  
Belgian Journal of Linguistics 1 (1986) 11-52

## 0. Introduction

Functional explanations of linguistic phenomena or processes of linguistic change are often made fun of by the more formally inclined among us. And indeed, many purported functional explanations are an easy target for ridicule, as they are often overly simplistic, retrodictive (predictive after the fact), and/or easily falsified. It is up to the functionalist, however, to provide a meaningful interpretation of what he understands by 'functional explanation'. In this paper I wish to take up this challenge by discussing at least some aspects of this metatheoretical problem. I will argue that in the most interesting cases functional explanations will have a rather complex structure, involving a number of distinct functional factors which may interact, and even counteract each other.<sup>1</sup>

### 1. Newmeyer's Simplified World

I start this discussion rather arbitrarily with some remarks on Newmeyer's recent monograph Grammatical theory; its limits and possibilities (1983). In this monograph Newmeyer describes a Simplified World in which there is the theory of Transformational Generative Grammar on the one hand, and the 'opposition' to this theory on the other. The opposition is easily characterized:

'Most opposing theories that allude to "grammar" at all do so with the explicit proviso that all (or most) grammatical constructs are but "convenient fictions" or a "shorthand abbreviation" for principles that are not specific to grammar' (ibid. 4).

By this characterization Newmeyer alludes to a mixed bag of 'opposing' linguists who, in his view, have in common the belief that grammatical phenomena can be directly reduced to, or deduced from, extragrammatical (i.e. communicative, pragmatic, cognitive, etc.) principles; the opponents agree in considering 'grammatical constructs as nothing more than diacritic devices for representing extragrammatical generalizations' (ibid. 5). Newmeyer then concludes:

'In the remainder of this book I will refer to the theory of "transformational generative grammar" simply as "grammatical theory". Given

that it seems to be the only nonreductionist approach to grammar, in doing so does not strike me as an unfair rhetorical ploy'. (ibid.)

I think it fair for me to reveal here, at the beginning of this discussion, that I am explicitly mentioned among that unfortunate bunch of intellectuals so astutely excommunicated from the field of grammatical theory by Newmeyer. Those who are only interested in grammatical theory as defined in Newmeyer's Simplified World may therefore stop reading here.

## 2. Why do clitics prefer the second position in the clause?

In chapter 4 of his book, Newmeyer argues in more detail against any attempt at explaining linguistic phenomena in terms of extragrammatical principles. One of his arguments runs as follows:

'Conversely, we find what appear to be true structural universals of language not admitting at all (as far as I can tell) to extragrammatical explanations. A good example is "Wackernagel's law" (....). Wackernagel was apparently the first to notice that sentential clitics always occur in "second position" - that is, after the first stressed word or phrase in the sentence (...).

There is no obvious reason why the exigencies of communication or the structure of the human perceptual apparatus would demand second position as opposed to third, fourth, or last; or, for that matter, why there should be sentential clitics at all. Those taking a reductionist approach to grammar have the obligation to demonstrate that, in some sense, the occurrence of sentential clitics in second position "reflects" the function of such clitics.' (ibid. 102-3).

McCawley (1985: 675) has already commented on the last statement in this argument by saying that:

'Several errors are packed into this short statement. A 'functional' account of grammatical phenomena need not be reductionist - e.g., an account in which functional considerations determine markedness relations among alternative linguistic forms is not, ipso facto, reductionist; functional accounts need not recognize such a thing as THE function of anything; and, in this case, it is not only the functions (N.B. plural) of clitics that come into the picture, but also the functions of all relevant details of constituent structure and

constituent order, such as the functions of 'second position'.

Further questions arise about this example of Newmeyer's. His message seems to be that the preference of clitics for second position in the clause, a preference which manifests itself in a great variety of genetically and typologically unrelated languages, is just an 'arbitrary' fact, which does not call for any explanation. For all we know, the preferred position of clitics might just as well have been third, fourth, or last.

This is a rather strange position to take, even for a generativist. Was not the highest goal for TGG to arrive at a level of explanatory adequacy, where a linguistic theory was called explanatorily adequate to the extent that it provides a principled basis for selecting the best from among a set of descriptively adequate grammars (Chomsky, 1965)? No doubt even the generativist would like to have a theory of grammar for which the preference of clitics for clause-second position is not an arbitrary fact, but a fact which 'falls out' naturally from more general principles and parameters.

For this reason, then, functionalist attempts at explaining a certain phenomenon cannot be answered by stating that the phenomenon has no explanation, but only by providing an explanation in non-functional terms, or by granting that our knowledge has not sufficiently advanced to provide such an explanation.

As a matter of fact, I happen to be on record for a theory of constituent ordering from which preferred second position for clitics falls out as a natural consequence (Dik 1978: ch. 8/9). Let us consider the relevant elements of this theory:

- (i) All languages use the first position in the clause (P1) for 'special purposes' including, at least, the placement of constituents with the pragmatic functions of Topic and/or Focus.
- (ii) All languages are sensitive to a general principle (LIPOC) which says that constituents are preferably ordered according to increasing complexity.
- (iii) Clitics count as the least complex constituents in terms of LIPOC ordering.
- (iv) P1 is not sensitive to LIPOC.

From these principles it follows that:

- (v) Clitics prefer the second position in the clause.

I take it that principles (i)-(iv) 'explain' fact (v) at least in the sense that if (i)-(iv) are part of the theory of constituent ordering, then (v) need not be separately stipulated; and also in the sense that, within a theory incorporating (i)-(iv), fact (v) is not an arbitrary fact. The question may now be posed to what extent the explanation provided by (i)-(iv) can be called a 'functional' explanation. This is dependent on the question to what extent principles (i)-(iv) can be understood as being interpretable within a functionally oriented theory of grammar. I will return below to what I understand by a functionally oriented theory of grammar.

Anticipating somewhat, I do indeed believe that the principles (i)-(iv) can be functionally interpreted, for the following reasons. Within the theory of Functional Grammar (FG), the pragmatic functions Topic and Focus are defined as relating to the organization of the information contained in the linguistic expression, as determined by the state of the pragmatic information of Speaker and Addressee at the moment of speaking. For different reasons, both Topic and Focus constituents are 'prominent' at this level of organization. Initial position is a natural position for placing communicatively prominent constituents. The universal use made of P1 in natural languages is thus a reflection of factors concerned with the functioning of linguistic expressions in communicative interaction.<sup>2</sup>

Principle (ii), as well, was hypothesized to be explainable in terms of a more general principle concerning ease of processing information (Dik 1978: 212):

- (1) It is easier to perceive, process, and store complex information when this information is presented in chunks of increasing internal complexity.

Principle (iv), finally, is explainable in functional terms as well: the assignment of Topic and Focus function to constituents is determined by factors which do not directly relate to the internal complexity of these constituents (although there may be some statistical correlation between Topic and 'lower degree of complexity').

Now it is important to stress that these various hypotheses have the merit of being potentially wrong. There is no doubt, however, about the basic functional character of the explanation: the apparently arbitrary fact of preferred second position for clitics is presented as the consequence of the interaction of several different principles, each of which can be understood in terms of either the communicative goals or the psychological constraints involved in verbal interaction.

### 3. The position and marking of conditional predications

For another example, consider the following, at first sight arbitrary, linguistic phenomena:

- (2) a. If you can't find a room, you can stay at my place
- b. You can stay at my place, if you can't find a room
- (3) a. If you can't find a room, then you can stay at my place
- b. \*Then you can stay at my place, if you can't find a room

The question is: why can then be added to the apodosis of a conditional construction when the protasis precedes, but not when it follows?

Again, one might want to give a non-answer to this question such as:

because English is the way it is, that's why. Again, I would prefer any attempt at explaining this fact over any non-attempt to do so. In this case, I have ventured the following explanation (Dik 1983):

- (i) With respect to the mutual ordering of Main and Subordinate predications, languages can be divided into those which have basic Main-Subordinate, and those which have Subordinate-Main ordering.
- (ii) Subordinate predications are typically marked by some subordinating element (either an independent subordinator or a subordinating affix). The subordinating element (sub) is a Relator, linking the Subordinate to the Main predication. Relators have their preferred position in between their two relata. Thus, we expect initial subordinators in Main-Subordinate patterns (Main-sub-Sub), and final subordinators in Subordinate-Main patterns (Sub-sub-Main).
- (iii) Conditional Subordinate predications prefer initial position independently of the basic MsS or SsM ordering of the language.
- (iv) In languages with basic MsS ordering, when no further measures are

taken, conditionals will have a sSM pattern which is non-preferred according to (ii).

- (v) We may thus expect a modification of the non-preferred pattern through (a) addition of a new Relator in medial position (sSrM), (b) loss of the initial subordinator (SM), (c) both (a) and (b) (SrM).

From this it follows that:

- (vi) then in (3 a) can be interpreted as a new relator r in the sense of (v).

Again, just like any theory, this theory may well be wrong. It has the merit, however, that - if true - it accounts for a number of at first sight arbitrary or inexplicable facts.

In what sense can this theory be interpreted as providing a 'functional explanation'? Surely not in the banal sense that the occurrence of then in (3 a) directly reflects THE function of then in terms of any extra-grammatical principle. Rather in the sense that the occurrence of then can be understood as the outcome of the interaction of a number of principles, each of which can be interpreted in terms of functional determinants of verbal interaction.

Let us consider this in more detail. Principle (i) is a specific instance of a more general principle which, roughly, says that languages basically either display Head-Modifier or Modifier-Head ordering. This choice makes for a general expectation pattern with respect to where, at different levels of linguistic organization, the Head of a construction is to be found. Given the importance of Head identification for the interpretation of linguistic expressions, this principle certainly appears to be functionally motivated. The preferred medial position of Relators (principle (ii)) can be subsumed under a more general principle of Iconicity, which says that constituents are preferably ordered in congruence with that which they express. If Relators function as linking devices between relata, then a position in between their relata is certainly the most natural position for them to take.<sup>4</sup> Principle (iii), first formulated by Greenberg (1963: 84), was interpreted by him as a specific instance of another type of Iconicity, according to which 'the order of elements in language parallels that in physical experience or the order of knowledge' (*ibid.* 103). This is based on the assumption that a Condition (protasis) conceptually precedes its



Conclusion (apodosis). This is similar to the principle which says that in temporal clauses the unmarked order language-independently reflects the sequence of the events described (a principle for which there is much psycholinguistic evidence, see Clark and Clark 1977: 78). Principle (v), finally, is based on the idea that if there are two conflicting principles A and B, such that A wins out, then therapeutic measures may be expected leading to an adaptation of the resulting construction pattern which is defined by B.

This last statement of course raises a potential objection: why should a language allow a principle A to work, although this may imply an output which is non-preferred with respect to principle B? This objection can be answered in terms of a general principle formulated by Lightfoot (1979), following Hermann Paul. According to this principle, languages may be said to apply 'local optimization' with respect to a given principle (cf. Ronneberger 1985), even though optimization with respect to one principle may lead to less highly evaluated structures in the light of some other principle. After that, therapeutic steps may be taken to remedy this outcome. In Lightfoot's terms (*ibid.* 149): 'Grammars practice therapy rather than prophylaxis'.<sup>5</sup> From this it follows that it is not incompatible with a functionalist conception of language to assume that a certain conjunction of factors may create linguistic expression types which are less functional or even downright dysfunctional with respect to a given functional principle. The prediction would then be that therapeutic measures may be expected to be taken wherever such situations arise.

#### 4. Towards a non-simplistic theory of functional explanation

What we learn from these examples is that a functional explanation of grammatical phenomena will typically not be based on an assumption of simple form-function correlations, but will instead involve a network of interacting requirements and constraints, each of which may be understood in functional terms itself, but which interact in complex ways and in a certain sense 'compete' for recognition and expression in the final design of linguistic expressions.

The idea that cultural artefacts can be viewed as drawn from a limited set of possible solutions to a certain functional problem has a long history in functionally oriented anthropology, sociology and linguistics. Depending on one's point of view, one may either stress the limitations imposed by the problem, or the non-deterministic freedom of finding solutions within these

limitations. Malinowski (as quoted in Merton 1967: 107) expressed the idea as follows:

'Given a definite cultural need, the means of its satisfaction are small in number, and therefore the cultural arrangement which comes into being in response to the need is determined within narrow limits'.

The theoretical sociologist Merton (1967 (=1949): 87) rather stressed the freedom of finding alternative solutions:

'Just as the same item may have multiple functions, so may the same function be diversely fulfilled by alternative items. Functional needs are here taken as permissive, rather than determinant, or specific social structures. Or, in other words, there is a range of variation in the structures which fulfill the function in question'.

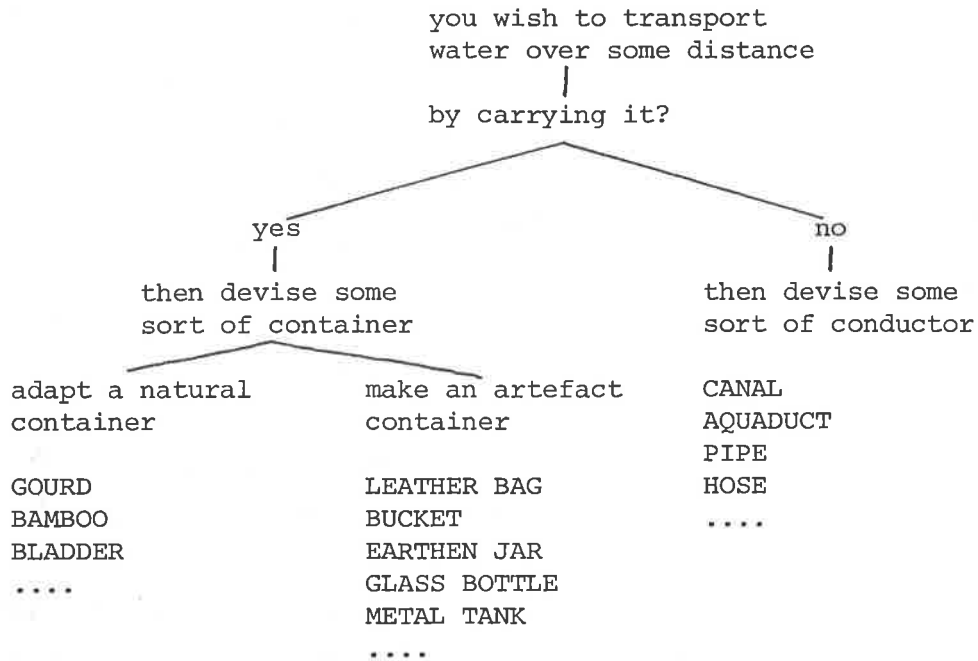
Greenberg (1957: 78) stressed the differences in the degree of efficiency of different functional alternatives, and the need to appeal to historic factors in explaining why one rather than another functional alternative is preferred at any given point of time:<sup>6</sup>

'The same function can always be performed by alternate structures, though only a limited variety of such structures is possible. Hence the mere fact that a particular function is performed is not sufficient to account for the specific structural arrangement. We therefore consider it in relation to its historical antecedents. We also investigate functional alternatives on a comparative basis, in order to discover the range of possible structures capable of fulfilling a given function.

The consideration of functional alternatives brings with it, inevitably, the related concept of functional efficiency. We can compare diverse structures with the same or similar functions with regard to the effectiveness with which the function is performed'.

Let us illustrate the idea of a limited set of solutions to a given functional problem by means of the following example. Probably all human cultures have devised certain instruments for transporting water over some distance. These instruments display a variety of forms and structures, and it is certainly not the case that these forms and structures are in any sense directly determined by their function. However, they all have certain minimal criterial features which make them fit for solving precisely this type of problem. Consider the following survey:

(4)



At first sight such diverse instruments as a GOURD, a TANK, and a HOSE have little or nothing in common. Still they belong together as members of a set of possible solutions to the water transport problem, a problem that defines a limited solution space. Suppose a new human tribe is discovered somewhere in the middle of the Amazon jungle. Suppose that (as is probable) this tribe has solved the water transport problem. Then we may almost be certain that that tribe will have one or more of the solutions of (4) among its cultural artefacts. Likewise, we may confidently assume that the tribe in question has solved the human communication problem and that, among its cultural artefacts, it will possess one variant of that intricate device called natural language.

The example presented above illustrates that the particular solution to a given cultural problem developed in any specific case is dependent on at least the following 'functional prerequisites':

- (a) the aims, goals, or purposes to be reached;
- (b) the means or materials used for implementing the solution;
- (c) the circumstances in which the solution is to be implemented.

These types of functional prerequisites will return in our attempt at formulating a non-simplistic theory of 'functional explanation'.

5. A non-simplistic explication of 'functional explanation'

We may now set up the following framework for explicating the notion 'functional

explanation':

- (I) A functional explanation of a (synchronic or diachronic) linguistic phenomenon is a statement in which that phenomenon is shown to follow from one or more principles which crucially refer to any of the functional prerequisites imposed on natural languages.
- (II) Functional prerequisites pertain to the extralinguistic requirements and constraints imposed on natural languages.
- (III) These requirements and constraints can be divided into:
  - (i) the aims or purposes for which natural language expressions are used;
  - (ii) the means by which natural languages are implemented;
  - (iii) the circumstances in which natural languages are used.
- (IV) The primary aim of natural languages is the establishment of inter-human communication; other aims are either secondary or derived.
- (V) The primary means by which natural languages are implemented is the human vocal-auditory tract; other means are either secondary or derived. The use of the vocal-auditory tract imposes physiological and psychological constraints on natural language design and change.
- (VI) The circumstances in which natural languages are used can be distinguished into:
  - (i) physical circumstances;
  - (ii) socio-cultural circumstances;
  - (iii) linguistic circumstances.
- (VII) The functional prerequisites imposed on natural languages allow for a limited range of possible solutions. This 'solution space' coincides with the notion 'possible natural language'.
- (VIII) Indirectly, the functional prerequisites define the notion 'possible linguistic change': a possible linguistic change must lead from one possible language to another possible language.
- (IX) Functional prerequisites may impose divergent requirements on linguistic design and linguistic change: a design/change feature which is positively evaluated with respect to prerequisite  $F_i$  may be negatively evaluated with respect to requirement  $F_j$ .
- (X) There is thus continuous competition between different functional prerequisites; the actual synchronic design of a language is a compromise solution, a precarious balance in efficacy with respect to different functional prerequisites.
- (XI) This means that languages may change, oscillating between different

- compromise solutions, even if the functional prerequisites remain constant. This may be called internally motivated change.
- (XII) In contrast to this, externally motivated change is involved when languages respond structurally to changes in the functional prerequisites.
- (XIII) As a consequence of the potential divergence of functional prerequisites, a change in favour of  $F_i$  which at the same time disfavors  $F_j$  may be followed by a further 'therapeutic change' repairing the damage done to  $F_j$ .
- (XIV) The operation of linguistic change is 'local' in the sense that it is unusual for a change favouring  $F_i$  to be blocked in order to avoid adverse effects with respect to  $F_j$ .

Linguists obviously differ in what importance they are prepared to assign to functional prerequisites in explaining linguistic design and change. Formally inclined theorists tend to downplay the role of functional explanations. Functionally inclined theorists will tend to maximize the role of such explanations. From such a functional point of view, we could add the following methodological principle to the above:

- (XV) Saying that a certain feature of linguistic design or change cannot be functionally explained is tantamount to saying that we have not yet been able to find a functional explanation for that feature.

In the following subsections I will further specify and discuss some of the principles formulated above.

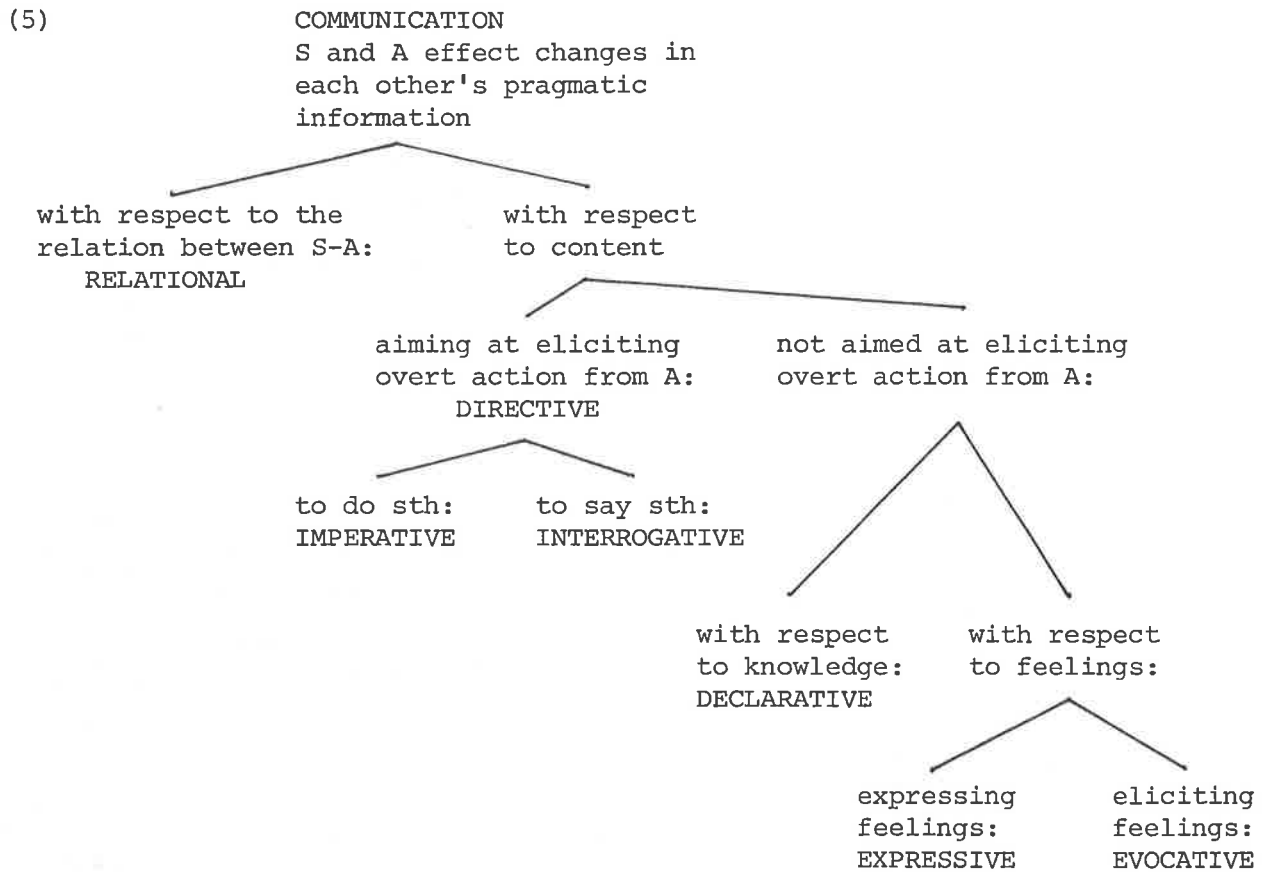
### 5.1. Aims and purposes

The aims and purposes for which natural languages are used are commonly discussed under the heading of 'functions' of language, and there is rather an extensive body of literature about what these functions precisely are, how many of them should be distinguished, and what relative importance they have with respect to each other. There is no opportunity here for reviewing this whole discussion. For recent analytic surveys, see Nuyts (1984, 1985). I will here restrict myself to some brief remarks on what, in my opinion, are the most important distinctions to be made in this area.

I start from the assumption that the establishing communication is the central or primary function of language. Other uses are either of secondary importance,

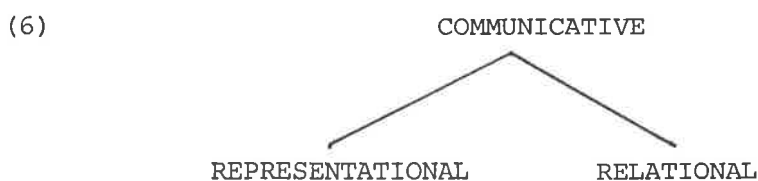
or can be understood as derived forms of communication. For example, it is often held that the function of language as an instrument of thought or symbolic representation is primary to its communicative function. But if this were true, it would certainly be a remarkable coincidence that different individuals belonging to the same social group use essentially the same instrument for these supposedly non-communicative functions. In other words, the essential identity of language across individuals within a social group can only be understood if the social, interactive use of language is taken as primary. There is, moreover, an interesting and plausible view according to which 'thinking in language' can be interpreted as a form of internalized social interaction of the self with the self (Mead 1934, Vygotsky 1962). In that view, the cognitive uses of language are regarded as derivative with respect to the communicative uses, rather than the other way around.

Communication can be defined as a social activity through which individuals mutually effect changes in each other's pragmatic information. By pragmatic information I understand the full body of knowledge, beliefs, assumptions, and feelings available to an individual. Pragmatic information can be divided into (i) general information: long-term information concerning the world and other possible worlds; (ii) situational information: information derived from what the individual perceives and otherwise experiences within the situation in which the interaction takes place; (iii) contextual information: information derived from the linguistic expressions which have been exchanged before any given moment in verbal interaction (see Dik 1978: 128 for further discussion). Within these general functions of effecting changes in each other's pragmatic information, a number of distinct aspects of communicative function may be distinguished, depending on what feature of pragmatic information these changes apply to and what further purposes these changes are directed at. One way of analysing these different aspects of communication is represented in the following diagram:



It is to be noted that the capitalized terminal nodes in this tree are to be regarded as aspects of communication which may well be conjunctively relevant (in different combinations) in one and the same linguistic expression. For example, S may well use an expression which is meant to add knowledge to A's pragmatic information (DECLARATIVE), while at the same time expressing S's emotion concerning this information (EXPRESSIVE), asking for A's help or sympathy (DIRECTIVE), and thus also changing the relation between A and S (RELATIONAL). In fact, it may be taken as axiomatic that most natural language expressions simultaneously function at several levels of the functionality tree displayed in (5).

Cutting up the cake in a somewhat different way, we can also make the following more global distinctions:<sup>7</sup>



In other words, any linguistic act has a representational aspect (it is 'about' something), and a relational aspect (it is a way in which S relates to A).

This distinction is important, since it is especially the representational potential of natural language which distinguishes it most clearly from other forms of human communication. We can say that any natural language has somehow solved the problem of how to represent an unlimited set of 'states of affairs', where by state of affairs I mean the conception of something which can be the case in some world. States of affairs are constructed out of entities by certain properties and relations. Entities are linguistically represented by terms, properties and relations by predicates, and states of affairs are therefore represented by predications, consisting of a predicate applied to one or more terms. Predications constitute the representational core of linguistic expressions in any natural language. They can be embedded within different types of relational frames which signal which modification S wishes to effect in A's pragmatic information with respect to the state of affairs designated by the predication. The essential point about the representational capacity of natural languages is that there is no limit to the types of state of affairs that can be represented: any entity and any property or relation that can be conceived of at all can also be expressed in natural language. This universal function is probably the most important determinant of the possible forms that natural languages can take.

## 5.2. Means

All natural languages use the human voice as their basic vehicle. Communicative systems which use visual rather than auditory signals are either derivative in relation to the vocal mode, or used in circumstances in which the vocal mode cannot be used, or both. The primarily vocal character of natural languages imposes a number of obvious physiological constraints on their organization. On the one hand, the forms of a language should be such that they can be produced by the articulatory system; on the other hand, they must be such that they can be clearly perceived by the auditory system. Less obviously, the vocal mode imposes limitations on the dimensionality of linguistic form: linguistic forms are subject to the principle of linearity (Saussure 1916) in the sense that their capacity for transmitting different forms of information is at the same time extremely limited. The fact of linearity in itself is not so exciting. What is significant, however, is the various ways in which the linear organiza-



tion of linguistic forms are exploited in the grammatical organization of natural languages.

More important than the physiological constraints are the psychological constraints imposed on a natural language as a communicative system using the vocal medium. These constraints can be divided into the following types:

- (7)           (i)     constraints on production
- (ii)    constraints on perception
- (iii)   constraints on comprehension
- (iv)    constraints on memorization
- (v)     constraints on acquisition

In other words, linguistic expressions should ideally be such that they are easily produced, easily perceived, easily comprehended, easily memorized and easily learned. By 'perception' I understand the interpretation of the incoming signal as a string of words ('What did you say?'); by 'comprehension' I understand the interpretation of that string of words in terms of communicative intensions ('What do you mean?'). These various psychological constraints on linguistic expressions do not always pull in the same direction. For example, a form which is to be easily produced should preferably be short and easy to perceive. In certain cases, then, the various psychological constraints may define divergent preferences between which some kind of compromise must be established.<sup>8</sup>

### 5.3. Circumstances

A third set of functional prerequisites depends<sup>d</sup> on the circumstances in which natural languages are used. We may here distinguish between:

- (i)     physical circumstances;
- (ii)    socio-cultural circumstances;
- (iii)   linguistic circumstances.

#### 5.3.1. Physical circumstances

The physical circumstances in which natural languages are used have a number of obvious effects on their organization, but as far as present evidence goes,

these effects would seem to be mainly confined to the vocabulary. Trivially, a language spoken by people who have never seen a technical device will have no word for that device. A little less trivially, a language spoken in circumstances where certain realia (like 'snow', or 'rice') are especially important will have many different words for subcategories of these realia. Even less trivially, there are some indications that deictic systems involving the vertical dimension ('up' vs. 'down') are confined to languages spoken in mountainous areas. Here, the physical circumstances seem to have some impact on what could be called the fringe of grammatical organization. There is little or no evidence, however, that physical circumstances have any influence on more basic features of grammatical organization. Note that physical circumstances shade off into cultural circumstances, since languages do not arbitrarily label any old feature of the outside physical world, but especially and primarily those which are of special cultural importance to their speakers.

### 5.3.2. Socio-cultural circumstances

Our ways of speaking are in many ways dependent on the socio-cultural circumstances in which we speak. These dependencies can be described, on the one hand, in terms of aims that we wish to reach in communication: whatever the ultimate communicative goal, our speaking necessarily also has a relational aspect: in speaking we aim, first, at establishing and maintaining contact with our Addressee; second, we signal the type of social relationship which (from our point of view) exists or should exist between ourselves and our Addressee. On the other hand, the socio-cultural determinants of linguistic behaviour can be described in terms of constraints imposed on that behaviour by the socio-cultural circumstances in which it is implemented: we cannot say just anything in any type of social environment.

Such social constraints on our ways of speaking could be described in terms of the following factors:

- (i) Politeness and deference;
- (ii) Solidarity;
- (iii) Prestige and stigmatization.

Politeness and deference can be seen as aspects of the social distance between Speaker and Addressee. This social distance is the resultant of the factors

'hierarchical relation between S and A' and 'familiarity of S and A' (see Brown and Gilman 1960). The greater the distance between S and A in terms of these factors, the more formal and polite will be their linguistic usage. The expression of formality and politeness can be characterized in terms of two main expression devices:

(a) Diminution vs. Augmentation

Many types of polite expression can be understood in terms of the metaphors in (8):<sup>9</sup>

- (8) a. Important is HIGH, unimportant is LOW
- b. Important is BIG, unimportant is SMALL

(b) Avoidance of directness

The greater the distance between S and A, the less 'direct' will be S's linguistic behaviour in a number of respects:

- uses of indirect forms of address,
- uses of indirect types of speech act,
- avoidance of direct mention of certain items,
- use of euphemisms to refer to certain items,
- avoidance of directly challenging the other's opinions.

The expression of politeness and deference thus typically leads to greater length and complexity of expression (cf. Haiman 1983). In a sense, then, the length and complexity of S's expression symbolizes the social distance between S and A.

Solidarity is involved in those cases in which certain forms of linguistic behaviour signal internal group identity and distinguish 'us' from 'them'. Solidarity, taken in this sense, thus leads to linguistic differentiation rather than to unification.

Prestige and stigmatization are involved in all those cases in which certain forms of linguistic behaviour are adopted or retained in order to acquire the social prestige attributed to these forms, or in order to avoid the stigmatization which these forms elicit within the social reference group.

### 5.3.3. Linguistic circumstances

By the linguistic circumstances in which a language is used I mean the other languages which are used in the same communicative community. This

involves all factors pertaining to language contact, substratum/superstratum/adstratum effects, code switching, bilingualism, and other forms of language interference. It is quite clear that such factors are extremely important in the shaping of linguistic design and in the instigation of linguistic change. But they are secondary to the other functional constraints, in that they only come into play when two or more languages come in touch with each other. All<sup>^</sup>̄pervading though such language contacts are, their effects should be distinguished from those which operate on any single language as such. In the context of the present paper I will not be further concerned with the effects of the linguistic circumstances on the design and change patterns of languages.

#### 6. Functionally motivated features of linguistic design and change

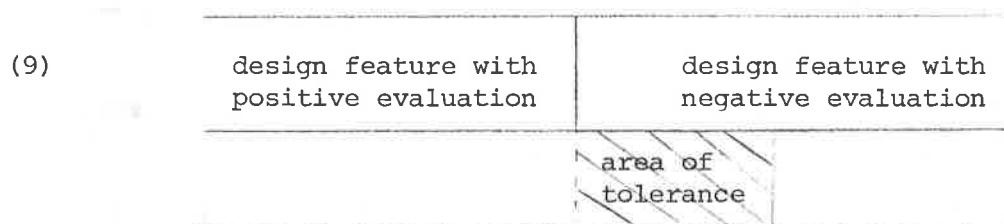
There is a great variety of aspects of linguistic design and change which can be plausibly accounted for by functional explanations. In fact, this variety is so extensive that I can hardly do justice to it in the context of the present paper. I will therefore restrict myself to an enumeration of the most important functionally motivated features at different levels of linguistic organization. It is expedient to discuss these features at two main levels of organization, corresponding to Martinet's concept of 'double articulation':<sup>10</sup>

- (i) the level of phonological articulation, where by 'phonological' I mean the conjunction of the levels of phonetic and phonemic organization of linguistic expressions;
- (ii) the level of morpho-semantic articulation, where by 'morpho-semantic' I mean anything pertaining to the mutual relations between linguistic forms and linguistic meanings.

In discussing the various functionally motivated features of linguistic design, I will often refer to features which are 'preferred', 'favoured', 'more efficient' or 'simpler' with respect to one or more functional prerequisites. It is important to note at the outset that none of these preferences has the value of an absolute law: in all cases, languages tolerate grammars or linguistic expressions with the non-preferred property up to a certain level of tolerance. The principle underlying this fact may be termed the Principle of Limited Tolerance. This principle makes it more difficult to formulate functional explanations in a hard and fast categorical form. There are hardly any situations in which we can say: expressions with

disfavoured property P are categorically excluded in natural languages. In most cases we will rather have to content ourselves with a statement to the effect that expressions with property P, although non-preferred, are tolerated up to a certain point. This means that when such expressions do occur in a language, we may expect measures to eliminate them, but only when they have cumulated up to a certain point approaching the tolerance limit. The matter is further complicated by the fact that different preferences may each have their own tolerance level, and that presumably trespassing into the non-preferred area of feature P may become more painful when accompanied by trespassing into the non-preferred area of feature Q.

The Principle of Limited Tolerance may be illustrated by the following diagram:



The shaded area of tolerance in this diagram symbolizes the degree to which non-preferred design features may occur in natural languages before measures arise to eliminate them. Although the Principle of Limited Tolerance greatly complicates any attempt at providing functional explanations for linguistic phenomena, the principle itself is easily recognized as one which has a much wider bearing on human behaviour. There are after all many everyday circumstances in which we tolerate a certain degree of unpleasant experience before we actually take measures to relieve ourselves of these experiences.

### 6.1. The phonological level<sup>11</sup>

Linguistic organization at the phonological level is monitored by at least the following functional principles which, in part, may operate at cross-purposes with each other:

- (i) ease or economy of articulation;
- (ii) degree of perceptibility;
- (iii) ease of acquisition and memorization;
- (iv) acquisition of prestige and avoidance of stigmatization.

6.1.1. Ease of articulation

Ease of articulation can be analyzed into the following three functional principles:

- (10) a. Shorter forms are preferred to longer forms;
- b. Those forms are preferred which require a simpler route through articulatory space;
- c. Those forms are preferred which require a lower degree of articulatory energy.

These three principles imply that, if phonological organization is left at the mercy of economy of articulation, linguistic forms will become shorter, more centralized in articulation, simpler with respect to articulatory movements, and less tense with respect to manner of articulation.

The operation of these principles is witnessed most clearly in two types of facts, which are probably related in a cause-and-effect relationship:

- (a) the forms used in allegro styles are simpler with respect to (10 a-c) than those used in lento styles;
- (b) many sound changes take the form of simplifications with respect to (10 a-c).

With respect to (b), the following processes of sound change may be mentioned:

- (i) shortening through loss of segments or syllables,
- (ii) cluster simplification through
  - resyllabification,
  - loss of segments from clusters,
  - vowel insertion.
- (iii) assimilation processes,
- (iv) weakening of
  - consonants, typically in intervocalic or word-final position,
  - vowels, typically in unstressed syllables.

These processes define the most important factors responsible for the continuous erosion of linguistic forms through time.

The relationship between (a) and (b) above can be established on the hypothesis that formal erosion starts in allegro styles. The relevant sequence

of events could be represented as follows (with  $f_o$  for the original form and  $f_r$  for the reduced form):

(11)		lento style	allegro style
	Stage 1	$f_o$	$f_o$
	Stage 2	$f_o$	$f_o, f_r$
	Stage 3	$f_o$	$f_r$
	Stage 4	$f_o, f_r$	$f_r$
	Stage 5	$f_r$	$f_r$

The essential steps would thus be a natural reduction from  $f_o$  to  $f_r$  in allegro style, a generalization to  $f_r$  in allegro style, and the adoption and subsequent generalization of  $f_r$  in lento style (cf. Greenberg 1966: 516, Koefoed 1978).

#### 6.1.2. Degree of perceptibility

In certain respects the factors 'ease of articulation' and 'ease of perception' go hand in hand: certain forms which are easier to produce are also easier to perceive, and vice versa. For example, cluster simplification may make a form easier to produce, and at the same time easier to perceive. More generally, a number of phonological processes would seem to have the common target of arriving at maximally perspicuous syllabic patterns built on 'preferred' syllable structures such as CV, CVCV, CVCVCV, ... etc.

Certain other processes monitored by 'ease of articulation', however, run counter to 'ease of perception'. This is especially so in the case of various shortening processes which may either be confined to or at least have their origin in allegro styles.<sup>12</sup> Consider for example the various lento to allegro forms of I do not know in English:

- (12) a. ai du nɔt now  
b. ai dɔwnt now  
c. ai dɔw̃ now  
d. ai dɔnow

It is clear that as we progress from (12 a) to (12 d), the phonetic exponent of negation becomes less and less perceptible until, in (12 d), it is very difficult to say which part of the form is responsible for the negative element at all. In such situations we may expect measures to be taken to restore the perspicuity of the forms in question.

Apart from the cases of syntagmatic perceptibility mentioned, ease of perception would also seem to be responsible for those paradigmatic aspects of phonological systems which have traditionally been described in terms of 'maximal degree of contrast', or 'maximal perceptual distinctness'. This factor explains such facts as that, when a language has only three vowels, these typically take the shape of /i/, /u/, and /a/; i.e. occupy the areas farthest apart in 'phonological space'. Since Jakobson (1941), a widely accepted theory has been that 'maximal contrast' not only explains the typological distribution of phoneme systems, but also the order in which the different parts of such systems are acquired in first language acquisition.

#### 6.1.3. Ease of acquisition and memorization

If Jakobson's theory is correct, then ease of acquisition will be positively loaded by ease of perception, and thus by ease of articulation to the extent that the latter does not work at cross-purposes with ease of perception. I do not know of any example of a phonological shape which is easy to acquire, even though it is difficult to produce and/or perceive. This would mean that 'ease of acquisition' would not play an independent role as a functional prerequisite at the phonological level.

Whether 'ease of memorization' plays such an independent role I do not know either. On the face of it, one would presume that certain phonological shapes are easier to remember than others, and that this would be a function of certain design features of such forms, which may or may not correlate with 'ease of perception'. I am not aware of any psychological research on quality of recall as a function of the phonological shape of the items to be recalled.

#### 6.1.4. Acquisition of prestige and avoidance of stigmatization

The importance of these social factors in explaining patterns of language change has been made abundantly clear in the work of Labov (1966 etc.). Let us just recall some of his most important explanatory concepts:

- (i) change from above: a certain feature of pronunciation originates in the higher social regions, thereby acquires prestige, is taken over by lower classes and thus finally spreads through the linguistic community.



- (ii) hypercorrection: a lower class (typically, higher middle class) 'overshoots' the mark of imitating the higher class in using prestigious form in higher frequency, more 'exaggerated' articulation, or wider/different contexts than the higher class itself.
- (iii) change from below: a certain 'natural' phonological change gradually spreads through the community from its place of origin, finally getting through without being noticed; or, if it comes from lower or less prestigious regions, it may be stigmatized, avoided by those in search of prestige, and thus blocked in its spread through the community.

It is important to note that changes of type (i) owe their effect to their being used by the more prestigious class, not to any intrinsic phonological property. This means that changes of this type may in fact favour such forms as are less positively valued with respect to parameters such as ease of production and perception. We thus find that at the phonological level, psychological prerequisites will tend towards simplification of phonological form (even though what is simpler for one purpose is not necessarily simpler for another), whereas social prerequisites may actually favour more complex phonological forms, if these have been invested with prestige through being used by some group which it is expedient to imitate.

## 6.2. The morpho-semantic level

At the morpho-semantic level, there are also a considerable number of aspects of linguistic design and change which can be accounted for in terms of the functional prerequisites imposed on natural languages. First of all, there is the obvious communicative requirement that any language should be able to fulfil both the representational and the relational functions as discussed in 5.1. above. For the morpho-semantic level this implies that this level should allow for the construction of a practically infinite set of complex expressions, to match all the (complexes of) states of affairs that need to be represented. On the other hand, the finiteness of the human organism requires that this infinite set of expressions be constructed from a finite set of basic elements through the application of a finite set of rules. All natural languages have solved this problem through recursively combining a finite set of basic elements into a finite set of complex expressions.

The set of basic content elements (the basic vocabulary) differs between languages, between dialects and jargons, and even between idiolects. What is expressed in one basic lexeme in one system may only be expressible through a combination of elements in another system. The size of the basic vocabulary is subject to various interactions between what Langacker (1977) has called 'signal complexity' and 'code complexity'.<sup>13</sup> Compare the following expressions in English and Dutch:

- ( 13) a. John was punished for speeding
- b. John was punished for driving too fast
- ( 14)     Jan werd bestraft voor te hard rijden

Through the availability of the basic lexeme speed = 'drive too fast' English has a choice between the shorter expression (13 a) and the longer expression (13 b). Dutch only has (14), corresponding to the latter choice in English. The signal simplicity of (13 a), however, is bought at the price of greater complexity of the lexical code, since English has one more basic lexeme than Dutch has in this area.

More generally we can say that a larger set of basic lexemes (in some semantic field) allows for more compact linguistic expressions, and conversely a less articulated lexical code leads to longer linguistic expressions. A larger vocabulary means a heavier load on memory; longer linguistic expressions imply a lower degree of communicative efficiency. There is probably an area of tolerated balance between these two, within which languages may vary between the one and the other extreme.

There are different indications that the size and the articulation of basic vocabulary is dependent on various functional prerequisites:

- (i)     the number of distinct basic lexemes in a given semantic field is a function of the communicative importance of that field for the (sub-) culture in question.
- (ii)    the length of the basic element is inversely proportionate to its frequency of use (Zipf's law). More specifically, (a) items with high text frequency tend to be monosyllabic, (b) longer expressions tend to get shortened as they get more frequently used.
- (iii)   typological hierarchies of lexical coding tend to correlate with frequency of usage of the relevant items.

The first two principles are well-known from the literature and need no further comment in the present context. Principle (iii) does need some explanation. By a 'hierarchy of lexical coding' I understand the type of hierarchy discovered by Berlin and Kay (1969) for basic colour terms:

$$(15) \quad \left\{ \begin{array}{l} \text{black} \\ \text{white} \end{array} \right\} > \text{red} > \left\{ \begin{array}{l} \text{green} \\ \text{yellow} \end{array} \right\} > \text{blue} > \text{brown} > \left\{ \begin{array}{l} \text{purple} \\ \text{pink} \\ \text{orange} \\ \text{grey} \end{array} \right\}$$

This hierarchy was given the following typological interpretation:

- (a) all languages have basic colour terms for 'black' and 'white';
- (b) if a language has a basic colour term for any colour in (15), then it also has basic colour terms for all the preceding ones;
- (c) no language has more basic colour terms than the eleven specified in (15).

Such hierarchies as (15), however, not only have a cross-linguistic typological significance; they can also be shown to be relevant for differentiating between the items concerned within a language. The general rule would seem to be that the later an item occurs in a hierarchy such as (15), the lower its text frequency within a language. Consider, for example, the relative text frequency of the Dutch colour terms corresponding to (15):<sup>14</sup>

$$(16) \quad \left\{ \begin{array}{l} 135 \text{ zwart} \\ 140 \text{ wit} \end{array} \right\} > 112 \text{ rood} > \left\{ \begin{array}{l} 66 \text{ groen} \\ 33 \text{ geel} \end{array} \right\} > 53 \text{ blauw} > 40 \text{ bruin} > \left\{ \begin{array}{l} 50 \text{ grijs} \\ 16 \text{ oranje} \\ 13 \text{ paars} \\ 13 \text{ rose} \end{array} \right\}$$

From (16) we see that the text frequency of the colour terms gradually decreases almost fully in accordance with the hierarchy. On the assumption that text frequency is a function of communicative need, the conclusion may be that lexical coding, i.e. the question whether a certain meaning will be coded in an arbitrary basic lexeme of its own, is (to some degree) co-determined by communicative need.

#### 6.2.1. Psychological factors

From one point of view it would be easy if each basic linguistic sign were an icon of its meaning, in the sense of directly signalling, via its form,

some essential feature of its meaning. However, given the requirement of universal representational capacity, the limitations imposed on the human ability to vocally mimick things meant, and the fact that many things meant cannot be mimicked at all, this is not a feasible solution. Therefore, the basic linguistic sign is in principle arbitrary (Saussure 1916), and those linguistic signs which have some degree of motivation (onomatop<sup>e</sup>ia, sound symbolism) form a distinct minority in any natural language.

As Saussure already stressed, however, complex linguistic expressions consisting of combinations of basic signs have a non-arbitrary, motivated structure. It may be an arbitrary fact of English that a book is called a book and a store a store, it is certainly not an arbitrary fact that a shop where books are sold is called a bookstore. Complex linguistic expressions are built on 'patterns', which help A decode the meaning of the complex expression in terms of the meanings of the component basic signs. Remarks on the non-arbitrary nature of complex linguistic signs may be found scattered through the literature ever since the analogists fought the anomalists in classical antiquity. Recently, new interest in the 'semiotic principles' underlying the motivation of complex signs has arisen in several foci of functionally oriented theory formation.

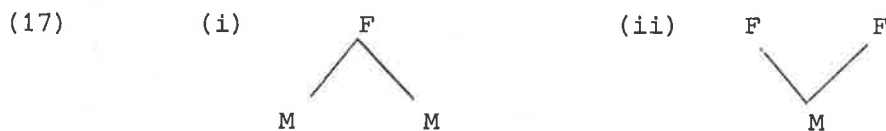
At the morphological level, we find this emphasis in the work of the initiators of 'Natural Morphology' (e.g. Mayerthaler 1981, Dressler 1983); at the syntactic level we find it in several studies by John Haiman (1980, 1983) on 'iconicity', leading up to a recent volume (Haiman 1985) devoted to 'iconicity in syntax'. Much of what follows is based on these two sources which, though more or less independently developed, have many points in common. In general we can say that comprehension, learning, and memorization are facilitated by the principles of Isomorphism and Patterning, where Patterning may be either Arbitrary or Iconic. Let us look at each of these principles in turn.

#### 6.2.1.1. Isomorphism

By Isomorphism (also called biuniqueness, transparency, or 'Humboldt's Principle') we understand the principle that each linguistic form should consistently express one and the same meaning, and that each meaning should be consistently expressed by one and the same linguistic form.<sup>15</sup> Note that

Isomorphism is not a 'law' of linguistic organization: every language has many form-meaning relations which violate Isomorphism one way or another. What Isomorphism says, then, is that form-meaning relations which conform to the one form - one meaning correlation will be favoured, and those which do not conform to this correlation will be disfavoured. By 'disfavoured' I mean that such relations will be avoided where possible, and preferably eliminated where they arise.

Isomorphism disfavours two types of form-meaning relation, each of which can be given a paradigmatic and a syntagmatic interpretation:



(i) One form is associated with more than one meaning.

- a. Paradigmatic interpretation: the form is ambiguous or homonymous. Isomorphism disfavours such ambiguity/homonymy.
- b. Syntagmatic interpretation: the form is the fused ( $p_{\text{r}}^{\text{t}}m_{\text{a}}\text{nteau}$ ) expression of two or more distinct semantic/grammatical units of meaning. Isomorphism disfavours such fused expression of distinct meanings.

(ii) One meaning is associated with two or more distinct forms.

- a. Paradigmatic interpretation: the forms are synonymous, or allomorphic realizations of one and the same unit of meaning. Isomorphism disfavours synonymy and allomorphy.
- b. Syntagmatic interpretation: a single meaning is spread over several distinct forms within the same linguistic expression. This is the case in idiomatic expressions: complex forms with a single unified meaning. Isomorphism disfavours such idiomatic expressions.

All languages tolerate a certain quantity of non-isomorphic form-meaning relations. How, then, can it be maintained that these are disfavoured?

The main point is that there are various diachronic processes which can be interpreted in terms of elimination of the disfavoured constellations or, in more exceptional cases, in terms of avoidance of such constellations. Corresponding to the four types of violation of Isomorphism distinguished above, we find the following restoration and avoidance strategies:

(A) When a form is ambiguous between two meanings, either one of the meanings may be eliminated, or a new form may be introduced as a carrier

of one of the meanings. In the latter case one may speak of the resolution of a 'homonymic clash', a phenomenon well-known since Gilliéron. Note that a high degree of ambiguity may be tolerated in natural languages, thanks to the disambiguating effect of context and situation.

(b) A single form which is the fused or synthetic expression of several distinct meaning components may be replaced by an innovative analytic expression which is more faithful to the principle of Isomorphism. The replacement of old synthetic forms by innovative analytic ones is one of the most common processes of language change.

(c) When two forms are synonymous, either one of the forms may be eliminated, or the meanings may drift apart and end up as distinct semantic units. Allomorphy may be eliminated through analogical extension of allomorphs through the paradigm (paradigmatic levelling).<sup>16</sup>

(d) Idioms may get fixed into phraseological units, then into unanalyzable basic signs.

Examples of the processes mentioned under (A)-(D) abound in the literature on historical linguistics. These processes receive a unified explanation when they are considered in the light of the Principle of Isomorphism, as long as this principle is interpreted within the framework of what we have called the Principle of Limited Tolerance.

#### 6.2.1.2. Arbitrary Patterning

It is a well-known fact that any form of patterning imposed on sets of individual items facilitates the perception and the memorization of these items. By patterning I understand any form of consistent organization imposed on linguistic items, as determined by some organizing principle. Patterning (and the organizing principle underlying it) may be either arbitrary or iconic. Arbitrary Patterning is involved when the underlying principle is not determined by the content of the linguistic items. Iconic Patterning is involved when the underlying principle is determined by the content of these items. In this section, we consider some examples of Arbitrary Patterning.

As we saw in section 3 most languages display a considerable degree of consistency in having either Modifier-Head or Head-Modifier ordering across different categories.<sup>17</sup> In itself, the choice of either the one or the other ordering would seem to be arbitrary. The Patterning lies in the consistency

with which it is applied. This consistency helps the Addressee in identifying the Head in any of the relevant constructions. This is essential for the Addressee to be able to decode the meanings of complex forms, as can be seen in the following examples:

- (18) a. drawing pen = 'pen which has something to do with drawing'  
b. pen drawing = 'drawing which has something to do with a pen'

The fact that speakers of English can correctly identify the meanings of these expressions, even if they have never heard them before, is due to the principle, arbitrary in itself, which says that the Modifier precedes the Head in this sort of combination in English. The Patterning is arbitrary in that another language, such as French, may have exactly the opposite ordering:

- (19) a. crayon à dessin  
b. dessin au crayon

For another example of Arbitrary Patterning, consider the position of the Verb in relation to its major arguments. The fact that we find VSO, SVO, and SOV languages shows that the position of the V is not determined by iconic or 'natural' considerations. The fact that individual languages are (to some extent) consistent in their positioning of V with respect to S-O is thus a matter of Arbitrary Patterning. It is arbitrary patterns of this type which allow the Addressee to apply 'perceptual strategies' in the sense of Bever (1970). These strategies might be more appropriately called 'interpretation strategies' or 'comprehension strategies', since their operation clearly goes beyond the perception of linguistic expressions in the more restricted sense in which we have used that term in the present paper.

Given the forms of Arbitrary Patterning mentioned above, speakers of English may apply the following interpretation strategies:

- (20) Interpret a sequence of N + N as Modifier + Head  
(21) Interpret a sequence of N + V + N as Subject + Predicate + Object

Arbitrary Patterning thus allows the Addressee to act on the assumption that certain standardized interpretation conditions will be fulfilled, and thus help him to quickly arrive at the intended interpretation.

### 6.2.1.3. Iconic Patterning

In the case of Iconic Patterning the underlying organizing principle is determined by the content of the linguistic expressions concerned. For example, the principle mentioned in section 3, according to which conditional constructions have the unmarked order Protasis + Apodosis in all languages, regardless of their basic ordering principles, is a matter of Iconic Patterning. It is clear that whereas Arbitrary Patterning is a matter which is fixed for each language separately, Iconic Patterning, if relevant at all, must be the same across all languages.

The following examples of Iconic Patterning (most of them due to Haiman 1980, 1983) may be mentioned:

#### (i) Reduplication

Reduplication is a matter of Iconic Patterning to the extent that the reduplicated sequence, say AA, signals plurality, repetition, or high degree of intensity of A.

#### (ii) Complexity

Formal complexity is a matter of Iconic Patterning to the extent that the formally more complex structure, say AB, is semantically more complex than either A or B.

#### (iii) Ordering

Constituent ordering is a matter of Iconic Patterning to the extent that the order of constituents, say AB, in some sense corresponds to the temporal or psychological order of the items A and B.

#### (iv) Degree of cohesion

By degree of cohesion of two elements we understand the relative tightness with which they are combined with each other. For example, the following series displays diminishing degrees of cohesion between A and B (cf.

Haiman 1983):

(22) a. A and B are two morphemes within one word: A + B

b. A and B are two words immediately juxtaposed: A # B

c. A and B are two words combined by means of some relator: A r B

Degree of cohesion is a matter of Iconic Patterning to the extent that the greater the degree of formal cohesion between A and B, the greater also their semantic cohesion.

#### (v) Explicitness of expression

In many grammatical situations, there is a possible choice between less explicit and more explicit expression forms. For example, many languages



have series of weak vs. strong personal or possessive pronouns. In addition, many languages allow a choice, in certain constructions, between zero vs. overt anaphorical elements. As Givón (1983) has argued, in such situations it is typically the case that, the less expected/predictable the entity to be referred to is, the more explicit the referring expression used. As an illustration, consider the following opposition:

- (23) a. John wants  $\emptyset$  to go home  
b. John wants him to go home

$\emptyset$  is chosen in the Equi construction (23a), where the Agent of go home can be reconstructed from the Subject of wants; him is chosen in (23b) where there is no coreference between the two participants, and the Agent of go home is therefore less or not reconstructible from the context. There are certainly languages in which even in the case of coreference the embedded Subject must be marked by an overt pronoun; but there are no languages on record in which constructions of type (23a) would have the non-coreferential interpretation, and constructions of type (23b) the coreferential one.

It has become clear, I hope, that there are a considerable number of design features of natural languages which can be interpreted as diminishing the basic arbitrariness of the linguistic sign, and thus injecting linguistic expressions with forms of consistency and patterning which serve to facilitate their comprehension and memorization.

#### 6.2.1.4. Markedness and Expressivity

I would like to add some remarks at this stage on the status of marked or expressive construction types at the morpho-semantic level. In one sense, such construction types can be subsumed under the heading of Iconic Patterning inasmuch as they are based on a principle which can be formulated as follows:

- (24) Marked features of content should be expressed in marked features of linguistic form.

By 'marked features of content' I mean those elements in the content of a linguistic expression which are communicatively most important, salient, or unexpected. These are in general the 'focal' parts of the information. 'Marked features of linguistic form' are such features as allow the Speaker to 'highlight' one or more constituents within the total structure of the

linguistic expression.

Suppose that the focal point of the content of some expression lies in a constituent X, then languages commonly apply the following strategies for highlighting X:

- (25) a. Assign extra stress and/or higher pitch to X
- b. Place X in a special, 'prominent' position
- c. Use a special construction by which X is made especially conspicuous.

In English, for example, these strategies can produce the following variations on a 'neutral' construction such as (26), on the condition that 'Mary' is the focal point of the information to be conveyed:

- (26)       John doesn't like Mary
- (27) a.     John doesn't like MARY
- b.     MARY John doesn't like
- c.     It is MARY that John doesn't like

Note that on the one hand the use of such marked construction types is a matter of Iconic Patterning, in that there is a correspondence between markedness of content and markedness of form. On the other hand, the marked constructions used typically diverge from the unmarked construction types which would be used in the 'neutral' condition, and which themselves might be defined by features of Arbitrary and Iconic Patterning.

Thus, the iconicity involved in markedness may lead to a divergence from the expected formal patterns, and thus contribute to complexity rather than to simplicity of the grammar. Markedness, interpreted in this sense, plays an important role in morpho-semantic change, because of the universal operation of the following principles:<sup>18</sup>

- (28) a.     There is a continuous need for marked forms of linguistic expression;
- b.     Marked forms of linguistic expression tend to gradually lose their marked character when they are regularly used;
- c.     Languages must therefore regularly renew their repertoire of marked expression forms.

### 6.2.2. Social factors

Most of the morpho-semantic patterning principles discussed so far define preferences for simpler, more transparent, and more consistent form-meaning correlations. There are also extragrammatical factors, however, which lead to complication rather than simplification of the form-meaning correspondence. This, as we saw above (section 5.3.2.) is generally the case with those expression types which serve to express politeness, deference, and social distance. Typically, these expressions are more complicated than their less polite counterparts. One reason for this is that polite expressions are typically less direct than their less polite counterparts, and that less direct linguistic expressions are typically more complicated than more direct ones.<sup>19</sup>

Compare:

- (29) a. Come here !  
b. Could you come here for a moment, please?
- (30) a. I want to see you for a moment.  
b. I wondered if I could possibly see you for a moment.

A second factor is that in certain languages and/or circumstances it is considered impolite to speak in sentence fragments, so that elliptic expression forms are avoided. Compare:

- (31) A: Where are you going?  
Ba: I'm going to London.  
Bb: To London.

Answers such as (31Bb) would be avoided in certain formal situations even in English. In other languages, however, the ban on speaking elliptically is even much stronger. Thus, Schaub (1985: 17 ff.) notes that in Babungo, a Grassfields Bantu language spoken in Cameroon, one is generally supposed to speak in full sentences, and elliptical expressions such as (31Bb) only occur exceptionally. Evidently, such politeness restrictions inhibit the operation of certain simplification processes which might otherwise be desirable for productive ease.

In a different way, the complicating effect of politeness and social constraints is vividly illustrated by Bird and Shopen (1979) in their comparison of what a person, fifteen minutes late for work, might say to a

friend he meets in an American or in a Maninka (West African) cultural setting. As a possible exchange in the American setting, they present the following 5 seconds conversation:

- (32) A: Hello Ed !  
B: Hi ! How are you ?  
A: Sorry, I'm in a hurry.  
B: Yeh, me too.  
A: See you on Saturday.

This dialogue could be translated into Maninka, but, the authors note, if it were used in comparable circumstances within Maninka culture, it would be viewed as 'unthinkably bad social behavior'. They then present the Maninka equivalent of (32), which consists of a 46 seconds conversation in which the two participants inquire elaborately about each other's health and well-being, as well as about that of their relatives. Bird and Shopen go on to argue that this is symbolic for the great importance attached to social ties and family relationships in Maninka society.

Politeness is not only coded in the indirectness and elaborateness of linguistic expressions used, it also comes out in the ways in which we address others or talk about third parties. Polite address may be monitored by augmentation of the other, by indirectness of approach, or by both:

- (33) Augmentation: use of plural forms in addressing single Addressees:  
vous vs. tu in French.  
Indirectness: use of third person forms of address:  
Has father slept well ?  
Both: use of exalting third person form:  
Your Highness, Your Majesty.

As Comrie (1975) has shown, these polite forms of address may lead to considerable complications in the grammar, because of the discrepancy between the grammatical form of address (2nd person plural/3rd person singular) and the pragmatic identity of the Addressee (2nd person singular). In languages with agreement between subject and predicate it is usually the case that the grammatical form of the subject determines agreement in certain cases, whereas the pragmatic identity of the Addressee is the determining factor in other cases. By comparing these patterns across languages, Comrie was able to set up a hierarchy of categories of the following form:

(34) grammatical agreement → VERB PARTICIPLE ADJECTIVE NOUN ← pragmatic agreement

In this hierarchy grammatical agreement will start at the left-hand side, pragmatic agreement at the right-hand side, and that the two will meet somewhere in the hierarchy, with possible overlap in a category. The workings of this mechanism may lead to such rather complicated patterns as the following:

- (35) a. You (2pl) have (2pl) been (sg) chosen (sg) president (sg)  
 b. Your Honour (3sg) are (2pl) very generous (sg)

For one example of how speaking about third parties may be monitored by features of social distance and respect, consider the use of nominal classifiers in Jacaltec (Mayan), as described in Craig (1979). Jacaltec has the following classifiers for Persons (including deities):

- (36) a. cumam he (deity)  
 b. cumi' she (deity)  
 c. ya' he/she (older generation)  
 d. unin he/she (infant)  
 e. ho' he (equal age, kin)  
 f. xo' she (equal age, kin)  
 g. naj he (equal age, non-kin)  
 h. ix she (equal age, non-kin)

Note that these classifiers can be used either attributively (as in naj Pel 'Peter') or pronominally (naj 'he') and that even in their objective usage they are already determined by a variety of social relations relevant to the 'distance' dimension. More interestingly, they can be dynamically or metaphorically used to convey special relational overtones. Thus, if ho'/xo' are used for non-kin, the effect is one of special affection, as if the person spoken of were drawn into the family; on the other hand, naj/ix can be used for kin to convey an aspect of estrangement. Further, when naj/ix are used for an older person (usually referred to by ya'), this counts as an insult.

We have seen that a number of examples in which the linguistic codification of social relationships leads to complications (sometimes considerable) in the linguistic system. This social factor thus runs counter to the drive for simplicity and perspicuity, which is defined by the psychological factors of productive, perceptual ease and ease of acquisition. This, I believe, provides an important

check on all-too-easy application of such principles as 'least effort' or 'simplicity' in the evaluation of grammatical systems. In all these cases, one might say, the amount of energy devoted to linguistic production rather directly symbolizes the degree of respect displayed in relation to the Addressee or the things and persons spoken about.

## 7. Conclusion

By way of a conclusion, I should like to ask the reader to re-read the theory of functional explanation set forth in section 5 of this paper. I believe I have provided ample evidence to demonstrate the feasibility and the interest of 'functional explanations' of linguistic phenomena in the sense of this theory.

Our main conclusion may be that functional explanations are never simple in the sense of directly accounting for a linguistic phenomenon X in terms of 'THE' function of X. On the contrary, functional explanations of any interest are complex in at least the following respects:

- (i) they typically involve the interaction of a number of different principles, each functionally motivated, but not necessarily converging in the same preferences with respect to linguistic organization;
- (ii) optimalization in one area of linguistic organization may entail less preferred consequences for other areas of the linguistic system;
- (iii) while many functional prerequisites determine a preference for greater simplicity, languages do not gain in over-all simplicity in the course of time, since (a) what is 'simpler' with respect to one functional requirement is not necessarily 'simpler' with respect to another (i.e. different functional requirements determine different kinds of simplicity); (b) greater simplicity in one area may entail complications in another area of linguistic organization (= (ii)); (c) certain functional requirements, especially those concerning social prestige and social distance, actually define a preference for greater complexity of expression.

Any natural language, in other words, can be seen as a temporary compromise created within the 'solution space' defined by the functional prerequisites imposed on human communication.

NOTES

1. For similar multifactor approaches to the explanation of language design and change, see Koefoed (1974), Langacker (1977), Koefoed (1978), Dik (1983), Hawkins (1983), the latter two with special reference to constituent ordering. Langacker (1977: 99, 128) appropriately uses the term 'vector sum': 'any syntactic change can be regarded as the vector sum of a variety of possibly conflicting pressures and tendencies'. The papers by Koefoed and Langacker have been especially helpful to me in writing this paper.
2. There is a considerable amount of psycholinguistic evidence pointing to the psychological relevance of the notions Topic, Focus and 'first position'. See, for example, Haviland and Clark (1974), Bates (1976), Clark and Clark (1977), and Clark and Haviland (1977).
3. For this principle, see Greenberg (1963), Vennemann (1973), Hawkins (1983).
4. See Kahrel (1985) and De Vries (1985) for evidence supporting this theory.
5. For the local nature of linguistic change and for the fact that the effects of one change may trigger another, see also Langacker (1977). Compare also Koefoed (1974: 280): 'There are changes that optimize a language in one respect but at the same time have the effect of rendering the language less simple in another respect'.
6. On the relative efficiency of different grammatical 'techniques', see also Dressler (1983).
7. The distinction between the representational and the relational aspects of communication is made by different authors in different terms. Bühler (1934) distinguished 'Darstellung' (representation) from 'Ausdruck' (expression) and 'Auslösung' (appeal), where the latter two are more 'relational' in character. Halliday (1970) opposes the 'ideational' and the 'inter-personal' (= relational) aspects of communication. Watzlawick et al. (1967) speak of content-oriented and relation-oriented features of communication.
8. For the divergent requirements imposed on natural languages by production and comprehension constraints, and for the role of these requirements in linguistic change, compare Bossuyt (1983). Our view differs from that of Bossuyt in that these performance constraints are seen as only two among a much wider set of 'functional prerequisites' imposed on natural languages.
9. Compare Lakoff and Johnson (1980) for the general nature of such metaphors.
10. See, for example, Martinet (1960).

11. For this section I have especially drawn on Hyman (1975) and Koefoed (1978).
12. Compare Bossuyt (1983) and the sources mentioned in footnote 1.
13. Martinet (1960: 182) uses the term 'syntagmatic' and 'paradigmatic' economy'.
14. The figures indicate the absolute frequency of the words in question in a Dutch corpus of 720.000 words used as a basis for the frequency data in Uit den Boogaart (1975).
15. The term isomorphism is used in Haiman (1980). Note, however, that I do not follow Haiman in subsuming isomorphism under 'iconicity'. The reason for this is that in my opinion an arbitrary linguistic form does not gain in 'iconicity' if it is consistently used in one and the same meaning. I reserve 'iconicity' for those cases in which the form in some sense symbolizes some aspect of its content. For the other terms used in the same sense as isomorphism, see Langacker (1977) - transparency; Koefoed (1978) - Humboldt's principle; and Dressler (1983) - bi-uniqueness. Note, further, that 'transparency' is also used, in the literature on generative phonology and in Lightfoot (1979), in a quite different meaning relating to the ease with which an underlying structure can be reconstructed on the basis of the surface form.
16. On reduction of allomorphy through paradigmatic levelling see also Skousen (1975), Koefoed (1978), Ronneberger (1985).
17. See the references in footnote 3. Hawkins appropriately coined the term 'Cross-categorial Harmony' for this phenomenon (Hawkins 1983).
18. Meillet formulated these principles as early as 1910. See Meillet (1982: 163ff). Compare also Lightfoot (1979: 384) on the explanatory value of 'expressivity' requirements, and of the demarking of originally marked forms.
19. For 'social deixis' and its influence on grammar, see Levinson (1983: 89ff). For examples such as (29)-(30), see ibid. 45.



REFERENCES

BATES, Elisabeth

- 1976 Language and context; the acquisition of pragmatics. New York: Academic Press.

BERLIN, Brent, and Paul KAY

- 1969 Basic color terms; their universality and evolution. Berkeley: University of California Press.

BEVER, Thomas G.

- 1970 The cognitive basis for linguistic structures. In: J.R. Hayes (ed.), Cognition and the development of language, 279-352. New York: Wiley.

BIRD, Charles, and Timothy SHOPEN

- 1979 Maninka. In: T. Shopen (ed.), 59-111.

BOSSUYT, Alain

- 1983 Historical Functional Grammar: an outline of an integrated theory of language change. In: Dik (ed.), 301-325.

BROWN, R., and A. GILMAN

- 1960 The pronouns of power and solidarity. In: T.A. Sebeok (ed.), Style in language, 253-276. Cambridge, MA : MIT Press.

BUHLER, Karl

- 1934 Sprachtheorie. Jena: Fischer.

CHOMSKY, Noam

- 1965 Aspects of the theory of syntax. Cambridge, MA : MIT Press.

CLARK, H.H., and E.V. CLARK

- 1977 Psychology and language; an introduction to psycholinguistics. New York: Harcourt, Brace, Jovanovich.

CLARK, H.H., and S.E. HAVILAND

- 1977 Comprehension and the given-new contrast. In: R.O. Freedle (ed.), Discourse production and comprehension, 1-40. Norwood, N.J.: Ablex

COMRIE, Bernard

- 1975 Polite plurals and predicate agreement. Language 51, 406-418.

CRAIG, Colette G.

- 1979 Jacalteco; field work in Guatemala. In: Shopen (ed.), 3-57.

DIK, Simon C.

- 1978 Functional Grammar. Amsterdam: North-Holland (3d impr. Dordrecht: Foris 1981).
- 1983 Two constraints on relators and what they can do for us. In: Dik (ed.), 267-298.

DIK, Simon C., ed.

1983 Advances in Functional Grammar. Dordrecht: Foris.

DRESSLER, Wolfgang U.

1983 On word formation in natural morphology. Proceedings of the XIIIth International Congress of Linguists, 172-182. Tokyo.

GIVON, Talmy, ed.

1983 Topic continuity in discourse: a quantitative cross-language study. Amsterdam: Benjamins.

GREENBERG, Joseph H.

1957 Essays in linguistics. Chicago: The University of Chicago Press.

1963 Some universals of language with particular reference to the order of meaningful elements. In: J.H. Greenberg (ed.), Universals of language, 73-113. Cambridge, MA : MIT Press.

1966 Synchronic and diachronic universals in phonology. Language 42, 508-517.

HAIMAN, John

1980 The iconicity of grammar. Language 56, 515-540.

1983 Iconic and economic motivation. Language 59, 781-819.

HAIMAN, John, ed.

1985 Iconicity in syntax. Amsterdam: Benjamins.

HALLIDAY, M.A.K.

1970 Language structure and language function. In: J. Lyons (ed.), New horizons in linguistics, 140-165. Harmondsworth: Penguin.

HAVILAND, S.E., and H.W. CLARK

1974 What's new? Acquiring new information as a process in comprehension. JVLVB 13, 512-521.

HAWKINS, John A.

1983 Word order universals. New York: Academic Press.

HYMAN, Larry M.

1975 Phonology; theory and analysis. New York: Holt, Rinehart and Winston.

JAKOBSON, Roman

1941 Kindersprache, Aphasie und allgemeine Lautgesetze. Uppsala (English translation: Child language, aphasia, and phonological universals. The Hague: Mouton 1968).

KAHREL, Peter

1985 Indirect questions and relators. In: A.M. Bolkestein et al. (eds.), Predicates and terms in Functional Grammar, 165-181. Dordrecht: Foris.

KOEFOED, Geert

- 1974 On formal and functional explanation: some notes on Kiparsky's 'Explanation in phonology'. In: J.M. Anderson and C. Jones (eds.), Historical linguistics, Vol. II, 276-293. Amsterdam: North-Holland.
- 1978 Taalverandering in het licht van taalverwerving en taalgebruik. In: G. Koefoed and J. van Marle (eds.), Aspecten van taalverandering, 11-70. Groningen: Wolters-Noordhoff.

LABOV, William

- 1966 The social stratification of English in New York City. Washington, D.C.

LAKOFF, G., and M. JOHNSON

- 1980 Metaphors we live by. Chicago: Chicago University Press.

LANGACKER, Ronald W.

- 1977 Syntactic reanalysis. In: C.N. Li (ed.), Mechanisms of syntactic change, 57-139. Austin: University of Texas Press.

LEVINSON, Stephen C.

- 1983 Pragmatics. Cambridge: Cambridge University Press.

LIGHTFOOT, David W.

- 1979 Principles of diachronic syntax. Cambridge: Cambridge University Press.

MARTINET, André

- 1960 Éléments de linguistique générale. Paris: Armand Colin.

MAYERTHALER, Willi

- 1981 Morphologische Natürlichkeit. Wiesbaden: Athenaion.

MCCAWLEY, James D.

- 1985 Review of Newmeyer 1983. Language 61, 668-679.

MEAD, G.H.

- 1934 Mind, self, and society; from the standpoint of a social behaviorist. Chicago: Chicago University Press.

MEILLET, Antoine

- 1982 Linguistique historique et linguistique générale. Paris: Champion.

MERTON, Robert K.

- 1967 On theoretical sociology; five essays, old and new. New York: Free Press.

NEWMAYER, Frederick J.

- 1983 Grammatical theory; its limits and possibilities. Chicago: The University of Chicago Press.

NUYTS, Jan

- 1984 Sprachfunktionen. Trier: LAUT A: 128.

- 1985 Funkties, funktionalisme en psychologische realiteit. Antwerp  
Papers in Linguistics 41. Antwerp: University of Antwerp.
- RONNEBERGER-SIBOLD, Elke
- 1985 A performance model for a natural theory of linguistic change  
Paper, University of Freiburg.
- SAUSSURE, Ferdinand de
- 1916 Cours de linguistique générale. Paris: Psyot.
- SCHAUB, Willi
- 1985 Babungo. London: Croom Helm.
- SHOPEN, Timothy, ed.
- 1979 Languages and their speakers. Cambridge, MA : Winthrop.
- SKOUSEN, Royal
- 1975 Substantive evidence in phonology. The Hague: Mouton.
- UIT DEN BOOGAART, P.C.
- 1975 Woordfrequenties in geschreven en gesproken Nederlands. Utrecht:  
Oosthoek, Scheltema en Holkema.
- VENNEMANN, Theo
- 1973 Explanation in syntax. In: J. Kimball (ed.), Syntax and Semantics  
2, 1-50. New York: Academic Press.
- VRIES, Lourens de
- 1985 The Wambon relator system. Paper, Yapper Seksi Linguistik,  
Mangelum, Irian Jaya.
- VYGOTSKY, L.S.
- 1962 Thought and language. Cambridge, MA : MIT Press.
- WATZLAWICK, Paul, Janet Helmick BEAVIN, and Don D. JACKSON
- 1967 Pragmatics of human communication. New York: Norton.